UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES CHIMIQUES

Laboratoire d'Innovation Thérapeutique (UMR7200)

THÈSE présentée par :

Luca CHIESA

soutenue le : **13 septembre 2024**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Chimie - informatique et théorique

Development of artificial intelligence methods to help the design of new ligands of G-protein coupled receptors

THÈSE dirigée par :

Mme KELLENBERGER Esther Professeur, Université de Strasbourg

RAPPORTEURS :

M. DE GRAAF Chris Senior director, Nxera Pharma

M. FIORUCCI Sebastien Maître de conférences, Université Côte d'Azur

---

AUTRES MEMBRES DU JURY :

Mme BELVISI Laura Professore associato, Università degli Studi di Milano

Leave this world a little better than you found it.

BP

Anche la quercia fu un tempo una ghianda.
Se mai disperi di non riuscire nei tuoi obbiettivi a causa dei tuoi modesti inizi, ricorda che anche la quercia, quell'albero grande e maestoso, fu un tempo una piccola ghianda giacente al suolo.

Contents

Acknowledgements **1**

General introduction **2**
References . . . 3

Résumé en Français **5**
Introduction . . . 5
Résultats et discussions . . . 8
Conclusions générale . . . 18
Références . . . 19

Publications and Communications **22**

1 Introduction **25**
1.1 G-protein coupled receptors . . . 26
1.1.1 Therapeutic relevance, structure, and biological role . . . 26
1.1.2 Allosteric regulation of intracellular signalling . . . 28
1.1.3 Affinity and efficacy . . . 29
References . . . 31
1.2 Efficacy-driven drug design . . . 33
1.2.1 Protein-ligand docking . . . 33
1.2.2 Docking post-processing . . . 34
1.2.3 Application of machine learning in GPCR drug discovery . . . 40
References . . . 43
1.3 Objectives and limits . . . 56

2 Efficacy driven drug discovery using docking and protein ligand interactions **58**
2.1 Analysis of efficacy specific differences in ligand and binding site structures . 59
2.1.1 Introduction . . . 59
2.1.2 Results . . . 59
2.1.3 Discussion . . . 70
2.1.4 Conclusions . . . 72
2.1.5 Materials and Methods . . . 72
References . . . 74
2.2 Are docking and protein ligand interactions enough for efficacy-driven drug discovery? . . . 77
2.2.1 Introduction . . . 77

2.2.2 Results . . . 78
2.2.3 Discussion . . . 82
2.2.4 Conclusions . . . 86
2.2.5 Materials and Methods . . . 86
References . . . 89
2.3 Critical evaluation . . . 92
References . . . 93

3 Development of a new tool for efficacy-driven drug discovery 96
3.1 One class classification for the detection of β2 adrenergic receptor agonists using single-ligand dynamic interaction data . . . 97
3.1.1 Introduction . . . 98
3.1.2 Results and Discussion . . . 99
3.1.3 Conclusions . . . 112
3.1.4 Materials and Methods . . . 114
References . . . 121
3.2 Efficacy driven virtual screening using one class classification models trained on dynamic binding modes . . . 128
3.2.1 Introduction . . . 128
3.2.2 Results . . . 129
3.2.3 Discussion . . . 134
3.2.4 Conclusions . . . 136
3.2.5 Materials and Methods . . . 136
References . . . 139
3.3 Critical evaluation . . . 142
References . . . 145

4 Design of new ligands targeting the C-X-C chemokine receptor type 4 148
4.1 Modelling of the CXCL12-CXCR4 complex . . . 149
4.1.1 Introduction . . . 149
4.1.2 Results . . . 150
4.1.3 Discussion . . . 159
4.1.4 Conclusions . . . 159
4.1.5 Materials and methods . . . 160
References . . . 161
4.2 Virtual screening . . . 165
4.2.1 Introduction . . . 165
4.2.2 Results . . . 166
4.2.3 Discussion . . . 170
4.2.4 Conclusions . . . 171
4.2.5 Materials and methods . . . 172
References . . . 174
4.3 Critical evaluation and perspectives . . . 180
References . . . 181

5 Virtual screening and pharmacological properties: the β2-adrenergic receptor case **183**
5.1 Predicting the duration of action of β2-adrenergic receptor agonists: ligand and structure-based approaches . . . 184
5.1.1 Introduction . . . 184
5.1.2 Results . . . 187
5.1.3 Discussion . . . 195
5.1.4 Conclusions . . . 196
5.1.5 Materials and methods . . . 197
References . . . 201
5.2 Predicting the duration of action of the agonist BI167107 . . . 209
5.2.1 Introduction . . . 209
5.2.2 Results and Discussion . . . 209
5.2.3 Conclusions . . . 210
5.2.4 Materials and Methods . . . 210
References . . . 211
5.3 Critical evaluation . . . 212

6 Multistate modelling of GPCRs using AlphaFold **214**
6.1 LIT-AlphaFold: a new tool for multistate modelling . . . 215
6.1.1 Scope and motivation . . . 215
6.1.2 State of the art . . . 215
6.1.3 The LIT-AlphaFold pipeline . . . 219
References . . . 223
6.2 Benchmarking AlphaFold-generated structures of chemokine – chemokine receptor complexes . . . 228
6.2.1 Introduction . . . 229
6.2.2 Methods . . . 232
6.2.3 Results . . . 234
6.2.4 Discussion . . . 238
6.2.5 Conclusions . . . 240
References . . . 241
6.3 Critical analysis and perspectives . . . 247
References . . . 247

7 General conclusions **249**

A Chapter 1.1.3 supporting information **251**
A.1 Section 2.1 supporting information . . . 251
A.2 Section 2.2 supporting information . . . 287

B Chapter 3 supporting information **333**
B.1 Section 3.2 supporting information . . . 333

C Chapter 4 supporting information **363**
C.1 Section 4.2 supporting information . . . 363

Acknowledgements

Now that my PhD journey has ended it is crucial to thank all of the people that directly or indirectly played a role in my life in the last three years.

First of all I would like to thank my supervisor, Prof. Esther Kellenberger. You welcomed me in the laboratory three years and a half ago as a intern, and kept me afterwards as a PhD student. Thank you for your guidance in these years, you were able to give a clear direction to my development as a scientist, while leaving me enough space to try my own ideas, make mistakes, and learn from them.

I warmly thank Dr Chris de Graaf and Dr Sebastien Fiorucci for their review work, you offered multiple interesting insights on my thesis, and highlighted various critical points. Your observations will improve the quality of future publications stemming from this thesis. I would like to thank Dr Laura Belvisi for being a part of my defence jury, and for her contribution, alongside the other organizers from Milan and Strasbourg, to the double degree program in Chemoinformatics between the two universities. If it weren't for the double degree I would not have had the occasion to continue my studies in Strasbourg, leading to this thesis.

The results presented in this thesis were often the fruit of a collaboration with other scientists, whose specific expertises have enriched the final research results. I would like to thank all the people whose work has been integrated inside this thesis: Dr Thiérry Durroux, Omolade Otun, Dr Emilie Sick, Dr François Daubeauf, and Lauri Urvas.

The present work was only possible thanks to the team in the Structural Chemogenomics group in the Laboratoire d'Innovation Thèrapeutique. I would like to thank the permanent members of the group Prof Esther Kellenberger, Dr Didier Rognan, Dr Celien Jacquemard, and Guillaume Bret for their help and guidance. A special mention goes to Guillaume, without whom we would have to do all of our calculations with pen and paper. My stay in the laboratory would not have been so profitable and enjoyable without the other members of the team with whom I shared most of this journey: Merveille, Mikhail, Julia, Lauri, Teodora, and François. Thank you for your help both from a scientific and a human point of view. I would also like to express my gratitude to all the other members of the group with whom I shared a shorter time: Viet, Augusta, Mariana, Anthony, Martin, Leonard, Vladimir, and in particular Dina, who I had the pleasure to co-supervise during her internship period.

Since one cannot live of science alone, a big thank you goes to all my the friends, both the new ones I have made in these three years in Strasbourg, and the friends of a lifetime back at home in Italy. Thanks for your support and the time we spent together.

Finally, a great thank you goes to my family, without whom all of this would have never been possible. I would like to thank my Mom and Dad, my brother Matteo, my grandmother Cesarina and my uncle Silvano, for everything they have done for me in the past twenty-seven years.

Grazie.

General introduction

The work described in this thesis is a continuation of the almost two decades of research on G-protein coupled receptors (GPCRs) conducted by my host laboratory, the Laboratoire d'Innovation Thérapeutique. Different aspects of computational drug discovery were previously covered: structure analysis and modelling[1], virtual screening[2], modelling of ligand binding or other underlying biological processes[3–5]. Chemokine receptors in particular have represented one of the most studied targets within the family[3–6].

The research topics covered in this thesis originated from the need to design a small molecule agonist targeting the C-X-C chemokine receptor type 4 (CXCR4), to use as a chemical probe to study the activation of the receptor in *in vitro* and *in vivo* experiments. Designing a small molecule agonist of CXCR4 was, and still is, a complex task due to multiple factors[7]. At the time no experimental structure of the CXCL12-CXCR4 complex was published (the first one was made public in march 2024[8]), and all available structures of CXCR4 were in a different conformational state compared to the one expected to bind CXCL12. At the beginning of the project a homology model of the CXCL12-CXCR4 complex [9] had been recently published, representing a validated hypothesis for CXCL12 recognition by CXCR4. The model however could not be used directly for virtual screening, since it might include incorrect detail in the binding site structure, or in the conformation of the bound chemokine. Molecular dynamics simulations were used to validate the proposed binding mode, and to obtain a better description of the complex. The validated *in silico* structure could be used as a starting point for structure-based virtual screening.

To guide the design of a GPCR agonist, the literature suggested the use of interaction similarity to selectively increase the hit rate for ligands with the desired pharmacological profile [10, 11]. Interaction similarity methods, however, were developed for crystallographic binding data, or high confidence modelled binding modes. Therefore, a new approach was developed, which leveraged the binding information extracted from molecular dynamics simulation to identify relevant docking poses. This new method used a one class classification algorithm, whose objective is to identify the common features from a dataset containing only positive examples, and use these information to determine if new instances are comparable to the training data or not. The method was developed and tested using the β2-adrenergic receptor as a proof-of-concept target. Once developed the method was applied to the design of new agonists of CXCR4, and to identify agonists of the β2-adrenergic receptor characterized by a specific duration of action.

Since structure-based drug design is limited by the availability of structures in the desired conformational state, a parallel line of research concerning the modelling of chemokine-chemokine receptors complexes and GPCRs using AlphaFold[12] and AlphaFold-Multimer[13] was established. The objective was to facilitate future medicinal chemistry projects targeting receptors without an available experimental structure.

This thesis is articulated in seven chapters: **Chapter 1** offers an overview on G-protein coupled receptors, and the current state of the art regarding GPCR drug discovery, with a particular attention to computational techniques aimed at developing ligands able to elicit a specific biological response. In **Chapter 2** the differences in molecular structure and properties between agonists and antagonists were compared, as well as the differences between the binding site conformations recognizing these ligands. These information were used to interpret the results from a retrospective virtual screening experiment comparing different methods for the selective discovery of agonists or antagonists. **Chapter 3** describes the development and evaluation of a new virtual screening tool called InteractionOCSVM, which uses machine learning and data from molecular dynamics simulations to identify new ligands with a similar binding mode as the reference. **Chapter 4 and 5** describe the application of InteractionOCSVM for the discovery of new ligands of CXCR4, and to identify agonists of the β2-adrenergic receptor with a desired duration of action. **Chapter 6** was dedicated to the development of LIT-AlphaFold, a tool to predict structures in multiple conformations for both monomeric and multimeric proteins, and its benchmarking on chemokine – chemokine receptor complexes. **Chapter 7** contains the general conclusions about the work described in this thesis.

References

[1] Jean-Sebastien Surgand, Jordi Rodrigo, Esther Kellenberger, and Didier Rognan. A chemogenomic analysis of the transmembrane binding cavity of human G-protein-coupled receptors. *Proteins: Structure, Function, and Bioinformatics*, 62(2):509–538, 2006.

[2] Esther Kellenberger, Jean-Yves Springael, Marc Parmentier, Muriel Hachet-Haas, Jean-Luc Galzi, and Didier Rognan. Identification of Nonpeptide CCR5 Receptor Agonists by Structure-based Virtual Screening. *Journal of Medicinal Chemistry*, 50(6):1294–1303, March 2007.

[3] Javier Garcia-Perez, Patricia Rueda, Jose Alcami, Didier Rognan, Fernando Arenzana-Seisdedos, Bernard Lagane, and Esther Kellenberger. Allosteric Model of Maraviroc Binding to CC Chemokine Receptor 5 (CCR5). *Journal of Biological Chemistry*, 286(38):33409–33421, September 2011.

[4] Jun Jin, Fanny Momboisse, Gaelle Boncompain, Florian Koensgen, Zhicheng Zhou, Nelia Cordeiro, Fernando Arenzana-Seisdedos, Franck Perez, Bernard Lagane, Esther Kellenberger, and Anne Brelot. CCR5 adopts three homodimeric conformations that control cell surface delivery. *Science Signaling*, 11(529):eaal2869, May 2018.

[5] Célien Jacquemard, Florian Koensgen, Philippe Colin, Bernard Lagane, and Esther Kellenberger. Modeling of CCR5 Recognition by HIV-1 gp120: How the Viral Protein Exploits the Conformational Plasticity of the Coreceptor. *Viruses*, 13(7):1395, July 2021.

[6] Muriel Hachet-Haas, Karl Balabanian, François Rohmer, Françoise Pons, Christel Franchet, Sandra Lecat, Ken Y. C. Chow, Rania Dagher, Patrick Gizzi, Bruno Didier, Bernard Lagane, Esther Kellenberger, Dominique Bonnet, Françoise Baleux, Jacques Haiech, Marc Parmentier, Nelly Frossard, Fernando Arenzana-Seisdedos, Marcel Hibert,

and Jean-luc Galzi. Small Neutralizing Molecules to Inhibit Actions of the Chemokine CXCL12 *. *Journal of Biological Chemistry*, 283(34):23189–23199, August 2008.

[7] Ilze Adlere, Birgit Caspar, Marta Arimont, Sebastian Dekkers, Kirsten Visser, Jeffrey Stuijt, Chris de Graaf, Michael Stocks, Barrie Kellam, Stephen Briddon, Maikel Wijtmans, Iwan de Esch, Stephen Hill, and Rob Leurs. Modulators of CXCR4 and CXCR7/ACKR3 Function. *Molecular Pharmacology*, 96(6):737–752, December 2019.

[8] Kei Saotome, Luke L. McGoldrick, Jo-Hao Ho, Trudy F. Ramlall, Sweta Shah, Michael J. Moore, Jee Hae Kim, Raymond Leidich, William C. Olson, and Matthew C. Franklin. Structural insights into CXCR4 modulation and oligomerization, February 2024.

[9] Tony Ngo, Bryan S. Stephens, Martin Gustavsson, Lauren G. Holden, Ruben Abagyan, Tracy M. Handel, and Irina Kufareva. Crosslinking-guided geometry of a complete CXC receptor-chemokine complex and the basis of chemokine subfamily selectivity. *PLOS Biology*, 18(4):e3000656, April 2020.

[10] Chris de Graaf and Didier Rognan. Selective Structure-Based Virtual Screening for Full and Partial Agonists of the $B2$ Adrenergic Receptor. *Journal of Medicinal Chemistry*, 51(16):4978–4985, August 2008.

[11] Albert J. Kooistra, Henry F. Vischer, Daniel McNaught-Flores, Rob Leurs, Iwan J. P. de Esch, and Chris de Graaf. Function-specific virtual screening for GPCR ligands using a combined scoring method. *Scientific Reports*, 6(1):28288, September 2016.

[12] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, August 2021.

[13] Richard Evans, Michael O'Neill, Alexander Pritzel, Natasha Antropova, Andrew Senior, Tim Green, Augustin Žídek, Russ Bates, Sam Blackwell, Jason Yim, Olaf Ronneberger, Sebastian Bodenstein, Michal Zielinski, Alex Bridgland, Anna Potapenko, Andrew Cowie, Kathryn Tunyasuvunakool, Rishub Jain, Ellen Clancy, Pushmeet Kohli, John Jumper, and Demis Hassabis. Protein complex prediction with AlphaFold-Multimer, March 2022.

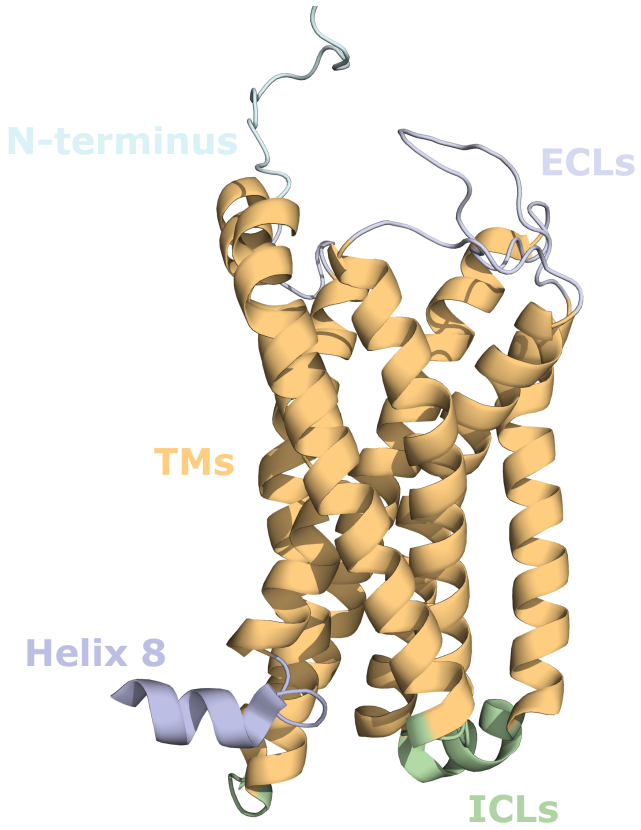
Résumé en Français

Introduction

Les récepteurs couplés aux protéines G (RCPG) constituent une large famille de protéines membranaires impliquées dans la transduction du signal en réponse à des stimuli externes. Le fonctionnement d'une majeure partie de ces récepteurs est régulé par la liaison d'un ligand, généralement dans une poche de liaison située dans leur domaine transmembranaire du côté extracellulaire. Les RCPG sont divisés en cinq classes basées sur la similarité de séquence, la classe A étant la plus grande[1]. Tous les membres de cette classe partagent une structure commune formée par sept domaines transmembranaires hélicoïdaux hydrophobes (TMs) connectés par deux ensembles de trois boucles flexibles situées du côté intracellulaire (ICLs) et extracellulaire (ECLs) de la membrane[2] (Figure 0.0.1). Les récepteurs présentent deux poches de liaison, un site de liaison extracellulaire qui lie les ligands diffusibles, et un site de liaison intracellulaire qui est responsable de la liaison des protéines G. La poche de liaison intracellulaire n'est exposée que lorsque le récepteur adopte une conformation particulière, appelée la conformation de l'état actif, tandis que généralement le récepteur libre se trouve dans une conformation de l'état inactif, qui n'est pas apte à lier une protéine G. La liaison d'une protéine G au récepteur déclenche une cascade de signalisation, ce qui provoque une réponse biologique de la cellule. Les ligands se lient de préférence à l'un de ces deux états et le stabilisent. Les ligands agonistes, qui entraînent une réponse fonctionnelle, se lient préférentiellement à une conformation de l'état actif, alors que les ligands antagonistes, qui n'ont pas d'effet direct sur la fonction du récepteur libre mais empêchent la liaison d'un ligand agoniste, se lient à la conformation de l'état inactif.

Le développement d'un nouveau médicament ciblant un RCPG nécessite de prendre en compte l'affinité du ligand pour le récepteur, qui correspond à sa capacité à se lier à la cible, et aussi son efficacité, qui traduit sa capacité à déclencher une réponse biologique spécifique[3]. La conception rationnelle de médicament a fait l'objet de nombreux travaux méthodologiques visant à aborder la question de l'affinité, tandis que celle de la fonction est moins prégnante. Le docking protéine-ligand compte parmi les techniques les plus utilisées pour la conception de médicaments par criblage de chimiothèque[4]. La méthode repose sur la génération de multiples poses d'un ligand dans la poche de liaison, puis leur classement à l'aide d'une fonction de score qui approxime l'affinité du ligand pour la protéine cible[5]. Le docking dans les structures à l'état actif d'un RCPG devrait privilégier les ligands agonistes, tandis que le docking dans une structure à l'état inactif devrait privilégier les antagonistes. Cette tendance a été observée dans de multiples expériences de criblage virtuel prospectif, qui ont identifié sélectivement des ligands avec une certaine efficacité en se basant sur les structures du récepteur dans une certaine conformation[6][7]. Cependant, il y a également eu plusieurs cas où des agonistes ont été découverts à partir de structures de l'état inactif,

Figure 0.0.1: Les domaines structuraux clés d'un RCPG: domaines transmembranaires hélicoïdaux (TMs), boucles flexibles situées du côté intracellulaire (ICLs), et boucles flexibles situées du côté extracellulaire (ECLs).



et des antagonistes à partir de structures de l'état actif [8]. Une étape de post-traitement peut améliorer les résultats de docking. Une approche utilisée pour identifier sélectivement des ligands agonistes ou antagonistes est basée sur la différence d'affinité d'une molécule pour l'état actif et l'état inactif du récepteur. En comparant les scores de docking de la même molécule dans différentes conformations du récepteur, il est possible d'identifier l'état d'activation préféré cette molécule, et par conséquent sa fonction. Cette méthode a été utilisée pour identifier sélectivement des agonistes du récepteur β2-adrénergique (ADRB2) [7]. L'autre méthode considère les interactions formées entre le ligand et la protéine (liaison hydrogène, liaison ionique, interaction $\pi - \pi$ et contact hydrophobe) et compare le mode de liaison prédit à celui observé dans une structure cristallographique de référence [9]. Cette approche s'est avérée utile pour la priorisation de ligands fonctionnels du ADRB2 : les interactions formées par une référence agoniste ont tendance à sélectionner un plus grand nombre d'agonistes que d'antagonistes, et vice versa.

Dans l'objectif d'apprécier la généralisation possible de ces observations, nous avons évalué ces deux approches sur sept autres RCPG : 5HT2A, AA2AR, ACM2, ADRB2, CNR1, DRD2, OPRK.

Pour chacun d'entre eux, nous avons préparé un ensemble de données de référence contenant des agonistes et des antagonistes bien caractérisés, ainsi que des structures du récepteur à l'état actif et inactif. L'analyse de ces données a révélé des différences significatives entre les structures de ligands agonistes et antagonistes, ainsi que des réarrangements dans la poche de liaison du récepteur sous l'effet de leur liaison.

Nous avons ensuite évalué les deux approches, en utilisant différentes méthodes pour comparer les interactions ainsi que plusieurs scores de docking. Pour chacun des récepteurs, nous avons considéré plusieurs structures cristallographiques, à l'état actif comme à l'état inactif, et nous les avons toutes utilisées pour le criblage virtuel rétrospectif d'une chimiothèque test ad hoc, comprenant des ligands agonistes, des ligands antagonistes et des leurres.

Les structures cristallographiques, en figeant un état conformationnel, peuvent donner une représentation partielle des interactions clés pour la fonction du ligand. La dynamique moléculaire, en simulant la flexibilité du complexe protéine-ligand, permet de compléter l'information de liaison, en révélant en particulier la persistance des interactions intermoléculaires. Il est cependant difficilement concevable d'utiliser tous les instantanés d'une trajectoire comme autant de références individuelles pour le post-traitement du docking dans un criblage virtuel. Pour les considérer dans leur ensemble, nous avons développé une nouvelle méthode d'apprentissage automatique appelée InteractionOCSVM[10]. Les modèles créés sont capables d'identifier, à partir des données de simulation, les interactions pertinentes. Ils utilisent cette information pour évaluer les poses de docking. Cette méthode a été initialement testée sur ADRB2, et s'est révélée efficace pour distinguer les agonistes des antagonistes. Le test de la méthode InteractionOCSVM a été étendu à différents RCPG, avec différentes représentations des interactions, dans le but de déterminer ses forces et ses limites dans des conditions de criblage virtuel.

La méthode InteractionOCSVM a également été appliquée de manière prospective dans le cadre de deux projets collaboratifs interdisciplinaires : pour l'identification de nouveaux ligands agonistes du récepteur aux chimiokines CXCR4, capables de reproduire le mode de liaison du ligand endogène CXCL12 ; et pour la prédiction de la durée d'action du ligand agoniste BI167107 du récepteur ADRB2[11].

Toutes les techniques évaluées nécessitent une structure du récepteur dans l'état

d'activation ciblé. Malheureusement, les deux états d'activation ne sont pas disponibles pour tous les RCPG. AlphaFold[12] est un outil d'apprentissage profond pour la prédiction des structures de protéines, qui peut être utilisé pour modéliser plusieurs conformations protéiques en utilisant des techniques spécifiques. Nous avons évalué la justesse et la précision des prédictions d'AlphaFold pour les RCPG, en utilisant comme référence plusieurs structures cristallographiques de complexes chimiokines-récepteurs de chimiokines[13].

Résultats et discussions

Analyse des liens entre la structure et la fonction

Une liste de 31 récepteurs avec des structures expérimentales à la fois dans l'état actif et inactif a été extraite de GPCRdb[14]. Cette analyse a identifié 31 RCPG de classe A, qui ont ensuite été filtrés en fonction de la disponibilité d'un grand nombre d'agonistes et d'antagonistes bien caractérisés[15]. Au total, 22 RCPG de classe A étaient bien caractérisés tant du point de vue structural que pharmacologique. Parmi eux, sept ont été sélectionnés afin de représenter diverses sous-familles de RCPG : récepteurs de la sérotonine, récepteurs de l'adénosine, récepteurs muscariniques, récepteurs adrénergiques, récepteurs cannabinoïdes, récepteurs de la dopamine et récepteurs aux opiacés. Pour chacun de ces sept récepteurs, nous avons analysé la structure chimique, les caractéristiques pharmacophoriques et les propriétés physicochimiques de ses ligands. La comparaison entre agonistes et antagonistes a montré des différences nettes pour cinq récepteurs (AA2AR, ACM2, ADRB2, CNR1, DRD2), mais dont l'ampleur dépend du récepteur considéré. Des similitudes entre les structures des ligands ayant la même efficacité n'ont été observées que pour quatre de ces récepteurs (ACM2, ADRB2, CNR1, DRD2). Ces différences de structure et de propriétés sont par conséquent susceptibles d'être exploitées pour entraîner des classificateurs d'apprentissage automatique à identifier l'efficacité d'un ligand uniquement sur la base de sa structure chimique ou de ses propriétés physicochimiques.

Ainsi, nous avons créé des modèles prédictifs à l'aide de trois algorithmes de classification différents (random forest, support vector machine et k nearest neighbours), en codant le structure sous la forme d'empreintes digitales de Morgan ou d'empreintes pharmacophoriques. Globalement, les modèles ont affiché en bonnes performances en validation croisée, avec des précisions équilibrées comprises entre 0,66 et 0,95. Cependant certains ligands étaient systématiquement mal classés, ce qui suggère que l'ensemble de données ne couvre pas uniformément tous les châssis moléculaires actifs possibles, ou bien que les modèles ne sont pas capables de généraliser à des châssis moléculaires plus diversifiés.

Nous avons également créée un classificateur random forest afin de prédire l'efficacité d'un ligand en se basant sur vingt-cinq propriétés physicochimiques. Le modèle entraîné sur les données 5HT2A montre une faible précision, de l'ordre de 0,65. Pour DRD2 et OPRK les modèles obtenus sont modérément précis, avec des précisions équilibrées comprise entre 0,74 et 0,73 respectivement. Enfin, les modèles pour les autres récepteurs sont plus précis, avec des valeurs au moins égale à 0,86.

Les informations apprises par les modèles entraînés sur les empreintes digitales moléculaires et sur les descripteurs physicochimiques ont été analysées en calculant les valeurs de Shapley de différentes caractéristiques. Cette analyse a identifié des liens entre la présence de certaines caractéristiques chimiques et l'efficacité d'un ligand. Par exemple,

une efficacité agoniste envers ADRB2 est associée à la présence de groupes hydroxyle, soit comme substituant en position ortho ou alpha dans un anneau aromatique, soit comme groupes donneurs de liaison hydrogène supplémentaires. Des biais structuraux et dans les propriétés des ligands ont été identifiés pour tous les récepteurs sauf 5HT2A, pour lequel les agonistes et antagonistes ont des structures et des propriétés physicochimiques similaires.

L'analyse des structures du récepteur lié à des agonistes ou antagonistes a mis en évidence des différences de conformation au niveau du site de liaison de ces ligands. Du point de vue quantitatif, la déviation quadratique moyenne (RMSD) entre les structures des états actif et inactif est faible à modérée, avec des valeurs comprises entre 1,20 Å (pour ACM2) et 2,98 Å (pour CNR1). Elle est néanmoins suffisante pour distinguer les états actifs des états inactifs par partitionnement, et ce pour toutes les cible sauf ADRB2. Pour certains des récepteurs (5HT2A et AA2AR), on observe en plus différentes conformations du site de liaison au sein du même état d'activation (Figure 0.0.2).

La présence de réarrangements structuraux dans les structures du site de liaison est à considérer avec attention pour des applications faisant appel au docking, puisque cette méthode considère généralement le récepteur comme un objet rigide.

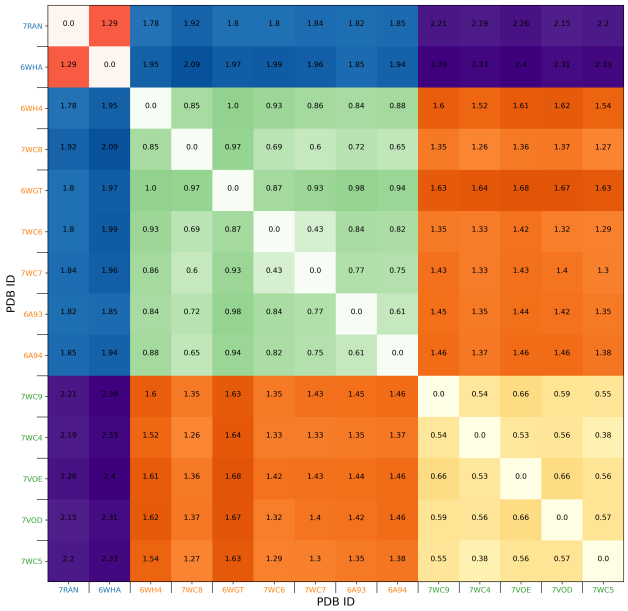
Evaluation de poses de docking par comparison d'interactions

Pour chacun des cibles RCPG représentatives identifiées, une chimiothèque test ad hoc, comprenant des ligands agonistes, des ligands antagonistes et des leurres, a été préparée. Les agonistes, antagonistes et les leurres ont été dockés dans plusieurs conformations des récepteurs, dans les états actifs et inactifs.

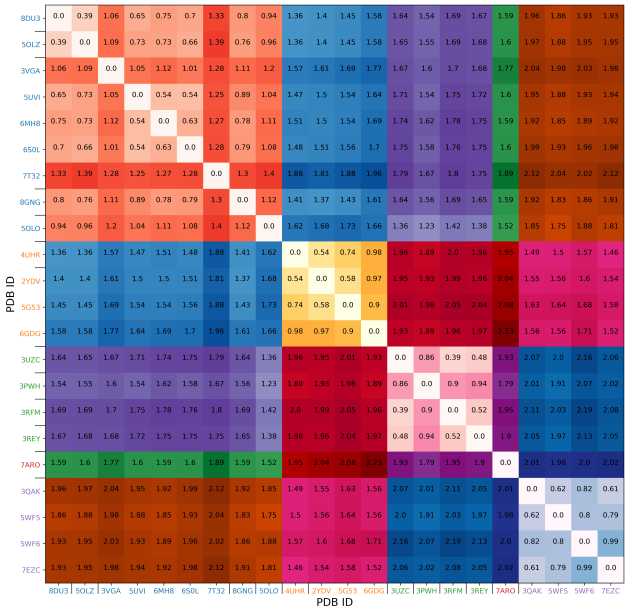
Le classement basé sur les scores de docking a mis en évidence l'existence d'un biais des ligands dans la fonction de score CHEMPLP utilisée[16], puisque les ligands ayant la même efficacité obtenaient de meilleurs scores que d'autres, indépendamment de la conformation du récepteur utilisée pour les calculs. Des biais similaires ont également été observés pour trois autres fonctions de score (GOLDSCORE[17], CHEMSCORE[17], GOLD PLP). Ce résultat est en contradiction avec les données de la littérature, qui montrent une variation de l'efficacité des composés les mieux classés selon la structure utilisée pour le criblage virtuel[7, 8, 18]. Par conséquent, la qualité de la fonction de score utilisée doit être soigneusement évaluée dans un test rétrospectif avant son utilisation dans le criblage virtuel.

Les scores de docking utilisées ne ségréguent pas les ligands en fonction de leur efficacité si on considère individuellement des résultats obtenus pour les différentes conformations du même récepteur. Par contre, la combinaison de résultats obtenus pour plusieurs conformations améliore la prédiction. Ainsi, la différence de score entre le docking dans une conformation à l'état actif d'un récepteur et le docking dans une conformation à l'état inactif de ce même récepteur a globalement bien séparé les agonistes et les antagonistes de l'ensemble des ligands testés. Les performances dans cet exercice ce classement sont appréciées par des valeurs ROCAUC allant de 0,77 pour ADRB2 à 0,98 pour ACM2 (valeurs moyennes entre toutes les paires de structures d'un même récepteur). Cette approche basée sur la différence de score n'est cependant pas efficace pour séparer entre les ligands des leurres. Ainsi si on compare cette approche par rapport au classement simple basé sur le score de docking, le rappel des agonistes n'est augmenté que pour 14 % des paires, et celui des antagonistes pour 11 % des paires. S'il permet de trier agonistes et antagonistes, le classement basé sur la différence de scores de docking, mais il priorise un plus petit nombre de molécules actives par rapport à l'utilisation directe de la fonction de score.

Figure 0.0.2: Représentation du regroupement des sites de liaison de 5HT2A (A) et AA2AR (B) basé sur le RMSD. Différents groupes sont mis en évidence par la couleur des étiquettes et par la couleur utilisée dans le graphique. (A) Le premier groupe représente les structures de l'état actif, le deuxième représente les structures de l'état inactif liées aux agonistes et aux antagonistes, et le troisième représente les structures de l'état inactif liées aux antagonistes. (B) Les premier et troisième groupes représentent deux conformations différentes de l'état inactif du site de liaison. Les deuxième et quatrième groupes représentent quant à eux deux conformations différentes de l'état actif du site de liaison.

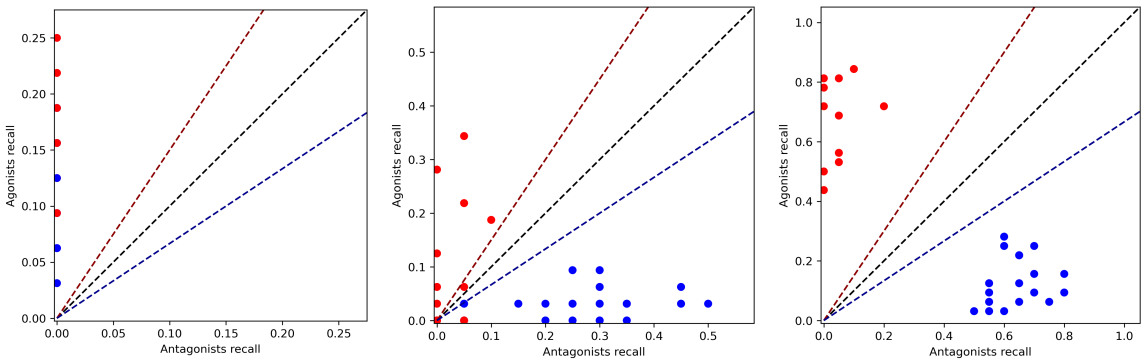


(A) 5HT2A



(B) AA2AR

Figure 0.0.3: Fraction d'agonistes et de antagonistes de ADRB2 avec un score meilleur que le 1% des leurres les mieux classés. (A) Résultats selon le docking score. (B) Résultats selon la similarite de IFP. (C) Résultats selon le GRIM score.



(A) Docking **(B)** IFP **(C)** GRIM

Les poses de docking ont ensuite été classées en fonction du degré de ressemblance entre les interactions prédites et les interactions observée dans un mode de liaison de référence, issue d'une structure cristallographiques. Dans l'ensemble, toutes les méthodes de comparaison (IFP[19], GRIM[20], graph kernel[10]) sont supérieures aux fonctions de score de docking pour identifier la fonction du ligand docké. Cependant, les résultats dépendent fortement du RCPG ciblé. Plus précisément, en utilisant des ligands et des leurres, le classement par similarité des interactions a amélioré l'enrichissement en molécules actives ayant la même efficacité que la référence par rapport à la fonction de score dans plusieurs cas : avec la méthode IFP, 28 % pour 5HT2A et CNR1, et 75 % pour ADRB2 ; avec la méthode Hyd IPAs et GRIM 57 % pour 5HT2A et 100 % pour ADRB2 ; avec la méthode graph kernel, de 28 % pour CNR1 à 72 % pour ADRB2 ; avec la méthode Newhyd IPAs / GRIM de 25 % pour ACM2 à 86 % pour ADRB2 ; avec la méthode graph kernel, de 0 % pour DRD2 à 75 % pour ADRB2. En général, les modes de liaison de référence plus complexes, comprenant de multiples interactions polaires, sont plus à même de prioriser les agonistes ou les antagonistes que ceux comprenant principalement des contacts hydrophobes. Par ailleurs, ce résultat est confirmé quelle que soit la représentation des interactions, c'est-à-dire par des empreintes ou bien des graphes tridimensionnels.

En général, les modes de liaison de référence plus complexes, comprenant de multiples interactions polaires, sont plus à même de prioriser les agonistes ou les antagonistes que ceux comprenant principalement des contacts hydrophobes. Par ailleurs, ce résultat est confirmé quelle que soit la représentation des interactions, c'est-à-dire par des empreintes ou bien des graphes tridimensionnels.

Mise au point et test de la méthode InteractionOCSVM

InteractionOCSVM est une nouvelle méthode développée pour comparer le mode de liaison prédit par docking à un ensemble d'interactions extraites de la simulation de dynamique moléculaire d'un complexe ligand-récepteur. Les informations du mode de liaison, encodées soit comme un graphe d'interaction, soit comme une empreinte digitale d'interactions, ont

été traitées par un algorithme de classification à une classe, spécifiquement un algorithme de machine à vecteurs de support à une classe (OCSVM), pour identifier les motifs d'interaction conservés dans l'ensemble de données. Les modèles entraînés sur ces données d'interactions ont été utilisés pour classifier les modes de liaison des structures générées par docking comme comparables aux modes de liaison de référence pertinents ou comme des modes de liaison alternatifs non pertinents.

InteractionOCSVM a été mis au point, dans une première version limitée aux graphes d'interaction, sur le récepteur ADRB2 à l'aide de trois groupes de modèles entraînés sur la base des interactions formées par trois agonistes de référence : l'épinéphrine, le HBI et le BI167107. Les modèles ont été testés pour leur capacité à récupérer sélectivement les ligands agonistes tout en écartant les antagonistes[10].

Les modèles InteractionOCSVM basés sur le HBI et le BI167107 ont clairement sélectionné les agonistes tout en écartant les antagonistes (Figure 0.0.4). En revanche, les modèles entraînés sur l'épinéphrine n'étaient pas prédictifs. L'analyse des simulations de dynamique moléculaire a révélé que le HBI et le BI167107 forment des complexes stables avec le récepteur ADRB2, contrairement à l'épinéphrine. Sur la base de ces résultats, nous avons conclu que la stabilité du complexe est un prérequis pour obtenir un modèle prédictif (Figure 0.0.4).

Pour corroborer et étendre les résultats initiaux obtenus sur ADRB2, nous avons conduit des expériences similaires sur six autres RCPGs, en utilisant à la fois des références agonistes et antagonistes. Les modèles InteractionOCSVM ont également été évalués dans des conditions de criblage virtuel en utilisant des leurres. Contrairement aux modèles initiaux pour ADRB2, plusieurs des modèles obtenus pour d'autres RCPG n'étaient pas en mesure de rejeter les ligands indésirables.

Une amélioration a été obtenue en combinant les prédictions de multiples modèles utilisant différentes représentations d'interactions ou basées sur des définitions différentes des interactions. Ces modèles d'ensemble ont souvent surpassé en performance les modèles individuels, conduisant à un enrichissement accru en molécules actives par rapport aux leurres. Malgré ces résultats encourageant, les modèles sélectionnent un nombre important de leurres.

En classant les poses sélectionnées par les modèles InteractionOCSVM en fonction de leur score de docking, nous observons un plus grand nombre d'agonistes et/ou d'antagonistes aux premières positions (Figure 0.0.5). Les poses des ligands agonistes (ou antagonistes) sélectionnés par les modèles ont des meilleurs scores de docking que les poses des leurres sélectionnés. Ainsi, nous pouvons conclure à l'utilité du modèle InteractionOCSVM pour filtrer les sorties de criblage virtuel de chimiothèque par docking sur un RCPG.

Le filtrage des poses de docking utilisant InteractionOCSVM, et le classement subséquent des poses sélectionnées en fonction de leur score de docking, ont montré des performances inférieures comparées au docking et classement utilisant une structure cristallographique du récepteur et un mode de liaison cristallographique. Le mode de liaison cristallographique d'un ligand semble donc représenter la référence optimale pour identifier des touches avec la même efficacité lors du criblage virtuel. Par conséquent, introduire du bruit et du flou dans le mode de liaison tend à abaisser la performance. L'analyse approfondie et détaillée de l'ensemble des résultats souligne l'importance de la qualité de la structure du récepteur utilisée en docking, et la robustesse du mode de liaison dynamique, suggérant l'utilité de l'approche InteractionOCSVM pour exploiter les informations structurales de modèles tridimentionnels (par exemple un docking sous contrainte d'un ligand de référence, ou un modèle par homologie), qui sont généralement moins fiables que les modèles expérimentaux.

Figure 0.0.4: Performances des modèles InteractionOCSVM pour distinguer les agonistes des antagonistes de ADRB2. Rappel : nombre d'agonistes sélectionnés par un modèle rapporté au nombre total d'agonistes. Précision : nombre d'agonistes sélectionnés par un modèle rapporté au nombre total de ligands sélectionnés Hyd, Newhyd, Polar : Différentes définitions des interactions protéine-ligand. Hyd, définition plus générique des contacts hydrophobes. Newhyd, définition plus spécifique des contacts hydrophobes. Polar, tout interactions sauf les contacts hydrophobes. Adapté de l'article de Chiesa et Kellenberger [10].

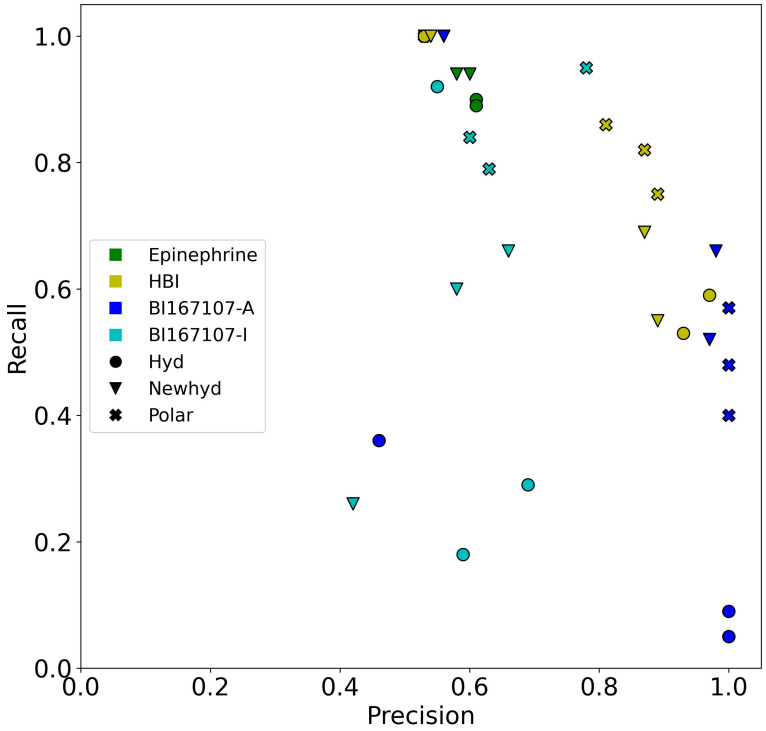
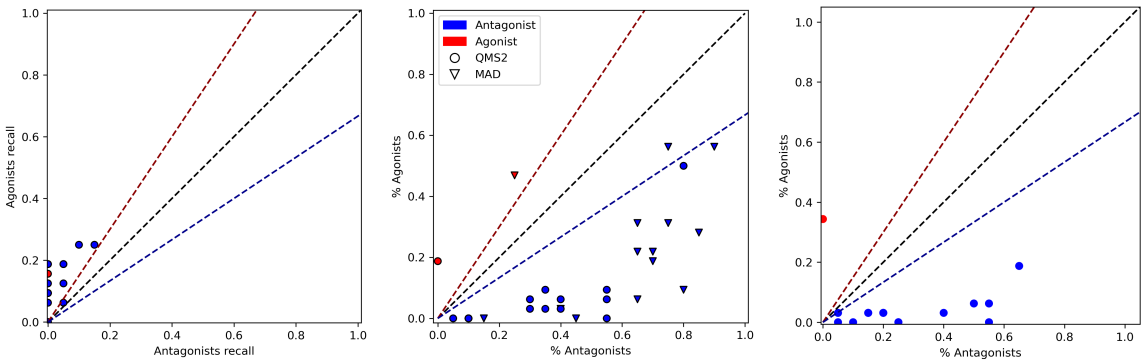


Figure 0.0.5: Fraction d'agonistes et de antagonistes de ADRB2 avec un score de docking meilleur que le 1% des leurres les mieux classés. (A) Résultats du docking en ensemble. (B) Résultats du docking en ensemble et filtrage par InteractionOCSVM. (C) Résultats du docking en ensemble et filtrage en ensemble par InteractionOCSVM.



(A) **(B)** **(C)**

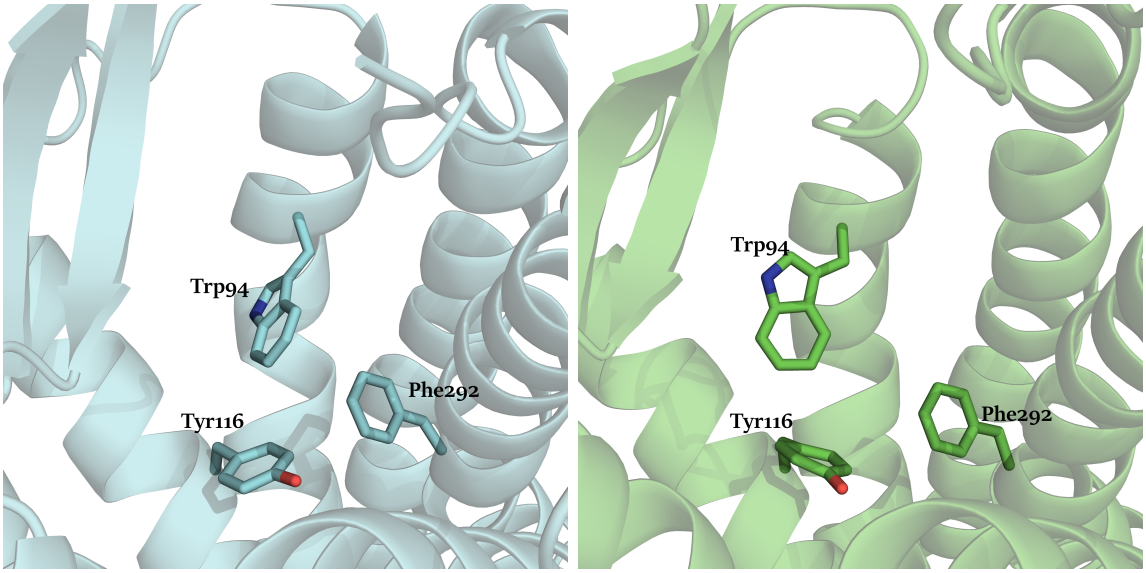
Decouvert de nouveau ligands du receptor chimiokine C-X-C type 4

Dans le but de concevoir des ligands agonistes non peptidiques du CXCR4, nous avons modélisé l'interaction de ce récepteur avec son agoniste endogène, c'est-à-dire la chimiokine CXCL12. La structure tridimensionnelle du complexe été construite à partir d'informations expérimentales [21], placée dans une bicouche lipidique, puis la stabilité du système a été évaluée par simulation de dynamique moléculaire en solvant explicit. A titre de comparaison, nous avons également évalué dans les mêmes conditions le récepteur lié à l'antagoniste IT1t (Figure 0.0.6). Pour chacun des deux systèmes, nous avons réalisé cinq trajectoires de 200 ns chacune. Leur analyse montre des complexes globalement stables. La comparaison des structures révèle également des différences dans la conformation du site de liaison du récepteur, selon qu'il soit lié à CXCL12 ou à IT1t.

Nous avons considéré plus précisément le mode de liaison entre CXCR4 et CXCL12, afin d'identifier les trajectoires exploitables pour construire les modèles InteractionOCSVM. Le mode de liaison entre le récepteur et son ligand est stationnaire dans trois des cinq réplicas, et ressemble au mode de liaison dans la structure initiale, c'est-à-dire celui observé dans le modèle construit par homologie, et est en accord avec les données de mutagenèse dirigée. En nous basant sur les données de la littérature décrivant l'effet de mutation ou de coupure dans la séquence de CXCL12 sur la fonction de cette chimiokine, nous avons foncalisé notre attention sur les trois premiers résidus qui sont critiques pour la liaison du récepteur et aussi pour son activation.

Les interactions entre le récepteur et les trois premiers résidus de la chimiokine ont été utilisées pour entraîner un modèle InteractionOCSVM. Le criblage virtuel d'une chimiothèque d'environ un million de composés par docking, suivi du filtrage par la méthode InteractionOCSVM, a identifié 1731 composés formant les interactions désirées. Parmi ces touches virtuelles, nous avons sélectionné 78 ligands en fonction de leur score de docking et de leur structure chimique. Les molécules sélectionnées ont été testées expérimentalement,

Figure 0.0.6: Structure des résidus Trp94, Tyr116 et Phe292 dans une trame extraite de la simulation IT1t-CXCR4 (A), et dans une trame extraite de la simulation CXCL12-CXCR4 (B).



(A) IT1t-CXCR4 **(B)** CXCL12-CXCR4

révélant quatre ligands avec une affinité micromolaire et une structure nouvelle par rapport aux composés connus (Tableau 0.0.1). Malheureusement, les tests fonctionnels ont montré que tous ces ligands étaient des antagonistes.

Ce résultat, bien que décevant, a pu être expliqué par une inexactitude du modèle tridimensionnel utilisé. En effet, une structure expérimentale du complexe entre CXCR4 et CXCL12 a été publiée après la réalisation de notre criblage virtuel[22]. Sa comparaison avec la structure par homologie utilisée pour ce criblage a mis en évidence des erreurs dans la structure de la poche de liaison et dans le mode de liaison de CXCL12. En revanche, un modèle alternatif du complexe généré à l'aide d'AlphaFold-Multimer[23] prédit correctement la structure du complexe, y compris l'interface entre les deux protéines (Figure 0.0.7). Une prochaine campagne de criblage virtuel pourra tirer parti de la structure expérimentale et du modèle in silico de haute qualité pour identifier un agoniste non peptidique de CXCR4.

Criblage virtuel et propriete pharmacologique: le cas du recepteur β2-adrenergique

Les modèles InteractionOCSVM entraînés pour ADRB2 ont été réévalués en tenant compte de la durée d'action des bronchodilatateurs ciblant ce récepteur[11]. Nous avons comparé les performances des modèles basés sur le mode de liaison du formotérol, un agoniste à longue durée d'action (LABA), avec celles des modèles basés sur le mode de liaison du salbutamol, un agoniste à courte durée d'action (SABA). L'étalonnage a été effectué à partir d'un ensemble de 22 agonistes de ADRB2 dont la durée d'action est connue. Nous avons observé que certains modèles présentaient un biais en faveur des ligands ayant la même durée d'action que la référence. En particulier, tous les modèles basés sur le salbutamol ont montré une tendance à sélectionner des SABA. Ce biais était moins marqué pour les modèles basés sur le formotérol,

Table 0.0.1: Affinité de liaison (IC50) et similarité moléculaire (similarité de Tanimoto) par rapport aux composés connus des quatre hits identifiés par criblage virtuel.

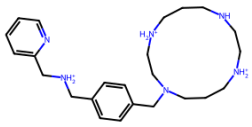
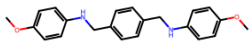
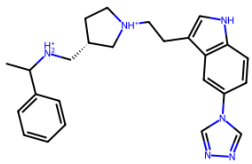
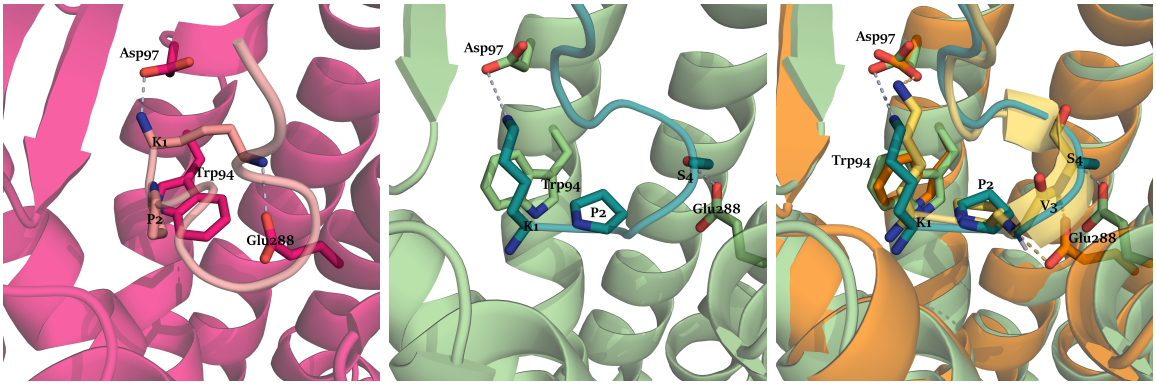
Molécule	IC50 (μM)	Ts	Molécule active la plus proche
Compound 1	2.066	0.23	
Compound 2	2.140	0.25	
Compound 3	2.119	0.27	
Compound 4	2.132	0.26	

Figure 0.0.7: Structure du site de liaison de CXCR4 lié à CXCL12. (A) Modèle d'homologie. (B) Structure expérimentale (PDB ID: 8U4O). (C) Structure AlphaFold-Multimer (jaune et orange) alignée sur la structure expérimentale (vert et bleu).



(A) **(B)** **(C)**

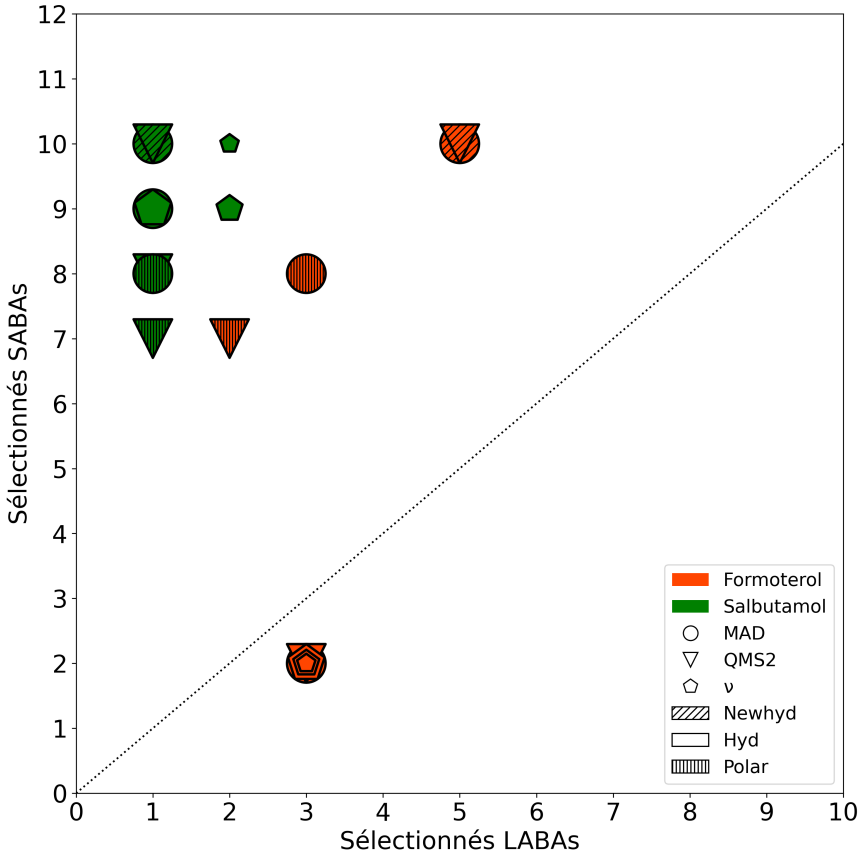
où seule une partie sélectionnait préférentiellement les LABA (Figure 0.0.8). L'analyse des propriétés des 22 agonistes de ADRB2 suggère que la durée d'action est fortement corrélée à la lipophilie, et que les LABA et les SABA présentent des différences structurales significatives. En se basant sur la similarité élevée entre les structures des molécules sélectionnées, nous supposons que le biais observé dans les modèles InteractionOCSVM est lié à la forme des ligands plutôt qu'à leurs propriétés physico-chimiques.

Enfin, nous avons utilisé ces modèles pour déterminer la durée d'action de l'agoniste BI167107, et nous avons conclu qu'il s'agit probablement d'un LABA. Cette prédiction a été confirmée par une approche de classification basée sur les propriétés physico-chimiques des 22 agonistes. Ensuite, la prédiction in silico a été validée expérimentalement sur un modèle murin d'asthme.

Modéliser des RCPGs et de leurs états conformationnels avec Alphafold

Pour modéliser de manière fiable différentes conformations d'une même protéine avec AlphaFold, nous avons développé un nouveau protocole intégrant plusieurs outils pour la modélisation multi-états. Ce protocole, appelé LIT-AlphaFold, a été testé lors d'une étude de référence visant à évaluer la qualité des structures des complexes chimiokine-récepteur de chimiokine[13]. Les structures générées sont de bonne qualité et reproduisent la plupart des caractéristiques observées dans les structures expérimentales. Tous les protocoles ont généré des structures avec un score moyen de DockQ de 0,70 ou 0,71, dont 13 sur 14 modèles (93%) étaient de qualité moyenne ou élevée. Comparé aux résultats d'autres études (variant entre 0,5 et 0,7 en utilisant différentes méthodes et jeux de données[23]), cela suggère que les complexes chimiokine-récepteur sont des cibles relativement accessibles pour AlphaFold-multimer. Un autre point important est la capacité à modéliser un état d'activation spécifique du récepteur. Différentes méthodes modifiant à la fois les informations de séquence et les modèles utilisés ont été étudiées afin de générer des structures du récepteur dans différents états d'activation, à la fois en tant que monomère et lié à une chimiokine. Nous confirmons que l'utilisation de modèles structuraux dans l'état d'activation souhaité est généralement suffisante pour obtenir une structure dans ce même état conformationnel pour

Figure 0.0.8: Performances des modèles InteractionOCSVM pour identifier les agonistes avec durée d'action longue (LABAs) ou courte (SABAs). MAD,QMS2,ν : différentes méthodes d'entrainement pour InteractionOCSVM. Hyd, Newhyd, Polar : Différentes définitions des interactions protéine-ligand. Hyd, définition plus générique des contacts hydrophobes. Newhyd, définition plus spécifique des contacts hydrophobes. Polar, tout interactions sauf les contacts hydrophobes. Adapté de l'article de Chiesa et al.[11].



le récepteur d'intérêt en tant que monomère. De plus, nous avons également réussi à forcer l'état conformationnel du récepteur d'intérêt en complexe avec une chimiokine, c'est-à-dire dans un complexe protéine-protéine.

Conclusions générale

La comparaison des modes d'interactions est un outil prometteur pour filtrer les résultats de docking et améliorer le taux de découvertes en criblage virtuel. Cependant, les performances de cette approche dépendent fortement de la cible et du mode de liaison de référence. InteractionOCSVM présente l'avantage d'une robustesse accrue par rapport aux méthodes existantes, mais reste tributaire de l'exactitude des données, comme illustré dans le cas de CXCR4. Malgré cela, les découvertes de molécules sont originales, ouvrant de nouvelles perspectives pour le développement de nouveaux ligands de CXCR4.

Notre expérience avec AlphaFold sur les récepteurs aux chimiokines a montré que les

structures générées constituent une bonne hypothèse de départ pour un projet de chimie médicinale.

En conclusion, pour concevoir de manière rationnelle des ligands fonctionnels des RCPGs, nous avons exploré les performances des outils in silico classiques, de l'apprentissage automatique et de l'apprentissage profond. Nos travaux proposent des stratégies pour leur application dans le filtrage de chimiothèques, le rééchantillonnage du docking et la prédiction de la structure des protéines.

Références

[1] Robert Fredriksson, Malin C. Lagerström, Lars-Gustav Lundin, and Helgi B. Schiöth. The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints. *Molecular Pharmacology*, 63(6):1256–1272, June 2003.

[2] Brian K. Kobilka. G protein coupled receptor structure and activation. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1768(4):794–807, April 2007.

[3] Brian K. Shoichet and Brian K. Kobilka. Structure-based drug screening for G-protein-coupled receptors. *Trends in Pharmacological Sciences*, 33(5):268–272, May 2012.

[4] Flavio Ballante, Albert J Kooistra, Stefanie Kampen, Chris de Graaf, and Jens Carlsson. Structure-Based Virtual Screening for Ligands of G Protein–Coupled Receptors: What Can Molecular Docking Do for You? *Pharmacological Reviews*, 73(4):1698–1736, October 2021.

[5] Douglas B. Kitchen, Hélène Decornez, John R. Furr, and Jürgen Bajorath. Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nature Reviews Drug Discovery*, 3(11):935–949, November 2004.

[6] Peter Kolb, Daniel M. Rosenbaum, John J. Irwin, Juan José Fung, Brian K. Kobilka, and Brian K. Shoichet. Structure-based discovery of $B2$-adrenergic receptor ligands. *Proceedings of the National Academy of Sciences*, 106(16):6843–6848, April 2009.

[7] Dahlia R. Weiss, SeungKirl Ahn, Maria F. Sassano, Andrew Kleist, Xiao Zhu, Ryan Strachan, Bryan L. Roth, Robert J. Lefkowitz, and Brian K. Shoichet. Conformation Guides Molecular Efficacy in Docking Screens of Activated β-2 Adrenergic G Protein Coupled Receptor. *ACS Chemical Biology*, 8(5):1018–1026, May 2013.

[8] David Rodríguez, Zhang-Guo Gao, Steven M. Moss, Kenneth A. Jacobson, and Jens Carlsson. Molecular Docking Screening Using Agonist-Bound GPCR Structures: Probing the A $_{2A}$ Adenosine Receptor. *Journal of Chemical Information and Modeling*, 55(3):550–563, March 2015.

[9] Chris de Graaf, Nicolas Foata, Ola Engkvist, and Didier Rognan. Molecular modeling of the second extracellular loop of G-protein coupled receptors and its implication on structure-based virtual screening. *Proteins: Structure, Function, and Bioinformatics*, 71(2):599–620, 2008.

[10] Luca Chiesa and Esther Kellenberger. One class classification for the detection of $B2$ adrenergic receptor agonists using single-ligand dynamic interaction data. *Journal of Cheminformatics*, 14(1):74, October 2022.

[11] Luca Chiesa, Emilie Sick, and Esther Kellenberger. Predicting the duration of action of $B2$-adrenergic receptor agonists: Ligand and structure-based approaches. *Molecular Informatics*, 42(12):e202300141, December 2023.

[12] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, August 2021.

[13] Lauri Urvas, Luca Chiesa, Guillaume Bret, Célien Jacquemard, and Esther Kellenberger. Benchmarking AlphaFold-Generated Structures of Chemokine–Chemokine Receptor Complexes. *Journal of Chemical Information and Modeling*, 64(11):4587–4600, June 2024.

[14] Albert J Kooistra, Stefan Mordalski, Gáspár Pándy-Szekeres, Mauricio Esguerra, Alibek Mamyrbekov, Christian Munk, György M Keserű, and David E Gloriam. GPCRdb in 2021: Integrating GPCR sequence, structure and function. *Nucleic Acids Research*, 49(D1):D335–D343, January 2021.

[15] Alexander S. Hauser, Misty M. Attwood, Mathias Rask-Andersen, Helgi B. Schiöth, and David E. Gloriam. Trends in GPCR drug discovery: New agents, targets and indications. *Nature Reviews Drug Discovery*, 16(12):829–842, December 2017.

[16] Oliver Korb, Thomas Stützle, and Thomas E. Exner. Empirical Scoring Functions for Advanced Protein-Ligand Docking with PLANTS. *Journal of Chemical Information and Modeling*, 49(1):84–96, January 2009.

[17] Marcel L. Verdonk, Jason C. Cole, Michael J. Hartshorn, Christopher W. Murray, and Richard D. Taylor. Improved protein–ligand docking using GOLD. *Proteins: Structure, Function, and Bioinformatics*, 52(4):609–623, 2003.

[18] Stefano Costanzi and Santiago Vilar. In Silico screening for agonists and blockers of the $B2$ adrenergic receptor: Implications of inactive and activated state structures. *Journal of Computational Chemistry*, 33(5):561–572, February 2012.

[19] Gilles Marcou and Didier Rognan. Optimizing Fragment and Scaffold Docking by Use of Molecular Interaction Fingerprints. *Journal of Chemical Information and Modeling*, 47(1):195–207, January 2007.

[20] Jérémy Desaphy, Eric Raimbaud, Pierre Ducrot, and Didier Rognan. Encoding Protein–Ligand Interaction Patterns in Fingerprints and Graphs. *Journal of Chemical Information and Modeling*, 53(3):623–637, March 2013.

[21] Tony Ngo, Bryan S. Stephens, Martin Gustavsson, Lauren G. Holden, Ruben Abagyan, Tracy M. Handel, and Irina Kufareva. Crosslinking-guided geometry of a complete CXC receptor-chemokine complex and the basis of chemokine subfamily selectivity. *PLOS Biology*, 18(4):e3000656, April 2020.

[22] Kei Saotome, Luke L. McGoldrick, Jo-Hao Ho, Trudy F. Ramlall, Sweta Shah, Michael J. Moore, Jee Hae Kim, Raymond Leidich, William C. Olson, and Matthew C. Franklin. Structural insights into CXCR4 modulation and oligomerization, February 2024.

[23] Richard Evans, Michael O'Neill, Alexander Pritzel, Natasha Antropova, Andrew Senior, Tim Green, Augustin Žídek, Russ Bates, Sam Blackwell, Jason Yim, Olaf Ronneberger, Sebastian Bodenstein, Michal Zielinski, Alex Bridgland, Anna Potapenko, Andrew Cowie, Kathryn Tunyasuvunakool, Rishub Jain, Ellen Clancy, Pushmeet Kohli, John Jumper, and Demis Hassabis. Protein complex prediction with AlphaFold-Multimer, March 2022.

Publications and Communications

Journal articles

1. *Chiesa, L.*; Kellenberger, E.
   One class classification for the detection of β2 adrenergic receptor agonists using single-ligand dynamic interaction data.
   Journal of Cheminformatics **2022**, 14, 74
   DOI : 10.1186/s13321-022-00654-z

2. Revillo Imbernon, J.; *Chiesa, L.*; Kellenberger, E.
   Mining the Protein Data Bank to inspire fragment library design.
   Frontiers in Chemistry **2023**, 11, 1089714
   DOI : 10.3389/fchem.2023.1089714

3. *Chiesa, L.*; Sick, E.; Kellenberger, E.
   Predicting the duration of action of β2-adrenergic receptor agonists: Ligand and structure-based approaches
   Molecular Informatics **2023**, 42, e202300141
   DOI : 10.1002/minf.202300141

4. Urvas L., *Chiesa, L.*, Bret G., Jaquemard C., Kellenberger E.
   Benchmarking AlphaFold-generated structures of chemokine – chemokine receptor complexes
   Journal of Chemical Information and Modelling **2024** 64, 11, 4587–4600
   DOI: 10.1021/acs.jcim.3c01835

Tutorials

1. *Chiesa, L.*, Urvas L., Bret G., Jaquemard C., Kellenberger E.
   Protein structure prediction with AlphaFold2 for drug discovery/Multistate modelling using LIT-AlphaFold.
   9^{th} Strasbourg Summer School in Chemoinformatics, Strasbourg, 24–28 June 2024

Conference talks

1. *Luca Chiesa* and Esther Kellenberger.
   One class classification and molecular dynamics identify ADRB2 ligands with specific pharmacological profiles using a single reference ligand.

Groupe graphisme et modélisation moléculaire (GGMM) 2023 - Young modellers conference, Toulouse, 15-17 May 2023

2. *Luca Chiesa* and Esther Kellenberger.
   Detection of $\beta 2$ adrenergic receptor agonists using single-ligand dynamic interaction data.
   11èmes Journées de la Société Française de Chémoinformatique, Caen, 5-6 October 2023

3. Lauri Urvas, *Luca Chiesa*, Guillaume Bret, Célien Jacquemard, Esther Kellenberger.
   Benchmarking AlphaFold-generated structures of chemokine – chemokine receptor complexes.
   3rd iGPCR-net meeting, Strasbourg – Illkirch-Graffenstaden, 25-27 October 2023

Conference posters

1. *Luca Chiesa* and Esther Kellenberger.
   A new machine learning based method for ADRB2 agonist detection using single-ligand dynamic interaction data.
   Journée du campus d'Illkirch, Illkirch-Graffenstaden, 23-24 May 2022

2. *Luca Chiesa* and Esther Kellenberger.
   A new machine learning based method for ADRB2 agonist detection using single-ligand dynamic interaction data.
   8th Strasbourg Summer School in Chemoinformatics, Strasbourg, 27 June – 1 July 2022

3. *Luca Chiesa* and Esther Kellenberger.
   A new machine learning based method for ADRB2 agonist detection using single-ligand dynamic interaction data.
   2nd iRCPG-net meeting, Wurzburg, 30 September - 1 October 2022

4. *Luca Chiesa* and Esther Kellenberger.
   Detection of $\beta 2$ adrenergic receptor agonists using single-ligand dynamic interaction data.
   3rd Transatlantic GPCR Symposium Early Career Investigator, on line, 7-8 September 2023

5. *Luca Chiesa* and Esther Kellenberger.
   Detection and classification of $\beta 2$ adrenergic receptor agonists using single-ligand dynamic interaction data.
   3rd iGPCR-net meeting, Strasbourg – Illkirch-Graffenstaden, 25-27 October 2023

6. *Luca Chiesa* and Esther Kellenberger.
   A new machine learning based method for ADRB2 agonist detection using single-ligand dynamic interaction data.
   9th French-Japanese Workshop on Computational Methods in Chemistry, Strasbourg, 24-25 April 2023

7. *Luca Chiesa* and Esther Kellenberger.
   Function-based virtual screening using machine learning and single ligand dynamic interaction data.
   European Workshop in Drug Design 2024, Certosa di Pontignano, 19-23 May 2024

8. *Chiesa, L.*, Urvas L., Bret G., Jaquemard C., Kellenberger E.
   LIT-AlphaFold: a new tool for multistate modelling.
   European Workshop in Drug Design 2024, Certosa di Pontignano, 19-23 May 2024

9. *Luca Chiesa* and Esther Kellenberger.
   Function-based virtual screening using machine learning and single ligand dynamic interaction data.
   9^{th} Strasbourg Summer School in Chemoinformatics, Strasbourg, 24 June – 28 June 2024

Chapter 1

Introduction

G-protein coupled receptors (GPCRs) represent one of the most important classes of drug targets, due to their role as mediators between the cell and the environment. These receptors, located on the cell membrane, are responsible for the regulation of intracellular signalling in response to external stimuli. The biological mechanism behind signalling regulation is intimately linked to the structure and plasticity of these proteins.

The first section of this chapter is dedicated to the description of the principal structural, thermodynamic, and pharmacological features of GPCRs, necessary to correctly understand the challenges behind the design of new bioactive molecules targeting them.

The second section is dedicated to an analysis of the most used techniques for GPCR drug discovery, with a particular focus on structure-based approaches. A particular attention is given to methods developed for the discovery of new ligands able to elicit a specific biological response upon binding.

1.1 G-protein coupled receptors

1.1.1 Therapeutic relevance, structure, and biological role

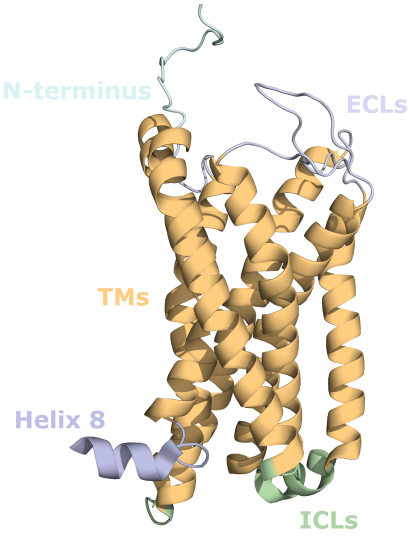
The G-protein coupled receptor (GPCR) family represent the largest family of membrane proteins in the human proteome. The human genome encodes ∼800 different GPCRs, with ∼400 being classified as sensory GPCRs, mediating olfaction, taste, light perception and pheromone signalling. The remaining 350 non-sensory GPCRs represent some of the most important targets for drug discovery, with an estimated ∼ 35% of all drugs approved until 2018 interacting with them [1]. Despite their popularity as therapeutic targets, only a third of all non-sensory GPCRs has been associated to an approved drug[1], making them still an attractive target for drug discovery.

Based on sequence similarity GPCRs are divided in five main classes: class A (rhodopsin-like), class B1 (secretin), class B2 (adhesion), class C (glutamate) and class F (frizzled/TAS2)[2]. This thesis will focus on the study of Class A GPCRs, the largest class with ∼700 receptors, encompassing multiple relevant drug discovery targets. Further observations about GPCRs will mostly concern members of Class A, even though some of them could apply also to other classes.

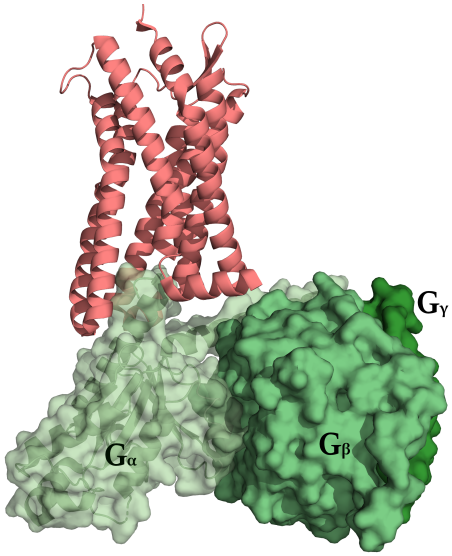
All GPCRs share a similar structure formed by seven hydrophobic transmembrane helical domains (TMs)[3], connected by two sets of loops located at the intracellular (ICLs), and at the extra-cellular (ECLs) side of the membrane, and an eight helix placed parallel to the membrane[4] (Figure 1.1.1A). GPCRs are characterized by the presence of two main binding pockets, located at the two sides of the cellular membrane. The extracellular cavity binds diffusible ligands of different biological (hormones, neurotransmitters, cytokines) and chemical (ions, small molecules, peptides, and proteins) nature [3]. The intracellular pocket mainly binds G-protein, a heterotrimeric protein formed by three subunits (α, β, γ). G-proteins are divided between four major families (G_s, $G_{i/o}$, $G_{q/11}$, and $G_{12/13}$) based on the type of α subunit, which forms the main interface with the receptor (Figure 1.1.1B). Around half of all GPCRs selectively bind two or more G-proteins. Unlike its extracellular counterpart, the intracellular binding pocket is observed only in one conformation of the receptor, called the active state. Structures without an exposed intracellular binding pocket are classified as inactive. The main structural difference between the two activation states concerns the position of TM6 compared to the rest of the helical bundle. In the inactive state TM3, TM5, and TM6 are packed together, and the structure is further stabilized by the effect of a transient ionic lock between TM3 and TM6. In this conformation the small size of the intracellular binding pocket is insufficient to allow G-protein coupling. A large outward movement of TM6 compared to the central TM3 opens the intracellular binding pocket, allowing the residues of TM3 and TM5, which were interacting with TM6, to interact with G-protein [3, 4] (Figure 1.1.1C).

The binding of a ligand in the extracellular binding pocket, either endogenous or xenobiotic, influences the binding of G-protein to the receptor. GPCRs act as mediators between the environment and the cell thanks to the relationship between the two binding pockets. The binding of G-protein to the receptor induces the separation of the G-protein subunits ($G\alpha$, and $G\beta\gamma$), which can interact with other effector proteins in the cytoplasm, triggering a signalling cascade inside the cell, and resulting in a biological response [5, 6].

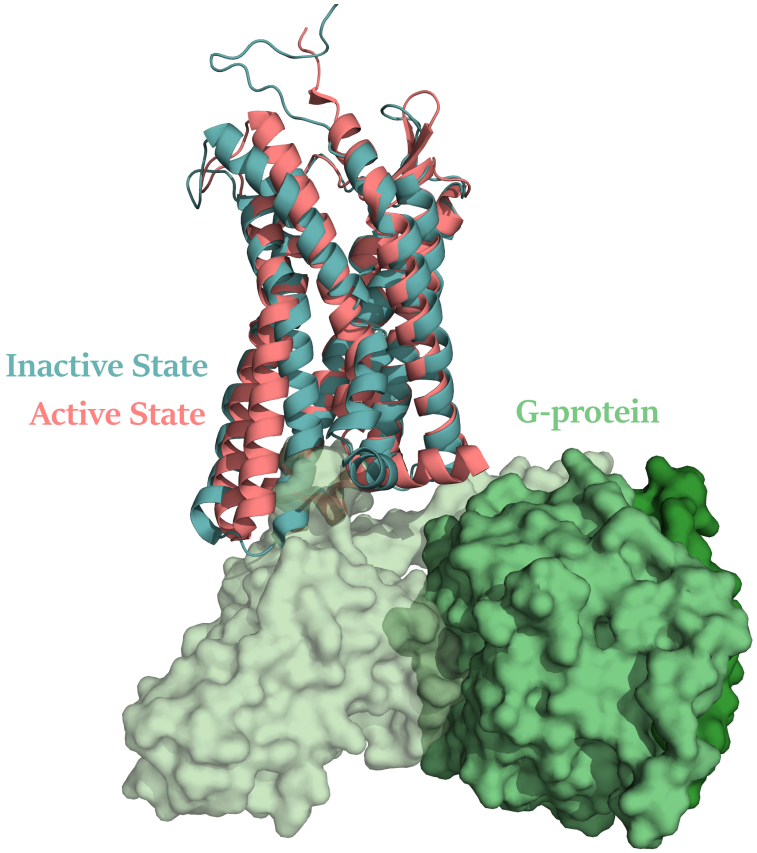
Figure 1.1.1: (A) Key structural domains of a GPCR (PDB ID: 6MEO). (B) Structure of the complex between a GPCR and the heterotrimeric G-protein (PDB ID: 7F1Q). (C) Aligned structures of the C-C chemokine receptor type 5 in the inactive (PDB ID: 6MEO) and active (PDB ID: 6MEO) state. The G_α subunit of the G-protein is shown as transparent to highlight the steric clash with the inactive state structure.



(A)



(B)



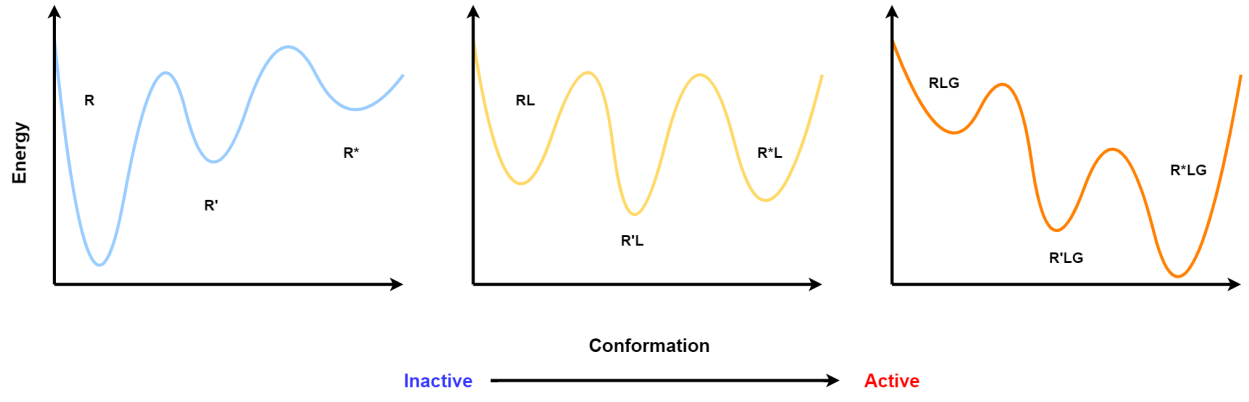
(C)

1.1.2 Allosteric regulation of intracellular signalling

Proteins can assume different conformations at physiological conditions, with different states coexisting in equilibrium condition. The relative population of each state is determined by a Boltzmann distribution, depending on their relative free energy. The probability of transition from one state to the other depends on the free energy of each conformation, and the energy barrier associated with the structural rearrangements necessary to pass from one state to the other [7–9]. Excluding few constitutively active GPCRs [10], the favourite conformation of a free receptor is the inactive state [4].

The binding of a ligand in the extracellular pocket influences the ability of the receptor to bind G-protein, or other signalling proteins, at the intracellular side, making them a classical example of allosteric proteins[4]. Allosteric regulation follows a conformational selection and equilibrium shift mechanism. Ligands selectively bind one conformation of the pocket and stabilize it, altering the free energy landscape compared to the apo receptor (Figure 1.1.2). Changes in the free energy profile lead to a shift in the relative population between different states. Agonist binding increases the population of receptors in an active state conformation, while the binding of an inverse agonist increases the population of receptors in an inactive state conformation. Neutral antagonists do not alter the free energy landscape and they maintain the inactive state conformation as the most populated state. The main biological role of antagonists is to block the binding of agonists and inverse agonsits to the extracellular pocket[11]. Ligands not binding in the orthosteric site, the pocket bound by the endogenous agonist, can still affect the conformational equilibrium, and subsequently regulate intracellular signalling. These molecules are called allosteric ligands, and their binding sites are called allosteric sites[12].

Figure 1.1.2: Free energy profile of a receptor in the inactive (R), active (R*), and an active-like (R') conformation. Binding of an agonist ligand to the receptor is indicated by L, and binding to G-protein by G.



Receptor activation is achieved by multiple conformational selection step[9, 13], with no strict coupling between agonist and G-protein binding[4]. Ligand recognition at the receptor follows the "conformational cafeteria" principle, with both agonists and G protein binding to the pocket conformation for which they have the highest affinity[6]. Agonists preferentially bind receptor structures in an intermediate activation state, with an active-like conformation of the extracellular pocket. Upon binding the receptor population distribution shifts towards

more active-like conformations[4, 14]. Some of these agonist-bound structures are able to couple to G-protein, which further stabilize the signalling complex (Figure 1.1.2). A receptor can assume different active state conformations depending on the bound G-protein subtype, or signalling protein. Different agonists stabilize different conformations of the receptor, facilitating the binding of specific proteins at the intracellular side.

Based on the free energy landscape model, a small portion of the receptors population is always in an active state conformation, even in the absence of a ligand bound to the extracellular binding pocket. These receptors in the active state can bind G-protein, and trigger intracellular signalling, leading to a low but constant level of signalling intensity. This phenomena is described as the basal activity of the receptor[11].

1.1.3 Affinity and efficacy

The interaction between a protein and a non-covalently bound ligand, being a small molecule, a peptide, or another protein, is generally measured in terms of binding affinity. The binding affinity of a ligand for a target describes the strength of the interaction between the two entities, and it is measured by the dissociation constant of the complex, as described by Equation 1.1 and Equation 1.2. Where R indicates the receptor, L the ligand, and [ ] indicates the activity of a species expressed as their concentrations in $mol \cdot L^{-1}$.

$$RL \rightleftharpoons R + L \tag{1.1}$$

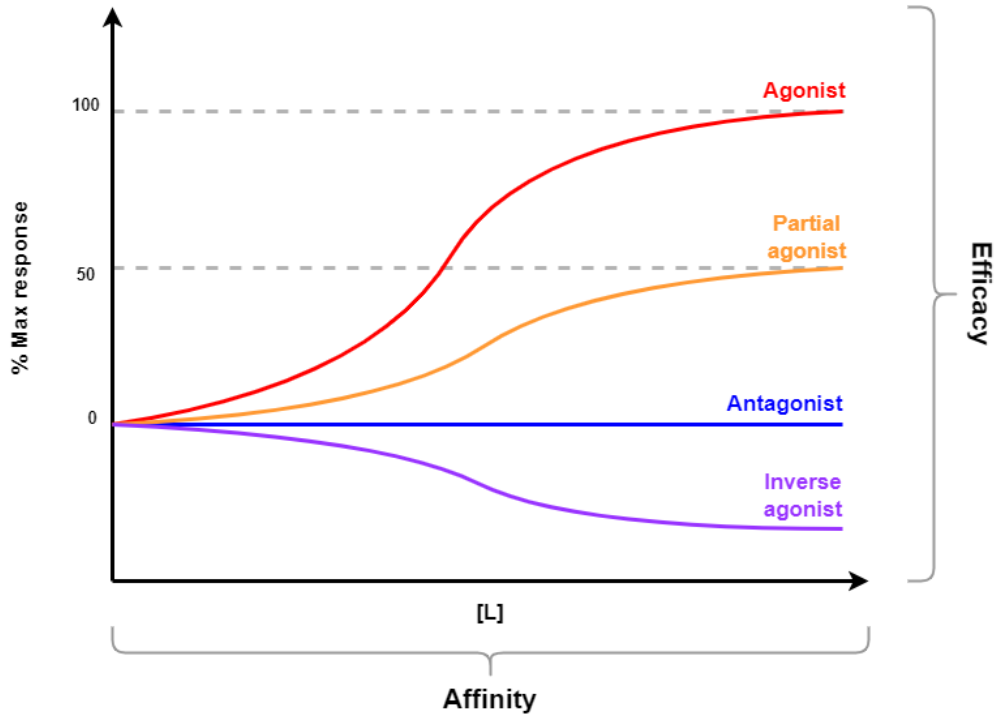
$$K_d = \frac{[R][L]}{[RL]} \tag{1.2}$$

Binding affinity describes how strong the interaction between a ligand and a receptor is, but it holds no information regarding the biological response caused by ligand binding. As shown in the previous sections, ligand binding can promote or not promote the coupling of G-protein, and as a consequence, it can trigger or inhibit intracellular signalling. To describe the relationship between the fraction of bound receptor and a measured biological response Stephenson introduced the concept of efficacy (Equation 1.3)[15, 16].

$$S_A = e_A p_{AR} \tag{1.3}$$

The measured biological response R is related a quantity called "stimulus" S_A by a non linear relationship defined as $R = f(S_A)$. The S_A value is directly proportional to the fraction of occupied receptors p_{AR}, by a constant e_A, called efficacy, which is specific for a ligand-receptor pair. The concept of efficacy could be used to explain the differences in biological response between ligands, even for molecules with the same binding affinity for the receptor (Figure 1.1.3). Efficacy is a vectorial quantity, with agonists having a positive efficacy, increasing activity from its basal level, and inverse agonists having a negative efficacy, decreasing activity from its basal level[17]. Efficacy not only determines the presence of a biological response, but also its intensity, with different ligands with the same affinity triggering a biological response with a different intensity. A reference molecule, often the endogenous ligand, is defined as a full agonist of the receptor, while other ligands with a lower efficacy are classified as partial agonists. Antagonists do not have an efficacy value based on this definition, since their binding does not elicit a biological response, maintaining activity at basal level. The role of antagonists is to occupy the extracellular pocket, preventing the binding of agonists and inverse agonists [9].

Figure 1.1.3: Effect of the binding of an agonist, a partial agonist, a neutral antagonist, and an inverse agonist on the maximal agonist response. 0% maximal response correspond to the receptor basal activity.



The affinity and efficacy of a ligand for a receptor are described by its potency, the concentration of ligand necessary to produce a specific response[15]. The potency of an agonist, or inverse agonist, is measured by the half maximal effective concentration (EC50), the ligand concentration producing 50% of the maximal possible effect[15]. The EC50 value however does not give any information about the maximal response triggered by ligand binding, which is instead represented by the maximal agonist effect[15]. Antagonist potency instead is measured by the half maximal inhibitory concentration (IC50), the antagonist concentration reducing the biological response by 50% at a given agonist concentration[15].

Efficacy measures the relationship between ligand binding and a biological response, therefore it depends on the evaluated property and the assay used to measure it. Historically, experiments aimed at measuring the biological response to a molecule administration focused on the effects at tissue level, using extracted tissue or whole-animal assays. Nowadays, more sophisticated assays allow to study different aspects of intracellular signalling, returning a more complete and nuanced picture of a ligand's pharmacology. Two of the most popular types of functional assays are direct signalling pathway assays, which measure the production of secondary messengers (cytoplasmic Ca^{2+} or cAMP levels), or whole cell assays, which measure a wider range of receptor behaviours, such as receptor internalization, or β-arrestin interaction[18]. The potency of a molecule for a receptor is intrinsically bound to the measured property. Moreover, the same response in a functional assay could be caused by the binding of different subtypes of G-protein, but with different efficacy values depending on the signalling protein [19]. Considering the different assays, signalling pathways, and G-protein subtypes, the efficacy of a ligand-receptor pair cannot be described by a single efficacy value,

but rather it is necessary to consider the "pluridimensionality of efficacy" [19].

Based on these observations GPCRs do not act as simple on/off switches of intracellular signalling, but rather as microprocessors [18], since the binding of a single ligand independently influences different aspects of intracellular signalling, with its regulation ranging from full activation to inhibition. Since a ligand does not present the same efficacy value for all signalling proteins, a bias towards a certain signalling pathway is generally observed. It is now well accepted that each ligands exist in a spectrum of signalling bias, ranging from unbiased to fully biased [20]. A signalling bias is measured by comparing the potencies of the tested ligand to the potency of the endogenous ligand, or to the potency of a reference compound designated as a balanced agonist or antagonist [20]. Signalling bias is not limited to ligands, but it also extends to receptor bias, where the signalling bias is a characteristic of the receptor itself, and system bias, where other environmental factors influence intracellular signalling[20].

References

[1] Krishna Sriram and Paul A. Insel. G Protein-Coupled Receptors as Targets for Approved Drugs: How Many Targets and How Many Drugs? *Molecular Pharmacology*, 93(4):251–258, April 2018.

[2] Robert Fredriksson, Malin C. Lagerström, Lars-Gustav Lundin, and Helgi B. Schiöth. The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints. *Molecular Pharmacology*, 63(6):1256–1272, June 2003.

[3] Brian K. Kobilka. G protein coupled receptor structure and activation. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1768(4):794–807, April 2007.

[4] William I. Weis and Brian K. Kobilka. The Molecular Basis of G Protein–Coupled Receptor Activation. *Annual Review of Biochemistry*, 87(1):897–919, June 2018.

[5] Daniel Hilger, Matthieu Masureel, and Brian K. Kobilka. Structure and dynamics of GPCR signaling complexes. *Nature Structural & Molecular Biology*, 25(1):4–12, January 2018.

[6] Terry Kenakin. Efficacy at g-protein-coupled receptors. *Nature Reviews Drug Discovery*, 1(2):103–110, February 2002.

[7] Arvind Ramanathan, Andrej Savol, Virginia Burger, Chakra S. Chennubhotla, and Pratul K. Agarwal. Protein Conformational Populations and Functionally Relevant Substates. *Accounts of Chemical Research*, 47(1):149–156, January 2014.

[8] Hans Frauenfelder, Stephen G. Sligar, and Peter G. Wolynes. The Energy Landscapes and Motions of Proteins. *Science*, 254(5038):1598–1603, December 1991.

[9] Brian K. Kobilka and Xavier Deupi. Conformational complexity of G-protein-coupled receptors. *Trends in Pharmacological Sciences*, 28(8):397–406, August 2007.

[10] Roland Seifert and Katharina Wenzel-Seifert. Constitutive activity of G-protein-coupled receptors: Cause of disease and common property of wild-type receptors. *Naunyn-Schmiedeberg's Archives of Pharmacology*, 366(5):381–416, November 2002.

[11] Xiao Jie Yao, Gisselle Vélez Ruiz, Matthew R. Whorton, Søren G. F. Rasmussen, Brian T. DeVree, Xavier Deupi, Roger K. Sunahara, and Brian Kobilka. The effect of ligand efficacy on the formation and stability of a GPCR-G protein complex. *Proceedings of the National Academy of Sciences*, 106(23):9501–9506, June 2009.

[12] Arthur Christopoulos, Jean-Pierre Changeux, William A. Catterall, Doriano Fabbro, Thomas P. Burris, John A. Cidlowski, Richard W. Olsen, John A. Peters, Richard R. Neubig, Jean-Philippe Pin, Patrick M. Sexton, Terry P. Kenakin, Frederick J. Ehlert, Michael Spedding, and Christopher J. Langmead. International Union of Basic and Clinical Pharmacology. XC. Multisite Pharmacology: Recommendations for the Nomenclature of Receptor Allosterism and Allosteric Ligands. *Pharmacological Reviews*, 66(4):918–947, October 2014.

[13] Ruth Nussinov, Buyong Ma, and Chung-Jung Tsai. Multiple conformational selection and induced fit events take place in allosteric propagation. *Biophysical Chemistry*, 186:22–30, February 2014.

[14] Peishen Zhao and Sebastian G. B. Furness. The nature of efficacy at G protein-coupled receptors. *Biochemical Pharmacology*, 170:113647, December 2019.

[15] Richard R. Neubig, Michael Spedding, Terry Kenakin, and Arthur Christopoulos. International Union of Pharmacology Committee on Receptor Nomenclature and Drug Classification. XXXVIII. Update on Terms and Symbols in Quantitative Pharmacology. *Pharmacological Reviews*, 55(4):597–606, December 2003.

[16] R. P. Stephenson. A Modification of Receptor Theory. *British Journal of Pharmacology and Chemotherapy*, 11(4):379–393, 1956.

[17] T Costa and A Herz. Antagonists with negative intrinsic activity at delta opioid receptors coupled to GTP-binding proteins. *Proceedings of the National Academy of Sciences*, 86(19):7321–7325, October 1989.

[18] Terry P. Kenakin. Cellular assays as portals to seven-transmembrane receptor-based drug discovery. *Nature Reviews Drug Discovery*, 8(8):617–626, August 2009.

[19] Ségolène Galandrin, Geneviève Oligny-Longpré, and Michel Bouvier. The evasive nature of drug efficacy: Implications for drug discovery. *Trends in Pharmacological Sciences*, 28(8):423–430, August 2007.

[20] Peter Kolb, Terry Kenakin, Stephen P. H. Alexander, Marcel Bermudez, Laura M. Bohn, Christian S. Breinholt, Michel Bouvier, Stephen J. Hill, Evi Kostenis, Kirill A. Martemyanov, Rick R. Neubig, H. Ongun Onaran, Sudarshan Rajagopal, Bryan L. Roth, Jana Selent, Arun K. Shukla, Martha E. Sommer, and David E. Gloriam. Community guidelines for GPCR ligand bias: IUPHAR review 32. *British Journal of Pharmacology*, 179(14):3651–3674, 2022.

1.2 Efficacy-driven drug design

The G-protein-coupled receptor (GPCR) family represent one of the most important targets in drug discovery. Ligands targeting GPCRs have been developed into drugs thanks to their ability to selectively regulate a biological function [1, 2]. For example, agonists and antagonists of the β-adrenergic receptor family are used in the clinical practice, but with different indications depending on their efficacy. Antagonists, known as β-blockers, are used to manage abnormal heart rhythms [3], thanks to their action on the β1-adrenergic receptor, mostly located in cardiac tissues. Agonists instead are used for the treatment of respiratory diseases such as asthma and chronic obstructive pulmonary disease (COPD)[4, 5], thanks to their activity at the β2-adrenergic receptor, mostly located in the trachea and lung tissues. Due to their high sequence and structure similarity the two receptors often bind the same molecules, although with different affinity. Since the two classes of medications have opposite effects, their off-target activity is linked to undesirable side-effects, and it restrict their prescription to patients. Ligands binding the same receptor can have a different therapeutic application depending on their efficacy.

Computer assisted drug design has facilitated the discovery of new bioactive molecules for multiple protein targets, in particular for GPCRs. Computational methods can be applied to perform a high-throughput virtual screening of a chemical library in order to identify potential binders of a target receptor. For GPCRs however, virtual screening should not focus only on binding affinity, but it should also predict a molecule's efficacy, to facilitate the design of new drugs with a desired mode of action. Currently the discovery of new ligands using structure-based drug design techniques has proven to be successful on multiple targets[6], while the discovery of new ligands with a desired efficacy has proven to be more challenging, despite offering interesting rewards for the development of new therapies[7]

1.2.1 Protein-ligand docking

Protein-ligand docking, or molecular docking, is an *in silico* drug design technique used to determine the structural and electronic complementarity between a molecule and the three dimensional structure of the protein, often used to identify potentially bioactive molecules [8, 9]. The docking process is based on a docking algorithm, which generates poses of the molecule in the protein's binding pocket, and a scoring function, which evaluates the fitness of the generated poses, and the binding affinity of the docked molecule [10]. Docking algorithms explore the conformational space of a protein-ligand complex to identify the pose which better complement the shape and electrostatic features of the binding site. Depending on the explored degrees of freedom, docking algorithms can be divided in three groups: rigid docking, where the algorithm intervenes on the relative position of the protein and the ligand, but not on their structures; ligand flexible docking, where the protein structure is maintained rigid, while different ligand conformation are explored; flexible docking, where both the ligand and the protein's sidechains are flexible[10]. A scoring function estimates the binding affinity of the molecule for the target, or the fitness of the pose based on the interactions formed with the pocket. Depending on the approximations, and the informations used to determine the docking score, scoring functions have traditionally been classified in three groups: force filed, empiric, and knowledge based[10]. More recently machine learning scoring functions [11] have emerged, which use machine learning models to estimate the binding affinity of a molecule based on information extracted from experimental protein-ligand complexes. Thanks to its

speed and relative accuracy, docking has evolved from a tool to determine the structure of a protein-ligand complex, to a high-throughput virtual screening method to identify potential bioactive molecules from chemical libraries [12].

Historically, the application of protein-ligand docking to GPCRs has been severely limited by the lack of experimental structures. The resolution of the 3D structure of bovine rhodopsin in 2000 kick-started structure-based drug design GPCRs [13]. Homology models based on the bovine rhodopsin structure were used for virtual screening application, leading in multiple cases to the discovery of new ligands[14–16]. While these models were effective in identifying new ligands, they had no control on the efficacy of these hits, with multiple instances of agonist identified from homology models based on the inactive state rhodopsin structure[14–16]. The first structure of a GPCR bound to a diffusible ligand was published in 2007, describing the complex between the β2-adrenergic receptor (ADRB2) and the antagonist carazolol[17]. Soon thereafter the first results from a virtual screening campaign using the experimental structure for docking were reported by Kolb et al. [18]. In the following years multiple structures of other GPCRs bound to antagonists or inverse agonists were published[19–22], and then used for docking-based virtual screening [23–29] (Figure 1.2.1). These early screening campaigns selectively identified new antagonists or inverse agonists, while the discovery of new agonists was lagging behind due to the lack of experimental active state structures. *In silico* models of active state structures were developed by adapting experimental inactive state structures to accommodate agonist binding. These structures often performed well in retrospective experiments [30, 31], but prospective applications often yielded antagonists or negative allosteric modulators[32–34]. In 2011 the active state structure of ADRB2 bound to the agonist BI167107[35, 36] was released, describing for the first time the structure of a GPCR in the active state. Agonists and partial agonists of ADRB2 were selectively identified by docking on this structure [37] (Figure 1.2.2).

Multiple virtual screening campaigns using docking and experimental GPCR structures showed a correlation between the activation state of the used structure and the efficacy of the hit molecules[1]. Docking in inactive state structures selectively identified antagonists and inverse agonists [18, 24–26, 28, 38–43], while agonists were identified by performing docking on active state structures [37, 44–46]. Selecting a structure in the correct activation state was a crucial step for the discovery of new ligands with a desired efficacy, but it did not guarantee efficacy selectivity. Multiple publications reported the discovery of agonists by docking on experimental inactive state structures[29, 33, 47–49], and antagonists from agonist bound inactive state structures[50] or active state structures[51].

1.2.2 Docking post-processing

While protein-ligand docking remains one of the most used techniques for GPCR drug discovery [6], the low accuracy of scoring functions requires further analysis of the top ranking compounds. Different post-processing protocols have been developed to improve the performance of docking based on better physical descriptions of the system, comparison to experimental data[52], or by relaying on the experience and capabilities of computational and medicinal chemists [53].

Post-processing methods aiming at the selective discovery of new ligands with a desired efficacy can be divided in two main categories: conformational selection methods, which relay on the difference in affinity of a molecule for different protein conformations [54]; and

Figure 1.2.1: Time-line describing the resolution of the first GPCR structures, and their first use in virtual screening.

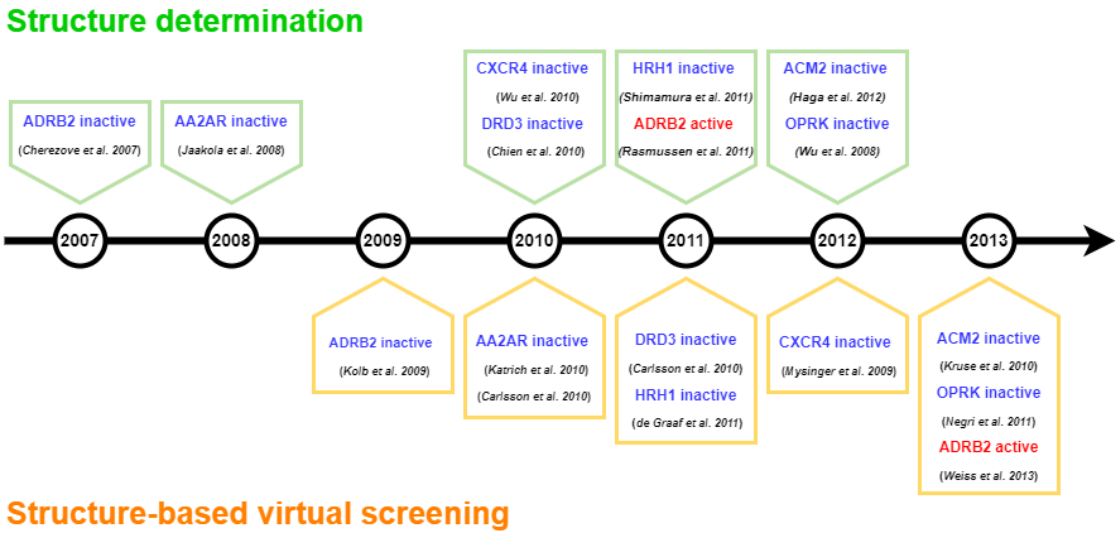
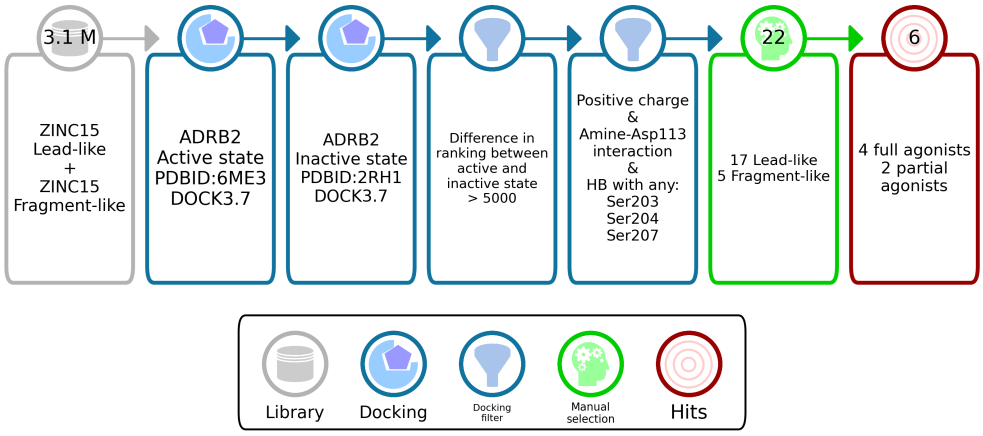


Figure 1.2.2: Schematic description of the virtual screening campaign conducted by Weiss et al. [37]



interaction similarity methods[55], which exploit the differences in binding mode between agonists and antagonists.

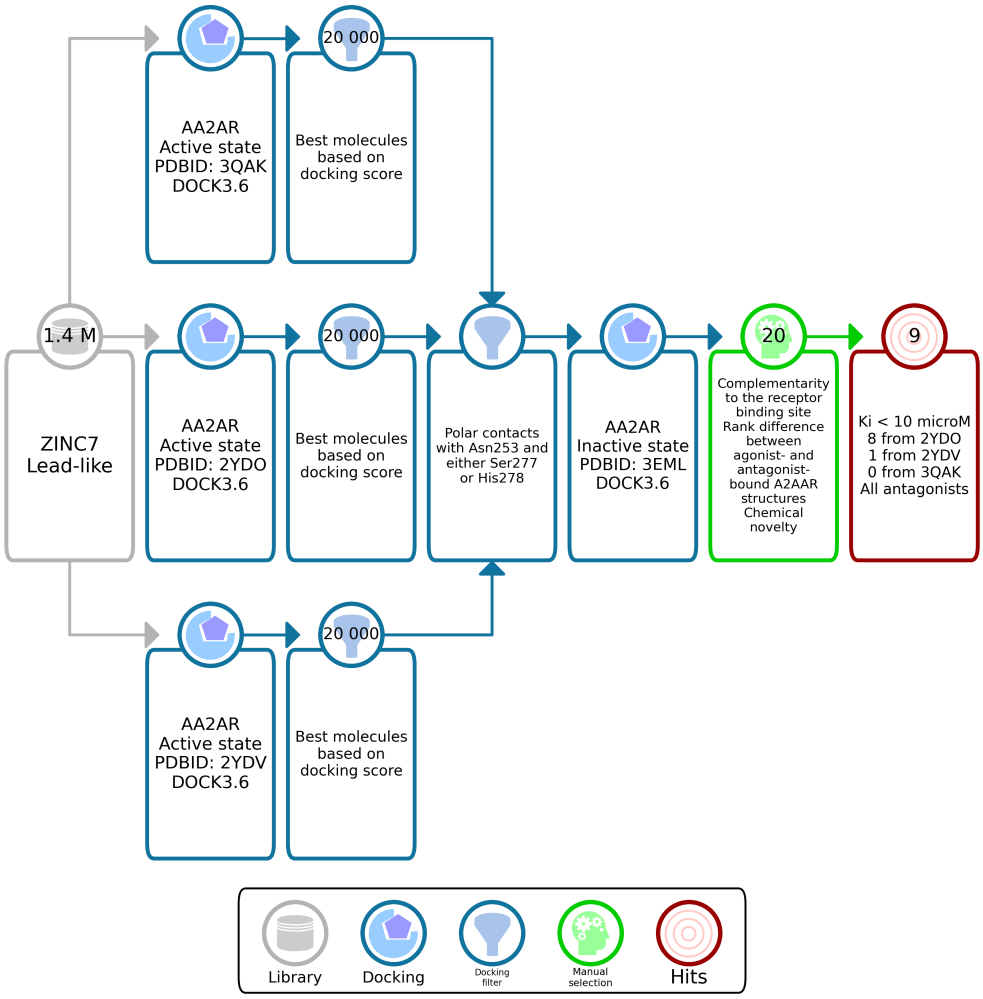
Conformational selection

Conformational selection methods exploit the difference in a ligand's binding affinity for two conformations of the same binding site, to identify its preferred conformation. For GPCRs the selectivity of a ligand towards a certain conformation determines its efficacy, with agonists preferentially binding active state conformations, while antagonists and inverse agonists prefer to bind inactive state conformations. Conformational selection methods are not based on the difference in binding affinity between agonists and antagonists for the same conformation, but rather on the difference in affinity of a ligand for the active and the inactive state[56].

Scoring functions In high-throughput virtual screening the docking score is often used as a proxy for a molecule's binding affinity, which depends on the structure used to perform the calculations[57, 58]. Multiple retrospective studies conducted on active and inactive state crystallographic structures of the β2-adrenergic receptor (ADRB2), and of the adenosine A2A receptor (AA2AR), have shown that docking on a structure in the "correct" activation state leads to selectively retrieve ligands with the desired efficacy [37, 51, 59]. The same trend was also observed in few virtual screening campaigns not relaying on additional post-processing tools other than manual selection [23, 24]. Since docking scores can roughly reproduce the binding affinity, or at least the affinity ranking[60], different prospective virtual screening campaigns have used the difference in rank to select virtual hits. For example Weiss et al. used the difference in docking ranking to selectively identify new agonists of ADRB2[37]. The top ranking molecules based on docking score for an active state structure were further filtered based on their difference in ranking after docking in an active and in an inactive state structure of the receptor (Figure 1.2.2). A similar protocol was used by Rodriguez et al. to identify new agonists of the adenosine A2A receptor (AA2AR)[51]. Molecules from commercial libraries were docked in an ensemble of three active state structures, and one reference inactive state structure. The difference in ranking between the active and the inactive state conformations was used as a criteria during manual selection (Figure 1.2.3). None of the confirmed hits selected using this method could promoted cAMP production in a functional assay, therefore no agonist ligand was identified. Scharf et al. used a "selective reranking" approach to identify new agonists of ADRB2[45]. Molecules were ranked based on the ratio between their ranking after docking on an inactive state structure, and on an active state structure. All the 500 compounds with the best selective reranking scores were discarded during visual inspection.

Free energy calculations Comparison of docking scores between different conformations allows to quickly estimate the difference in binding affinity, however these values lack in accuracy, and depend on the scoring power of the scoring function[60]. A more accurate estimate of binding free energy can be computed using molecular mechanics. Alchemical free energy calculations can measure the difference in free energy between two states using a physically unrealistic pathway, by transforming one molecule into another (relative free energy) of by progressive decoupling of the molecule from the receptor (absolute free energy)[61]. Between the different alchemical free energy methods free energy perturbation

Figure 1.2.3: Schematic description of the virtual screening campaign described by Rodriguez et al. [51]



(FEP) has been often used on GPCRs to determine the relative free energy of binding of congeneric ligands [62], to guide the optimization of a selective adenosine A1 receptor (A1AR) ligand [63], or to determine the effect of mutations on the affinity of AA2AR ligands [64]. By using a carefully built thermodynamic cycle, it is possible to determine the difference in binding free energy of a ligand for the active and the inactive state of the receptor, making FEP a suitable method for efficacy prediction.

Jespers et al. [64] used FEP to measure the relative binding free energy between analogous agonist/antagonist pairs of A1AR ligands. Not only FEP correctly determined the agonists' higher affinity for the active state, but a higher difference in relative binding free energy was measured for the full agonist NECA, compared to the partial agonist adenosine. Vogele et al. [65] used absolute free energy calculations to study the difference in binding affinity between the active and the inactive state of the receptor for a panel of eight GPCRs. The results showed a clear trend, with agonists having a higher affinity for active state structures compared to inactive state structures, and vice versa. Despite the promising results FEP has yet to be applied to post-process the results of a high-throughput virtual screening. The closest example to a true application of FEP for efficacy-driven drug discovery is the work of Panel et al. [66], where FEP was used to evaluate the initial hits from a virtual screening campaign on a library of ADRB2 agonists analogues. The difference in binding free energy to the active and the inactive state was measured, identifying seven potential agonsits or partial agonists. Binding and functional assays confirmed that of the seven virtual hits four had agonist efficacy, one was an antagonist, and two did not bind the receptor. Furthermore, two of the agonist hits were modified to disrupt or facilitate known key interactions with the receptor. FEP calculations matched the variations in affinity and efficacy measured in biological assays for these derivatives.

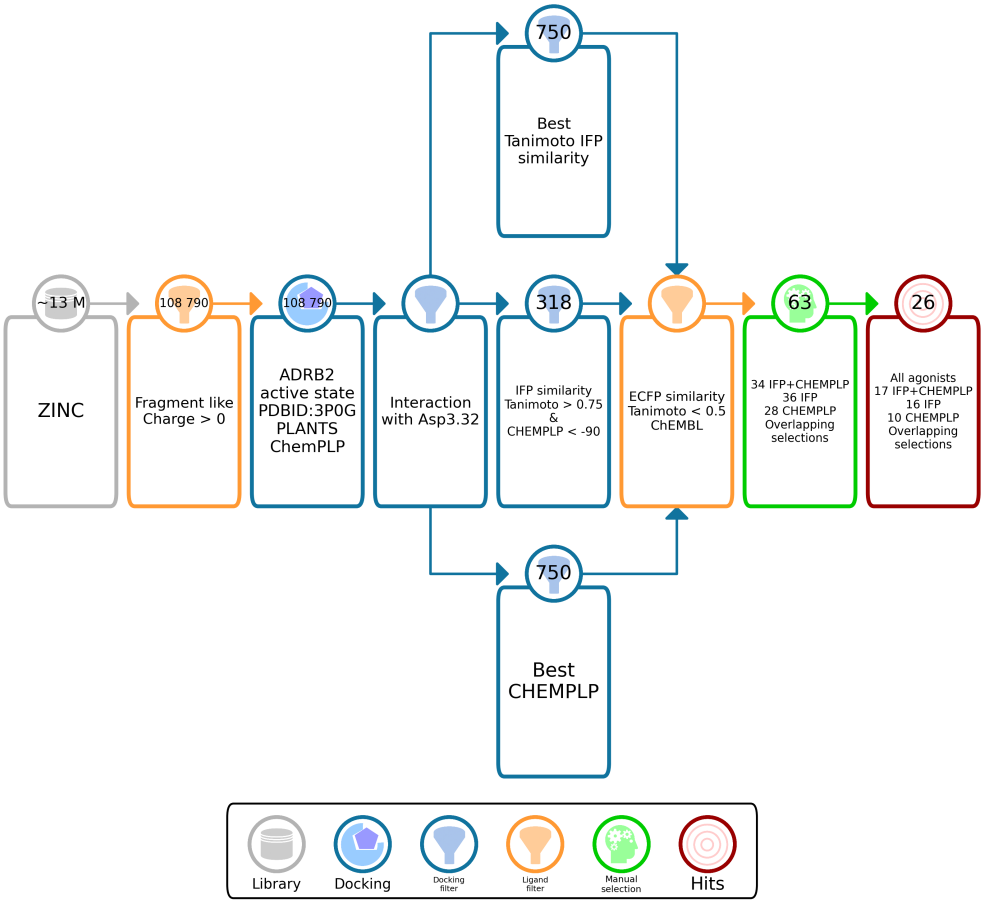
Interaction similarity

The analysis of the binding mode of the generated poses has always been a key part of the evaluation of docking results. Even in the first screening campaigns based on homology models the presence of certain known key interactions was used to select relevant docking poses [14]. Filtering based on the presence of specific interactions, either automatically or manually, has been used in multiple virtual screening campaigns targeting GPCRs [26, 28, 29, 34, 37, 40, 41, 44, 46–48, 50, 51, 66–68].

Compared to interaction filtering, interaction similarity does not consider the presence of specific non-covalent interactions, but rather it relays on the comparison of the entire binding mode after converting it into a computer readable representation[55, 69]. Interaction fingerprints (IFPs) represent the most common type of interaction representation. Interactions are represented as bit vectors describing the interactions at residue level[52, 70, 71] or atom level[72, 73], or by using hashed vectors of finite size [74–76]. Each fingerprint encodes specific non-bonded interactions [52, 70], or all possible forms of protein-ligand contacts[75]. The similarity between two binding modes can be measured by Tanimoto similarity between binary fingerprints, or by using more complex similarity scores[74, 77]. Other than fingerprints, graphs have also been used to represent protein-ligand interactions [78].

The link between interaction similarity and efficacy bias in virtual screening was first shown by de Graaf and Rognan, who obtained a selective increase in recall for ADRB2 agonists or antagonists depending on the efficacy of the reference ligand[30]. A selective

Figure 1.2.4: Schematic description of the virtual screening campaign described by Kooistra et al. [44]



enrichment in antagonists was achieved using a crystallographic structure of the inactive state receptor, and the binding mode of the co-crystalized ligand carazolol, while selective enrichment in agonists was achieved using a modelled active state structure, and a docking-generated reference binding mode. Similar results were obtained by Kooistra et al. [79] in a retrospective virtual screening study using active and inactive state crystallographic structures of ADRB2 and ADRB1. Ranking based on interaction similarity has been used to selectively identify antagonists of HRH1[27] , CXCR4[80], and OPRK[43], as well as agonists of ADRB2 (Figure 1.2.4)[44].

Conformational selection vs Interaction similarity

Conformational selection and interaction similarity represent two different approaches to solve the problem of efficacy-driven drug discovery. Conformational selection is a physics-based method, since it relays on the thermodynamics of ligand recognition and signalling regulation. Interaction similarity instead represents a pharmacophore and knowledge-based approach. IFPs indirectly encode information about the presence of pharmacophoric features in the reference ligand, as well as their relative spatial location based on the residues forming the interactions. Interaction filtering instead only selects ligands with chemical features able

to form the desired interactions. At the same time, knowledge from experimental receptor structures or from mutagenesis experiments is used to establish a link between the efficacy of a ligand and its ability to form a certain interaction or binding mode.

The two approaches represent orthogonal methods which evaluate the same biological property, each with its own strength and limitations. Conformational selection requires an active and an inactive state structure of the receptor, limiting the analysis to few targets. Additionally, it is necessary to strike the correct balance between speed and accuracy in order to select the best method to measure the difference in binding affinity. Interaction similarity only requires a structure of the receptor with a ligand bound to the extracellular pocket. However, the number of retrieved hits depends on the reference binding mode, and small differences in the formed interactions can lead to different hit lists in a high-throughput context [79].

Based on literature data there is no best approach for efficacy-driven drug discovery, but rather it is necessary to evaluate the problem on a case by case basis, depending on data availability (known ligands, crystallographic structures), and on the results obtained in retrospective studies mimicking the real conditions of virtual screening, regarding both the structure, the computational methods, and the evaluated chemical library [51]. Despite their differences, the two approaches are not mutually exclusive and they have already been combined in the same virtual screening pipeline to selectively retrieve ligands with a desired efficacy [37, 51, 66]. As an example, Panel et al.[66] used interaction filters to select the docking poses used for FEP calculations.

1.2.3 Application of machine learning in GPCR drug discovery

Machine learning methods represent a new resource for drug discovery, with many "classical" computational tools such as docking algorithms, scoring functions, library generation methods [81–84], quantitative structure activity relationship, and protein modelling[85–87], being replaced or supplemented by machine learning. GPCR drug discovery has benefited from this revolution with both generic tools, and new GPCR-specific models and methods [88]. In particular two specific needs of GPCR drug discovery have been addressed: efficacy prediction, and the modelling of unresolved structures in different conformational states.

Efficacy and affinity prediction

The objective of virtual screening is to identify new ligands binding a target receptor. Machine learning models can achieve this task by directly predicting GPCR-ligand association, without however determining the strength of this association, or by predicting the binding affinity of the evaluated molecule for the target. Once a relationship between the molecule and the protein has been established, the efficacy of the pair can be predicted.

Based on these two aspects of ligand-receptor interaction two main categories of machine learning models have been developed: affinity prediction models, and efficacy prediction models.

Affinity prediction Affinity models predict the association between a molecule and a receptor by classifying the pair as interacting or not interacting [89–92], or by predicting the binding affinity of an evaluated molecule for a target receptor[93–96]. Both target specific models, which predict the interaction between molecules and a single receptor, and

universal models, which predict the interaction between a molecule and a selected target receptor, have been developed. Target specific models only relay on ligand descriptors to predict the interaction between the evaluated molecule and the receptor[90, 93, 94, 96], while universal models combine molecular and protein descriptors to predict molecule-receptor association[89, 91, 95]. GPCR_LigandClassify.py, developed by Ahmed et al.[92], represents a middle point between the two approaches, since it uses only ligand descriptors for its predictions, but at the same time it predicts the association of a ligand to a panel of multiple GPCR targets, without determining which receptor is bound by a positively classified molecule.

A special case of affinity prediction is binding selectivity. The relative binding affinity of a molecule between two targets, rather than the absolute binding affinity, is determined using these models. As described before, the task can be tackled by regression, to determine the selectivity ratio [97], or as a classification problem, determining which target the molecule is more likely to bind [90, 94, 98, 99].

Efficacy prediction Due to the complexity of data regarding a ligand's efficacy, predictions are generally tackled as a classification problem. A ligand's efficacy is described as a binary value: agonist or antagonists[100, 101], agonist or not agonist [102], G-protein-biased or β-arrestin-biased[103]. Regardless of the two classes considered for prediction, both target specific and universal models have been developed.

Target-specific models predict the efficacy of a molecule for one GPCR. Models of this type have been trained using both molecular descriptors and binding mode descriptors. Bushdid et al. [102] trained a classifier on molecular descriptors to distinguish between agonists and not agonist of the human odorant receptor OR51E1. The model was tested on a library of 258 compounds, which was filtered to exclude molecules outside of the applicability domain. 32 molecules were described as within the applicability domain, of these 4 were predicted as agonist, and 28 as not agonist. *in vitro* tests confirmed that two of the four selected molecules were agonists, while all tested non-agonists did not have an efficacy for the target. The same method was used by Jabeen and Ranganathan[104] to build a similar model for OR1G1. Sakamuru et al. [105] trained a model on molecular fingerprints based on quantitative high throughput screening data. Six models for three different opioid receptor were trained to predict agonist and not-agonist efficacy, or antagonist and not-antagonist efficacy. While the models performed poorly in cross-validation (balanced accuracies between 0.58 and 0.73), they were able to identify new agonists and antagonists. El-Atawneh and Goldblum[106] trained an ensemble of models for different GPCRs located in the central nervous system by using known agonists or antagonists to define one class, and randomly selected decoys to represent the other class. The ensemble of models could predict the pharmacological fingerprint of a compound for a diverse panel of related targets. Interaction-based models use the interaction fingerprint generated by protein-ligand docking to determine the efficacy of a ligand. This approach was developed by Jimenez-Roses et al. [101] to classify known ligands of ADRB2 as agonists or antagonists. The model was trained using binding modes predicted by docking on an ensemble of active and inactive state structures of the receptor. Using docking-generated binding modes allows to overcome the limit caused by the lack of experimental structures, increasing the number of possible targets. Additionally, by using explainable machine learning the contribution of each interaction towards the predicted efficacy could be identified. The same method was applied by Gossen et al. [107] in a virtual screening campaign to identify new antagonists of AA2AR. While multiple hits were

identified, no functional assay was reported confirming their efficacy.

Universal models predict either the relationship between a specific ligand-receptor pair, or they consider the classification of the ligand compared to a panel of different targets. Oh et al [100] built a prediction pipeline formed by an affinity classifier, separating molecules between ligands and not ligands, and an efficacy classifier, labelling a ligand as agonist or antagonist. The models were trained by labelling ligands as agonist, antagonists, or non-binder irrespective of the receptor towards which they have an efficacy. Sanchez et al. [103] trained BiasNet, a model predicting the signalling bias of a molecule, classifying it as G-protein-biased or β arrestin-biased. Since the model was trained using only known ligands a companion model was trained to flag a compound as a likely to bind a GPCR or not. Both the model of Oh et al. and BiasNet do not consider towards which target the molecule has the predicted efficacy. The use and biochemical interpretability of such models is debatable, since they disregard the promiscuity of GPCR ligands, and their functional selectivity, the same ligand can act as an agonist for one receptor, and as an antagonist for another. Other models instead combine molecular descriptors with protein descriptors, in order to predict the efficacy of a receptor-molecule pair. Kumar et al. [108] have developed GPCR-IPL score, a model able to predict the efficacy of a ligand for any target receptor. Other than receptor and ligand descriptors, the experimental, for the training set, and the docking generated binding mode, for evaluation, were considered. Since the model was trained on data from crystallographic structures, its dataset is small and inherently unbalanced, often leading to poor performances on a test set containing known agonists and antagonists of different receptors (balanced accuracies ~ 0.5). Cai et al. [109] developed DeepREAL, a model predicting both the binding and the efficacy of a molecule for a target, using a language model trained on the multiple sequence alignment of the target receptor, and the evaluated molecule's SMILES. While the model performed well in the proposed test sets and in cross-validation, no information regarding its performances on specific receptor subfamilies, or in a virtual screening-like context were given.

AlphaFold2

As previously stated, structure-based drug design targeting GPCRs has been greatly impaired by the lack of experimental receptor structures. While homology models have been used both in the past[14], and in more recent years [110] as substitutes for experimental structures, they generally under-perform in virtual screening applications compared to experimental structures.

In the last four years the field of protein structure modelling has been revolutionized by the development of deep learning models for protein structure prediction. This revolution was heralded by the success of AlphaFold2 at CASP14[111], and by the subsequent release of the algorithm, code, and models' weights in 2021[85]. AlphaFold2 is an end-to-end prediction pipeline, able to generate a protein structure using only its amino acid sequence as input. The generated structures can reach near experimental accuracy for backbone structure prediction[111], and the method has established itself as the new go-to method for the modelling of unresolved protein structures. Despite the successes of AlphaFold, and other structure prediction tools[87], the modelling of alternative conformations of a protein still remained an unsolved problem. This is especially important for GPCRs, whose conformational landscape is characterized by the presence of two main conformational states. For class A GPCRs, the predicted AlphaFold structures are generally in an inactive state

conformation[112, 113]. To overcome this limit, multiple techniques have been developed to model different conformational state of the same protein, with GPCRs often being selected as proof-of-concept targets [112, 114, 115].

While AlphaFold2 brought a revolution in structure prediction, its impact on drug design has been limited. Multiple benchmarking studies have evaluated the use of AlphaFold structures for docking, measuring their suitability in crossdocking experiments [112, 116–118], or in retrospective virtual screening against decoys [118–123]. The performances of the AlphaFold-predicted structures have been compared to holo-structures, apo-structures, and to automatically generated homology models of the target proteins. AlphaFold structures generally performed better or had comparable performances to homology models, similar performances as apo experimental structures, and clearly underperformed compared to holo experimental structures. Comparisons between AlphaFold generated structure, and the experimental references generally showed a high accuracy in binding site structure prediction, but also small discrepancies concerning the rotameric state of few sidechains, which impair the correct docking of the ligands.

Lyu et al.[124] evaluated the use of AlphaFold structures in a prospective test, since the previous benchmarking experiments might have presented a protein conformation bias. The co-crystalized ligand influences the binding site structure used as a benchmark, while the tested ligands might have been designed to bind the studied conformation of the binding pocket. Two different proteins, one of which was the GPCR 5HT2A, were selected as targets using both an experimental and the AlphaFold structure for virtual screening using docking. Similar hit rates were reported for both structures, and even more interesting, the resolved crystallographic structure of 5HT2A bound to one of the AlphaFold hits was more similar to the in silico model than to other known experimental structures. Despite the success, the study was only limited to two targets, therefore only future research could properly asses the use of AlphaFold structures in drug discovery.

References

[1] Brian K. Shoichet and Brian K. Kobilka. Structure-based drug screening for G-protein-coupled receptors. *Trends in Pharmacological Sciences*, 33(5):268–272, May 2012.

[2] Joel D. A. Tyndall and Radhika Sandilya. GPCR Agonists and Antagonists in the Clinic. *Medicinal Chemistry*, 1(4):405–421, June 2005.

[3] Sripal Bangalore, Franz H. Messerli, John B. Kostis, and Carl J. Pepine. Cardiovascular Protection Using Beta-Blockers. *Journal of the American College of Cardiology*, 50(7):563–572, August 2007.

[4] Charlotte K. Billington, Raymond B. Penn, and Ian P. Hall. $B2$ Agonists. In Clive P. Page and Peter J. Barnes, editors, *Pharmacology and Therapeutics of Asthma and COPD*, volume 237, pages 23–40. Springer International Publishing, Cham, 2016.

[5] Mario Cazzola, Paola Rogliani, and Maria Gabriella Matera. Ultra-LABAs for the treatment of asthma. *Respiratory Medicine*, 156:47–52, September 2019.

[6] Flavio Ballante, Albert J Kooistra, Stefanie Kampen, Chris de Graaf, and Jens Carlsson. Structure-Based Virtual Screening for Ligands of G Protein–Coupled

Receptors: What Can Molecular Docking Do for You? *Pharmacological Reviews*, 73(4):1698–1736, October 2021.

[7] Zoran Rankovic, Tarsis F. Brust, and Laura M. Bohn. Biased agonism: An emerging paradigm in GPCR drug discovery. *Bioorganic & Medicinal Chemistry Letters*, 26(2):241–250, January 2016.

[8] Irwin D. Kuntz, Jeffrey M. Blaney, Stuart J. Oatley, Robert Langridge, and Thomas E. Ferrin. A geometric approach to macromolecule-ligand interactions. *Journal of Molecular Biology*, 161(2):269–288, October 1982.

[9] Luca Pinzi and Giulio Rastelli. Molecular Docking: Shifting Paradigms in Drug Discovery. *International Journal of Molecular Sciences*, 20(18):4331, January 2019.

[10] Douglas B. Kitchen, Hélène Decornez, John R. Furr, and Jürgen Bajorath. Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nature Reviews Drug Discovery*, 3(11):935–949, November 2004.

[11] Jin Li, Ailing Fu, and Le Zhang. An Overview of Scoring Functions Used for Protein–Ligand Interactions in Molecular Docking. *Interdisciplinary Sciences: Computational Life Sciences*, 11(2):320–328, June 2019.

[12] Brian K. Shoichet. Virtual screening of chemical libraries. *Nature*, 432(7019):862–865, December 2004.

[13] Krzysztof Palczewski, Takashi Kumasaka, Tetsuya Hori, Craig A. Behnke, Hiroyuki Motoshima, Brian A. Fox, Isolde Le Trong, David C. Teller, Tetsuji Okada, Ronald E. Stenkamp, Masaki Yamamoto, and Masashi Miyano. Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor. *Science*, 289(5480):739–745, August 2000.

[14] Esther Kellenberger, Jean-Yves Springael, Marc Parmentier, Muriel Hachet-Haas, Jean-Luc Galzi, and Didier Rognan. Identification of Nonpeptide CCR5 Receptor Agonists by Structure-based Virtual Screening. *Journal of Medicinal Chemistry*, 50(6):1294–1303, March 2007.

[15] Outi M. H. Salo, Katri H. Raitio, Juha R. Savinainen, Tapio Nevalainen, Maija Lahtela-Kakkonen, Jarmo T. Laitinen, Tomi Järvinen, and Antti Poso. Virtual Screening of Novel CB2 Ligands Using a Comparative Model of the Human Cannabinoid CB2 Receptor. *Journal of Medicinal Chemistry*, 48(23):7166–7171, November 2005.

[16] Irina G. Tikhonova, Chi Shing Sum, Susanne Neumann, Stanislav Engel, Bruce M. Raaka, Stefano Costanzi, and Marvin C. Gershengorn. Discovery of Novel Agonists and Antagonists of the Free Fatty Acid Receptor 1 (FFAR1) Using Virtual Screening. *Journal of Medicinal Chemistry*, 51(3):625–633, February 2008.

[17] Vadim Cherezov, Daniel M. Rosenbaum, Michael A. Hanson, Søren G. F. Rasmussen, Foon Sun Thian, Tong Sun Kobilka, Hee-Jung Choi, Peter Kuhn, William I. Weis, Brian K. Kobilka, and Raymond C. Stevens. High-Resolution Crystal Structure of an Engineered Human β_2 -Adrenergic G Protein–Coupled Receptor. *Science*, 318(5854):1258–1265, November 2007.

[18] Peter Kolb, Daniel M. Rosenbaum, John J. Irwin, Juan José Fung, Brian K. Kobilka, and Brian K. Shoichet. Structure-based discovery of $B2$-adrenergic receptor ligands. *Proceedings of the National Academy of Sciences*, 106(16):6843–6848, April 2009.

[19] Veli-Pekka Jaakola, Mark T. Griffith, Michael A. Hanson, Vadim Cherezov, Ellen Y. T. Chien, J. Robert Lane, Adriaan P. IJzerman, and Raymond C. Stevens. The 2.6 Angstrom Crystal Structure of a Human A2A Adenosine Receptor Bound to an Antagonist. *Science*, 322(5905):1211–1217, November 2008.

[20] Beili Wu, Ellen Y. T. Chien, Clifford D. Mol, Gustavo Fenalti, Wei Liu, Vsevolod Katritch, Ruben Abagyan, Alexei Brooun, Peter Wells, F. Christopher Bi, Damon J. Hamel, Peter Kuhn, Tracy M. Handel, Vadim Cherezov, and Raymond C. Stevens. Structures of the CXCR4 Chemokine GPCR with Small-Molecule and Cyclic Peptide Antagonists. *Science*, 330(6007):1066–1071, November 2010.

[21] Ellen Y. T. Chien, Wei Liu, Qiang Zhao, Vsevolod Katritch, Gye Won Han, Michael A. Hanson, Lei Shi, Amy Hauck Newman, Jonathan A. Javitch, Vadim Cherezov, and Raymond C. Stevens. Structure of the Human Dopamine D3 Receptor in Complex with a D2/D3 Selective Antagonist. *Science*, 330(6007):1091–1095, November 2010.

[22] Tatsuro Shimamura, Mitsunori Shiroishi, Simone Weyand, Hirokazu Tsujimoto, Graeme Winter, Vsevolod Katritch, Ruben Abagyan, Vadim Cherezov, Wei Liu, Gye Won Han, Takuya Kobayashi, Raymond C. Stevens, and So Iwata. Structure of the human histamine H1 receptor complex with doxepin. *Nature*, 475(7354):65–70, July 2011.

[23] Vsevolod Katritch, Veli-Pekka Jaakola, J. Robert Lane, Judy Lin, Adriaan P. IJzerman, Mark Yeager, Irina Kufareva, Raymond C. Stevens, and Ruben Abagyan. Structure-Based Discovery of Novel Chemotypes for Adenosine A $_{2A}$ Receptor Antagonists. *Journal of Medicinal Chemistry*, 53(4):1799–1809, February 2010.

[24] Jens Carlsson, Lena Yoo, Zhan-Guo Gao, John J. Irwin, Brian K. Shoichet, and Kenneth A. Jacobson. Structure-Based Discovery of A2A Adenosine Receptor Ligands. *Journal of Medicinal Chemistry*, 53(9):3748–3755, May 2010.

[25] Jens Carlsson, Ryan G. Coleman, Vincent Setola, John J. Irwin, Hao Fan, Avner Schlessinger, Andrej Sali, Bryan L. Roth, and Brian K. Shoichet. Ligand discovery from a dopamine D3 receptor homology model and crystal structure. *Nature Chemical Biology*, 7(11):769–778, November 2011.

[26] Michael M. Mysinger, Dahlia R. Weiss, Joshua J. Ziarek, Stéphanie Gravel, Allison K. Doak, Joel Karpiak, Nikolaus Heveker, Brian K. Shoichet, and Brian F. Volkman. Structure-based ligand discovery for the protein–protein interface of chemokine receptor CXCR4. *Proceedings of the National Academy of Sciences*, 109(14):5517–5522, April 2012.

[27] Chris de Graaf, Albert J. Kooistra, Henry F. Vischer, Vsevolod Katritch, Martien Kuijer, Mitsunori Shiroishi, So Iwata, Tatsuro Shimamura, Raymond C. Stevens, Iwan J. P. de Esch, and Rob Leurs. Crystal Structure-Based Virtual Screening for

Fragment-like Ligands of the Human Histamine H1 Receptor. *Journal of Medicinal Chemistry*, 54(23):8195–8206, December 2011.

[28] Andrew C. Kruse, Dahlia R. Weiss, Mario Rossi, Jianxin Hu, Kelly Hu, Katrin Eitel, Peter Gmeiner, Jürgen Wess, Brian K. Kobilka, and Brian K. Shoichet. Muscarinic Receptors as Model Targets and Antitargets for Structure-Based Ligand Discovery. *Molecular Pharmacology*, 84(4):528–540, October 2013.

[29] Ana Negri, Marie-Laure Rives, Michael J. Caspers, Thomas E. Prisinzano, Jonathan A. Javitch, and Marta Filizola. Discovery of a Novel Selective Kappa-Opioid Receptor Agonist Using Crystal Structure-Based Virtual Screening. *Journal of chemical information and modeling*, 53(3):521–526, March 2013.

[30] Chris de Graaf and Didier Rognan. Selective Structure-Based Virtual Screening for Full and Partial Agonists of the $B2$ Adrenergic Receptor. *Journal of Medicinal Chemistry*, 51(16):4978–4985, August 2008.

[31] Santiago Vilar, Giulio Ferino, Sharangdhar S. Phatak, Barkin Berk, Claudio N. Cavasotto, and Stefano Costanzi. Docking-based virtual screening for ligands of G protein-coupled receptors: Not only crystal structures but also in silico models. *Journal of molecular graphics & modelling*, 29(5):614–623, February 2011.

[32] J. Robert Lane, Pavel Chubukov, Wei Liu, Meritxell Canals, Vadim Cherezov, Ruben Abagyan, Raymond C. Stevens, and Vsevolod Katritch. Structure-Based Ligand Discovery Targeting Orthosteric and Allosteric Pockets of Dopamine Receptors. *Molecular Pharmacology*, 84(6):794–807, December 2013.

[33] Zhong Zheng, Xi-Ping Huang, Thomas J. Mangano, Rodger Zou, Xin Chen, Saheem A. Zaidi, Bryan L. Roth, Raymond C. Stevens, and Vsevolod Katritch. Structure-Based Discovery of New Antagonist and Biased Agonist Chemotypes for the Kappa Opioid Receptor. *Journal of Medicinal Chemistry*, 60(7):3070–3081, April 2017.

[34] Anat Levit Kaplan, Danielle N. Confair, Kuglae Kim, Ximena Barros-Álvarez, Ramona M. Rodriguiz, Ying Yang, Oh Sang Kweon, Tao Che, John D. McCorvy, David N. Kamber, James P. Phelan, Luan Carvalho Martins, Vladimir M. Pogorelov, Jeffrey F. DiBerto, Samuel T. Slocum, Xi-Ping Huang, Jain Manish Kumar, Michael J. Robertson, Ouliana Panova, Alpay B. Seven, Autumn Q. Wetsel, William C. Wetsel, John J. Irwin, Georgios Skiniotis, Brian K. Shoichet, Bryan L. Roth, and Jonathan A. Ellman. Bespoke library docking for 5-HT2A receptor agonists with antidepressant activity. *Nature*, 610(7932):582–591, October 2022.

[35] Søren G. F. Rasmussen, Brian T. DeVree, Yaozhong Zou, Andrew C. Kruse, Ka Young Chung, Tong Sun Kobilka, Foon Sun Thian, Pil Seok Chae, Els Pardon, Diane Calinski, Jesper M. Mathiesen, Syed T. A. Shah, Joseph A. Lyons, Martin Caffrey, Samuel H. Gellman, Jan Steyaert, Georgios Skiniotis, William I. Weis, Roger K. Sunahara, and Brian K. Kobilka. Crystal structure of the $B2$ adrenergic receptor–Gs protein complex. *Nature*, 477(7366):549–555, September 2011.

[36] Søren G. F. Rasmussen, Hee-Jung Choi, Juan Jose Fung, Els Pardon, Paola Casarosa, Pil Seok Chae, Brian T. DeVree, Daniel M. Rosenbaum, Foon Sun Thian, Tong Sun

Kobilka, Andreas Schnapp, Ingo Konetzki, Roger K. Sunahara, Samuel H. Gellman, Alexander Pautsch, Jan Steyaert, William I. Weis, and Brian K. Kobilka. Structure of a nanobody-stabilized active state of the $B2$ adrenoceptor. *Nature*, 469(7329):175–180, January 2011.

[37] Dahlia R. Weiss, SeungKirl Ahn, Maria F. Sassano, Andrew Kleist, Xiao Zhu, Ryan Strachan, Bryan L. Roth, Robert J. Lefkowitz, and Brian K. Shoichet. Conformation Guides Molecular Efficacy in Docking Screens of Activated β-2 Adrenergic G Protein Coupled Receptor. *ACS Chemical Biology*, 8(5):1018–1026, May 2013.

[38] Vsevolod Katritch and Ruben Abagyan. GPCR agonist binding revealed by modeling and crystallography. *Trends in Pharmacological Sciences*, 32(11):637–643, November 2011.

[39] Yu Zhou, Jing Ma, Xingyu Lin, Xi-Ping Huang, Kaichun Wu, and Niu Huang. Structure-Based Discovery of Novel and Selective 5-Hydroxytryptamine 2B Receptor Antagonists for the Treatment of Irritable Bowel Syndrome. *Journal of Medicinal Chemistry*, 59(2):707–720, January 2016.

[40] Arman A. Sadybekov, Rebecca L. Brouillette, Egor Marin, Anastasiia V. Sadybekov, Aleksandra Luginina, Anastasiia Gusach, Alexey Mishin, Élie Besserer-Offroy, Jean-Michel Longpré, Valentin Borshchevskiy, Vadim Cherezov, Philippe Sarret, and Vsevolod Katritch. Structure-Based Virtual Screening of Ultra-Large Library Yields Potent Antagonists for a Lipid GPCR. *Biomolecules*, 10(12):1634, December 2020.

[41] Magdalena M. Scharf, Mirjam Zimmermann, Florian Wilhelm, Raimond Stroe, Maria Waldhoer, and Peter Kolb. A Focus on Unusual ECL2 Interactions Yields β_2-Adrenergic Receptor Antagonists with Unprecedented Scaffolds. *ChemMedChem*, 15(10):882–890, May 2020.

[42] Pierre Matricon, Anh TN. Nguyen, Duc Duy Vo, Jo-Anne Baltos, Mariama Jaiteh, Andreas Luttens, Stefanie Kampen, Arthur Christopoulos, Jan Kihlberg, Lauren Therese May, and Jens Carlsson. Structure-based virtual screening discovers potent and selective adenosine A1 receptor antagonists. *European Journal of Medicinal Chemistry*, 257:115419, September 2023.

[43] Leslie Salas-Estrada, Davide Provasi, Xing Qiu, Husnu Ümit Kaniskan, Xi-Ping Huang, Jeffrey F. DiBerto, João Marcelo Lamim Ribeiro, Jian Jin, Bryan L. Roth, and Marta Filizola. De Novo Design of κ-Opioid Receptor Antagonists Using a Generative Deep-Learning Framework. *Journal of Chemical Information and Modeling*, 63(16):5056–5065, August 2023.

[44] Albert J. Kooistra, Henry F. Vischer, Daniel McNaught-Flores, Rob Leurs, Iwan J. P. de Esch, and Chris de Graaf. Function-specific virtual screening for GPCR ligands using a combined scoring method. *Scientific Reports*, 6(1):28288, September 2016.

[45] Magdalena M. Scharf, Moritz Bünemann, Jillian G. Baker, and Peter Kolb. Comparative Docking to Distinct G Protein–Coupled Receptor Conformations Exclusively Yields Ligands with Agonist Efficacy. *Molecular Pharmacology*, 96(6):851–861, December 2019.

[46] Elissa A. Fink, Jun Xu, Harald Hübner, Joao M. Braz, Philipp Seemann, Charlotte Avet, Veronica Craik, Dorothee Weikert, Maximilian F. Schmidt, Chase M. Webb, Nataliya A. Tolmachova, Yurii S. Moroz, Xi-Ping Huang, Chakrapani Kalyanaraman, Stefan Gahbauer, Geng Chen, Zheng Liu, Matthew P. Jacobson, John J. Irwin, Michel Bouvier, Yang Du, Brian K. Shoichet, Allan I. Basbaum, and Peter Gmeiner. Structure-based discovery of nonopioid analgesics acting through the α2A-adrenergic receptor. *Science*, 377(6614):eabn7065, September 2022.

[47] Aashish Manglik, Henry Lin, Dipendra K. Aryal, John D. McCorvy, Daniela Dengler, Gregory Corder, Anat Levit, Ralf C. Kling, Viachaslau Bernat, Harald Hübner, Xi-Ping Huang, Maria F. Sassano, Patrick M. Giguère, Stefan Löber, Da Duan, Grégory Scherrer, Brian K. Kobilka, Peter Gmeiner, Bryan L. Roth, and Brian K. Shoichet. Structure-based discovery of opioid analgesics with reduced side effects. *Nature*, 537(7619):185–190, September 2016.

[48] Sheng Wang, Daniel Wacker, Anat Levit, Tao Che, Robin M. Betz, John D. McCorvy, A. J. Venkatakrishnan, Xi-Ping Huang, Ron O. Dror, Brian K. Shoichet, and Bryan L. Roth. D $_4$ dopamine receptor high-resolution structures enable the discovery of selective agonists. *Science*, 358(6361):381–386, October 2017.

[49] Jiankun Lyu, Sheng Wang, Trent E. Balius, Isha Singh, Anat Levit, Yurii S. Moroz, Matthew J. O'Meara, Tao Che, Enkhjargal Algaa, Kateryna Tolmachova, Andrey A. Tolmachev, Brian K. Shoichet, Bryan L. Roth, and John J. Irwin. Ultra-large library docking for discovering new chemotypes. *Nature*, 566(7743):224–229, February 2019.

[50] Reed M. Stein, Hye Jin Kang, John D. McCorvy, Grant C. Glatfelter, Anthony J. Jones, Tao Che, Samuel Slocum, Xi-Ping Huang, Olena Savych, Yurii S. Moroz, Benjamin Stauch, Linda C. Johansson, Vadim Cherezov, Terry Kenakin, John J. Irwin, Brian K. Shoichet, Bryan L. Roth, and Margarita L. Dubocovich. Virtual discovery of melatonin receptor ligands to modulate circadian rhythms. *Nature*, 579(7800):609–614, March 2020.

[51] David Rodríguez, Zhang-Guo Gao, Steven M. Moss, Kenneth A. Jacobson, and Jens Carlsson. Molecular Docking Screening Using Agonist-Bound GPCR Structures: Probing the A $_{2A}$ Adenosine Receptor. *Journal of Chemical Information and Modeling*, 55(3):550–563, March 2015.

[52] Gilles Marcou and Didier Rognan. Optimizing Fragment and Scaffold Docking by Use of Molecular Interaction Fingerprints. *Journal of Chemical Information and Modeling*, 47(1):195–207, January 2007.

[53] André Fischer, Martin Smieško, Manuel Sellner, and Markus A. Lill. Decision Making in Structure-Based Drug Discovery: Visual Inspection of Docking Results. *Journal of Medicinal Chemistry*, 64(5):2489–2500, March 2021.

[54] Terry Kenakin. Efficacy at g-protein-coupled receptors. *Nature Reviews Drug Discovery*, 1(2):103–110, February 2002.

[55] Márton Vass, Albert J Kooistra, Tina Ritschel, Rob Leurs, Iwan JP de Esch, and Chris de Graaf. Molecular interaction fingerprint approaches for GPCR drug discovery. *Current Opinion in Pharmacology*, 30:59–68, October 2016.

[56] Renata Del Carmine, Caterina Ambrosio, Maria Sbraccia, Susanna Cotecchia, Adriaan P Ijzerman, and Tommaso Costa. Mutations inducing divergent shifts of constitutive activity reveal different modes of binding among catecholamine analogues to the $B2$-adrenergic receptor. *British Journal of Pharmacology*, 135(7):1715–1722, 2002.

[57] Thomas Scior, Andreas Bender, Gary Tresadern, José L. Medina-Franco, Karina Martínez-Mayorga, Thierry Langer, Karina Cuanalo-Contreras, and Dimitris K. Agrafiotis. Recognizing Pitfalls in Virtual Screening: A Critical Review. *Journal of Chemical Information and Modeling*, 52(4):867–881, April 2012.

[58] Gregory L. Warren, C. Webster Andrews, Anna-Maria Capelli, Brian Clarke, Judith LaLonde, Millard H. Lambert, Mika Lindvall, Neysa Nevins, Simon F. Semus, Stefan Senger, Giovanna Tedesco, Ian D. Wall, James M. Woolven, Catherine E. Peishoff, and Martha S. Head. A Critical Assessment of Docking Programs and Scoring Functions. *Journal of Medicinal Chemistry*, 49(20):5912–5931, October 2006.

[59] Stefano Costanzi and Santiago Vilar. In Silico screening for agonists and blockers of the $B2$ adrenergic receptor: Implications of inactive and activated state structures. *Journal of Computational Chemistry*, 33(5):561–572, February 2012.

[60] Tiejun Cheng, Xun Li, Yan Li, Zhihai Liu, and Renxiao Wang. Comparative Assessment of Scoring Functions on a Diverse Test Set. *Journal of Chemical Information and Modeling*, 49(4):1079–1093, April 2009.

[61] Michail Papadourakis, Hryhory Sinenka, Pierre Matricon, Jérôme Hénin, Grace Brannigan, Laura Pérez-Benito, Vineet Pande, Herman van Vlijmen, Chris de Graaf, Francesca Deflorian, Gary Tresadern, Marco Cecchini, and Zoe Cournia. Alchemical Free Energy Calculations on Membrane-Associated Proteins. *Journal of Chemical Theory and Computation*, 19(21):7437–7458, November 2023.

[62] Eelke B. Lenselink, Julien Louvel, Anna F. Forti, Jacobus P. D. van Veldhoven, Henk de Vries, Thea Mulder-Krieger, Fiona M. McRobb, Ana Negri, Joseph Goose, Robert Abel, Herman W. T. van Vlijmen, Lingle Wang, Edward Harder, Woody Sherman, Adriaan P. IJzerman, and Thijs Beuming. Predicting Binding Affinities for GPCR Ligands Using Free-Energy Perturbation. *ACS Omega*, 1(2):293–304, August 2016.

[63] Pierre Matricon, Duc Duy Vo, Zhan-Guo Gao, Jan Kihlberg, Kenneth A. Jacobson, and Jens Carlsson. Fragment-based design of selective GPCR ligands guided by free energy simulations. *Chemical Communications*, 57(92):12305–12308, November 2021.

[64] Willem Jespers, Laura H. Heitman, Adriaan P. IJzerman, Eddy Sotelo, Gerard J. P. van Westen, Johan Åqvist, and Hugo Gutiérrez-de-Terán. Deciphering conformational selectivity in the A2A adenosine G protein-coupled receptor by free energy simulations. *PLOS Computational Biology*, 17(11):e1009152, November 2021.

[65] Martin Vögele, Bin W. Zhang, Jonas Kaindl, and Lingle Wang. Is the Functional Response of a Receptor Determined by the Thermodynamics of Ligand Binding? *Journal of Chemical Theory and Computation*, 19(22):8414–8422, November 2023.

[66] Nicolas Panel, Duc Duy Vo, Nour Aldin Kahlous, Harald Hübner, Stephanie Tiedt, Pierre Matricon, Jody Pacalon, Oliver Fleetwood, Stefanie Kampen, Andreas Luttens, Lucie Delemotte, Jan Kihlberg, Peter Gmeiner, and Jens Carlsson. Design of Drug Efficacy Guided by Free Energy Simulations of the β_2-Adrenoceptor. *Angewandte Chemie*, 135(22):e202218959, May 2023.

[67] Inbar Fish, Anne Stößel, Katrin Eitel, Celine Valant, Sabine Albold, Harald Huebner, Dorothee Möller, Mary J. Clark, Roger K. Sunahara, Arthur Christopoulos, Brian K. Shoichet, and Peter Gmeiner. Structure-Based Design and Discovery of New M_2 Receptor Agonists. *Journal of Medicinal Chemistry*, 60(22):9239–9250, November 2017.

[68] Nilkanth Patel, Xi Ping Huang, Jessica M Grandner, Linda C Johansson, Benjamin Stauch, John D McCorvy, Yongfeng Liu, Bryan Roth, and Vsevolod Katritch. Structure-based discovery of potent and selective melatonin receptor agonists. *eLife*, 9:e53779, March 2020.

[69] Eelke B. Lenselink, Willem Jespers, Herman W. T. van Vlijmen, Adriaan P. IJzerman, and Gerard J. P. van Westen. Interacting with GPCRs: Using Interaction Fingerprints for Virtual Screening. *Journal of Chemical Information and Modeling*, 56(10):2053–2060, October 2016.

[70] Zhan Deng, Claudio Chuaqui, and Juswinder Singh. Structural Interaction Fingerprint (SIFt): A Novel Method for Analyzing Three-Dimensional Protein-Ligand Binding Interactions. *Journal of Medicinal Chemistry*, 47(2):337–344, January 2004.

[71] Cédric Bouysset and Sébastien Fiorucci. ProLIF: A library to encode molecular interactions as fingerprints. *Journal of Cheminformatics*, 13(1):72, September 2021.

[72] Matthew D. Kelly and Ricardo L. Mancera. Expanded Interaction Fingerprint Method for Analyzing Ligand Binding Modes in Docking and Structure-Based Drug Design. *Journal of Chemical Information and Computer Sciences*, 44(6):1942–1951, November 2004.

[73] Chidochangu P. Mpamhanga, Beining Chen, Iain M. McLay, and Peter Willett. Knowledge-Based Interaction Fingerprint Scoring: A Simple Method for Improving the Effectiveness of Fast Scoring Functions. *Journal of Chemical Information and Modeling*, 46(2):686–698, March 2006.

[74] C. Da and D. Kireev. Structural Protein–Ligand Interaction Fingerprints (SPLIF) for Structure-Based Virtual Screening: Method and Benchmark Study. *Journal of Chemical Information and Modeling*, 54(9):2555–2561, September 2014.

[75] Maciej Wójcikowski, Michał Kukiełka, Marta M Stepniewska-Dziubinska, and Pawel Siedlecki. Development of a protein–ligand extended connectivity (PLEC) fingerprint and its application for binding affinity predictions. *Bioinformatics*, 35(8):1334–1341, April 2019.

[76] Alexandre V. Fassio, Laura Shub, Luca Ponzoni, Jessica McKinley, Matthew J. O'Meara, Rafaela S. Ferreira, Michael J. Keiser, and Raquel C. de Melo Minardi. Prioritizing Virtual Screening with Interpretable Interaction Fingerprints. *Journal of Chemical Information and Modeling*, 62(18):4300–4318, September 2022.

[77] Anita Rácz, Dávid Bajusz, and Károly Héberger. Life beyond the Tanimoto coefficient: Similarity measures for interaction fingerprints. *Journal of Cheminformatics*, 10(1):48, October 2018.

[78] Jérémy Desaphy, Eric Raimbaud, Pierre Ducrot, and Didier Rognan. Encoding Protein–Ligand Interaction Patterns in Fingerprints and Graphs. *Journal of Chemical Information and Modeling*, 53(3):623–637, March 2013.

[79] Albert J. Kooistra, Rob Leurs, Iwan J. P. de Esch, and Chris de Graaf. Structure-Based Prediction of G-Protein-Coupled Receptor Ligand Function: A β-Adrenoceptor Case Study. *Journal of Chemical Information and Modeling*, 55(5):1045–1061, May 2015.

[80] I. Adlere, S. Sun, A. Zarca, L. Roumen, M. Gozelle, C. Perpiñá Viciano, B. Caspar, M. Arimont, J. P. Bebelman, S. J. Briddon, C. Hoffmann, S. J. Hill, M. J. Smit, H. F. Vischer, M. Wijtmans, C. de Graaf, I. J. P. de Esch, and R. Leurs. Structure-based exploration and pharmacological evaluation of N-substituted piperidin-4-yl-methanamine CXCR4 chemokine receptor antagonists. *European Journal of Medicinal Chemistry*, 162:631–649, January 2019.

[81] Xuhan Liu, Kai Ye, Herman W. T. van Vlijmen, Adriaan P. IJzerman, and Gerard J. P. van Westen. An exploration strategy improves the diversity of de novo ligands using deep reinforcement learning: A case for the adenosine A2A receptor. *Journal of Cheminformatics*, 11(1):35, May 2019.

[82] Xuhan Liu, Kai Ye, Herman W. T. van Vlijmen, Michael T. M. Emmerich, Adriaan P. IJzerman, and Gerard J. P. van Westen. DrugEx v2: De novo design of drug molecules by Pareto-based multi-objective reinforcement learning in polypharmacology. *Journal of Cheminformatics*, 13(1):85, November 2021.

[83] Morgan Thomas, Robert T. Smith, Noel M. O'Boyle, Chris de Graaf, and Andreas Bender. Comparison of structure- and ligand-based scoring functions for deep generative models: A GPCR case study. *Journal of Cheminformatics*, 13(1):39, May 2021.

[84] Xuhan Liu, Kai Ye, Herman W. T. van Vlijmen, Adriaan P. IJzerman, and Gerard J. P. van Westen. DrugEx v3: Scaffold-constrained drug design with graph transformer-based reinforcement learning. *Journal of Cheminformatics*, 15(1):24, February 2023.

[85] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, August 2021.

[86] Richard Evans, Michael O'Neill, Alexander Pritzel, Natasha Antropova, Andrew Senior, Tim Green, Augustin Žídek, Russ Bates, Sam Blackwell, Jason Yim, Olaf Ronneberger, Sebastian Bodenstein, Michal Zielinski, Alex Bridgland, Anna Potapenko, Andrew Cowie, Kathryn Tunyasuvunakool, Rishub Jain, Ellen Clancy, Pushmeet Kohli, John Jumper, and Demis Hassabis. Protein complex prediction with AlphaFold-Multimer, March 2022.

[87] Rohith Krishna, Jue Wang, Woody Ahern, Pascal Sturmfels, Preetham Venkatesh, Indrek Kalvet, Gyu Rie Lee, Felix S. Morey-Burrows, Ivan Anishchenko, Ian R. Humphreys, Ryan McHugh, Dionne Vafeados, Xinting Li, George A. Sutherland, Andrew Hitchcock, C. Neil Hunter, Alex Kang, Evans Brackenbrough, Asim K. Bera, Minkyung Baek, Frank DiMaio, and David Baker. Generalized biomolecular modeling and design with RoseTTAFold All-Atom. *Science*, 0(0):eadl2528, March 2024.

[88] Wei Chen, Chi Song, Liang Leng, Sanyin Zhang, and Shilin Chen. The Application of Artificial Intelligence Accelerates G Protein-Coupled Receptor Ligand Discovery. *Engineering*, 32:18–28, January 2024.

[89] Xingyue Gu, Junkai Liu, Yue Yu, Pengfeng Xiao, and Yijie Ding. MFD–GDrug: Multimodal feature fusion-based deep learning for GPCR–drug interaction prediction. *Methods*, 223:75–82, March 2024.

[90] Krzysztof Rataj, Ádám Andor Kelemen, José Brea, María Isabel Loza, Andrzej J. Bojarski, and György Miklós Keserű. Fingerprint-Based Machine Learning Approach to Identify Potent and Selective 5-HT2BR Ligands. *Molecules*, 23(5):1137, May 2018.

[91] Sangmin Seo, Jonghwan Choi, Soon Kil Ahn, Kil Won Kim, Jaekwang Kim, Jaehyuck Choi, Jinho Kim, and Jaegyoon Ahn. Prediction of GPCR-Ligand Binding Using Machine Learning Algorithms. *Computational and Mathematical Methods in Medicine*, 2018(1):6565241, 2018.

[92] Marawan Ahmed, Horia Jalily Hasani, Subha Kalyaanamoorthy, and Khaled Barakat. GPCR_LigandClassify.py; a rigorous machine learning classifier for GPCR targeting compounds. *Scientific Reports*, 11(1):9510, May 2021.

[93] Jiansheng Wu, Ben Liu, Wallace K B Chan, Weijian Wu, Tao Pang, Haifeng Hu, Shancheng Yan, Xiaoyan Ke, and Yang Zhang. Precise modelling and interpretation of bioactivities of ligands targeting G protein-coupled receptors. *Bioinformatics*, 35(14):i324–i332, July 2019.

[94] Xiang S. Wang, Hao Tang, Alexander Golbraikh, and Alexander Tropsha. Combinatorial QSAR Modeling of Specificity and Subtype Selectivity of Ligands Binding to Serotonin Receptors 5HT1E and 5HT1F. *Journal of Chemical Information and Modeling*, 48(5):997–1013, May 2008.

[95] Haiping Zhang, Hongjie Fan, Jixia Wang, Tao Hou, Konda Mani Saravanan, Wei Xia, Hei Wun Kan, Junxin Li, John Z H Zhang, Xinmiao Liang, and Yang Chen. Revolutionizing GPCR–ligand predictions: DeepGPCR with experimental validation for high-precision drug discovery. *Briefings in Bioinformatics*, 25(4):bbae281, July 2024.

[96] Jiansheng Wu, Qiuming Zhang, Weijian Wu, Tao Pang, Haifeng Hu, Wallace K B Chan, Xiaoyan Ke, and Yang Zhang. WDL-RF: Predicting bioactivities of ligand molecules acting with G protein-coupled receptors by combining weighted deep learning and random forest. *Bioinformatics*, 34(13):2271–2282, July 2018.

[97] Song-Bing He, Ben Hu, Zheng-Kun Kuang, Dong Wang, and De-Xin Kong. Predicting Subtype Selectivity for Adenosine Receptor Ligands with Three-Dimensional Biologically Relevant Spectrum (BRS-3D). *Scientific Reports*, 6(1):36595, November 2016.

[98] Rafał Kurczab, Vittorio Canale, Paweł Zajdel, and Andrzej J. Bojarski. An Algorithm to Identify Target-Selective Ligands – A Case Study of 5-HT7/5-HT1A Receptor Selectivity. *PLOS ONE*, 11(6):e0156986, June 2016.

[99] Chao Ma, Lirong Wang, Peng Yang, Kyaw Z. Myint, and Xiang-Qun Xie. LiCABEDS II. Modeling of Ligand Selectivity for G-Protein-Coupled Cannabinoid Receptors. *Journal of Chemical Information and Modeling*, 53(1):11–26, January 2013.

[100] Jooseong Oh, Hyi-thaek Ceong, Dokyun Na, and Chungoo Park. A machine learning model for classifying G-protein-coupled receptors as agonists or antagonists. *BMC Bioinformatics*, 23(S9):346, August 2022.

[101] Mireia Jiménez-Rosés, Bradley Angus Morgan, Maria Jimenez Sigstad, Thuy Duong Zoe Tran, Rohini Srivastava, Asuman Bunsuz, Leire Borrega-Román, Pattarin Hompluem, Sean A. Cullum, Clare R. Harwood, Eline J. Koers, David A. Sykes, Iain B. Styles, and Dmitry B. Veprintsev. Combined docking and machine learning identify key molecular determinants of ligand pharmacological activity on $B2$ adrenoceptor. *Pharmacology Research & Perspectives*, 10(5):e00994, 2022.

[102] C. Bushdid, C. A. de March, S. Fiorucci, H. Matsunami, and J. Golebiowski. Agonists of G-Protein-Coupled Odorant Receptors Are Predicted from Chemical Features. *The Journal of Physical Chemistry Letters*, 9(9):2235–2240, May 2018.

[103] Jason E. Sanchez, Govinda B. KC, Julian Franco, William J. Allen, Jesus David Garcia, and Suman Sirimulla. BiasNet: A Model to Predict Ligand Bias Toward GPCR Signaling. *Journal of Chemical Information and Modeling*, 61(9):4190–4199, September 2021.

[104] Amara Jabeen and Shoba Ranganathan. Applications of machine learning in GPCR bioactive ligand discovery. *Current Opinion in Structural Biology*, 55:66–76, April 2019.

[105] Srilatha Sakamuru, Jinghua Zhao, Menghang Xia, Huixiao Hong, Anton Simeonov, Iosif Vaisman, and Ruili Huang. Predictive Models to Identify Small Molecule Activators and Inhibitors of Opioid Receptors. *Journal of Chemical Information and Modeling*, 61(6):2675–2685, June 2021.

[106] Shayma El-Atawneh and Amiram Goldblum. Activity Models of Key GPCR Families in the Central Nervous System: A Tool for Many Purposes. *Journal of Chemical Information and Modeling*, 63(11):3248–3262, June 2023.

[107] Jonas Goßen, Rui Pedro Ribeiro, Dirk Bier, Bernd Neumaier, Paolo Carloni, Alejandro Giorgetti, and Giulia Rossetti. AI-based identification of therapeutic agents targeting GPCRs: Introducing ligand type classifiers and systems biology. *Chemical Science*, 14(32):8651–8661, August 2023.

[108] Surendra Kumar, Mahesh K Teli, and Mi-hyun Kim. GPCR-IPL score: Multilevel featurization of GPCR–ligand interaction patterns and prediction of ligand functions from selectivity to biased activation. *Briefings in Bioinformatics*, 25(2):bbae105, March 2024.

[109] Tian Cai, Kyra Alyssa Abbu, Yang Liu, and Lei Xie. DeepREAL: A deep learning powered multi-scale modeling framework for predicting out-of-distribution ligand-induced GPCR activity. *Bioinformatics*, 38(9):2561–2570, April 2022.

[110] Alessandro Nicoli, Verena Weber, Carlotta Bon, Alexandra Steuer, Stefano Gustincich, Raul R. Gainetdinov, Roman Lang, Stefano Espinoza, and Antonella Di Pizio. Structure-Based Discovery of Mouse Trace Amine-Associated Receptor 5 Antagonists. *Journal of Chemical Information and Modeling*, 63(21):6667–6680, November 2023.

[111] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Applying and improving AlphaFold at CASP14. *Proteins: Structure, Function, and Bioinformatics*, 89(12):1711–1721, 2021.

[112] Lim Heo and Michael Feig. Multi-state modeling of G-protein coupled receptors at experimental accuracy. *Proteins: Structure, Function, and Bioinformatics*, 90(11):1873–1885, 2022.

[113] Diego del Alamo, Davide Sala, Hassane S Mchaourab, and Jens Meiler. Sampling alternative conformational states of transporters and receptors with AlphaFold2. *eLife*, 11:e75751, March 2022.

[114] Davide Sala, Peter W. Hildebrand, and Jens Meiler. Biasing AlphaFold2 to predict GPCRs and kinases with user-defined functional or structural properties. *Frontiers in Molecular Biosciences*, 10, 2023.

[115] Richard A. Stein and Hassane S. Mchaourab. SPEACH_AF: Sampling protein ensembles and conformational heterogeneity with Alphafold2. *PLOS Computational Biology*, 18(8):e1010483, August 2022.

[116] Matthew Holcomb, Ya-Ting Chang, David S. Goodsell, and Stefano Forli. Evaluation of AlphaFold2 structures as docking targets. *Protein Science*, 32(1):e4530, 2023.

[117] Masha Karelina, Joseph J Noh, and Ron O Dror. How accurately can one predict drug binding modes using AlphaFold models? *eLife*, 12:RP89386, December 2023.

[118] Sumin Lee, Seeun Kim, Gyu Rie Lee, Sohee Kwon, Hyeonuk Woo, Chaok Seok, and Hahnbeom Park. Evaluating GPCR modeling and docking strategies in the era of deep learning-based protein structure prediction. *Computational and Structural Biotechnology Journal*, 21:158–167, January 2023.

[119] Valeria Scardino, Juan I. Di Filippo, and Claudio N. Cavasotto. How good are AlphaFold models for docking-based virtual screening? *iScience*, 26(1), January 2023.

[120] Anna M. Díaz-Rovira, Helena Martín, Thijs Beuming, Lucía Díaz, Victor Guallar, and Soumya S. Ray. Are Deep Learning Structural Models Sufficiently Accurate for Virtual Screening? Application of Docking Algorithms to AlphaFold2 Predicted Structures. *Journal of Chemical Information and Modeling*, 63(6):1668–1674, March 2023.

[121] Yuqi Zhang, Marton Vass, Da Shi, Esam Abualrous, Jennifer M. Chambers, Nikita Chopra, Christopher Higgs, Koushik Kasavajhala, Hubert Li, Prajwal Nandekar, Hideyuki Sato, Edward B. Miller, Matthew P. Repasky, and Steven V. Jerome. Benchmarking Refined and Unrefined AlphaFold2 Structures for Hit Discovery. *Journal of Chemical Information and Modeling*, 63(6):1656–1667, March 2023.

[122] Shukai Gu, Yuwei Yang, Yihao Zhao, Jiayue Qiu, Xiaorui Wang, Henry Hoi Yee Tong, Liwei Liu, Xiaozhe Wan, Huanxiang Liu, Tingjun Hou, and Yu Kang. Evaluation of AlphaFold2 Structures for Hit Identification across Multiple Scenarios. *Journal of Chemical Information and Modeling*, April 2024.

[123] Christian Kersten, Steven Clower, and Fabian Barthels. Hic Sunt Dracones: Molecular Docking in Uncharted Territories with Structures from AlphaFold2 and RoseTTAfold. *Journal of Chemical Information and Modeling*, 63(7):2218–2225, April 2023.

[124] Jiankun Lyu, Nicholas Kapolka, Ryan Gumpper, Assaf Alon, Liang Wang, Manish K. Jain, Ximena Barros-Álvarez, Kensuke Sakamoto, Yoojoong Kim, Jeffrey DiBerto, Kuglae Kim, Isabella S. Glenn, Tia A. Tummino, Sijie Huang, John J. Irwin, Olga O. Tarkhanova, Yurii Moroz, Georgios Skiniotis, Andrew C. Kruse, Brian K. Shoichet, and Bryan L. Roth. AlphaFold2 structures guide prospective ligand discovery. *Science*, 384(6702):eadn6354, May 2024.

1.3 Objectives and limits

In this thesis different aspects of drug discovery on G-protein coupled receptors (GPCRs) will be tackled, with a particular focus on the development of methods to selectively discover new ligands with the desired efficacy.

"Classical" computational techniques (docking, interaction similarity, molecular dynamics) were combined with different machine learning methods to predict the properties of known ligands, to evaluate the quality of docking poses in virtual screening, and to model different conformational states of a target protein. The different known and newly developed methods were tested to determine their application in drug discovery at different steps of the pipeline, going from preliminary studies to the fine-tuning of a hit's pharmacological properties.

The regulation of intracellular signalling through ligand binding is a complex and multifaceted problem, which could offer great opportunities for the development of new drugs able to selectively regulate specific signalling pathways. Unfortunately, due to the approximate nature of the investigated computational tools, and the lack of large and well curated experimental datasets, many aspects of the problem were approximated to ensure their tractability. These approximations mainly concern the definition of different conformational states of the receptor, and the definition of ligand efficacy.

As described in Section 1.1, the study of signalling regulation requires to consider multiple axis to obtain a complete view of the problem. Different biological properties and signalling pathways are regulated in an independent manner, and each ligand presents its own characteristic efficacy fingerprint describing the response elicited upon binding. At the same time, the efficacy and the maximal agonist response do not represent a simple binary system with activation or inhibition of a specific signalling pathway, but rather a continuum of efficacies (full agonist, partial agonist, antagonist, and inverse agonist) exists for orthosteric ligands. This multiplicity of signalling regulations cannot be treated in a quantitative way, since it would limit any analysis to small datasets of molecules tested using the same assay in the same experimental conditions. In order to perform a broad comparison using multiple references, a ligand's efficacy was modelled as a binary label, with known molecules being classified as either agonist (full agonist and partial agonist) or antagonist (antagonist or inverse agonist). Labelling was based on a ligand's efficacy towards G-protein signalling, without considering any possible subtype bias or alternative signalling pathway. Allosteric ligands were not considered for analysis, despite the great therapeutic interest in their development, due to a lack of data regarding their binding site. This is especially relevant for ligands binding at the interface between the transmembrane domains and the cellular membrane, since their modelling would require additional considerations compared to orthosteric ligands.

GPCRs can assume a variety of conformations, which are generally classified between an active and an inactive state, with few well characterized intermediate states. While the true active state of the receptor can only be observed for the complex formed by the receptor, an agonist, and a bound G-protein, multiple experimental structures bound to stabilizing nanobodies and orthosteric agonists are also described as in the active state. In this work the conformational state of a structure was determined based on the labelling used in GPCRdb for experimental structure, or by measuring the distance between pairs of conserved residues located at the intracellular side of TM2-TM6 and TM3-TM7. More subtle differences between activation states, like the rearrangements of microswitches, or the role of the ionic lock, were

not studied due to the lack of experimental reference structures in both active and inactive state, or because their analysis was outside of the scope of the study.

Chapter 2

Efficacy driven drug discovery using docking and protein ligand interactions

The design of new ligands targeting members of the G-protein-coupled receptor family requires to take into account both the affinity of the molecule for the receptor, as well as the biological response triggered by ligand binding. Structure-based drug design using protein-ligand docking has represented one of the most important approaches for drug discovery. Ligands with a desired efficacy profile have been designed by exploiting the differences in binding site structure between the active and inactive state conformation of the receptor, as well as differences between the binding mode of agonists and antagonists.

The first section of the chapter is dedicated to an analysis of the realtionship between structure and efficacy, concerning both ligands and receptors. Known agonists and antagonists were compared to identify differences in their molecular structure or physicochemical properties which could justify their differences in efficacy. At the same time the structure of agonist-bound and antagonist-bound conformations of seven representative GPCRs were compared to identify structural rearrangements regulating the selectivity of each conformation towards ligands with a specific efficacy.

The second section is dedicated to a benchmark study, comparing different methods for high-throughput hit discovery. Both methods based on conformational selection and on interaction similarity were used to screen a test library containing known agonists, antagonists, and property matched decoys. The analysis of the results gave an insight on the strengths and weaknesses of the investigated methods, as well as on the delicate balance between conformational selection and interaction similarity in efficacy-driven drug discovery.

2.1 Analysis of efficacy specific differences in ligand and binding site structures

Abstract

Ligands of G-protein coupled receptors can be divided in two categories depending on the biological response elicited upon binding. Agonists preferentially bind active state conformations of a receptor, and trigger intracellular signalling, while antagonists preferentially bind inactive state conformations. The structure and physicochemical properties of both agonists and antagonists of seven well characterized receptors were analysed to identify potential links between a ligand structure and its efficacy. An analysis of eighty-six experimental structures of the seven receptors bound to small molecule ligands were used to determine the main conformational differences between the binding pockets in the two activation states. Multiple differences in both ligand structure, and receptor conformation were identified, which could be exploited to guide drug discovery.

2.1.1 Introduction

G-protein coupled receptors (GPCRs) act as mediators between the cell and the environment by regulating the intracellular signalling based on environmental stimuli. Generally, the binding of a ligand in the extracellular cavity regulates the signalling of the target receptor by influencing its ability to bind G-protein. Ligands facilitating the coupling to G-protein, or other signalling proteins, are called agonists. Ligands binding the receptor without changing its affinity for G-protein instead are called antagonists. Agonists and antagonists are known to bind different conformation of the same receptor. These differences in affinity could be linked to differences in both ligand shape and properties.

In this section agonist and antagonists ligands, and agonist and antagonist bound structures of seven representative GPCRs were compared to identify differences and similarities between the two groups. Ligand analysis focused on the structure of small molecule ligands, and on their physicochemical properties. Specific features relating structure to function were identified using different computational and chemoinformatics tools. Machine learning models were trained on ligand data to identify specific structure activity relationships, as well as to evaluate if ligand based models could be used to predict the efficacy of new ligands. The analysis of the receptors structures instead focused on the differences between different binding site conformations, identifying both global and local differences related to the binding of agonists and antagonists. The effect of these rearrangements on the quality of the poses generated by protein-ligand docking was evaluated by re-docking and cross-docking experiments.

Both analysis identified specific structural determinants, both from the ligand and receptor side, which could be used to guide the development of new potential agonists or antagonists.

2.1.2 Results

Dataset description

From GPCRdb (version released the 19th June 2023)[1][2] 31 class A GPCRs with at least one active and one inactive state structure co-crystallized with a small molecule ligand bound

Table 2.1.1: Number of experimental structures of the seven selected GPCR targets, and their distributions between active state structures, agonist-bound active state structures, and antaognist bound inactive state structures.

Target	Structures	Active-agonist	Inactive-antagonist	Inactive-agonist
5HT2A	14	2	3	9
AA2AR	22	8	13	1
ACM2	8	2	6	-
ADRB2	29	11	18	-
CNR1	7	5	2	-
DRD2	4	2	2	-
OPRK	2	1	1	-
Total	86	31	45	10

to the extracellular binding pocket were extracted. From the same database 58 GPCR targets had at least one agonist and one antagonist ligand reported as a drug or in clinical trials [3]. Only 22 receptors were present in both lists, with most of them being part of well-studied families like serotonin receptors (5HT1B, 5HT2A, 5HT2C), adenosine receptors (AA1R, AA2AR), muscarinic receptors (ACM1, ACM2, ACM4), adrenergic receptors (ADA2A, ADRB1, ADRB2), cannabinoid receptors (CNR1, CNR2), dopamine receptors (DRD2, DRD3), and opioid receptors (OPRK, OPRD). Receptors not part of these families either had few reported ligands (CCKAR, LT4R1, PE2R4), or a large efficacy imbalance between them (HRH1, S1PR1). A representative target from each subfamily was selected based on data availability: 5-hydroxytryptamine receptor 2A (5HT2A), adenosine A2A receptor (AA2AR), muscarinic receptor M2 (ACM2), β2-adrenergic receptor (ADRB2), cannabinoid receptor 1 (CNR1), dopamine receptor 2 (DRD2), and κ opioid receptor (OPRK).

Receptor structures were further filtered to exclude structures with covalently bound ligands, bound allosteric ligands, or complexes with water mediated interactions. In total 86 experimental structures were selected as representative of the seven targets (Tables 2.1.1, A.1.1, A.1.2). Inactive-state structures bound to an antagonist ligand represent half of the dataset (52 %), followed by active state-structures bound to an agonist (36 %), and by inactive-state agonist bound structures, which were observed only for 5HT2A and AA2AR (12 %). ADRB2 is the most represented receptor with 29 structures, and OPRK the least with only two structures. The 86 structures describe the binding mode of 55 unique ligands, with 14 ligands being co-crystalized multiple times bound to the same receptor. No ligand was co-crystalized bound to different receptors.

By combining the list of known drugs and clinical candidates with the co-crystalized ligands, the ligand dataset contains 323 entries, representing 278 unique ligands (Table 2.1.2). The 39 non-selective ligands bind to different aminergic receptors, generally 5HT2A and DRD2. Antagonists represent around two thirds of all entries, and they represent the majority of ligands for 5HT2A, AA2AR, ACM2, CNR1, and DRD2, while ADRB2 and OPRK present a larger number of agonists.

Table 2.1.2: Total number of ligands, number of agonists, and number of antagonists for the seven selected GPCR targets.

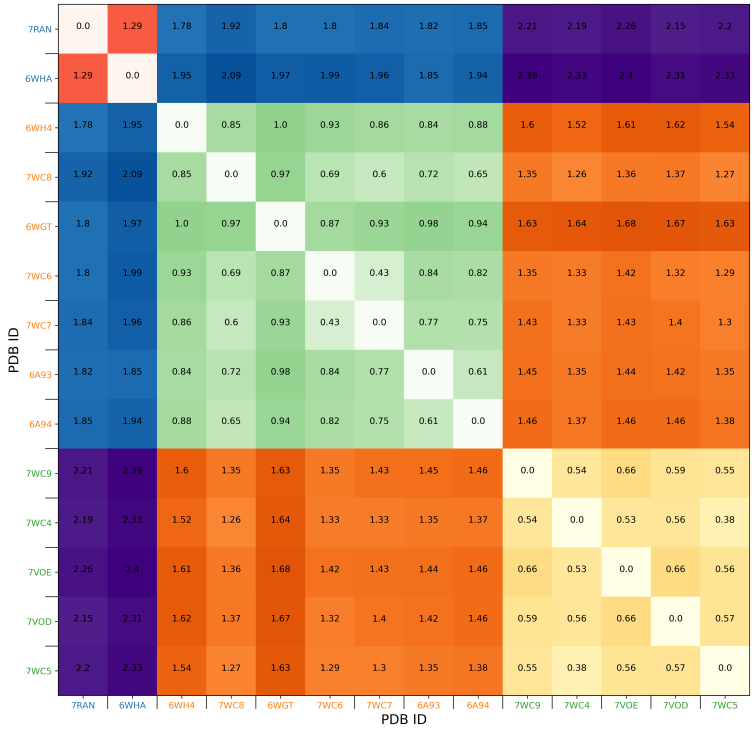
Target	Ligands	Agonists	Antagonists	% Agonists
5HT2A	69	10	59	17
AA2AR	36	10	26	28
ACM2	43	5	38	12
ADRB2	52	32	20	62
CNR1	15	6	9	40
DRD2	77	19	58	27
OPRK	31	23	8	74
Total	323	105	218	

Binding site analysis

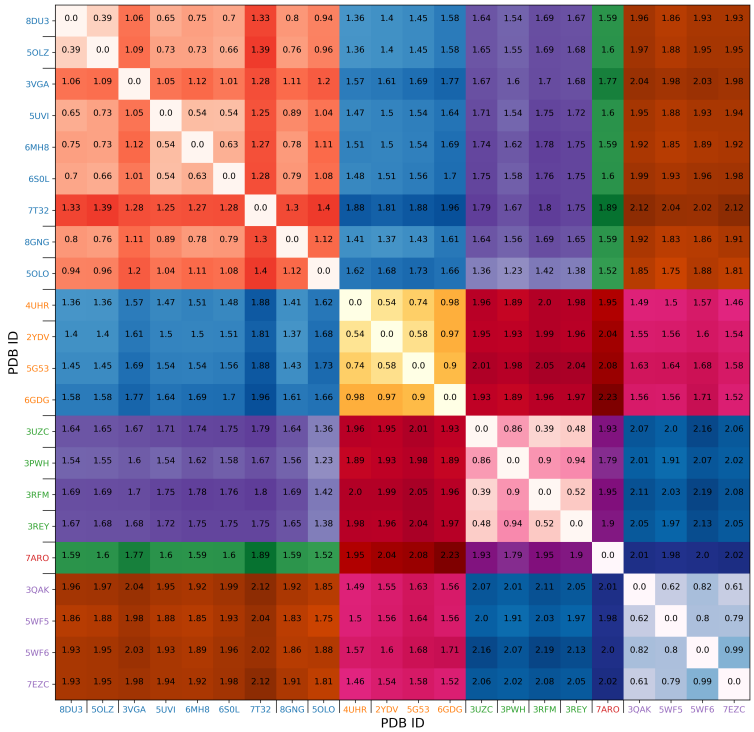
Binding sites structures were clustered using the root mean squared deviation (RMSD) of all non-hydrogen atoms of residues interacting with at least one ligand in any crystallographic structure (Figure A.1.1A-A.1.7A). Between two and five distinct conformations were identified for each target, with most clusters being homogeneous in activation state(Figure A.1.1B-A.1.6B), except ADRB2 (Figure A.1.4B). Multiple conformations within the same activation state were observed for 5HT2A and AA2AR (Figure 2.1.1). Agonist bound inactive state structures of 5HT2A were split between a homogeneous cluster, and a heterogeneous cluster formed with antagonist bound inactive state structures(Figure 2.1.1A). For AA2AR instead five different conformational states were identified: two agonist bound active state conformations, two antagonist bound inactive state conformations, and one agonist bound inactive state conformation(Figure 2.1.1B).

The analysis of the binding site structures using RMSD gives an average overview of their structural differences, however ligand binding can be influenced by small rearrangements of a single sidechain. The per-residue RMSD value for all structure was measured, and the local similarity relationships were analysed using the graph-based redundancy removal method developed by Mohammadi et al.[4]. Structures were represented as nodes in a graph, and an edge was established between two structures if they were identical. Two structures were defined as identical if the RMSD of each residue forming the binding site was below 2 Å. The resulting graphs roughly reproduced the results from clustering (Figure A.1.1C-A.1.6C). Thanks to its more fine-grained comparison, graph analysis could detect unique conformations which were not identified using RMSD clustering. For example, one inactive state structure of ACM2 (PDB ID: 5ZKB) presents a unique conformation of the sidechain of Tyr426, while the conformation of the binding pocket on average matches other inactive state structures (RMSD $\sim$ 1 Å) (Figure 2.1.2). This difference was caused by the binding of the selective antagonist AF-DX 384, whose larger molecular volume is not compatible with the other inactive state binding site conformation. Similar local differences were observed for CNR1 (Figure A.1.5C) and DRD2 (Figure A.1.6C). While useful, graph analysis also identified structural differences not relevant for ligand binding, as observed for AA2AR structures bound to UK-432097 (Figure A.1.2C). Of the four structures only two were considered as identical, while the other two were defined as unique due to rearrangements of residues His264

Figure 2.1.1: Representation of the binding site clustering of 5HT2A (A), and AA2AR (B). Different clusters are highlighted by the colour of the labels, and by the color used in the plot. (A) The first cluster represent the active state structures, the second represents agonists and antagonist bound inactive state structures, ans the third represent antagonist bound inactive state structures. (B) The first and third cluster represent two different inactive state conformations of the binding pocket. The second and fourth cluster instead represent two different conformations of the active state binding pocket.



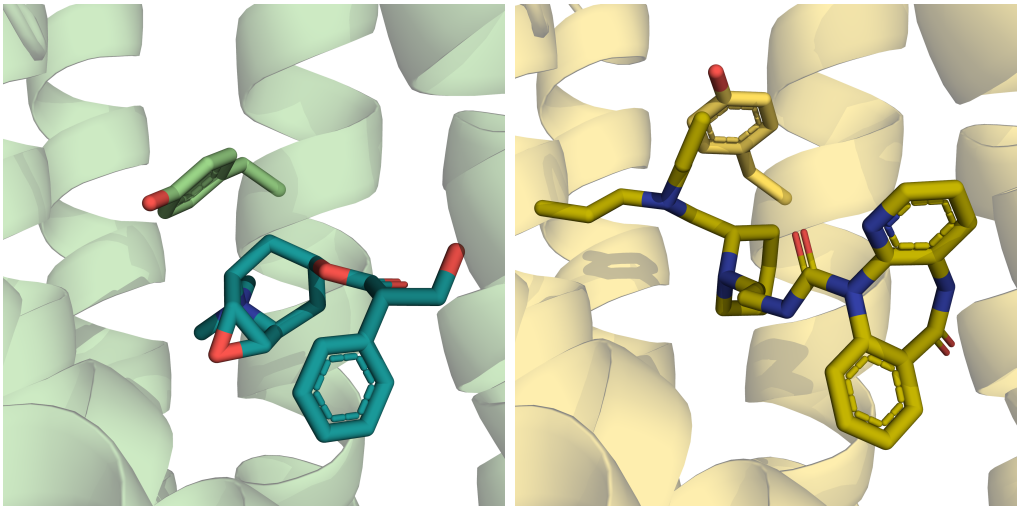
(A) 5HT2A



(B) AA2AR

and Gln89, which do not directly influence ligand binding.

Figure 2.1.2: Alternative conformations of the Tyr426 sidechain in inactive state ACM2 structures identified by graph analysis. (A) ACM2 bound to the antagonist N-methyl scopolamine (PDB ID: 5ZK8). (B) ACM2 bound to the selective antagonist AF-DX 384 (PDB ID: 5ZKB).



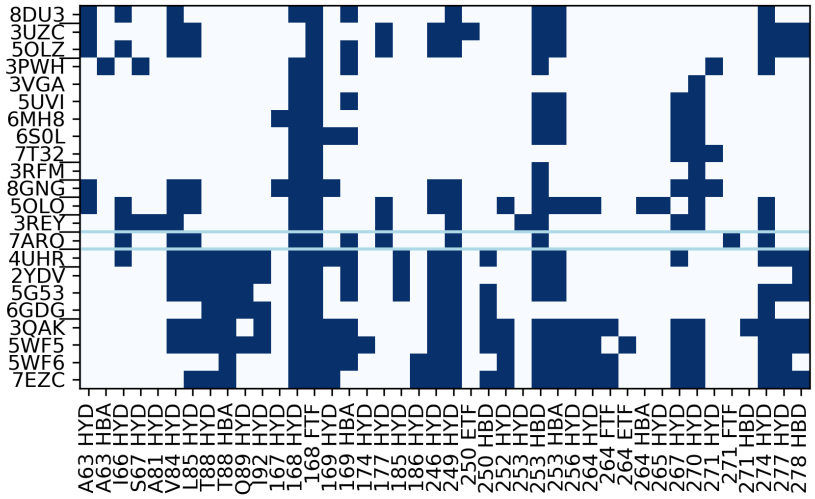
(A) **(B)**

Binding mode analysis

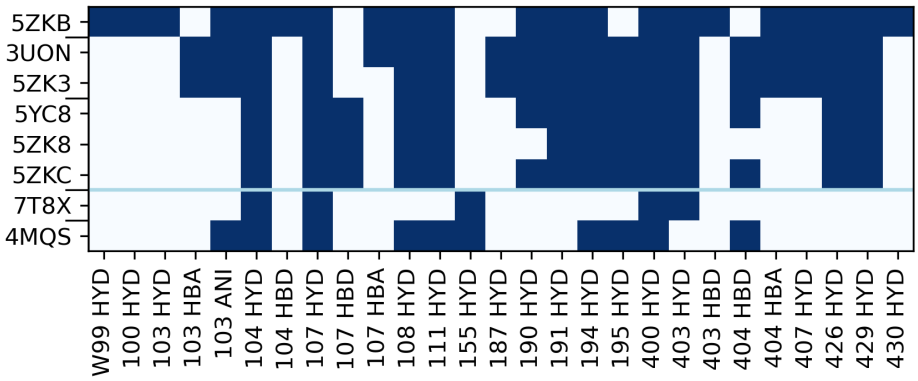
Multiple shared and efficacy specific interactions were identified by comparison of interaction fingerprints (IFP)[5] extracted from the experimental structures. A conserved ionic bond with $Asp^{3.32}$ (Ballesteros-Weinstein notation [6]) was observed in almost all aminergic receptors structures (Asp115 5HT2A, Asp103 ACM2, Asp113 ADRB2, Asp114 DRD2), regardless of the efficacy of the bound ligand. Certain hydrogen bonds formed by ligands with AA2AR, ACM2, and ADRB2 could be associated with an agonist or antagonist efficacy. AA2AR agonists interact with $Tyr88^{3.36}$ and $His250^{6.52}$ (Figure 2.1.3A), while ADRB2 agonists interact with $Ser207^{5.46}$(Figure 2.1.4). Only antagonists of ACM2 formed a hydrogen bond with $Ser107^{3.36}$(Figure 2.1.3B). Despite being non-directional, multiple efficacy specific or conformation specific hydrophobic contacts were identified. Contacts between a ligand and $Leu228^{45.51}$, and $Leu229^{45.52}$ of 5HT2A were observed only in inactive state structures (Figure 2.1.5A). In ACM2 structures only antagonists formed contacts with $Tyr426^{7.39}$ and $Cys429^{7.42}$(Figure 2.1.3B). Agonists of CNR1 were observed to selectively form contacts with $Ser173^{2.60}$(Figure 2.1.5B). For DRD2 instead both agonists and antagonists formed exclusive interactions, with the former interacting with $Val190^{5.39}$ and $Ser193^{5.42}$, and the latter with $Phe198^{5.47}$ and $Phe382^{6.44}$ (Figure 2.1.5C).

Other than specific interactions, the entire binding modes were compared using the Tanimoto similarity between IFPs (Figure A.1.8-A.1.14). A clear difference in binding mode between agonist and antagonist was observed for all targets except 5HT2A (Figure A.1.8B) and OPRK(Figure A.1.14). While the binding modes of agonists and antagonists were different from each other, ligands with the same efficacy did not necessarily share the

Figure 2.1.3: Interaction fingerprints extracted from the crystallographic structures of (A) AA2AR, (B) ACM2, (C) ADRB2. Structures are grouped based on their activation state (antagonist bound inactive, agonist bound inactive, active), and the bound ligand. Different conformations are separated by the horizontal light blue line. Different ligands are separated by the horizontal lines on the PDB ID axis. The darker color indicates the presence of an interaction between a ligand with the designated residue.



(A) AA2AR



(B) ACM2

Figure 2.1.4: Interaction fingerprints extracted from the crystallographic structures of ADRB2. Structures are grouped based on their activation state (inactive and active), and the bound ligand. Different conformations are separated by the horizontal light blue line. Different ligands are separated by the horizontal lines on the PDB ID axis. The darker color indicates the presence of an interaction between a ligand with the designated residue.

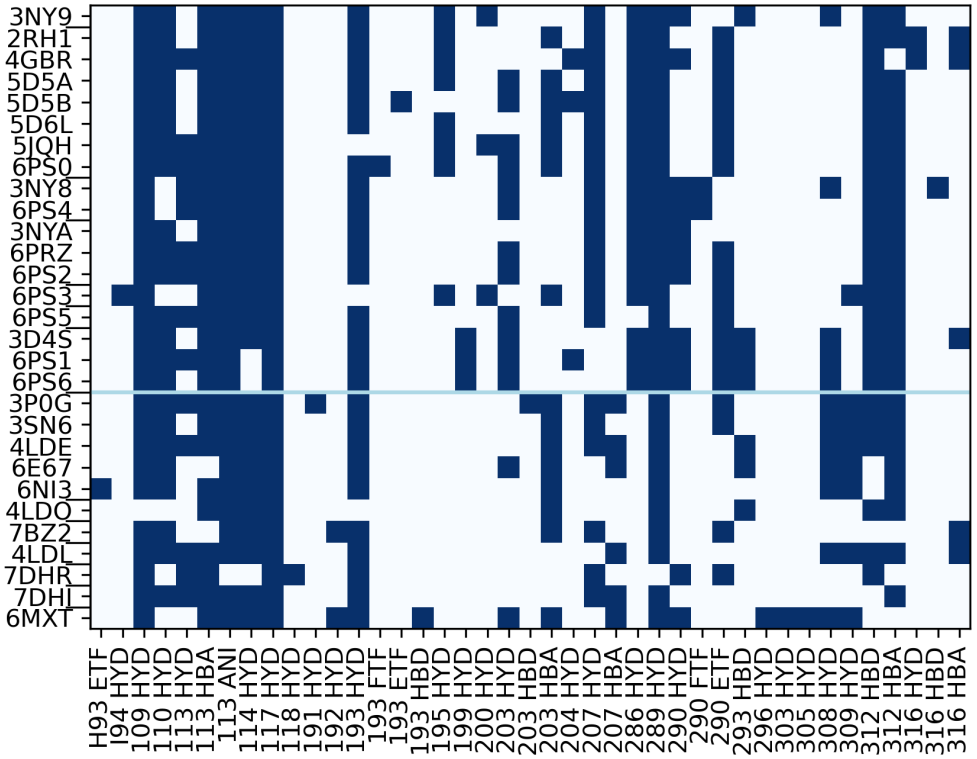
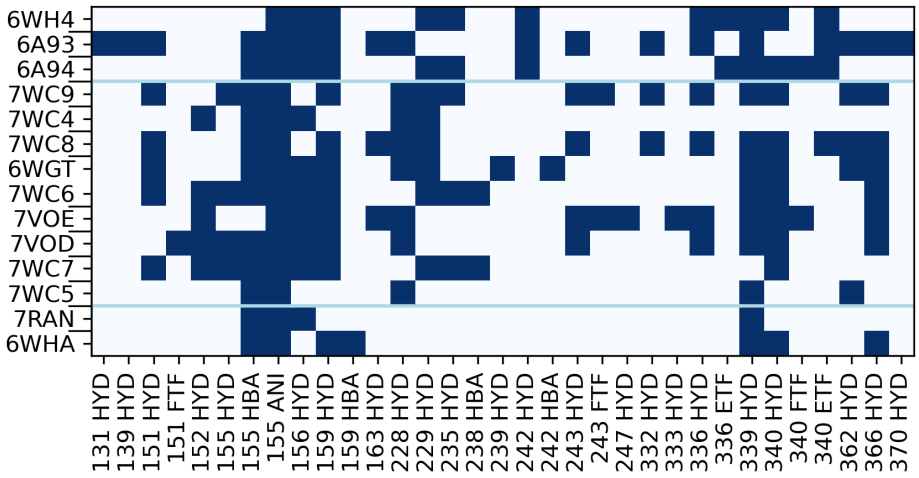
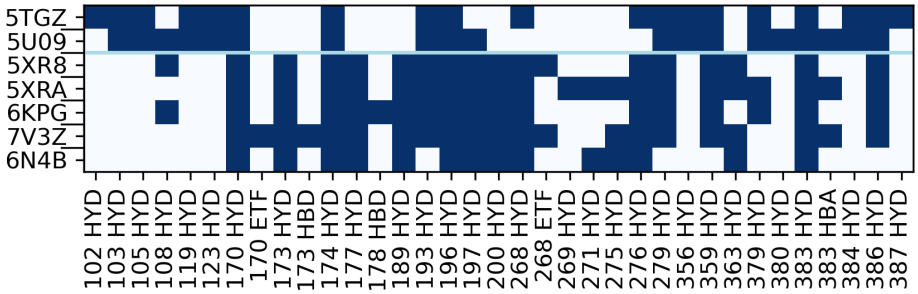


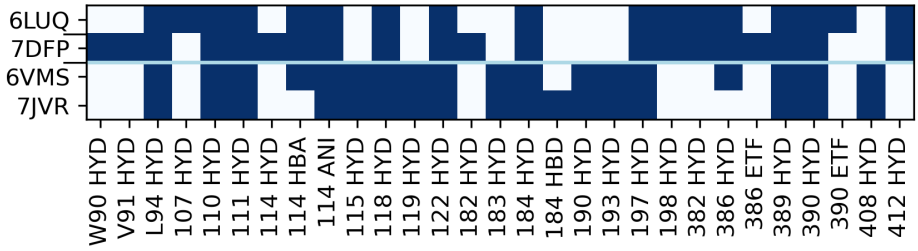
Figure 2.1.5: Interaction fingerprints extracted from the crystallographic structures of (A) 5HT2A, (B) CNR1, (C) DRD2. Structures are grouped based on their activation state (antagonist bound inactive, agonist bound inactive, active), and the bound ligand. Different conformations are separated by the horizontal light blue line. Different ligands are separated by the horizontal lines on the PDB ID axis. The darker color indicates the presence of an interaction between a ligand with the designated residue.



(A) 5HT2A



(B) CNR1



(C) DRD2

same interactions. The similarity between all binding modes of AA2AR's antagonists were comparable to their similarity to agonist binding modes (Figure A.1.9B). The same trend was also observed for ACM2 and ADRB2 agonists (Figure A.1.10B, A.1.11B). Only CNR1 and DRD2 ligands had two clearly defined binding modes for agonists and antagonists, although this might just reflect the small number of analysed structures (Figure A.1.12B, A.1.13B).

Clustering using IFP similarity identified from two to seven distinct binding modes for each target (Figure A.1.8C-A.1.13C), often highlighting clear differences between the interactions formed by ligands with the same efficacy. 5HT2A ligands form distinct interactions (Figure A.1.8C), with seven different binding modes being identified. The largest cluster combined the binding modes of both agonists and antagonists bound to inactive state structures. AA2AR antagonists present two distinct binding modes, separating caffeine and DB08770 from the other antagonists (Figure A.1.9C). The ligands of ACM2 present highly conserved binding modes between different structures, with clustering separating the different ligands (Figure A.1.10C). Antagonists of ADRB2 present a homogeneous binding mode, while agonists were separated between one cluster describing the binding mode of the structurally similar BI167107, HBI, and salbutamol, and four singletons (Figure A.1.11C).

Ligands' structures and properties

The known drugs, clinical candidates, and crystallographic ligands were compared based on their structures and physicochemical properties, to analyse the relationship between structure and efficacy.

The similarity between ligands was measured using binary Morgan fingerprints of radius 2 (MFP2), and two different pharmacophoric fingerprints, the default RdKit pharmacophoric fingerprints (Pharma)[7], and the pharmacophoric fingerprints developed by Gobbi and Poppinger (Gobbi)[8]. Three scenarios were observed by comparing the molecular similarities of ligands with the same efficacy or with different efficacies (Figure A.1.15-A.1.21): agonists and antagonists of ACM2, ADRB2, CNR1, and DRD2 were different from each other and similar between themselves. Agonists and antagonists of AA2AR were structurally distinct from each other, but only the agonists presented similarities between them (Figure A.1.16). It was not possible to distinguish between agonists and antagonists of 5HT2A and OPRK by molecular similarity alone.

The ranking power of molecular similarity to a reference for efficacy prediction was measured using the ROCAUC score(Figure A.1.22-A.1.28 and Table A.1.3). In multiple instances values above 0.70 were recorded, indicating a clear structure bias in the dataset. Most of the previously observed relationships between ligands were confirmed, with reference ligands of both efficacies performing well (ACM2, ADRB2, CNR1, DRD2), only references for one efficacy (AA2AR, OPRK), and structurally unbiased datasets (5HT2A).

Three different machine learning models were trained using molecular or pharmacophoric fingerprints to classify a ligand based on its efficacy: random forest (RF), k nearest neighbours (kNN), and support vector machine (SVM). RF is a non-linear classification method, SVM instead is a linear classification method in the transformed feature space, while kNN is a lazy learning method, which assigns a label based on the most similar training examples. The models were evaluated based on their balanced accuracy in leave one out cross validation (LOO-CV). All trained classifiers were predictive (balanced accuracy between 0.66 to 1.00), with only few ligands being consistently misclassified (Table 2.1.3).

The physicochemical properties of known agonists and antagonists were compared based

Table 2.1.3: Performance of the different machine learning models trained on molecular and pharmacophoric fingerprints

Target	Fingerprint	RF	kNN (k)	SVM	Frequent errors
5HT2A	ECFP4	0.76	0.88 (1)	0.8	terguride
	Pharma	0.66	0.82 (1)	0.84	terguride, lisuride, Lysergide, 25CN-NBOH
	Gobbi	0.76	0.89 (1)	0.85	terguride, lisuride, 25CN-NBOH
AA2AR	ECFP4	0.94	0.92 (2)	0.94	pentoxifylline
	Pharma	0.94	0.92 (8)	0.92	pentoxifylline
	Gobbi	0.94	0.92 (1)	0.92	pentoxifylline
ACM2	ECFP4	0.9	0.89 (1)	0.8	pilocarpine
	Pharma	0.9	0.90 (1)	0.8	pilocarpine
	Gobbi	0.9	0.90 (1)	0.9	pilocarpine
ADRB2	ECFP4	0.94	0.94 (3)	0.92	celiprolol, pindolol
	Pharma	0.94	0.92 (3)	0.94	celiprolol, pindolol, sotalol
	Gobbi	0.97	0.92 (3)	0.94	celiprolol, pindolol
CNR1	ECFP4	0.93	0.93 (1)	0.93	MDMB-Fubinaca
	Pharma	0.93	0.86 (2)	0.93	MDMB-Fubinaca
	Gobbi	0.93	0.93 (1)	0.93	MDMB-Fubinaca
DRD2	ECFP4	0.89	0.96 (3)	0.95	-
	Pharma	0.86	0.93 (3)	0.94	acp-104, bifeprunox
	Gobbi	0.89	0.91 (1)	0.93	cariprazine
OPRK	ECFP4	0.81	0.85 (1)	0.79	JDTic
	Pharma	0.79	0.83 (1)	0.87	nalbuphine, JDTic
	Gobbi	0.77	0.79 (8)	0.77	naloxone

Table 2.1.4: Performance of the random forest models trained on physicochemical descriptors

Target	Balanced accuracy
5HT2A	0.65
AA2AR	0.94
ACM2	1
ADRB2	0.91
CNR1	0.86
DRD2	0.74
OPRK	0.73

on 23 calculated descriptors (see Materials and Methods). For all targets, except 5HT2A, at least one property was significantly different between agonists and antagonists. These differences in properties were further exploited by training a RF classifier. Based on the balanced accuracy calculated in LOO-CV, the model trained on 5HT2A data had a low accuracy of 0.65, for DRD2 and OPRK the models were moderately accurate, with balanced accuracies of 0.74 and 0.73 respectively, while the remaining models had accuracies of at least 0.86 (Table 2.1.4).

Agonists and antagonists present clear differences in their structures and properties. In order to determine which features are correlated to a ligand's efficacy SHAP analysis [9][10] was performed on the RF models trained on physicochemical properties, and MFP2. 5HT2A was the target with the worse performances between all models, with the only insights behind the mechanism regulating a ligands' efficacy being a link between agonism and the presence of a secondary or aromatic amine (Figure A.1.29). OPRK and DRD2 could be well classified based on their structures (balanced accuracy 0.81 and 0.89), while a lower accuracy was observed using properties descriptors (balanced accuracy 0.73 and 0.74). The AA2AR, ACM2, ADRB2, and CNR1 ligand datasets present clear biases regarding the structures (balanced accuracy 0.94, 0.90, 0.94, 0.93) and properties (0.94, 1.00, 0.91, 0.86). For AA2AR both biases could be reconducted to the presence of an adenosine-like scaffold in agonist ligands. Key structural features for classification represented fragments of the ribose or the adenine groups, while the larger number of hydrogen bond donor and acceptor groups, oxygens, and chiral centres, could be explained by the presence of a ribose-like moiety(Figure A.1.30). ACM2 agonism was associated to smaller molecules, without aromatic rings, and with a positive charge located on a quaternary amine (Figure A.1.31). An agonist efficacy for ADRB2 was associated with the presence of hydroxyl groups, either as a substituent in ortho or alpha position in an aromatic ring, or as additional hydrogen bond donor groups(Figure A.1.32). CNR1 ligands present a correlation between the presence of a chlorine atom in their structure and antagonism, likely caused by an external bias in the dataset rather than a specific contribution to ligand binding (Figure A.1.33).

Redocking and crossdocking

The role of receptor conformation in binding mode prediction was evaluated via redocking and crossdocking experiments. Redocking of the crystallographic ligands in their native

structures generated at least one pose with a root mean squared deviation (RMSD) below 2 Å in 86% of cases, 93% within 2.5 Å (Figure 2.1.6A). It was not possible to reproduce the crystallographic binding mode only for 5HT2A and AA2AR ligands. For 5HT2A this was caused by a difference in conformation between the co-crystalized ligand and the in-silico generated structure (6WH4), or by the original binding mode being largely solvent exposed (7WC4, 7WC5). For AA2AR the only ligand which could not be properly redocked was the large agonist UK-432097. The ligand was co-crystallized four times, and in two instances a pose within 2.5 Å was generated, while in the other cases a higher RMSD was measured (3QAK, 7EZC). The CHEMPLP scoring function ranked poses with RMSD below 2 Å as the best only in 59% of cases. However, for ADRB2, CNR1, and DRD2 the best scored pose had an RSMD below 2 Å in at least 85% of cases, while poor performances were observed for 5HT2A (21%), AA2AR (36%), and ACM2 (50%) (Figure 2.1.6B). The quality of the generated poses was evalauted based on the Tanimoto similarity between the crystallographic and docking-generated IFP. A similarity of at least 0.70 was observed for 69% of poses, increasing to 88% with a similarity of at least 0.60 (Figure 2.1.6C). The pose with the best docking score had an IFP similarity of at least 0.70 only in 44% of cases, and of at least 0.60 in 62% of cases (Figure 2.1.6C).

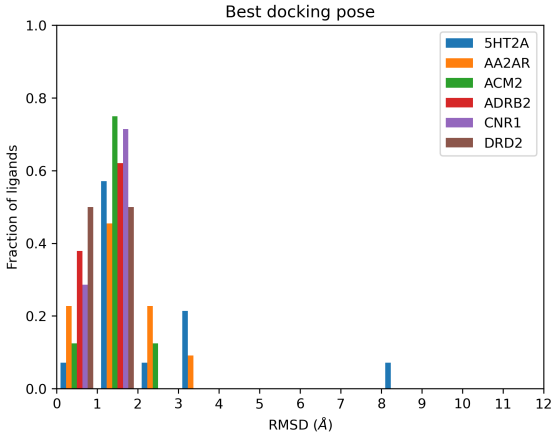
Crossdocking was performed to determine the influence of structural rearrangements on the quality of the docking poses (Figure A.1.36-A.1.42). The distribution of RMSDs between the different combinations of ligand efficacies (agonist, antagonist), and receptor activation states (active state, agonist bound inactive state, antagonist bound inactive state) showed a consistent quality for agonist-active state, and the antagonists-antagonist bound inactive state pairings. The correct binding modes were recovered for most targets, with 5HT2A, and AA2AR being the two main exceptions. Crossdocking between "mismatched" pairs produced three different scenarios: crossdocking was not possible (CNR1), crossdocking was possible only for agonists or antagonists, without an increase in RMSD (ADRB2 agonists), crossdocking was possible only for agonists or antaognists, but with an increase in RMSD (ACM2 agonist, DRD2 agonist, AA2AR antagonists). Comparable results were observed using IFP similarity. The poor performances of 5HT2A, and AA2AR could be explained by the greater plasticity of their binding pockets. By re-analysing the results grouping the structures based on the binding site conformation clustering, better results were observed for crossdocking between structures clustered together, than for structures in the same activation state (Figure A.1.43).

2.1.3 Discussion

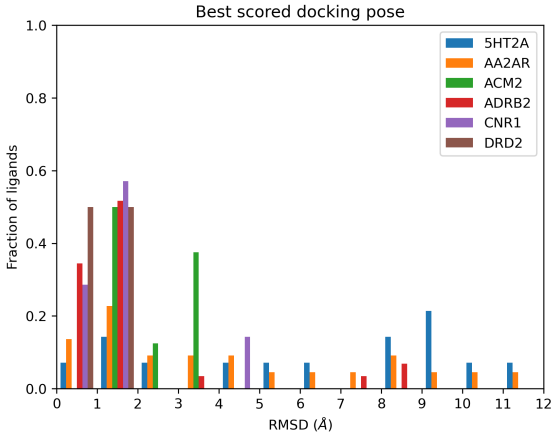
While agonists and antagonists are different from each other, ligands with the same efficacy do not necessarily share the same scaffold or structural features. For multiple targets only few ligands with the same efficacy could be described as part of the same chemical series, while other ligands were structurally diverse from each other. Even if no global structural differences could be identified depending on ligand efficacy, local similarities and key features could be exploited by machine learning. The trained classifier were discriminative, however, the algorithm often identified key features corresponding to only one of the two classes, while the other was defined by exclusion. For example, AA2AR agonists were identified by the presence of adenosine-like features, while ligand with a non-adenosine-like scaffold were automatically classified as antagonists, as was the case for pentoxifylline.

The binding site of the receptors assumed different conformations depending on its

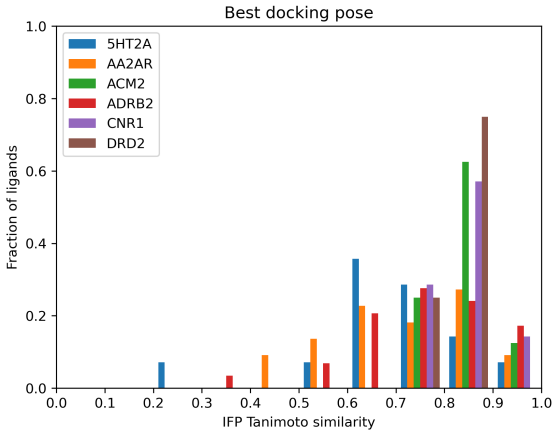
Figure 2.1.6: Normalized distribution of RMSD values and IFP Tanimoto similarities from the redocking experiment. (A) Pose with the lowest RMSD. (B) RMSD of the pose with the best docking score. (C) Pose with the highest IFP similarity. (D) IFP similarity of the pose with the best docking score.



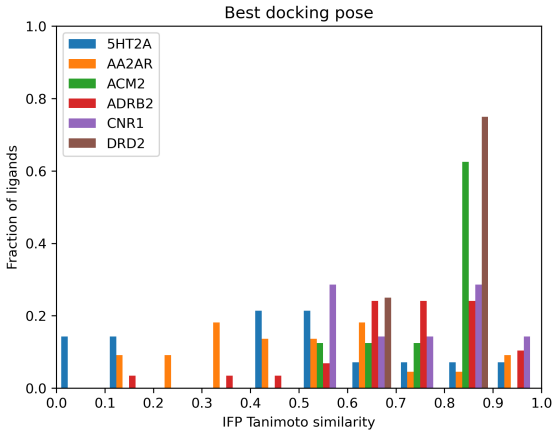
(A) Lowest RMSD poses



(B) RMSD best scoring poses



(C) Highest IFP similarity poses



(D) IFP similarity best scoring poses

activation state. Both clustering and graph analysis generally identified two main conformational states, roughly separating the active and inactive state structures. The magnitude of these differences varied greatly, from an average RMSD of 1.20 Å for ACM2, to 2.98 Å for CNR1. The activation state of the receptor was not the only factor regulating the conformation of the binding site, since multiple pocket conformations could be identified between structures bound to ligands with the same efficacy. These differences were both large, influencing the structure of the entire pocket, or small, being limited to the conformation of one sidechain. Even small differences in pocket conformation were able to impact the ability of a ligand to bind the receptor, or to be more precise, they impacted the ability of docking in reproducing the crystallographic binding modes in a different structures. Cross-docking showed a general increase in RMSD compared to the crystallographic structure, since docking was performed on a different conformation of the binding pocket. However, it was not always possible to predict the outcome of these calculations, since similar binding pockets were unable to accommodate the same ligand, while structurally different pockets could bind the same molecule with the same pose. For example, no significant differences between the active and inactive state conformations of ADRB2's binding pockets could be identified by clustering and graph analysis, however it was not possible to reproduce the crystallographic binding mode of the antagonists in the active state structures, while for agonists the opposite was possible. For ACM2, whose active and inactive state binding pockets were identified as structurally different by clustering, crossdocking between "mismatched" efficacy-conformation pairs often achieved a RMSD $<$ 2.5 Å.

2.1.4 Conclusions

Agonists and antagonists are structurally different molecules, which bind different conformations of the same protein. These difference however are not trivial and require a careful analysis on a case-by-case scenario, since the magnitude of these differences and their specific relationships vary depending on the target. Regardless, both ligand and receptor information could be used during virtual screening to guide the design of ligands with a target efficacy.

2.1.5 Materials and Methods

Dataset preparation

Data regarding the available crystallographic structures of GPCRs, and an annotated list of approved drugs or clinical candidates were obtained from GPCRdb (version 19th June 2023)[2][1][3].

The GPCRdb entries for receptor structures were filtered based on their annotation based on different criteria:

- A bound small molecule ligand in the extracellular pocket.
- Receptor not classified as orphan receptor.
- Structure in the active or inactive state.

The remaining structures were further filtered to remove receptors crystallized in a single activation state. The remaining entries were filtered manually based on the number of drugs targeting the receptor. For each receptor subfamily a representative target was selected for analysis, and all its available crystallographic structures were downloaded from the protein data bank (PDB). The structures were evaluated manually to remove entries with the receptor bound to an allosteric ligand, as well as structures with water-mediated protein-ligand interactions. Small molecules other than the ligand were removed from the structures (water, lipids, crystallization cofactors), missing atoms were reconstructed, and protonation of the system was performed using MOE 2020.

The list of drugs and clinical candidates was filtered to determine receptor targets with both known agonists and antagonists. The list was filtered to include only the targets selected for analysis based on the availability of crystallographic structures. The ligands co-crystallized with the receptor were added to the list of known drugs and clinical candidates. The SMILES of the selected molecules were retrieved from PubChem based on their name. Duplicates and specific stereoisomers of racemic drugs were identified as pairs of molecules with Tanimoto similarity between RdKit's MFP2 of 1. The label describing the efficacy (agonist or antagonist) was manually checked against literature data, changing incorrectly labelled molecules, and removing entries with an unclear labelling.

Binding site analysis

Protein-ligand interactions in the selected crystallographic structures were detected using IChem's IFP module[11]. The binding site of each target was defined by including all residues forming an interaction with a ligand in at least one crystallographic structure. Structures were aligned on the rigid transmembrane domains of the receptor based on the definition given by GPCRdb. Pairwise RMSD between all non-hydrogen atoms forming the binding site was calculated using pytraj[12, 13]. The pairwise RMSD matrix was used to perform clustering, using the agglomerative clustering algorithm with average linkage. The optimal number of clusters was determined based on the distribution of silhouette scores.

Further analysis of the binding sites was performed using the graph-based redundancy removal method developed by Mohammadi et al.[4]. The structural similarity between different structures was represented using a graph, with each node representing one structure, and edges connecting identical structures. Two structures were considered as identical if for all residues symmetry adjusted RMSD calculated on all non-hydrogen atoms was below 2.0 Å. The symmetry adjusted RMSD value was measured using the symm_rmsd funtion from pytraj.

Ligand analysis

The protonation state of the identified ligands was determined using FILTER 2.5.1.4 (OpenEye, Cadence Design Systems, Inc.). The physicochemical properties, MFP2, RdKit pharamcophoric fingerprints, and the Gobbi pharmacophoric fingerprints were calculated for each molecule based on its SMILES using RDKit [7].

For each molecule 23 different physicochemical properties were calculated: logP, molecular weight, number of hydrogan bond donor groups, number of hydrogen bond acceptor groups, number of rotatable bonds, total charge, number of aromatic rings, number of heavy atoms, number of atoms, number of rings, topological polar surface area, number of chiral centres,

Model	Parameter	Values
kNN	k neighbours	1 to 11
	metric	Tanimoto, cosine
SVM	kernel	Linear
	class weights	Balanced, not balanced
RF	Bootstrap	Yes or no
	Number of estimators	50,100,150,200,250,300
	Split criterion	Gini, entropy
	Maximum number of features	square root, log2

quantitative estimate of drug-likeness (QED) score, total positive charge, total negative charge, and atom counts (C, N, O, F, Cl, Br, I, S, P). logP was calculated using the Crippen method [14]. The distribution of physicochemical properties between agonists and antagonists were compared for each target using the Kolmogorov-Smirnov test, considering the two distributions as different for $p < 0.05$.

A classification model to predict a ligand's efficacy was trained on the calculated descriptors using the random forest algorithm implemented in scikit-learn[15]. The performance of the classifier was evaluated in leave one out cross validation (LOO-CV), using the balanced accuracy value. Shaply values were calculated during cross validation using the SHAP python module[9][10].

The molecular and pharmacophoric fingerprints were used as descriptors to train three different efficacy classification models based on the RF, SVM, and kNN algorithms. For all algorithms their scikit-learn implementation was used. The models were optimized via grid search in LOO-CV, based on their balanced accuracy. The random forest models were analysed using SHAP.

Redocking and crossdocking

Initial conformations of the crystallographic ligands were generated using CORINA, while maintaining the original sterochemistry observed in the experimental structure. Docking was performed using PLANTS with the scoring function CHEMPLP at search speed 1 [16], generating 10 poses per ligand conformation. The pocket was defined using a 12 Å radius around the centre of mass of the binding pocket.

The quality of docking was measured by calculating the symmetry adjusted RMSD of the non-hydrogen atoms between the docking poses and the crystallographic structures. Calculations were performed after alignment of the structures on the binding site residues. Alignment and RMSD were calculated using pytraj.

Protein ligand interactions were detected using IChem's IFP module. Interaction similarity was measured based on the Tanimoto similarity between the two interaction fingerprints, calculated using the jaccard distance module from scipy[17].

References

[1] Albert J Kooistra, Stefan Mordalski, Gáspár Pándy-Szekeres, Mauricio Esguerra, Alibek Mamyrbekov, Christian Munk, György M Keserű, and David E Gloriam. GPCRdb in

2021: Integrating GPCR sequence, structure and function. *Nucleic Acids Research*, 49(D1):D335–D343, January 2021.

[2] Vignir Isberg, Bas Vroling, Rob van der Kant, Kang Li, Gert Vriend, and David Gloriam. GPCRDB: An information system for G protein-coupled receptors. *Nucleic Acids Research*, 42(D1):D422–D425, January 2014.

[3] Alexander S. Hauser, Misty M. Attwood, Mathias Rask-Andersen, Helgi B. Schiöth, and David E. Gloriam. Trends in GPCR drug discovery: New agents, targets and indications. *Nature Reviews Drug Discovery*, 16(12):829–842, December 2017.

[4] Sara Mohammadi, Zahra Narimani, Mitra Ashouri, Rohoullah Firouzi, and Mohammad Hossein Karimi-Jafari. Ensemble learning from ensemble docking: Revisiting the optimum ensemble size problem. *Scientific Reports*, 12(1):410, January 2022.

[5] Gilles Marcou and Didier Rognan. Optimizing Fragment and Scaffold Docking by Use of Molecular Interaction Fingerprints. *Journal of Chemical Information and Modeling*, 47(1):195–207, January 2007.

[6] Juan A. Ballesteros and Harel Weinstein. [19] Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. In Stuart C. Sealfon, editor, *Methods in Neurosciences*, volume 25 of *Receptor Molecular Biology*, pages 366–428. Academic Press, January 1995.

[7] Greg Landrum, Paolo Tosco, Brian Kelley, Ric, Sriniker, Gedeck, Riccardo Vianello, NadineSchneider, Eisuke Kawashima, Andrew Dalke, Dan N, David Cosgrove, Brian Cole, Matt Swain, Samo Turk, AlexanderSavelyev, Gareth Jones, Alain Vaucher, Maciej Wójcikowski, Ichiru Take, Daniel Probst, Kazuya Ujihara, Vincent F. Scalfani, Guillaume Godin, Axel Pahl, Francois Berenger, JLVarjo, Strets123, JP, and DoliathGavid. Rdkit/rdkit: 2022_03_2 (Q1 2022) Release. Zenodo, April 2022.

[8] Alberto Gobbi and Dieter Poppinger. Genetic optimization of combinatorial libraries. *Biotechnology and Bioengineering*, 61(1):47–54, 1998.

[9] Scott M. Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1):56–67, January 2020.

[10] Scott M Lundberg and Su-In Lee. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[11] Franck Da Silva, Jeremy Desaphy, and Didier Rognan. IChem: A Versatile Toolkit for Detecting, Comparing, and Predicting Protein–Ligand Interactions. *ChemMedChem*, 13(6):507–510, 2018.

[12] Daniel R. Roe and Thomas E. Cheatham. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation*, 9(7):3084–3095, July 2013.

[13] Hai Nguyen, Jason Swails, Daniel R. Roe, and Cody. Amber-Md/Pytraj: V2.0.0. Zenodo, March 2017.

[14] Scott A. Wildman and Gordon M. Crippen. Prediction of Physicochemical Parameters by Atomic Contributions. *Journal of Chemical Information and Computer Sciences*, 39(5):868–873, September 1999.

[15] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85):2825–2830, 2011.

[16] Oliver Korb, Thomas Stützle, and Thomas E. Exner. Empirical Scoring Functions for Advanced Protein-Ligand Docking with PLANTS. *Journal of Chemical Information and Modeling*, 49(1):84–96, January 2009.

[17] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, SciPy 1.0 Contributors, Aditya Vijaykumar, Alessandro Pietro Bardelli, Alex Rothberg, Andreas Hilboll, Andreas Kloeckner, Anthony Scopatz, Antony Lee, Ariel Rokem, C. Nathan Woods, Chad Fulton, Charles Masson, Christian Häggström, Clark Fitzgerald, David A. Nicholson, David R. Hagen, Dmitrii V. Pasechnik, Emanuele Olivetti, Eric Martin, Eric Wieser, Fabrice Silva, Felix Lenders, Florian Wilhelm, G. Young, Gavin A. Price, Gert-Ludwig Ingold, Gregory E. Allen, Gregory R. Lee, Hervé Audren, Irvin Probst, Jörg P. Dietrich, Jacob Silterra, James T Webber, Janko Slavič, Joel Nothman, Johannes Buchner, Johannes Kulick, Johannes L. Schönberger, José Vinícius de Miranda Cardoso, Joscha Reimer, Joseph Harrington, Juan Luis Cano Rodríguez, Juan Nunez-Iglesias, Justin Kuczynski, Kevin Tritz, Martin Thoma, Matthew Newville, Matthias Kümmerer, Maximilian Bolingbroke, Michael Tartre, Mikhail Pak, Nathaniel J. Smith, Nikolai Nowaczyk, Nikolay Shebanov, Oleksandr Pavlyk, Per A. Brodtkorb, Perry Lee, Robert T. McGibbon, Roman Feldbauer, Sam Lewis, Sam Tygier, Scott Sievert, Sebastiano Vigna, Stefan Peterson, Surhud More, Tadeusz Pudlik, Takuya Oshima, Thomas J. Pingel, Thomas P. Robitaille, Thomas Spura, Thouis R. Jones, Tim Cera, Tim Leslie, Tiziano Zito, Tom Krauss, Utkarsh Upadhyay, Yaroslav O. Halchenko, and Yoshiki Vázquez-Baeza. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, March 2020.

2.2 Are docking and protein ligand interactions enough for efficacy-driven drug discovery?

Abstract

Protein-ligand docking and interaction similarity were used in a virtual screening exercise, to identify known agonists and antagonists against property matched decoys. Ranking based on docking scores, comparative docking scores, and interaction similarity were used to determine the performance of each method. Scoring functions generally were not sufficiently precise to convert the difference in binding site structures into a bias in ligand ranking. Interaction similarity identified ligands with the same efficacy as the reference, however the performances depended on the conformational state of the structure used for docking, and on the quality of the reference. Despite its simplicity interaction similarity paired with docking proved to be a useful tool for efficacy-driven drug design.

2.2.1 Introduction

The design of ligands targeting members of the G-protein coupled receptor family has always been conditioned by the need to take into account the biological response caused by ligand binding, its efficacy[1]. Efficacy-driven virtual screening has been achieved often times by the use of a structure in a relevant conformational state for docking. In particular docking on inactive state structures generally identified hits with antagonist efficacy[2], while docking on active state structures generally identified hits with agonist efficacy [3][4]. However multiple exceptions to this trend have been observed[5], leading to the development of new methods and tools to design ligands with a given efficacy. Efficacy selectivity in virtual screening has been achieved by comparing the ranking or docking score of potential hits between active and inactive state structures of the same receptor [6][3][7], or by comparing the binding modes generated by docking with a reference experimental binding mode [8][9][4]. More recent approaches instead have proposed to combine docking with machine learning in order to determine the efficacy of a ligand based on its docking-generated binding mode [10].

This section will focus on the analysis and the comparison of these three different approaches: docking score comparison, interaction similarity, and machine learning models. These methods have been previously validated in retrospective virtual screening exercises[6][8][9][10], and in few cases they were further tested in prospective virtual screening campaigns [3][4]. However, these analysis were often limited to a single receptors, and considered only a limited number of experimental structures. Seven different targets representing some of the most studied GPCR subfamilies were selected to evaluate the performance of different docking post-processing methods. The methods were used to classify known ligands as agonist or antagonists against each other, as well as their performance in efficacy selective enrichment against a library of property matched decoys.

This analysis confirmed the effectiveness of interaction similarity methods, while also highlighting its limitations and its inconsistency in performance between different target receptors. Based on these results it was possible to determine some of the characteristics behind the success of scoring based methods and interaction similarity methods. Both approaches could identify the favourite receptor conformation of a ligand in the "conformational cafeteria"[11], leading to an efficacy bias in the final selection.

2.2.2 Results

Docking-based virtual screening

The seven datasets of known ligands were docked in the corresponding crystallographic structures using PLANTS. Ranking the compounds using the CHEMPLP scoring function [12] showed a clear ligand bias regardless of the structure used for docking. Ligands with one efficacy, varying depending on the target, consistently obtained better or similar scores than other ligands in all structures. Rescoring of the docking poses using three different scoring functions (GOLDSCORE[13], CHEMSCORE[13], and GOLD PLP), and normalizing the CHEMPLP score by the number of heavy atoms did not remove the bias (Figure A.2.1-A.2.7).

The docking scores and rankings obtained in different receptor structures were used for comparative docking, based on the rescoring schemes proposed by Scharf et al.[7]: dual reranking, which evaluates the consensus ranking obtained in two structures in the same conformational state, and selective reranking, which used the difference in docking score, or the ratio of ranks, between pairs of structures in different activation states. More information about the two comparative docking methods is available in the Materials and Methods section. Dual reranking did not improve the results compared to simple docking, often conserving the efficacy bias of the scoring function. On the other hand, selective reranking often yielded ROCAUC values above 0.70, clearly prioritizing ligands with the desired efficacy (Figure A.2.8-A.2.14). On average, ranking using the difference in docking score between active and inactive state structures led to high ROCAUC values from 0.77 for ADRB2, to 0.98 for ACM2. Despite the good performance in ranking, the difference in docking score did not always yield a physically relevant value. Using the sign of the difference between docking scores to classify a ligand as an agonist or antagonist yielded varying performances depending on the target. The average balanced accuracy of these predictions was poor for 5HT2A (0.51), acceptable for ADRB2 (0.69), and CNR1 (0.72), and good for ACM2 (0.95). For AA2AR the results varied greatly, from non-predictive (~ 0.4) to accurate (~ 0.8).

Property matched decoys were docked in the available structures. The screening power of docking, and the corresponding comparative docking methods was evaluated based on the fraction of agonists and the fraction of antagonists, ranked higher than the top 1% of decoys (Figure A.2.15). The scoring function retrieved multiple known ligands, up to 13% of all active molecules, however for 17 out of 86 structures no active molecule was identified. Only for CNR1 the retrieved active molecules had the same efficacy as the co-crystalized ligand (Figure A.2.15E), while an efficacy bias towards agonists was observed for ADRB2 and OPRK(Figure A.2.15D and A.2.15G), and towards antagonists for 5HT2A and DRD2(Figure A.2.15A and A.2.15F), regardless of the activation state of the structure used for docking. Dual reranking increased the number of selected active molecules only in 8% of cases compared to docking, while an increase in the number of ligands with the desired efficacy was observed only for 4% of all pairs. Selective reranking led to an increased recall towards agonists for 14% of pairs, and towards antagonists for 11 % of pairs.

Interaction-based virtual screening

The generated docking poses were rescored based on their interaction similarity to binding mode of the co-crystallized ligand. Two different interaction representations were considered: interaction fingerprints (IFPs), and a 3D representation based on triplets of interaction pseudo-atoms (IPAs), placed on the interacting atom of the ligand, the interacting atom

of the protein, and in the middle point between them. For the IPA representation two different definitions of hydrophobic contacts were used: the Hyd definition, based only on atom types (Hyd), and the stricter Newhyd criterion, considering the nature of the surrounding environment. Similarity between IFPs was measured using the Tanimoto similarity score, while for IPAs similarity was computed using a maximum subgraph scoring function (GRIM)[14], and by using a graph kernel based on an approximated representation of the IPAs as a weighted graph[15].

Considering only the known active molecules, interaction similarity prioritized ligands with the same efficacy as the reference. The results were highly dependent on both the nature of the ligand, the interaction representation, and the target (Figure A.2.16 - A.2.22). Certain similarity scores could prioritize both efficacies (ROCAUC > 0.70), like IFPs and graph kernels for ACM2, GRIM for ADRB2, GRIM and graph kernels for CNR1. In other cases, only ligands with one efficacy could be prioritized by interaction similarity: graph kernels for 5HT2A antagonists, IFPs or GRIM for AA2AR agonists, GRIM-Hyd for ACM2 antagonists, IFPs and graph kernels for ADRB2 antagonists, IFPs for CNR1 agonists. The lack of an efficacy bias was rare but nonetheless observed for 5HT2A and IFP, AA2AR and graph kernel-Hyd, ACM2 and GRIM Newhyd. Generally, poor performances were always observed for DRD2 and OPRK, likely due to the small number of reference binding modes.

In the presence of decoys, interaction similarity improved the enrichment in active molecules with the same efficacy as the reference compared to the docking scoring function. IFP had better performances from 28% of times for 5HT2A and CNR1 to 75% for ADRB2 . Using Hyd IPAs and GRIM from 57% of times for 5HT2A to 100% for ADRB2 , while with graph kernel from 28% of times for CNR1 to 72% for ADRB2. Newhyd IPAs with GRIM from 25% of times for ACM2 to 86% of cases for ADRB2, while the graph kernel from 0% of times for DRD2 to 75% for ADRB2. Regardless of the target or the representation, interaction similarity achieved efficacy selectivity in a large majority of cases (Table 2.2.1). CNR1 was the only significantly difficult target, with interaction similarity yielding no enrichment in the majority of cases using IFPs and Newhyd IPAs with a graph kernel, or in three cases out of seven using Hyd IPAs with a graph kernel, or Newhyd IPAs with GRIM. Considering only the fraction of identified ligands with the desired efficacy, GRIM using Hyd IPAs was the best performing method, with its best per target results ranging from the 14% recall for 5HT2A and DRD2 antagonists, to the 84% of ADRB2 agonists(Figure A.2.24), while with Newhyd IPAs the recall ranged from 10% for DRD2 agonists to the 84% of ADRB2 agonists(Figure A.2.25). Graph kernels using Hyd IPAs still performed well selecting from 10% of 5HT2A and DRD2 antagonists to 70% for ADRB2 antagonists (Figure A.2.26), as well as by using Newhyd IPAs, going from the 5% of DRD2 antagonists to 70% of ADRB2 antagonists(Figure A.2.27). IFP similarity was the worst performing method, retrieving from 5% of DRD2 agonists, to 60% of ACM2 agonists (Figure A.2.23).

The relationship between a binding mode reference and the structure used for docking was examined by rescoring the docking poses using all the available binding modes as references (Figure A.2.28-A.2.63). Generally, pairing binding modes and structures sharing the same activation state yielded similar performances compared to the experimental couple. “Mismatched” pairs instead presented a more complex situation, with a decrease in performance only observed for certain pairings and efficacies, while in the other case the fractions of recovered ligands was comparable to the ”correct” pairing. Clear drops in recall were observed for AA2AR agonists, ACM2 antagonists, and ADRB2 antagonists by using a reference binding mode with the same efficacy, but a structure in the “incorrect” activation

Table 2.2.1: Efficacy bias of the different method.

Methods	Efficacy biased	Unbiased	Inverse efficacy biased	Non enriching
IFP	55	6	9	14
GRIM Hyd	75	3	4	2
GRIM Newhyd	55	5	15	9
Kernel hyd	65	6	3	10
Kernel newhyd	58	4	7	15

state. An interesting phenomenon was observed for AA2AR and ADRB2, regarding the recall of ligands obtained using a binding mode reference with the opposite efficacy. For both receptors a higher number of antagonists was retrieved using an agonist biding mode as reference on docking poses from inactive state structures, than on docking poses from an active state structure (Figure 2.2.1A). Only for ADRB2 instead the number of agonists identified using an antagonist binding mode as reference was higher using docking poses in an inactive state structure than in the active state(Figure 2.2.1B). A clear links between the activation state of the receptor and the used reference binding mode could not be established for other targets due to the poor performances of the method (5HT2A), lack of sufficient structural data (DRD2, OPRK), or both (CNR1).

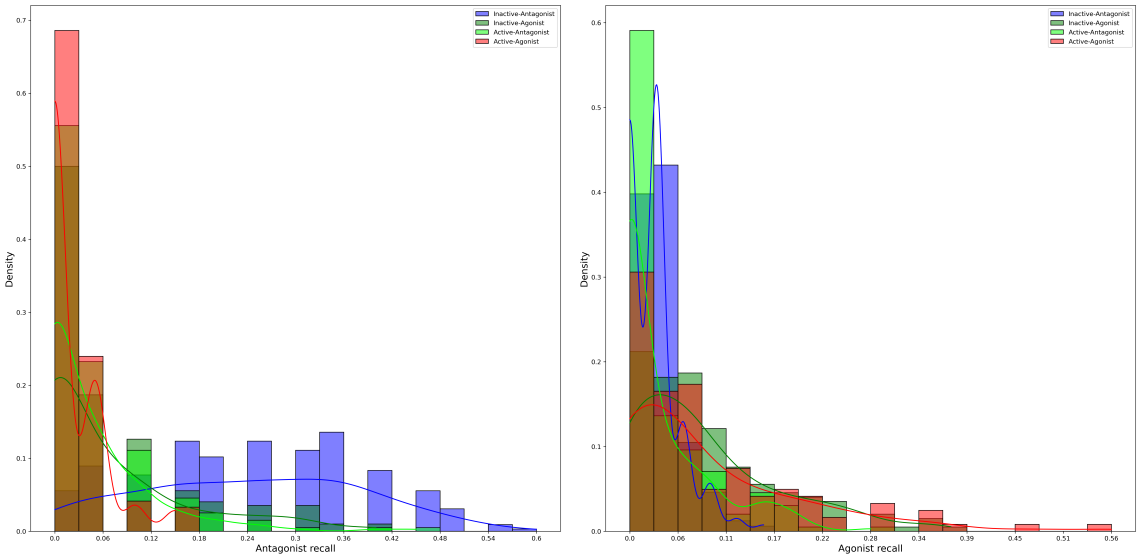
Binding mode classification

The binding modes of the known ligands generated with docking were used to train multiple random forest models, to classify the binding mode as agonist-like or antagonist-like. The quality of the models was evaluated using leave one ligand out cross validation (LOLO-CV). For each experimental structure models were trained using its corresponding docking poses. Other models instead were trained using an ensemble of docking poses obtained on structures in different activation states. For single structure models, all generated docking poses were used for training, as well as only the docking poses with the best docking score. Models trained using two sets of docking poses instead were trained by selecting for each ligand the poses with the best docking scores for each structure.

The trained models ranged from predictive with low accuracy to accurate (Table 2.2.2). On average 5HT2A, DRD2 and OPRK models were not accurate (balanced accuracy ~ 0.60), even though in some cases moderate accuracies could be reached (0.75 for DRD2, 0.75 for OPRK, and 0.71 for 5HT2A). ACM2, ADRB2, and CNR1 models were able to perform well in LOLO-CV, with average balanced accuracies above 0.80. For AA2AR both high accuracy and moderately accurate models, as well as non-predictive models were obtained. Pairing different datasets of docking poses extracted from different structures did not improve the performances in LOLO-CV (Table 2.2.3).

SHAP analysis was performed on all sufficiently accurate models (balanced accuracy $>$ 0.70) to identify the most relevant interactions used for classification. The number of times each interaction appeared within the ten most important features, and its relationship to efficacy was recorded. The results were aggregated based on the activation state of the structure used for docking (active, inactive, pair), and the type of docking poses used as training set (all, best score). Models trained only on the poses with the best docking

Figure 2.2.1: Distribution of ADRB2 antagonist (A) and agonist (B) recall after selecting the top 1% of the decoys with the highest Tanimoto IFP similarity. Results are grouped based on the efficacy of the reference ligand, and the activation state of the structure used for docking.



(A) Antagonist recall **(B)** Agonist recall

Table 2.2.2: Balanced accuracy of the RF models trained on docking generated interaction fingerprints

Target	Mean all poses	Mean best pose	Min all poses	Min best pose	Max all poses	Max best pose
5HT2A	0.59	0.59	0.54	0.59	0.63	0.71
AA2AR	0.72	0.68	0.66	0.48	0.77	0.85
ACM2	0.85	0.85	0.81	0.7	0.88	1
ADRB2	0.81	0.84	0.77	0.71	0.85	0.91
CNR1	0.87	0.89	0.78	0.77	0.92	0.93
DRD2	0.6	0.67	0.58	0.55	0.61	0.75
OPRK	0.6	0.69	0.59	0.62	0.6	0.75

Table 2.2.3: Balanced accuracy of the RF models trained on docking generated interaction fingerprints

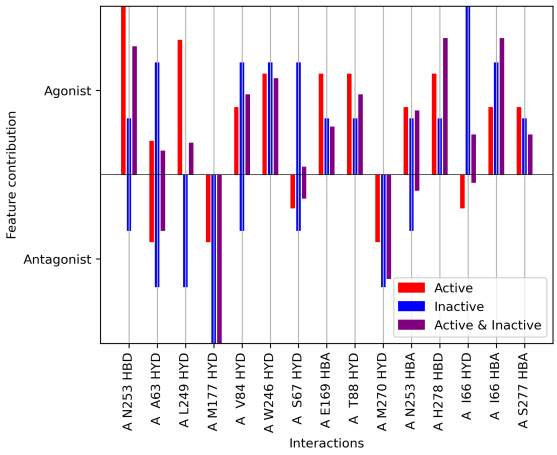
Target	Mean	Minimum	Maximum
5HT2A	0.6	0.54	0.65
AA2AR	0.69	0.48	0.89
ACM2	0.86	0.75	0.95
ADRB2	0.82	0.75	0.9
CNR1	0.84	0.77	0.92
DRD2	0.6	0.65	0.67
OPRK	0.67	0.67	0.67

scores were able to identify key features for ligand binding. However, depending on the activation state of the structure, different interactions were identified as efficacy determinant. Interestingly, in multiple cases the same interaction contributed to a different efficacy depending on the activation state of the structure used for docking. For example, the presence of a hydrogen bond with the amide nitrogen of Asn312 in ADRB2 was linked to agonism for models trained on active state structures, while it was linked to antagonism for models trained on inactive state structures. Models trained on pairs of structures often presented an intermediate result, for example the hydrogen bond with the nitrogen of Asn312 in ADRB2 was associated to both agonism and antagonism depending on the pair of structures used for docking (Figure 2.2.2C). The same trend was observed for the hydrophobic contact with Leu249 in ACM2(Figure 2.2.2B), or for the hydrophobic contact with Phe174 in CNR1(Figure 2.2.2D). On the other hand, certain features were identified as relevant only by a specific subset of models: The hydrogen bond between a ligand and Ser207 of ADRB2 was important to classify it as an agonist only for models trained on active state structures. The same was observed for the hydrophobic contacts with Tyr190 for ACM2 antagonists, identified only using data from inactive state structures, as well as the importance of a hydrophobic contact with either Thr197 or Leu276 for agonist efficacy at CNR1, determined only using data from active state structures. The interpretation of the models also identified certain interactions not observed in crystallographic structures. For example, a hydrophobic contact with Lys192 for CNR1 antagonists, the presence of a hydrogen bond with Ile66 for AA2AR agonists (Figure 2.2.2A), or a hydrophobic contact with Phe194 for ADRB2 antagonists. Models trained on all available docking poses were more consistent in the association between an interaction and a ligand's efficacy, but certain features were only identified for models trained one one ensemble of structure, or were associated to a different efficacy depending on the training set (Figure A.2.64).

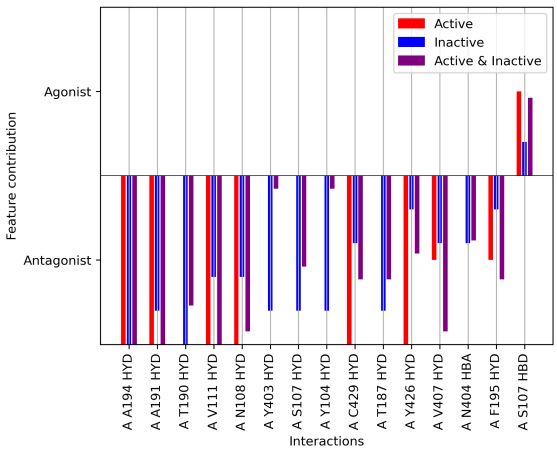
2.2.3 Discussion

Two different approaches for efficacy-driven drug discovery were evaluated in this analysis: conformational selection, and interaction similarity. Conformational selection aims at using the docking score, or the difference in docking scores, to identify ligands with a certain efficacy. Interaction similarity instead aims at leveraging the difference between the binding modes of agonists and antagonists, to introduce efficacy selectivity.

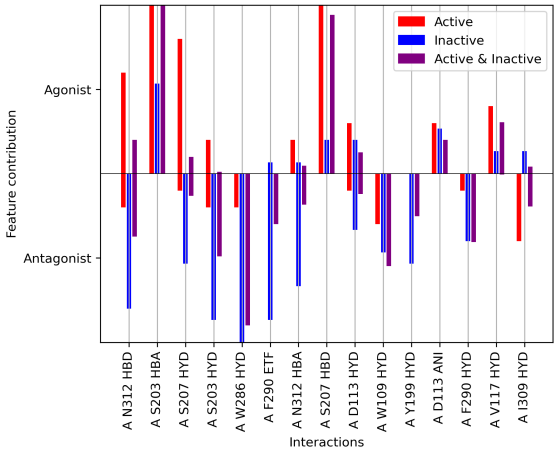
Figure 2.2.2: Frequency of an interaction being between the ten most important features for efficacy classification for models trained using the docking pose with the best docking score. Frequencies are relative to the activation state of the structure used for docking (active or inactive), or by using two structures in different activation states. Results are indicated for the four best performing targets: (A)AA2AR, (B)ACM2, (C)ADRB2, (D)CNR1



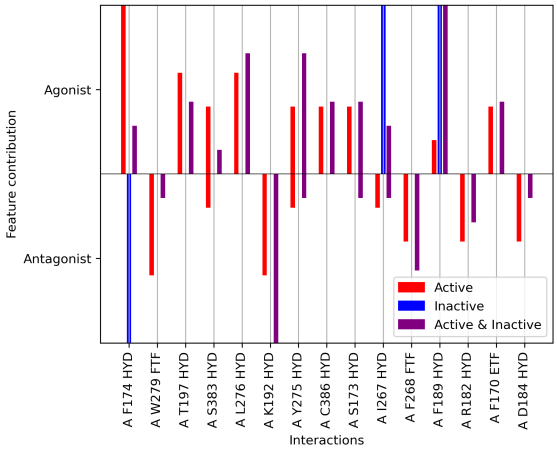
(A) AA2AR



(B) ACM2



(C) ADRB2



(D) CNR1

The evaluated scoring functions present an undesired ligand bias, always ranking ligands with one efficacy higher than others, regardless of the receptor conformation. The ligand ranking, efficacy wise, was slightly influenced by the activation state of the used receptor structure, but it was not sufficient to clearly separate the two classes. In the literature other scoring function (Glide SP score [6],DOCK3.6 [3, 5]) were shown to selectively identify ligands based on efficacy depending on the receptor structure used for docking. These three studies were conducted on ADRB2 and AA2AR, and showed an efficacy bias depending on the structure used for virtual screening, either from comparison between known ligands[6], or by evaluating the enrichment against decoys[3, 5]. The discrepancies between the obtained results and the literature data might be caused by a difference in the accuracy of the investigated scoring functions, or by differences in the composition of the dataset. The dataset used by Costanzi and Vilar[6], and by Weiss et al. [3], was formed by 30 agonists, and 30 antaognists of ADRB2, and it considered different enantiomers of the same ligand as separate entries. Only 30 entries are shared between this dataset and the ligands used in our study, with 23 common unique ligands. However, conformational selection is not based on agonists having a higher affinity for active state structures than antagonists, but on their higher affinity for active state conformations of the receptor rather than for inactive state conformations, and vice versa. Selective reranking, using the difference between docking scores from pairs of structures in different conformational states, generally could prioritize ligands with the desired efficacy given a library of known active molecules. In this case the results reflected the trend observed in the literature [6]. The difference in docking score seems to be able to remove some of the ligand dependent bias in the scoring functions. While this approach proved to be useful on a small scale, selective reranking was inconclusive in a virtual screening like scenario, as already shown by Scharf et al. in a prospective virtual screening campaign against ADRB2[7]. Comparison of docking scores could still be used to help identify ligands with a desired efficacy after restricting the analysis to a smaller subset of promising candidates[3].

Interaction similarity was confirmed to be a more reliable approach than docking scoring functions for both hit enrichment, and efficacy-selective enrichment. Hit rates increased compared to docking regardless of the used interaction representation, and an efficacy bias was observed in almost the totality of cases. Between the different methods GRIM, the most complex similarity score, clearly outperformed the others, followed by the graph kernel, and finally by the simpler IFP similarity. Despite the successes, the recall values varied greatly depending on the target, with ADRB2 being the easiest target, retrieving on average 11% of all agonists and 28% of all antagonists, and DRD2 being the most difficult target, retrieving on average 5% of all agonists and 3% of all antagonists.

The small number of available GPCR crystallographic structures limits the use of experimental binding modes for virtual screening to interaction similarity, since more complex approaches, like machine learning models would require a larger dataset. Jimenez-roses et al. proposed to train an efficacy classification model using *in silico* binding modes generated using docking[10]. Such models were tested based on their ability to discriminate between agonist-like and antagonist-like binding modes, as well as to gain a better insight on efficacy specific interactions through model interpretation. For four of the seven evaluated targets (AA2AR, ACM2, ADRB2, CNR1) the classifiers performed well in cross-validation, while for the remaining three targets (5HT2A, DRD2, and OPRK) it was not possible to consistently train predictive models. While based on the same principles, the models developed in this work are different from similar works presented in the literature [10, 16] concerning the size

of the dataset, the docking software used to generate the poses, and the type of fingerprint used to describe the binding mode. Given these differences it is not possible to draw a direct comparison between the performances of the different models, however by comparing the relevant features identified for ADRB2 and AA2AR, most of the same key interactions were identified in both works. However, the results show that such models should only be trained on docking poses obtained from a single crystallographic structure, or only from ensembles of crystallographic structures in a similar conformational state. Separating the data from structures in different activation states does not increase the performance of the models, but assures a better consistency in what the models effectively learn, since the same interaction was observed to be associated with an agonist efficacy for models trained on active state structures, and to an antagonist efficacy for models trained on inactive state structures. This separation is relevant also for perspective applications, since the evaluated docking poses would be obtained only for one conformation of the receptor, and not for the entire ensemble.

The use of structure-based approaches for efficacy-driven drug discovery hinges on a complex balance between efficacy specific interactions and conformational selection. Both in this work and in the literature a difference in docking scores for the same ligand docked on structures in different conformational states has often been reported [3, 5, 6],. These differences, however, do not automatically translate in an efficacy bias between the top ranking compounds. Interaction similarity acts as an efficient scoring function measuring the fitness of a ligand for a certain protein conformation, as well as the presence of specific chemical features in the ligand's structure. Ranking by binding mode similarity of docking poses obtained in an "incorrect" receptor conformation led to a decrease in recall for AA2AR agonists, and ADRB2 and ACM2 antagonists, while no significant decrease was observed in the opposite cases. The results from the crossdocking experiments from Section 2.1, showed a complex relationship regulating ligand recognition by different binding site conformations. Docking could not reproduce the crystallographic binding mode in structures in a different activation state for certain ligands, including AA2AR agonists, and ADRB2 and ACM2 antagonists, while AA2AR antagonists, and ADRB2 and ACM2 agonists could be docked correctly in all structures. Interaction similarity not only determines the presence or absence of efficacy specific interactions, but it also measures the fitness of the ligand for the binding pocket. A ligand is able to form certain interactions, either efficacy specific or shared interactions, only when it is docked in a compatible conformation of the receptor, while the same binding mode cannot be reproduced, or it can only be reproduced partially, in other structures. This interpretation might explain why a higher fraction of ADRB2 agonists was identified using an antagonist binding mode reference in inactive state structures, than in active state structures. ADRB2 agonists can be correctly docked in both active and inactive state conformations of the receptor, but their binding mode is more antagonist-like in inactive state structures, than when docked in active state structures. At the same time a higher fraction of ADRB2 and AA2AR antagonists is identified by comparison to an agonist binding mode after docking in inactive state structures of the receptor, than after docking in active state structures. This means that the binding mode of such antagonists is more agonist-like in inactive state structures, than the respective binding modes obtained in active state structures. These results could also explain the changes in feature importance for the efficacy classifiers trained on docking poses. Shared interactions in crystallographic structures are associated to a different efficacy depending on the activation state of the structure used for docking, while other relevant feature are only observed on docking poses obtained in certain

conformations of the receptor. Some interactions can only be formed by agonist ligands when docked in an active state structure, while the same interactions are not observed, or they are observed more rarely if the same molecules are docked in an inactive state structure, and vice versa. Therefore, the success of interaction similarity also includes a conformational selection contribution, rather than just considering pharmacophoric information. It is important to note that in this case conformational selection is not seen in reference to binding affinity, but rather it is linked to shape complementarity between the molecule and the binding pocket.

Both interaction similarity and conformational selection methods require to have access to a structure in the correct conformational state to identify relevant ligands. While this can be roughly summed up to the need for an active state structure for agonist discovery, or of an inactive state structure for antagonist design, sometimes more nuance is necessary. As shown in Section 2.1 even within the same activation state different binding site conformation might exist, which influence the final hit list, as recently highlighted by the work of Lyu et al.[17] comparing experimental and *in silico* generated conformations of 5HT2A.

2.2.4 Conclusions

To answer to question "Are docking and protein ligand interactions enough for efficacy driven drug discovery?", despite their simplicity both are still effective tools to identify new ligands with a desired efficacy. Docking could generate a pose with a similar binding mode as a known reference for a fraction of known ligands, and interaction similarity could correctly prioritize it as more relevant than decoys' poses. The activation state of the receptor plays a focal role in the quality of the results, since many interactions critical for ligand efficacy can only be formed in one conformation of the receptor. While useful, these methods are not universally successful and had lackluster performances for certain target-reference pairs. This indicates the need for further developments in this field to fully exploit the potential of these two resources, or to identify alternative and more effective approaches.

2.2.5 Materials and Methods

Ligand preparation

The three dimensional structures of the protonated molecules generated in Section 2.1 were prepared using CORINA. The rc function was used explore alternative ring conformations, capping the number of ring conformers to 2. If present stereochemistry indications for a ligand were available they were maintained.

Decoy generation

Decoys were generated using the DeepCoy generative model trained on DUDE data [18]. The generated molecules were filtered to remove entries containing rings with more than 8 members. The protonation state of the generated molecules was recalculated using OpenEye FILTER to remove eventual protonation errors, and to assure consistency with the ligand. For each active molecule 30 decoys were selected considering ALL descriptors. The quality of the decoys was determined by the built-in scores of DeepCoy (Table A.2.1). The random forest models trained on the matched physicochemical properties of 5HT2A, DRD2, and OPRK had high ROCAUC scores (0.74,0.79,0.70). The script was modified to include SHAP analysis on the model. 5HT2A and DRD2 decoys seem to have differences in synthetic

accessibility, topological polar surface area, and number of stereocenters compared to known actives, while OPRK decoys tend to have a smaller number of rings and a higher number of rotatable bonds. Despite the presence of a bias the decoy sets were kept, since these properties should not influence the virtual screening results.

Docking

Docking was performed using the same protocol described in Section 2.1 for redocking and crossdocking.

The docking poses generated using PLANTS[12] for known agonists and antagonists were scored using three different scoring functions available in the GOLD docking software: GOLD PLP, CHEMSCORE[13], GOLDSCORE[13]. Docking pose optimization was allowed during rescoring to remove potential clashes or bad scores caused by differences in definitions between the docking software.

Consensus and relative docking

The results from docking were post-processed using the dual reranking[19] and the selective reranking[7] rescoring schemes. For dual reranking a new score D was calculated based on the ranking of each ligand in two receptor structures in the same activation state, according to Equation 2.1. The compounds were ranked in ascending order based on the value of D.

$$R_{rel,i} = \frac{rank_i - 1}{total\ molecules_i - 1}$$

$$D = \frac{(R_{rel,1} - R_{rel,2})^2 + R_{rel,1} + R_{rel,2}}{2} \tag{2.1}$$

Relative scoring instead was calculated using the difference in docking score between pairs of structures in different activation states. The new score was calculated as the ratio between the rankings, as described by Equation 2.2, or by the difference between the docking scores, as described by Equation 2.3. By using the ratio between ranking compounds were ranked in ascending order based on S to identify agonists, and in descending order to identify antagonists. For the difference in docking score the order on ranking depended on the evaluated docking score.

$$S = \frac{rank_{Active}}{rank_{Inactive}} \tag{2.2}$$

$$\Delta Score = Docking\ score_{Active} - Docking\ score_{Inactive} \tag{2.3}$$

Both rescoring schemes were used with all four evaluated scoring functions to evaluate the the classification of known agonists and antagonists. For the screening against decoys only the CHEMPLP scoring function was considered for evaluation.

Interaction similarity ranking

Protein ligand interactions were detected using IChem's[20] IFP module[21], storing the binding mode as a binary vector, and using IChem's Ints module[14], which represents the interactions as triplets of interaction pseudoatoms (IPAs). A IPA triplet is represented by two IPAs placed in the same position as the interaction atom of the ligand, and of the

protein, and by a third IPA placed in the middle point between the two. Each IPA is labelled based on its position (protein, ligand, center), and the type of interaction it encodes (hydrophobic contact, hydrogen bond with the protein acting as acceptor, hydrogen bond with the protein acting as donor, ionic bond with the protein acting as anion, ionic bond with the protein acting as cation, edge to face $\pi - \pi$, face to face $\pi - \pi$). The Ints module uses two different definitions of hydrophobic contacts, one considering only the atom type of the interacting atoms (Hyd), and a stricter one allowing hydrophobic contacts only for atoms in a hydrophobic environment (NewHyd).

The IPAs were converted into approximated interaction graphs to facilitate comparison[15]. A weighted complete graph was generated using the IPAs as nodes. Each node was connected to all other nodes in the graph, and the weight of the edge corresponded to the distance between the two points approximated to the nearest Å. In case of two overlapping IPAs, an edge with a weight of 0.5 was generated.

Interactions were detected for the docking poses, and for the crystallographic structures described in Section 2.1. The similarity between the binding mode of a reference ligand, and the binding mode of the docked molecule was measured using:

- The Tanimoto similarity between the binary IFPs
- The similarity computed using the shortest path graph kernel for the IGs built from IPAs[15].
- The GRIM score between two groups of IPAs[14].

GRIM is an empirical scoring function for interaction similarity, based on the maximum common subgraph obtained by aligning the IPAs. More information about the shortest path kernel for graph similarity is given in the Materials and Methods of Section 3.1.

Docking poses were ranked in descending order based on the interaction similarity score.

Evaluation of virtual screening results

Screening of the library containing only known agonists and antagonists was evaluated using the ROCAUC score. Docking and dual reranking considered as positive class the efficacy of the co-crystallized ligand in the experimental structure used for docking. For selective reranking agonists were considered as the positive class.

The quality of the virtual screening against the property matched decoys was measured based on the fraction of recovered agonists, and the fraction of recovered antagonists, after selecting the top 1% of the ranked decoys.

IFP-based classifiers

The IFPs of the docked known agonists and antagonists were used to train random forest model to predict ligand efficacy. The IFP were preprocessed to remove not relevant interactions. Ionic bonds with the protein acting as the cation were considered only for Arg, His and Lys, ionic bonds with the protein acting as the anion for Asp and Glu, and $\pi - \pi$ interactions for His, Phe, Trp, and Tyr. Additionally residues not forming any interactions with the docked ligands were also removed from the IFP.

Hyperparameter tuning was performed using a grid search algorithm in leave one ligand out cross validation (LOLO-CV) based on the balanced accuracy of the model. In LOLO-CV

all docking poses of a molecule, including all stereoisomers and conformers, are removed from the training dataset, and used as a test set. Since docking might generate multiple poses with a similar binding mode for the same ligand, not separating all the poses of the same ligand between the training and the test set might artificially increase the measured balanced accuracy of the model.

The models were trained using IFPs generated after docking on a single structure, and by pairing IFPs generated after docking on structures in different activation states. SHAP analysis was performed on all models with a LOLO-CV balanced accuracy > 0.70, to identify the ten most important interactions for classification, and their correlation to either agonist or antagonist efficacy. The frequency with which each interaction appear between the ten most important features, and its correlation to ligand efficacy, was measured for the ensemble of models trained on the same number of docking poses, and from structures in the same activation state.

References

[1] Brian K. Shoichet and Brian K. Kobilka. Structure-based drug screening for G-protein-coupled receptors. *Trends in Pharmacological Sciences*, 33(5):268–272, May 2012.

[2] Peter Kolb, Daniel M. Rosenbaum, John J. Irwin, Juan José Fung, Brian K. Kobilka, and Brian K. Shoichet. Structure-based discovery of $B2$-adrenergic receptor ligands. *Proceedings of the National Academy of Sciences*, 106(16):6843–6848, April 2009.

[3] Dahlia R. Weiss, SeungKirl Ahn, Maria F. Sassano, Andrew Kleist, Xiao Zhu, Ryan Strachan, Bryan L. Roth, Robert J. Lefkowitz, and Brian K. Shoichet. Conformation Guides Molecular Efficacy in Docking Screens of Activated β-2 Adrenergic G Protein Coupled Receptor. *ACS Chemical Biology*, 8(5):1018–1026, May 2013.

[4] Albert J. Kooistra, Henry F. Vischer, Daniel McNaught-Flores, Rob Leurs, Iwan J. P. de Esch, and Chris de Graaf. Function-specific virtual screening for GPCR ligands using a combined scoring method. *Scientific Reports*, 6(1):28288, September 2016.

[5] David Rodríguez, Zhang-Guo Gao, Steven M. Moss, Kenneth A. Jacobson, and Jens Carlsson. Molecular Docking Screening Using Agonist-Bound GPCR Structures: Probing the A $_{2A}$ Adenosine Receptor. *Journal of Chemical Information and Modeling*, 55(3):550–563, March 2015.

[6] Stefano Costanzi and Santiago Vilar. In Silico screening for agonists and blockers of the $B2$ adrenergic receptor: Implications of inactive and activated state structures. *Journal of Computational Chemistry*, 33(5):561–572, February 2012.

[7] Magdalena M. Scharf, Moritz Bünemann, Jillian G. Baker, and Peter Kolb. Comparative Docking to Distinct G Protein–Coupled Receptor Conformations Exclusively Yields Ligands with Agonist Efficacy. *Molecular Pharmacology*, 96(6):851–861, December 2019.

[8] Chris de Graaf and Didier Rognan. Selective Structure-Based Virtual Screening for Full and Partial Agonists of the $B2$ Adrenergic Receptor. *Journal of Medicinal Chemistry*, 51(16):4978–4985, August 2008.

[9] Albert J. Kooistra, Rob Leurs, Iwan J. P. de Esch, and Chris de Graaf. Structure-Based Prediction of G-Protein-Coupled Receptor Ligand Function: A β-Adrenoceptor Case Study. *Journal of Chemical Information and Modeling*, 55(5):1045–1061, May 2015.

[10] Mireia Jiménez-Rosés, Bradley Angus Morgan, Maria Jimenez Sigstad, Thuy Duong Zoe Tran, Rohini Srivastava, Asuman Bunsuz, Leire Borrega-Román, Pattarin Hompluem, Sean A. Cullum, Clare R. Harwood, Eline J. Koers, David A. Sykes, Iain B. Styles, and Dmitry B. Veprintsev. Combined docking and machine learning identify key molecular determinants of ligand pharmacological activity on $B2$ adrenoceptor. *Pharmacology Research & Perspectives*, 10(5):e00994, 2022.

[11] Terry Kenakin. Efficacy at g-protein-coupled receptors. *Nature Reviews Drug Discovery*, 1(2):103–110, February 2002.

[12] Oliver Korb, Thomas Stützle, and Thomas E. Exner. Empirical Scoring Functions for Advanced Protein-Ligand Docking with PLANTS. *Journal of Chemical Information and Modeling*, 49(1):84–96, January 2009.

[13] Marcel L. Verdonk, Jason C. Cole, Michael J. Hartshorn, Christopher W. Murray, and Richard D. Taylor. Improved protein–ligand docking using GOLD. *Proteins: Structure, Function, and Bioinformatics*, 52(4):609–623, 2003.

[14] Jérémy Desaphy, Eric Raimbaud, Pierre Ducrot, and Didier Rognan. Encoding Protein–Ligand Interaction Patterns in Fingerprints and Graphs. *Journal of Chemical Information and Modeling*, 53(3):623–637, March 2013.

[15] Luca Chiesa and Esther Kellenberger. One class classification for the detection of $B2$ adrenergic receptor agonists using single-ligand dynamic interaction data. *Journal of Cheminformatics*, 14(1):74, October 2022.

[16] Jonas Goßen, Rui Pedro Ribeiro, Dirk Bier, Bernd Neumaier, Paolo Carloni, Alejandro Giorgetti, and Giulia Rossetti. AI-based identification of therapeutic agents targeting GPCRs: Introducing ligand type classifiers and systems biology. *Chemical Science*, 14(32):8651–8661, August 2023.

[17] Jiankun Lyu, Nicholas Kapolka, Ryan Gumpper, Assaf Alon, Liang Wang, Manish K. Jain, Ximena Barros-Álvarez, Kensuke Sakamoto, Yoojoong Kim, Jeffrey DiBerto, Kuglae Kim, Tia A. Tummino, Sijie Huang, John J. Irwin, Olga O. Tarkhanova, Yurii Moroz, Georgios Skiniotis, Andrew C. Kruse, Brian K. Shoichet, and Bryan L. Roth. AlphaFold2 structures template ligand discovery, December 2023.

[18] Fergus Imrie, Anthony R Bradley, and Charlotte M Deane. Generating property-matched decoy molecules using deep learning. *Bioinformatics*, 37(15):2134–2141, August 2021.

[19] Denis Schmidt, Viachaslau Bernat, Regine Brox, Nuska Tschammer, and Peter Kolb. Identifying Modulators of CXC Receptors 3 and 4 with Tailored Selectivity Using Multi-Target Docking. *ACS Chemical Biology*, 10(3):715–724, March 2015.

[20] Franck Da Silva, Jeremy Desaphy, and Didier Rognan. IChem: A Versatile Toolkit for Detecting, Comparing, and Predicting Protein–Ligand Interactions. *ChemMedChem*, 13(6):507–510, 2018.

[21] Gilles Marcou and Didier Rognan. Optimizing Fragment and Scaffold Docking by Use of Molecular Interaction Fingerprints. *Journal of Chemical Information and Modeling*, 47(1):195–207, January 2007.

2.3 Critical evaluation

Docking and interaction similarity methods have been described in the literature as two important tools for the design of new ligands targeting G-protein coupled receptors, with a desired efficacy profile. In this chapter similar experiments as the ones performed by de Graaf and Rognan[1], Costanzi and Vilar [2], and Kooistra et al.[3] were performed on a diverse set of targets, considering different interaction representations, and rescoring schemes. The objective of this study was to determine if interaction similarity was a sufficient criteria to guide the selection of ligands with a desired efficacy in docking-based virtual screening. Interaction similarity has been described as efficacy-biased based on the results obtained on the β2-adrenergic receptor (ADRB2)[1, 3, 4], but few information about its performance on other targets are available. The results described in this Chapter are in accordance with the data from the literature, showing a similar trend for other class A GPCRs. However, the magnitude of the hit rate was often significantly different between the targets, with the reference ADRB2 presenting significantly better hit rates than the other receptors.

ADRB2 represent one of the most important GPCR targets for in silico modelling and virtual screening, due to its wealth of structural and ligand data. Historically, ADRB2 has been greatly studied since it represented the first non-rhodopsin GPCR with a crystallographic structure [5], and the first GPCR with a crystallographic structure in the active state [6, 7]. Currently, multiple experimental structures of the receptor in both the active and the inactive state are deposited in the protein data bank (PDB). At the same time ADRB2 represents an important target for drug discovery, with many well characterized agonists and antagonists. Unlike other well studied targets its agonists and antagonists have both common and efficacy-specific structural features, making a ligand-based classification simple, but not trivial. Considering all of these aspects ADRB2 has represented the perfect proof of concept target to test new methods for GPCR drug discovery, especially concerning efficacy prediction [3, 4, 8–11]. However, its performances in the described benchmark study compared to other receptors, suggest that ADRB2 might not be a good representative target for efficacy prediction, leading to overoptimistic results for the tested methods. While future works should still use ADRB2 as a proof-of-concept target, a more thorough testing using more "difficult" receptors should be considered, to gain a better insight on the method's real performances.

Despite considering multiple GPCR targets, ligands, scoring and classification methods, the study presented clear limits which should be addressed by future research.

Concerning the tested in silico tools, this work mostly focused on interaction similarity methods[12, 13], while neglecting the comparison of different scoring functions and docking algorithms. Multiple works in the literature have reported an efficacy bias in docking scores depending on the activation state of the structure used for calculation[2, 14, 15]. The same result was not observed in this work using a different docking algorithm and scoring functions[16][17]. A more systematic analysis comparing different docking algorithms and scoring functions should be performed to determine whether the observed bias was specific of the studied scoring functions, or the results from the literature were dataset dependent.

The analysis of GPCR structures in different activation states brought some interesting insights regarding the role of the binding site conformation in docking, however, the role of crystallographic water molecules was completely neglected. For certain targets, especially the adenosine A2A receptor (AA2AR), water mediated interactions play a key role to correctly determine the affinity, and the efficacy, of a ligand [18]. Unfortunately, the study of the

crystallographic water molecules in the binding site would require to carefully determine their stability and role, as well as considering their influence and eventual displacement in docking calculations[19]. Both of these aspects were defined as outside the scope of this study, therefore structures presenting water mediated interactions between the receptor and the ligand were removed from the dataset, rather than including them after removing the water molecules. Future studies could focus on the role of these conserved crystallographic waters in both hit finding and efficacy selectivity.

Virtual screening was performed on a small dataset of drugs, clinical candidates, and crystallographic ligands, to reduce the risk of including mislabelled molecules in the dataset. This was done to assure consistency within the dataset, since the classification of a ligand as agonist or antagonist depends on the functional assay used to monitor biological response[20]. While results from public bioassays are available on ChEMBL[21], a consistent classification can only be achieved by comparing molecules tested using the same assay (Ca^{2+} mobilization, cAMP levels, receptor internalization, G-protein or β arrestin interaction, ecc.) and in the same experimental conditions (cell type, reference ligand, ligand concentration, time between measurements). Additionally, functional assay results might not distinguish between orthosteric or allosteric ligand, polluting the results for structure based models. The Guide to Pharmacology Database (GtoPDB)[22] is an expert curated database, with more reliable pharmacological labelling than data from assays, however the criteria for ligand inclusion in the dataset are unclear. To assure a consistency in labelling only drug, clinical candidates, and crystallographic ligands were considered for analysis. Known drugs and clinical candidates were extracted from the manually curated GPCRdb [23], which still presented few incorrectly labelled molecules. Future efforts for efficacy prediction using in silico tools should focus first on creating a high quality datasets of known agonists and antagonists, which could be used as references to build future prediction pipelines.

References

[1] Chris de Graaf and Didier Rognan. Selective Structure-Based Virtual Screening for Full and Partial Agonists of the $B2$ Adrenergic Receptor. *Journal of Medicinal Chemistry*, 51(16):4978–4985, August 2008.

[2] Stefano Costanzi and Santiago Vilar. In Silico screening for agonists and blockers of the $B2$ adrenergic receptor: Implications of inactive and activated state structures. *Journal of Computational Chemistry*, 33(5):561–572, February 2012.

[3] Albert J. Kooistra, Rob Leurs, Iwan J. P. de Esch, and Chris de Graaf. Structure-Based Prediction of G-Protein-Coupled Receptor Ligand Function: A β-Adrenoceptor Case Study. *Journal of Chemical Information and Modeling*, 55(5):1045–1061, May 2015.

[4] Albert J. Kooistra, Henry F. Vischer, Daniel McNaught-Flores, Rob Leurs, Iwan J. P. de Esch, and Chris de Graaf. Function-specific virtual screening for GPCR ligands using a combined scoring method. *Scientific Reports*, 6(1):28288, September 2016.

[5] Vadim Cherezov, Daniel M. Rosenbaum, Michael A. Hanson, Søren G. F. Rasmussen, Foon Sun Thian, Tong Sun Kobilka, Hee-Jung Choi, Peter Kuhn, William I. Weis, Brian K. Kobilka, and Raymond C. Stevens. High-Resolution Crystal Structure of an Engineered Human β $_2$ -Adrenergic G Protein–Coupled Receptor. *Science*, 318(5854):1258–1265, November 2007.

[6] Søren G. F. Rasmussen, Brian T. DeVree, Yaozhong Zou, Andrew C. Kruse, Ka Young Chung, Tong Sun Kobilka, Foon Sun Thian, Pil Seok Chae, Els Pardon, Diane Calinski, Jesper M. Mathiesen, Syed T. A. Shah, Joseph A. Lyons, Martin Caffrey, Samuel H. Gellman, Jan Steyaert, Georgios Skiniotis, William I. Weis, Roger K. Sunahara, and Brian K. Kobilka. Crystal structure of the $B2$ adrenergic receptor–Gs protein complex. *Nature*, 477(7366):549–555, September 2011.

[7] Søren G. F. Rasmussen, Hee-Jung Choi, Juan Jose Fung, Els Pardon, Paola Casarosa, Pil Seok Chae, Brian T. DeVree, Daniel M. Rosenbaum, Foon Sun Thian, Tong Sun Kobilka, Andreas Schnapp, Ingo Konetzki, Roger K. Sunahara, Samuel H. Gellman, Alexander Pautsch, Jan Steyaert, William I. Weis, and Brian K. Kobilka. Structure of a nanobody-stabilized active state of the $B2$ adrenoceptor. *Nature*, 469(7329):175–180, January 2011.

[8] Chris de Graaf, Nicolas Foata, Ola Engkvist, and Didier Rognan. Molecular modeling of the second extracellular loop of G-protein coupled receptors and its implication on structure-based virtual screening. *Proteins: Structure, Function, and Bioinformatics*, 71(2):599–620, 2008.

[9] Mireia Jiménez-Rosés, Bradley Angus Morgan, Maria Jimenez Sigstad, Thuy Duong Zoe Tran, Rohini Srivastava, Asuman Bunsuz, Leire Borrega-Román, Pattarin Hompluem, Sean A. Cullum, Clare R. Harwood, Eline J. Koers, David A. Sykes, Iain B. Styles, and Dmitry B. Veprintsev. Combined docking and machine learning identify key molecular determinants of ligand pharmacological activity on $B2$ adrenoceptor. *Pharmacology Research & Perspectives*, 10(5):e00994, 2022.

[10] Luca Chiesa and Esther Kellenberger. One class classification for the detection of $B2$ adrenergic receptor agonists using single-ligand dynamic interaction data. *Journal of Cheminformatics*, 14(1):74, October 2022.

[11] Nicolas Panel, Duc Duy Vo, Nour Aldin Kahlous, Harald Hübner, Stephanie Tiedt, Pierre Matricon, Jody Pacalon, Oliver Fleetwood, Stefanie Kampen, Andreas Luttens, Lucie Delemotte, Jan Kihlberg, Peter Gmeiner, and Jens Carlsson. Design of Drug Efficacy Guided by Free Energy Simulations of the β_2 -Adrenoceptor. *Angewandte Chemie*, 135(22):e202218959, May 2023.

[12] Márton Vass, Albert J Kooistra, Tina Ritschel, Rob Leurs, Iwan JP de Esch, and Chris de Graaf. Molecular interaction fingerprint approaches for GPCR drug discovery. *Current Opinion in Pharmacology*, 30:59–68, October 2016.

[13] Eelke B. Lenselink, Willem Jespers, Herman W. T. van Vlijmen, Adriaan P. IJzerman, and Gerard J. P. van Westen. Interacting with GPCRs: Using Interaction Fingerprints for Virtual Screening. *Journal of Chemical Information and Modeling*, 56(10):2053–2060, October 2016.

[14] Dahlia R. Weiss, SeungKirl Ahn, Maria F. Sassano, Andrew Kleist, Xiao Zhu, Ryan Strachan, Bryan L. Roth, Robert J. Lefkowitz, and Brian K. Shoichet. Conformation Guides Molecular Efficacy in Docking Screens of Activated β-2 Adrenergic G Protein Coupled Receptor. *ACS Chemical Biology*, 8(5):1018–1026, May 2013.

[15] David Rodríguez, Zhang-Guo Gao, Steven M. Moss, Kenneth A. Jacobson, and Jens Carlsson. Molecular Docking Screening Using Agonist-Bound GPCR Structures: Probing the A $_{2A}$ Adenosine Receptor. *Journal of Chemical Information and Modeling*, 55(3):550–563, March 2015.

[16] Oliver Korb, Thomas Stützle, and Thomas E. Exner. Empirical Scoring Functions for Advanced Protein-Ligand Docking with PLANTS. *Journal of Chemical Information and Modeling*, 49(1):84–96, January 2009.

[17] Marcel L. Verdonk, Jason C. Cole, Michael J. Hartshorn, Christopher W. Murray, and Richard D. Taylor. Improved protein–ligand docking using GOLD. *Proteins: Structure, Function, and Bioinformatics*, 52(4):609–623, 2003.

[18] Jonathan S. Mason, Andrea Bortolato, Dahlia R. Weiss, Francesca Deflorian, Benjamin Tehan, and Fiona H. Marshall. High end GPCR design: Crafted ligand design and druggability analysis using protein structure, lipophilic hotspots and explicit water networks. *In Silico Pharmacology*, 1(1):23, December 2013.

[19] Niu Huang and Brian K. Shoichet. Exploiting Ordered Waters in Molecular Docking. *Journal of Medicinal Chemistry*, 51(16):4862–4865, August 2008.

[20] Terry P. Kenakin. Cellular assays as portals to seven-transmembrane receptor-based drug discovery. *Nature Reviews Drug Discovery*, 8(8):617–626, August 2009.

[21] Anna Gaulton, Anne Hersey, Michał Nowotka, A. Patrícia Bento, Jon Chambers, David Mendez, Prudence Mutowo, Francis Atkinson, Louisa J. Bellis, Elena Cibrián-Uhalte, Mark Davies, Nathan Dedman, Anneli Karlsson, María Paula Magariños, John P. Overington, George Papadatos, Ines Smit, and Andrew R. Leach. The ChEMBL database in 2017. *Nucleic Acids Research*, 45(D1):D945–D954, January 2017.

[22] Simon D Harding, Jane F Armstrong, Elena Faccenda, Christopher Southan, Stephen P H Alexander, Anthony P Davenport, Michael Spedding, and Jamie A Davies. The IUPHAR/BPS Guide to PHARMACOLOGY in 2024. *Nucleic Acids Research*, 52(D1):D1438–D1449, January 2024.

[23] Alexander S. Hauser, Misty M. Attwood, Mathias Rask-Andersen, Helgi B. Schiöth, and David E. Gloriam. Trends in GPCR drug discovery: New agents, targets and indications. *Nature Reviews Drug Discovery*, 16(12):829–842, December 2017.

Chapter 3

Development of a new tool for efficacy-driven drug discovery

Interaction similarity methods, or the selection of relevant docking poses based on the presence of key interactions, represent a cornerstone of structure-based drug design. Such approaches have been used for the discovery of new ligands with a desired efficacy profile, making them an attractive tool for GPCR drug discovery. The reliance on crystallographic data however is a limit of structure-based drug design since it represents the protein structure and a ligand's binding mode as static, while in reality they both have a dynamic nature. In this chapter a new method for the post-processing of docking results, called InteractionOCSVM, is described. This new approach uses machine learning to identify ligands whose binding mode is comparable with an ensemble of references extracted from molecular dynamics simulations. Given the link between binding mode and efficacy for GPCRs, the developed method was evaluated as a tool for efficacy-driven drug design.

The first section of the chapter is dedicated to the description and development of the method. A small proof-of-concept study aimed at the selective detection of known agonists of the β2-adrenergic receptor against known antagonists, or experimentally confirmed non-binders is described. Different aspects of model training, concerning the quality of the interaction data, the selection of the correct interaction representation, or the role of interactions sampling were tackled in order to determine guidelines for the training of new models for prospective virtual screening applications.

The second section is dedicated to the benchmark of the method using the dataset developed in Chapter 2. InteractionOCSVM was tested against a panel of different receptors, and using reference ligands with different efficacies. This second study aims at overcoming the limits of the original proof-of-concept evaluation. While the method generally proved to be insufficiently precise in discarding decoys against active molecules, it performed well in the identification of meaningful docking poses, which could then be ranked using an orthogonal method, like scoring functions.

3.1 One class classification for the detection of β2 adrenergic receptor agonists using single-ligand dynamic interaction data

This section was previously published in Journal of Cheminformatics(*J Cheminform* 14, 74 (2022)). All supporting informations are available on-line at: `https://doi.org/10.1186/s13321-022-00654-z`

Chiesa and Kellenberger
Journal of Cheminformatics (2022) 14:74
https://doi.org/10.1186/s13321-022-00654-z

Journal of Cheminformatics

RESEARCH Open Access

One class classification for the detection of β2 adrenergic receptor agonists using single-ligand dynamic interaction data

Luca Chiesa and Esther Kellenberger*

Abstract

G protein-coupled receptors are involved in many biological processes, relaying the extracellular signal inside the cell. Signaling is regulated by the interactions between receptors and their ligands, it can be stimulated by agonists, or inhibited by antagonists or inverse agonists. The development of a new drug targeting a member of this family requires to take into account the pharmacological profile of the designed ligands in order to elicit the desired response. The structure-based virtual screening of chemical libraries may prioritize a specific class of ligands by combining docking results and ligand binding information provided by crystallographic structures. The performance of the method depends on the relevance of the structural data, in particular the conformation of the targeted site, the binding mode of the reference ligand, and the approach used to compare the interactions formed by the docked ligand with those formed by the reference ligand in the crystallographic structure.

Here, we propose a new method based on the conformational dynamics of a single protein-ligand reference complex to improve the biased selection of ligands with specific pharmacological properties in a structure-based virtual screening exercise. Interactions patterns between a reference agonist and the receptor, here exemplified on the β2 adrenergic receptor, were extracted from molecular dynamics simulations of the agonist/receptor complex and encoded in graphs used to train a one-class machine learning classifier. Different conditions were tested: low to high affinity agonists, varying simulation duration, considering or ignoring hydrophobic contacts, and tuning of the classifier parametrization. The best models applied to post-process raw data from retrospective virtual screening obtained by docking of test libraries effectively filtered out irrelevant poses, discarding inactive and non-agonist ligands while identifying

agonists. Taken together, our results suggest that consistency of the binding mode during the simulation is a key to the success of the method.

3.1.1 Introduction

G protein-coupled receptors (GPCRs) constitute the largest family of membrane proteins in humans. GPCRs mediate the cellular response to outside stimuli by initiating specific signaling transduction pathways [1]. As key player in many physiological processes, GPCRs have been largely exploited as drug targets [2]. The pharmacological profile of a GPCR ligand, whether endogenous or xenobiotic, is characterized according to the response it induces in cell [3]: agonist triggers signaling, unlike antagonist and inverse agonist (henceforth referred to as antagonist too, for simplicity). The computer-assisted discovery of GPCR ligands for therapeutic purposes therefore requires the development of predictive models of the pharmacological profile. Machine learning well suits this purpose by extracting complex relationship from experimental data and classification algorithms are typically adapted to distinguish agonists from antagonists [4, 5]. However, their application is limited to well-studied protein targets with many known agonists and antagonists. Methods based on the three-dimensional structure of the protein-ligand complex can circumvent this limitation. Since the resolution of the first structure of a GPCR bound to an antagonist in 2007 [6] and coupled to G-protein in 2011 [7], the continuous elucidation of new protein-ligand structures has led to a golden age for structure based drug design projects [8–10]. The abundant structural information also highlighted the GPCR's structural plasticity. GPCRs adopt a wide spectrum of conformational states, ranging from inactive to active ones.

In the search for novel and original ligands of GPCRs, ligand/protein docking accounts for the most common tool for virtual screening of chemical libraries [11]. Docking predicts the binding mode and scores the solutions. The score can be interpreted in terms of ligand binding, but not ligand function.

While agonist ligands are known to have a higher binding affinity for the active state, and antagonists for the inactive state[12], the structures of the binding pocket in both states tend to be very similar [11]. Agonist ligands have thus been identified by virtual screening using an inactive state of the receptor[13–17] and antagonists using an active state [18]. The analysis of the binding mode, by comparison with reference crystallographic structures, better accounts for the pharmacophoric specificity of agonists and antagonists [12, 19]. For example Kooistra et al. [12] investigated the use of experimental reference binding modes to functionally bias the screening against the β-adrenoreceptor family. Their retrospective analysis showed that it is possible to improve the retrieval of agonist or antagonist ligands from a test library containing active molecules and decoys by assessing the similarity between the binding mode generated by docking with the crystallographic binding mode, here by the comparison of interaction fingerprints (IFP). By considering the enrichment factor at 1% false positive rate, a reference agonist IFP increased the proportion of agonists among the true positives, and similarly, a reference antagonist IFP increased the proportion of antagonists. Despite the overall improvement different reference ligands, having different IFPs, yielded different hit lists. Even the binding mode of the same ligand in multiple crystallographic structures can vary [12, 20]. The changes in binding mode can be explained by the inherent flexibility of GPCRs[3] or by different conditions of crystallization (e.g., thermostabilizing mutations, coupled nanobodies) [21]. Selecting a reference structure containing the relevant interaction patterns plays a key role in virtual screening [12]. The best reference can be

identified by retrospective virtual screening experiments, yet such approach requires data from known agonists, antagonists, and inactive molecules [22].

The β2 adrenergic receptor (ADRB2) has been the first receptor crystalized bound to a ligand and coupled to G-protein. At the time of writing the structure of the receptor has been elucidated by thirty-seven crystallographic structures. ADRB2 has been thoroughly studied due to the crucial role of both its agonist and antagonists in drug discovery. The receptor is present mostly in lung tissues, where its signaling regulates smooth muscle relaxation and bronchodilation. Agonists are used to treat respiratory disorders[23] such as asthma[24] and chronic obstructive pulmonary disease (COPD)[25]. Antagonists, more commonly known as β-blockers, are used to treat cardiovascular diseases by targeting the β1 adrenergic receptors in heart tissues. The action of β-blockers on ADRB2 is mostly linked to undesirable side-effects [26]. Thanks to the large amount of information regarding both its structure and pharmacology, ADRB2 has been used in computational works as a model system [12, 22].

Key interactions for GPCR activation are inferred from the analysis of experimental structures of the receptor bound to different agonists, provided experimental structures are available for several agonist/GPCR complexes[27]. Considering a single agonist-GPCR structure, molecular dynamics (MD) calculations can evaluate intermolecular interactions, based on the principle that the key interactions are better conserved during simulation [27, 28]. Here, we consider a single agonist-GPCR structure and propose to combine MD simulation and machine learning to bias the search towards agonist ligands by virtual screening of chemical libraries using docking. The proof of concept was performed on ADRB2, for which we constructed two suitable test sets: A library containing ADRB2-targeting drugs or clinical candidates with a well characterized pharmacological profile, and a library composed of virtual hits selected by docking into ADRB2 orthosteric site and validated/invalidated for agonist activity [29]. A one class support vector machine (OCSVM) [30] model was trained on the ADRB2-agonist interactions extracted from a single MD trajectory and represented by graphs, allowing a binding mode definition based on interactions conservation. The classifier was used to filter ligand–ADRB2 docking generated poses. The ability of each classifier to distinguish between agonist and not agonist was determined considering the fraction of true agonists between the selected molecules (precision) and the fraction of true agonists selected from the test dataset (recall). The effect of simulation length, interaction types, combinations of multiple references, and training method were taken into consideration by training multiple classifiers, with different combinations of parameters. The results of the analysis can be used as guidelines for the training of new models targeting different GPCRs. The proposed methodology can be applied to find new agonists for targets with a limited number of known reference ligands.

3.1.2 Results and Discussion

Introduction

GPCRs classically mediate responses to agonist binding by coupling to effector, typically a G-protein. The activation mechanism and the resulting signaling are intimately connected to the conformation of the receptor, which itself is stabilized or induced by the interactions formed with the agonist. The process is well characterized for ABDR2 activation by the endogenous agonists epinephrine and norepinephrine. For example, hydrogen bonding to

Ser207 has been proposed as key determinant of agonist binding [31]. In the present work, we consider the relationships between structure and function to be unknown and assess whether a machine learning model trained on the structural information of a single ligand is able to determine key agonist-related interactions, assuming that the patterns of interactions that are the most conserved during MD simulation of the receptor-agonist complex are also the most relevant for signaling. The machine learning model is applied to assess the compatibility of the interaction patterns predicted by docking for any ligand of the receptor with the common definition determined from the training data on the reference agonist. A ligand evaluated positively thus potentially exhibits a mode of binding similar to that of the reference agonist and is supposed to trigger the same pharmacological response.

Method development

Reference agonists and MD simulations Three different agonists were selected for generating the training sets from published or publicly available molecular dynamics simulations [32, 33]: Epinephrine, hydroxybenzyl isoproterenol (HBI) and BI167107 (Figure 3.1.1A). Epinephrine is a low affinity agonist of ABDR2 [34]. As a member of the endogenous neurotransmitters of catecholamine family, it contains the 1,2-dihydroxybenzene group (i.e., catechol) and a side-chain amine. The side-chain amine is part of a 2-aminoethanol group common to ADRB2 agonists. HBI is a high potency binder of ADRB2 [34]. It is a superstructure of epinephrine that features a phenol group at the end of the sidechain. BI167107 is an ultra-high affinity binder of ADRB2 [34]. Its structure shares with epinephrine and HBI the 2-aminoethanol group, but not the catechol which is replaced by a different functional group, with one of the two hydroxyl groups replaced by a 3-morpholinone like moiety.

The crystallographic structures of ADBR2 suggest, for the three agonists, a strong anchoring of their 2-aminoethanol group by hydrogen bonding to residues in TM2 (Asp113) and TM7 (Asn312), as well as an interaction between the hydroxyl of catechol, or its substitute in BI167107, with a serine of TM5 (Figure 3.1.1A) [7, 34]. The interactions formed between epinephrine and ADRB2 are not strictly conserved during the MD simulation (Figure 3.1.1B and S1). Moreover, new interactions are observed, including a hydrogen bond with Ser207. Binding mode variations are also marked in the HBI-ADRB2 MD trajectory, involving additional receptor residues in the interactions formed with the agonist. While epinephrine-ADBR2 and HBI-ADRB2 MD trajectories both describe the active state of the receptor, the MD trajectory of BI167107-ADRB2 [33] simulates the transition of the receptor from the active state to the inactive state, and therefore was split to distinguish the binding of BI167107 to the active state (BI167107-A) and to the inactive state (BI167107-I), thus giving two separated datasets. Importantly, hydrogen bonding to Ser207 is lost in BI167107-I. The shift from the active to the inactive state was observed despite the receptor being bound to BI167107, an agonist. The absence of an intracellular binding partner, which stabilizes the active state conformation, explains the transition[35]. The fully inactive state of the receptor was observed only after microseconds of simulation, a longer timescale than the one used in the other selected simulations (Table 3.1.1). The study of BI167107 bound to both the active and inactive state of the receptor was used to investigate the differences between models built using structural information close and far from the active state crystallographic structure.

A total of four training sets were used in this work. The Table 3.1.1 characterises the MD trajectories with the number of simulations, the simulations length, and the sampling

Table 3.1.1: MD simulations used for the training dataset. The simulation length refers to a single replica and not the entire trajectory.

Reference trajectory	Number of replicas	Number of frames	Simulation length (ns)	Interval between writing structures (ns)
Epinephrine	3	2500	500	0.200
HBI	3	2500	500	0.200
BI167107-A	1	1944	350	0.180
BI167107-I	1	7500	1350	0.180

frequency.

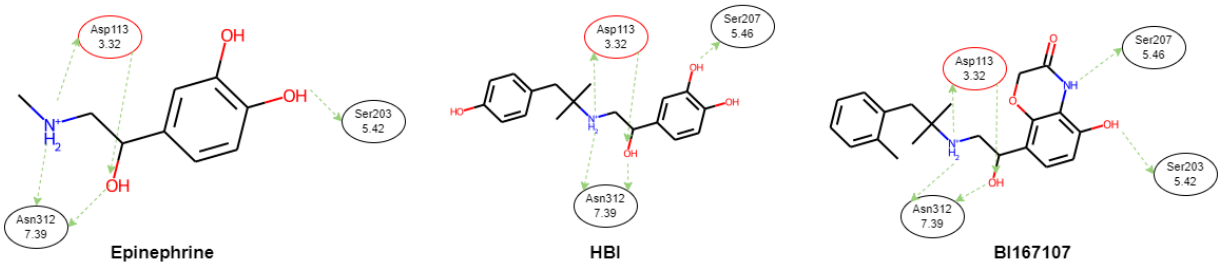
Interaction graph For each of the four reference trajectories (epinephrine, HBI, BI167107-A and BI167107-I), the following interactions were detected between ADRB2 and the reference agonist using IChem [36] : Hydrogen bonds, ionic bonds, aromatic stacking and hydrophobic contacts. Each interaction is represented by a triplet of interactions pseudo-atoms (IPA), one placed on each of the two atoms involved in the interaction (ligand IPA and protein IPA), and a third one in the middle point of the segment they define (centre IPA). More details about IChem detection and encoding of interactions are given in the Material and Methods section.

The average number of IPAs is shown on Figure 3.1.2. Aromatic interactions are rare in the four reference trajectories. The other polar interactions, hydrogen and ionic bonds, represents together ca 13 (for epinephrine) to ca 25 IPAs (for HBI) per interactions pattern. The number of IPAs corresponding to hydrophobic contacts is equal to or greater than the number of IPAs corresponding to polar interactions, except in the epinephrine trajectory. Since the description of hydrophobic contacts does not rely on a consensual physical model, unlike that of polar interactions, we tested three conditions: the two IChem definitions of hydrophobic contacts and the exclusion of hydrophobic contacts. Hence, for each of the four reference trajectories, three sets of IPAs were considered: all interactions with default definition of hydrophobic contacts (Hyd), all interaction with stricter definition of hydrophobic contacts (Newhyd), and polar interactions only (Polar).

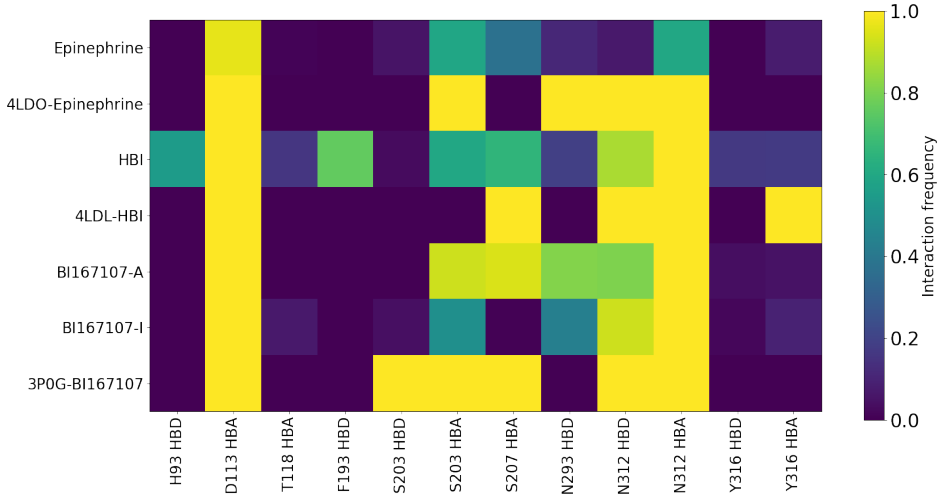
Interaction graphs (IGs) were generated from the detected interactions, with the IPAs representing the nodes. To include the largest number of topological information all nodes were connected to each other, forming a complete graph. The nodes were labelled according to the IChem interaction type and the represented position (protein, ligand, centre). The edges were described by a weighted adjacency matrix, the weights being equal to the Euclidean distances between the IPAs rounded to the nearest Å. The distance between two overlapping IPAs was set to 0.1, to establish an edge between the nodes.

Graph similarity assessment is a key problem in graph theory, which is why different types of graph kernels have been developed [37]. In this work graph similarity was evaluated using the shortest path kernel[38]. Graphs similarity is based on the common occurrence in the graphs of identical shortest paths between two nodes with the same label. In practice, a

Figure 3.1.1: (A) Structure of the three reference agonists showing the hydrogen bonds observed in the crystallographic structures. The ADRB2 residues are identified by their sequence number and by Ballesteros-Weinstein notation. (B) IFPs with hydrogen bonds formed between the agonist and ABDR2 in the trajectory and crystallographic structures. HBA indicates interactions with the protein acting as the hydrogen bond acceptor, HBD indicates interactions with the protein acting as the hydrogen bond donor. Crystallographic structure IFPs are indicated by their PDBID.

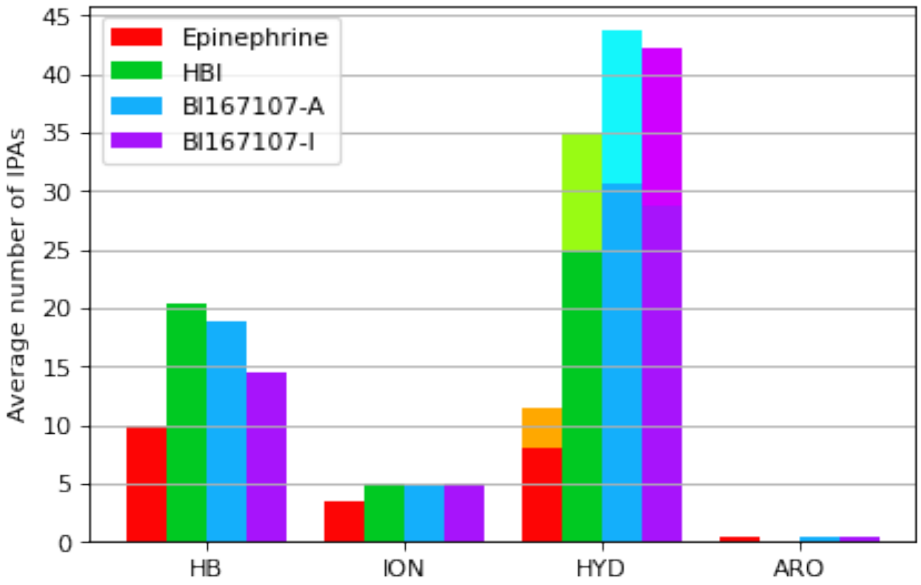


(A)



(B)

Figure 3.1.2: Average distribution of the interaction pseudo-atoms types during the MD trajectories. HB: hydrogen bonds, ION: ionic interactions, HYD: hydrophobic contacts, ARO: $\pi - \pi$ interactions. The lighter-coloured and legend-coloured parts of HYD bars corresponds to the default (Hyd) and stricter (Newhyd) definition of hydrophobic contacts, respectively.



graph was converted into a fingerprint containing the frequency of each shortest path. Two fingerprints were compared by calculating the cosine similarity and the vector dot product, yielding a normalized and a non-normalized (NN) score, respectively. Thus, for each of the three interaction definitions (Hyd, NewHyd and Polar), both normalized and non-normalized similarity scores were calculated.

Model training Each MD trajectory contains information on a single class. To process data from a single class it is necessary to use algorithms built to handle limited information. One class classification or outlier detection [39, 40] is a family of semi-supervised classification algorithms whose focus is not to determine the boundary between two classes, but it is to learn a common definition describing the training instances. Such algorithms have already been proposed in ligand-based drug discovery [41], detection of druggable protein pockets [42], and in other fields of chemistry [43] characterized by a low availability of experimental data. Multiple outlier detection algorithms are available based on different well known machine learning models: OCSVM[30], isolation forest[44], k-nearest neighbours (KNN).

The OCSVM [30] algorithm was used to find a criterion to define conditions for the inclusion of a new graph into the ensemble of interactions observed during the reference trajectory. OCSVM, like other support vector machine-based classification methods, defines a class using a support vector (SV). In the case of OCSVM the SV represents the hyperplane with the maximum distance from the origin in the transformed feature space.

Training the OCSVM algorithm requires the user to set the hyperparameter ν to a value between 0 and 1. Two key quantities correspond to ν: The upper bound for the fraction of training instances classified as outliers during training, and the lower bound for the fraction of

instances defining the SV. Since the effective tuning of the ν value is not possible in absence of an external dataset, we followed heuristic approaches focusing on removing outliers or estimating the percentage of outliers in the training set.

Outliers in the training set were identified using the average shortest path kernel similarity value to the KNN, with K corresponding to 3% of the total number of training instances [45]. The first approach called quick model selection 2 (QMS2)[46], removes from the training set both outliers and instances on the boundaries of the distribution. Since the new training set should contain only true inliers the fraction of classification errors during training (ν) is set to a small value of 0.01. The second approach uses the median absolute deviation (MAD) to determine the number of outliers in the training set. The upper bound to the fraction of training errors is set to be equal to the number of instances with a distance from the median KNN similarity larger than three times the MAD divided by the total number of instances in the training set.

Models using normalized kernels were trained using both QMS2 and MAD, while the models based on NN scores were trained using QMS2 only. In short, for each of the four reference trajectories (epinephrine, HBI, BI167107-A and BI167107-I), nine models were constructed, combining three different sets of interactions (HYD or NEWHYD or POLAR) and three parametrizations (QSM2 or MAD or NN).

Method validation

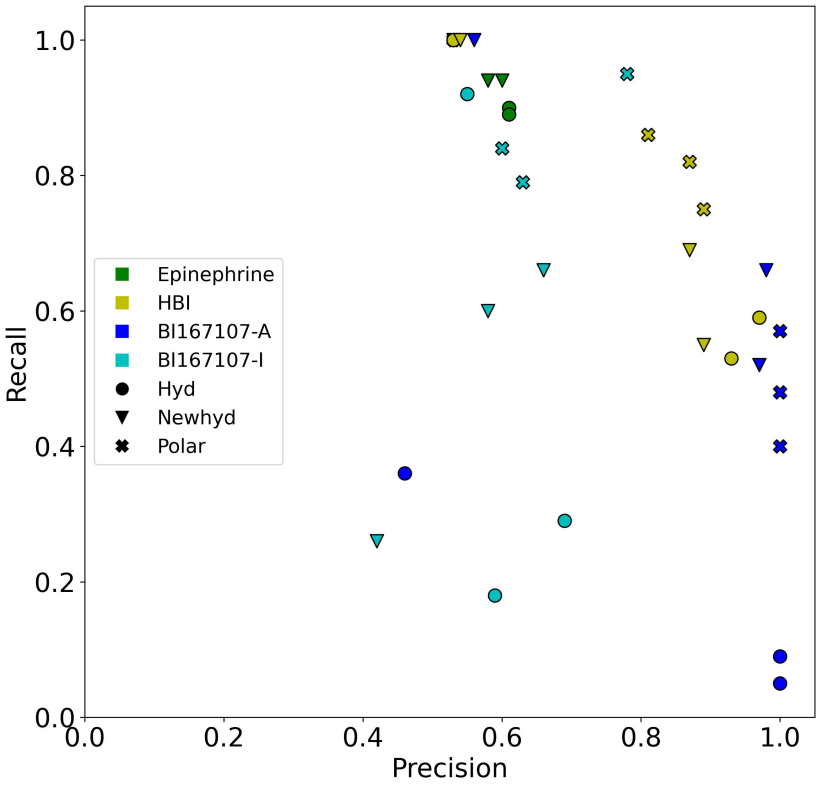
Introduction The thirty-six trained OCSVM models were used to screen two small chemical libraries to evaluate their ability in recognizing agonist ligands. One library contains known agonists and antagonists, while the second contains experimentally validated agonists and inactive molecules.

For each molecule in the libraries, the docking into the three crystallographic structures of ABDR2 issued ten poses each. The calculations were repeated for up to nine representative structures of the receptor agonist binding site. Site definition and clustering of epinephrine-ADRB2, HBI-ADRB2, and BI167017-ADRB2 trajectories are detailed in Materials and Methods. Each model was used for post-processing only on the docking poses generated using a protein structure extracted from the respective training set. If the docking calculations generated at least one binding mode selected by an OCSVM model, then the molecule is classified as an agonist. Therefore, the performances discussed below apply to the entire virtual screening process, not only the classifiers.

Validation with agonist/antagonist dataset The dataset contains 19 agonists and 17 antagonists extracted from the literature (SI Table 1). All 36 molecules have been largely studied and their pharmacology is well known. The dataset was also curated to contain compounds with different chemical structures. A total of 24 different scaffolds are represented in the dataset, four scaffolds are shared by multiple compounds. Benzene is the only common scaffold between agonists and antagonists. All the molecules present the 2-ethanolamine group, except the agonist dobutamine. The agonists all present a benzene ring with a functional group able to form hydrogen bonds, like the catechol group of epinephrine, except tulobuterol whose benzene ring is substituted with a chlorine atom.

The models performed differently depending on the definition of interactions, the normalization of the kernel, and the trajectory used as reference (Figure 3.1.3 and SI Table 3-4). However, it is possible to observe some general trends: Models based on epinephrine

Figure 3.1.3: Average performance of OCSVM models obtained on the agonist/antagonist dataset from docking using the representative structures.



are all inefficient, predicting most to all molecules as agonists. HBI-based models correctly identified most of the true agonists and misclassified a limited number of antagonists. Training with the BI167107-A dataset generated models with a high precision, but with a tendency to discard many true agonists. Including hydrophobic contacts in the IGs was detrimental to NN models, causing all ligands to be classified as agonists. Likewise, the performances of models based on HBI and BI167107 overall improved if only polar interactions were included in the IGs.

The four best models for each reference trajectory are presented on Table 3.1.2. They were selected according to the average F1 measure calculated from docking poses using each representative structure, after discarding models which selected all ligands as agonists. For each model, the average performance over a set of representative conformations, selected from the respective training set, gives an estimate of the classifier behaviour. The four best HBI-based models retrieved three quarters or more of the agonists, while showing very few false positives. The best BI167107-based models discriminated between agonists and antagonists, provided they were trained from the portion of the trajectory simulating the active state of the receptor. BI167107-I-based models indeed incorrectly predicted many antagonists as agonists. The decrease of precision between models based on BI167107-A

and BI167107-I is consistent with the vanishing of the key interaction with Ser207 upon deactivation of the receptor during the BI167107-ADRB2 simulation.

Table 3.1.2: Performance of the best four models obtained for each reference trajectory on the agonist/antagonist dataset from docking data obtained using the representative structures or the crystallographic structure

Reference trajectory	Model	Average			Crystallographic			Ensemble		
		Precision	Recall	Size	Precision	Recall	Size	Precision	Recall	Size
Epinephrine	Hyd–MAD	0.61	0.90	28/36	0.62	0.95	29/36	0.53	0.95	34/36
	Hyd–QMS2	0.61	0.89	28/36	0.58	0.95	31/36	0.55	0.95	33/36
	Newhyd–MAD	0.58	0.94	31/36	0.53	1.00	36/36	0.53	1.00	36/36
	Newhyd–QMS2	0.60	0.94	30/36	0.54	1.00	35/36	0.53	1.00	36/36
HBI	Polar–NN	0.87	0.82	18/36	0.94	0.79	16/36	0.73	1.00	26/36
	Polar–MAD	0.81	0.86	20/36	1.00	0.74	14/36	0.68	1.00	28/36
	Polar–QMS2	0.89	0.75	16/36	1.00	0.74	14/36	0.76	1.00	25/36
	Newhyd–MAD	0.87	0.69	15/36	1.00	0.68	13/36	0.72	0.95	25/36
BI167107-A	Polar–NN	1.00	0.57	11/36	1.00	0.89	17/36	1.00	0.84	16/36
	Polar–MAD	1.00	0.40	8/36	1.00	0.26	5/36	1.00	0.74	14/36
	Newhyd–QMS2	0.97	0.52	10/36	–	0.00	0/36	0.94	0.89	18/36
	Newhyd–MAD	0.98	0.66	13/36	–	0.00	0/36	0.94	0.89	18/36
BI167107-I	Polar–NN	0.78	0.95	23/36				0.79	1.00	24/36
	Polar–QMS2	0.60	0.84	27/36				0.59	0.89	29/36
	Polar–MAD	0.63	0.79	24/36				0.57	0.84	28/36
	Newhyd–MAD	0.66	0.66	19/36				0.67	0.84	24/36

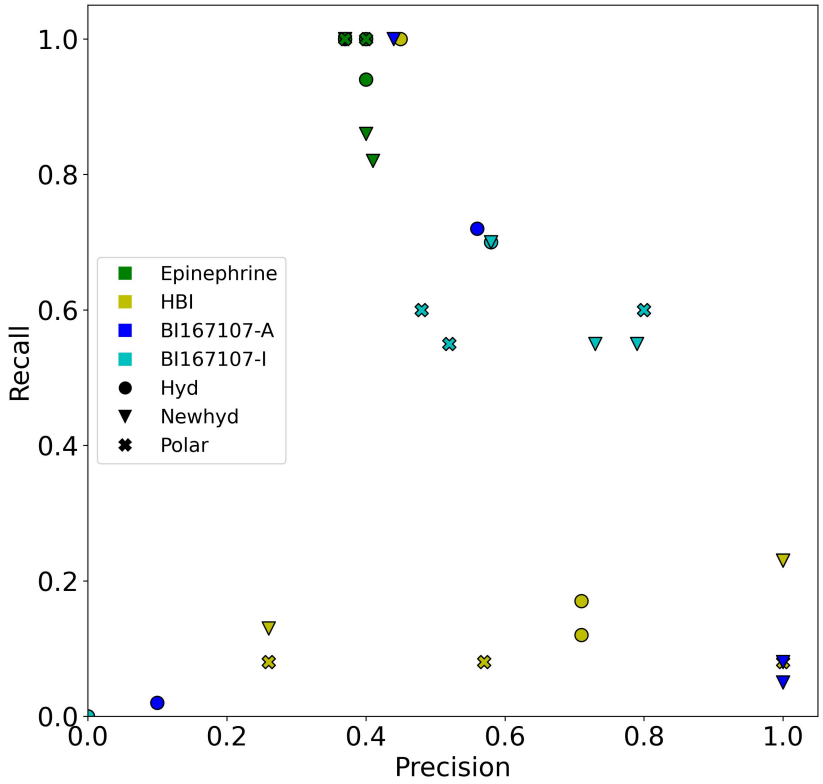
Size refers to the fraction of molecules classified as agonists. Average refers to the mean of the values obtained for the representative structures considered individually. Ensemble refers to the value obtained by merging data from all representative structures. Models discarding all docking poses are indicated as -

For epinephrine and HBI reference trajectories, the results were comparable whether the considered docking poses use the crystallographic structure or a representative structure, on average. Similar findings were made for BI167107-A models based on polar interactions. The BI167107-A models including hydrophobic contacts did not select any molecule. An inappropriate conformation of the hydrophobic sub-pockets was assumed based on the two following evidences: The hydrophobic contacts between BI167107 and ADBR2 in the crystallographic structure and during the MD simulation differ (SI Figure 1); the binding site underwent significant structural changes during the simulation (SI Figure 2), with an average root mean squared deviation (RMSD) from the initial crystallographic structure above 3 Å while the average pairwise RMSD between the frames is below 2.5 Å (SI Figure 3 – 4). For the sake of comparison, the RMSD values for epinephrine-ADBR2 and HBI-ADBR2 complexes are below 2.5 Å (SI Figures 5,6).

Since the performance of a model depends on the receptor's structure, we also considered ensemble docking, by merging the docking poses obtained using all representative structures of the reference trajectory (Table 3.1.2). Overall, using all the representative structures together produced F1-measures comparable to or better than the average obtained by considering the representative structures separately. Increasing the number of evaluated poses improved recall, yet often at the expense of precision. The best improvement concerns BI167107-A high precision models, which showed a significant increase in recall (e.g. from 0.57 to 0.84 for the Polar-NN model) with no loss of precision (e.g., 1.00 for the Polar-NN model). Models characterized by a looser definition of agonist binding mode, such as HBI-based models and to a lesser extent BI167107-I, tend to include a larger number of false positives.

Validation with agonist/inactive dataset To test the ability of the OCSVM models in correctly identifying inactive molecules as non-agonist a dataset of experimentally validated

Figure 3.1.4: Average performance of OCSVM models on the agonist/inactive dataset from docking using the representative structures. Note that one Hyd and the three Polar models of BI167107-A are not visible since no molecules were predicted as active.



true agonists and true inactive molecules was selected from the literature [29]. The library consists of ten agonists and 17 inactive molecules. All the 27 molecules are virtual hits of a structure-based screening of the ZINC12 library. They represent 17 different scaffolds, including 3 which are specific for agonists and 14 are specific for inactive molecules. All active molecules and two inactive molecules contain the 2-aminoethanol group which is present in ADRB2 ligands and allows their anchoring to both TM3 and TM7. The agonists share a second common characteristic: all but one contain a halogenated aromatic ring instead of the catechol of epinephrine. The halogen atoms are assumed to interact with the subpocket formed by TM5, TM4, and TM3, for receptor activation upon agonist binding. With halogen interactions not being taken into account by IChem, the present dataset is highly challenging for the models. Overall, the performance of the models is worse on the agonist/inactive dataset than on the agonist/antagonist dataset (Figure 3.1.3 and SI Table 5-6). Again, epinephrine-based models are not predictive, classifying all or nearly all molecules of the agonist/inactive dataset as agonists.

The four best models for each reference trajectory are presented on Table 3.1.3. They were selected according to the average F1 measure calculated from docking poses using each

representative structure, after discarding models which select all ligands as agonists. For each model, the average performance over a set of representative conformations, selected from the respective training set, gives an estimate of the classifier behaviour. The best HBI-based models retrieved on average a quarter of the agonists, yet they well classified the inactive molecules as non-agonists. The best BI16707-A models showed better recall, but at the expense of more false positives. For these two reference trajectories, models ignoring hydrophobic interactions missed even more agonists, likely due to the absence of any hydrogen bond to TM5 (Ser203 and Ser207) as a consequence of replacing the aromatic hydrogen bond donor of the canonical ADRB2 agonists with a halogen atom. However, these polar models successfully classified the inactive molecules as non-agonists. The BI167107-I-based best models show the best statistics on the agonist/inactive dataset, although they performed worse than on the agonist/antagonist set. Since the BI167107-I training set structures do not include the hydrogen bond between the reference agonist and ADRB2 Ser207, we can assume that the classifier is less impacted by the unusual binding mode of the test agonists which likely involves a halogen-protein or fluorine-protein interaction. Taken together, the results suggest that polar interactions patterns allow the classification of non-agonists, yet that the interaction with serine residues in TM5, which constitute a key pharmacophore, must be correctly appreciated by the model to identify agonists.

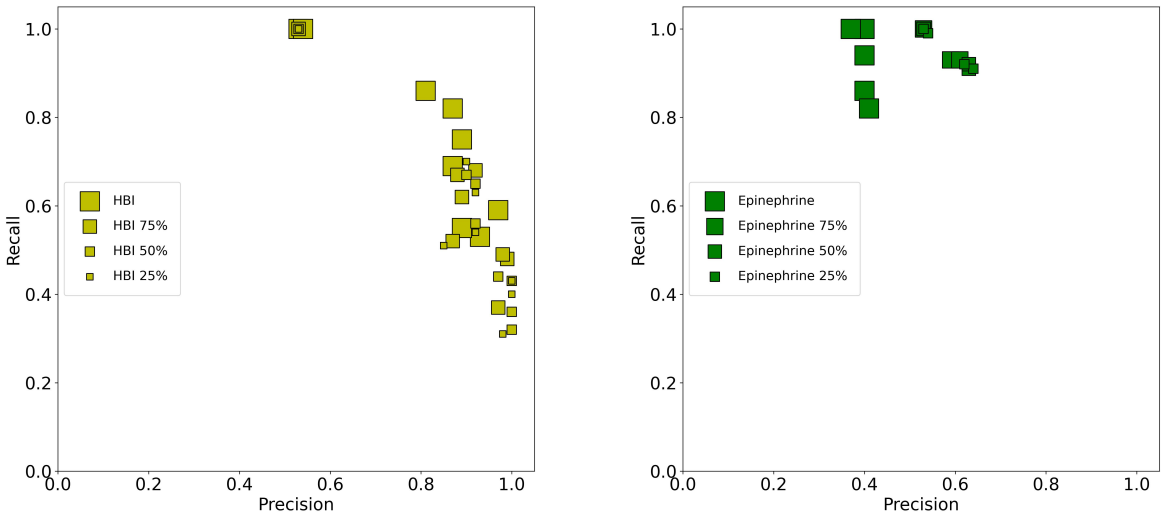
Table 3.1.3: Performance of the four best models obtained for each reference trajectory on the active/inactive dataset from docking data obtained using the representative structures

Reference trajectory	Model	Average			Ensemble		
		Precision	Recall	Size	Precision	Recall	Size
Epinephrine	Polar–MAD	0.40	1.00	25/27	0.37	1.00	27/27
	Hyd–QMS2	0.40	0.94	24/27	0.38	1.00	26/27
	Polar–QMS2	0.40	1.00	25/27	0.37	1.00	27/27
	Hyd–MAD	0.40	1.00	25/27			
HBI	Newhyd–QMS2	1.00	0.23	2/27	1.00	0.80	8/27
	Newhyd–MAD	1.00	0.23	2/27	1.00	0.70	7/27
	Hyd–NN	0.45	1.00	22/27	0.40	1.00	25/27
	Newhyd–NN	0.40	1.00	25/27	0.38	1.00	26/27
BI167107-A	Hyd–NN	0.56	0.72	13/27	0.56	1.00	18/27
	Newhyd–NN	0.44	1.00	23/27	0.42	1.00	24/27
	Newhyd–MAD	1.00	0.08	1/27	1.00	0.30	3/27
	Newhyd–QMS2	1.00	0.05	1/27	1.00	0.20	2/27
BI167107-I	Polar–NN	0.80	0.60	8/27	0.75	0.60	8/27
	Newhyd–NN	0.58	0.70	12/27	0.62	0.80	13/27
	Hyd–NN	0.58	0.70	12/27	0.62	0.80	13/27
	Newhyd–MAD	0.79	0.55	7/27	0.78	0.70	9/27

Size refers to the fraction of molecules classified as agonists. Average refers to the mean of the values obtained for the representative structures considered individually. Ensemble refers to the value obtained by merging data from all representative structures

As for the agonist/antagonist dataset the performance on the ensemble docking results was also considered. An important effect was observed for hydrophobic NN models, which under these conditions tend to select all molecules. By contrast, significant increase in recall with no loss in precision was observed for the best models based on HBI and BI167170-A. The better sampling of ligand bound conformations generated more poses containing hydrophobic

Figure 3.1.5: Average performance of the OCSVM models obtained for HBI, and epinephrine reference trajectories of variable length on the agonist/antagonist dataset from docking using the representative structures. Symbols size is proportional to the simulation length, ranging from 125ns (25%) to 500ns (100%).



contacts similar to the reference.

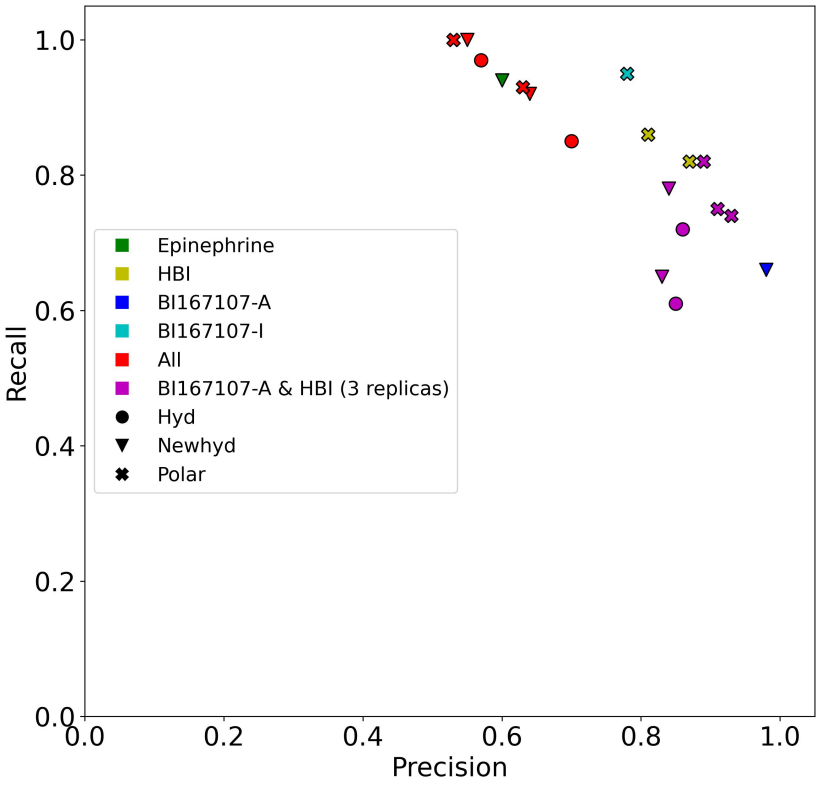
Importance of Sampling in Characterizing the Binding Mode

Introduction As for all machine learning models the characteristics of the training dataset reflect on the quality of the predictions. We questioned two aspects of binding mode sampling: the length of the MD simulation and the relevance of combining information from multiple references. Their effect on the classifiers' performances on the antagonist/agonist dataset is discussed below.

Molecular dynamics simulation length We repeated the OCSVM training using only the first 25%, 50%, and 75% of the frames of the MD simulations of epinephrine-ADRB2 and HBI-ABDR2 trajectories. These two trajectories were generated using the same experimental conditions. Both are formed by three different replicas, each one formed by 2500 frames, for a total of 7500 frames. Therefore, since they are the longest simulations and the only ones formed by more than a single replica, they were used to study the effect of the number of replicas and simulation length on the classifiers' performances.

The truncated simulation data was obtained by merging the selected frames in the three replicates. Applying the newly trained models on the agonist/antagonist dataset revealed that the number of selected molecules decreases as the MD simulation length is shortened, which is reflected by an increase in precision and a decrease in recall (Figure 3.1.4 and SI Table 7-10). The effect is much more marked if HBI is used as a reference, thus confirming that the epinephrine-ADRB2 trajectory does not highlight the interaction patterns that are crucial for the activation of the receptor.

Figure 3.1.6: Average performance on the agonist/antagonist dataset of OCSVM models obtained from docking using the representative structures from all merged reference trajectories or HBI and BI167107-A. For the sake of comparison are also shown the best models obtained for epinephrine, HBI , BI167107-A, and BI167107-I, as given in Table 3.1.2.



Combination of multiple references Although the proposed workflow was developed to use a single agonist as reference, the effect of merging multiple reference trajectories was investigated. We tested two combinations of datasets (Figure 3.1.5 and SI Table 11-12): All trajectories, BI167107-A and HBI. The models based on all the reference trajectories are characterized by a high recall and a relatively low precision. These poor performances are comparable to those obtained with the least relevant individual reference trajectory (that of epinephrine), suggesting that the determined agonist definition is too broad to be useful. The inclusion of BI167107-A data alongside the HBI trajectory increased the recall of the models including hydrophobic contacts. The overall effect of merging the two trajectories was however limited considering the performances of all models. This result further suggests that agonist-ADRB2 interactions in HBI and BI167107-A trajectories define a consistent binding mode.

Comparison with known methods

The comparison between the binding modes of a reference ligand with the docking poses has already shown to be an effective method to improve the performances of virtual screening. To evaluate the performances of the proposed model with the state of the art we performed a rescoring of the docking poses of the agonist/antagonist and agonist/inactive datasets using IFP similarity, GRIM, and structure-based 3D pharmacophores. Here, given the docking poses generated using an ADRB2 structure extracted from a MD trajectory, the agonist binding mode observed in the selected frame was used as a reference for the three methods to rescore the corresponding set of docking poses.

GRIM and OCSVM models are based on the same description of interactions. GRIM scores docking poses using an empirical function evaluating the maximum common subgraph between the reference IG and the docking pose's IG. For GRIM four different score thresholds have been proposed separating similar from dissimilar binding modes: 0.59, 0.65[36], 0.70, 1.00 [47]. None of the four thresholds yielded satisfying results overall, with only a small increase in precision over selecting all ligands as agonist (eg average precision around 0.60 at threshold 0.65 for HBI and BI167107), or an extremely low recall (eg average recall of 0.10 at threshold 1.00 for HBI and BI167107-A) on both the dataset containing agonists and antagonists (Table 3.1.4 and SI Table 13-16), and the dataset containing agonists and true inactive molecules (SI Tables 17-20). The difficulty in identifying a suitable GRIM score threshold to distinguish agonists is valid regardless of the set of docking poses considered (average results using a single representative structure, results using all representative structures, and results using the crystallographic structure).

Table 3.1.4: GRIM results on the agonists-antagonist dataset from docking data obtained using the representative structures or the crystallographic structure

Ligand	Threshold	Average		Crystallographic		Ensemble	
		Precision	Recall	Precision	Recall	Precision	Recall
Epinephrine	0.59	0.55	0.97	0.53	1.00	0.53	1.00
	0.65	0.70	0.62	0.63	1.00	0.56	1.00
	0.70	0.84	0.30	0.86	1.00	0.63	0.89
	1.00		0.00		0.00		0.00
HBI	0.59	0.53	1.00	0.53	1.00	0.53	1.00
	0.65	0.60	0.99	0.59	1.00	0.53	1.00
	0.70	0.67	0.98	0.68	1.00	0.53	1.00
	1.00	1.00	0.10	–	0.00	1.00	0.37
BI167107 active	0.59	0.53	1.00	0.53	1.00	0.53	1.00
	0.65	0.58	1.00	0.59	1.00	0.53	1.00
	0.70	0.64	0.99	0.68	1.00	0.58	1.00
	1.00	1.00	0.10	1.00	0.47	1.00	0.26
BI167107 inactive	0.59	0.53	1.00			0.53	1.00
	0.65	0.61	0.94			0.58	1.00
	0.70	0.69	0.84			0.70	1.00
	1.00	1.00	0.02			1.00	0.05

The standard definition of hydrophobic contacts is applied. Thresholds discarding all docking poses are indicated as –

Table 3.1.5: Optimal average scoring threshold for GRIM, IFP similarity, and 3D pharmacophore based on the agonist–antagonist dataset

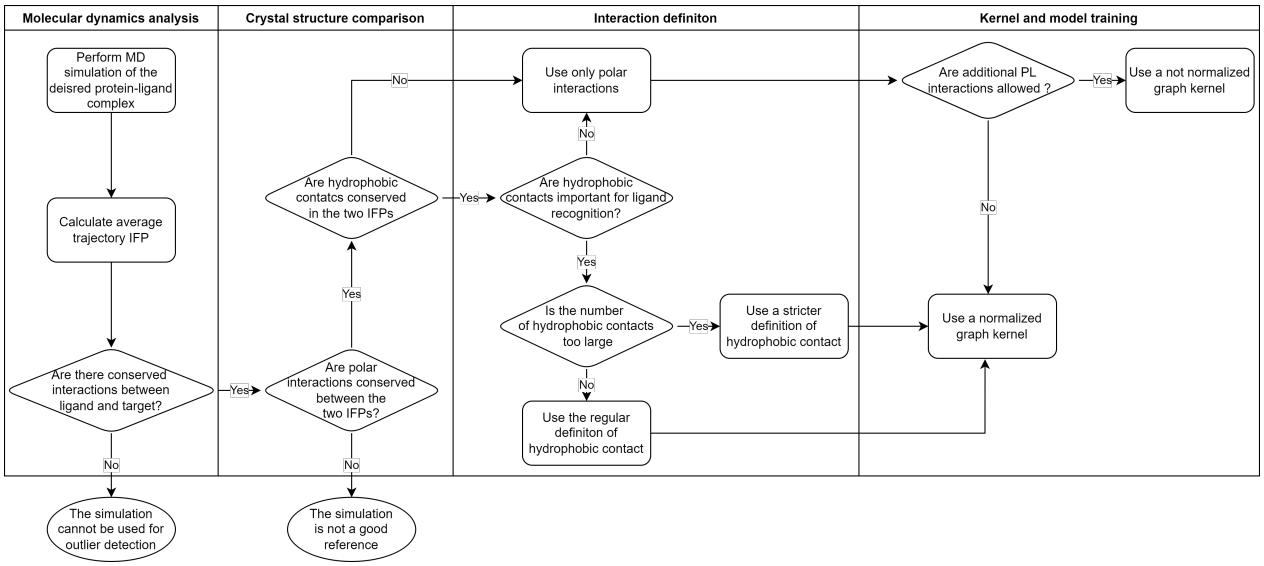
Ligand	Model	Threshold	Precision	Recall
Epinephrine	GRIM	0.63	0.66	0.93
	GRIM–Newhyd	0.63	0.66	0.90
	IFP	0.38	0.64	0.93
	IFP–Polar	0.48	0.65	0.94
	FitValue	0.59	0.61	0.92
	Pharmtype	1.43	0.60	0.92
HBI	GRIM	0.76	0.82	0.91
	GRIM–Newhyd	0.75	0.81	0.91
	IFP	0.58	0.76	0.89
	IFP–Polar	0.57	0.71	0.92
	FitValue	2.00	0.87	0.80
	Pharmtype	4.00	0.79	0.80
BI167107 active	GRIM	0.82	0.91	0.92
	GRIM–Newhyd	0.77	0.89	0.92
	IFP	0.60	0.79	0.86
	IFP–Polar	0.58	0.80	0.89
	FitValue	2.16	0.93	0.84
	Pharmtype	4.00	0.76	0.88
BI167107 inactive	GRIM	0.73	0.75	0.95
	GRIM–Newhyd	0.71	0.74	0.97
	IFP	0.64	0.73	0.88
	IFP–Polar	0.67	0.76	0.84
	FitValue	1.52	0.74	0.93
	Pharmtype	4.50	0.78	0.78

The different methods, GRIM, IFP, and pharmacophore search, were tested to determine the existence of a ligand independent threshold for classification. The optimal scoring value separating agonists and antagonists was determined for each ligand (Table 3.1.5 and SI Table 21), as the threshold maximizing the F1 measure for the agonist class. All methods can be used to bias the screening towards agonists over antagonists, but no method specific score threshold can be identified from the results. The overall performances depend on the reference agonist rather than the method used for rescoring, but even for the same reference agonist it is difficult to determine a specific threshold since each receptor structure used for docking was characterized by its own optimal threshold. In summary, the three methods well performed in prioritizing agonists, however it is not possible to clearly define the threshold separating agonists from antagonists a priori.

3.1.3 Conclusions

OCSVM models were developed from graphs encoding intermolecular interactions extracted from MD simulations of three reference complexes, between ADBR2 and epinephrine, HBI, and BI167107. Models based on HBI performed well in identifying agonists of the

Figure 3.1.7: Flowchart for model development



antagonist/agonist test set (best model average precision: 0.87), and filtering out inactive molecules of the agonist/inactive test set (best model average precision: 1.00). Failure in identifying agonists from the agonist/inactive test set was related to the substitution of hydrogen bond donor groups with chlorine and fluorine atoms. Similar results were obtained for BI167107-based models if they were trained from the portion of the MD trajectory that describes the active state of ADRB2. Performance deteriorated for models trained with the portion of the MD trajectory that describes the inactive state of ADRB2. Epinephrine-based models displayed low precision, tending to predict as agonists all molecules from both datasets. This poor performance was explained by an inappropriate sampling of the agonist binding mode, with epinephrine showing variable interactions patterns during simulation, probably including unbinding initiation. The importance of agonist binding mode sampling was also revealed by the effect of the MD simulation length on the models' performances. The OCSVM models were compared to established approaches for the post-processing of virtual screening results that search for the reference binding mode in docking poses: matching of interactions graphs using GRIM, comparing interaction fingerprints using IFP, and searching for 3D pharmacophores. All methods prioritized agonist ligands over inactive molecules and antagonists, yet only OCSVM models have a built-in threshold separating the two classes. With the three other methods ranking molecules in the dataset, a test set is required to properly define a scoring threshold, thus restraining their application to already well-studied GPCRs.

The proposed method can be applied to any GPCR provided that a three-dimensional structure is available between this GPCR and an agonist, and the corresponding MD simulation correctly samples the binding mode. Basic guidelines on how to build an appropriate model are given on Figure 3.1.7 and can be applied to other targets.

3.1.4 Materials and Methods

Molecular dynamics simulations

The MD trajectories of ADRB2 bound to epinephrine and HBI were downloaded from GPCRmd [32]. The ADRB2-Epinephrine (ID 117) simulation is based on the PDBID:4LDO [34] crystallographic structure. The protein is embedded in a lipid bilayer formed by POPC, the system is solvated by an aqueous solution of Na+ (158 mM) and Cl- (184 mM).

The ADRB2-HBI (ID 115) simulation is based on the PDBID:4LDL [34] crystallographic structure. The protein is embedded in a lipid bilayer formed by POPC, the system is solvated by an aqueous solution of Na+ (156 mM) and Cl- (182 mM).

Both simulations were performed using ACEMD with GPUGRID, the used forcefield is CHARMM36m. Three replicas of 0.5 μs were performed for each system with a timestep of 4.0 fs and a gap between the saved frames of 0.2 ns. For each system a dataset of 7500 protein-ligand structures is available.

The ADRB2-BI167107 simulation was performed by Dror et al.[33] using as starting structure the PDBID:3P0G [7] crystallographic structure. The protein is embedded in a lipid bilayer formed by POPC; the system is solvated by an aqueous solution of NaCl (0.15 M). The simulation was performed on Anton, the used forcefield is CHARMM27.

The entire trajectory (trajectory number 11 in the original publication) represents 10 μs of simulation, with frames saved every 180 ps. The part of the trajectory representing the active state receptor bound to the ligand corresponds to the first 350 ns of simulation. The part of trajectory used to represent the inactive state receptor begins after 3.60 μs of simulation and ends at 4.95 μs. The active state receptor dataset corresponds to 1944 frames, while the inactive state receptor dataset corresponds to 7500 frames.

Crystallographic structure preparation

The three reference crystallographic structures (PDBID: 3P0G, 4LDO, 4LDL) were downloaded from the PDB online repository. The structures were protonated and converted to mol2 files using MOE 2020. Protonation state and tautomer state of the residues were selected to match the one observed in the corresponding MD simulations.

Binding mode analysis

The agonist-ADRB2 binding modes depicted on the top of Figure 1 are a representation of the interactions detected using MOE 2020.

The binding of each ligand during the simulation, as shown on Figure 1, was analysed by computing IFP using IChem. The average IFP over each trajectory was calculated. Interactions observed in at least 10% of the frames of a single simulation, or observed in the crystallographic structure, are shown in the IFP graphical representation.

Interaction detection

Protein-ligand interactions were computed using IChem v5.2.9 [36]. IChem differentiates interactions according to the following seven types: Hydrogen bonds with the protein acting as acceptor, hydrogen bond with the protein acting as donor, ionic interactions, separated depending on the charge of the protein and the ligand, aromatic stacking, grouping together edge-to-face and face-to-face interactions, hydrophobic contacts, and metal chelation. An

interaction is detected between two atoms depending on their types and on topological constrains. Hydrogen bonds and aromatic stacking have both a distance and an angle constrain, while the remaining interaction depend only on the distance [36]. IChem presents two definitions of hydrophobic contacts, Hyd and Newhyd. The Hyd definition considers hydrophobic contacts between all atoms with the correct atom types, while the Newhyd definition considers only atoms with the correct atom types and in a hydrophobic environment.

IChem requires as input mol2 files with precise characteristics. The MD trajectory was converted in a series of pairs of mol2 files containing the protein and the ligand, removing lipids, water molecules, and ions. Such conversion was performed using a Python script based on pytraj [48, 49] for mol2 file generation. The generated files were corrected by a Python script to be compatible with IChem: CHARMM atom types were converted to SYBYL atom types, and the bond order between atoms was determined based on the atom types. Interaction detection was performed using the default definition of hydrophobic contacts and a stricter definition (-Newhyd). The definition of aromatic interactions was changed increasing the threshold distance to 5.0 A ($-D_Ar5.0$).

Model building

Graph kernel Graph similarity was evaluated using the shortest path kernel implemented in GraKel[50]. The shortest path kernel generates from the input graph the corresponding shortest path graph. The shortest path graph is a fully connected graph, with edges labelled according to the length of the shortest path between two nodes in the original graph. The shortest path graph is encoded as a fingerprint. Each element of the fingerprint corresponds to a pair of node labels and the label of the edge connecting them. The fingerprint contains the frequency of each path in the graph. There is no predefined fingerprint to describe the graphs, the fingerprint is learned from the training dataset to contain the observed paths. The fingerprint is temporarily updated during kernel application on new graphs to include previously unseen paths. The update in the fingerprint is relevant only for normalized similarity score calculations (Fig. 3.1.8).

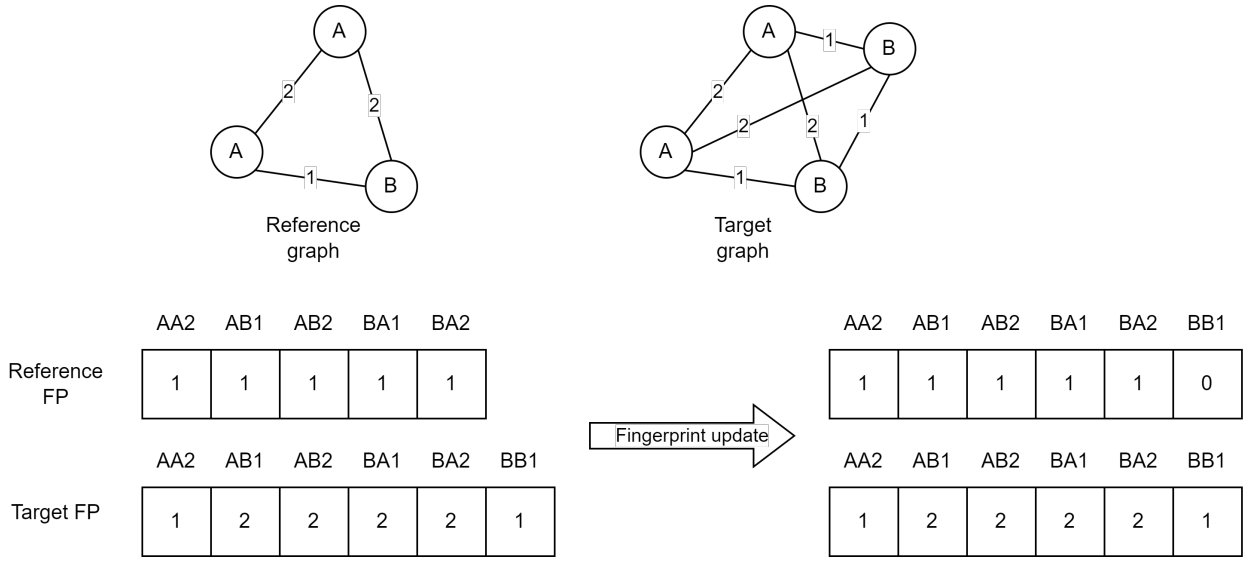
Model training The Scikit-learn[51] implementation of the OCSVM algorithm was used for the classifier. The model was trained using two different heuristics: QMS2 and a MAD based method. For both methods the average KNN similarity was calculated, with K corresponding to 3% of the number of instances in the dataset.

For QMS2 the KNN-similarity values were sorted in ascending order. A function was interpolated linking the rank of each point to its similarity score. Outliers are defined as points before the first major knee. For such points an increase in rank is associated with a rapid increase of KNN similarity. The knee was determined using the kneedle algorithm[52], selecting the first concave point. Kneedle sensibility parameter was manually selected to identify the true global knee of the function. Points below the selected knee were discarded and $\nu = 0.01$.

For training based on MAD, the median and MAD values were calculated using SciPy[53]. The fraction of outliers in the training set was determined as the fraction of points with a distance from the median greater than 3 times MAD.

A Python script, available at `https://github.com/LIT-CCM-lab/OCSVM-ADRB2`, was developed for the training of the kernel and the determination of OCSVM parameters using

Figure 3.1.8: Fingerprint generation and update for shortest path kernel similarity calculation



both methods.

Evaluation metrics The performances of the presented classification methods are evaluated according to their precision, recall, and F1 measure[39, 55]. Precision and recall are defined as:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

With TP (true positive) indicating agonist ligands correctly classified as agonist, FP (false positives) indicating antagonist or inactive molecules classified as agonist, and FN indicating agonist molecules not classified as agonist.

The classifier was trained to identify instances from a single class, the F1 score calculated on the positive class (agonist) was selected as a metric to evaluate the overall performance of the model. The F1 measure is defined as:

$$F1 = 2\frac{Precision * Recall}{Precision + Recall}$$

Selection of representative conformations

Binding site definition Two definitions of the receptor binding site have been used in this work, a system specific definition and a shared definition. The system specific binding site is obtained by selecting all non-hydrogen atoms at less than 4.5 Å from the non-hydrogen atoms of the ligand for at least 10% of the trajectory frames. The shared definition includes all residues containing at least one atom present in one of the four different system specific binding site definitions.

Trajectory clustering Similarity between receptor conformations was evaluated using the RMSD of interacting atoms.

$$RMSD = \sqrt{\frac{\sum_i^N (X_i - Y_i)^2}{N}}$$

With N indicating the selected interacting atoms, X_i and Y_i the coordinates of each atom i in two compared structures. The pairwise RMSD between structures in the same trajectory was computed, on atoms in the system specific binding site definition, using pytraj. The pairwise distances were represented by a histogram to determine the presence of a clusterable data structure[54].

Only the ADRB2-HBI trajectory presents multiple clearly distinct conformation of the binding site based on the RMSD. Agglomerative hierarchical clustering with average linkage was performed on the HBI-ADRB2 dataset using the scikit-learn implementation of the algorithm. The optimal number of clusters was evaluated considering different geometrical criteria, as already proposed by De Paris et al. [55]: Silhouette score, Dunn index, and Davies-Bouldin index. The optimal number of clusters was estimated at two, with a major cluster (6881 elements) and a minor cluster (619 elements).

The remaining trajectories and the major cluster of the HBI-ADRB2 complex were clustered using the K-medoids algorithm [56], implemented in scikit-learn-extra. The optimal number of clusters was estimated according to geometrical criteria: Silhouette score, Dunn index, and Davies-Bouldin index. From HBI-main-cluster 9 clusters were obtained, 8 for epinephrine, 2 for BI167107-I, and 4 for BI167107-A. For each cluster the medoid, the member of the cluster with the lowest distance from all other members of the same cluster, was selected as the representative structures.

Comparison to the crystallographic structure The RMSD between the starting crystal structure and the trajectory was evaluated using the shared definition of binding site. The RMSD was calculated using pytraj. The shared definition of binding site was used to have comparable values between the trajectories.

Pairwise comparison of the trajectories The pairwise RMSD of the representative structures and of the trajectories was calculated using the common definition of the binding site. For the comparison between the trajectories one every five saved frames was kept for analysis.

Agonist-antagonist validation set preparation

Initial ligand selection The initial set of molecules was selected from the ADRB2 (CHEMBL210) target page on ChEMBL[57], in the section "Drugs And Clinical Candidates", 33 agonists, 2 partial agonists and 13 antagonists were retrieved. An additional literature search was performed to increase the number of antagonists in the dataset, with the inclusion of 7 new molecules, all explicitly described as ADRB2 antagonist. Bupranolol, tertatolol, IPS339, spirendolol, ICI118551 were selected from an article comparing the electronic structure of β-receptor agonist and antagonist[58]. Compound CHEMBL355038 (15a) [59] and CHEMBL3228930 (4)[60] were retrieved from articles describing their synthesis. Ligands with "Mechanism of action" antagonist from GPCRdb [61] "Drugs target and Indications"

targeting ADRB2 were selected. Only new entries with literature reference were considered. Three ligands (Alprenolol, Carteolol, Carvedilol) were added to the dataset. From the IUPHAR/BPS Guide to Pharmacology page on ADRB2 [62] four new antagonist ligands were added to the dataset (Butoxamine, CGP-12177, SR59230A, NIP). The initial dataset contains 29 agonists, 27 antagonists, and 2 partial agonists.

Dataset cleaning Duplicates corresponding to specific enantiomers of already present racemic ligands were discarded.

Three ligands with unknown structure were removed (LAS190792, AZD3199, GSK159802). The two ligands described as partial agonist (pindolol, celiprolol) were also discarded from the dataset. The number of heavy atoms of each ligand was calculated using RDKit[63]. Molecules with a significantly larger number of heavy atoms (Z-score > 1.5) were discarded to have a dataset of molecules with homogeneous size.

Morgan's structural fingerprints were calculated for the remaining molecules with a radius of 2 and a fingerprint length of 2048 bits using RDKit. The pairwise Tanimoto distance (1 – Tanimoto similarity) was computed and used to perform agglomerative hierarchical clustering. A distance threshold of 0.5 was set for the clustering. From each cluster the molecule with the highest affinity for the target was selected. The highest affinity was determined by comparing results from assays performed in similar conditions. In case no comparable assay results were found, the best characterized molecule was selected.

The scaffold distribution between agonist and antagonists was evaluated by retrieving the Bemis-Murcko scaffolds with RDKit.

Ligand preparation

The ligands, represented as SMILES, were ionized using the FILTER program from OpenEye (Filter 2.5.1.4 OpenEye Scientific Software, Santa Fe, NM, USA). The ligands' 3D structures were generated using Corina 3.40 (Molecular Networks GmbH, Nürnberg, Germany)[64]. The rc option was used to generate multiple ring conformation. For ligands presenting stereocenters all stereoisomers were generated.

Docking

Rigid protein ligand docking was performed using PLANTS v1.2, using the CHEMPLP scoring function and search speed 1 (highest accuracy) [65]. The cavity centre of each protein was identified as the centre of mass of the interacting atoms in the system specific definition of the binding site. The cavity radius was set to 12 Å for all structures. For each ligand stereoisomer 10 poses were saved.

Docking post-processing

IFP Interaction fingerprints were calculated using IChem. Comparison between fingerprints was performed using a Python script. Interaction similarity was defined as the Tanimoto similarity between the two binary IFPs.

GRIM Protein-ligand interactions were detected using IChem, with both available definitions of hydrophobic contacts. For each docking pose the generated IPAs were saved

in mol2 files. The definition of aromatic interactions was changed increasing the threshold distance to 5.0 A ($-D_Ar5.0$). The GRIM score between the docking IPAs and the reference IPAs was computed using IChem.

3D pharmacophore Interaction 3D pharmacophores were generated automatically using DiscoveryStudio (Dassault Systemes BIOVIA), the settings for generation were changed to include all the detected interactions between ligand and protein as pharmacophoric features.

The software citest was used for evaluation of the docking pose by comparing the alignment to the pharmacophoric query without performing fitting. Two different scores were used to evaluate the poses: FitScore, which is a sum of the fitting quality to each feature, and Pharmtype, which measures the number of features of the query matched by the pose.

OCSVM evaluation The IPA files generated from the MD trajectory were used to train the shortest path graph kernel and the OCSVM classification model. The docking poses IPAs were converted to IGs. The node labels were extracted from the residue name of the IPAs. The adjacency matrix was calculated as the pairwise Euclidean distances between the IPAs, approximated to the nearest A. A distance of 0.1 was set for overlapping IPAs. Polar IGs were generated from regular interaction files by removing hydrophobic IPAs. The similarity to the graphs in the training set was calculated using the trained shortest path graph kernel. The obtained Gram matrix is used by the OCSVM model to score the IGs. Negative scores indicate an outlier, while positive scores indicate an inlier. Inliers are considered as active molecules with the same pharmacological effect as the reference.

Abbreviations

Abbreviations ADRB2: β2 Adrenergic receptor; COPD: Chronic obstructive pulmonary disease; MD: Molecular dynamics; OCSVM: One class support vector machine; KNN: K-nearest neighbors; HBI: Hydroxybenzyl isoproterenol; BI167107-A: BI167107 bound to the active state of ADRB2; BI167107-I: BI167107 bound to the inactive state of ADRB2; Hyd: Regular definition of hydrophobic contacts in IChem; Newhyd: Stricter definition of hydrophobic contacts in IChem; IG: Interaction graph; NN: Not normalized; SV: Support vector; QMS2: Quick model selection 2; MAD: Median absolute deviation; IFP: Interaction fingerprints; GRIM: Graph matching of interaction patterns; RMSD: Root mean square deviation.

Supplementary information

The online version contains supplementary material available at `https://doi.org/10.1186/s13321-022-00654-z`.

Additional file 1: **Figure S1**. Average and crystallographic interaction fingerprints for the three reference agonists. **Figure S2**. RMSD of the binding site residues during the simulation of the BI167107-ADRB2 complex. **Figure S3**. Pairwise RMSD of the four reference trajectories. **Figure S4**. Pairwise RMSD of the representative structures used for docking. **Figure S5**. RMSD of the binding site residues during the simulation of the epinephrine-ADRB2 complex. **Figure S6**. RMSD of the binding site residues during the simulation of the HBI-ADRB2 complex. **Table S1**. Annotated

list of molecules from the agonist/antagonist dataset. **Table S2**. Annotated list of molecules from the agonist/inactive dataset. **Table S3**. Average performances of the OCSVM models on the agonist/antagonist dataset, from docking data obtained using the representative structures. **Table S4**. Performances of the OCSVM models on the agonist/antagonist dataset, from docking data obtained using crystallographic structures. **Table S5**. Average performances of the OCSVM models on the agonist/inactive dataset, from docking data obtained using the representative structures. **Table S6**. Performances of the OCSVM models on the agonist/inactive dataset, from docking data obtained using crystallographic structures. **Table S7**. Average performances of shorter trajectories of epinephrine-based OCSVM models on the agonist-antagonist dataset, from docking data obtained using the representative structures. **Table S8**. Performances of shorter trajectories of epinephrine-based OCSVM models on the agonist-antagonist dataset, from docking data obtained using crystallographic structures. **Table S9**. Average performances of shorter trajectories of HBI-based OCSVM models on the agonist-antagonist dataset, from docking data obtained using the representative structures. **Table S10**. Performances of shorter trajectories of HBI-based OCSVM models on the agonist-antagonist dataset, from docking data obtained using crystallographic structures. **Table S11**. Average performances of the combined dataset OCSVM models on the agonist-antagonist dataset, from docking data obtained using the representative structures. **Table S12**. Performances of the combined dataset OCSVM models on the agonist-antagonist dataset, from docking data obtained using crystallographic structures. **Table S13**. Average performance of GRIM at different thresholds on the agonist-antagonist dataset, from docking data obtained using the representative structures. Default definition of hydrophobic contacts. **Table S14**. Performance of GRIM at different thresholds on the agonist-antagonist dataset, from docking data obtained using crystallographic structures. Default definition of hydrophobic contacts. **Table S15**. Average performance of GRIM at different thresholds on the agonist-antagonist dataset, from docking data obtained using the representative structures. Newhyd definition of hydrophobic contacts. **Table S16**. Performance of GRIM at different thresholds on the agonist-antagonist dataset, from docking data obtained using crystallographic structures. Newhyd definition of hydrophobic contacts. **Table S17**. Average performance of GRIM at different thresholds on the agonist-inactive dataset, from docking data obtained using the representative structures. Default definition of hydrophobic contacts. **Table S18**. Performance of GRIM at different thresholds on the agonist-inactive dataset, from docking data obtained using crystallographic structures. Default definition of hydrophobic contacts. **Table S19**. Average performance of GRIM at different thresholds on the agonist-inactive dataset, from docking data obtained using the representative structures. Newhyd definition of hydrophobic contacts. **Table S20**. Performance of GRIM at different thresholds on the agonist-inactive dataset, from docking data obtained using crystallographic structures. Newhyd definition of hydrophobic contacts. **Table S21**. Average optimal threshold for agonist/antagonist classification for GRIM, IFP, and 3D pharmacophore. **Table S22**. Interaction graphs statistics from the four reference trajectories. **Table S23**. Interaction graphs statistics form the docking poses of the agonist-antagonist dataset and agonist-inactive datasets.

Acknowledgements

The authors acknowledge the High Performance Computing Center of the French National Institute for Nuclear and Particle Physics (IN2P3), E Sick for critical comments on ADRB2 function, C Jacquemard for the technical support, and RO Dror for sharing data.

Authors' contributions

Project coordination: EK; design of study: LC and EK; data preparation: LC ; data analysis: LC and EK; preparation of manuscript: LC and EK. All authors read and approved the final manuscript.

Funding

This work was supported by the Agence Nationale de Recherches (ANR) (grant number 2019 CE14 OCHRE to E.K.), Centre National de la Recherche Scientifique (CNRS), Institut du Médicament de Strasbourg (IMS), and Université de Strasbourg.

Availability of data and materials

The training data (IPA, IG) and the test datasets (annotated structures and docking poses) are accessible via an online repository (https://seafile.unistra.fr/d/548138136ff14e3889c1). The code and protocol required to develop the model starting from the MD trajectory is available at: https://github.com/LIT-CCM-lab/OCSVM-ADRB2.

References

[1] Brian K. Kobilka. G protein coupled receptor structure and activation. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1768(4):794–807, April 2007.

[2] Krishna Sriram and Paul A. Insel. G Protein-Coupled Receptors as Targets for Approved Drugs: How Many Targets and How Many Drugs? *Molecular Pharmacology*, 93(4):251–258, April 2018.

[3] Daniel Hilger, Matthieu Masureel, and Brian K. Kobilka. Structure and dynamics of GPCR signaling complexes. *Nature Structural & Molecular Biology*, 25(1):4–12, January 2018.

[4] José Jiménez-Luna, Francesca Grisoni, Nils Weskamp, and Gisbert Schneider. Artificial intelligence in drug discovery: Recent advances and future perspectives. *Expert Opinion on Drug Discovery*, 16(9):949–959, September 2021.

[5] Amara Jabeen and Shoba Ranganathan. Applications of machine learning in GPCR bioactive ligand discovery. *Current Opinion in Structural Biology*, 55:66–76, April 2019.

[6] Vadim Cherezov, Daniel M. Rosenbaum, Michael A. Hanson, Søren G. F. Rasmussen, Foon Sun Thian, Tong Sun Kobilka, Hee-Jung Choi, Peter Kuhn, William I. Weis, Brian K. Kobilka, and Raymond C. Stevens. High-Resolution Crystal Structure of an Engineered Human β_2 -Adrenergic G Protein–Coupled Receptor. *Science*, 318(5854):1258–1265, November 2007.

[7] Søren G. F. Rasmussen, Hee-Jung Choi, Juan Jose Fung, Els Pardon, Paola Casarosa, Pil Seok Chae, Brian T. DeVree, Daniel M. Rosenbaum, Foon Sun Thian, Tong Sun Kobilka, Andreas Schnapp, Ingo Konetzki, Roger K. Sunahara, Samuel H. Gellman, Alexander Pautsch, Jan Steyaert, William I. Weis, and Brian K. Kobilka. Structure of a nanobody-stabilized active state of the $B2$ adrenoceptor. *Nature*, 469(7329):175–180, January 2011.

[8] Brian K. Shoichet and Brian K. Kobilka. Structure-based drug screening for G-protein-coupled receptors. *Trends in Pharmacological Sciences*, 33(5):268–272, May 2012.

[9] Albert J. Kooistra, Rob Leurs, Iwan J. P. de Esch, and Chris de Graaf. From Three-Dimensional GPCR Structure to Rational Ligand Discovery. In Marta Filizola, editor, *G Protein-Coupled Receptors - Modeling and Simulation*, volume 796, pages 129–157. Springer Netherlands, Dordrecht, 2014.

[10] Miles Congreve, Chris de Graaf, Nigel A. Swain, and Christopher G. Tate. Impact of GPCR Structures on Drug Discovery. *Cell*, 181(1):81–91, April 2020.

[11] Flavio Ballante, Albert J Kooistra, Stefanie Kampen, Chris de Graaf, and Jens Carlsson. Structure-Based Virtual Screening for Ligands of G Protein–Coupled Receptors: What Can Molecular Docking Do for You? *Pharmacological Reviews*, 73(4):1698–1736, October 2021.

[12] Albert J. Kooistra, Rob Leurs, Iwan J. P. de Esch, and Chris de Graaf. Structure-Based Prediction of G-Protein-Coupled Receptor Ligand Function: A β-Adrenoceptor Case Study. *Journal of Chemical Information and Modeling*, 55(5):1045–1061, May 2015.

[13] Andrew C. Kruse, Dahlia R. Weiss, Mario Rossi, Jianxin Hu, Kelly Hu, Katrin Eitel, Peter Gmeiner, Jürgen Wess, Brian K. Kobilka, and Brian K. Shoichet. Muscarinic Receptors as Model Targets and Antitargets for Structure-Based Ligand Discovery. *Molecular Pharmacology*, 84(4):528–540, October 2013.

[14] Aashish Manglik, Henry Lin, Dipendra K. Aryal, John D. McCorvy, Daniela Dengler, Gregory Corder, Anat Levit, Ralf C. Kling, Viachaslau Bernat, Harald Hübner, Xi-Ping Huang, Maria F. Sassano, Patrick M. Giguère, Stefan Löber, Da Duan, Grégory Scherrer, Brian K. Kobilka, Peter Gmeiner, Bryan L. Roth, and Brian K. Shoichet. Structure-based discovery of opioid analgesics with reduced side effects. *Nature*, 537(7619):185–190, September 2016.

[15] Sheng Wang, Daniel Wacker, Anat Levit, Tao Che, Robin M. Betz, John D. McCorvy, A. J. Venkatakrishnan, Xi-Ping Huang, Ron O. Dror, Brian K. Shoichet, and Bryan L. Roth. D $_4$ dopamine receptor high-resolution structures enable the discovery of selective agonists. *Science*, 358(6361):381–386, October 2017.

[16] Jiankun Lyu, Sheng Wang, Trent E. Balius, Isha Singh, Anat Levit, Yurii S. Moroz, Matthew J. O'Meara, Tao Che, Enkhjargal Algaa, Kateryna Tolmachova, Andrey A. Tolmachev, Brian K. Shoichet, Bryan L. Roth, and John J. Irwin. Ultra-large library docking for discovering new chemotypes. *Nature*, 566(7743):224–229, February 2019.

[17] Esther Kellenberger, Jean-Yves Springael, Marc Parmentier, Muriel Hachet-Haas, Jean-Luc Galzi, and Didier Rognan. Identification of Nonpeptide CCR5 Receptor Agonists by Structure-based Virtual Screening. *Journal of Medicinal Chemistry*, 50(6):1294–1303, March 2007.

[18] Reed M. Stein, Hye Jin Kang, John D. McCorvy, Grant C. Glatfelter, Anthony J. Jones, Tao Che, Samuel Slocum, Xi-Ping Huang, Olena Savych, Yurii S. Moroz, Benjamin Stauch, Linda C. Johansson, Vadim Cherezov, Terry Kenakin, John J. Irwin, Brian K. Shoichet, Bryan L. Roth, and Margarita L. Dubocovich. Virtual discovery of melatonin receptor ligands to modulate circadian rhythms. *Nature*, 579(7800):609–614, March 2020.

[19] Chris de Graaf and Didier Rognan. Selective Structure-Based Virtual Screening for Full and Partial Agonists of the $B2$ Adrenergic Receptor. *Journal of Medicinal Chemistry*, 51(16):4978–4985, August 2008.

[20] Antonella Ciancetta, Alberto Cuzzolin, and Stefano Moro. Alternative Quality Assessment Strategy to Compare Performances of GPCR-Ligand Docking Protocols: The Human Adenosine A $_{2A}$ Receptor as a Case Study. *Journal of Chemical Information and Modeling*, 54(8):2243–2254, August 2014.

[21] Thomas Coudrat, Arthur Christopoulos, Patrick Michael Sexton, and Denise Wootten. Structural features embedded in G protein-coupled receptor co-crystal structures are key to their success in virtual screening. *PLOS ONE*, 12(4):e0174719, April 2017.

[22] Albert J. Kooistra, Henry F. Vischer, Daniel McNaught-Flores, Rob Leurs, Iwan J. P. de Esch, and Chris de Graaf. Function-specific virtual screening for GPCR ligands using a combined scoring method. *Scientific Reports*, 6(1):28288, September 2016.

[23] Charlotte K. Billington, Raymond B. Penn, and Ian P. Hall. $B2$ Agonists. In Clive P. Page and Peter J. Barnes, editors, *Pharmacology and Therapeutics of Asthma and COPD*, volume 237, pages 23–40. Springer International Publishing, Cham, 2016.

[24] Nicola A Hanania, Burton F Dickey, and Richard A Bond. Clinical implications of the intrinsic efficacy of beta-adrenoceptor drugs in asthma: Full, partial and inverse agonism. *Current Opinion in Pulmonary Medicine*, 16(1):1–5, January 2010.

[25] Stan Ejiofor and Alice M. Turner. Pharmacotherapies for COPD. *Clinical Medicine Insights: Circulatory, Respiratory and Pulmonary Medicine*, 7:CCRPM.S7211, January 2013.

[26] Jillian G Baker. The selectivity of β -adrenoceptor antagonists at the human β 1, β 2 and β 3 adrenoceptors: β -Blockers and β -adrenoceptor selectivity. *British Journal of Pharmacology*, 144(3):317–322, February 2005.

[27] H. C. Stephen Chan, Slawomir Filipek, and Shuguang Yuan. The Principles of Ligand Specificity on beta-2-adrenergic receptor. *Scientific Reports*, 6(1):34736, December 2016.

[28] Mariona Torrens-Fontanals, Tomasz Maciej Stepniewski, David Aranda-García, Adrián Morales-Pastor, Brian Medel-Lacruz, and Jana Selent. How Do Molecular Dynamics

Data Complement Static Structural Data of GPCRs. *International Journal of Molecular Sciences*, 21(16):5933, August 2020.

[29] Magdalena M. Scharf, Moritz Bünemann, Jillian G. Baker, and Peter Kolb. Comparative Docking to Distinct G Protein–Coupled Receptor Conformations Exclusively Yields Ligands with Agonist Efficacy. *Molecular Pharmacology*, 96(6):851–861, December 2019.

[30] Bernhard Schölkopf, John C. Platt, John Shawe-Taylor, Alex J. Smola, and Robert C. Williamson. Estimating the Support of a High-Dimensional Distribution. *Neural Computation*, 13(7):1443–1471, July 2001.

[31] Yiran Wu, Liting Zeng, and Suwen Zhao. Ligands of Adrenergic Receptors: A Structural Point of View. *Biomolecules*, 11(7):936, June 2021.

[32] Ismael Rodríguez-Espigares, Mariona Torrens-Fontanals, Johanna K. S. Tiemann, David Aranda-García, Juan Manuel Ramírez-Anguita, Tomasz Maciej Stepniewski, Nathalie Worp, Alejandro Varela-Rial, Adrián Morales-Pastor, Brian Medel-Lacruz, Gáspár Pándy-Szekeres, Eduardo Mayol, Toni Giorgino, Jens Carlsson, Xavier Deupi, Slawomir Filipek, Marta Filizola, José Carlos Gómez-Tamayo, Angel Gonzalez, Hugo Gutiérrez-de-Terán, Mireia Jiménez-Rosés, Willem Jespers, Jon Kapla, George Khelashvili, Peter Kolb, Dorota Latek, Maria Marti-Solano, Pierre Matricon, Minos-Timotheos Matsoukas, Przemyslaw Miszta, Mireia Olivella, Laura Perez-Benito, Davide Provasi, Santiago Ríos, Iván R. Torrecillas, Jessica Sallander, Agnieszka Sztyler, Silvana Vasile, Harel Weinstein, Ulrich Zachariae, Peter W. Hildebrand, Gianni De Fabritiis, Ferran Sanz, David E. Gloriam, Arnau Cordomi, Ramon Guixà-González, and Jana Selent. GPCRmd uncovers the dynamics of the 3D-GPCRome. *Nature Methods*, 17(8):777–787, August 2020.

[33] Ron O. Dror, Daniel H. Arlow, Paul Maragakis, Thomas J. Mildorf, Albert C. Pan, Huafeng Xu, David W. Borhani, and David E. Shaw. Activation mechanism of the β_2-adrenergic receptor. *Proceedings of the National Academy of Sciences*, 108(46):18684–18689, November 2011.

[34] Aaron M. Ring, Aashish Manglik, Andrew C. Kruse, Michael D. Enos, William I. Weis, K. Christopher Garcia, and Brian K. Kobilka. Adrenaline-activated structure of $B2$-adrenoceptor stabilized by an engineered nanobody. *Nature*, 502(7472):575–579, October 2013.

[35] Rie Nygaard, Yaozhong Zou, Ron O. Dror, Thomas J. Mildorf, Daniel H. Arlow, Aashish Manglik, Albert C. Pan, Corey W. Liu, Juan José Fung, Michael P. Bokoch, Foon Sun Thian, Tong Sun Kobilka, David E. Shaw, Luciano Mueller, R. Scott Prosser, and Brian K. Kobilka. The Dynamic Process of $B2$-Adrenergic Receptor Activation. *Cell*, 152(3):532–542, January 2013.

[36] Jérémy Desaphy, Eric Raimbaud, Pierre Ducrot, and Didier Rognan. Encoding Protein–Ligand Interaction Patterns in Fingerprints and Graphs. *Journal of Chemical Information and Modeling*, 53(3):623–637, March 2013.

[37] Nils M. Kriege, Fredrik D. Johansson, and Christopher Morris. A survey on graph kernels. *Applied Network Science*, 5(1):6, December 2020.

[38] K.M. Borgwardt and H. Kriegel. Shortest-Path Kernels on Graphs. In *Fifth IEEE International Conference on Data Mining (ICDM'05)*, pages 74–81, Houston, TX, USA, 2005. IEEE.

[39] Shehroz S. Khan and Michael G. Madden. One-class classification: Taxonomy of study and review of techniques. *The Knowledge Engineering Review*, 29(3):345–374, June 2014.

[40] Naeem Seliya, Azadeh Abdollah Zadeh, and Taghi M. Khoshgoftaar. A literature review on one-class classification and its potential applications in big data. *Journal of Big Data*, 8(1):122, December 2021.

[41] Pavel V. Karpov, Dmitry I. Osolodkin, Igor I. Baskin, Vladimir A. Palyulin, and Nikolay S. Zefirov. One-class classification as a novel method of ligand-based virtual screening: The case of glycogen synthase kinase 3β inhibitors. *Bioorganic & Medicinal Chemistry Letters*, 21(22):6728–6731, November 2011.

[42] Riccardo Aguti, Erika Gardini, Martina Bertazzo, Sergio Decherchi, and Andrea Cavalli. Probabilistic Pocket Druggability Prediction via One-Class Learning. *Frontiers in Pharmacology*, 13:870479, June 2022.

[43] Aikaterini Vriza, Angelos B. Canaj, Rebecca Vismara, Laurence J. Kershaw Cook, Troy D. Manning, Michael W. Gaultois, Peter A. Wood, Vitaliy Kurlin, Neil Berry, Matthew S. Dyer, and Matthew J. Rosseinsky. One class classification as a practical approach for accelerating π–π co-crystal discovery. *Chemical Science*, 12(5):1702–1719, 2021.

[44] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation Forest. In *2008 Eighth IEEE International Conference on Data Mining*, pages 413–422, Pisa, Italy, December 2008. IEEE.

[45] Andrew F. Emmott, Shubhomoy Das, Thomas Dietterich, Alan Fern, and Weng-Keen Wong. Systematic construction of anomaly detection benchmarks from real data. In *Proceedings of the ACM SIGKDD Workshop on Outlier Detection and Description - ODD '13*, pages 16–21, Chicago, Illinois, 2013. ACM Press.

[46] Zahra Ghafoori, Sarah M. Erfani, Sutharshan Rajasegarar, James C. Bezdek, Shanika Karunasekera, and Christopher Leckie. Efficient Unsupervised Parameter Estimation for One-Class Support Vector Machines. *IEEE Transactions on Neural Networks and Learning Systems*, 29(10):5057–5070, October 2018.

[47] Inna Slynko, Franck Da Silva, Guillaume Bret, and Didier Rognan. Docking pose selection by interaction pattern graph similarity: Application to the D3R grand challenge 2015. *Journal of Computer-Aided Molecular Design*, 30(9):669–683, September 2016.

[48] Hai Nguyen, Jason Swails, Daniel R. Roe, and Cody. Amber-Md/Pytraj: V2.0.0. Zenodo, March 2017.

[49] Daniel R. Roe and Thomas E. Cheatham. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation*, 9(7):3084–3095, July 2013.

[50] Giannis Siglidis, Giannis Nikolentzos, Stratis Limnios, Christos Giatsidis, Konstantinos Skianis, and Michalis Vazirgiannis. GraKeL: A Graph Kernel Library in Python. *Journal of Machine Learning Research*, 21(54):1–5, 2020.

[51] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85):2825–2830, 2011.

[52] Ville Satopaa, Jeannie Albrecht, David Irwin, and Barath Raghavan. Finding a "Kneedle" in a Haystack: Detecting Knee Points in System Behavior. In *2011 31st International Conference on Distributed Computing Systems Workshops*, pages 166–171, June 2011.

[53] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, SciPy 1.0 Contributors, Aditya Vijaykumar, Alessandro Pietro Bardelli, Alex Rothberg, Andreas Hilboll, Andreas Kloeckner, Anthony Scopatz, Antony Lee, Ariel Rokem, C. Nathan Woods, Chad Fulton, Charles Masson, Christian Häggström, Clark Fitzgerald, David A. Nicholson, David R. Hagen, Dmitrii V. Pasechnik, Emanuele Olivetti, Eric Martin, Eric Wieser, Fabrice Silva, Felix Lenders, Florian Wilhelm, G. Young, Gavin A. Price, Gert-Ludwig Ingold, Gregory E. Allen, Gregory R. Lee, Hervé Audren, Irvin Probst, Jörg P. Dietrich, Jacob Silterra, James T Webber, Janko Slavič, Joel Nothman, Johannes Buchner, Johannes Kulick, Johannes L. Schönberger, José Vinícius de Miranda Cardoso, Joscha Reimer, Joseph Harrington, Juan Luis Cano Rodríguez, Juan Nunez-Iglesias, Justin Kuczynski, Kevin Tritz, Martin Thoma, Matthew Newville, Matthias Kümmerer, Maximilian Bolingbroke, Michael Tartre, Mikhail Pak, Nathaniel J. Smith, Nikolai Nowaczyk, Nikolay Shebanov, Oleksandr Pavlyk, Per A. Brodtkorb, Perry Lee, Robert T. McGibbon, Roman Feldbauer, Sam Lewis, Sam Tygier, Scott Sievert, Sebastiano Vigna, Stefan Peterson, Surhud More, Tadeusz Pudlik, Takuya Oshima, Thomas J. Pingel, Thomas P. Robitaille, Thomas Spura, Thouis R. Jones, Tim Cera, Tim Leslie, Tiziano Zito, Tom Krauss, Utkarsh Upadhyay, Yaroslav O. Halchenko, and Yoshiki Vázquez-Baeza. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, March 2020.

[54] Andreas Adolfsson, Margareta Ackerman, and Naomi C. Brownstein. To cluster, or not to cluster: An analysis of clusterability methods. *Pattern Recognition*, 88:13–26, April 2019.

[55] Renata De Paris, Christian V. Quevedo, Duncan D. Ruiz, Osmar Norberto de Souza, and Rodrigo C. Barros. Clustering Molecular Dynamics Trajectories for Optimizing Docking Experiments. *Computational Intelligence and Neuroscience*, 2015:e916240, March 2015.

[56] Hae-Sang Park and Chi-Hyuck Jun. A simple and fast algorithm for K-medoids clustering. *Expert Systems with Applications*, 36(2, Part 2):3336–3341, March 2009.

[57] Anna Gaulton, Anne Hersey, Michał Nowotka, A. Patrícia Bento, Jon Chambers, David Mendez, Prudence Mutowo, Francis Atkinson, Louisa J. Bellis, Elena Cibrián-Uhalte, Mark Davies, Nathan Dedman, Anneli Karlsson, María Paula Magariños, John P. Overington, George Papadatos, Ines Smit, and Andrew R. Leach. The ChEMBL database in 2017. *Nucleic Acids Research*, 45(D1):D945–D954, January 2017.

[58] Nabil El Tayar, Pierre Alain Carrupt, Han Van de Waterbeemd, and Bernard Testa. Modeling of .beta.-adrenoceptors based on molecular electrostatic potential studies of agonists and antagonists. *Journal of Medicinal Chemistry*, 31(11):2072–2081, November 1988.

[59] Marie Christiane Carre, Alphonse Youlassani, and Paul Caubere. Synthesis of a novel series of (aryloxy)propanolamines: New selective .beta.2-blocking agents. *Journal of Medicinal Chemistry*, 27(6):792–799, June 1984.

[60] C. Richard Crooks, Jeremy Wright, Patrick S. Callery, and J. Edward Moreton. Synthesis and preliminary biological studies of 4- and 5-[2-hydroxy-3-(isopropylamino)propoxy]benzimidazoles: Selective $B2$ adrenergic blocking agents. *Journal of Medicinal Chemistry*, 22(2):210–214, February 1979.

[61] Albert J Kooistra, Stefan Mordalski, Gáspár Pándy-Szekeres, Mauricio Esguerra, Alibek Mamyrbekov, Christian Munk, György M Keserű, and David E Gloriam. GPCRdb in 2021: Integrating GPCR sequence, structure and function. *Nucleic Acids Research*, 49(D1):D335–D343, January 2021.

[62] Katrin Altosaar, Poornima Balaji, Richard A. Bond, David B. Bylund, Susanna Cotecchia, Dominic Devost, Van A. Doze, Douglas C. Eikenburg, Sarah Gora, Eugénie Goupil, Robert M. Graham, Terry Hébert, J. Paul Hieble, Rebecca Hills, Shahriar Kan, Gayane Machkalyan, Martin C. Michel, Kenneth P. Minneman, Sergio Parra, Dianne Perez, Rory Sleno, Roger Summers, and Peter Zylbergold. Adrenoceptors in GtoPdb v.2021.3. *IUPHAR/BPS Guide to Pharmacology CITE*, 2021(3), September 2021.

[63] Greg Landrum, Paolo Tosco, Brian Kelley, Ric, Sriniker, Gedeck, Riccardo Vianello, NadineSchneider, Eisuke Kawashima, Andrew Dalke, Dan N, David Cosgrove, Brian Cole, Matt Swain, Samo Turk, AlexanderSavelyev, Gareth Jones, Alain Vaucher, Maciej Wójcikowski, Ichiru Take, Daniel Probst, Kazuya Ujihara, Vincent F. Scalfani, Guillaume Godin, Axel Pahl, Francois Berenger, JLVarjo, Strets123, JP, and DoliathGavid. Rdkit/rdkit: 2022_03_2 (Q1 2022) Release. Zenodo, April 2022.

[64] Jens Sadowski, Johann Gasteiger, and Gerhard Klebe. Comparison of Automatic Three-Dimensional Model Builders Using 639 X-ray Structures. *Journal of Chemical Information and Computer Sciences*, 34(4):1000–1008, July 1994.

[65] Oliver Korb, Thomas Stützle, and Thomas E. Exner. Empirical Scoring Functions for Advanced Protein-Ligand Docking with PLANTS. *Journal of Chemical Information and Modeling*, 49(1):84–96, January 2009.

3.2 Efficacy driven virtual screening using one class classification models trained on dynamic binding modes

Abstract

InteractionOCSVM is a new computational tool to identify ligands with a compatible binding mode to an ensemble of references extracted from a molecular dynamics simulation. The method has been previously tested as a tool to identify agonists of the β2-adrenergic receptor against a curated dataset of known antagonists, and true inactives. In this section the method was evaluated as a tool for the detection of ligands with a desired efficacy, either agonist or antagonists, against a panel of six diverse G-protein coupled receptors. The tool was evaluated not only based on its ability to select ligands with a certain efficacy over others, but also for its use in a virtual-screening like context. Filtering of the docking poses using InteractionOCSVM led to an increase in enrichment towards ligands with the desired efficacy compared to docking. While the method could not consistently outperform interaction similarity to a crystallographic reference, it proved to be more robust, often improving the results of poorly performing reference experimental structures.

3.2.1 Introduction

The previous section discussed the development of a new method for the identification of ADRB2 agonists based on the dynamic binding mode of a single reference agonist[1]. These models were called interaction graph one class support vector machine (IG-OCSVM)[2], based on the chosen interaction representation and machine learning algorithm. IG-OCSVMs were able to identify agonists of ADRB2 against known antagonists and experimentally tested inactive molecules.

Despite some promising results, the proof-of-concept work on ADRB2 was limited by the narrow evaluation conditions. ADRB2 is a well-known target, with a well-documented history of success for efficacy-driven drug discovery guided by protein-ligand interactions [3][4][5], which was further confirmed by the analysis described in Chapter 2. IG-OCSVM has been only applied to identify agonist ligands, disregarding whether the same results could be achieved for antagonists. While thoroughly studied, only the pseudo three-dimensional representation of interactions through IG was considered, neglecting simpler descriptions such as interaction fingerprints[6]. Additionally, the analysis focused on the application of the method for the post processing of a small number of ligands, without testing its application in a virtual screening-like context, measuring its selectivity against decoy molecules. This section aims at addressing these issues by testing the method on a more diverse set of targets, using both agonist and antagonist references, and measuring the selectivity against decoy molecules to assess its application in a virtual screening pipeline.

The IG-OCSVM approach was expanded to include interaction fingerprints (IFP)[6], calling this new approach IFP-OCSVM. IFPs represent a complementary interaction representation compared to IGs, since they forgo information about the 3D arrangement of interaction features, relying instead on an explicit labelling of the interacting residues. The IG-COSVM and IFP-OCSVM models will further be indicated as special case of a larger approach called InteractionOCSVM, which combines dynamic interactions data and one class classification to identify docking poses with a relevant binding mode. The InteractionOCSVM

approach was tested on six different GPCR targets: 5HT2A, AA2AR, ACM2, ADRB2, CNR1, and DRD2. These correspond to the same targets used in Chapter 1.1.3 to study the use of binding mode information from crystallographic structures to guide efficacy driven drug discovery, allowing for a comparison between the two approaches. 31 publicly available molecular dynamics (MD) simulations hosted on the GPCRmd platform [7] were used to extract relevant receptor conformations for structure based virtual screening, and for the training of different InteractionOCSVM models. This dataset contains both agonist and antagonist ligands, as well as references with a varying complexity of binding mode, ranging from the complex network of polar interactions of ADRB2, to the mostly hydrophobic binding modes observed for CNR1. The performances of the models in a virtual screening-like environment were tested based on the ability of the methods in selecting active molecules with the same efficacy as the reference ligands compared to property matched decoys.

3.2.2 Results

MD simulations analysis

Most of the simulations represent a stable complex between the receptor and its ligand, with no significant shifts in ligand position or binding mode, and RMSDs between 1.0 and 5 Å. In three cases (6MH8, 5G53, 6LUQ) shifts in binding modes linked to ligand unbinding were observed in one replica of the simulation, leading to the removal of those replicas from the training dataset. Between the analysed complexes two structures, one for AA2AR and one for CNR1, presented significant structural changes. In the complex between AA2AR and the xanthine XAC (PDB ID: 3REY), the ligand changes its position in the binding pocket in all three replicas at the beginning of the simulation, with an average RMSD from the initial structure between 8 and 12 Å. The structure of the binding site varied significantly in two replicas, while in the third it remained similar to the initial structure, excluding minor rearrangements (RMSD = 2 Å). The replica with the most stable conformation of the binding site, which also presented the most stable binding mode, was kept for model training. The complex between taranabant and CNR1 (PDB ID: 5U09), presents an interaction between the N-terminus of the receptor and the ligand, acting as a lid for the binding pocket. However, the N-terminus is truncated in both the crystallographic structure and in the simulation, where it was capped with an acetyl group. In two of the three replicas the N-terminal residue interacts directly with the ligand, which does not correspond to the natural behaviour of the complex. Only the structure with a stable conformation of the N-terminus was kept for model training.

The comparison between the average binding mode from the simulations with the corresponding experimental IFP, highlighted small differences in the formed interactions (Figure B.1.1-B.1.3). During the simulation certain interactions were systematically broken, mainly $\pi - \pi$ interactions, but also hydrogen bonds, as was the ones formed by Tyr316 of ADRB2. On the other hand, the simulations revealed certain known key interactions which were missing in the crystallographic structures, as the hydrogen bonds between HBI and Ser203 in ADRB2 (PDB ID: 4LDL), XAC and Glu169 in AA2AR (PDB ID: 3REY), and Asn404 in ACM2. Additionally, the simulations were able to smooth out the small differences in binding mode observed between different structures of the same protein-ligand complex.

Representative structures of the receptor were selected using the graph-based redundancy removal method developed by Mohammadi et al.[8]. The results were compared with binding

site clustering based on all non-hydrogen atoms RMSD. The graph analysis identified the same macro-differences in binding site conformation identified by clustering. The fitness of the extracted structures for protein-ligand docking was assessed by re-docking of the reference ligand, classifying a structure as acceptable if at least a docking pose with a RMSD of 2 Å or less could be generated. Only in few cases a representative structure was rejected due to redocking issues: AA2AR (4UHR, 5UVI,6MH8), ADRB2 (4LDL,6PS3,5U09), and DRD2 (6LUQ). For six of the 31 simulations only one structure was selected to represent the entire conformational ensemble due to the low diversity in binding site conformation, or the poor performance in redocking of the representative structures.

Ensemble docking

The agonists, antagonists, and decoys were docked in the ensemble of receptor conformations extracted from the MD simulations. The best docking score obtained by a ligand in the ensemble of structures extracted from one MD simulation was used to determine the final ranking. The performances of ensemble docking were evaluated based on the fraction of recovered agonists, and the fraction of recovered antagonists, after selecting the top 1% of the best ranked decoys.

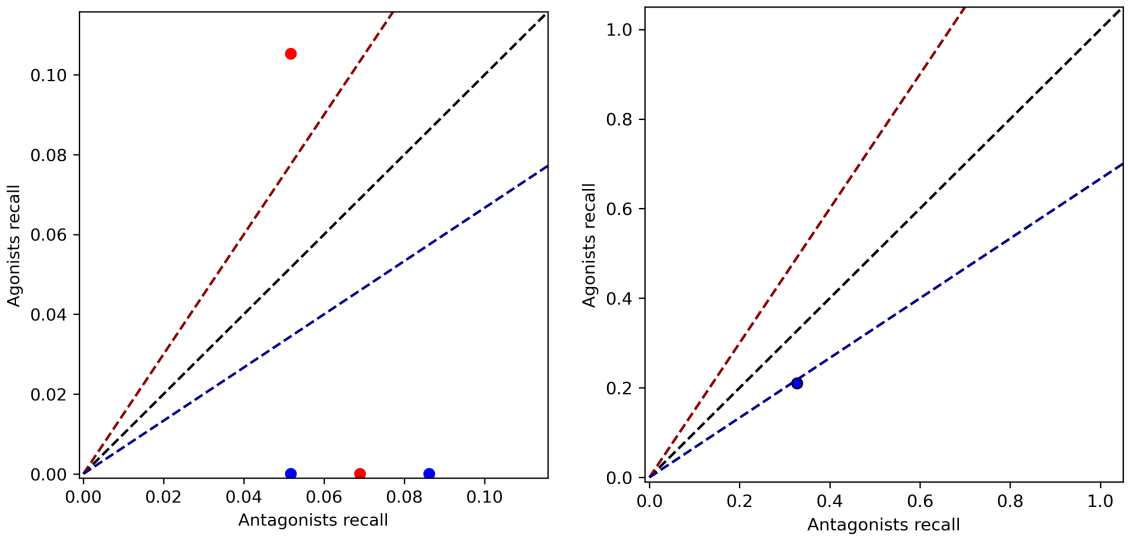
Ensemble docking led to an increase in the fraction of selected ligands with the desired efficacy compared to docking in a crystallographic structure (Table B.1.1) in 11 out of 31 structures, the same number of ligands was found in 14 cases, and only in six cases the performance decreased. In 16 cases ensemble docking did not identify any ligand with the same efficacy as the co-crystallized ligand, while for crystallographic structures it was 19. Significant increases in performance were observed for the 6A94 structure of 5HT2A, the 6LUQ structure of DRD2 (Figure 3.2.1), and the 3REY structure of AA2AR, for which the fraction of selected antagonists went from 5%, 0%, and 5%, to 31%, 19%, and 33% respectively. For six inactive state ADRB2 structural ensembles, docking could identify antagonist ligands, which were never identified by docking in a crystallographic structure. In the remaining two cases, the 5ZKB structure of ACM2 and the 4LDL structure from ADRB2, only small increases in recall were observed. The results on these new ensembles of structures confirm the results from Chapter 2, with the CHEMPLP scoring function [9] still presenting a bias, leading to a larger enrichment for ligands with one efficacy over the others, regardless of the structure used for docking.

InteractionOCSVM classification

The docking poses of the known agonists, antagonists, and decoys were evaluated using the trained InteractionOCSVM models. Poses obtained on an ensemble of structures were evaluated using the models trained on data from the same MD simulation. A total of 186 models were tested, based on 31 MD simulations, three different interaction representation (IFP[6], IG-Hyd[10], IG-Newhyd), and two different model training heuristics (QMS2[11] and MAD[1]).

The efficacy bias of each model was determined according to the ratio between the fraction of selected agonists and the fraction of selected antagonists (Figure B.1.6-B.1.10). The models were classified as biased, unbiased, inverse biased, and non-enriching, depending on this value. 109 models were efficacy biased, selecting ligands with the same efficacy as the reference molecule at a higher rate than other active molecules. 70 models were unbiased, selecting

Figure 3.2.1: Comparison between the fraction of retrieved agonist and antagonist of DRD2 using the experimental structures (A), and the structure extracted from a MD simulation (B).



(A) Experimental structure

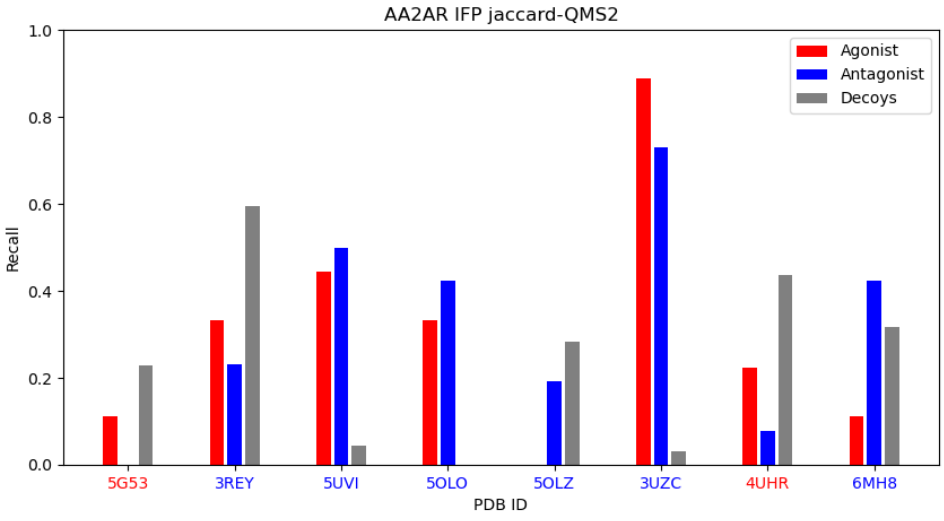
(B) MD structure

agonists and antagonists at a similar rate. Only the two IFP models describing the binding of taranabant to CNR1 (PDB iD: 5U09), and one of the models describing the binding of carvedilol to ADRB2 (PDB ID: 6PS3), had an inverse bias, selecting ligands with a different efficacy than the refence at a higher rate than ligands with the same efficacy. The four IG-OCSVM models trained on the binding mode of AF-DX 384 to ACM2 (PDB ID: 5ZKB) were non-enriching, since they did not identify any docking pose of known active molecules as reference-like. Based on the average efficacy bias ratio, IFP models were more selective towards ligands with the desired efficacy than the others, followed by IG-OCSVM models trained using the Hyd definition. Models trained using Newhyd interactions were often unbiased.

The precision of the models against decoys was determined by the fraction of selected decoys (Figure 3.2.2, B.1.6-B.1.10). Only 19 models selected less than 1% of all decoys, and 33 models less than 5%. Most of the models were not highly precise, with 133 of them selecting more than 10% of all decoys. Despite the lack of precision, an enrichment in ligands with the desired efficacy was observed in 142 cases compared to the initial dataset, with an enrichment factor of at least 2 for 94 models.

To overcome the low selectivity of the models against decoys, a consensus approach was used, considering a docking pose selected only if multiple InteractionOCSVM models identified its binding mode as reference-like. Four different ensembles of models were used (IFP-OCSVM and Hyd-IG-OCSVM, IFP-OCSVM and Newhyd-IG-OCSVM, Hyd-IG-OCSVM and Newhyd-IG-OCSVM, and all three models) for consensus classification (Figure B.1.11-B.1.16). Each group was only formed by models trained using the same heuristic. A total of 248 different models were examined, eight for each of the 31 MD simulations. The consensus models were more efficacy biased than the single models in

Figure 3.2.2: Fraction of agonist, antagonists, and decoys selected by IFP-OCSVM models trained on the binding modes of AA2AR ligands. The efficacy of the reference ligand is indicated by the color of the PDB ID label, red for agonists, and blue for antagonists. The AA2AR models present almost all of the different cases of InteractionOCSVM selectivity, with efficacy-biased and selective against decoys models (6MH8), models selective against decoys but not efficacy biased (5UVI, 5OLO, 3UZC), efficacy biased model not selective against decoys (5G53, 5OLZ, 4UHR), and models which were neither biased nor selective (3REY)



185 cases. An efficacy bias was observed for 190 models, no bias for 46, 2 presented an inverse efficacy bias, and 10 models were non-enriching. All the ensemble models were more selective against decoys, with 75 of them selecting less than 1% of all decoys, 120 less than 5% of all decoys, and only 78 models selected more than 10% of all decoys. Consequently, 161 consensus classifiers reported a higher enrichment than the corresponding single models, and 27 models yielded the same enrichment factor as the most enriching single model. An enrichment in ligands with the same efficacy as the reference molecule was observed for 200 models, with 169 of them yielding an enrichment factor of at least 2.

InteractionOCSVM filtering

The docking poses selected by the different InteractionOCSVM models, and consensus models, were ranked based on their docking score. The performance was evaluated based on the fraction of recovered agonists, and the fraction of recovered antagonists, after selecting 1% of all decoys (Figure B.1.17-B.1.22). In case less than 1% of decoys was selected by the InteractionOCSVM models, the fractions of agonists and antagonists identified by the model were reported. The effect of interaction filtering was evaluated by comparing the results with ensemble docking.

Filtering of docking poses using a single InteractionOCSVM model improved the results compared to ensemble docking in 129 cases, in 34 cases the results were unchanged, ad in 23 cases filtering led to the selection of a smaller number of relevant ligands. IFP-OCSVM models yield a higher average enrichment than IG-OCSVM models. ADRB2 models were generally the best performing ones, selecting between 20 to 30% of relevant ligands. InteractionOCSVM filtering and ranking yielded 122 cases of biased selection, 16 cases of unbiased selection, 18 cases of inverse bias, and in 30 cases the ranking was not enriching.

Filtering using the consensus InteractionOCSVM models led to better results than docking in 212 cases, to the same results in 15 cases, and to worse results in 21 cases. Model ensembles formed by all three models retrieved on average the smallest fraction of ligands, followed by models pairing IFP-OCSVM and IG-OCSVM. The pairing of the two IG-OCSVM models led to the best results. Good performances were recorded for all targets, selecting between 10 to 60% of all relevant ligands. Consensus classification and docking score ranking led to 194 efficacy biased selections, 5 inverse biased, 38 unbiased, and 11 non enriching. Compared to the single InteractionOCSVM classifiers, consensus models had better performances in 115 cases, maintained the same results in 36 cases, and performed worse in 97 cases. These performances however were not homogeneous between all the ensembles of models. While IFP-OCSVM+IG-OCSVM outperformed their corresponding single models in around a third of cases (48 out of 124), the pairing of the two IG-OCSVM models improved the performance in more than two third of cases (48 out of 62). Combining all the models together had the worst results, resulting in an improvement in less than a third of cases (19 out of 62).

Comparison to crystallographic references

The results obtained using InteractionOCSVM and ranking based on docking score were compared to the results from Chapter 1.1.3, which used experimental structures of the receptor for docking, and the binding modes of the co-crystallized ligands as references. Comparisons were made only between methods using the same interaction representation and interaction similarity measure, therefore IFP-OCSVM models were compared to IFP

similarity, while IG-OCSVM models were compared to graph kernel similarity using the same definition of hydrophobic contacts. Consensus models were compared to both of their corresponding similarity measures, considering the best performance as the benchmark.

IFP-OCSVM filtering paired with docking score ranking identified a larger number of ligands with the desired efficacy in 23 cases out of 62, compared to docking in the crystallographic structure and IFP similarity. IG-OCSVM using Hyd interactions outperformed the ranking based on graph kernel similarity in 13 cases, while with Newhyd interactions only in 16 cases. InteractionOCSVM models rarely could improve the results obtained using docking and interaction similarity based on the crystallographic structure. Only for few targets an improvement in performances was observed for more than one interaction representation: 5HT2A (6A94), AA2AR (3REY), ADRB2 (4GBR, 4LDL, 3NY8, 3NY9), CNR1 (5U09), and DRD2 (6LUQ).

For the consensus models, the IFP-OCSVM+Hyd IG-OCSVM outperformed simple interaction similarity in 26 out of 62 cases, IFP-OCSVM+Newhyd IG-OCSVM in 30 cases, and the two IG-OCSVM models outperformed the graph kernels in 45 cases. Combining all models yielded the worst performances, giving better results than interaction similarity only in 21 cases. Consensus models consistently improved the results over interaction similarity for all the studied targets: 5HT2A (6A94), AA2AR (3REY, 5UVI, 5OLO, 3UZC, 6MH8), ACM2 (4MQS, 3UON), ADRB2(6PS4, 3D4S, 3NY8, 6PS1, 3NYA, 4LDL), CNR1(5U09, 5XRA), DRD2(6LUQ).

3.2.3 Discussion

The InteractionOCSVM approach, expanded to include an IFP representation of interactions, was tested in a virtual screening-like environment, using reference ligands with different efficacies, and a panel of different GPCR targets for evaluation.

Ensemble docking using structures extracted from MD simulations yielded better or equivalent results than docking on crystallographic structures in most cases. For three protein-ligand complexes the simulation structures were significantly better for virtual screening compared to the respective crystallographic structures. This was especially important for the 6A94 structure of 5HT2A, and the 6LUQ structure of DRD2, since only a single structure was used for docking. Improvements in docking results were not necessarily caused by the larger number of considered structures, but rather by the better quality of the simulated structures for docking. Outside these specific cases there was no significant improvement in efficacy bias using multiple structures with the proposed scoring scheme, indicating either that the differences between the conformations were not relevant, or the used scoring function is not able to properly determine the effect of these structural differences.

The trained InteractionOCSVM models were used to classify the generated docking poses of known agonists, antagonists, and decoys. The models generally showed a bias towards ligands with the same efficacy as the reference, compared to other active molecules. Based on a qualitative comparison between the selectivity of the model and the binding mode of the reference, it seems that the presence or absence of polar interactions associated to a certain efficacy is one of the key determinants behind the specificity of the models. Unlike what was described in the original implementation[1], InteractionOCSVM models often lacked in selectivity against decoys. The models were unable to discard decoy binding modes, making it impossible to directly use the classification results for virtual hit selection. Interestingly, the only simulation used in both benchmarking studies (PDB ID: 4LDL) generated some of

the best performing models, with a clear efficacy bias and a high selectivity against decoys, selecting only ∼5% of them. The lack of precision of the single InteractionOCSVM models can be alleviated by combining multiple models trained on different interaction representations. The consensus models drastically decreased the number of selected decoys' docking poses, while at the same time increasing the enrichment factor for ligands with the same efficacy as the training reference. It appears that while the single models have broad definitions of an acceptable binding mode, their overlap is narrow, mostly discarding decoys or ligands with a different efficacy.

InteractionOCSVM models were not trained on ligand data, but on binding mode data, therefore they do not evaluate the affinity or efficacy of a ligand, but rather the fitness of its binding mode. Ligands were considered as selected if docking could generate a pose compatible with the training set binding modes, however no information about the quality of the docking pose outside of the formed interactions were considered. The classification from InteractionOCSVM models should be used to select relevant docking poses, which need to be ranked based on an alternative score. Filtering of the docking poses improved the enrichment compared to ranking based on a scoring function. Interestingly, interaction filtering removed the efficacy bias observed for the CHEMPLP scoring function. Docking can generate a pose resembling the reference binding mode for both true ligands and decoys, but only molecules with the same efficacy as the reference can also achieve a good docking score for that pose. The results from InteractionOCSVM were further improved by the consensus InteractionOCSVM models. The models obtained by paring together the two IG-OCSVM models performed better than the ones containing IFP-OCSVM models, despite the letters often performing better than the other two models. It is possible that the more precise IFP-OCSVM models perform better alone since they can already discard most of the decoys and irrelevant poses, but combining them with another model leads to a diminishing return in performance.

Unfortunately, the dynamic interactions extracted from MD simulation could not consistently outperform the results obtained using the crystallographic structure. In a few cases all single InteractionOCSVM models trained on a reference simulation outperformed the crystallographic binding mode. This success might be linked to two contributions, one being ensemble docking, and the other InteractionOCSVM filtering. Ensemble docking alone could already greatly improve the enrichment in true ligands compared to docking in the respective crystallographic structures for 5HT2A (6A94), DRD2 (6LUQ), and AA2AR (3REY). While InteractionOCSVM could further improve their efficacy specific recalls, the largest contribution to their performance is caused by the improved docking results. In the other cases, InteractionOCSVM generally outperformed interaction similarity if the corresponding experimental structure and binding mode reference led to a low enrichment in active molecules.

The lack of consistent improvement in performances suggest that the crystallographic binding mode represents the optimal binding mode reference for virtual screening. Therefore, the inclusion of noise, or a fuzzy definition of the binding mode by use of an ensemble leads to a suboptimal description of the complex. On the other hand, a poor reference binding mode benefits from InteractionOCSVM, since the ensemble definition better represent the optimal binding mode. Based on the analysis of the few cases where InteractionOCSVM outperforms interaction similarity, without a significant contribution from docking, it was not possible to determine which aspects might determine an increase in performance. For example, for ADRB2 the average binding modes and the crystallographic binding modes are similar, with

only few differences between them. The most relevant differences concern hydrogen bonds, like the interaction with Tyr316, observed in multiple crystallographic structures but never observed as stable during simulations, or the known critical residue Ser203, which interacted with the agonist HBI only in the MD simulations. While the introduction of these interactions in the binding mode could increase the fraction of recovered ligands, it is not possible to determine if this was the only relevant contribution. Other differences in binding mode, especially hydrophobic contacts and $\pi - \pi$ interactions, were observed, but they do not seem to influence the outcome in most cases. Since no clear a priori marker of success exists, it is recommended to evaluate both approaches in a retrospective manner using known actives and decoys, selecting the best performing method between the two.

3.2.4 Conclusions

Binding mode information extracted from MD simulations proved to be a useful tool to identify ligands with a similar efficacy as the reference. The developed InteractionOCSVM pipeline can leverage both an ensemble of binding site conformations for docking, as well as an analysis of the binding mode with a built-in threshold separating relevant from non-relevant interactions. Unfortunately, the InteractionOCSVM approach could not improve the performances of high-quality crystallographic data, but the models were more robust to missing interactions or unstable contacts compared to experimental references.

3.2.5 Materials and Methods

Molecular dynamics simulations

MD simulations based on the crystallographic structures of GPCR-ligand complexes described in Chapter 2 were downloaded from the GPCRmd webpage[7]. Of the 86 considered structures, only 31 were available for download (Table 3.2.1). Out of these the simulation of the 4LDO structure of ADRB2 was discarded given the poor performances of the trained models described in Section 3.1[1]. The simulation of the 4LDE structure of ADRB2 was also discarded since the critical residue Asp113 was simulated in an incorrect protonation state. The 5WF5 structure describing AA2AR bound to the agonist UK-432097 was discarded since it represent a constitutively active mutant of the receptor, and due to the large size of the ligand (MW = 777.9 Da) compared to other molecules in the dataset.

Molecular dynamics simulations analysis

The MD simulations were analysed using pytraj[12, 13]. The simulations were converted into pairs of mol2 files containing the receptor and the ligand. These files were used to compute protein ligand interactions using IChem's IFP and Ints modules[14]. The average ligand binding modes in each replica, and in the full simulation were computed.

The trajectories were aligned on the reference crystallographic structures based on the positions of the Cαs of their transmembrane domains, according to GPCRdb annotation[15][16]. The full receptor RMSD on the Cαs was computed to identify large structural rearrangement in receptor's conformation. Ligand RMSD was computed to identify large shifts in ligand binding mode, or to detect the initiation of ligand unbinding. Based on the analysis of the ligand's RMSD and the average IFP, replicas presenting ligand unbinding

Table 3.2.1: Selected, missing, and discarded MD simulation from GPCRmd based on the structures evaluated in Section 2.1.

Target	Selected simulations	Missing simulations	Discarded simulations
5HT2A	6A94	6WGT, 6WH4, 6WHA, 7RAN, 7VOD, 7VOE, 7WC4, 7WC5, 7WC6, 7WC7, 7WC8, 7WC9	6A93
AA2AR	3REY, 3UZC, 4UHR, 5G53, 5OLO, 5OLZ, 5UVI, 6MH8	2YDV, 3PWH, 3QAK, 3RFM, 3VGA, 5WF6, 6GDG, 6S0L, 7ARO, 7EZC, 7T32, 8DU3, 8GNG	5WF5, 3RFM
ACM2	3UON, 4MQS, 5ZK3, 5ZKB, 5ZKC	5YC8, 5ZKC, 7T8X	5ZK8
ADRB2	3D4S, 3NY8, 3NY9, 3NYA, 4GBR, 4LDL, 5D6L, 5JQH, 6PRZ, 6PS0, 6PS1, 6PS2, 6PS3, 6PS4, 6PS6	2RH1, 3P0G, 3SN6, 5D5A, 5D5B, 6E67, 6MXT, 6NI3, 7BZ2, 7DHI, 7DHR	4LDE, 4LDO, 6PS5
CNR1	5U09, 5XRA	5TGZ, 5XR8, 6KPG, 6N4B, 7V3Z	
DRD2	6LUQ	6VMS, 7DFP, 7JVR	
OPRK		6B73, 6VI4	

or other undesired phenomena, as described by the guidelines in Section 3.1, were removed from the dataset.

Binding site structure clustering

The binding site of the receptor was defined as being formed by all residues interacting with the ligand in at least 10% of the saved frames in the trajectory.

The structures were aligned on the receptor residues, and the non-hydrogen atoms RMSD compared to the initial structure, and the pairwise RMSD, were computed. Binding site clustering based on pairwise RMSD was performed using the agglomerative clustering algorithm with average linkage.

The graph-based redundancy removal method developed by Mohammadi et al.[8] was used to select representative conformations of the binding site. Once the graph structure was determined, the node with the highest degree (number of edges connected to that node) was selected as the most representative. The node corresponding to the most representative structure, and all of its neighbouring nodes were removed from the graph. The procedure was repeated iteratively until all nodes were removed. The three structures corresponding to the nodes with the highest degree, number of other connected nodes, were selected as representative structures. A smaller number of representative structures was selected they represented at least 90% of all structures in the ensemble.

Redocking

The fitness of the selected structures for docking was determined by the ability of docking in reproducing the binding mode observed in the representative frame with a non-hydrogen atoms RMSD < 2 Å. In case re-docking of the ligand was not possible the structure was discarded. Docking was performed using PLANTS with the scoring function CHEMPLP at search speed 1 [9].

Model training

IG-OCSVM models were trained using the same training work-flow described in Section 3.1 (Figure 3.1.7)[1], considering the adjustment described in Section 5.1 (Figure 5.1.7)[2]. To allow a clear and unbiased evaluation of the method both definition of hydrophobic contacts were used, as well as the two main training heuristics (QMS2[11] and MAD[1]) were used. IFP-OCSVM models used a Tanimoto similarity kernel for prediction, by computing the distance between fingerprints using the jaccard metric from scipy[17]. IG-OCSVM models instead used the shortest path kernel to compute the similarity between IGs [18].

Docking

The same active molecules and decoys from Section 2.2 were docked in the ensemble of structures, using the same docking protocol. Molecules were ranked based on the best docking score obtained between all the structures in each ensemble.

Performance evaluation

A molecule was considered as selected by an InteractionCOSVM model if at least one of its generated docking poses was positively classified. The performance of the InteractionOCSVM

models in selecting ligands with a desired efficacy was measured based on the ratio between the fraction of identified agonists, and the fraction of identified antagonists. The models were classified as agonist biased if the ratio value was > 1.5, antagonists biased if the ratio value was < 0.67, and unbiased otherwise. InteractionOCSVM models not selecting any ligands were classified as non-enriching.

In the virtual screening against a library containing property matched decoys the InteractionOCSVM models were used to filter the docking poses, removing all poses with a binding mode identified as different from the training data by the InteractionOCSVM model. From the filtered poses the performances were measured based on the ratio between the fraction of agonists, and the fraction of antagonists with a higher docking score than the top 1% of all ranked decoys. The same thresholds for efficacy bias used for the screening of the agonist-antagonist library were used.

References

[1] Luca Chiesa and Esther Kellenberger. One class classification for the detection of $B2$ adrenergic receptor agonists using single-ligand dynamic interaction data. *Journal of Cheminformatics*, 14(1):74, October 2022.

[2] Luca Chiesa, Emilie Sick, and Esther Kellenberger. Predicting the duration of action of $B2$-adrenergic receptor agonists: Ligand and structure-based approaches. *Molecular Informatics*, 42(12):e202300141, December 2023.

[3] Chris de Graaf, Nicolas Foata, Ola Engkvist, and Didier Rognan. Molecular modeling of the second extracellular loop of G-protein coupled receptors and its implication on structure-based virtual screening. *Proteins: Structure, Function, and Bioinformatics*, 71(2):599–620, 2008.

[4] Albert J. Kooistra, Rob Leurs, Iwan J. P. de Esch, and Chris de Graaf. Structure-Based Prediction of G-Protein-Coupled Receptor Ligand Function: A β-Adrenoceptor Case Study. *Journal of Chemical Information and Modeling*, 55(5):1045–1061, May 2015.

[5] Albert J. Kooistra, Henry F. Vischer, Daniel McNaught-Flores, Rob Leurs, Iwan J. P. de Esch, and Chris de Graaf. Function-specific virtual screening for GPCR ligands using a combined scoring method. *Scientific Reports*, 6(1):28288, September 2016.

[6] Gilles Marcou and Didier Rognan. Optimizing Fragment and Scaffold Docking by Use of Molecular Interaction Fingerprints. *Journal of Chemical Information and Modeling*, 47(1):195–207, January 2007.

[7] Ismael Rodríguez-Espigares, Mariona Torrens-Fontanals, Johanna K. S. Tiemann, David Aranda-García, Juan Manuel Ramírez-Anguita, Tomasz Maciej Stepniewski, Nathalie Worp, Alejandro Varela-Rial, Adrián Morales-Pastor, Brian Medel-Lacruz, Gáspár Pándy-Szekeres, Eduardo Mayol, Toni Giorgino, Jens Carlsson, Xavier Deupi, Slawomir Filipek, Marta Filizola, José Carlos Gómez-Tamayo, Angel Gonzalez, Hugo Gutiérrez-de-Terán, Mireia Jiménez-Rosés, Willem Jespers, Jon Kapla, George Khelashvili, Peter Kolb, Dorota Latek, Maria Marti-Solano, Pierre Matricon, Minos-Timotheos Matsoukas, Przemyslaw Miszta, Mireia Olivella, Laura Perez-Benito, Davide Provasi, Santiago Ríos, Iván R. Torrecillas, Jessica Sallander, Agnieszka

Sztyler, Silvana Vasile, Harel Weinstein, Ulrich Zachariae, Peter W. Hildebrand, Gianni De Fabritiis, Ferran Sanz, David E. Gloriam, Arnau Cordomi, Ramon Guixà-González, and Jana Selent. GPCRmd uncovers the dynamics of the 3D-GPCRome. *Nature Methods*, 17(8):777–787, August 2020.

[8] Sara Mohammadi, Zahra Narimani, Mitra Ashouri, Rohoullah Firouzi, and Mohammad Hossein Karimi-Jafari. Ensemble learning from ensemble docking: Revisiting the optimum ensemble size problem. *Scientific Reports*, 12(1):410, January 2022.

[9] Oliver Korb, Thomas Stützle, and Thomas E. Exner. Empirical Scoring Functions for Advanced Protein-Ligand Docking with PLANTS. *Journal of Chemical Information and Modeling*, 49(1):84–96, January 2009.

[10] Jérémy Desaphy, Eric Raimbaud, Pierre Ducrot, and Didier Rognan. Encoding Protein–Ligand Interaction Patterns in Fingerprints and Graphs. *Journal of Chemical Information and Modeling*, 53(3):623–637, March 2013.

[11] Zahra Ghafoori, Sarah M. Erfani, Sutharshan Rajasegarar, James C. Bezdek, Shanika Karunasekera, and Christopher Leckie. Efficient Unsupervised Parameter Estimation for One-Class Support Vector Machines. *IEEE Transactions on Neural Networks and Learning Systems*, 29(10):5057–5070, October 2018.

[12] Daniel R. Roe and Thomas E. Cheatham. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation*, 9(7):3084–3095, July 2013.

[13] Hai Nguyen, Jason Swails, Daniel R. Roe, and Cody. Amber-Md/Pytraj: V2.0.0. Zenodo, March 2017.

[14] Franck Da Silva, Jeremy Desaphy, and Didier Rognan. IChem: A Versatile Toolkit for Detecting, Comparing, and Predicting Protein–Ligand Interactions. *ChemMedChem*, 13(6):507–510, 2018.

[15] Albert J Kooistra, Stefan Mordalski, Gáspár Pándy-Szekeres, Mauricio Esguerra, Alibek Mamyrbekov, Christian Munk, György M Keserű, and David E Gloriam. GPCRdb in 2021: Integrating GPCR sequence, structure and function. *Nucleic Acids Research*, 49(D1):D335–D343, January 2021.

[16] Vignir Isberg, Bas Vroling, Rob van der Kant, Kang Li, Gert Vriend, and David Gloriam. GPCRDB: An information system for G protein-coupled receptors. *Nucleic Acids Research*, 42(D1):D422–D425, January 2014.

[17] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, SciPy 1.0

Contributors, Aditya Vijaykumar, Alessandro Pietro Bardelli, Alex Rothberg, Andreas Hilboll, Andreas Kloeckner, Anthony Scopatz, Antony Lee, Ariel Rokem, C. Nathan Woods, Chad Fulton, Charles Masson, Christian Häggström, Clark Fitzgerald, David A. Nicholson, David R. Hagen, Dmitrii V. Pasechnik, Emanuele Olivetti, Eric Martin, Eric Wieser, Fabrice Silva, Felix Lenders, Florian Wilhelm, G. Young, Gavin A. Price, Gert-Ludwig Ingold, Gregory E. Allen, Gregory R. Lee, Hervé Audren, Irvin Probst, Jörg P. Dietrich, Jacob Silterra, James T Webber, Janko Slavič, Joel Nothman, Johannes Buchner, Johannes Kulick, Johannes L. Schönberger, José Vinícius de Miranda Cardoso, Joscha Reimer, Joseph Harrington, Juan Luis Cano Rodríguez, Juan Nunez-Iglesias, Justin Kuczynski, Kevin Tritz, Martin Thoma, Matthew Newville, Matthias Kümmerer, Maximilian Bolingbroke, Michael Tartre, Mikhail Pak, Nathaniel J. Smith, Nikolai Nowaczyk, Nikolay Shebanov, Oleksandr Pavlyk, Per A. Brodtkorb, Perry Lee, Robert T. McGibbon, Roman Feldbauer, Sam Lewis, Sam Tygier, Scott Sievert, Sebastiano Vigna, Stefan Peterson, Surhud More, Tadeusz Pudlik, Takuya Oshima, Thomas J. Pingel, Thomas P. Robitaille, Thomas Spura, Thouis R. Jones, Tim Cera, Tim Leslie, Tiziano Zito, Tom Krauss, Utkarsh Upadhyay, Yaroslav O. Halchenko, and Yoshiki Vázquez-Baeza. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, March 2020.

[18] K.M. Borgwardt and H. Kriegel. Shortest-Path Kernels on Graphs. In *Fifth IEEE International Conference on Data Mining (ICDM'05)*, pages 74–81, Houston, TX, USA, 2005. IEEE.

3.3 Critical evaluation

Structure-based virtual screening relays on the three-dimensional structure of a protein to identify new molecules able to bind one of its pockets. Despite proteins being flexible entities in reality, they are generally treated as rigid in structure-based virtual screening. Most of the routinely used docking algorithms treat the protein as a rigid structure, while the ligand is treated as flexible, identifying the pose which better complement the shape and features of the binding pocket. More complex algorithms can consider the protein as a flexible entity, although this is only limited to its sidechains conformations[1]. Larger structural rearrangements could be considered using ensemble docking, which uses multiple structures of the receptor, either experimental or extracted from molecular dynamics (MD) simulations, to perform docking[2]. The binding mode of a ligand is also flexible, with rearrangements of the molecule in the binding site, and more and less conserved interactions. However, the use of dynamic interaction data in structure-based virtual screening has been generally limited. Ensembles of protein-ligand interactions extracted from multiple experimental structures, or from MD simulations, have been mostly used to develop structure-based pharmacophores [3]. Dynophores [3] combine binding mode information extracted from a MD simulation of the protein-ligand complex in a single pharmacophore query. Each pharmacophoric features is formed by a three-dimensional density clouds, and it is represented by a probability density function. By using the probability density it is possible to use the dynophores for virtual screening. The Common Hits Approach[4] instead extracts multiple pharmacophores from a MD simulation and pools them to avoid redundancy. The ensemble of queries is used for virtual screening, by evaluating a molecule based on the number of pharmacophores selecting it as active. MYSHAPE[5] instead combines shared interactions formed by different ligands in MD simulations to determine a common pharmacophore used for screening, or to apply constrains during docking based on the location of specific interacting features. GRAIL[6] instead is a grid-based method, which identifies binding hotspots from MD simulations of protein-ligand complexes. LID [7] is another grid-based method which relays on an ensembles of experimental structures of the target protein bound to fragment ligands, which act as probes identifying relevant interaction hotspots. Unlike the other methods LID is not based on MD data, however it could be adapted to build interaction grids using data from simulations.

The described pharmacophore and grid methods either consider the frequency of single interactions, or aim at identifying a consensus score between the extracted binding information. InteractionOCSVM represent an alternative approach compared to the described methods. The evaluated binding mode, representing the pharmacophore of the molecule, is not ranked based on a consensus or ensemble scores, or based on the relative importance of single features, but rather it is classified as comparable to the references or not. A binding mode is defined as comparable to the references if it is not distinguishable from another binding mode extracted from the reference MD simulation. Classification is achieved by using a one class classification algorithm[8, 9], from the outlier detection family of machine learning. One class classification algorithms, in this case a one class support vector machine (OCSVM)[10], have been developed to classify new instances as being extracted from the same distribution originating the training set, or representing a novel object. These models are trained on a dataset containing only examples of the target class, in a process called semi-supervised learning. In the described application, a model was trained on MD generated binding modes of one ligand, which should represent a realistic depiction of the true

binding mode of the molecule. The trained models were applied to evaluate docking-generated binding modes as comparable to the ensemble of references, or as different from the references. A similar classification could be achieved by combining interaction similarity with a similarity threshold value, separating relevant from not relevant poses[11–15]. InteractionOCSVM presents two advantages over this simpler approach, thanks to its use of ensemble information. Classification is based on the presence of specific interactions or interaction patterns, with the most stable interactions having a greater weight compared to unstable contacts, rather than considering all interactions as contributing equally. The threshold used for classification is determined based on the training data, rather than on an a priori threshold value, which does not take into account the specific properties of the reference or the target.

InteractionOCSVM identifies molecules which can potentially bind the target protein with a specific binding mode, making it a perfect tool to identify new ligands of G-protein coupled receptors with a desired efficacy. A relationship between a ligand's binding mode and its efficacy is well established in the literature [16–18]. Additionally, since the trained models are not limited to interaction similarity, but they determine if a pose is compatible with the target binding mode or not, they should be robust against biased chemical libraries[19]. A small number of hits selected by InteractionOCSVM, or a complete lack thereof, might indicate that the fraction of explored chemical space does not contain molecules with the desired pharmacophoric features.

InteractionOCSVM was initially used to identify agonists of the β2-adrenergic receptor (ADRB2) against antagonists and experimentally validated inactives. Other than the potentialities of the method, this study highlighted the key role of a careful analysis of the training data. Initial guidelines on trajectory analysis and parameter selection where given, however they are of qualitative nature, since it was not possible to determine all the aspects influencing the quality of the reference binding mode. In the end, the evaluation of the training set will have to relay on expert knowledge of the target receptor, to determine if significant shifts in binding mode, or ligand unbinding initiation, are observed in the simulation.

The application of the method on a panel of six diverse GPCRs, using both agonist and antagonist references, allowed to obtained a better picture of the methods' performances. The models were tested in a virtual screening-like scenario, using property matched decoys generated based on known agonists and antagonists. Unlike what was previously observed for ADRB2, the models were not necessarily selective towards agonists or antagonists, and furthermore they were not always able to discard decoys. While InteractionOCSVM could not efficiently discard all decoy molecules, ranking the positively classified docking poses based on their docking score increased both the hit rate, and the efficacy selectivity of the hits, compared to docking. Despite the possible advantages of InteractionOCSVM filtering and ranking using a scoring function, it had worse performances than docking in the crystallographic structure and ranking based on interaction similarity to the experimental binding mode. The crystallographic binding mode seems to often represent the optimal reference for ranking based on interaction similarity. The introduction of noise or fuzziness from the ensemble of binding modes leads to poorer performances than the original reference. However, InteractionOCSVM models were able to clearly improve the results compared to experimental references with a low hit retrieval. The dynamic binding mode extracted from MD simulations represent a better approximation of the "optimal" binding mode, than a poor quality crystallographic structure.

Since InteractionOCSVM uses binding information from MD simulations for its

predictions, it has been described as a particularly adapt tool to use data from in silico models of protein-ligand complexes for virtual screening. Since the binding mode of a single reference ligand is the only information required to perform predictions, the method can be easily applied to study targets with limited available experimental data. While applications of the method on in silico structures are described in Chapters 4 and 5, no systematic benchmarking was performed. The limit of such analysis is the creation of plausible models in a high-throughput fashion, since building accurate structural models requires a careful analysis of the initial structure, the MD simulation, and specific knowledge about a target's structure and biology. Additionally, producing a large number of in silico models for the evaluation of InteractionOCSVM would introduce a third variable in method evaluation, other than the quality of the InteractionOCSVM model, and the docking power of the used docking software or representative structure, regarding the quality of the used model, as it will be further explored in Chapter 4.

Despite the in-depth analysis on InteractionOCSVM parameters, and the ligand references, two aspects have remained understudied: the influence of docking and structure selection, and the algorithm used for one class classification. In the initial study on ADRB2 agonists detection, the recall of the models was evaluated assuming that docking could generate a "correct" docking pose for all agonists, regardless of the used receptor conformation. In reality docking might not generate a "correct" pose for all relevant molecules due to limits in its docking algorithm, or because the ligand binds a different conformation of the receptor than the one used for docking (Chapter 1.1.3). In all of the performed studies PLANTS with the scoring function CHEMPLP was the only used docking protocol [20]. Different combinations of docking algorithms and scoring functions should be evaluated in order to determine the best protocol. InteractionOCSVM, as the name suggests, is based on the one class support vector machine (OCSVM) algorithm. OCSVM was selected over other traditional one class classification or outlier detection methods, like local outlier factor [21] or isolation forest[22], for two reasons: it is not limited to the analysis of tabular data, and the model automatically determines the threshold used for classification, without requiring to exactly estimate the fraction of outliers in the training set. While deep learning methods for outlier detection exist[9], they were not used to develop similar models, outside of few not reported preliminary studies. Deep one class classification presents two disadvantages compared to InteractionOCSVM regarding outlier classification and representation learning. Most of the models available in the literature do not classify instances as outliers or inliers, which represented one of the most interesting features of InteractionOCSVM, but rather determined a classification score which should be further interpreted[23]. InteractionOCSVM uses two interaction representations and similarity scores which were known to prioritize active molecules with a desired efficacy, over decoys. The interaction representation learned by the deep learning model, generally determined using an autoencoders for parameter pretraining [23], is not guaranteed to be meaningful. While feature learning on the tabular interaction fingerprint data is relatively straightforward, interaction graphs have been defined using arbitrary criteria. The generated learned representations should be tested, to evaluate if the distance between the embedding vectors maintain the same screening power as the original representations.

While the application of other one class classification algorithms could improve the performance and the efficiency of the method, the more interesting development might be the use of alternative datasets for model training. InteractionOCSVM was developed to use data extracted from MD simulations, however models could be trained using any ensemble

of binding modes, regardless of their origins. For examples, binding modes from multiple experimental structures, or binding mode generated by docking of known ligands[24][25] could be used as training data. The use of these alternative data sources will not require to adapt the InteractionOCSVM algorithm, but rather to build a new data cleaning and analysis pipeline, able to generate datasets containing ensembles of meaningful binding modes.

References

[1] Douglas B. Kitchen, Hélène Decornez, John R. Furr, and Jürgen Bajorath. Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nature Reviews Drug Discovery*, 3(11):935–949, November 2004.

[2] Rommie E. Amaro, Jerome Baudry, John Chodera, Özlem Demir, J. Andrew McCammon, Yinglong Miao, and Jeremy C. Smith. Ensemble Docking in Drug Discovery. *Biophysical Journal*, 114(10):2271–2278, May 2018.

[3] David Schaller, Dora Šribar, Theresa Noonan, Lihua Deng, Trung Ngoc Nguyen, Szymon Pach, David Machalz, Marcel Bermudez, and Gerhard Wolber. Next generation 3D pharmacophore modeling. *WIREs Computational Molecular Science*, 10(4):e1468, 2020.

[4] Marcus Wieder, Arthur Garon, Ugo Perricone, Stefan Boresch, Thomas Seidel, Anna Maria Almerico, and Thierry Langer. Common Hits Approach: Combining Pharmacophore Modeling and Molecular Dynamics Simulations. *Journal of Chemical Information and Modeling*, 57(2):365–385, February 2017.

[5] Ugo Perricone, Marcus Wieder, Thomas Seidel, Thierry Langer, Alessandro Padova, Anna Maria Almerico, and Marco Tutone. A Molecular Dynamics–Shared Pharmacophore Approach to Boost Early-Enrichment Virtual Screening: A Case Study on Peroxisome Proliferator-Activated Receptor α. *ChemMedChem*, 12(16):1399–1407, 2017.

[6] Doris A. Schuetz, Thomas Seidel, Arthur Garon, Riccardo Martini, Markus Körbel, Gerhard F. Ecker, and Thierry Langer. GRAIL: GRids of phArmacophore Interaction fieLds. *Journal of Chemical Theory and Computation*, 14(9):4958–4970, September 2018.

[7] Célien Jacquemard, Viet-Khoa Tran-Nguyen, Malgorzata N. Drwal, Didier Rognan, and Esther Kellenberger. Local Interaction Density (LID), a Fast and Efficient Tool to Prioritize Docking Poses. *Molecules*, 24(14):2610, January 2019.

[8] Shehroz S. Khan and Michael G. Madden. One-class classification: Taxonomy of study and review of techniques. *The Knowledge Engineering Review*, 29(3):345–374, June 2014.

[9] Naeem Seliya, Azadeh Abdollah Zadeh, and Taghi M. Khoshgoftaar. A literature review on one-class classification and its potential applications in big data. *Journal of Big Data*, 8(1):122, December 2021.

[10] Bernhard Schölkopf, John C. Platt, John Shawe-Taylor, Alex J. Smola, and Robert C. Williamson. Estimating the Support of a High-Dimensional Distribution. *Neural Computation*, 13(7):1443–1471, July 2001.

[11] Jérémy Desaphy, Eric Raimbaud, Pierre Ducrot, and Didier Rognan. Encoding Protein–Ligand Interaction Patterns in Fingerprints and Graphs. *Journal of Chemical Information and Modeling*, 53(3):623–637, March 2013.

[12] Inna Slynko, Franck Da Silva, Guillaume Bret, and Didier Rognan. Docking pose selection by interaction pattern graph similarity: Application to the D3R grand challenge 2015. *Journal of Computer-Aided Molecular Design*, 30(9):669–683, September 2016.

[13] Chris de Graaf, Albert J. Kooistra, Henry F. Vischer, Vsevolod Katritch, Martien Kuijer, Mitsunori Shiroishi, So Iwata, Tatsuro Shimamura, Raymond C. Stevens, Iwan J. P. de Esch, and Rob Leurs. Crystal Structure-Based Virtual Screening for Fragment-like Ligands of the Human Histamine H1 Receptor. *Journal of Medicinal Chemistry*, 54(23):8195–8206, December 2011.

[14] Albert J. Kooistra, Henry F. Vischer, Daniel McNaught-Flores, Rob Leurs, Iwan J. P. de Esch, and Chris de Graaf. Function-specific virtual screening for GPCR ligands using a combined scoring method. *Scientific Reports*, 6(1):28288, September 2016.

[15] I. Adlere, S. Sun, A. Zarca, L. Roumen, M. Gozelle, C. Perpiñá Viciano, B. Caspar, M. Arimont, J. P. Bebelman, S. J. Briddon, C. Hoffmann, S. J. Hill, M. J. Smit, H. F. Vischer, M. Wijtmans, C. de Graaf, I. J. P. de Esch, and R. Leurs. Structure-based exploration and pharmacological evaluation of N-substituted piperidin-4-yl-methanamine CXCR4 chemokine receptor antagonists. *European Journal of Medicinal Chemistry*, 162:631–649, January 2019.

[16] Yiran Wu, Liting Zeng, and Suwen Zhao. Ligands of Adrenergic Receptors: A Structural Point of View. *Biomolecules*, 11(7):936, June 2021.

[17] Chris de Graaf and Didier Rognan. Selective Structure-Based Virtual Screening for Full and Partial Agonists of the $B2$ Adrenergic Receptor. *Journal of Medicinal Chemistry*, 51(16):4978–4985, August 2008.

[18] Albert J. Kooistra, Rob Leurs, Iwan J. P. de Esch, and Chris de Graaf. Structure-Based Prediction of G-Protein-Coupled Receptor Ligand Function: A β-Adrenoceptor Case Study. *Journal of Chemical Information and Modeling*, 55(5):1045–1061, May 2015.

[19] David Rodríguez, Zhang-Guo Gao, Steven M. Moss, Kenneth A. Jacobson, and Jens Carlsson. Molecular Docking Screening Using Agonist-Bound GPCR Structures: Probing the A $_{2A}$ Adenosine Receptor. *Journal of Chemical Information and Modeling*, 55(3):550–563, March 2015.

[20] Oliver Korb, Thomas Stützle, and Thomas E. Exner. Empirical Scoring Functions for Advanced Protein-Ligand Docking with PLANTS. *Journal of Chemical Information and Modeling*, 49(1):84–96, January 2009.

[21] Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. LOF: Identifying density-based local outliers. *SIGMOD Rec.*, 29(2):93–104, May 2000.

[22] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation Forest. In *2008 Eighth IEEE International Conference on Data Mining*, pages 413–422, Pisa, Italy, December 2008. IEEE.

[23] Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft. Deep One-Class Classification. In *Proceedings of the 35th International Conference on Machine Learning*, pages 4393–4402. PMLR, July 2018.

[24] Mireia Jiménez-Rosés, Bradley Angus Morgan, Maria Jimenez Sigstad, Thuy Duong Zoe Tran, Rohini Srivastava, Asuman Bunsuz, Leire Borrega-Román, Pattarin Hompluem, Sean A. Cullum, Clare R. Harwood, Eline J. Koers, David A. Sykes, Iain B. Styles, and Dmitry B. Veprintsev. Combined docking and machine learning identify key molecular determinants of ligand pharmacological activity on $B2$ adrenoceptor. *Pharmacology Research & Perspectives*, 10(5):e00994, 2022.

[25] Tomohiro Sato, Teruki Honma, and Shigeyuki Yokoyama. Combining Machine Learning and Pharmacophore-Based Interaction Fingerprint for in Silico Screening. *Journal of Chemical Information and Modeling*, 50(1):170–185, January 2010.

Chapter 4

Design of new ligands targeting the C-X-C chemokine receptor type 4

The C-X-C chemokine receptor type 4 (CXCR4) is a member of the chemokine receptor subfamily of G-protein coupled receptors. Due to its importance as modulator of inflammatory response CXCR4 represent an interesting and underexploited target for drug development. This Chapter describes the development of virtual screening protocol for the discovery of a small molecule ligand with the same efficacy for the receptor as its endogenous ligand, the chemokine CXCL12. The described modelling and virtual screening pipeline was developed before the release of an experimental structure of the CXCL12-CXCR4 complex, therefore the rational design of new ligands was guided by information extracted from *in silico* models.

The first section describes the modelling of the CXCL12-CXCR4 complex based on an experiment-guided homology model. The in silico structure was simulated using molecular dynamics to validate the proposed binding mode, and to gather relevant data regarding the chemokine's binding mode. The homology model strcutre was compared to an alternative *in silico* model of the complex generated using AlphaFold-Multimer. An experimental structure of the complex was released after the development of the two *in silico* structural models, allowing to compare the quality of the predicted structures against the experimental reference.

The second section describes the use of the dynamic data extracted from the simulated homology model to guide the design of new small molecule ligands of the receptor. The virtual screening campaign applied the InteractionOCSVM method developed in Chapter 3. Four new ligands with a novel scaffold and a low micromolar potency for CXCL12 binding inhibition were identified. The identified ligands were not able to promote G-protein coupling, like the reference CXCL12, classifying them as potential antagonists of the receptor.

This chapter is the product of a collaboration with Dr Thiérry Durroux and Omolade Otun from the Institut de Génomique Fonctionnelle in Montpellier, who performed the experimental testing on the identified hit compounds.

Since both chemokines and chemokine receptors are proteins, in this chapter one letter codes for amino acids are used to reference chemokines' residues, while three letter codes are used to reference chemokine recepotrs' residues.

4.1 Modelling of the CXCL12-CXCR4 complex

Abstract

The interaction between the C-X-C chemokine receptor type 4 and its endogenous ligand, the chemokine CXCL12, plays an important role in immune response regulation. In order to study the recognition mechanism, and the allosteric effects caused by chemokine binding to the receptor structure, two *in silico* models of the complex were developed. An homology model based on experimental data was validated by performing molecular dynamics simulations of the complex, and an alternative model was generated using AlphaFold-Multimer, a deep learning based protein structure prediction tool. A comparison to the recently released structure of the complex confirmed the prediction from AlphaFold-Multimer, while the homology model resulted incorrect. The information regarding the chemokine's binding mode and the pocket's structure were used as the basis of a virtual screening campaign using protein-ligand docking.

4.1.1 Introduction

The C-X-C chemokine receptor type 4 (CXCR4) is a member of the Class A of the G-protein-coupled receptor family. CXCR4 is a member of the chemokine receptor subfamily, formed by GPCRs whose endogenous ligand is a chemokine, small soluble proteins responsible for immune response regulation. Chemokine receptors are further separated in subfamilies depending on the presence of a specific conserved disulphide bridge motif in the sequence of the bound chemokine.

As all class A GPCRs the structure of CXCR4 is formed by a bundle of seven hydrophobic transmembrane α-helices (TMs) connected by two sets of three flexible loops at the intracellular (ICLs) and extracellular side (ECLs), and with a long flexible N-terminus[1]. The main structural difference between chemokine receptors and other GPCRs is their larger and more open extracellular binding pocket, due to the nature of the bound ligands. This extracellular binding pocket is formed by two main subpockets, the major, and the minor subpockets[2]. Chemokines are small soluble proteins, whose structure is formed by a rigid core, stabilized by the presence of two disulfide bridges, and by a flexible N-terminus. The interaction between a chemokine and the receptor has been generally described using a two site model, based on two distinct chemokine recognition sites (CRS) at the protein-protein interface. CRS1 is formed by the interface between the flexible N-terminus of the receptor, and the rigid core of the chemokine. CRS1 has been described as the initiator of chemokine binding to the receptor, and it has been observed to regulate the selectivity of chemokine-chemokine receptors pairings. CRS2 instead is formed by the interface between the extracellular binding site of the receptor, and the flexible N-terminus of the chemokine. The efficacy of a chemokine fora receptor is linked by its binding to CRS2 alone, since truncations or mutations of a chemokine N-terminus led to a loss in efficacy, while peptides formed by a chemokine's N-terminus alone were instead observed to maintain their efficacy[3]. The analysis of the multiple recently resolved structures of chemokine-chemokine receptor complexes has highlighted the existance of two other recognition sites, named CRS1.5 [4] and CRS3[5]. An additional recognition site, called CRS0.5, was identified by mutagenesis and crosslinking experiments for CXCR4[6] and the homologous atypical chemokine receptor type 3 (ACKR3). This additional binding site is hypothesized to be more commonly observed for

members of the C-X-C receptor family, than for the CC receptor family, due to the relative orientation of the chemokine in respect to the receptor [2].

Chemokine receptors generally bind different chemokines, contributing to the complexity of the signalling system. CXCR4 however, represents an exception, since its only known endogenous ligand is the chemokine CXCL12.

Until 2024 only three crystallographic structure of CXCR4 were available, representing the receptor in the inactive state conformation bound to the small molecule IT1t[7], the peptide CVX15[7], and the viral chemokine vCCL2[4]. Despite the lack of an experimental structure, different putative binding and activation mechanism models were proposed for the CXCL12-CXCR4 complex based on mutagensis data. The work of Wescott et al. [8] was particularly notable since they performed a full-shot mutagenesis experiment on the receptor, identifying multiple residues responsible for chemokine engagement, signal initiation, signal propagation, microswitch activation, and G-protein coupling. The interactions between the N-terminus of the receptor and the chemokine core were determined through crosslinking experiments by Ngo et al. [6]. By combining the experimental data describing the interactions between CXCL12 and CXCR4 at CRS1 and CRS2, Ngo et al. were able to propose an experiment-driven homology model of the full length complex[6]. Further mutagenesis experiments were performed to validate the interactions described in the structural model[9]. Given the abundance of experimental data used for its construction, and its validation by a series of mutagenesis data, the homology model of Ngo et al. represented at the time the state-of-the-art description of the CXCL12-CXCR4 complex.

The release of AlphaFold in 2021[10] and of AlphaFold-Multimer in 2022[11] have revolutionized the field of protein modelling, simplifying the determination of the structure of both monomeric proteins, and of protein-protein complexes. As will be further described in Section 6.2, AlphaFold-Multimer performed well in a benchmark study on the prediction of chemokine-chemokine receptor structures[12], with all models presenting a correct chemokine orientation in respect to the receptor, and a plausible description of CRS2.

In this work the experiment-guided homology model developed by Ngo et al.[6] was simulated using molecular dynamics to determine the stability of the interactions formed by CXCL12 at CRS2. The binding site structure of the simulated CXCL12-bound receptor was compared to the binding site structures obtained by simulating the IT1t-CXCR4 complex, to identify key structural rearrangements suggesting a conformational shift towards an active or active-like conformation of the pocket. An alternative CXCL12-CXCR4 structure was modelled using AlphaFold-Multimer, and the accuracy of the prediction was evaluated by comparing the *in silico* structure with orthogonal data from mutagensis experiments.

In March 2024 the first experimental structure of the CXCL12-CXCR4 complex was published[13]. The resolution of another structure of the same complex was announced few months earlier, however the results are not yet available to the public [14]. The two *in silico* models of the chemokine-chemokine receptor complex were compared to the experimental structure to determine their accuracy.

4.1.2 Results

Critical evaluation of the homology model

The homology model proposed by Ngo et al. [6] describes the full-length complex between CXCL12 and CXCR4. The placement of the receptor's flexible N-terminus around the

chemokine core was determined using cross-linking data. The interactions between the chemokine's N-terminus and the receptor binding pocket were instead determined based on a previous model proposed by Wescott et al.[8]. Unlike the previous model CXCL12 is shown as engaging both the major and the minor binding pocket of the receptor. The binding of the N-terminus of CXCL12 to the extracellular binding pocket of CXCR4 describe the chemokine as entering through the major binding pocket and interacting mainly with the minor binding pocket. The chemokine is anchored in place by K1 forming two salt bridges with the critical residues Asp97 and Glu288, the former through the charged N-terminal amine, while the latter using the residue's sidechain. Removing the two positive charges from the first residue by truncation or alanine mutation [15] turns CXCL12 into an antagonist, while mutations to amino acids with a similar pharmacophore to Lys, like Arg[16] or Orn[17], maintained the agonist efficacy. The N-terminus extends deeply inside the pocket, with P2 interacting with Trp94 at the bottom of the binding cavity. The loss of this interaction by mutating P2 to Gly turns CXCL12 into an antagonist[17]. While the model can clearly explain the effect of the mutations of K1 and P2 on the chemokine's efficacy, it is not able to explain the lack of changes in efficacy upon the addition of an extra residue in N-terminal position[17–19]. While a rearrangement of the chemokine in the pocket to accommodate the extra residue might be possible, the effect of such mutations was not properly addressed by the proposed structure. Both the entering of CXCL12 through the major binding pocket, and its deep binding mode have been observed in other structures of chemokine-CKR complexes. As noted by Urvas and Kellenberger, chemokines from the CC and CX3C subfamilies enter the receptors' binding site through the minor pocket, while the CXC chemokine CXCL8 enters the binding site of CXCR2 through the major binding pocket[2]. CXCL8 however presents a shallow binding mode, interacting with the residues at the bottom of the pocket through the long sidechain of K3, while in the proposed structure the chemokine binds deeply into the pocket, as observed in the CXCL12-ACKR3 complex, and for other chemokines[2]. However, a deep binding mode is generally paired with a turn in the N-terminus backbone structure, placed in the proximity of $Glu^{7.39}$ (Ballesteros-Weinstein notation [20]). The backbone nitrogens of the residues forming this turn generally form hydrogen bonds with the carboxylic group of $Glu^{7.39}$[2]. In the analysed structure the backbone of CXCL12's N-terminus is mostly linear, without any turn. The proposed binding mode is a hybrid between the common deep binding mode, due to the direct interaction with residues situated at the bottom of the cavity, and the lack of a turn in the N-terminus structure, more common for shallow binding modes.

Molecular dynamics simulations

All-atom MD simulations of the receptor in complex with the agonist CXCL12 and bound to the antagonist IT1t were performed. The simulations describe the receptor inserted in a lipid membrane, with the system solvated using explicit water molecules. The aim was to evaluate the stability of the proposed binding mode, as well as possible induced-fit effects caused by the binding of CXCL12 to the inactive state receptor binding pocket structure. The simulation of the receptor bound to an antagonist was used as a control reference, to identify relevant differences in ligand binding mode or receptor structure. The binding mode of the antagonist IT1t with the receptor was not stable during the simulation, with the initial structure during the production trajectories already presenting a RMSD of 3.1 Å compared to the crystallographic pose (Figure C.1.1). During the simulation the ligand underwent through further rearrangements, either returning towards a more similar pose compared

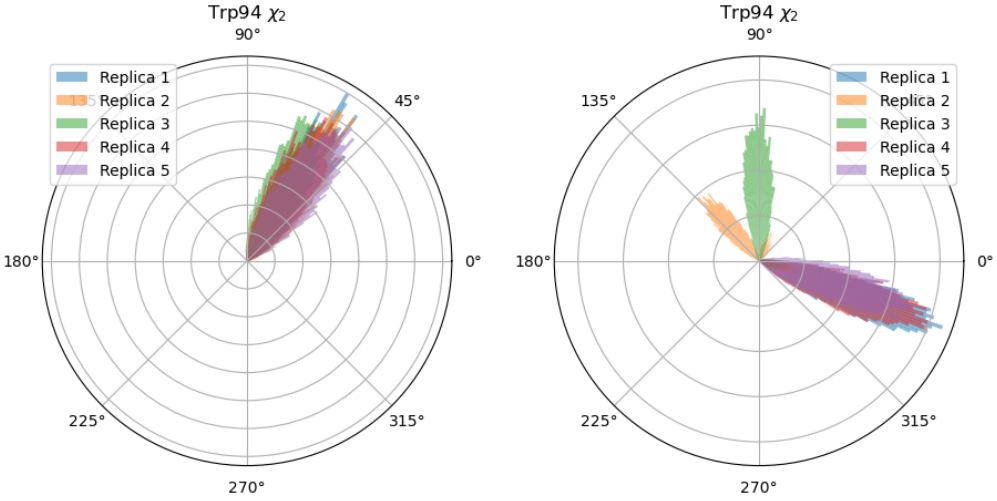
to the crystallographic structure, or moving towards a completely different binding mode. The link between structural rearrangements and the formed protein-ligand interactions was evaluated through an analysis of the average interaction fingerprint (IFP) in each replica. The largest shifts in binding mode were a consequence of the breaking of the salt bridge between the imidazothiazole ring and Glu288. The interaction was replaced either by an unstable ionic bond with Asp187 or Glu32, or not replaced at all (Figure C.1.2).

The simulation of the CXCL12-CXCR4 complex presents different alternative orientations of the chemokine compared to the receptor, with the structure in the first frame of the simulation having a 12.5 Å RMSD compared to the homology model structure (Figure C.1.3B). This rearrangement mostly involved the chemokine's core, while the N-terminus remained mostly stable, with an initial RMSD value of 3.9 Å(Figure C.1.3D). Despite the large rearrangement observed during equilibration, the position of the chemokine core continued to change during the entire simulation, without reaching a stable conformation (Figure C.1.3C). The structure of CRS2 was studied focusing on the interactions formed by the three distal residues of CXCL12 (K1-P2-V3) and the minor binding pocket, which was shared by CXCL12 and the antagonist IT1t. The position of the three residues in the binding pocket remained consistent in three of the five replicas forming the full trajectory (RMSD between 1 and 3 Å), while in the remaining two replicas significant rearrangements were observed (RMSD between 3 and 7 Å). The interactions formed with the minor binding pocket were analysed considering the average IFP of each replica(Figure C.1.4), which identified three distinct binding modes, coherently with the results from RMSD analysis. The original binding mode proposed in the homology model was the most stable, being observed in three of the five replicas, while two alternative binding modes were observed in the remaining replicas. In both cases the shift in chemokine binding was initiated by the breaking of the salt bridge between the N-terminal nitrogen of K1 and Asp97. In one replica the distal N-terminus moved towards the major binding pocket, with the N-terminal nitrogen forming a salt bridge with Glu288, while K1's sidechain formed an unstable interaction with Asp262. The other identified binding mode does not present a large rearrangement in the N-terminus position, but the salt bridge with Asp97 is broken and not reformed during the simulation.

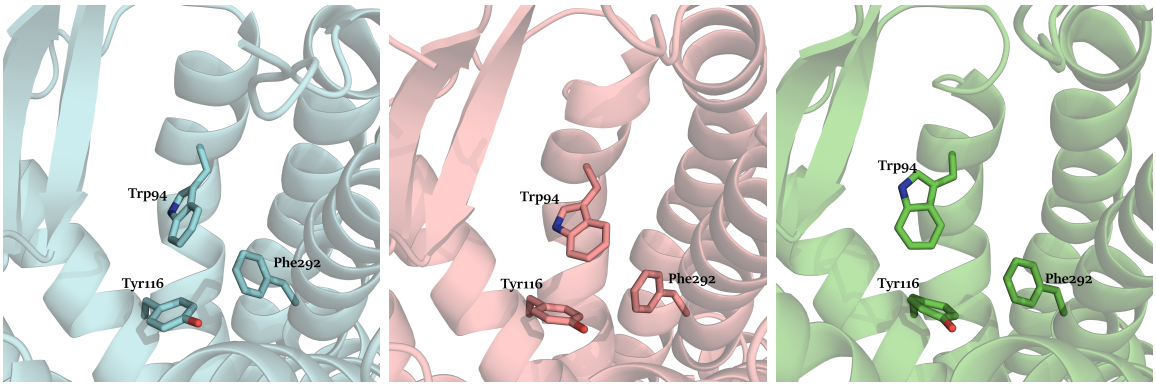
Despite their differences in structure and efficacy, the binding modes of IT1t and CXCL12 share few critical interactions, with the most important being the two salt bridges with Asp97 and Glu288. The main difference between the two binding modes is the stability of these interactions. The salt bridge between IT1t and Asp97 was always conserved for almost the entirety of the simulation, while the interaction with Glu288 was mostly unstable. For CXCL12 instead the interaction with Glu288 was always observed, even though it was formed by different atoms, while the salt bridge with Asp97 was unstable. The two simulated complexes were compared based on the conformation of the residues placed at the bottom of the binding pocket, in particular three of the residues classified by Wescott et al [8] as responsible for signalling initiation (Tyr45, Trp94, and Tyr116), since hydrophobic contacts with such residues were formed by CXCL12 but not IT1t. The conformation of the three residues were compared based on their rotameric state according to the χ_1 angles, and Trp94's χ_2 angle. No significant shift in rotameric state was observed for Tyr45 and Tyr116, while Trp94 assumes a different conformation depending on the bound ligand (Figure 4.1.1B, 4.1.1A). In the IT1t bound structure χ_2 angle values between 45° and 90° were observed (Figure 4.1.1A), while in the three replicas with a stable CXCL12 binding mode the χ_2 angle had a more than 45° difference compared to the IT1t-bound reference. The contact formed by P2 with Trp94 pushed the residue towards the bottom of the binding pocket,

assuming a new stable conformation(Figure 4.1.1E). In the IT1t bound structures instead Trp94 was placed at the bottom of the binding pocket (Figure 4.1.1C). The effect of this rearrangement on the conserved aromatic subpocket formed by Trp94, Tyr116, and Phe292 was measured using the residue residue contact score (RRCS) developed by Zhou et al. to describe inter-helical packing [21] (Figure C.1.8). In the antagonist bound structure the three residues remained packed together, with average contact values around 2. On the other hand, the stable CXCL12 binding mode abolished the contacts of Trp94 and Tyr116 with Phe292, indicating a rearrangement in sidechain packing between the signalling initiation and signalling propagation residues.

Figure 4.1.1: Histogram of Trp94 $\chi 2$ angle distribution during the molecular dynamics simulations of CXCR4 bound to IT1t (A) and to CXCL12 (B). Structure of the residues Trp94, Tyr116, and Phe292 in a frame extracted from the IT1t-CXCR4 simulation (C), in the homology model developed by Ngo et al. (D) and ina frame extracted from the CXCL12-CXCR4 simulation (E).



(A) IT1t-CXCR4 **(B)** CXCL12-CXCR4



(C) IT1t-CXCR4 **(D)** CXCL12-CXCR4 **(E)** CXCL12-CXCR4 MD

Modelling using AlphaFold – Multimer

This section is adapted from the article "Benchmarking structures of chemokine – chemokine receptor complexes generated by AlphaFold" published in the Journal of Chemical Information and Modelling (*J. Chem. Inf. Model.* 64, 11, 4587–4600 (2024)).

The AlphaFold-Multimer structure of the CXCL12-CXCR4 complex resembles a typical chemokine-CKR complex. The N-terminus of the receptor binds the rigid core of the chemokine forming the expected CRS1 and CRS1.5, and CRS3. The N-terminus of the chemokine was placed inside the transmembrane cavity of CXCR4, forming CRS2. The CXCL12 N-terminus was predicted with a high mean pLDDT confidence score of 90.0. Residues 2-5 of the 8-residue long N-terminus had pLDDT scores over 90, suggesting that the prediction for this segment is likely to reproduce the physiological binding mode, as observed in a benchmark exercise of chemokine-chemokine receptor structures (Section 6.2). CXCL12 enters the cavity through the major binding pocket, and it mainly occupies the minor binding pocket, where it extensively forms hydrophobic contacts and electrostatic interactions (Figure 4.1.2). The binding mode exhibits characteristics found in experimental chemokine – CKR complexes: the folding of the N-terminus backbone resembles that of CCL2, and a cyclic residue binding to the bottom of the minor binding pocket has been observed in several complexes[2].

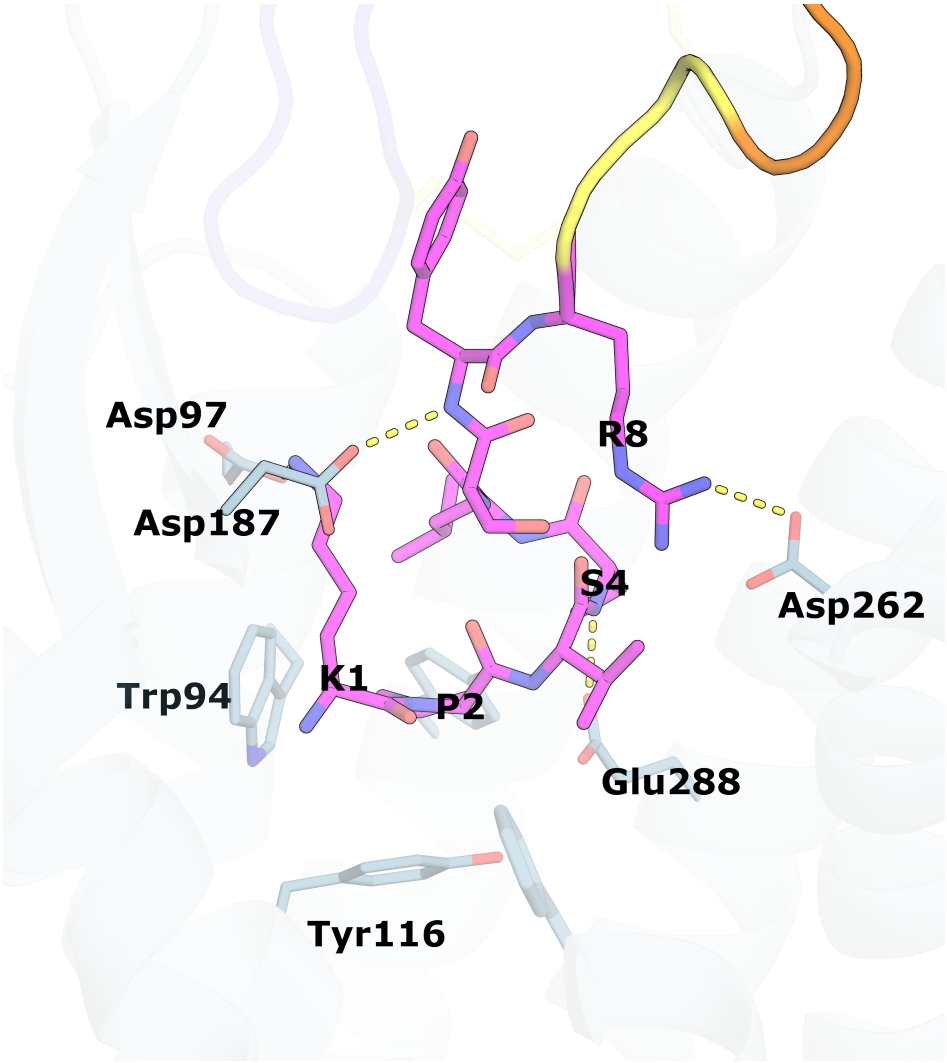
The binding mode predicted by AlphaFold-Multimer fully agrees with the results of mutagenesis experiments performed on both the receptor and chemokine. In the AlphaFold-Multimer structure CXCL12 forms interactions with residues described as being involved in chemokine engagement (Asp262, Asp97, Asp187), signal initiation (Trp94 and Tyr116, and Glu288) or signal propagation (Phe292) based on mutagenesis results on CXCR4 [8]. The sidechain of the chemokine's K1 forms an ionic bond with Asp97, while the positively charged N-terminal nitrogen does not form any specific interaction with the receptor. As discussed for the homology model, the truncation of K1 [17] or its mutation to alanine (K1A)[15], turns CXCL12 into an antagonist, while the K1R[16] and K1Orn[17] mutants maintain their agonist activity. The addition of an extra residue to the CXCL12 N-terminus instead does not influence the chemokine's agonist activity [17–19]. Mutation of P2 to glycine (P2G) abolishes agonist activity[17], probably due to the absence of the observed interaction with the conserved hydrophobic sub-pocket formed by Trp94, Tyr116 and Phe292.

Comparison between the experimental and the in silico structures

An experimental structure of the CXCL12-CXCR4 complex [13] was released after the development of the *in silico* models. The homology model, the MD trajectories, and the AlphaFold-Multimer model were compared to the experimental reference to determine the accuracy of each approach.

The homology model proposed by Ngo et al. presents clear differences to the experimental structure. The receptor structure has a 4.1 Å RMSD for the Cαs, with the largest differences being at the N-terminus (RMSD = 7.1 Å), TM6 (RMSD = 3.5 Å), ICL1 (RMSD = 6.0 Å), and ICL3 (RMSD = 6.0 Å) domains. The differences in TM6 and ICL3 structure were caused by the different activation state of the receptor in the two structures. The homology model presents the receptor in an inactive state conformation, while the experimental structure represents the active state receptor bound to G-protein. The differences in ICL1 conformation instead could be caused by the steric hindrance from Helix 8. The chemokine is placed

Figure 4.1.2: Close-up view of the CXCL12 N-terminus binding mode in the AlphaFold-Multimer model. Hydrogen bonds and ionic contacts are shown as yellow dashes. (Figure adapted from [12])

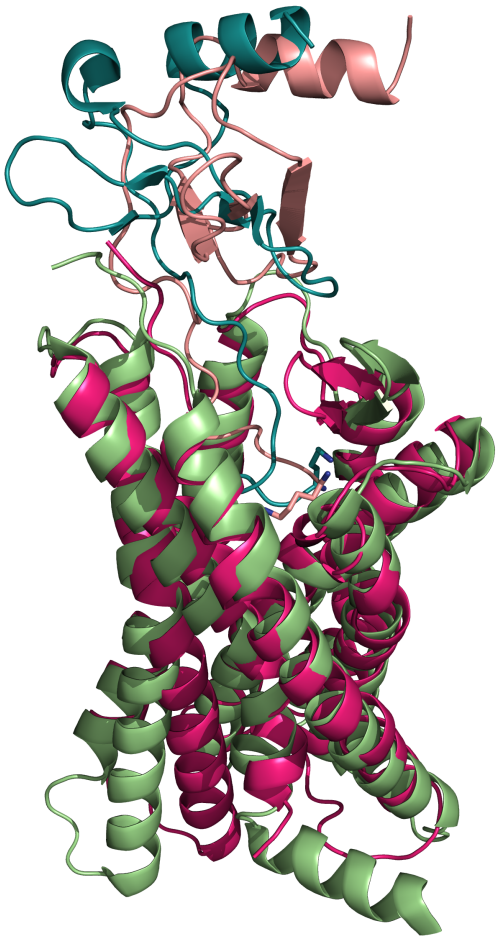


incorrectly in respect to the receptor in the homology model, with an RMSD of 9.2 Å (Figure 4.1.3A). At CRS2 the interface between the chemokine and the binding pocket is also predicted incorrectly, with high RMSD values for the chemokine N-terminus (6.7 Å), and the three distal residues (4.7 Å) (Figure 4.1.3Band Figure 4.1.3B). Considering the individual residues forming the binding site, a per residue RMSD value above 2 Å was observed for Trp94 and Glu288, two key residues for chemokine engagement and signalling initiation.

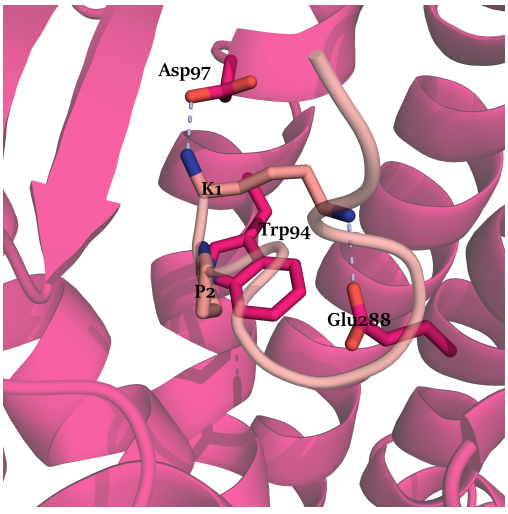
Performing molecular dynamics simulations of the CXCL12-CXCR4 homology model could not shift the receptor's conformation closer to the experimental active state conformation (Cα-RMSD between 3 and 4.6 Å) (Figure C.1.9A). Some improvements were observed regarding the position of the chemokine relative to the receptor, with RMSD values fluctuating between 3 and 12 Å(Figure C.1.9B), after aligning the structures on the receptor's rigid transmembrane domains. The simulation could sample some alternative conformations of the complex which were closer to the experimentally observed chemokine orientation, however they did not represent stable states of the complex. No improvement was instead observed at CRS2, with a Cα-RMSD between 6 and 7.5 Å for the full chemokine's N-terminus (Figure C.1.9C), while the distal three residues had a Cα-RMSD between 5 and 8 Å(Figure C.1.9D). The structural rearrangements observed in the binding site during the simulation did not reflect the real pocket conformation bound by CXCL12. In the experimental structure Trp94 assumes a unique rotameric state, with χ1 being -58°, and the χ2 being 133°, with the aromatic nitrogen pointing towards the bottom of the binding pocket. Some of the conserved interactions observed during the simulations were also observed in the crystallographic structure: the salt bridge with Asp97, the hydrogen bonds with Cys186, Asp187, and Glu288, as well as the hydrophobic contacts with Trp94 and Ile185. While the same interactions were observed, they were not formed by the same pairs of residues: The salt bridge with Asp97 was formed by the sidechain of K1 in the experimental structure, rather than by the terminal nitrogen. The contact with Trp94 was still formed by P2, but rather than pushing the residue towards the bottom of the cavity, the triptophane was pushed towards both the bottom of the cavity and TM2. The hydrogen bond with Glu288 was formed by the neutrally charged sidechain of S4, rather than the positively charged sidechain of K1.

Comparing the experimental structure to the AlphaFold-Multimer complex, the position of the chemokine N-terminus in the binding pocket was predicted correctly with an RMSD of 1.56 Å for the full N-terminus, and 0.73 Å for the first three residues. The relative position of the chemokine compared to the receptor was correctly predicted with a RMSD of 4.05 Å, mostly caused by differences in the conformation of the chemokine's flexible loops. The structure of the receptor was correctly predicted, with few differences in the structures of the flexible loops ICL1 and ICL3. The binding site presented two significantly different sidechain conformations (RMSD $>$ 2 Å) for Tyr116 and Glu288. Almost all the interactions observed in the AlphaFold-Multimer model for the minor binding pocket were also observed in the experimental structure. The key differences were the hydrogen bonds formed with Glu288, which were formed by the backbone nitrogens of S4 and V3 in the AlphaFold model, and by the sidechain of S4 in the experimental structure.

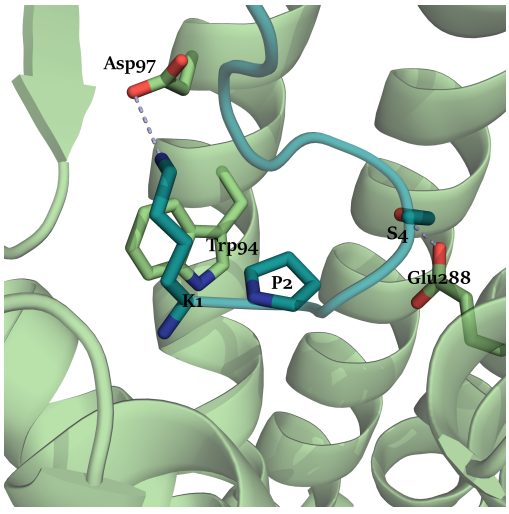
Figure 4.1.3: (A) Homology model of the CXCL12-CXCR4 complex (magenta and pink) aligned on the experimental structure PDB ID: 8U4O (green and blue). Detail of the binding site of CXCR4, and of the binding mode of CXCL12 in the homology model (B), and in the experimental structure (C).



(A)

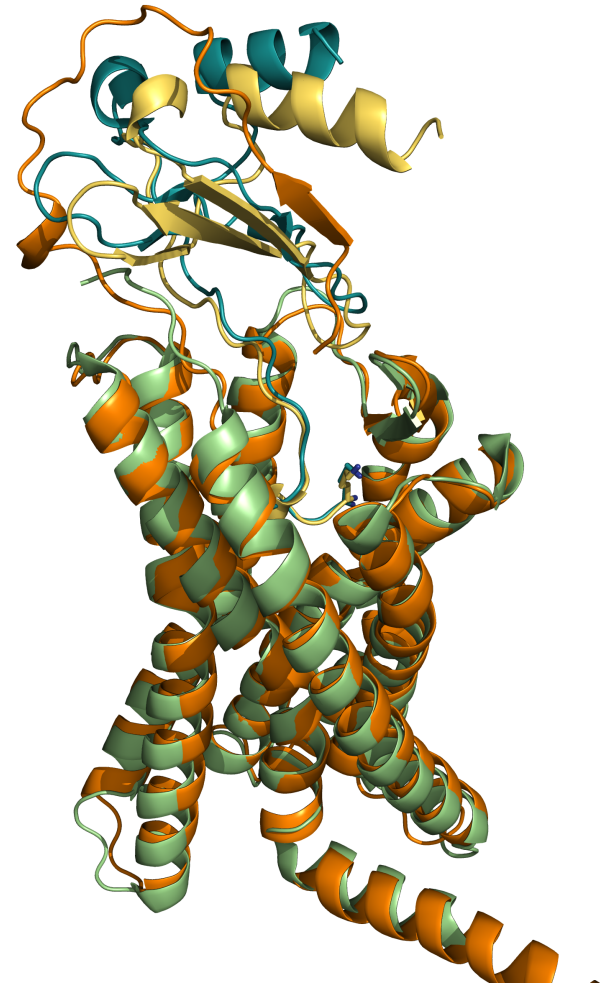


(B)

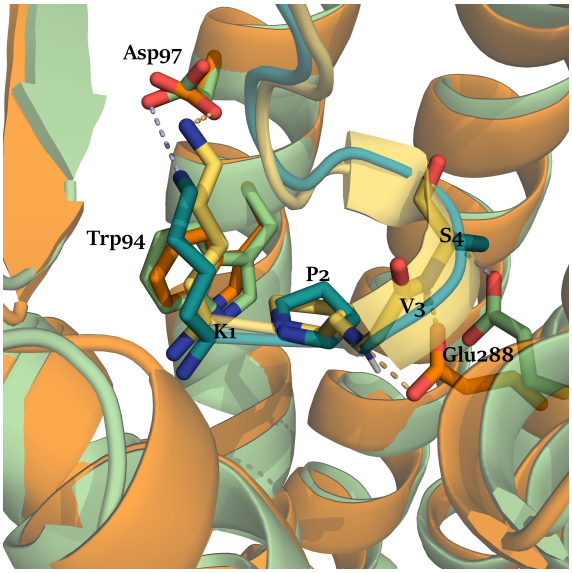


(C)

Figure 4.1.4: (A) AlphaFold-Multimer model of the CXCL12-CXCR4 complex (orange and yellow) aligned on the experimental structure PDB ID: 8U4O (green and blue). (B) Detail of the binding site in the two aligned complexes



(A)



(B)

4.1.3 Discussion

The homology model developed by Ngo et al. could explain most of the mutagenesis data available at the time of its publication, except the lack of changes in CXCL12 efficacy upon introduction of an additional residue in the N-terminal position. Unfortunately, the binding mode of the chemokine's N-terminus did not match the interaction features observed in experimental chemokine-chemokine receptor structures released after its publication. The CXCL12-CXCR4 complex was simulated to capture the small induced-fit rearrangements in the binding site conformation caused by CXCL12 binding. While agonist binding in GPCRs follows a conformation selection model, it was assumed that the inactive state structure proposed in the homology model was similar to the real pocket conformation bound by CXCL12. The aim was to determine the optimized binding pocket conformation describing the interface between the distal N-terminus of CXCL12 and CXCR4 at CRS2. The MD simulations of the CXCL12-CXCR4 homology seemed to validate the structure as plausible, due to the conformational rearrangements observed in simulated CXCL12-bound structure but not in the simulated It1t-CXCR4 complex. These structural rearrangements were compatible with the activation mechanism proposed by Wescott et al.[8], suggesting the validity of both the binding mode and the activation mechanism.

The structure of the CXCL12-CXCR4 complex generated using AlphaFold-Multimer presented an alternative binding mode of the chemokine to the receptor compared to the homology model. The generated binding mode could explain the results from multiple mutagenesis experiments, as well as being compatible with the binding features observed in other chemokine-chemokine receptor structures.

Unfortunately, neither the homology model, nor the results from MD simulation could truly reproduce the experimental structure of the CXCL12-CXCR4 complex. The homology model presents an incorrect placement of the chemokine core compared to the receptor, and an incorrect folding of the chemokine's N-terminus in CRS2. The MD simulation could slightly improve the orientation of the chemokine core, but it was not possible to improve the incorrect binding mode in CRS2. On the other hand, the AlphaFold-Multimer model reproduced the experimental structure with high accuracy, with the largest differences concerning flexible protein domains not directly involved in the protein-protein interface. The position of the distal N-terminus of CXCL12 was reproduced with the highest accuracy, and most of the predicted interactions matched the experimental binding mode. Despite the accuracy, small differences in the receptor's sidechains conformations were observed, which might impact the use of the structural model for drug discovery.

4.1.4 Conclusions

Two different in-silico models of the CXCL12-CXCR4 complex were analysed:a homology model, based on mutagenesis and crosslinking experiment data, and a deep learning generated model based only on co-evolutionary information. The homology model was validated using MD simulations, and it presented a unique conformation of the binding site compared to the receptor bound to the antagonist IT1t. The two models were compared to an experimental structure of the CXCL12-CXCR4 complex. The AlphaFold-Multimer model correctly reproduced the experimentally binding mode of CXCL12, while the homology model could not reproduce neither the orientation of the chemokine in respect to the receptor, nor the interface between CXCL12's N-terminus and CXCR4's extracellular binding pocket.

4.1.5 Materials and methods

Structure preparation

The structure of CXCR4 bound to IT1t (PDB ID: 3ODU)[7] was downloaded from the protein data bank (PDB), while the structure of the CXCL12-CXCR4 complex was obtained from the supplementary materials given by Ngo et al.[6]. From the IT1t-CXCR4 structure point mutations were reverted to match the wild type sequence, and the lysozyme insertion in place of ICL3 was removed. The structure of ICL3 from the CXCL12-CXCR4 structure was grafted on the IT1t-CXCR4 to reconstruct the missing loop in a consistent manner. Water, ions, and lipids were removed from the experimental structure. Protonation was performed using MOE 2018.01.

The complexes were inserted in a lipid bilayer and solvated using CHARMM-GUI. The bilayer had a composition of 1-palmitoyl-2-oleoyl-phosphatidylethanolamine (POPE), 1-palmitoyl-2-oleoyl-phosphatidylcholine (POPC) and cholesterol in a 2:2:1 ratio. The bilayer system was solvated using an aqueous solution containing 0.15 M of KCl, the ions were positioned using the Monte-Carlo method. The initial box length was set to a=b=100 A. The name of the lipids, water and ions was changed to AMBER's naming convention using the charmmlipid2amber.py script (from AmberTools). Residues Tyr7 and Tyr21 were manually mutated to sulfotyrosine in the CXCL12 bound structure using the mutation tool of MOE 2018.01.

Topology and coordinate were created for the ligand IT1t. The structure of IT1t was obtained from its crystallographic structure (PDBID: 3ODU)[7]. Charges and atom types were determined using antechamber[22]. The AM1-BCC model was used to assign the charges considering IT1t as doubly protonated. Atom types were obtained from the GAFF force field[23]. Atomic charges and atom types were manually checked to correct possible errors.

Parameters were determined using parmchk, the results were manually inspected to evaluate the proposed parameters. Topology and coordinate files were obtained using the tleap program and the Amber's force fields: ff14SB for proteins[24], lipid14 for the lipid bilayer[25], TIP3 for water[26], GAFF for the small molecule ligand[23]. The antechamber, parmchk, prepgen and tleap versions used are the ones available in the Amber2016 suite[27][28].

Molecular dynamics simulations

The starting structures were minimized in 15000 steps, 10000 using the steepest descent algorithm and 5000 using the conjugate gradient algorithm. Each system was heated from 0 K to 300 K in 75 ps with different restrains applied to each component of the system. The restrains were progressively removed in 175 ps, using a time-step of 0.001 ps. After heating, the IT1t-CXCR4 system was equilibrated without constrains for 20 ns, while the CXCR4-CXCL12 complex required an equilibration step of 70 ns. For the production trajectory each system was simulated for 5 replicas of 200 ns each, restarting the simulation every 10 ns. The production and simulation steps were performed at constant temperature and pressure conditions, with a 0.002 ps time-step.

Trajectory analysis

The evolution of the system during the simulation was studied using pytraj[29, 30]. The structures were aligned on the Cα of the receptor's transmembrane domains. After alignment the RMSD and RMSF of the receptor on the Cα was computed. RMSD for IT1t and the first three residues of CXCL12 were also computed. The RMSD of the full chemokine during the simulation was computed.

$$RMSD = \sqrt{\frac{\sum_i^N (X_i - Y_i)^2}{N}}$$

$$RMSF_i = \sqrt{\frac{1}{T}\sum_j^T (x_{i(j)} - \langle x_i \rangle)^2}$$

The $\chi 1$ and $\chi 2$ angles of relevant residues for ligand interactions were measured, to determine the contribution of ligand binding to their conformation. The interactions between different receptor residues were measured using the residue-residue contact score developed by Zhou et al.[21]. Protein ligand interactions between CXCR4 and the two ligands were computed using IChem[31], after converting the trajectories into pairs of mol2 files. The conversion was performed using a python script, handling both format and atom type conversion.

Modelling with AlphaFold-Multimer

A structure of the CXCL12-CXCR4 complex was generated using AlphaFold-Multimer v2.3.1[11]. Template structures generated until the 22^{nd} March 2023 were used, as well as the full genetic databases used by AlphaFold default pipeline. Calculations were performed five times with different random seeds using LIT-AlphaFold[12]. The structure with the highest confidence score was selected for analysis.

Comparison to the experimental structure

The experimental structure of the CXCL12-CXCR4 complex was downloaded from the protein data bank (PDB ID: 8U4O)[13]. The experimental structure was used as reference to align the in silico models on the Cα of the TMs, based on the GPCRdb annotation. RMSD between the Cα of the full structure and of the chemokine was measured for the homology model and the AlphaFold model. All atom RMSD between the distal N-terminus of CXCL12, and the full receptor was measured only compared to the AlphaFold model. Alignment and RMSD calculations were performed using MOE 2020 for the single structures, while pytraj was used to compare the structure to the molecular dynamics simulation.

References

[1] Brian K. Kobilka. G protein coupled receptor structure and activation. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1768(4):794–807, April 2007.

[2] Lauri Urvas and Esther Kellenberger. Structural Insights into Molecular Recognition and Receptor Activation in Chemokine–Chemokine Receptor Complexes. *Journal of Medicinal Chemistry*, 66(11):7070–7085, June 2023.

[3] Pius Loetscher, Jiang-Hong Gong, Beatrice Dewald, Marco Baggiolini, and Ian Clark-Lewis. N-terminal Peptides of Stromal Cell-derived Factor-1 with CXC Chemokine Receptor 4 Agonist and Antagonist Activities. *Journal of Biological Chemistry*, 273(35):22279–22283, August 1998.

[4] Ling Qin, Irina Kufareva, Lauren G. Holden, Chong Wang, Yi Zheng, Chunxia Zhao, Gustavo Fenalti, Huixian Wu, Gye Won Han, Vadim Cherezov, Ruben Abagyan, Raymond C. Stevens, and Tracy M. Handel. Crystal structure of the chemokine receptor CXCR4 in complex with a viral chemokine. *Science*, 347(6226):1117–1122, March 2015.

[5] Zhehua Shao, Qingya Shen, Bingpeng Yao, Chunyou Mao, Li-Nan Chen, Huibing Zhang, Dan-Dan Shen, Chao Zhang, Weijie Li, Xufei Du, Fei Li, Honglei Ma, Zhi-Hua Chen, H. Eric Xu, Songmin Ying, Yan Zhang, and Huahao Shen. Identification and mechanism of G protein-biased ligands for chemokine receptor CCR1. *Nature Chemical Biology*, 18(3):264–271, March 2022.

[6] Tony Ngo, Bryan S. Stephens, Martin Gustavsson, Lauren G. Holden, Ruben Abagyan, Tracy M. Handel, and Irina Kufareva. Crosslinking-guided geometry of a complete CXC receptor-chemokine complex and the basis of chemokine subfamily selectivity. *PLOS Biology*, 18(4):e3000656, April 2020.

[7] Beili Wu, Ellen Y. T. Chien, Clifford D. Mol, Gustavo Fenalti, Wei Liu, Vsevolod Katritch, Ruben Abagyan, Alexei Brooun, Peter Wells, F. Christopher Bi, Damon J. Hamel, Peter Kuhn, Tracy M. Handel, Vadim Cherezov, and Raymond C. Stevens. Structures of the CXCR4 Chemokine GPCR with Small-Molecule and Cyclic Peptide Antagonists. *Science*, 330(6007):1066–1071, November 2010.

[8] Melanie P. Wescott, Irina Kufareva, Cheryl Paes, Jason R. Goodman, Yana Thaker, Bridget A. Puffer, Eli Berdougo, Joseph B. Rucker, Tracy M. Handel, and Benjamin J. Doranz. Signal transmission through the CXC chemokine receptor 4 (CXCR4) transmembrane helices. *Proceedings of the National Academy of Sciences*, 113(35):9928–9933, August 2016.

[9] Bryan S. Stephens, Tony Ngo, Irina Kufareva, and Tracy M. Handel. Functional anatomy of the full-length CXCR4-CXCL12 complex systematically dissected by quantitative model-guided mutagenesis. *Science Signaling*, 13(640):eaay5024, July 2020.

[10] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, August 2021.

[11] Richard Evans, Michael O'Neill, Alexander Pritzel, Natasha Antropova, Andrew Senior, Tim Green, Augustin Žídek, Russ Bates, Sam Blackwell, Jason Yim, Olaf Ronneberger, Sebastian Bodenstein, Michal Zielinski, Alex Bridgland, Anna Potapenko, Andrew Cowie, Kathryn Tunyasuvunakool, Rishub Jain, Ellen Clancy, Pushmeet Kohli, John Jumper, and Demis Hassabis. Protein complex prediction with AlphaFold-Multimer, March 2022.

[12] Lauri Urvas, Luca Chiesa, Guillaume Bret, Célien Jacquemard, and Esther Kellenberger. Benchmarking AlphaFold-Generated Structures of Chemokine–Chemokine Receptor Complexes. *Journal of Chemical Information and Modeling*, 64(11):4587–4600, June 2024.

[13] Kei Saotome, Luke L. McGoldrick, Jo-Hao Ho, Trudy F. Ramlall, Sweta Shah, Michael J. Moore, Jee Hae Kim, Raymond Leidich, William C. Olson, and Matthew C. Franklin. Structural insights into CXCR4 modulation and oligomerization, February 2024.

[14] Yezhou Liu, Aijun Liu, Xinyu Li, Qiwen Liao, Weijia Zhang, Lizhe Zhu, and Richard D. Ye. Cryo-EM structure of monomeric CXCL12-bound CXCR4 in active state, December 2023.

[15] Eui Kyung Ryu, Tae Gyun Kim, Taek Hun Kwon, In Duk Jung, Dowook Ryu, Yeong-Min Park, Junhong Kim, Kyo Han Ahn, and Changill Ban. Crystal structure of recombinant human stromal cell-derived factor-1α. *Proteins: Structure, Function, and Bioinformatics*, 67(4):1193–1197, March 2007.

[16] Agnieszka Jaracz-Ros, Guillaume Bernadat, Pasquale Cutolo, Carmen Gallego, Martin Gustavsson, Erika Cecon, Françoise Baleux, Irina Kufareva, Tracy M. Handel, Françoise Bachelerie, and Angélique Levoye. Differential activity and selectivity of N-terminal modified CXCL12 chemokines at the CXCR4 and ACKR3 receptors. *Journal of Leukocyte Biology*, 107(6):1123–1135, June 2020.

[17] M. P. Crump. Solution structure and basis for functional activity of stromal cell-derived factor-1; dissociation of CXCR4 activation from binding and inhibition of HIV-1. *The EMBO Journal*, 16(23):6996–7007, December 1997.

[18] Otto O. Yang, Stephen L. Swanberg, Zhijian Lu, Michelle Dziejman, John McCoy, Andrew D. Luster, Bruce D. Walker, and Steven H. Herrmann. Enhanced Inhibition of Human Immunodeficiency Virus Type 1 by Met-Stromal-Derived Factor 1β Correlates with Down-Modulation of CXCR4. *Journal of Virology*, 73(6):4582–4589, June 1999.

[19] Vincent F.M. Segers, Tomotake Tokunou, Luke J. Higgins, Catherine MacGillivray, Joseph Gannon, and Richard T. Lee. Local Delivery of Protease-Resistant Stromal Cell Derived Factor-1 for Stem Cell Recruitment After Myocardial Infarction. *Circulation*, 116(15):1683–1692, October 2007.

[20] Juan A. Ballesteros and Harel Weinstein. [19] Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. In Stuart C. Sealfon, editor, *Methods in Neurosciences*, volume 25 of *Receptor Molecular Biology*, pages 366–428. Academic Press, January 1995.

[21] Qingtong Zhou, Dehua Yang, Meng Wu, Yu Guo, Wanjing Guo, Li Zhong, Xiaoqing Cai, Antao Dai, Wonjo Jang, Eugene I Shakhnovich, Zhi-Jie Liu, Raymond C Stevens, Nevin A Lambert, M Madan Babu, Ming-Wei Wang, and Suwen Zhao. Common activation mechanism of class A GPCRs. *eLife*, 8:e50279, December 2019.

[22] Junmei Wang, Wei Wang, Peter A. Kollman, and David A. Case. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling*, 25(2):247–260, October 2006.

[23] Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case. Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25(9):1157–1174, 2004.

[24] James A. Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E. Hauser, and Carlos Simmerling. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*, 11(8):3696–3713, August 2015.

[25] Callum J. Dickson, Benjamin D. Madej, Åge A. Skjevik, Robin M. Betz, Knut Teigen, Ian R. Gould, and Ross C. Walker. Lipid14: The Amber Lipid Force Field. *Journal of Chemical Theory and Computation*, 10(2):865–879, February 2014.

[26] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, July 1983.

[27] Andreas W. Götz, Mark J. Williamson, Dong Xu, Duncan Poole, Scott Le Grand, and Ross C. Walker. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *Journal of Chemical Theory and Computation*, 8(5):1542–1555, May 2012.

[28] Romelia Salomon-Ferrer, Andreas W. Götz, Duncan Poole, Scott Le Grand, and Ross C. Walker. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *Journal of Chemical Theory and Computation*, 9(9):3878–3888, September 2013.

[29] Daniel R. Roe and Thomas E. Cheatham. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation*, 9(7):3084–3095, July 2013.

[30] Hai Nguyen, Jason Swails, Daniel R. Roe, and Cody. Amber-Md/Pytraj: V2.0.0. Zenodo, March 2017.

[31] Franck Da Silva, Jeremy Desaphy, and Didier Rognan. IChem: A Versatile Toolkit for Detecting, Comparing, and Predicting Protein–Ligand Interactions. *ChemMedChem*, 13(6):507–510, 2018.

4.2 Virtual screening

Abstract

Structural and binding information extracted from molecular dynamics simulations were used to perform virtual screening to identify a small molecule agonist of the C-X-C chemokine receptor type 4. A library of ∼1 million of molecules was evaluated using docking and a machine learning model trained on the dynamic binding mode of the receptor's endogenous ligand, the chemokine CXCL12. Four compounds with micromolar affinity for CXCR4 were identified. Unfortunately, the hit compounds had all antagonist efficacy. The lack of agonists between the identified hits was explained based on an incorrect hypothesis regarding the structure of the CXCL12-CXCR4 complex.

4.2.1 Introduction

Chemokine receptors are responsible for the regulation of immune response in humans, with the binding of chemokines to the receptor orchestrating immune cell migration. Unlike other members of the chemokine family the C-X-C chemokine receptor type 4 (CXCR4) selectively binds the chemokine CXCL12, its only known endogenous ligand. The regulation of CXCR4 signalling is not mediated only by its interaction with CXCL12, but it is also influenced by the ability of the receptor in forming homo-[1] and hetero-multimers, in particular with the atypical chemokine receptor type 3 (ACKR3)[2]. While the presence of these dimers is known to affect intracellular signalling, there is currently no information about the link between specific multimeric complexes and biological responses[3]. ACKR3 not only regulates CXCR4 signalling by forming multimers on the cell surface, but it can also bind CXCL12. Despite being a member of the chemokine receptor family, ACKR3 does not signal through G-protein coupling [4] but it acts as a scavenger for CXCL12, regulating its concentration in solution.

Due to its key role in immune response regulation, the CXCL12-CXCR4 signalling axis has been linked to autoimmune and inflammatory diseases[5, 6], as well as playing a role in the proliferation of cancer cells[7]. Mutations in the genes encoding CXCR4 are known to cause WHIM syndrome[8], a congenital immune deficiency disease[9]. Furthermore CXCR4 is also a known auxiliary co-receptor for HIV-1 entry in host cells [10]. Since the most relevant pathological aspects of the CXCL12-CXCR4 signalling are linked to an over-expression of CXCL12, or to the interaction between a viral protein and the receptor, most of the therapeutic efforts towards CXCR4 have been focused on the development of antagonist ligands [11], with 8 molecules reaching clinical trials[12]. Two notable examples are AMD3100 [13] and AMD11070[14], which were originally developed to block HIV infection, but reached clinical trials for the treatment of cancer for AMD3100, and of HIV and WHIM syndrome for AMD11070[12]. Despite multiple efforts only AMD3100 has been approved as a drug under the name of Plerixafor [15, 16]. The design of CXCR4 agonists remains an underdeveloped area, even though it holds therapeutic potential to facilitate dibaetic wound healing[17, 18], with two studies reporting the discovery of positive allosteric modulators or allosteric agonists able to improve diabetic wound healing in a diabetic wound mouse model [19, 20]. The literature currently reports the existence of one small molecule agonist of CXCR4, NUCC-390[21], however G protein binding assays performed by our collaborators showed a lack of G-protein coupling in response to NUCC-390 binding, indicating a lack of agonist efficacy towards G-protein (*results unpublished*).

The objective of this work was the design of a small molecule agonist of CXCR4, which could be used as a probe to study receptor activation in vitro, and the influence of receptor dimerization on signalling. The new ligand should able to trigger the same biological response upon binding as the endogenous ligand CXCL12. A structure-based approach was selected due to the lack of other experimentally validated small molecule agonists, and because the flexible nature of the known protein and peptide agonists did not allow to rationalize their binding mode or shared pharmacophoric features. At the time of the project all experimental structures of CXCR4 represented the receptor in an inactive state conformation bound to an antagonist[22, 23]. Previous virtual screening results in the literature reported a higher hit rate for agonists by performing docking on active state structures[24–27], or in silico models built from inactive state structures[28]. A representative structure was extracted from the molecular dynamics simulation of the CXCL12-CXCR4 experiment-guided homology model [29], which was described (at the time) as having an active-like conformation of the binding site. Molecules able to mimic the dynamic binding mode of the three distal residues of CXCL12 were identified using an IG-OCSVM model, using the method described in Chapter 3 [30]. Virtual screening identified four ligands with low micromolar potency for CXCL12 binding inhibition, however they were not able to elicit G-protein coupling upon binding.

4.2.2 Results

Model training

The complex between CXCL12 and CXCR4 presents a large protein-protein interface formed by the N-terminus of the receptor with the rigid core of the chemokine, and by the N-terminus of the chemokine binding in the extracellular pocket of the receptor. The agonist activity of CXCL12 is linked to the binding of the chemokine's N-terminus to CRS2, as the truncation of the N-terminus caused a loss of efficacy (Table 4.2.1), while peptides formed by the N-terminus conserve their efficacy [31, 32],. Targeting the extracellular pocket of CXCR4 should be sufficient to develop an agonist ligand. The full binding interface formed by the N-terminus includes both the major and the minor binding pocket of the receptor, which is too large to be efficiently reproduce by a small molecule ligand. Based on the results from mutagenesis experiments performed on the N-terminus of the chemokine (Table 4.2.1) a structure activity relationship for CXCL12's agonist activity could be proposed. The interactions formed by K1 and P2 seems to determine the efficacy of the ligand, therefore it was assumed that a ligand mimicking the binding mode of K1-P2 should elicit the same biological response as CXCL12. The analysis of the molecular dynamics simulation of the CXCL12-CXCR4 complex also highlighted the presence of a stable hydrogen bond formed by the backbone nitrogen of V3. The interactions formed by the three distal residues of the N-terminus were selected to train an IG-OCSVM model. The stricter NewHyd definition of hydrophobic contact of IChem [33] was selected since the binding cavity is largely solvent exposed, and the NewHyd hydrophobic contacts definition mostly considered the interaction

with the bottom of the cavity.

Table 4.2.1: Mutagenesis data of CXCL12 extracted from the literature.

Mutant		Sequence								Efficacy	Ref
		1	2	3	4	5	6	7	8		
CXCL12		K	P	V	S	L	S	Y	R	Agonist	
CXCL12-Gly	G	•	•	•	•	•	•	•	•	Agonist	[34]
CXCL12-Met	M	•	•	•	•	•	•	•	•	Agonist	[35]
CXCL12-Ser	S	•	•	•	•	•	•	•	•	Agonist	[36]
CXCL12 2-16			•	•	•	•	•	•	•	Antagonist	[32]
CXCL12 3-16				•	•	•	•	•	•	Antagonist	[32]
CXCL12 4-67					•	•	•	•	•	Antagonist	[37]
CXCL12 5-67						•	•	•	•	Antagonist	[37]
CXCL12 K1R		R	•	•	•	•	•	•	•	Partial agonist	[34, 37]
CXCL12 K1A		A	•	•	•	•	•	•	•	Antagonist	[38]
CXCL12 K1Orn		Orn	•	•	•	•	•	•	•	Agonist	[34]
CXCL12 P2G		•	G	•	•	•	•	•	•	Antagonist	[31, 34, 37]
CXCL12 V3I		•	•	I	•	•	•	•	•	Agonist	[34]
CXCL12 Y7H		•	•	•	•	•	•	H	•	Agonist	[34]
CXCL12 Y7A		•	•	•	•	•	•	A	•	Agonist	[34]
CXCL12 R8Q		•	•	•	•	•	•	•	Q	Agonist	[39]
CXCL12 R8K		•	•	•	•	•	•	•	K	Agonist	[34]
CXCL12 R8A		•	•	•	•	•	•	•	A	Agonist	[40]

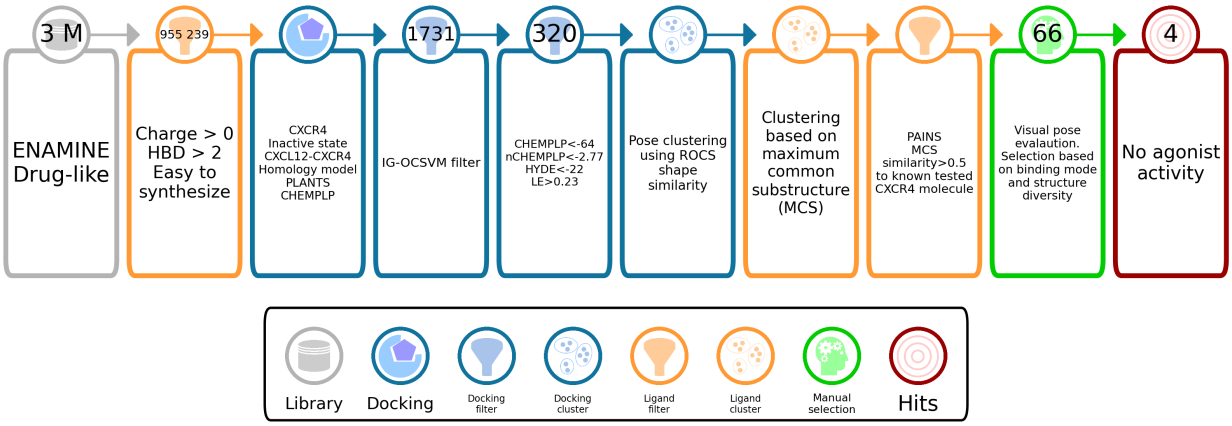
CXCL12 mutants with modifications placed in the first eight residues of the protein. The first sequence (CXCL12) represent the wild type sequence of CXCL12 at residues 1-8. • in the table indicates the lack of mutation at that position in the sequence. Blank spaces indicate a truncation of the corresponding residues. Truncated proteins are annotated based on their true size.

Virtual screening

A library of 955239 molecules was docked on a representative CXCR4 structure extracted from the molecular dynamics simulation of the CXCL12-CXCR4 complex from Section 4.1 (Figure 4.2.1). Filtering using the trained IG-OCSVM model identified 2180 valid docking poses, corresponding to 1731 different molecules. The selected docking poses were rescored using the HYDE scoring function. Poses were filtered based on a threshold value obtained by calculating the docking score of the K1-P2-V3 peptide from CXCL12. In total 347 docking poses satisfied both the docking scores and IG-OCSVM criteria, corresponding to 320 molecules.

The selected poses and molecules were clustered based on structure similarity and docking pose similarity. Clustering based on maximum common substructure identified 178 clusters, with the largest one containing 12 molecules. Clustering of the docking poses using shape similarity identified 140 different poses, with different occupancy of the binding site. The largest cluster was formed by 15 molecules, while 37 poses were observed only for single molecules. The clustered molecules were further filtered to remove PAINS and molecules with a maximum common substructure Tanimoto similarity greater than 0.5 compared to

Figure 4.2.1: Schematic representation of the virtual screening workflow used in this work.



known compounds tested against CXCR4.

78 molecules were manually selected based on a visual analysis of the docking pose, trying to select different scaffolds and binding modes. Of the selected molecules 66 were synthesized by ENAMINE.

Experimental validation

The binding of the virtual hits to CXCR4 was measured using a TR-FRET assay against labelled CXCL12 at 10 μM and 1 μM concentrations. Four ligands able to inhibit the binding of CXCL12 by more than 50% at 10 μM were selected to determine their dose response curve. Additionally, five ligands with relevant structural features and inhibiting close to 50% of CXCL12 binding at 10 μM were also used to determine the dose response curve. Of the tested ten molecules four present an IC50 value below 10 μM (Table 4.2.2). The four ligands with low μM affinity were tested in a BRET functional assay to measure G-protein activation upon ligand binding. The four ligands do not act as either agonist or inverse agonist of the receptor. Since the four molecules bind in a dose-dependent manner, they can be expected to be either antagonists or negative allosteric modulators. A functional assay to measure the ability of the ligands to inhibit G-protein coupling to CXCR4 in the presence of CXCL12 will be performed in the future to confirm the antagonist efficacy of the ligands.

Hits evaluation

The four hit compounds present novel scaffolds compared to other known ligands. While all five ligands form two salt bridges with Asp97 and Glu288, they form different networks of hydrophobic contacts with the minor binding pocket. The shape overlap between the ligands in their corresponding docking poses is small, with the maximum shape Tanimoto score being 0.52, between compound 2 and compound 4. The hit compounds are significantly different from known compounds binding CXCR4 in both their structure, and their physicochemical properties (Table 4.2.3, Figure 4.2.2). The identified compounds have a lower molecular weight and logP, compared to known binders.

Table 4.2.2: Results of the binding assays for the hit compounds

Molecule	% inhibition 10 μM	% inhibition 1 μM	IC50 (μM)
Compound 1	61	19	2.066
Compound 2	58	20	2.140
Compound 3	53	17	2.119
Compound 4	54	11	2.132
Compound 5	52	18	> 10
Compound 6	42	26	> 10
Compound 7	49	13	> 10
Compound 8	49	8	> 10
Compound 9	48	8	> 10

Table 4.2.3: Binding affinity, expressed in IC50, and Tanimoto similarity using Morgan fingerprints (r=2) to the most similar known CXCR4 ligands, for the four hit compounds.

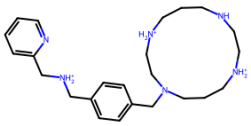
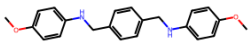
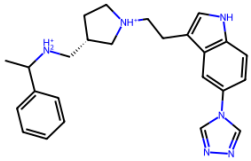
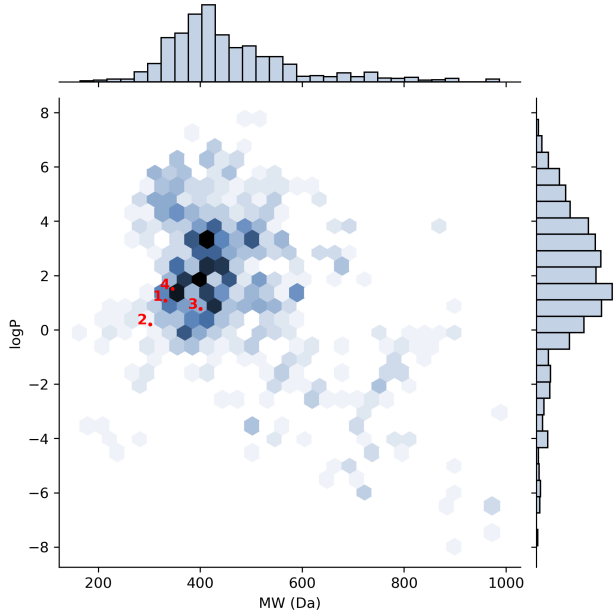
Molecule	IC50 (μM)	Similarity	Closest active molecule
Compound 1	2.066	0.23	
Compound 2	2.140	0.25	
Compound 3	2.119	0.27	
Compound 4	2.132	0.26	

Figure 4.2.2: Comparison between logP and molecular weight (MW) of the four hit compounds, indicated by the red dots, and the known ligands of CXCR4.



The four active molecules were docked in the AlphaFold-Multimer generated CXCL12-CXCR4 complex structure[41], in the experimental structure of the CXCL12-CXCR4 complex[42], and in the inactive state experimental structure bound of the IT1t-CXCR4 complex [22]. Crossdocking in these structures could not reproduce the binding mode selected by the IG-OCSVM within a 2 Å root mean squared deviation (RMSD). The only exception was Compound 1, which could be correctly docked in the AlphaFold-Multimer structure. Based on the crossdocking results it seems that the four hit compounds selectively bind the structure extracted from the molecular dynamics simulation, and not the active state conformation, or the IT1t-bound inactive state conformation of the binding site.

4.2.3 Discussion

The developed virtual screening pipeline, leveraging binding data extracted from a molecular dynamics simulation through a one class classification model, was able to identify four new ligands with micromolar affinity for CXCR4. Unfortunately, no hit molecule had agonist efficacy. The lack of success in identifying a small molecule agonist could be attributed to two main assumptions formulated during the virtual screening process:

- The binding mode of CXCL12 in the homology model represent the native binding mode.
- The trained IG-OCSVM model is selective for both affinity and efficacy.

The incorrect modelling of CXCL12 might explain the lack of agonist efficacy, while the limits of the screening method might explain the low hit rate.

As described in Section 4.1, the homology model used to describe the CXCL12-CXCR4 complex did not match the experimental structure, with significant differences concerning both the binding site structure and the chemokine's binding mode. The structure used for docking represent a unique conformation of the receptor, which was different from both the active-state and the IT1t-bound inactive state structure. The predicted binding modes of the four hit ligands could not be reproduced in other receptor structures. Based on the cross-docking results the used virtual screening pipeline was able to identify four ligands binding a specific conformation of the receptor with a desired binding mode. Unfortunately the bound structure does not seem to represent a relevant conformation of the receptor, and the reference binding mode did not contain any efficacy relevant information other than information already available from crystallographic structures [22] or mutagensis experiments [43].

The screening protocol was designed based on the initial results regarding IG-OCSVM screening power from Section 3.1. While 99.98% of the evaluated molecules was discarded, the selected docking poses were not necessarily of good quality. IG-OCSVM only evaluates the presence of specific interaction patterns in the docking-generated binding mode, but the model is blind to the strength of the detected interactions, the presence of protein-ligand clashes, or ligand torsional strains. The quality of the docking poses was evaluated only based on the docking score from two different scoring function, and from a visual analysis of the docking poses. A more thorough analysis of the docking poses, considering the strain of the torsional angles, the relative energy of the ligand conformation, or the presence of unsatisfied interactions, might have increased the final hit rate by removing low quality predictions[44].

Regardless, the virtual screening campaign had a 6% hit rate, confirming CXCR4 as a difficult target for docking-based virtual screening[45]. A previous docking campaign using the IT1t-bound CXCR4 structure (PDB ID: 3ODU) had a 17% hit rate[46], but with IC50 values from 55 to 70 μM, while a different group reported a high 50% hit rate on the same structure[47]. The screening of a focused library of 52.500 molecules with two cationic centres had a 17% hit rate, identifying fragment-like ligands with affinities in the micromolar range[11]. Other virtual screening campaigns obtained higher hit rates by combining docking with ligand-based drug design. The screening of a library of molecules with high ROCS shape similarity to IT1t (TanimotoShape > 0.70) had a 19% hit rate[48], while by using molecules with a similar threedimensional shape and pharmacophoric features as the known ligand AMD3329 (TanimotoCombo > 0.72) a 57% hit rate was obtained[49]. Compared to these references, the identified ligands either have a higher binding affinity ($\sim 2\ \mu M$) or they present an original structure compared to known ligands. Compounds 2, 3, and 4 have a ROCS similarity based on TanimotoCombo below 0.72 compared to AMD3329, and all of them have a Tanimoto shape similarity below 0.70 compared to IT1t.

4.2.4 Conclusions

Interaction-guided virtual screening on CXCR4, using as reference the dynamic binding mode of the agonist CXCL12 from a homology model, identified four new ligands with an original scaffold and low micromolar affinities for CXCR4. Unfortunately, the discovered ligands do not elicit G-protein coupling upon binding, unlike the reference CXCL12. The lack of agonists between the identified hits, as well as the low hit rate, might have been caused by the virtual screening protocol using data from an incorrect structure of the CXCL12-CXCR4 complex, which present both a different binding mode and a different binding site conformation

compared to the experimental structure.

4.2.5 Materials and methods

Model training

An IG-OCSVM model [30] was trained using the interactions formed by the first three residues of CXCL12 with the receptor in three of the five replicas of the simulation described in the previous section. Interactions were detected with IChem's Ints module [33], using the stricter NewHyd definition for hydrophobic contacts, and represented as triplet of interaction pseudo atoms. The Quick Model Selection 2 (QMS2) [50] heuristic was used for the training of the model.

Structure selection

The molecular dynamics simulations of the CXCL12-CXCR4 complex described in Section 4.1 was analysed to identify a representative conformation for ligand docking. Clustering was performed based on the RMSD value of the binding site. The binding site was defined as all receptor residues within 4.5 Å of the three distal CXCL12 residues in at least 10% of the analysed frames. Clustering was performed using the agglomerative clustering algorithm with average linkage, as implemented in scikit-learn [51]. The structure with the lowest mean RMSD compared to other members of the same cluster was selected as the representative. The structure representing the larger cluster was selected for docking.

Ligand preparation

The diversity druglike subset of ENAMINE REAL formed by around 3 million compounds was selected for screening. The molecules were screened using OpenEye FILTER, removing molecules with structural alerts, non-drug-like properties, or more than two stereocenters. Two additional requirements were included based on the characteristics of the binding pocket: the presence of a net positive charge, and more than two hydrogen bond donor groups. 2.6 million molecules respected the requested properties. The remaining molecules were further filtered based on their synthesizability, according to ENAMINE indications. Only easily synthesizable molecules were selected, for a total of $\sim$1 million compounds.

The three-dimensional structure of the molecules was generated using CORINA. Molecules with more than two stereocenters were removed to facilitate virtual screening. The additional filtering was required since CORINA and FILTER have different definitions of stereocenters. 955239 compounds were selected for virtual screening.

Docking and post processing

The ligands were docked in a representative structure of the receptor extracted from the MD simulation through clustering based on binding site RMSD. Docking was performed using PLANTS at search speed 1, using the CHEMPLP scoring function[52]. Protein ligand interactions were detected using IChem with the NewHyd definition of hydrophobic contacts. The binding mode was evaluated using the trained IG-OCSVM model. Poses with positive scores were selected as representing relevant binding modes, while poses with negative values

were discarded. After filtering 2180 poses were selected, representing 1731 compounds. The selected docking poses were rescored using the HYDE scoring function [53].

The docking score of the three-residue peptide of CXCL12 bound to the reference structure was computed using CHEMPLP and HYDE. The docking scores and normalized docking scores were used to filter the remaining docking poses. (Thresholds: HYDE < -22 kJ/mol, LE < 0.23, CHEMPLP < -64, N-CHEMPLP < 2.77).

Pairwise structural similarity between the selected molecules was computed based on their maximum common substructure. The Tanimoto similarity score was defined as:

$$Ts = \frac{C_{bonds}}{A_{bonds} + B_{bonds} - C_{bonds}} \tag{4.1}$$

With C_{bonds} representing the number of bonds in the maximum common substructure, and A_{bonds} and B_{bonds} the number of bonds in the compared molecules.

Similar molecules were clustered together using the agglomerative clustering algorithm with average linkage. The optimal number of clusters was determined by analysis of the silhouette scores. The scikit-learn implementation of the clustering algorithm and for silhouette scores was used.

The similarity between docking poses was measured by computing the volume overlap between the ligands without structural alignment, since all ligands bind the same conformation of CXCR4 in the same reference coordinates. Shape similarity was computed using OpenEye ROCS, using the Tanimoto shape similarity measure, which is defined as:

$$Tanimoto_{r,q} = \frac{O_{r,q}}{I_r + I_q - O_{r,q}}$$

With r indicating the reference molecule, q the molecule compared to the reference, O the volume overlap between two entities, I the self-volume overlap of an entity.

Docking poses were clustered using the agglomerative cluster algorithm with average linkage, selecting the optimal number of clusters by analysis of silhouette scores.

Similarity to known compounds

A list of compounds experimentally tested against CXCR4 was obtained from ChEMBL[54]. Inactive molecules were removed by the dataset by filtering the entries according to three criteria:

- Standard Relation between the values was '='.
- Standard Value > 0
- Standard Units was 'nM'

Similarity between the hit molecules and the known binders was measured using the Tanimoto similarity between their Morgan fingerprints, with radius 2, calculated with rdkit[55].

The molecular weight and logP of the hit compounds and of the ligand recovered from ChEMBL was calculated with rdkit, using the Crippen method to determine logP [56].

Experimental validation

Ligand binding to CXCR4 was measured based on the ability of the molecules in inhibiting CXCL12 binding with a TR-FRET assay. Halo tagged CXCR4 was expressed in HEK cells and labelled with Terbium cryptate (Donor). This was incubated with CXCL12-d2 (acceptor) with either 10 μM or 1 μM of candidate compound. Compounds that were able to inhibit maximum binding by 50% at 10 μM were used for dose-response competition binding experiment, as well as other ligands with relevant structural properties.

G-protein coupling was measured using a BRET assay. Halo-CXCR4 was expressed in HEK cells with rLuc-tagged $G_{\alpha i1}$, β subunit and venus-tagged γ subunit. Activation of G-protein by the receptor therefore results in the dissociation of the G-protein heterotrimer and a decrease in the BRET signal. BRET signal was followed for 15 mins after treatment with compounds. Experiments were performed in triplicate.

References

[1] Ali Işbilir, Jan Möller, Marta Arimont, Vladimir Bobkov, Cristina Perpiñá-Viciano, Carsten Hoffmann, Asuka Inoue, Raimond Heukers, Chris de Graaf, Martine J. Smit, Paolo Annibale, and Martin J. Lohse. Advanced fluorescence microscopy reveals disruption of dynamic CXCR4 dimerization by subpocket-specific inverse agonists. *Proceedings of the National Academy of Sciences of the United States of America*, 117(46):29144–29154, November 2020.

[2] Angélique Levoye, Karl Balabanian, Françoise Baleux, Françoise Bachelerie, and Bernard Lagane. CXCR7 heterodimerizes with CXCR4 and regulates CXCL12-mediated G protein signaling. *Blood*, 113(24):6085–6093, June 2009.

[3] Philip M. Murphy and Lauren Heusinkveld. Multisystem multitasking by CXCL12 and its receptors CXCR4 and ACKR3. *Cytokine*, 109:2–10, September 2018.

[4] Françoise Bachelerie, Gerard J. Graham, Massimo Locati, Alberto Mantovani, Philip M. Murphy, Robert Nibbs, Antal Rot, Silvano Sozzani, and Marcus Thelen. New nomenclature for atypical chemokine receptors. *Nature Immunology*, 15(3):207–208, March 2014.

[5] Eva M. García-Cuesta, César A. Santiago, Jesús Vallejo-Díaz, Yasmina Juarranz, José Miguel Rodríguez-Frade, and Mario Mellado. The Role of the CXCL12/CXCR4/ACKR3 Axis in Autoimmune Diseases. *Frontiers in Endocrinology*, 10, August 2019.

[6] Stephanie Hummel, Hugo Van Aken, and Alexander Zarbock. Inhibitors of CXC chemokine receptor type 4: Putative therapeutic approaches in inflammatory diseases. *Current Opinion in Hematology*, 21(1):29, January 2014.

[7] Yi Shi, David J. Riese, and Jianzhong Shen. The Role of the CXCL12/CXCR4/CXCR7 Chemokine Axis in Cancer. *Frontiers in Pharmacology*, 11, December 2020.

[8] Paolo A. Hernandez, Robert J. Gorlin, John N. Lukens, Shoichiro Taniuchi, Jože Bohinjec, Fleur Francois, Mary E. Klotman, and George A. Diaz. Mutations in the

chemokine receptor gene CXCR4 are associated with WHIM syndrome, a combined immunodeficiency disease. *Nature Genetics*, 34(1):70–74, May 2003.

[9] Toshinao Kawai and Harry L. Malech. WHIM Syndrome: Congenital Immune Deficiency Disease. *Current opinion in hematology*, 16(1):20–26, January 2009.

[10] Yu Feng, Christopher C. Broder, Paul E. Kennedy, and Edward A. Berger. HIV-1 Entry Cofactor: Functional cDNA Cloning of a Seven-Transmembrane, G Protein-Coupled Receptor. *Science*, 272(5263):872–877, May 1996.

[11] Ilze Adlere, Birgit Caspar, Marta Arimont, Sebastian Dekkers, Kirsten Visser, Jeffrey Stuijt, Chris de Graaf, Michael Stocks, Barrie Kellam, Stephen Briddon, Maikel Wijtmans, Iwan de Esch, Stephen Hill, and Rob Leurs. Modulators of CXCR4 and CXCR7/ACKR3 Function. *Molecular Pharmacology*, 96(6):737–752, December 2019.

[12] Jun Yang, Erkang Tian, Li Chen, Zihang Liu, Yijiu Ren, Wuyu Mao, Yiwen Zhang, and Jifa Zhang. Development and therapeutic perspectives of CXCR4 antagonists for disease therapy. *European Journal of Medicinal Chemistry*, 275:116594, September 2024.

[13] E De Clercq, N Yamamoto, R Pauwels, J Balzarini, M Witvrouw, K De Vreese, Z Debyser, B Rosenwirth, P Peichl, and R Datema. Highly potent and selective inhibition of human immunodeficiency virus by the bicyclam derivative JM3100. *Antimicrobial Agents and Chemotherapy*, 38(4):668–674, April 1994.

[14] Renato T. Skerlj, Gary J. Bridger, Al Kaller, Ernest J. McEachern, Jason B. Crawford, Yuanxi Zhou, Bem Atsma, Jonathon Langille, Susan Nan, Duane Veale, Trevor Wilson, Curtis Harwig, Sigrid Hatse, Katrien Princen, Erik De Clercq, and Dominique Schols. Discovery of Novel Small Molecule Orally Bioavailable C-X-C Chemokine Receptor 4 Antagonists That Are Potent Inhibitors of T-Tropic (X4) HIV-1 Replication. *Journal of Medicinal Chemistry*, 53(8):3376–3388, April 2010.

[15] Jingzhe Wang, Bakhos A. Tannous, Mark C. Poznansky, and Huabiao Chen. CXCR4 antagonist AMD3100 (plerixafor): From an impurity to a therapeutic agent. *Pharmacological Research*, 159:105010, September 2020.

[16] Erik De Clercq. The AMD3100 story: The path to the discovery of a stem cell mobilizer (Mozobil). *Biochemical Pharmacology*, 77(11):1655–1664, June 2009.

[17] Huating Chen, Gongchi Li, Yiqiong Liu, Shuaifei Ji, Yan Li, Jiangbing Xiang, Laixian Zhou, Huanhuan Gao, Wenwen Zhang, Xiaoyan Sun, Xiaobing Fu, and Binghui Li. Pleiotropic Roles of CXCR4 in Wound Repair and Regeneration. *Frontiers in Immunology*, 12, May 2021.

[18] Dustin M. Bermudez, Junwang Xu, Benjamin J. Herdrich, Antoneta Radu, Marc E. Mitchell, and Kenneth W. Liechty. Inhibition of stromal cell-derived factor-1α further impairs diabetic wound healing. *Journal of Vascular Surgery*, 53(3):774–784, March 2011.

[19] Satyamaheshwar Peddibhotla, Karly Caples, Alka Mehta, Qi-Yin Chen, Junyi Hu, Shaquia Idlett-Ali, Liping Zhang, Carlos Zgheib, Junwang Xu, Kenneth W. Liechty, and

Siobhan Malany. Triazolothiadiazine derivative positively modulates CXCR4 signaling and improves diabetic wound healing. *Biochemical Pharmacology*, 216:115764, October 2023.

[20] Junwang Xu, Junyi Hu, Shaquia Idlett-Ali, Liping Zhang, Karly Caples, Satyamaheshwar Peddibhotla, Morgan Reeves, Carlos Zgheib, Siobhan Malany, and Kenneth W. Liechty. Discovery of Small Molecule Activators of Chemokine Receptor CXCR4 That Improve Diabetic Wound Healing. *International Journal of Molecular Sciences*, 23(4):2196, January 2022.

[21] Rama K. Mishra, Andrew K. Shum, Leonidas C. Platanias, Richard J. Miller, and Gary E. Schiltz. Discovery and characterization of novel small-molecule CXCR4 receptor agonists and antagonists. *Scientific Reports*, 6(1):30155, July 2016.

[22] Beili Wu, Ellen Y. T. Chien, Clifford D. Mol, Gustavo Fenalti, Wei Liu, Vsevolod Katritch, Ruben Abagyan, Alexei Brooun, Peter Wells, F. Christopher Bi, Damon J. Hamel, Peter Kuhn, Tracy M. Handel, Vadim Cherezov, and Raymond C. Stevens. Structures of the CXCR4 Chemokine GPCR with Small-Molecule and Cyclic Peptide Antagonists. *Science*, 330(6007):1066–1071, November 2010.

[23] Ling Qin, Irina Kufareva, Lauren G. Holden, Chong Wang, Yi Zheng, Chunxia Zhao, Gustavo Fenalti, Huixian Wu, Gye Won Han, Vadim Cherezov, Ruben Abagyan, Raymond C. Stevens, and Tracy M. Handel. Crystal structure of the chemokine receptor CXCR4 in complex with a viral chemokine. *Science*, 347(6226):1117–1122, March 2015.

[24] Dahlia R. Weiss, SeungKirl Ahn, Maria F. Sassano, Andrew Kleist, Xiao Zhu, Ryan Strachan, Bryan L. Roth, Robert J. Lefkowitz, and Brian K. Shoichet. Conformation Guides Molecular Efficacy in Docking Screens of Activated β-2 Adrenergic G Protein Coupled Receptor. *ACS Chemical Biology*, 8(5):1018–1026, May 2013.

[25] Albert J. Kooistra, Henry F. Vischer, Daniel McNaught-Flores, Rob Leurs, Iwan J. P. de Esch, and Chris de Graaf. Function-specific virtual screening for GPCR ligands using a combined scoring method. *Scientific Reports*, 6(1):28288, September 2016.

[26] Inbar Fish, Anne Stößel, Katrin Eitel, Celine Valant, Sabine Albold, Harald Huebner, Dorothee Möller, Mary J. Clark, Roger K. Sunahara, Arthur Christopoulos, Brian K. Shoichet, and Peter Gmeiner. Structure-Based Design and Discovery of New M $_2$ Receptor Agonists. *Journal of Medicinal Chemistry*, 60(22):9239–9250, November 2017.

[27] Magdalena M. Scharf, Moritz Bünemann, Jillian G. Baker, and Peter Kolb. Comparative Docking to Distinct G Protein–Coupled Receptor Conformations Exclusively Yields Ligands with Agonist Efficacy. *Molecular Pharmacology*, 96(6):851–861, December 2019.

[28] Chris de Graaf, Nicolas Foata, Ola Engkvist, and Didier Rognan. Molecular modeling of the second extracellular loop of G-protein coupled receptors and its implication on structure-based virtual screening. *Proteins: Structure, Function, and Bioinformatics*, 71(2):599–620, 2008.

[29] Tony Ngo, Bryan S. Stephens, Martin Gustavsson, Lauren G. Holden, Ruben Abagyan, Tracy M. Handel, and Irina Kufareva. Crosslinking-guided geometry of a complete CXC

receptor-chemokine complex and the basis of chemokine subfamily selectivity. *PLOS Biology*, 18(4):e3000656, April 2020.

[30] Luca Chiesa and Esther Kellenberger. One class classification for the detection of $B2$ adrenergic receptor agonists using single-ligand dynamic interaction data. *Journal of Cheminformatics*, 14(1):74, October 2022.

[31] Pius Loetscher, Jiang-Hong Gong, Beatrice Dewald, Marco Baggiolini, and Ian Clark-Lewis. N-terminal Peptides of Stromal Cell-derived Factor-1 with CXC Chemokine Receptor 4 Agonist and Antagonist Activities. *Journal of Biological Chemistry*, 273(35):22279–22283, August 1998.

[32] Nikolaus Heveker, Mônica Montes, Lothar Germeroth, Ali Amara, Alain Trautmann, Marc Alizon, and Jens Schneider-Mergener. Dissociation of the signalling and antiviral properties of SDF-1-derived small peptides. *Current Biology*, 8(7):369–376, March 1998.

[33] Franck Da Silva, Jeremy Desaphy, and Didier Rognan. IChem: A Versatile Toolkit for Detecting, Comparing, and Predicting Protein–Ligand Interactions. *ChemMedChem*, 13(6):507–510, 2018.

[34] M. P. Crump. Solution structure and basis for functional activity of stromal cell-derived factor-1; dissociation of CXCR4 activation from binding and inhibition of HIV-1. *The EMBO Journal*, 16(23):6996–7007, December 1997.

[35] Otto O. Yang, Stephen L. Swanberg, Zhijian Lu, Michelle Dziejman, John McCoy, Andrew D. Luster, Bruce D. Walker, and Steven H. Herrmann. Enhanced Inhibition of Human Immunodeficiency Virus Type 1 by Met-Stromal-Derived Factor 1β Correlates with Down-Modulation of CXCR4. *Journal of Virology*, 73(6):4582–4589, June 1999.

[36] Vincent F.M. Segers, Tomotake Tokunou, Luke J. Higgins, Catherine MacGillivray, Joseph Gannon, and Richard T. Lee. Local Delivery of Protease-Resistant Stromal Cell Derived Factor-1 for Stem Cell Recruitment After Myocardial Infarction. *Circulation*, 116(15):1683–1692, October 2007.

[37] Agnieszka Jaracz-Ros, Guillaume Bernadat, Pasquale Cutolo, Carmen Gallego, Martin Gustavsson, Erika Cecon, Françoise Baleux, Irina Kufareva, Tracy M. Handel, Françoise Bachelerie, and Angélique Levoye. Differential activity and selectivity of N-terminal modified CXCL12 chemokines at the CXCR4 and ACKR3 receptors. *Journal of Leukocyte Biology*, 107(6):1123–1135, June 2020.

[38] Eui Kyung Ryu, Tae Gyun Kim, Taek Hun Kwon, In Duk Jung, Dowook Ryu, Yeong-Min Park, Junhong Kim, Kyo Han Ahn, and Changill Ban. Crystal structure of recombinant human stromal cell-derived factor-1α. *Proteins: Structure, Function, and Bioinformatics*, 67(4):1193–1197, March 2007.

[39] James W. Murphy, Yoonsang Cho, Aristidis Sachpatzidis, Chengpeng Fan, Michael E. Hodsdon, and Elias Lolis. Structural and Functional Basis of CXCL12 (Stromal Cell-derived Factor-1α) Binding to Heparin. *Journal of Biological Chemistry*, 282(13):10018–10027, March 2007.

[40] Joshua J. Ziarek, Andrew B. Kleist, Nir London, Barak Raveh, Nicolas Montpas, Julien Bonneterre, Geneviève St-Onge, Crystal J. DiCosmo-Ponticello, Chad A. Koplinski, Ishan Roy, Bryan Stephens, Sylvia Thelen, Christopher T. Veldkamp, Frederick D. Coffman, Marion C. Cohen, Michael B. Dwinell, Marcus Thelen, Francis C. Peterson, Nikolaus Heveker, and Brian F. Volkman. Structural basis for chemokine recognition by a G protein–coupled receptor and implications for receptor activation. *Science Signaling*, 10(471):eaah5756, March 2017.

[41] Lauri Urvas, Luca Chiesa, Guillaume Bret, Célien Jacquemard, and Esther Kellenberger. Benchmarking AlphaFold-Generated Structures of Chemokine–Chemokine Receptor Complexes. *Journal of Chemical Information and Modeling*, 64(11):4587–4600, June 2024.

[42] Kei Saotome, Luke L. McGoldrick, Jo-Hao Ho, Trudy F. Ramlall, Sweta Shah, Michael J. Moore, Jee Hae Kim, Raymond Leidich, William C. Olson, and Matthew C. Franklin. Structural insights into CXCR4 modulation and oligomerization, February 2024.

[43] Melanie P. Wescott, Irina Kufareva, Cheryl Paes, Jason R. Goodman, Yana Thaker, Bridget A. Puffer, Eli Berdougo, Joseph B. Rucker, Tracy M. Handel, and Benjamin J. Doranz. Signal transmission through the CXC chemokine receptor 4 (CXCR4) transmembrane helices. *Proceedings of the National Academy of Sciences*, 113(35):9928–9933, August 2016.

[44] André Fischer, Martin Smieško, Manuel Sellner, and Markus A. Lill. Decision Making in Structure-Based Drug Discovery: Visual Inspection of Docking Results. *Journal of Medicinal Chemistry*, 64(5):2489–2500, March 2021.

[45] Marta Arimont, Shan-Liang Sun, Rob Leurs, Martine Smit, Iwan J. P. de Esch, and Chris de Graaf. Structural Analysis of Chemokine Receptor–Ligand Interactions. *Journal of Medicinal Chemistry*, 60(12):4735–4779, June 2017.

[46] Michael M. Mysinger, Dahlia R. Weiss, Joshua J. Ziarek, Stéphanie Gravel, Allison K. Doak, Joel Karpiak, Nikolaus Heveker, Brian K. Shoichet, and Brian F. Volkman. Structure-based ligand discovery for the protein–protein interface of chemokine receptor CXCR4. *Proceedings of the National Academy of Sciences*, 109(14):5517–5522, April 2012.

[47] Denis Schmidt, Viachaslau Bernat, Regine Brox, Nuska Tschammer, and Peter Kolb. Identifying Modulators of CXC Receptors 3 and 4 with Tailored Selectivity Using Multi-Target Docking. *ACS Chemical Biology*, 10(3):715–724, March 2015.

[48] Debananda Das, Kenji Maeda, Yasuhiro Hayashi, Navnath Gavande, Darshan V. Desai, Simon B. Chang, Arun K. Ghosh, and Hiroaki Mitsuya. Insights into the Mechanism of Inhibition of CXCR4: Identification of Piperidinylethanamine Analogs as Anti-HIV-1 Inhibitors. *Antimicrobial Agents and Chemotherapy*, 59(4):1895–1904, March 2015.

[49] Yoon Hyeun Oum, Steven A. Kell, Younghyoun Yoon, Zhongxing Liang, Pieter Burger, and Hyunsuk Shim. Discovery of novel aminopiperidinyl amide CXCR4 modulators through virtual screening and rational drug design. *European Journal of Medicinal Chemistry*, 201:112479, September 2020.

[50] Zahra Ghafoori, Sarah M. Erfani, Sutharshan Rajasegarar, James C. Bezdek, Shanika Karunasekera, and Christopher Leckie. Efficient Unsupervised Parameter Estimation for One-Class Support Vector Machines. *IEEE Transactions on Neural Networks and Learning Systems*, 29(10):5057–5070, October 2018.

[51] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85):2825–2830, 2011.

[52] Oliver Korb, Thomas Stützle, and Thomas E. Exner. Empirical Scoring Functions for Advanced Protein-Ligand Docking with PLANTS. *Journal of Chemical Information and Modeling*, 49(1):84–96, January 2009.

[53] Nadine Schneider, Gudrun Lange, Sally Hindle, Robert Klein, and Matthias Rarey. A consistent description of HYdrogen bond and DEhydration energies in protein–ligand complexes: Methods behind the HYDE scoring function. *Journal of Computer-Aided Molecular Design*, 27(1):15–29, January 2013.

[54] Anna Gaulton, Anne Hersey, Michał Nowotka, A. Patrícia Bento, Jon Chambers, David Mendez, Prudence Mutowo, Francis Atkinson, Louisa J. Bellis, Elena Cibrián-Uhalte, Mark Davies, Nathan Dedman, Anneli Karlsson, María Paula Magariños, John P. Overington, George Papadatos, Ines Smit, and Andrew R. Leach. The ChEMBL database in 2017. *Nucleic Acids Research*, 45(D1):D945–D954, January 2017.

[55] Greg Landrum, Paolo Tosco, Brian Kelley, Ric, Sriniker, Gedeck, Riccardo Vianello, NadineSchneider, Eisuke Kawashima, Andrew Dalke, Dan N, David Cosgrove, Brian Cole, Matt Swain, Samo Turk, AlexanderSavelyev, Gareth Jones, Alain Vaucher, Maciej Wójcikowski, Ichiru Take, Daniel Probst, Kazuya Ujihara, Vincent F. Scalfani, Guillaume Godin, Axel Pahl, Francois Berenger, JLVarjo, Strets123, JP, and DoliathGavid. Rdkit/rdkit: 2022_03_2 (Q1 2022) Release. Zenodo, April 2022.

[56] Scott A. Wildman and Gordon M. Crippen. Prediction of Physicochemical Parameters by Atomic Contributions. *Journal of Chemical Information and Computer Sciences*, 39(5):868–873, September 1999.

4.3 Critical evaluation and perspectives

The design of an orthosteric small molecule agonist of the C-X-C chemokine receptor type 4 (CXCR4) presented two relevant challenges: no small molecule agonist of CXCR4 was known, and at the time the structure of the complex between CXCR4 and its endogenous ligand, the chemokine CXCL12, was unknown. Due to the absence of comparable ligand references a structure-based drug design approach was selected. As highlighted in Section 2.2 and by the literature [1], it is advisable to use an active state structure of the receptor for docking to identify new agonists. The use of an agonist binding mode to filter the docking poses was also shown to greatly increase the chance of identifying a ligand with the desired efficacy[2, 3]. While no active state structures of the receptor was available at the time, a homology model of the CXCL12-CXCR4 complex had been recently published at the beginning of the project[4]. To validate the structure, and to allow for induce-fit structural rearrangements to optimize the protein-protein interface, molecular dynamics simulations of the complex were performed. Ligand recognition for GPCRs is described as following a conformational selection mechanism, however small induced-fit rearrangements are expected to be observed, optimizing the interface between the receptor and the ligand. The molecular dynamics simulation of the complex showed an instability of both the chemokine core orientation, as well as an instability of the chemokine's binding mode in the extracellular binding pocket. The observed fluctuations were considered as acceptable, probably due to conformational strains in the homology model structure. Binding site rearrangements stabilized by the presence of the bound CXCL12 were determined by comparing the conformation of critical residues between the simulated CXCL12-CXCR4 structure, and a reference simulation of the IT1t-CXCR4 complex. The binding information extracted from the simulation was used to train an InteractionOCSVM model, specifically and IG-OCSVM model, to identify other ligands able to form the same interactions as the three distal residues of the chemokine with the minor binding pocket. The InteractionOCSVM method was developed specifically to extract information from molecular dynamics simulation of receptor-ligand complexes obtained through homology models or other in silico approaches, with this virtual screening campaign representing its first prospective application. The IG-OCSVM model filtered out most of the compounds in the screened chemical library, probably due to the presence of a conserved double salt bridge in the binding mode. Unfortunately of the tested 66 virtual hits, only four had a binding affinity for CXCR4 below 10 μM. While the hit rate of the screening campaign was low, 6%, multiple compounds were selected for testing just because they represented an interesting scaffold or binding mode, rather than due to the quality of the docking pose, which might have decreased the final hit rate. Despite the low hit rate, the four identified compounds represent novel scaffolds for CXCR4 ligands. Of the four hits none had agonist efficacy. Initially it was not clear whether this was due to an error in the initial homology model, a lack of selectivity of the InteractionOCSVM model, or due to the fact that CXCR4 agonists would need to also occupy the major binding pocket to stabilize the active state, as observed for CXCL12 in the model structure.

While doubts on the quality of the CXCL12-CXCR4 model were already raised by the release of new experimental structures of chemokine - chemokine receptor complexes, and further confirmed by the analysis of an alternative CXCL12-CXCR4 structure generated using AlphaFold-Multimer, the release of the first experimental structure of the complex confirmed the inadequacy of the homology model. Both the binding pocket structure and the binding mode described by the homology model did not reflect the native conformation of

the CXCL12-CXCR4 structure. Chapter 2 described the success of docking and interaction similarity in efficacy-driven drug discovery as depending on two factors: the reference binding mode containing relevant pharmacophoric features, and the recognition of relevant ligands by the binding pocket conformation used for docking. Since both the pharmacophore described by the binding mode, and the binding pocket conformation used for docking did not match the data from the experimental structure, the discovery of an orthosteric agonist would have been likely caused by chance, rather than by the predictive power of the virtual screening pipeline.

While the receptor structure used for docking is in a different conformation than the CXCL12 bound structure, it also represents an alternative conformation compared to the structure bound to the small molecule antagonist IT1t. The considered structure might represent an unstable or metastable pocket conformation, corresponding to a poorly populated state, which could explain both the low hit rate, and the relatively low affinity of the hits. Since the four identified compounds seem to selectively bind this unique conformation of the binding site, it would be interesting to further study the stability of their predicted binding modes using molecular dynamics simulations. If the compounds were to be confirmed to bind and stabilize a unique conformation of the receptor, further experimental studies regarding their effect on CXCR4's signalling, or on receptor dimerization, could determine whether the evaluated conformation plays any relevant biological role[5].

While this initial attempt at designing a small molecule orthosteric agonist of CXCR4 failed, future attempts could use the data from the experimental structure for docking and poses post-processing. The structure predicted using AlphaFold-Multimer almost perfectly matched the experimental structure, excluding few small differences regarding the conformation of certain key residues sidechains, and at the protein-protein interface in the extracellular binding pocket. Therefore given its similarity to the experimental reference the AlphaFold-Multimer structure should also be used for docking calculations, in order to evaluates its use in a prospective virtual screening campaign. By using both the experimental and the AlphaFold-Multimer structure in the same pipeline developed for the homology model, and by screening the same chemical library, it would be possible to determine the effect of the binding site conformation and of the reference binding mode on the final virtual hits list.

References

[1] Dahlia R. Weiss, SeungKirl Ahn, Maria F. Sassano, Andrew Kleist, Xiao Zhu, Ryan Strachan, Bryan L. Roth, Robert J. Lefkowitz, and Brian K. Shoichet. Conformation Guides Molecular Efficacy in Docking Screens of Activated β-2 Adrenergic G Protein Coupled Receptor. *ACS Chemical Biology*, 8(5):1018–1026, May 2013.

[2] Albert J. Kooistra, Rob Leurs, Iwan J. P. de Esch, and Chris de Graaf. Structure-Based Prediction of G-Protein-Coupled Receptor Ligand Function: A β-Adrenoceptor Case Study. *Journal of Chemical Information and Modeling*, 55(5):1045–1061, May 2015.

[3] Albert J. Kooistra, Henry F. Vischer, Daniel McNaught-Flores, Rob Leurs, Iwan J. P. de Esch, and Chris de Graaf. Function-specific virtual screening for GPCR ligands using a combined scoring method. *Scientific Reports*, 6(1):28288, September 2016.

[4] Tony Ngo, Bryan S. Stephens, Martin Gustavsson, Lauren G. Holden, Ruben Abagyan,

Tracy M. Handel, and Irina Kufareva. Crosslinking-guided geometry of a complete CXC receptor-chemokine complex and the basis of chemokine subfamily selectivity. *PLOS Biology*, 18(4):e3000656, April 2020.

[5] Ali Işbilir, Jan Möller, Marta Arimont, Vladimir Bobkov, Cristina Perpiñá-Viciano, Carsten Hoffmann, Asuka Inoue, Raimond Heukers, Chris de Graaf, Martine J. Smit, Paolo Annibale, and Martin J. Lohse. Advanced fluorescence microscopy reveals disruption of dynamic CXCR4 dimerization by subpocket-specific inverse agonists. *Proceedings of the National Academy of Sciences of the United States of America*, 117(46):29144–29154, November 2020.

Chapter 5

Virtual screening and pharmacological properties: the β2-adrenergic receptor case

Chapter 2 and Chapter 3 tackled the problem of efficacy-driven drug discovery, however other aspects of a ligand's pharmacology might be relevant during the drug development process. The duration of action of β2-adrenergic receptor agonists defines their therapeutic application as drugs. In this chapter both structure-based and ligand-based approaches were proposed to guide virtual screening towards the identification of β2-agonists with a desired duration of action.

In the first section ligand-based methods were used to identify the structural and physicochemical properties regulating the mechanism behind the duration of action of β2-agonists. While ligand-based approaches could classify known ligands based on their duration of action, they were unable to identify ligands with agonist efficacy for the β2-adrenergic receptor. The InteractionOCSVM method developed in Chapter 3 was tested to determine the presence of a bias regarding the duration of action of the ligands selected by the models.

In the second section the developed prediction pipeline was used to predict the duration of action of the known agonist BI167107. The predictions were confirmed by measuring the duration of action of BI167107 in an in vivo assay on a mice model.

The work of this chapter is the outcome of an interdisciplinary collaboration with Dr Emilie Sick, from the Laboratoire de Chémo Biologie Synthétique Thérapeutique - Université de Strasbourg, which aided in the pharmacological characterization of the known agonists based on literature data, and Dr François Daubeauf from the Plateforme de Chimie Biologique Intégrative de Strasbourg, who developed the mice model and performed the experiment to measure the duration of action of β2-agonists.

5.1 Predicting the duration of action of β2-adrenergic receptor agonists: ligand and structure-based approaches

This section was previously published in Molecular Informatics(*Mol Inform* 42, e202300141 (2023)). All supporting informations are available on-line at: `https://doi.org/10.1002/minf.202300141`

Received: 11 June 2023 | Revised: 17 October 2023 | Accepted: 23 October 2023
DOI: 10.1002/minf.202300141

RESEARCH ARTICLE

molecular informatics

Predicting the duration of action of β2-adrenergic receptor agonists: Ligand and structure-based approaches

Luca Chiesa[1] | **Emilie Sick**[2] | **Esther Kellenberger**[1]

Abstract

Agonists of the β2 adrenergic receptor (ADRB2) are an important class of medications used for the treatment of respiratory diseases. They can be classified as short acting (SABA) or long acting (LABA), with each class playing a different role in patient management. In this work we explored both ligand-based and structure-based high-throughput approaches to classify β2-agonists based on their duration of action.

A completely in-silico prediction pipeline using an AlphaFold generated structure was used for structure-based modelling. Our analysis identified the ligands' 3D structure and lipophilicity as the most relevant features for the prediction of the duration of action. Interaction-based methods were also able to select ligands with the desired duration of action, incorporating the bias directly in the structure-based drug discovery pipeline without the need for further processing.

5.1.1 Introduction

The β2 adrenergic receptor (ADRB2) is one of the most studied targets in the G protein coupled receptors family (GPCRs), with multiple resolved crystallographic structures and many known ligands with different pharmacological profiles. GPCRs play a key biological role mediating the interaction between the cell and the environment via signalling transduction pathways. The binding of a molecule in the extracellular binding pocket regulates the signalling by enhancing it (agonist), or inhibiting it (antagonists, and inverse agonist). ADRB2 agonists are an important class of drugs for the treatment of respiratory diseases such as asthma and chronic obstructive pulmonary disease (COPD) [1, 2]. Based on their duration of action they have been classified as short acting (SABA), long acting (LABA), and ultra-long acting (Ultra-LABA) β-agonists [3]. Each class has a specific therapeutic role: SABAs are used as quick relief medications, while LABAs are used as long-term control medications[4]. In the past twenty years pharmaceutical companies have developed new

LABAs and Ultra-LABAs[3, 5], aiming for a once-daily dosing agonist, whose duration of action is optimal for patient management [6].

The exact mechanism regulating the duration of action of β2-agonists is still unclear. The most supported hypothesis is the diffusion microkinetic theory[7, 8], which links the duration of action to the partition of the ligand between the aqueous phase and the lipid membrane. More lipophilic compounds accumulate in the membrane, forming a ligand's reserve, increasing the probability of rebinding and consequently the duration of their effect. Other mechanisms could explain the difference in duration of action: The binding kinetics of the ligands was proposed as a determining factor, however experiments examining this aspect gave conflicting results [9, 10]. The existence of an additional binding site, called exosite, was proposed to explain the long duration of action of salmeterol and its derivatives [11]. A secondary binding pocket accommodating the ligand long aliphatic chain was identified by crystallography in the region between the extracellular ends of TM6 and TM7, and ECL2 [12]. Mutations in this region however only affected the binding affinity of salmeterol but not its duration of action[13].

In this work we propose to explore different computer aided drug design (CADD) methods, compatible with high-throughput virtual screening, to qualitatively predict the duration of action of 22 well characterized β2-agonists: 10 LABAs (including both LABAs and Ultra-LABAs) and 12 SABAs, (Figure 5.1.1, Table S1-2). The evaluated approaches can be divided in two categories: structure-based, and ligand-based. Ligand based drug design (LBDD) uses information from known binders to select new potential active molecules. In this work we explore different molecular representations of increasing complexity: calculated physicochemical properties, molecular and pharmacophoric fingerprints, and 3D ligand structures. The aim is to identify the properties and structural features regulating the duration of action. Structure based drug design (SBDD) uses the 3D structure of the target protein to guide drug discovery. The most used technique is protein-ligand docking, which predicts the pose of a molecule in the target's binding site and evaluates its binding affinity using a scoring function. Pharmacophore models, based on the interactions observed in crystallographic structures[14] or molecular dynamic simulations[15], were often used to post-process docking results, especially to identify ligands with a desired pharmacological profile. SBDD strategies were applied using both crystallographic structures of the receptor-agonist complex[16, 17], and structures generated by an *in silico* pipeline combining AlphFold2 (AF)[18] and molecular docking. The development of deep learning methods for protein structure prediction carries the promise of access to SBDD for any target, although the use of such structures in virtual screening is still limited by the quality of the predictions[19–22]. Protein-ligand docking, and molecular dynamics were used to model the complex between ADRB2 and the agonists formoterol and salbutamol. The modelled and crystallographic binding modes were used for pose rescoring in a retrospective docking exercise using the dynamic representation obtained combining interaction graphs (IGs) and one class support vector machine (IG-OCSVM)[15].

The objective is to determine the key molecular properties regulating the duration of action of known β2-agonists and evaluate different strategies for its prediction. This analysis was performed considering different descriptors, different machine learning methods, as well as various degrees of information (one reference, two references with different duration, a curated dataset). We also explored the use of interaction-based methods in SBDD for such task, since these methods were already successfully applied for the detection of ligands with a specific pharmacological profile, even in the presence of limited experimental

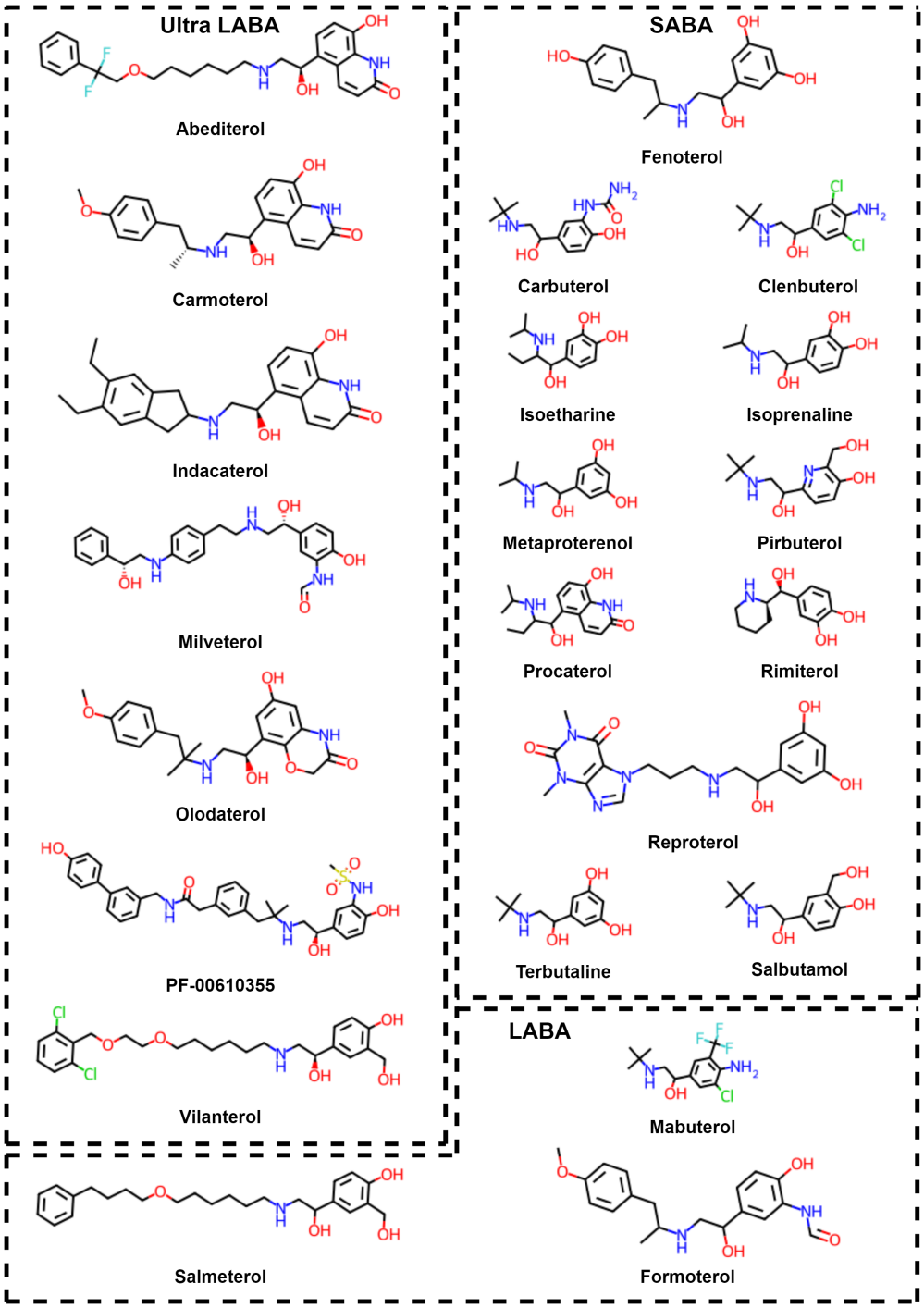
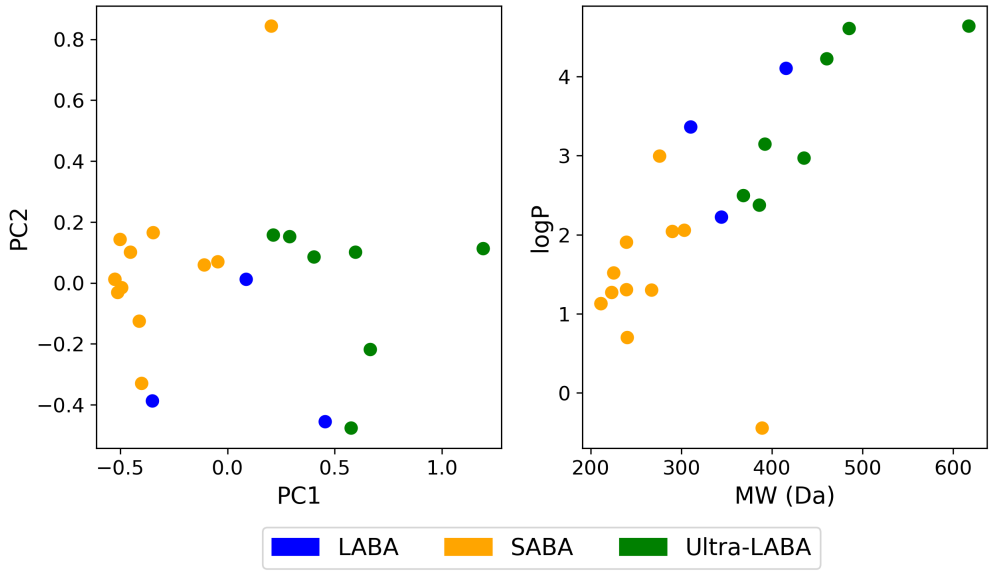


Figure 5.1.1: Molecular structures of the β2-agonists dataset containing 12 SABAs and 10 LABAs+Ultra-LABAs

Figure 5.1.2: Plot of the physicochemical properties of the molecules in the β2-agonists dataset according to the first two principal components, and molecular weight vs. logP plot.



data[15]. The presence of a bias towards ligands with a specific duration of action in SBDD interaction-based methods could be linked to the used representation indirectly encoding some ligand properties. Comparing the results of the two approaches might explain the origin of such bias.

5.1.2 Results

Physicochemical properties

Eight common descriptors were calculated from the ligands' 2D structures (Table S3): logP, molecular weight, topological polar surface area (TPSA), number of hydrogen bond acceptors (HBA), number of hydrogen bond donors (HBD), number of rotatable bonds, number of rings, number of aromatic rings. Principal component analysis (PCA) was performed on the calculated descriptors. The first two principal components represent respectively 66% and 20% of the variance. The first principal component (PC1) has positive contributions from all the descriptors. Excluding HBD, all descriptors are positively correlated ($R > 0.5$) to the molecular weight, therefore PC1 was identified as describing the "size" of the ligand. The second principal component (PC2) presents negative contributions from logP and the number of rotatable bonds, and positive contributions from the TPSA, HBA, the number of rings, and the number of aromatic rings. The number of rotatable bonds is positively correlated ($R = 0.71$) with the logP, while the number of rings ($R = 0.70$), and the number of aromatic rings ($R = 0.78$) are positively correlated with the TPSA. PC2 was identified as describing the polarity of a ligand, with low values associated to more lipophilic molecules, and higher values to more hydrophilic molecules. Both the PCA, and the molecular weight vs. logP plots separate LABAs and SABAs (Figure 5.1.2), with LABAs and Ultra-LABAs being heavier and more lipophilic than SABAs.

A classification tree (CT) trained on the calculated descriptors, evaluated in leave one

out cross validation (LOO-CV), separated the two classes with a mean accuracy of 0.86. An analysis of feature importance using SHAP values [23, 24] identified the molecular weight and logP as the most important properties linked to duration of action. Heavier and more lipophilic molecules were more likely to be long acting or ultra-long acting. Despite the good performances, these results should be taken with caution, excluding outliers like the lipophilic SABA clenbuterol (logP=3.00) or the heavy SABA reproterol (MW=389 Da), the difference in logP between LABAs (formoterol = 2.22, carmoterol = 2.50), and SABAs (fenoterol = 2.06, procaterol = 2.04) and the differences in molecular weight (LABA mabuterol = 310 Da, SABA fenoterol = 303 Da) are small. While the calculated physicochemical properties play a role in determining a ligand's duration of action, such results seem to be limited just to a qualitative analysis. LABAs and Ultra-LABAs do not present a significant difference in logP or molecular weight, indicating the contribution of other factors to the duration of action.

Molecular similarity

Molecular similarity is one of the most used concepts in cheminformatics, including different representations and similarity measures[25]. In this work we explored the use of 2D molecular fingerprints, 2D pharmacophoric fingerprints, and 3D similarity. Morgan fingerprints were used to evaluate molecular similarity. Pharmacophore fingerprints were generated using RdKit [26] with the default pharmacophore definition (Pharma), and the one proposed by Gobbi and Poppinger [27] (Gobbi). 3D similarity was computed using the rapid overlay of chemical shapes (ROCS) method. ROCS measures molecular similarity based on the shape of the molecule, its pharmacophoric features (color), and a combination of the two.

Three different training sets were evaluated for model development: a single reference ligand (the LABA formoterol, or the SABA salbutamol), one reference ligand for each class (formoterol and salbutamol), and the entire β2-agonists dataset. k-nearest neighbours (kNN) classifiers were trained on datasets containing multiple references, while single references were used to rank the ligands by molecular similarity.

Ranking alone was not always sufficient to prioritize one class over the other (Table 5.1.1). 2D pharmacophoric fingerprints could prioritize one class over the other (ROCAUC $\geq$ 0.71), while Morgan fingerprints did not (ROCAUC $\leq$ 0.69). 3D similarity performed as well or better than 2D pharmacophoric fingerprints, depending on the used similarity measure: Salbutamol prioritized SABAs based on Tanimoto similarity (ROCAUC $\geq$ 0.79) while Tversky similarity did not give consistent results. Formoterol prioritized LABAs with Tversky similarity (ROCAUC $\geq$ 0.76) but not with Tanimoto similarity (ROCAUC $\leq$ 0.64).

kNN classifiers generally discriminated between LABAs and SABAs, regardless of the selected similarity measures. The size of the dataset influenced the quality of the results, with models trained using the β2-agonists dataset performing better (accuracy between 0.82 and 0.91) than models based only on formoterol and salbutamol (accuracy between 0.62 and 0.85) (Figure 5.1.3, Table S4). Molecular fingerprints performed well only in the presence of a sufficiently large training dataset. Descriptions based on pharmacophoric features, both 2D and 3D, were generally predictive even with a limited number of references, with 3D molecular similarity being the best approach with only two reference molecules.

Random forest (RF) and CT models were trained using binary and count Morgan fingerprints, and 2D pharmacophoric fingerprints to identify key structural and pharmacophoric features regulating the duration of action. CT and RF models trained

Table 5.1.1: ROCAUC values based on molecular similarity rankings with formoterol and salbutamol as references. The positive class correspond to the same class as the reference, LABAs for formoterol, and SABAs for salbutamol.

Similarity measure	Formoterol (LABA)	Salbutamol (SABA)
TanimotoMorgan FP	0.55	0.69
TanimotoPharma FP	0.85	0.82
TanimotoGobbi FP	0.71	0.77
Tanimoto ROCS color	0.57	0.79
Tversky ROCS color	0.76	0.79
Tanimoto ROCS shape	0.61	0.88
Tversky ROCS shape	0.90	0.45
Tanimoto combo ROCS color	0.64	0.83
Tversky combo ROCS shape	0.85	0.66

on count fingerprint were more accurate (accuracy 0.83 and 0.86) than the ones trained on binary fingerprints (accuracy 0.71 and 0.77). Pharmacophore-based models performed better than the ones based on molecular fingerprints (Pharma: CT accuracy = 0.80 RF accuracy = 0.86, Gobbi: CT accuracy = 0.90 RF accuracy = 0.95) (Figure 5.1.3, Table S5).

Since molecular similarity was sufficient to separate LABAs and SABAs we analysed the dataset by performing clustering using the different measures to obtain a better insight on the results. Structural differences between the two classes could be observed, with members of the same cluster often sharing the same label (Figure 5.1.4). Few molecules were consistently clustered with members of the opposite class (the SABA fenoterol, and the LABA mabuterol) or as singlets (PF-00610355 and reproterol). All similarity measures could separate the two classes, with the ROCS combo score and Pharma fingerprints giving the best results. The former yielded one singlet (PF-00610355) and three well defined clusters: one containing most of the SABAs, one containing most of the LABAs, and one containing three LABAs (abediterol, vilanterol and salmeterol) characterized by a long aliphatic chain in their structures. The latter identified two clusters corresponding roughly to the two classes, except for three molecules (mabuterol, fenoterol, and reproterol). Clustering grouped together compounds with specific structural features: Morgan fingerprints identified specific structural features in the dataset like molecules with an isopropane-like (fenoterol, isoetharine, isoprenaline, and metaproterenol) or tertbutyl (clenbuterol, mabuterol, pirbuterol, salbutamol, and terbutaline) group bound to the protonated amine, pairs of molecules with one being the substructure of the other (formoterol and carmoterol, and salmeterol and vilanterol), or molecules sharing the same "head" group (albediterol, indacaterol and procaterol). ROCS color similarity grouped together molecules with similar "head" groups like catechol or catechol-like (clenbuterol, mabuterol, carbuterol, isoetharine, pirbuterol, rimiterol, salbutamol), resorcinol (fenoterol, metaproterenol, terbutaline), or 2,8-quinolinediol or 2,8-quinolinediol-like (abediterol, carmoterol, indacaterol, formoterol, procaterol).

Molecular similarity and molecular features were sufficient to distinguish between the two classes in the unsupervised contexts of clustering and ranking, and for training supervised

Figure 5.1.3: Accuracies of classification models trained on molecular similarity, Morgan fingerprints, and 2D pharmacophoric fingerprints. The colours in the legend indicate the training dataset and the used machine learning method, the descriptors are indicated as labels and by the hatching on the bars.

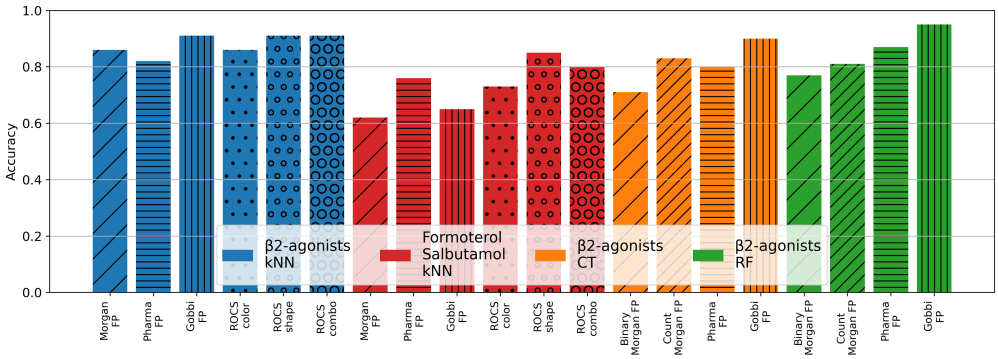
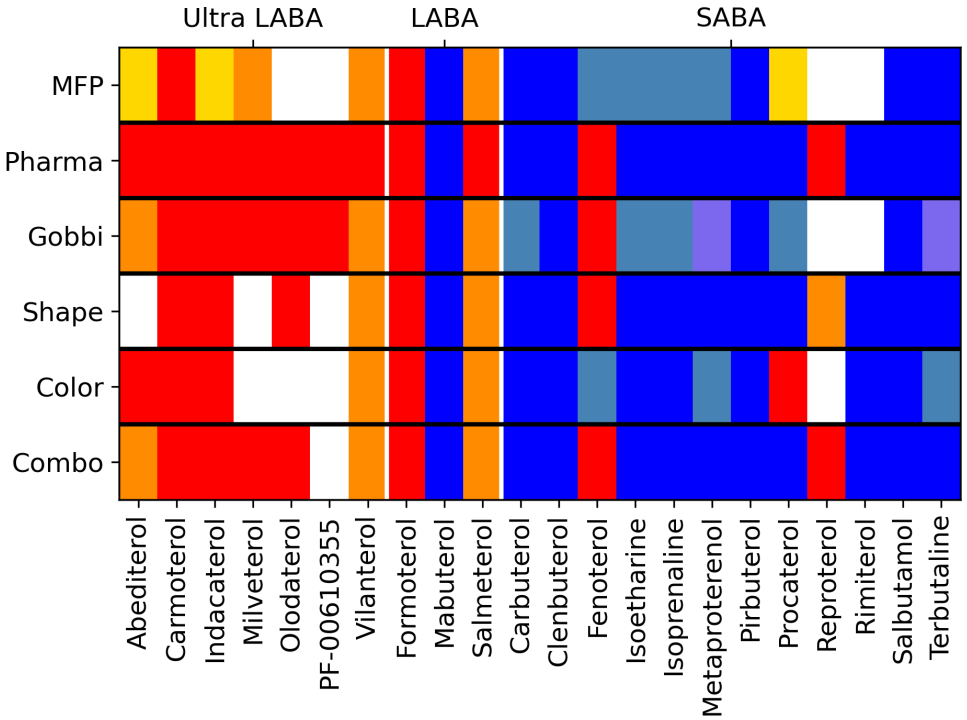


Figure 5.1.4: Clustering of the molecules in the b2-agonists dataset based on six different similarity measures. Molecules in the same cluster are represented by the same colour, white indicates singlets. Clustering was performed using Tanimoto similarity between Morgan fingerprints (MFP), 2D pharmacophoric fingerprints (Pharma, Gobbi), and the three ROCS similarity measures (Shape, Color, Combo). Molecules are shown grouped together based on their label.



machine learning classifiers. Molecular fingerprints generally performed worse than methods based on pharmacophoric features or 3D shape, indicating that pharmacophores are a better molecular representation to predict the duration of action.

Modelling of the ADRB2-agonist complex

With the release of AF [18], deep learning models have become the go to approaches for the prediction of protein structure. AF-generated structures have already been used to study protein-ligand interactions [28–30], and their use for virtual screening has been benchmarked. The active state structure of ADRB2 was generated using AlphaFold2-Multistate by Heo and Feig [19]. The active state conformation of the receptor was selected since the two studied reference ligands, formoterol and salbutamol, are agonists and therefore have a higher affinity for the active state conformation compared to the inactive state one [31]. The AF structure is comparable to the available active state crystallographic structures (Table S6-7), without large changes in the backbone conformation (Cα-RMSD $<$ 1 Å). Some differences in the position of the sidechains in the orthosteric binding pocket were observed, with the most notable being the flipped Asn312.

Models trained using data obtained from simulations based on the AF structure and crystallographic structures of the ADRB2-salbutamol and ADRB2-formoterol complexes were employed, to compare the results of prediction pipelines based on experimental or *in silico* structures.

The structures of the *in silico* complexes were determined by protein-ligand docking. The pose forming most of the experimentally determined key interactions (Table 5.1.2) was manually selected as the starting structure. Relevant interactions were identified from articles published before the release of the first ADRB2 crystallographic structure in 2007, to assess the possibility of generating relevant models in absence of reference structural data. Poses matching most of the expected interactions for both ligands were generated by protein-ligand docking. No interaction was observed with Asn312, located in the vicinity of the ligands and Asp113, due to its conformation. Since AF generates unreliable sidechains conformations, we modified the structure of Asn312 to facilitate its interaction with the ligand. Docking in the modified structure yielded poses forming hydrogen bonds with Asn312, without losing any of the previously observed interactions. The same rearrangement of Asn312 was observed in the early steps of MD simulation of the unmodified receptor systems. Since the observed rearrangements match the proposed correction, no further simulations were performed using the original receptor's structure, only data from the modified structure was used for the two complexes.

All-atom MD simulations of the receptor-agonist complexes embedded in a lipid bilayer with explicit solvent were performed. Each simulation consisted of 5 replicas of 100 ns, for a total simulation time of 500 ns for each system. During the trajectory significant structural rearrangements were observed in correspondence of the receptor's loop regions and N-terminus. In the AF structures the third intracellular loop, which is missing or replaced by lysozyme insertion in the available crystallographic structures, presented the largest conformational changes based on root mean square fluctuation (RMSF) (Figure S1). Similar rearrangements were observed in the corresponding truncated region in the simulations based on crystallographic structures (Figure S2). The structure of the binding pocket generated by AF was not changed significantly by the binding of salbutamol (RMSD $<$ 2.0 Å), while formoterol led to larger rearrangements (RMSD $<$ 3.0 Å) (Figure S3). The

Table 5.1.2: Critical residues for agonist binding and signalling initiation. All interactions are referred to the endogenous ligand epinephrine.

Residue	Notes	Ref
Asp113	Binds the ligands' protonated amine	[32, 33]
Ser203	Forms HB with the meta OH of the catechol	[34, 35]
Ser204	Forms HB with the meta OH of the catechol	[36, 37]
Ser207	Forms HB with the para OH of the catechol	[36, 37]
Phe289	Forms π-π interactions with the catechol	[36]
Phe290	Forms π-π interactions with the catechol	[36]
Asn293	Responsible for stereospecificity	[38]
Asn312	Mostly involved in antagonist binding. Its mutation also affects the affinity of agonists	[39]
Tyr308	Relevant for affinity of selective agonists	[40–42]

binding pocket RMSD in the simulations of crystallographic structures was below 2.0 Å for the formoterol complex, and below 2.5 Å for the salbutamol complex (Figure S4).

Protein-ligand interactions were detected and represented as interaction fingerprints (IFP)[43] and triplets of interaction pseudoatoms (IPAs)[44]. All simulations present conserved interactions corresponding to the ones determined from experiments (Figure 5A). The contribution of each interaction type (hydrogen bond, ionic bond, π-π interaction, metal chelation, and hydrophobic contact) to the binding mode was determined by the average number of IPAs used to describe it (Figure 5B). Polar interactions represented between a third and a half of all interactions: 3 to 5 IPAs for hydrogen bonds with the protein acting as the donor. Slightly less than 15 IPA for hydrogen bonds with the protein acting as the acceptor. 5 IPAs describing the ionic bond with Asp113. π-π interactions were mostly absent, except in the crystallographic complex with formoterol. Hydrophobic contacts were determined using the default (Hyd) and stricter (NewHyd) definitions available in IChem [45]. Formoterol in the AF structure presents a larger number of Hyd IPAs than the others, while all complexes present a similar number of NewHyd IPAs.

Structure-based approaches

The structure-based pipeline is formed by two steps: docking and pose rescoring. Ensemble docking was performed using multiple representative conformations of the binding pocket extracted from the trajectories. Pose rescoring was performed using IG-OCSVM.

IG-OCSVM is a machine learning model trained using interaction data from a MD simulation between the target and a reference ligand. It classifies the docking-generated binding modes as comparable or not to the interactions in the training data. A ligand was considered as selected by IG-OCSVM if at least one of its docking poses obtained a positive score. Models were trained using the simulations of the salbutamol-ADRB2, and formoterol-ADRB2 complexes. Models trained on simulations of the AF generated structure could present a clear bias towards one class of agonists. Salbutamol-based IG-OCSVMs were selective towards SABAs regardless of the used definition of interactions (precision between 0.82 and 0.91), while only Hyd-formoterol models were selective towards LABAs (precision 0.80) (Table S8). IG-OCSVMs selected specific groups of ligands (Table S9-10)

Figure 5.1.5: (A) Hydrogen bond average IFP formed by formoterol and salbutamol with ADRB2 during the MD simulation. (B) Average number of IPAs used to describe the binding mode of formoterol and salbutamol during the MD simulation. HBD hydrogen bonds with the protein acting as the donor, HBA hydrogen bonds with the protein acting as the acceptor, ION ionic interactions, HYD hydrophobic contacts, ARO $\pi - \pi$ interactions. The lighter-coloured and legend-coloured parts of HYD bars corresponds to the default (Hyd) and stricter (Newhyd) definition of hydrophobic contacts, respectively. (C) and (D) Structure of the ADRB2-salbutamol and ADRB2-formoterol complex based on the AF protein structure extracted from the MD simulation. Hydrogen bonds are represented by the yellow dotted lines. Protein residues shown correspond to the ones found in the literature search or forming conserved hydrogen bonds during the MD simulation. TM4 was removed to improve the visibility of the binding cavity.

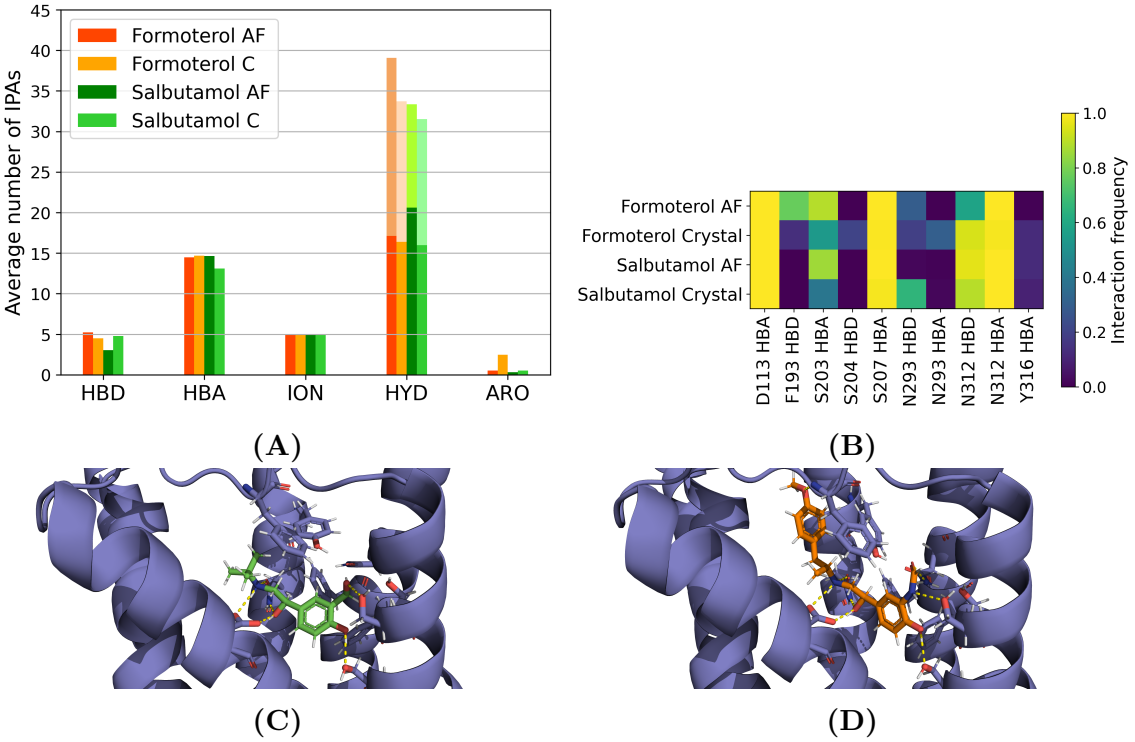
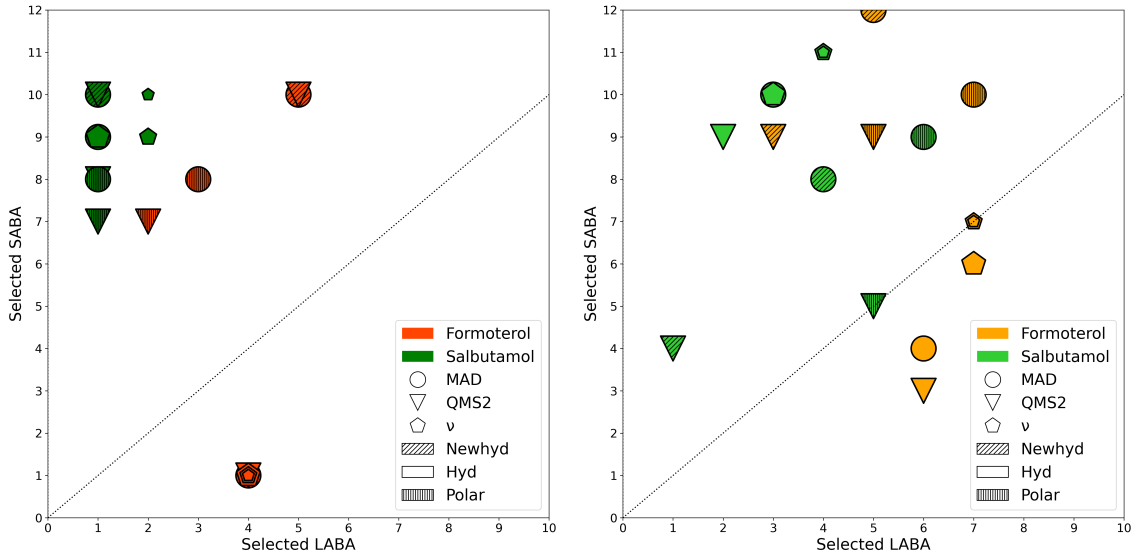


Figure 5.1.6: Number of LABAs and SABAs selected by the trained IG-OCSVM models. Each model is characterized by the reference ligand (colour), training method (shape), and type of interactions (hatching). (A) Results from models trained on simulations based on AF structures, Hyd-formoterol models all selected the same molecules; therefore, the points are overlapping in the plot. (B) Results from models trained on simulations based on crystallographic structures. Hyd salbutamol model trained using MAD and QMS2 select the same molecules, therefore the points are overlapping in the plot.



depending on the reference and interaction definition. Nine of the eleven SABAs and the LABAs mabuterol, and carmoterol were selected by Salbutamol models. Hyd-formoterol models selected four LABAs, and the SABA fenoterol. Newhyd-formoterol models selected the same LABAs as the Hyd-formoterol ones, while also including most of the SABAs and the LABA mabuterol (Figure 6A). Models based on crystallographic structures selected a larger number of molecules, presenting a similar behaviour as their *in silico* counterparts, excluding for the overall decrease in precision (Figure 6B). Salbutamol-based models were less biased towards SABAs (precision values between 0.82 and 0.50) compared to salbutamol AF models, adding formoterol and olodaterol to the previously incorrectly detected mabuterol. As previously observed formoterol models trained on Hyd interactions were biased towards LABAs (precision between 0.50 and 0.67), selecting abediterol, carmoterol, indacaterol, olodaterol, formoterol, salmeterol, and the SABAs clenbuterol, fenoterol, isoprenaline, and salbutamol.

IGs contain relevant information linked to the duration of action of β2-agonists and IG-OCSVM identifies a similarity threshold for classification, therefore only the LABAs and SABAs forming the most similar interactions to the reference were selected, while discarding the remaining molecules, mainly from the opposite class. Most of the SABAs were identified using salbutamol as the reference, while with formoterol only a small number of LABAs (carmoterol, olodaterol, formoterol, and indacaterol) were selected by both AF and crystallographic models.

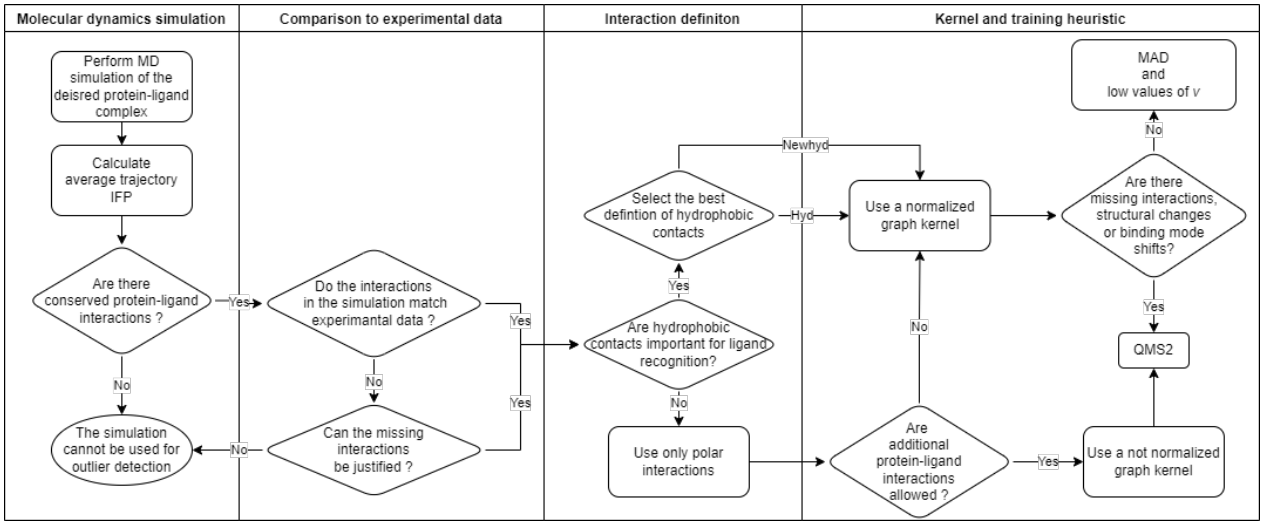
5.1.3 Discussion

In this work we explored different approaches to classify β2-agonists based on their duration of action, or to rank them to favour a specific class.

The lipophilicity of a ligand, as described by logP, was identified as one of the driving forces regulating the duration of action of β2-agonists. This result is in accordance with the microkinetic theory mechanism, which links the duration of action of β2-agonists to the partition coefficient between the aqueous phase and the membrane. The role of logP could explain the behaviour of some compounds with unusual structure or molecular weight. Mabuterol has a similar structure and molecular weight (310 Da) to SABAs, but its lipophilicity (3.36) classifies it as a LABA. The opposite was observed for reproterol, a SABA with a higher molecular weight (389 Da) but also the only hydrophilic molecule in the dataset (logP = -0.44). While logP can qualitatively separate β2-agonists between LABAs and SABAs, it cannot be used to explain the difference in duration of action between LABAs and Ultra-LABAs. One reason might be the descriptor itself, the partition coefficient between water and octanol is often used as a stand-in for the partition coefficient between water and the lipid membrane, but the membrane around GPCRs is characterized by the presence of rafts rich in cholesterol[46], whose physicochemical properties are different from octanol. The difference in affinity for the raft-rich regions of the membrane has already been used to explain the Ultra-LABA character of indacaterol[5, 47, 48]. Other mechanisms and factors, such as the binding kinetics, might also play a role in determining the duration of action of β2-agonist.

Approaches based on the ligands' structure could discriminate between LABAs and SABAs. The structural differences between the two classes already emerged from the clustering of the β2-agonists dataset. With few notable exceptions (mabuterol, fenoterol) members of the same cluster shared the same label. The clustering also highlighted the structural differences within each class: SABAs were generally more similar to each other, being grouped in few clusters, while LABAs were split between multiple clusters or presented as singlets. These differences within each class might explain the considerable drop in performance for models trained on a smaller dataset. Models trained on the β2-agonists dataset performed well in cross validation, regardless of the similarity metric, while limiting the number of representative ligands to formoterol and salbutamol led to a drop in performances. Molecular fingerprints kNN classifiers required a larger number of training references than 2D pharmacophoric fingerprints or 3D similarity models to distinguish between the two classes. Based on the analysis of SHAP values in the CT and RF models trained on molecular and 2D pharmacophoric fingerprints, the most relevant features for classification were the nature of the substituent bound to the amine, or the relative position of aromatic features. The duration of action seems to be linked to the number of aromatic rings in the scaffold, and therefore the shape of the molecule, rather than a specific functional group. Ranking based on 3D molecular similarity could favour ligands with the desired duration of action, but the results were highly dependent on the reference ligand and the used similarity measure. Salbutamol performed better with Tanimoto similarity, and formoterol with Tversky similarity. Since LABAs are generally larger than SABAs and could be described as "superstructures" of the letters, they are better discriminated in comparison to salbutamol using a symmetric similarity measure. On the other hand, many LABAs and Ultra-LABAs are larger than formoterol, therefore the asymmetric Tversky similarity gives better scores to such molecules, while still penalizing the smaller SABAs.

Figure 5.1.7: Updated flowchart for IG-OCSVM training.



The results from interaction-based methods appeared to be linked to ROCS results evaluating 3D molecular similarity, regardless of the origin of the training data. The more selective IG-OCSVMs using Hyd interactions selected mostly molecules clustered with the reference ligands based on shape similarity: Formoterol-based IG-OCSVMs selected carmoterol, indacaterol, olodaterol, and fenoterol, while salbutamol-based IG-OCSVMs selected all the SABAs, except for reproterol and fenoterol, and the LABA mabuterol. Comparing the results from the interaction and ligand-based models, IGs seems to include some information regarding the 3D shape of the ligand, leading to similar results between the two approaches. IG-OCSVM, thanks to its built-in threshold, was able to identify ligands with relevant binding pattern, while discarding less interesting docking poses, leading to more meaningful results than simple similarity ranking. The results however were highly dependent on the parameters and data used for model training, with different simulations presenting different biases in the selected ligands.

The original training workflow of IG-OCSVM [15] was updated to include the current work's results (Figure 7). Originally the average MD binding mode was compared to the reference crystallographic structure to determine the quality of the simulation data, to include targets without a resolved experimental structure we propose to use any available experimental data, such as mutagenesis experiment, to evaluate the results. Newhyd hydrophobic contacts were previously recommended over Hyd to reduce the number of hydrophobic IPAs in the IGs. However, the larger number of hydrophobic IPAs was necessary to bias the search towards LABAs for formoterol-based IG-OCSVMs. The best definition of interaction should be considered on a case-by-case basis, considering the specifics of the complex. Additional consideration for the selection of the best training heuristic and the introduction of a new method were also added in the workflow.

5.1.4 Conclusions

Lipophilicity, and by extension the ability of a ligand to partition between the aqueous phase and the lipid bilayer, was identified as the most relevant physiochemical property separating

LABAs from SABAs. While analysis of the logP is sufficient for qualitative classification, the development of quantitative models, or methods to distinguish LABAs from Ultra-LABAs, would require a better understanding of the underlying mechanism regulating the duration of action, a larger and more diverse training set, and better descriptors.

Interaction-based methods, which have already been used to identify ligands with a desired function, were biased towards ligands with a specific duration of action despite the binding mode not being directly related to it. IGs include some information regarding the 3D ligand structure, leading to results comparable to the ones obtained from molecular similarity. IG-OCSVM models introduced a bias towards LABAs or SABAs directly in the SBDD virtual screening pipeline, limiting the need to further filter the results based on ligand properties.

Combining the filtering of chemical libraries based on molecular weight and logP, or other measures of lipophilicity, with the proposed interaction-based pipeline should yield the best overall results. The first step reduces the number of molecules to evaluate based on their property, while the second one identifies agonists by including an additional shape-based bias to the selection.

5.1.5 Materials and methods

Preparation of the β2-agonists dataset

A list of known β2-agonists was retrieved from publicly available sources: Chembl[49, 50], IUPHAR guide to pharmacology[51], DrugBank[52], and from the analysis of Ritchie et al. on respiratory drugs [53]. Ligands without a publicly available structure, specific enantiomers of racemic drugs, pro-drugs (bambuterol, bitolterol, dipivefrin, droxidopa) and bivalent ligands (batefentrol and navafenterol) were removed from the dataset. Ligands were labelled as SABA, LABA or Ultra-LABA based on a literature search. Ligands without a clear indication of their duration of action (arbutamine, bedoradrine, dilevalol, dobutamine, ephedrine, epinephrine, etafedrine, isoxsuprine, norepinephrine, phenylpropanolamine, protokylol, ractopamine, ritodrine, tretoquinol, zinterol) were discarded from the dataset. Tulobuterol (LABA) was discarded since it is administered using patches [54], without indications on whether the duration of action depends on molecular properties or on the administration route.

Physicochemical properties

Descriptors of the ligands were calculated using RDKit. logP was calculated using the Crippen method[55]. The Lipinksi definition of hydrogen bond donor and acceptor groups was used. The descriptors were scaled using the maximum absolute scaling method before performing PCA. The CT model was trained and evaluated in LOO-CV. Scikit-Learn was used for data scaling, PCA and CT.

Molecular similarity

Morgan binary fingerprints (r = 2), and 2D pharmacophoric fingerprints were computed using RdKit[26]. Pairwise molecular similarity between the fingerprints was calculated using SciPy[56], with metric Jaccard.

The initial 3D structures of the ligands were generated using CORINA 3.40 (Molecular Networks GmbH, Nurnberg, Germany) [57]. For ligands with undefined stereochemistry

Table 5.1.3: Definition of transmembrane domains used for protein alignment.

Structure	First residue	Last residue
TM1	Q26	F61
TM2	T66	K97
TM3	G102	S137
TM4	T146	H172
TM5	N196	E237
TM6	S262	Q299
TM7	R304	R328

all possible stereoisomers were generated. For each stereoisomer 50 conformations were generated using OMEGA (OpenEye, Cadence Design Systems, Inc.). 3D pairwise similarity was computed with ROCS v3.4.1 (OpenEye, Cadence Design Systems, Inc.), by using all ligands as both query and database. Given a pair of ligands, the best similarity score was selected for analysis. ROCS similarity was measured using two scores: Tanimoto and Tversky.

$$Tanimoto_{r,q} = \frac{O_{r,q}}{I_r + I_q - O_{r,q}}$$

$$Tversky_{r,q} = \frac{O_{r,q}}{0.95 * I_r + 0.05 * I_q}$$

With r indicating the reference molecule, q the molecule compared to the reference, O the volume overlap between two entities, I the self-volume overlap of an entity. For Tversky similarity a larger weight was given to the features of the reference ligand.

Clustering was performed using the hierarchical agglomerative clustering algorithm implemented in SciKit-Learn[58], with average linkage. The optimal number of clusters was determined by analysis of the silhouette scores.

The SciKit-Learn implementation of kNN, CT, and RF were used to implement the classifiers. For models trained on the entire β2-agonists dataset the optimal parameters, and the accuracy of the models was determined using leave one out cross validation (LOO-CV). The accuracy of the models trained using formoterol and salbutamol was evaluated using the entire β2-agonists dataset as a test set. For kNN models the predictions were obtained by weighting the contribution of each neighbour by its distance from the target. Since CT results depend on the random state used for feature shuffling, models were trained using 100 random seeds and the average result was reported. RF models were formed by 100 CTs, training was performed with and without data bootstrapping.

AlphaFold structure evaluation

The active state AF structure of human ADRB2 was downloaded from GRPCdb. Residues corresponding to low confidence regions in the unfolded N-terminus and C-terminus were removed, with the final structure consisting of residues from 21 to 341.

The definition of transmembrane domains (TMs) was taken from the GPCRdb entry for ADRB2 (Table 5.1.3).

The residues forming the binding cavity were identified from the apo structure using the cavity detection tool of MOE 2020. Only cavities corresponding to the orthosteric binding site were selected. Four additional residues were included in the binding site definition after simulation of the systems since they form an interaction with either of the ligands for at least 10 % of the analysed frames of a single trajectory.

The list of available ADRB2 crystallographic structures was obtained from GPCRdb. The structures were grouped based on the annotated: Ligand, insertion, and antibodies. From each group the structure with the best resolution was selected.

The analysis of the reference crystallographic structures and AF structure was performed using MOE 2020. The RMSD between the structures was computed after aligning the structures on the TMs.

$$RMSD = \sqrt{\frac{\sum_i^N (X_i - Y_i)^2}{N}}$$

With N indicating the number of selected atoms, X_i and Y_i the coordinates of the atom in the two compared structures.

ADRB2-agonist modelling and simulation

The preparation of the AF and crystallographic structures (PDB-IDs: 7DHI[16],7BZ2[17]) was performed using MOE 2020. The protein builder tool was used to cap truncated N-terminus and C-terminus with acetyl and methylamine groups respectively. In the AF structure multiple rotamers of Asn312 were generated using the same tool, with the best rotamer presenting the amide's oxygen and nitrogen in inverted positions being selected.

Docking of the two ligands was performed in the AF structure. The structures were protonated and charged using the structure preparation tool. 3D conformations of the enantiomers with higher affinity ((R)-salbutamol, and (R,R)-formoterol) were generated using Corina 3.40. PLANTS v1.2 was used to place the ligands in the binding site, with the CHEMPLP [59] scoring function at search speed 1. The search space was set to a radius 10 Å from the centre of mass of the residues forming the orthosteric cavity. For each ligand ten different docking poses were generated. Interactions between the docked ligands and the receptor were computed using the IFP module of IChem v5.2.9 [45]. The pose presenting the most relevant interactions between the receptor and the ligand was manually selected.

Using CHARMM-GUI the structures were inserted in a lipid bilayer formed by POPC, POPE, and cholesterol in a 2:2:1 ratio and solvated using a 0.15 M aqueous solution of KCl. Molecular dynamics parameters for the small molecules were obtained from the generic atom force field (GAFF)[60], using Antechamber[61].The receptor was parametrized using the ff14SB[62] force field, the lipid membrane using Lipid14[63], and the water using tip3p[64].

Molecular dynamics simulations were performed using Amber2018[65, 66]. The initial system underwent minimization for 10000 steps using the steepest gradient descent method, then for an additional 5000 steps using the conjugated gradient method. Minimization was performed at constant volume conditions. The system was heated from 0 to 300 K in 1.00 ns. Temperature was regulated by the Langevin thermostat. The different components of the systems were kept in their original position using harmonic restrains. Heating from 0 to 300 K was performed at constant volume to allow for the density of the system to equilibrate. Positional restrains were gradually removed in 1800 ps at constant pressure using anisotropic

pressure scaling. Heating and release of restrains were performed with a 0.001 ps timestep. Equilibration was performed for 20 ns with a 0.002 ps timestep. The equilibrated system was simulated for five replicas of 100 ns with a 0.002 ps timestep. The trajectory was restarted, without reading previous velocities, every 10 ns.

For each trajectory the RMSD from the initial AF structure, and RMSF were computed using pytraj[67, 68].

$$RMSF_i = \sqrt{\frac{1}{T}\sum_{j}^{T}(x_{i(j)} - \langle x_i \rangle)^2}$$

With i indicating the residue, T the number of examined frames, xi(j) the position of residue i in frame j, and <xi> the average position of residue i.

From each trajectory one frame every 0.1 ns was converted to pairs of mol2 files containing the ligand and the receptor using a python script. Interaction fingerprints and IPAs were generated using IChem v5.2.9. Only interactions present in at least 10% of the analysed frames of a single simulation were visualized in the graphical representation of the IFP (Figure 5A).

Protein-ligand docking and rescoring

Protonation of the ligands in the β2-agonists dataset was performed using FILTER 2.5.1.4 (OpenEye, Cadence Design Systems, Inc.). Ligand conformations were generated using Corina 3.40. For ligands with an undefined stereochemistry all stereoisomers were generated. Multiple ring conformation were generated using the rc option.

From each MD simulation representative structures were selected for ensemble docking. Hierarchical agglomerative clustering, with average linkage, was performed using the pairwise RMSD values calculated on non-hydrogen atoms in the binding site, after performing structure alignment on the pocket residues. The optimal number of clusters was evaluated based on the average silhouette score and the distribution of sample silhouette scores. Clusters containing single structures were discarded from the analysis. From each cluster the structure with the lowest RMSD compared to the others was selected as the most representative.

Docking was performed using PLANTS v1.2, with the CHEMPLP scoring function and search speed 1. The search space was defined by a radius of 12 Å from the centre of mass of the binding pocket residues. Interactions were detected as triplets of IPAs using IChem v 5.9.2, using both defintion of hydrophobic contacts: Hyd, based only on the atom type and the distance between atoms. Newhyd, based on the atom type, the distance between atoms, and the hydrophobic nature of the surrounding environment. The IPA representation of interaction is used by both GRIM and IG-OCSVM for pose rescoring.

The IG-OCSVM scripts were downloaded and modified from the GitHub repository (https://github.com/LIT-CCM-lab/OCSVM-ADRB2). Models were trained using an update version of the training workflow proposed in the original publication (Figure 7). The training heuristics median absolute deviation (MAD) and quick model selection 2 (QMS2), were used alongside an additional approach based on setting the ν parameter to an arbitrary small value (ν= 0.01, 0.005, 0.001). This new approach was applied only for the training of Hyd models. The interactions from the ADRB2-salbutamol and ADRB2-formoterol simulations were used to train two groups of models, one for each ligand.

The trained IG-COSVMs were used to post-process the docking results. The poses' IGs were classified as either inliers (positive score) or outliers (negative score). A ligand was considered as selected by a model if at least one of the evaluated docking poses had a positive score.

Model evaluation

The performance of classification models was evaluated using the accuracy score, which measures the fraction of correctly predicted instances over the total number of evaluated instances.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

For the one class classification models (IG-OCSVM), the performances were evaluated using precision, recall, and F1-score, since one class classification methods should be evaluated based on the results on the positive class, rather than the entire dataset.

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = 2\frac{Precision * Recall}{Precision + Recall}$$

TP (true positives) and TN (true negatives) indicate the correctly predicted instances, FP (false positives) and FN (false negatives) indicate the incorrectly predicted instances.

For ranking methods, the quality of the predictions was evaluated using the area under the receiver operator curve (ROCAUC) score.

Acknowledgements

The authors acknowledge the High-Performance Computing Center of the French National Institute for Nuclear and Particle Physics (IN2P3). This work was supported by the Agence Nationale de Recherches (ANR) (Grant Number 2019 CE14 OCHRE to E.K.), Centre National de la Recherche Scientifique (CNRS), Institut du Médicament de Strasbourg (IMS), and Université de Strasbourg.

References

[1] Charlotte K. Billington, Raymond B. Penn, and Ian P. Hall. *B*2 Agonists. In Clive P. Page and Peter J. Barnes, editors, *Pharmacology and Therapeutics of Asthma and COPD*, volume 237, pages 23–40. Springer International Publishing, Cham, 2016.

[2] Mario Cazzola, Paola Rogliani, and Maria Gabriella Matera. Ultra-LABAs for the treatment of asthma. *Respiratory Medicine*, 156:47–52, September 2019.

[3] Mario Cazzola, Maria Gabriella Matera, and Jan Lötvall. Ultra long-acting β_2 -agonists in development for asthma and chronic obstructive pulmonary disease. *Expert Opinion on Investigational Drugs*, 14(7):775–783, July 2005.

[4] Stacy Gelhaus Wendell, Hao Fan, and Cheng Zhang. G Protein–Coupled Receptors in Asthma Therapy: Pharmacology and Drug Action. *Pharmacological Reviews*, 72(1):1–49, January 2020.

[5] Lorraine Murphy, Stephen Rennard, James Donohue, Mathieu Molimard, Ronald Dahl, Kai-Michael Beeh, Juergen Dederichs, Hans-Jürgen Fülle, Mark Higgins, and David Young. Turning a Molecule into a Medicine: The Development of Indacaterol as a Novel Once-Daily Bronchodilator Treatment for Patients with COPD. *Drugs*, 74(14):1635–1657, September 2014.

[6] Maria Gabriella Matera and Mario Cazzola. Ultra-Long-Acting $B2$-Adrenoceptor Agonists. *Drugs*, 67(4):503–515, March 2007.

[7] G.P. Anderson, A. Lindén, and K.F. Rabe. Why are long-acting beta-adrenoceptor agonists long-acting? *European Respiratory Journal*, 7(3):569–578, March 1994.

[8] A Szczuka, M Wennerberg, A Packeu, and G Vauquelin. Molecular mechanisms for the persistent bronchodilatory effect of the β_2 -adrenoceptor agonist salmeterol. *British Journal of Pharmacology*, 158(1):183–194, September 2009.

[9] David A Sykes and Steven J Charlton. Slow receptor dissociation is not a key factor in the duration of action of inhaled long-acting $B2$-adrenoceptor agonists: Exploring $B2$-adrenoceptor kinetics of LABAs. *British Journal of Pharmacology*, 165(8):2672–2683, April 2012.

[10] Paola Casarosa, Ines Kollak, Tobias Kiechle, Angela Ostermann, Andreas Schnapp, Ralf Kiesling, Michael Pieper, Peter Sieger, and Florian Gantner. Functional and Biochemical Rationales for the 24-Hour-Long Duration of Action of Olodaterol. *Journal of Pharmacology and Experimental Therapeutics*, 337(3):600–609, June 2011.

[11] R. A. Coleman, M. Johnson, A. T. Nials, and C. J. Vardey. Exosites: Their current status, and their relevance to the duration of action of long-acting beta 2-adrenoceptor agonists. *Trends in Pharmacological Sciences*, 17(9):324–330, September 1996.

[12] Matthieu Masureel, Yaozhong Zou, Louis-Philippe Picard, Emma van der Westhuizen, Jacob P. Mahoney, João P. G. L. M. Rodrigues, Thomas J. Mildorf, Ron O. Dror, David E. Shaw, Michel Bouvier, Els Pardon, Jan Steyaert, Roger K. Sunahara, William I. Weis, Cheng Zhang, and Brian K. Kobilka. Structural insights into binding specificity, efficacy and bias of a β2AR partial agonist. *Nature Chemical Biology*, 14(11):1059–1066, November 2018.

[13] Jillian G. Baker, Richard G. W. Proudman, and Stephen J. Hill. Salmeterol's Extreme $B2$ Selectivity Is Due to Residues in Both Extracellular Loops and Transmembrane Domains. *Molecular Pharmacology*, 87(1):103–120, January 2015.

[14] Chris de Graaf and Didier Rognan. Selective Structure-Based Virtual Screening for Full and Partial Agonists of the $B2$ Adrenergic Receptor. *Journal of Medicinal Chemistry*, 51(16):4978–4985, August 2008.

[15] Luca Chiesa and Esther Kellenberger. One class classification for the detection of $B2$ adrenergic receptor agonists using single-ligand dynamic interaction data. *Journal of Cheminformatics*, 14(1):74, October 2022.

[16] Dehua Yang, Qingtong Zhou, Viktorija Labroska, Shanshan Qin, Sanaz Darbalaei, Yiran Wu, Elita Yuliantie, Linshan Xie, Houchao Tao, Jianjun Cheng, Qing Liu, Suwen Zhao, Wenqing Shui, Yi Jiang, and Ming-Wei Wang. G protein-coupled receptors: Structure- and function-based drug discovery. *Signal Transduction and Targeted Therapy*, 6(1):1–27, January 2021.

[17] Yanan Zhang, Fan Yang, Shenglong Ling, Pei Lv, Yingxin Zhou, Wei Fang, Wenjing Sun, Longhua Zhang, Pan Shi, and Changlin Tian. Single-particle cryo-EM structural studies of the β2AR–Gs complex bound with a full agonist formoterol. *Cell Discovery*, 6(1):45, July 2020.

[18] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, August 2021.

[19] Lim Heo and Michael Feig. Multi-state modeling of G-protein coupled receptors at experimental accuracy. *Proteins: Structure, Function, and Bioinformatics*, 90(11):1873–1885, 2022.

[20] Valeria Scardino, Juan I. Di Filippo, and Claudio N. Cavasotto. How good are AlphaFold models for docking-based virtual screening? *iScience*, 26(1), January 2023.

[21] Matthew Holcomb, Ya-Ting Chang, David S. Goodsell, and Stefano Forli. Evaluation of AlphaFold2 structures as docking targets. *Protein Science*, 32(1):e4530, 2023.

[22] Sumin Lee, Seeun Kim, Gyu Rie Lee, Sohee Kwon, Hyeonuk Woo, Chaok Seok, and Hahnbeom Park. Evaluating GPCR modeling and docking strategies in the era of deep learning-based protein structure prediction. *Computational and Structural Biotechnology Journal*, 21:158–167, January 2023.

[23] Scott M Lundberg and Su-In Lee. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[24] Scott M. Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. From local

explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1):56–67, January 2020.

[25] Gerald Maggiora, Martin Vogt, Dagmar Stumpfe, and Jürgen Bajorath. Molecular Similarity in Medicinal Chemistry. *Journal of Medicinal Chemistry*, 57(8):3186–3204, April 2014.

[26] Greg Landrum, Paolo Tosco, Brian Kelley, Ric, Sriniker, Gedeck, Riccardo Vianello, NadineSchneider, Eisuke Kawashima, Andrew Dalke, Dan N, David Cosgrove, Brian Cole, Matt Swain, Samo Turk, AlexanderSavelyev, Gareth Jones, Alain Vaucher, Maciej Wójcikowski, Ichiru Take, Daniel Probst, Kazuya Ujihara, Vincent F. Scalfani, Guillaume Godin, Axel Pahl, Francois Berenger, JLVarjo, Strets123, JP, and DoliathGavid. Rdkit/rdkit: 2022_03_2 (Q1 2022) Release. Zenodo, April 2022.

[27] Alberto Gobbi and Dieter Poppinger. Genetic optimization of combinatorial libraries. *Biotechnology and Bioengineering*, 61(1):47–54, 1998.

[28] Albert Ros-Lucas, Nieves Martinez-Peinado, Jaume Bastida, Joaquim Gascón, and Julio Alonso-Padilla. The Use of AlphaFold for In Silico Exploration of Drug Targets in the Parasite Trypanosoma cruzi. *Frontiers in Cellular and Infection Microbiology*, 12, 2022.

[29] Matteo Rossi Sebastiano, Giuseppe Ermondi, Shinji Hadano, and Giulia Caron. AI-based protein structure databases have the potential to accelerate rare diseases research: AlphaFoldDB and the case of IAHSP/Alsin. *Drug Discovery Today*, 27(6):1652–1660, June 2022.

[30] Rui Tian, Yufei Li, Xiaofeng Wang, Jiajun Li, Yingqian Li, Shaosheng Bei, and Huashan Li. A Pharmacoinformatics Analysis of Artemisinin Targets and de novo Design of Hits for Treating Ulcerative Colitis. *Frontiers in Pharmacology*, 13, 2022.

[31] Albert J. Kooistra, Rob Leurs, Iwan J. P. de Esch, and Chris de Graaf. Structure-Based Prediction of G-Protein-Coupled Receptor Ligand Function: A β-Adrenoceptor Case Study. *Journal of Chemical Information and Modeling*, 55(5):1045–1061, May 2015.

[32] C D Strader, I S Sigal, R B Register, M R Candelore, E Rands, and R A Dixon. Identification of residues required for ligand binding to the beta-adrenergic receptor. *Proceedings of the National Academy of Sciences*, 84(13):4384–4388, July 1987.

[33] C. D. Strader, I. S. Sigal, M. R. Candelore, E. Rands, W. S. Hill, and R. A. Dixon. Conserved aspartic acid residues 79 and 113 of the beta-adrenergic receptor have different roles in receptor function. *The Journal of Biological Chemistry*, 263(21):10267–10271, July 1988.

[34] Takayuki Sato, Hiroyuki Kobayashi, Taku Nagao, and Hitoshi Kurose. Ser203 as well as Ser204 and Ser207 in fifth transmembrane domain of the human $B2$-adrenoceptor contributes to agonist binding and receptor activation. *British Journal of Pharmacology*, 128(2):272–274, 1999.

[35] George Liapakis, Juan A. Ballesteros, Stavros Papachristou, Wai Chi Chan, Xun Chen, and Jonathan A. Javitch. The Forgotten Serine: A CRITICAL ROLE FOR Ser-2035.42

IN LIGAND BINDING TO AND ACTIVATION OF THE *B*2-ADRENERGIC RECEPTOR *. *Journal of Biological Chemistry*, 275(48):37779–37788, December 2000.

[36] R. a. F. Dixon, I. S. Sigal, and C. D. Strader. Structure-Function Analysis of the β-Adrenergic Receptor. *Cold Spring Harbor Symposia on Quantitative Biology*, 53:487–497, January 1988.

[37] C. D. Strader, M. R. Candelore, W. S. Hill, I. S. Sigal, and R. A. Dixon. Identification of two serine residues involved in agonist activation of the beta-adrenergic receptor. *The Journal of Biological Chemistry*, 264(23):13572–13578, August 1989.

[38] K Wieland, H M Zuurmond, C Krasel, A P Ijzerman, and M J Lohse. Involvement of Asn-293 in stereospecific agonist recognition and in activation of the beta 2-adrenergic receptor. *Proceedings of the National Academy of Sciences*, 93(17):9276–9281, August 1996.

[39] S. Suryanarayana and B. K. Kobilka. Amino acid substitutions at position 312 in the seventh hydrophobic segment of the beta 2-adrenergic receptor modify ligand-binding specificity. *Molecular Pharmacology*, 44(1):111–114, July 1993.

[40] Hideo Kikkawa, Masafumi Isogaya, Taku Nagao, and Hitoshi Kurose. The Role of the Seventh Transmembrane Region in High Affinity Binding of a *B*2-Selective Agonist TA-2005. *Molecular Pharmacology*, 53(1):128–134, January 1998.

[41] Masafumi Isogaya, Yoko Yamagiwa, Shigeo Fujita, Yoshiyuki Sugimoto, Taku Nagao, and Hitoshi Kurose. Identification of a Key Amino Acid of the *B*2-Adrenergic Receptor for High Affinity Binding of Salmeterol. *Molecular Pharmacology*, 54(4):616–622, October 1998.

[42] Masafumi Isogaya, Yoshiyuki Sugimoto, Ryuji Tanimura, Rie Tanaka, Hideo Kikkawa, Taku Nagao, and Hitoshi Kurose. Binding Pockets of the β_1 - and β_2 -Adrenergic Receptors for Subtype-Selective Agonists. *Molecular Pharmacology*, 56(5):875–885, November 1999.

[43] Gilles Marcou and Didier Rognan. Optimizing Fragment and Scaffold Docking by Use of Molecular Interaction Fingerprints. *Journal of Chemical Information and Modeling*, 47(1):195–207, January 2007.

[44] Jérémy Desaphy, Eric Raimbaud, Pierre Ducrot, and Didier Rognan. Encoding Protein–Ligand Interaction Patterns in Fingerprints and Graphs. *Journal of Chemical Information and Modeling*, 53(3):623–637, March 2013.

[45] Franck Da Silva, Jeremy Desaphy, and Didier Rognan. IChem: A Versatile Toolkit for Detecting, Comparing, and Predicting Protein–Ligand Interactions. *ChemMedChem*, 13(6):507–510, 2018.

[46] Van Anthony M. Villar, Santiago Cuevas, Xiaoxu Zheng, and Pedro A. Jose. Chapter 1 - Localization and signaling of GPCRs in lipid rafts. In Arun K. Shukla, editor, *Methods in Cell Biology*, volume 132 of *G Protein-Coupled Receptors*, pages 3–23. Academic Press, January 2016.

[47] Dario Lombardi, Bernard Cuenoud, and Stefanie D. Krämer. Lipid membrane interactions of indacaterol and salmeterol: Do they influence their pharmacological properties? *European Journal of Pharmaceutical Sciences*, 38(5):533–547, December 2009.

[48] Mario Cazzola, Luigino Calzetta, and Maria Gabriella Matera. *B*2-adrenoceptor agonists: Current and future direction. *British Journal of Pharmacology*, 163(1):4–17, 2011.

[49] Anna Gaulton, Anne Hersey, Michał Nowotka, A. Patrícia Bento, Jon Chambers, David Mendez, Prudence Mutowo, Francis Atkinson, Louisa J. Bellis, Elena Cibrián-Uhalte, Mark Davies, Nathan Dedman, Anneli Karlsson, María Paula Magariños, John P. Overington, George Papadatos, Ines Smit, and Andrew R. Leach. The ChEMBL database in 2017. *Nucleic Acids Research*, 45(D1):D945–D954, January 2017.

[50] Mark Davies, Michał Nowotka, George Papadatos, Nathan Dedman, Anna Gaulton, Francis Atkinson, Louisa Bellis, and John P. Overington. ChEMBL web services: Streamlining access to drug discovery data and utilities. *Nucleic Acids Research*, 43(W1):W612–W620, July 2015.

[51] Katrin Altosaar, Poornima Balaji, Richard A. Bond, David B. Bylund, Susanna Cotecchia, Dominic Devost, Van A. Doze, Douglas C. Eikenburg, Sarah Gora, Eugénie Goupil, Robert M. Graham, Terry Hébert, J. Paul Hieble, Rebecca Hills, Shahriar Kan, Gayane Machkalyan, Martin C. Michel, Kenneth P. Minneman, Sergio Parra, Dianne Perez, Rory Sleno, Roger Summers, and Peter Zylbergold. Adrenoceptors in GtoPdb v.2021.3. *IUPHAR/BPS Guide to Pharmacology CITE*, 2021(3), September 2021.

[52] David S Wishart, Yannick D Feunang, An C Guo, Elvis J Lo, Ana Marcu, Jason R Grant, Tanvir Sajed, Daniel Johnson, Carin Li, Zinat Sayeeda, Nazanin Assempour, Ithayavani Iynkkaran, Yifeng Liu, Adam Maciejewski, Nicola Gale, Alex Wilson, Lucy Chin, Ryan Cummings, Diana Le, Allison Pon, Craig Knox, and Michael Wilson. DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Research*, 46(D1):D1074–D1082, January 2018.

[53] Timothy J. Ritchie, Christopher N. Luscombe, and Simon J. F. Macdonald. Analysis of the Calculated Physicochemical Properties of Respiratory Drugs: Can We Design for Inhaled Drugs Yet? *Journal of Chemical Information and Modeling*, 49(4):1025–1032, April 2009.

[54] Yoichi Nakamura, Jun Tamaoki, Hiroyuki Nagase, Masao Yamaguchi, Takahiko Horiguchi, Soichiro Hozawa, Masakazu Ichinose, Takashi Iwanaga, Rieko Kondo, Makoto Nagata, Akihito Yokoyama, and Yuji Tohda. Japanese guidelines for adult asthma 2020. *Allergology International*, 69(4):519–548, October 2020.

[55] Scott A. Wildman and Gordon M. Crippen. Prediction of Physicochemical Parameters by Atomic Contributions. *Journal of Chemical Information and Computer Sciences*, 39(5):868–873, September 1999.

[56] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan

Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, SciPy 1.0 Contributors, Aditya Vijaykumar, Alessandro Pietro Bardelli, Alex Rothberg, Andreas Hilboll, Andreas Kloeckner, Anthony Scopatz, Antony Lee, Ariel Rokem, C. Nathan Woods, Chad Fulton, Charles Masson, Christian Häggström, Clark Fitzgerald, David A. Nicholson, David R. Hagen, Dmitrii V. Pasechnik, Emanuele Olivetti, Eric Martin, Eric Wieser, Fabrice Silva, Felix Lenders, Florian Wilhelm, G. Young, Gavin A. Price, Gert-Ludwig Ingold, Gregory E. Allen, Gregory R. Lee, Hervé Audren, Irvin Probst, Jörg P. Dietrich, Jacob Silterra, James T Webber, Janko Slavič, Joel Nothman, Johannes Buchner, Johannes Kulick, Johannes L. Schönberger, José Vinícius de Miranda Cardoso, Joscha Reimer, Joseph Harrington, Juan Luis Cano Rodríguez, Juan Nunez-Iglesias, Justin Kuczynski, Kevin Tritz, Martin Thoma, Matthew Newville, Matthias Kümmerer, Maximilian Bolingbroke, Michael Tartre, Mikhail Pak, Nathaniel J. Smith, Nikolai Nowaczyk, Nikolay Shebanov, Oleksandr Pavlyk, Per A. Brodtkorb, Perry Lee, Robert T. McGibbon, Roman Feldbauer, Sam Lewis, Sam Tygier, Scott Sievert, Sebastiano Vigna, Stefan Peterson, Surhud More, Tadeusz Pudlik, Takuya Oshima, Thomas J. Pingel, Thomas P. Robitaille, Thomas Spura, Thouis R. Jones, Tim Cera, Tim Leslie, Tiziano Zito, Tom Krauss, Utkarsh Upadhyay, Yaroslav O. Halchenko, and Yoshiki Vázquez-Baeza. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, March 2020.

[57] Jens Sadowski, Johann Gasteiger, and Gerhard Klebe. Comparison of Automatic Three-Dimensional Model Builders Using 639 X-ray Structures. *Journal of Chemical Information and Computer Sciences*, 34(4):1000–1008, July 1994.

[58] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85):2825–2830, 2011.

[59] Oliver Korb, Thomas Stützle, and Thomas E. Exner. Empirical Scoring Functions for Advanced Protein-Ligand Docking with PLANTS. *Journal of Chemical Information and Modeling*, 49(1):84–96, January 2009.

[60] Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case. Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25(9):1157–1174, 2004.

[61] Junmei Wang, Wei Wang, Peter A. Kollman, and David A. Case. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling*, 25(2):247–260, October 2006.

[62] James A. Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E. Hauser, and Carlos Simmerling. ff14SB: Improving the Accuracy of Protein

Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*, 11(8):3696–3713, August 2015.

[63] Callum J. Dickson, Benjamin D. Madej, Åge A. Skjevik, Robin M. Betz, Knut Teigen, Ian R. Gould, and Ross C. Walker. Lipid14: The Amber Lipid Force Field. *Journal of Chemical Theory and Computation*, 10(2):865–879, February 2014.

[64] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, July 1983.

[65] Andreas W. Götz, Mark J. Williamson, Dong Xu, Duncan Poole, Scott Le Grand, and Ross C. Walker. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *Journal of Chemical Theory and Computation*, 8(5):1542–1555, May 2012.

[66] Romelia Salomon-Ferrer, Andreas W. Götz, Duncan Poole, Scott Le Grand, and Ross C. Walker. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *Journal of Chemical Theory and Computation*, 9(9):3878–3888, September 2013.

[67] Hai Nguyen, Jason Swails, Daniel R. Roe, and Cody. Amber-Md/Pytraj: V2.0.0. Zenodo, March 2017.

[68] Daniel R. Roe and Thomas E. Cheatham. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation*, 9(7):3084–3095, July 2013.

5.2 Predicting the duration of action of the agonist BI167107

Abstract

The computational methods based on ligand physicochemical properties, and based on protein ligand interactions described in Section 5.1 were used to predict the duration of action of the known ligand BI167107. Both ligand and structure based models predicted the molecule as long acting. The predictions were confirmed by measuring the duration of action of the molecules in a mouse model, compared to the reference ligands formoterol (LABA), and salbutamol (SABA).

5.2.1 Introduction

The previous section reviewed both ligand-based and structure-based methods to select ADRB2 agonists with a specific duration of action. The retrospective examination of the twenty-two well characterized beta-2 adrenergic agonists identified lipophilicity as the crucial factor affecting the duration of action. Surprisingly, it also revealed that the 3D structure of the agonist when bound to the receptor offered a reliable indication of the duration of action. Despite the LABA formoterol and the SABA salbutamol displaying several common interactions, the small differences in the binding mode led the IG-OCSVM models to be biased towards ligands within the same class as the reference (LABA or SABA). It was suggested that the interaction graphs used to represent the ligand binding mode might include some information about the shape of the ligand, which was observed to be correlated with the duration of action. The structure-based approach formed by docking and IG-OCSVM, and the ligand-based analysis of physicochemical properties represent two orthogonal methods which can be used to identify b2-agonists with a specific duration of action. In this section these two approaches were applied to determine the duration of action of the known agonist BI167107. BI167107 is a high affinity non-selective agonist of ADRB2, whose main application is as a crystallization tool for the study of β-adrenergic receptors[1–3]. To confirm the *in-silico* prediction an asthma mouse model was used to measure the duration of action of BI167107 against the reference ligands formoterol (LABA) and salbutamol (SABA).

5.2.2 Results and Discussion

The agonist efficacy of BI167107 was evaluated computationally using a structure-based pipeline combining docking and IG-OCSVM. In the previous section it was determined that such approach includes a bias towards ligands with a similar duration of action as the reference, allowing to formulate an initial hypothesis on the nature of a ligand. A classification tree model trained on the calculated physicochemical properties of twenty-two well characterized β2-agonists was used to confirm the results. The structure-based pipeline correctly identified BI167107 as an agonist of ADRB2, with the ligand being selected more often by formoterol-based models than salbutamol-based ones. Three out of the five biased formoterol models selected BI167107, while only the least biased salbutamol models selected it. These models were trained by setting a small arbitrary value for ν, rather than inferring the hyperparameter from the training data. By using only IG-OCSVM models trained using either the quick model selection 2 (QMS2) or the median absolute deviation (MAD) heuristics, BI167107 would only be selected by formoterol IG-OCSVM, clearly classifying it

Model	Results
Physicochemical properties classification tree	BI167107 is a LABA
Formoterol IG-OCSVM	7/9 models select BI167107 (4/60 poses passed)
Salbutamol IG- OCSVM	3/9 models select BI167107 (4/60 poses passed)

as a LABA. The classification tree model trained on physicochemical properties identified it as a LABA, due to its molecular weight, and lipophilicity expressed as logP. The consensus between the two orthogonal approaches labelled BI167107 as a LABA.

An asthmatic mouse model was used to measure the duration of action of BI167107, and the two reference ligands salbutamol and formoterol. BI167107 elicited a longer response in mice than any of the two references. The ligand has a long-acting duration of action, confirming the in-silico predictions.

5.2.3 Conclusions

The predictions from the semi supervised structure-based approach, and the supervised ligand-based model complemented each other, allowing to predict the duration of action of a known b2-agonist. As shown in chapter 3 ADRB2 IG-OCSVM models can identify agonists while discarding antagonists and decoys. The structure-based approach identified BI167107 as a likely agonist of ADRB2, and based on the models giving a positive classification it was predicted to be a LABA rather than a SABA. The model trained on physicochemical properties instead could not determine whether BI167107 was an agonist of ADRB2, but it could clearly identify it as a LABA. The in silico predictions were confirmed experimentally using a mouse asthma model.

5.2.4 Materials and Methods

The protocols and data described in Section 5.1 were employed to predict the efficacy and the duration of action of BI167107.

Docking and Rescoring

The protonation state of BI167107 was determined using FILTER 2.5.1.4 (OpenEye, Cadence Design Systems, Inc.). Three different conformations of the ligand were generated using Corina 3.40, with the multiple ring conformation option rc. Docking was performed using PLANTS v1.2 with the CHEMPLP scoring function and search speed 1 on 4 representative structures of the receptor extracted from molecular dynamics simulations of the AlphaFold-based ADRB2-salbutamol and ADRB2-formoterol complexes, two for each simulation (see the previous section for more details). For each ligand conformation 10 poses were generated, for a total of 30 poses for each receptor structure. Interactions were detected and described as triplets of interaction pseudo-atoms using IChem v 5.9.2[4]. The docking results were post-processed using the IG-OSVM models trained using the dynamic binding modes of formoterol and salbutamol extracted from molecular dynamics simulations. The models assign a positive score to selected poses, and a negative score to outliers. BI167107 was considered as selected by a model if at least one of its docking poses had a positive score. Nine

models were trained using each reference, considering different definitions of interactions and different training heuristics. The selectivity of the models for a specific class was previously tested using 22 well characterized LABAs and SABAs: Five formoterol-IG-OCSVM models were selective towards LABAs, while all salbutamol models were generally selective towards SABAs.

Physicochemical properties

The duration of action was predicted using the classification tree model trained on the physicochemical properties of 22 well characterized ligands (12 SABAs and 10 LABAs). The scikit-learn implementation of the classification tree algorithm was used. The physicochemical properties of BI167107 and the reference ligands were calculated using RDKit (logP, molecular weight, topological polar surface area, number of hydrogen bond acceptors, number of hydrogen bond donors, number of rotatable bonds, number of rings, number of aromatic rings). logP was calculated using the Crippen method[5]. The Lipinski definition of hydrogen bond donor and acceptor groups was used.

References

[1] Søren G. F. Rasmussen, Brian T. DeVree, Yaozhong Zou, Andrew C. Kruse, Ka Young Chung, Tong Sun Kobilka, Foon Sun Thian, Pil Seok Chae, Els Pardon, Diane Calinski, Jesper M. Mathiesen, Syed T. A. Shah, Joseph A. Lyons, Martin Caffrey, Samuel H. Gellman, Jan Steyaert, Georgios Skiniotis, William I. Weis, Roger K. Sunahara, and Brian K. Kobilka. Crystal structure of the $B2$ adrenergic receptor–Gs protein complex. *Nature*, 477(7366):549–555, September 2011.

[2] Søren G. F. Rasmussen, Hee-Jung Choi, Juan Jose Fung, Els Pardon, Paola Casarosa, Pil Seok Chae, Brian T. DeVree, Daniel M. Rosenbaum, Foon Sun Thian, Tong Sun Kobilka, Andreas Schnapp, Ingo Konetzki, Roger K. Sunahara, Samuel H. Gellman, Alexander Pautsch, Jan Steyaert, William I. Weis, and Brian K. Kobilka. Structure of a nanobody-stabilized active state of the $B2$ adrenoceptor. *Nature*, 469(7329):175–180, January 2011.

[3] Aaron M. Ring, Aashish Manglik, Andrew C. Kruse, Michael D. Enos, William I. Weis, K. Christopher Garcia, and Brian K. Kobilka. Adrenaline-activated structure of $B2$-adrenoceptor stabilized by an engineered nanobody. *Nature*, 502(7472):575–579, October 2013.

[4] Franck Da Silva, Jeremy Desaphy, and Didier Rognan. IChem: A Versatile Toolkit for Detecting, Comparing, and Predicting Protein–Ligand Interactions. *ChemMedChem*, 13(6):507–510, 2018.

[5] Scott A. Wildman and Gordon M. Crippen. Prediction of Physicochemical Parameters by Atomic Contributions. *Journal of Chemical Information and Computer Sciences*, 39(5):868–873, September 1999.

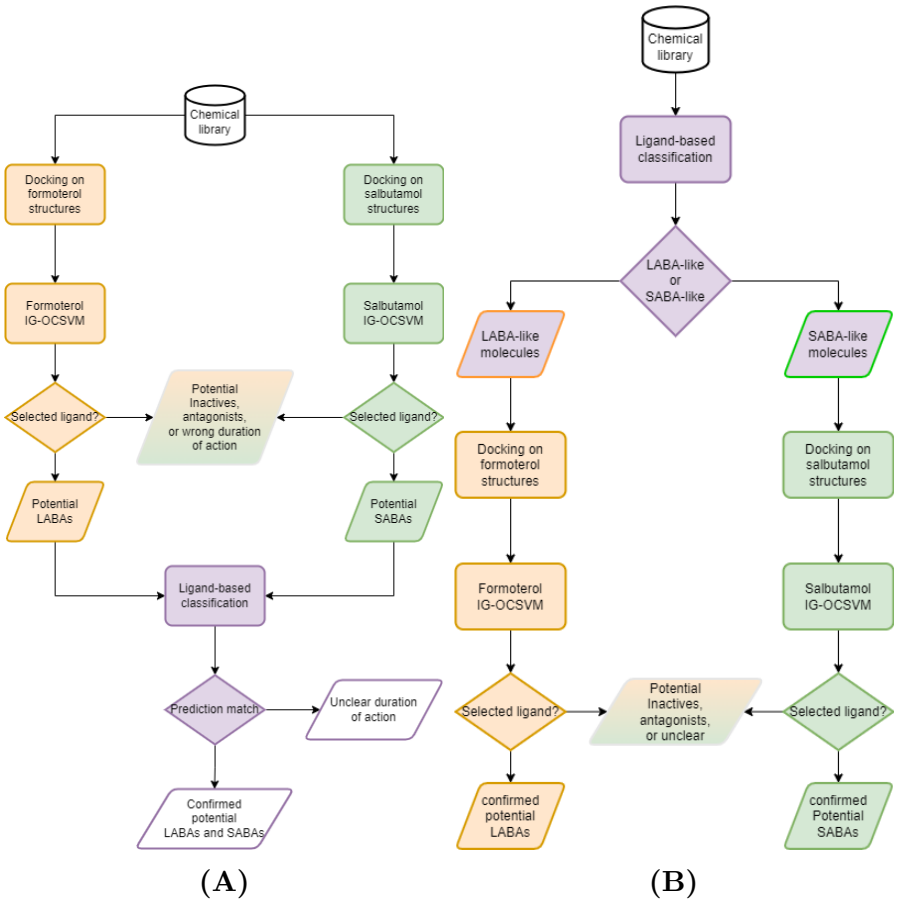
5.3 Critical evaluation

The main objective of this thesis was the development of new computational methods for the discovery of ligand of G-protein coupled receptors with a desired efficacy, however the pharmacological profile of a compound is not limited to its efficacy. Due to its relevance for therapeutical applications the duration of action of agonists of the β2-adrenergic receptor was selected as a relevant additional property to study.

In a retrospective study the semi-supervised IG-OCSVM model was used to evaluate both long acting and short acting agonists, in order to determine whether a bias in the selected ligands could be observed. Furthermore, the structural and physicochemical properties of the evaluated molecules were studied, to determine which ligand-based features regulated the duration of action. The interpretation of the trained models confirmed that the lipophilicity of a ligand was the driving force behind its duration of action, as supported by the reservoir model. However, it was also highlighted that the current descriptors and data can only give a superficial description of the phenomena, with more complex questions, such as the mechanism behind the longer duration of Ultra-LABA, still unanswered. The analysis of the results of the structure-based approach showed that while a bias could be observed for certain models, its magnitude and the overall specificity were reference dependent and interaction representation dependent. The results from Chapter 3 showed that an efficacy bias can generally be observed for InteractionOCSVM models, however additional biases can only be determined by retrospectively analysing known ligands, as described in this chapter. Interaction models developed for other targets might present a different form of bias, therefore only by performing an a posteriori interpretation of the results on known ligand it would be possible to identify if and towards which property a bias exist. Additionally, the known ligands can be used to train ligand-based models, which have the double function of being a complementary tool for prediction when paired with a structure-based approach, as well as giving a potential insight behind the mechanism regulating the studied property. Even if no additional biases were detected, InteractionOCSVM models could still be used in tandem with a ligand-based model. In this scenario the structure-based screening will only identify potential active molecules with the desired efficacy, while the task of selecting molecules with a fine-tuned pharmacological profile will only be a responsibility of the ligand-based models.

As a proof-of-concept the developed in silico pipeline was used to predict the duration of action of the known agonist BI167107. The ligand was docked in structures describing the formoterol-ADRB2 and salbutamol-ADRB2 complexes, and the generated binding modes were evaluated using the corresponding IG-OCSVM models. The classification model trained on physicochemical properties was used to confirm the prediction from the structure-based methods once BI167107 was identified as an ADRB2 agonist and an initial hypothesis regarding its duration of action was proposed. The structure-based analysis represents the most computationally expensive step of the pipeline. While in the analysed case computational efficiency is not relevant, this could seriously limit the application of this approach on larger chemical libraries. The proposed protocol requires to dock the same chemical library twice, once for each structural ensemble, while the ligand-based model is only used to confirm the predictions. A more efficient protocol could use the physicochemical properties classifier to filter all the molecules in the library, separating them in molecules with LABA-like properties or SABA-like properties. Based on the classification results one could decide to remove the molecules with undesired properties and just dock a smaller subset of the entire library, or the classification could be used to select which receptor structure and

Figure 5.3.1: Virtual screening pipelines for the discovery of new β2-agonists with a desired duration of action. (A) The pipeline developed in Section 5.2. (B) A more computationally efficient pipeline based on the same methodology.



IG-OCSVM models are more likely to identify the molecule as a potential agonist. This alternative pipeline should at least halve the screening time. The downside would be that molecules with a conflicting classification between the structure-based and the ligand-based approach will not be identified, classifying them as inactive. The existence of additional biases in InteractionOCSVM models might shed a light on the results from Chapter 3. Most trained models were unable to identify all ligands with the same efficacy as the reference. Previous explanations of this behaviour were linked to either the docking algorithm not being able to generate a relevant pose in the selected receptor's conformation, or by considering the existence of alternative binding modes. Here we propose the hypothesis that small differences in binding mode, like the ones observed between salbutamol and formoterol, might generate models with significantly different interaction biases, with subtle changes in the ligand binding mode generating a different hit list. This feature could be seen as an advantage if the bias could be linked to a relevant property, or as a disadvantage, if specific pharmacological properties outside of efficacy or affinity should not be considered.

Chapter 6

Multistate modelling of GPCRs using AlphaFold

The bottleneck of structure-based drug design is the availability of experimental structures for the desired target. In recent years deep learning-based methods for protein structure prediction have revolutionized the filed of protein structure modelling. However these methods are severely limited by their inability to model different conformations of the same protein targets. This limit is especially relevant for G-protein-coupled receptors, since the receptors' plasticity is one of the key factors behind the differences in ligand efficacy.

The first section of this chapter is dedicated to describe the problem of modelling different protein conformations using AlphaFold, and the different methods developed to tackle this issue. To simplify the modelling of protein structure in different conformation a new input preparation and prediction pipeline called LIT-AlphaFold was developed, with a particular attention to methods developed for G-protein-coupled receptors.

The second section instead describes part of a benchmarking study aimed at assessing the quality of chemokine - chemokine receptor complexes predicted using AlphaFold-Multimer. Other than evaluating the quality of the generated structures, the effect of different prediction parameter on the conformational state of the receptor were investigated. This work represent the first description of methods for guided alternative conformation generation for multimeric targets.

The benchmarking study on chemokine - chemokine receptors is the outcome of a collaboration with Lauri Urvas, PhD student in the Laboratoire d'Innovation Therapeutique - Université de Strasbourg, who compared the available crystallographic structures with the structural models generated by AlphaFold and AlphaFold-Multimer.

6.1 LIT-AlphaFold: a new tool for multistate modelling

6.1.1 Scope and motivation

Since the release of AlphaFold[1, 2] multiple pipelines have been developed to leverage its predictive power to model multiple conformational states of protein targets[3]. The current state of the art methods achieve multistate modelling by modifying the input information extracted from genetic and template databases, or by acting on the prediction parameters. The neural network architecture and its weights are not directly modified; therefore, these different approaches could be combined in a single pipeline. Unfortunately, each method was developed independently from the others, making their comparison and use cumbersome. In an effort to simplify and streamline this process LIT-AlphaFold was developed. LIT-AlphaFold is a new input preparation and prediction pipeline for AlphaFold, which includes all necessary tools for multistate modelling in a single environment. The tools can be used to reproduce existing methods, or to help the development of new protocols.

The code of LIT-AlphaFold is available on GitHub at: `https://github.com/LIT-CCM-lab/LIT-AlphaFold`

6.1.2 State of the art

The advent of AlphaFold2 (AlphaFold)[1] and other deep learning models for protein structure prediction has transformed the field of molecular modelling. Thanks to these tools the structures of the entire proteome[4, 5] have become accessible, since good quality structural model can be built just by knowing the amino acid sequence of the target. AlphaFold-Multimer[2], based on the same architecture as AlphaFold but trained on multichain structures, has simplified the complex task of modelling protein-protein interfaces, for both large complexes and peptide docking. The structure of a protein complex can be studied just by knowing the amino acid sequences of its components and their stoichiometry. The success of AlphaFold is built on top of more than thirty years of research on the relationship between coevolution and structure in proteins[6]. Phylogenetic information is extracted from the analysis of a multiple sequence alignment (MSA)[7], algorithmically aligned homologous sequences extracted from genetic databases based on an initial query. By analysing conserved and mutated residues in the MSA, it is possible to identify pattern of residues that tend to mutate co-operatively, called correlated mutations[8]. The presence of correlated mutations indicates some form of relationship between two residues, since a negative point mutation on one is compensated by the mutation of the other. The link between coevolutionary signals in the MSA and spatial proximity in a protein's three dimensional structure was established based on the observation that correlated residue pairs which are far from each other in the protein sequence often corresponded to residue-residue contacts in experimental structures[9–11]. AlphaFold uses deep learning, through a multiheaded attention mechanism, to extract spatial relationships from the MSA built using the input sequence. Other than the MSA, homologous templates are used for prediction, however the entire process is mainly guided by sequence data. Of the five models forming AlphaFold only two utilize any template information. AlphaFold-Multimer instead employs template data to predict the structure of the individual monomeric units, while the interface between the protein chains is inferred only from coevolutionary data. While the

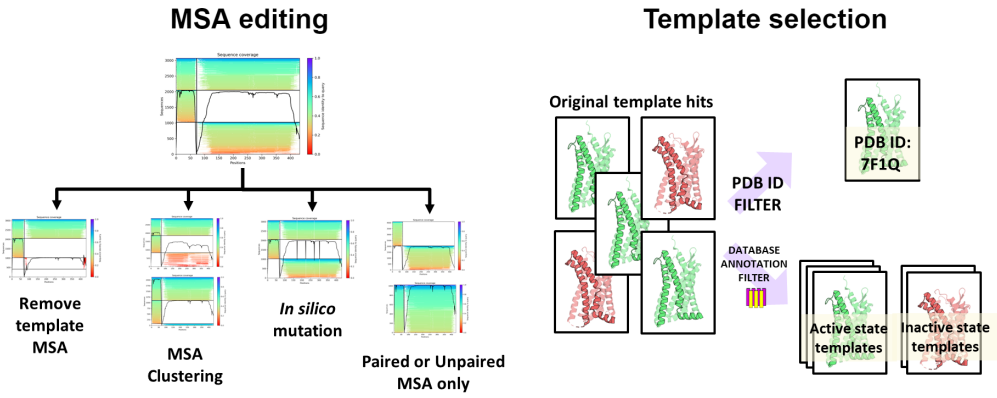
analysis of the MSA can be used to build an initial structural hypothesis, the strength of AlphaFold relays in its ability to refine this initial hypothesis using its predicted confidence scores. The predicted local distance difference test (plDDT), the predicted template modelling score (pTM), and the interface predicted template modelling score (ipTM) are based on state-of-the-art measures used for protein models evaluation. lDDT is a local quality measure, it represent the fraction of conserved interatomic distances within a threshold between a structural model and a reference structure[12]. The TM score instead is a global similarity measure between two protein structures[13]. The ipTM score was developed for AlphaFold-Multimer, it is based on the TM score but it is used to describe the global quality of protein-protein interfaces[2]. According to the analysis of Roney and Ovchinnikov[14] these confidence measures represent a learned "energy function" describing the "folding energy" of the predicted structure. The term "energy function" is used to describe a function whose optima correspond to the native structure of the target protein. This "energy" is independent from the MSA, or the templates used for calculation, and it is only a function of the structure. By combining the information extracted from the MSA with the learned "energy function", AlphaFold efficiently identifies a structure representing a native conformation of the target protein. Sequence data are used to build an initial hypothesis in the proximity of the "energy minimum", then the structure is iteratively refined by descending along the energy surface. This interpretation explains how AlphaFold is able to quickly and correctly identify the native fold between all possible protein conformations, and why MSA information plays such a crucial role for prediction, despite a protein's fold being determined only by its sequence[15]. Despite its remarkable achievements AlphaFold is unable to model different conformations of multistate proteins[16]. Proteins are flexible systems, whose energy landscape is often characterized by the presence of multiple local minima, with each representing a different stable state at physiological conditions[17]. Protein function is often intimately linked to the presence of multiple states separated by energy barriers, with shifts between conformations playing a crucial role in many biological processes[18]. Being able to effectively sample the conformational space of a protein is essential to understand the intricate relationship between structure and function. Based on the interpretation given by Roney and Ovchinnikov, this could be caused by either the "energy function" not recognizing one state as an energy minimum, or by the information extracted from the MSA only identifying one conformational state. Xie and Huang showed, by working on membrane transporters, that AlphaFold's "energy function" can present multiple local minima[19]. Membrane transporters assume two main conformations, an inward facing conformation and an outward facing conformation. For the same target, the two conformations have a similar AlphaFold combined confidence score. AlphaFold recognizes both the inward facing and outward facing conformations as stable states, therefore the learned "energy function" can reproduce the two energy minima of membrane transporters' conformational landscapes. Despite the "energy function" not being biased towards any conformation, the inward facing conformation was predicted more often than the outward-facing one. The limiting factor therefore are the residue-residue contacts inferred from correlated mutations. MSAs however are known to encode spatial information describing different protein confromations[20, 21], or structural rearrangements induced by ligand binding[22] or multimerization[23]. Information regarding alternative conformations might be available in the MSA, but AlphaFold seems unable to deconvolve these signals. Current evidence suggests that for each protein there is a preferred conformation which is observed in most of the predicted structures[16, 19]. The exact origin of this phenomena is unclear, Heo and Feig suggested an influence due to biases in the training set[16], while

Xie and Huang suggested an influence from the targets' biological function[19]. Regardless of its origin, it is possible to overrule this tendency by preprocessing the MSA before prediction[3], directly altering the encoded coevolutionary information. Before addressing the different methods used for multistate modelling, it is important to clearly define what exactly multistate modelling means. The problem of generating alternative conformations encompasses two distinct meaning, depending on whether one is studying a single chain target or a multichain target. For monomers, or single chains within a multimeric complex, the aim is to generate structures representing different arrangements of the target's domains. This may involve sampling the entire conformational landscape or generating structures representing a specific conformation of interest. For multichain targets instead the interest lies in generating structures showcasing different protein-protein interfaces. This is done either to identify the "correct" conformation of the multimer by sampling a larger portion of the conformational space, or to identify alternative interfaces associated with distinct biological functions[24]. Methods developed for single chain multistate modelling can be separated in two main groups: sampling methods, and state specific methods. Methods in the first group are used to explore the entire conformational space of a target, or to generate a structure in a specific state by unbiased sampling of the conformational space. For the second group of methods instead the aim is to reliably generate structures in a specific conformational state, without the need for sampling. af_conformations[25] was the first published method for conformational sampling using AlphaFold. While originally developed for membrane transporters and G-protein coupled receptors, it has also been applied to study other protein families[26–28]. Different conformations are generate through subsampling of the MSA, using AlphaFold's data preparation pipeline. Even at default settings the full MSA is not used directly for calculation, instead predictions are performed using two randomly selected subsets of sequences[1]. The sequences in the first group, which always includes the input sequence, represent the clusters centers. The remaining sequences are assigned to a cluster based on their distance from the centers. From the sequences not selected as cluster centers, a second subset is selected and labelled as extra sequences. At default settings the pipeline yields homogeneous predictions despite its stochastic nature due to the large number of sequences and extra sequences (512 and 1024 or 5120 for AlphaFold, 508 and 2048 or 1152 for AlphaFold-Multimer). The original implementation of af_conformations uses a significantly smaller number of sequences and extra sequences (16 and 32). Randomly picked shallow MSAs might encode different protein contacts than the ones extracted at default settings, therefore generating alternative target conformations. The full sampling of the conformational space is achieved by repeating calculations multiple times, selecting a different subset of sequences at each iteration. By applying this approach, it is possible to identify most of the stable states, as well as possible intermediate states between them. While this method was generally effective in generating diverse conformations, it struggled for targets with a single representative conformation in the training set. SPEACH_AF[29] achieves multistate modelling by "mutating" the MSA in specific regions. Since a residue-residue contact is inferred if a correlated mutation is observed between a pair of amino acids, if one position were to be mutated in all sequences to alanine, then the coevolutionary signal would be lost. By masking certain state specific contacts AlphaFold is forced to rely on other information to build the structure. Key sequence regions are identified through a sliding window mutagenesis scan, often guided by experimental data or by an initial AlphaFold run. Unlike af_conformations SPEACH_AF is deterministic, since a mutated MSA would always yield a specific conformation; however, it is not possible to establish a priori which

mutation will trigger a conformational change. AF-Cluster[30] uses clustering to achieve a shallower MSA, grouping together similar sequences. The objective is to deconvolve the different coevolutionary signals by using small subsets of similar sequences for prediction. Different protein conformations should be generated if each MSA cluster contains different contact information. While this method was successfully applied to correctly model both stable states of fold-switch proteins, it is not possible to identify a priori which cluster is associated to a specific conformation. The three described sampling methods[25, 29, 30] achieve multistate modelling by directly modifying the initial coevolutionary information encoded in the MSA. To avoid any bias from known structures, template information is removed during calculation, relaying only on sequence data. While this approach can be used to obtain structures in alternative conformations, it requires to perform multiple calculations which are not guaranteed to yield the desired outcome[19]. State specific methods instead use structural templates to stir the predictions towards a desired conformation. This approach was first pioneered by the DeepMind team during CASP14 to predict the structure of the target T1024[31]. The active transporter LmrP can assume an outward-facing or an inward-facing conformation, with AlphaFold only being able to model one of them. The alternative conformation was obtained using state specific templates, and by removing the MSA or by limiting it to the best thirty sequences. A previous attempt which used state specific templates without MSA editing did not yield the alternative conformation. Pairing state specific templates with dampened coevolutionary information allows to generate structures in a similar conformation as the templates. Two methods are based on this approach: AlphaFold-Multistate[16] and AF2_GPCR_kinase[32]. AlphaFold-Multistate was developed to overcome the bias of AlphaFold in GPCR structure prediction. Amino acid positions in the MSA covered by at least one template were replaced with a gap character, removing most of the information encoded in the MSA. This approach can be described as "homology modelling" through AlphaFold, since most of the MSA is masked, and the prediction is based on template information. AF2_GPCR_kinase instead combines template selection with the stochastic MSA subsampling of af_conformations. This approach weakens the coevolutionary signal without completely removing it. The method was benchmarked on GPCRs and kinases structures released after AlphaFold's training set cutoff date. AF2_GPCR_kinase reported similar predictions as AlphaFold-Multistate for rigid domains but produced better predictions for flexible loop regions. Regardless, both methods can be applied to generate state specific alternative conformation for a target protein. For multichain targets unfortunately no systematic study has been reported on the modelling of specific conformational states of their monomeric units. Heo and Feig used AlphaFold-Multimer to predict the complex between the peptide-agonist neuropeptide Y and the neuropeptide Y receptor type 1 (NPY1R), as part of GPCR Dock 2021[16]. An active state structure of NPY1R generated with AlphaFold-Multistate was used as template for AlphaFold-Multimer. The predicted structure correctly reproduced both the receptor activation state, and the peptide binding mode observed in the experimental reference structure[33]. Unlike for monomers, multistate modelling using AlphaFold-Multimer does not relay on subsampling or editing of the MSA, but instead uses dropout to randomly modify the neural network's weights. Dropout is a technique used for the training of deep-learning models, which consists in randomly ignoring (dropping) some of the nodes in the neural network[34]. By applying dropout during inference a model would generate a different output compared to the unperturbed results, with a minimal drop in accuracy[35]. Since dropout includes a stochastic component, it is generally paired with extended sampling, by repeating calculations multiple

times with different random seeds[36, 37]. AFsample[36, 38], the best performing method at CASP15 for multimer prediction, used six different dropout-based inference protocols, repeating calculations multiple times with a median number of 4810 predictions per target. Currently this approach has been used only to generate better multimer structures by increasing the size of the explored conformational space. No systematic study on the ability of AlphaFold-Multimer to model different protein-protein interfaces has been published. The DockQ[39] score evaluates the quality of predicted protein-protein interfaces, ranking them from incorrect, when the generated interface does not correspond to the reference, to high quality, when the model reproduces all aspects of the reference. The analysis of the results from CASP15 highlighted multiple instances of high confidence structures describing an incorrect interface, and high-quality structures with high confidence scores[36, 37]. While the accuracy of the "energy function" for AlphaFold-Multimers has not been studied, one might assume that it is characterized by the presence of multiple minima. Unfortunately, it is unknown if these alternative protein-protein interfaces just represent computational artefacts, or a proof that AlphaFold-Multimer can predict biologically relevant alternative conformations of protein-protein complexes.

6.1.3 The LIT-AlphaFold pipeline



LIT-AlphaFold is an input preparation and prediction pipeline for AlphaFold and AlphaFold-Multimer developed to simplify the access to multistate modelling. Tools for template selection, MSA editing, as well as an advanced access to prediction parameters are all implemented in a single environment, with their use extended to AlphaFold-Multimer. LIT-AlphaFold is based on two important software contributions from the research community: AlphaPulldown[40] and ColabFold[41]. AlphaPulldown is a python package developed for efficient protein-protein interaction screening using AlphaFold-Multimer. The end-to-end calculation of AlphaFold is split between a CPU-based and GPU-based phase. The CPU is used to query genetic databases to retrieve the MSA, and to identify structural templates, while predictions are performed using GPU. In AlphaPulldown the sequence and template information of a single protomer are stored in a python object (MonomericObject), which can be saved for storage and future use. At inference time a single MonomericObject can be used for prediction using AlphaFold, or multiple monomers can be combined to create an input for AlphaFold-Multimer (MultimericObject). This approach was developed to optimize protein-protein interaction screening, since the same monomeric unit can be reused in different complexes. ColabFold is an efficient online tool

to perform predictions using AlphaFold. The MSA is generated using MMseqs2[42, 43], which significantly speeds up calculations compared to the default pipeline. Despite its user-friendly interface ColabFold allows to modify multiple input preparation and prediction parameters, like the number of sequences and extra-sequences, enabling dropout during inference, or using user-generated templates. LIT-AlphaFold uses ColabFold to easily gain access to its parameter customization features, to its use of the MMseqs2 webserver for MSA generation, its integration of a user generated template database, as well as the additional functionalities developed for data visualization (plDDT plot, PAE plot, 3D structure visualization, etc.). The LIT-AlphaFold pipeline, like AlphaPulldown, is split in two main stages, input preparation (Figure 6.1.1) and structure prediction (Figure 6.1.2). The first stage can be further separated in two main steps "data gathering", and "data editing". During the data gathering step the MSA is generated, and homologous templates are identified. Depending on the user preferences either the default AlphaFold pipeline or the MMseqs2 webserver can be used. The data are further refined during the "data editing" step by selecting templates based on their PDB ID, or according to their structural annotations in GPCRdb[44]. The processed data from the MSA and the templates is stored in a compressed pickle file, corresponding to the MonomericObject from AlphaPulldown. While not implemented directly in the main pipeline, MSA clustering can be performed using an additional script ($LIT-AF_Cluster.ipynb$). The structure prediction stage is formed by a data preparation step, followed by the prediction step. During data preparation further editing of the MSA can be applied (remove template MSA, MSA mutations, template shuffling, etc.). For multimeric targets the data from the single monomeric units are combined to form the input for AlphaFold-Multimer. The processed inputs are fed to AlphaFold and AlphaFold-Multimer for structure prediction, with the user defined prediction parameters. The prediction step can be automatically repeated multiple times to achieve a better sampling of the conformational space.

All necessary information about a protein target is stored as a LIT-AlphaFold MonomericObject, like with AlphaPulldown. Compared to their original implementation, LIT-AlphaFold includes additional object functions to edit both MSA and template information:

- `remove_template_from_msa`, this tool removes MSA positions covered by at least one of the available templates[16].
- `remove_msa_features`, the MSA of the single monomer is removed.
- `mutate_msa`, perform in silico mutation of the entire MSA in a specific position[29]. All non-gap residues can be replaced with the desired amino acid. Mutations can be performed on both the paired and unpaired MSA.
- `remove_templates`, replace template information with mock template information[41].
- `remove_msa_region`, mutate the MSA in multiple position to a gap residue.
- `cluster_msa`, perform density-based clustering[45, 46], of the MSA, using the given ϵ and minimum number of samples[30].
- `shuffle_templates`, randomly shuffle the templates identified by the pipeline to change which will be selected during prediction[32].

- `filter_templates`, filter the identified homologous templates based on their PDB ID. For GPCR and G-protein structures filtering can be performed using the structural annotations from GPCRdb[32, 44].

Figure 6.1.1: Input preparation pipeline of LIT-AlphaFold. The script `create_individual_features.py` searches for the relevant data in genetic and template databases. The Jupyter Notebook `LIT-AlphaFold_Cluster.ipynb` instead is used to cluster the MSA as described by [30].

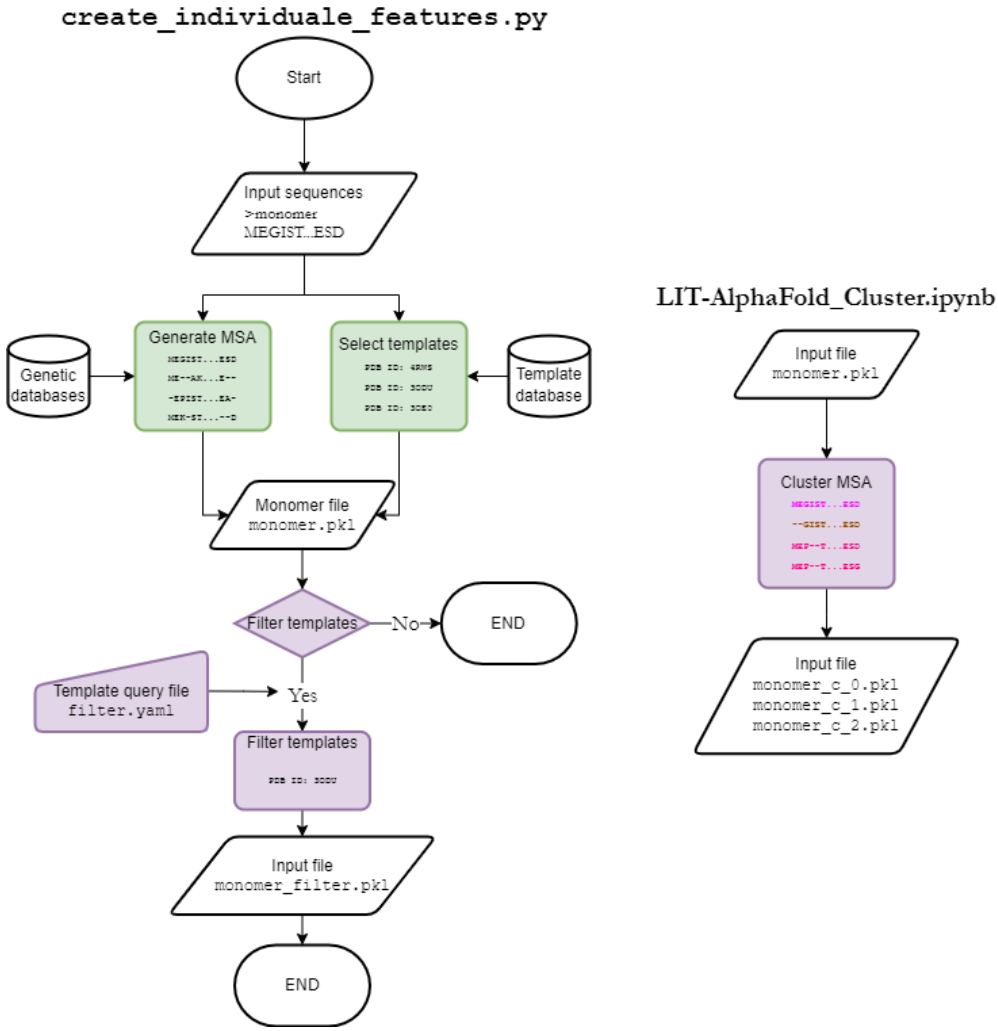
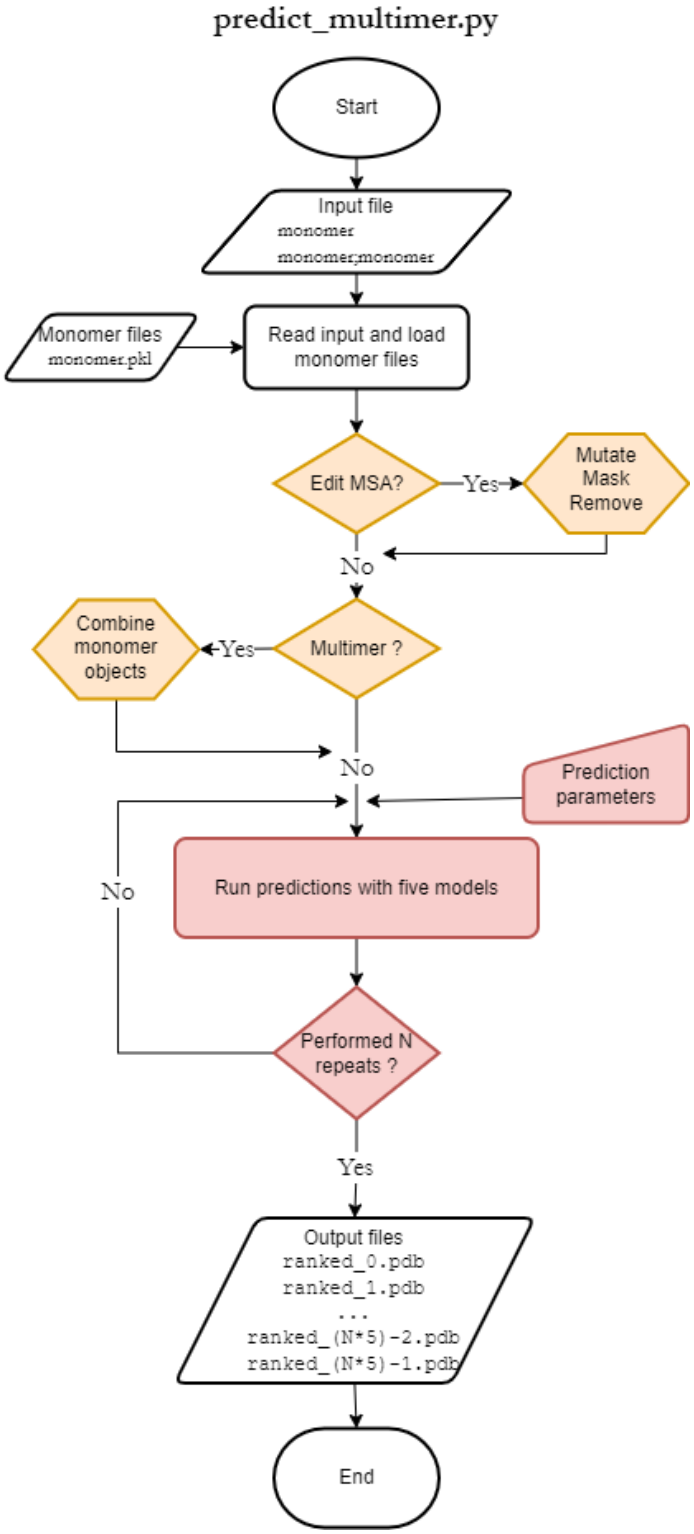


Figure 6.1.2: Structure prediction pipeline of LIT-AlphaFold. The prediction parameters include the number of sequences to use for calculation, the number of time calculation should be repeated with a different random seed (N), the use of dropout, the number of recycles, ecc.



References

[1] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, August 2021.

[2] Richard Evans, Michael O'Neill, Alexander Pritzel, Natasha Antropova, Andrew Senior, Tim Green, Augustin Žídek, Russ Bates, Sam Blackwell, Jason Yim, Olaf Ronneberger, Sebastian Bodenstein, Michal Zielinski, Alex Bridgland, Anna Potapenko, Andrew Cowie, Kathryn Tunyasuvunakool, Rishub Jain, Ellen Clancy, Pushmeet Kohli, John Jumper, and Demis Hassabis. Protein complex prediction with AlphaFold-Multimer, March 2022.

[3] D. Sala, F. Engelberger, H. S. Mchaourab, and J. Meiler. Modeling conformational states of proteins with AlphaFold. *Current Opinion in Structural Biology*, 81:102645, August 2023.

[4] Mihaly Varadi, Stephen Anyango, Mandar Deshpande, Sreenath Nair, Cindy Natassia, Galabina Yordanova, David Yuan, Oana Stroe, Gemma Wood, Agata Laydon, Augustin Žídek, Tim Green, Kathryn Tunyasuvunakool, Stig Petersen, John Jumper, Ellen Clancy, Richard Green, Ankur Vora, Mira Lutfi, Michael Figurnov, Andrew Cowie, Nicole Hobbs, Pushmeet Kohli, Gerard Kleywegt, Ewan Birney, Demis Hassabis, and Sameer Velankar. AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, 50(D1):D439–D444, January 2022.

[5] Kathryn Tunyasuvunakool, Jonas Adler, Zachary Wu, Tim Green, Michal Zielinski, Augustin Žídek, Alex Bridgland, Andrew Cowie, Clemens Meyer, Agata Laydon, Sameer Velankar, Gerard J. Kleywegt, Alex Bateman, Richard Evans, Alexander Pritzel, Michael Figurnov, Olaf Ronneberger, Russ Bates, Simon A. A. Kohl, Anna Potapenko, Andrew J. Ballard, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Ellen Clancy, David Reiman, Stig Petersen, Andrew W. Senior, Koray Kavukcuoglu, Ewan Birney, Pushmeet Kohli, John Jumper, and Demis Hassabis. Highly accurate protein structure prediction for the human proteome. *Nature*, 596(7873):590–596, August 2021.

[6] David de Juan, Florencio Pazos, and Alfonso Valencia. Emerging methods in protein co-evolution. *Nature Reviews Genetics*, 14(4):249–261, April 2013.

[7] Maria Chatzou, Cedrik Magis, Jia-Ming Chang, Carsten Kemena, Giovanni Bussotti, Ionas Erb, and Cedric Notredame. Multiple sequence alignment modeling: Methods and applications. *Briefings in Bioinformatics*, 17(6):1009–1023, November 2016.

[8] Florencio Pazos, Manuela Helmer-Citterich, Gabriele Ausiello, and Alfonso Valencia. Correlated mutations contain information about protein-protein interaction 1. *Journal of Molecular Biology*, 271(4):511–523, August 1997.

[9] D. Altschuh, A. M. Lesk, A. C. Bloomer, and A. Klug. Correlation of co-ordinated amino acid substitutions with function in viruses related to tobacco mosaic virus. *Journal of Molecular Biology*, 193(4):693–707, February 1987.

[10] I.N. Shindyalov, N.A. Kolchanov, and C. Sander. Can three-dimensional contacts in protein structures be predicted by analysis of correlated mutations? *Protein Engineering, Design and Selection*, 7(3):349–358, March 1994.

[11] Ulrike Göbel, Chris Sander, Reinhard Schneider, and Alfonso Valencia. Correlated mutations and residue contacts in proteins. *Proteins: Structure, Function, and Bioinformatics*, 18(4):309–317, 1994.

[12] Valerio Mariani, Marco Biasini, Alessandro Barbato, and Torsten Schwede. lDDT: A local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics*, 29(21):2722–2728, November 2013.

[13] Yang Zhang and Jeffrey Skolnick. Scoring function for automated assessment of protein structure template quality. *Proteins: Structure, Function, and Bioinformatics*, 57(4):702–710, 2004.

[14] James P. Roney and Sergey Ovchinnikov. State-of-the-Art Estimation of Protein Model Accuracy Using AlphaFold. *Physical Review Letters*, 129(23):238101, November 2022.

[15] Christian B. Anfinsen. Principles that Govern the Folding of Protein Chains. *Science*, 181(4096):223–230, July 1973.

[16] Lim Heo and Michael Feig. Multi-state modeling of G-protein coupled receptors at experimental accuracy. *Proteins: Structure, Function, and Bioinformatics*, 90(11):1873–1885, 2022.

[17] Katherine Henzler-Wildman and Dorothee Kern. Dynamic personalities of proteins. *Nature*, 450(7172):964–972, December 2007.

[18] Arvind Ramanathan, Andrej Savol, Virginia Burger, Chakra S. Chennubhotla, and Pratul K. Agarwal. Protein Conformational Populations and Functionally Relevant Substates. *Accounts of Chemical Research*, 47(1):149–156, January 2014.

[19] Tengyu Xie and Jing Huang. Can Protein Structure Prediction Methods Capture Alternative Conformations of Membrane Transporters? *Journal of Chemical Information and Modeling*, 64(8):3524–3536, April 2024.

[20] Thomas A. Hopf, Lucy J. Colwell, Robert Sheridan, Burkhard Rost, Chris Sander, and Debora S. Marks. Three-Dimensional Structures of Membrane Proteins from Genomic Sequencing. *Cell*, 149(7):1607–1621, June 2012.

[21] Ludovico Sutto, Simone Marsili, Alfonso Valencia, and Francesco Luigi Gervasio. From residue coevolution to protein conformational ensembles and functional dynamics. *Proceedings of the National Academy of Sciences*, 112(44):13567–13572, November 2015.

[22] Faruck Morcos, Biman Jana, Terence Hwa, and José N. Onuchic. Coevolutionary signals across protein lineages help capture multiple protein conformations. *Proceedings of the National Academy of Sciences*, 110(51):20533–20538, December 2013.

[23] Guido Uguzzoni, Shalini John Lovis, Francesco Oteri, Alexander Schug, Hendrik Szurmant, and Martin Weigt. Large-scale identification of coevolution signals across homo-oligomeric protein interfaces by direct coupling analysis. *Proceedings of the National Academy of Sciences*, 114(13):E2662–E2671, March 2017.

[24] Jun Jin, Fanny Momboisse, Gaelle Boncompain, Florian Koensgen, Zhicheng Zhou, Nelia Cordeiro, Fernando Arenzana-Seisdedos, Franck Perez, Bernard Lagane, Esther Kellenberger, and Anne Brelot. CCR5 adopts three homodimeric conformations that control cell surface delivery. *Science Signaling*, 11(529):eaal2869, May 2018.

[25] Diego del Alamo, Davide Sala, Hassane S Mchaourab, and Jens Meiler. Sampling alternative conformational states of transporters and receptors with AlphaFold2. *eLife*, 11:e75751, March 2022.

[26] Bodhi P. Vani, Akashnathan Aranganathan, Dedi Wang, and Pratyush Tiwary. AlphaFold2-RAVE: From Sequence to Boltzmann Ranking. *Journal of Chemical Theory and Computation*, 19(14):4351–4354, July 2023.

[27] Artur Meller, Soumendranath Bhakat, Shahlo Solieva, and Gregory R. Bowman. Accelerating Cryptic Pocket Discovery Using AlphaFold. *Journal of Chemical Theory and Computation*, 19(14):4355–4363, July 2023.

[28] Gabriel Monteiro da Silva, Jennifer Y. Cui, David C. Dalgarno, George P. Lisi, and Brenda M. Rubenstein. High-throughput prediction of protein conformational distributions with subsampled AlphaFold2. *Nature Communications*, 15(1):2464, March 2024.

[29] Richard A. Stein and Hassane S. Mchaourab. SPEACH_AF: Sampling protein ensembles and conformational heterogeneity with Alphafold2. *PLOS Computational Biology*, 18(8):e1010483, August 2022.

[30] Hannah K. Wayment-Steele, Adedolapo Ojoawo, Renee Otten, Julia M. Apitz, Warintra Pitsawong, Marc Hömberger, Sergey Ovchinnikov, Lucy Colwell, and Dorothee Kern. Predicting multiple conformations via sequence clustering and AlphaFold2. *Nature*, 625(7996):832–839, January 2024.

[31] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Applying and improving AlphaFold at CASP14. *Proteins: Structure, Function, and Bioinformatics*, 89(12):1711–1721, 2021.

[32] Davide Sala, Peter W. Hildebrand, and Jens Meiler. Biasing AlphaFold2 to predict GPCRs and kinases with user-defined functional or structural properties. *Frontiers in Molecular Biosciences*, 10, 2023.

[33] Chaehee Park, Jinuk Kim, Seung-Bum Ko, Yeol Kyo Choi, Hyeongseop Jeong, Hyeonuk Woo, Hyunook Kang, Injin Bang, Sang Ah Kim, Tae-Young Yoon, Chaok Seok, Wonpil Im, and Hee-Jung Choi. Structural basis of neuropeptide Y signaling through Y1 receptor. *Nature Communications*, 13(1):853, February 2022.

[34] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958, 2014.

[35] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pages 6405–6416, Red Hook, NY, USA, December 2017. Curran Associates Inc.

[36] Björn Wallner. Improved multimer prediction using massive sampling with AlphaFold in CASP15. *Proteins: Structure, Function, and Bioinformatics*, 91(12):1734–1746, 2023.

[37] Isak Johansson-Åkhe and Björn Wallner. Improving peptide-protein docking with AlphaFold-Multimer using forced sampling. *Frontiers in Bioinformatics*, 2, 2022.

[38] Björn Wallner. AFsample: Improving multimer prediction with AlphaFold using massive sampling. *Bioinformatics*, 39(9):btad573, September 2023.

[39] Sankar Basu and Björn Wallner. DockQ: A Quality Measure for Protein-Protein Docking Models. *PLOS ONE*, 11(8):e0161879, August 2016.

[40] Dingquan Yu, Grzegorz Chojnowski, Maria Rosenthal, and Jan Kosinski. AlphaPulldown—a python package for protein–protein interaction screens using AlphaFold-Multimer. *Bioinformatics*, 39(1):btac749, January 2023.

[41] Milot Mirdita, Konstantin Schütze, Yoshitaka Moriwaki, Lim Heo, Sergey Ovchinnikov, and Martin Steinegger. ColabFold: Making protein folding accessible to all. *Nature Methods*, 19(6):679–682, June 2022.

[42] Milot Mirdita, Martin Steinegger, and Johannes Söding. MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinformatics*, 35(16):2856–2858, August 2019.

[43] Martin Steinegger and Johannes Söding. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, 35(11):1026–1028, November 2017.

[44] Gáspár Pándy-Szekeres, Jimmy Caroli, Alibek Mamyrbekov, Ali A Kermani, György M Keserű, Albert J Kooistra, and David E Gloriam. GPCRdb in 2023: State-specific structure models using AlphaFold2 and new ligand resources. *Nucleic Acids Research*, page gkac1013, November 2022.

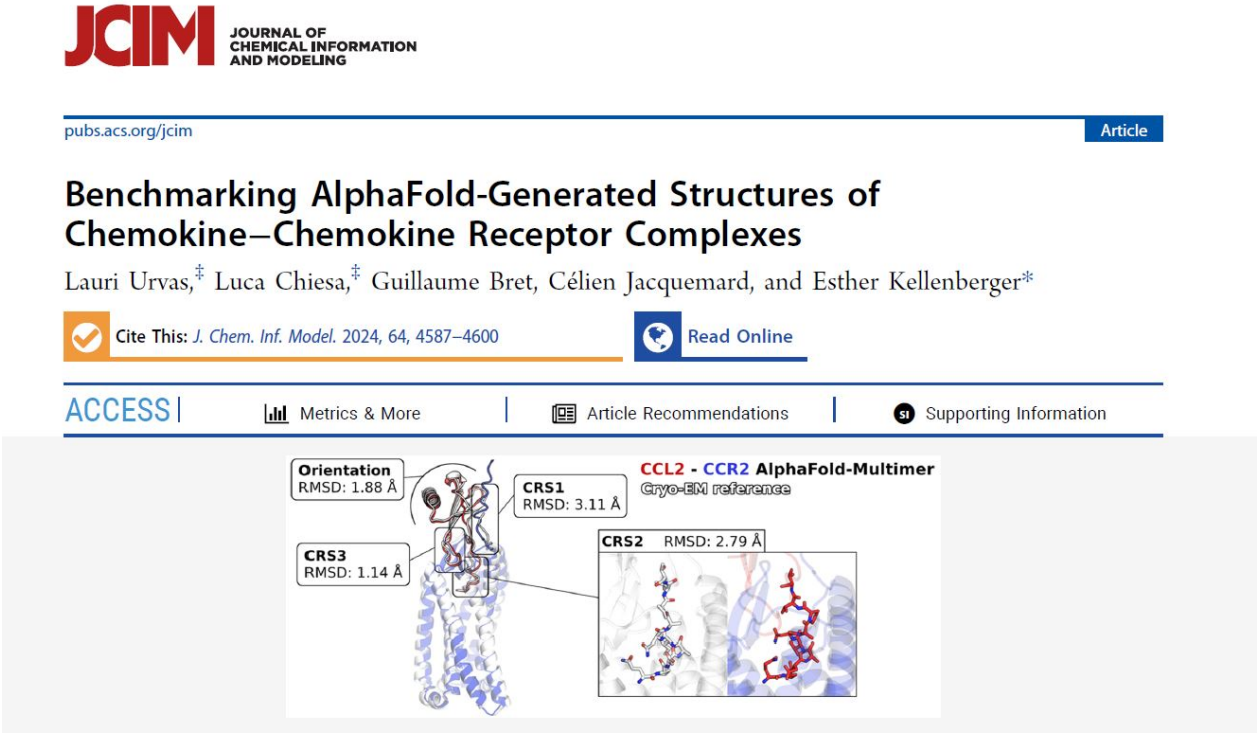
[45] Erich Schubert, Jörg Sander, Martin Ester, Hans Peter Kriegel, and Xiaowei Xu. DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN. *ACM Transactions on Database Systems*, 42(3):19:1–19:21, July 2017.

[46] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD'96, pages 226–231, Portland, Oregon, August 1996. AAAI Press.

6.2 Benchmarking AlphaFold-generated structures of chemokine – chemokine receptor complexes

This section is adapted from the article "Benchmarking AlphaFold-generated structures of chemokine – chemokine receptor complexes" published in the Journal of Chemical Information and Modelling (*J. Chem. Inf. Model.* 64, 11, 4587–4600 (2024)). The content of the article has been modified to fit the topics of this thesis, and to isolate as much as possible the two main contributing authors. The analysis of the generated models, as well as the bibliographical work regarding chemokine - chemokine receptor complexes were performed by Lauri Urvas.

The full text of the article as well as the Supporting Information are available at: `https://doi.org/10.1021/acs.jcim.3c01835`



Abstract

AlphaFold and AlphaFold-Multimer have become two essential tools for the modelling of unknown structures of proteins and protein complexes. In this work, we extensively benchmarked the quality of chemokine – chemokine receptor structures generated by AlphaFold-Multimer against experimentally determined structures. Our analysis considered both the global quality of the model, as well as key structural features for chemokine recognition. To study the effects of template and multiple sequence alignment parameters on the results, a new prediction pipeline called LIT-AlphaFold (`https://github.com/LIT-CCM-lab/LIT-AlphaFold`) was developed, allowing extensive input customization. The accuracy of chemokine N-terminus binding mode prediction varied between complexes, but the confidence score permitted the distinguishing of residues that were very likely well positioned.

6.2.1 Introduction

Chemokine system

Chemokine receptors (CKRs) play a central role in guiding the directional migration of immune cells upon stimulation by small extracellular proteins known as chemokines. Approximately 50 chemokines have been identified to date, and they are divided in four subfamilies according to the arrangement of conserved cysteine residues in their N-terminal region [1, 2]. There are two major subfamilies CC and CXC, and two minor subfamilies CX3C and XC. A total of 23 CKRs have been identified, and they include typical CKRs, atypical CKRs (ACKRs) and viral CKRs. The typical CKRs are further categorized as CC/CXC/CX3C/XC receptors according to the subfamily of their endogenous chemokine ligands. ACKRs lack G protein signaling, and they can bind to chemokines from one or multiple subfamilies.

CKRs offer promising prospects for drug development due to their involvement in a range of diseases, such as inflammatory diseases and cancer [3–6]. Details on the structures of these receptors and the interactions that take place upon chemokine binding are of great utility for rational design of therapeutics. Such knowledge has started to emerge through advanced techniques like X-ray diffraction and cryo-electron microscopy (cryo-EM), which have enabled the visualization of the complexes formed by chemokines and their corresponding receptors. To date, at least 22 experimental structures of chemokine – CKR complexes have been solved, an achievement that reflects significant effort[7–19]. These structures involve 15 different chemokines and nine different CKRs. However, due to the vastness of the chemokine system, there remains much more to uncover.

Chemokine – chemokine receptor interactions

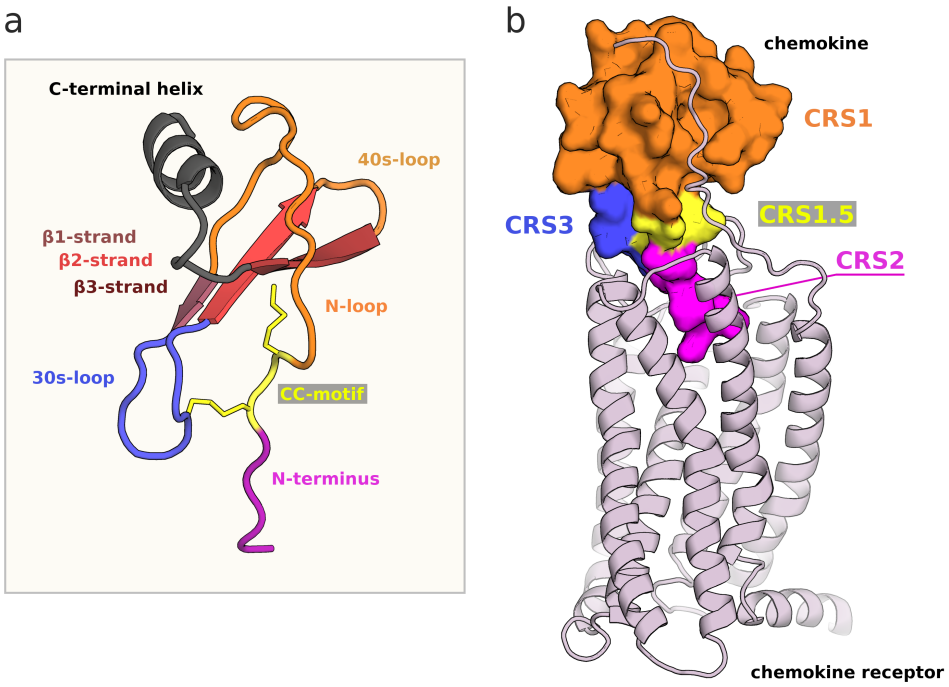
Chemokines have a common tertiary structure, featuring a flexible N-terminus, an N-loop, a core region defined by three anti-parallel β-strands interconnected by two loops called 30s- and 40s-loops, and a C-terminal α-helix (Figure 1a). The cysteine motif that defines the chemokine subfamilies is located in the N-loop, which is thereby anchored to the chemokine core via disulfide bridges to the 30s-loop and the β3-strand.

CKRs are part of the rhodopsin-like class A group of G protein-coupled receptors (GPCRs). These receptors feature the signature seven transmembrane (7TM) structure, which includes three intracellular and three extracellular loops, known as ICLs 1-3 and ECLs 1-3 The binding cavity for the chemokine N-terminus is located on the extracellular side of CKRs between the transmembrane helices.

The typical chemokine – CKR interface exhibits multifaceted interactions that are commonly divided into several chemokine recognition sites (CRSs) (Figure 6.2.1.b). The most prominent sites are CRS1 and CRS2, involving the N-terminal domains of the receptor and the chemokine, respectively. Several review articles discuss the chemokine – CKR interactions [20–25].

In CRS1, the receptor N-terminus binds to a groove formed between the N-loop and the 40s-loop of the chemokine. In CRS2, the chemokine N-terminus inserts into the 7TM cavity of the receptor and promotes receptor activation by favoring activating conformational changes in binding site residues.

Figure 6.2.1: a) Tertiary structure of chemokines, shown with CXCL8 (PDB: 6LFO). b) Composition of the typical chemokine-chemokine receptor complex, with chemokine recognition sites CRS1 (orange), CRS1.5 (yellow), CRS2 (magenta), and CRS3 (blue) highlighted, shown with the CXCL8-CXCR2 complex (PDB: 6LFO).



AlphaFold and AlphaFold-Multimer

AlphaFold (AlphaFold v2.0)[26] is a deep learning tool for protein structure prediction which can reliably reach near experimental accuracy[27]. AlphaFold generates structures by leveraging both coevolutionary information extracted from multiple sequence alignments (MSA) and structural information from experimental templates. This information is processed by an attention-based mechanism, called the Evoformer, which extracts spatial relationships from the MSA and the templates, while the final 3D representation is generated by the Structure module.

Based on the same architecture, AlphaFold-Multimer[28] was trained to predict multimeric protein structures, simplifying multichain target modelling[29, 30] and peptide docking[31]. Coevolutionary information for multimers is extracted from the MSAs of the single protomers (unpaired MSA), and the MSA obtained by merging sequences from different chains based on species annotation (paired MSA).

While representing a great achievement for protein structure determination, some key limitations are still present. Proteins are flexible systems and can assume multiple conformational states, which AlphaFold is not able to fully reproduce. GPCRs are characterized by their plasticity, assuming active and inactive state conformations. While AlphaFold is not able to reliably model a specific activation state, different approaches have been developed to increase the size of the explored conformational space. More diverse structures have been obtained by subsampling[32, 33], mutating[34] or masking[35] the MSA, while by selecting templates with desired features it was possible to generate structures with a specific conformation[33, 35]. Based on these two principles different multistate modelling protocols have been developed[36] AlphaFold-Multistate[35], af2_conformations[32], SPEACH-AF[34], AF2_GPCR_Kinase[33], AF-Cluster[37]. Such methods, however, were developed for monomers, and they have yet to be applied to multimers. The exploration of a larger conformational space for multichain targets has been achieved using extended sampling, by repeating calculations multiple times with different random seeds, and by enabling dropout during prediction. Dropout is a technique used for the training of deep learning models; it consists of randomly ignoring certain nodes in the neural network during training. By enabling dropout during inference, AlphaFold-Multimer generates more diverse conformations of the target complex, leading to better results compared to the default prediction protocol[31, 38, 39].

In this work, we developed LIT-AlphaFold, a new pipeline for multistate modelling. It includes multiple tools for MSA editing (subsampling, clustering, mutation, masking) and template selection, and extends their use to AlphaFold-Multimer predictions. The pipeline was used to generate structure of known chemokine – CKR complexes.

Objectives

This work has two main objectives: benchmark the quality of chemokine - CKR structures predicted by AlphaFold–Multimer against experimentally resolved references, and determine if LIT-AlphaFold could generate chemokine - CKR structures in different receptor activation states. Our results showed that AlphaFold-Multimer can generate accurate models of chemokine – CKR complexes, and that multistate modelling of GPCR – protein ligand complexes could be achieved by using template selection, MSA editing, and dropout.

Table 6.2.1: Input preparation and model generation settings used with LIT-Alphafold.

protocol	default	1	2	3	4	5	6	7	8
5 repeats		×	×			×	×	×	×
dropout			×					×	×
state-specific templates				×	×	×	×	×	×
no unpaired MSA					×		×		×

6.2.2 Methods

AlphaFold model generation

The input sequences for the wild-type receptors and chemokines were obtained from UniProt[40], while the sequences of engineered chemokines and the mutated sequences used for solving the experimental complexes, were obtained from the Protein Data Bank (PDB). The benchmarking was carried out with wild-type models. Additionally, we generated models using the mutated sequences and compared them to the wild-type models and experimental structures to examine the impact of mutations in the input sequences.

For each chemokine and CKR, the input MSA and templates were identified using the default AlphaFold databases. Templates were filtered to remove all PDB entries describing chemokine – CKR complexes. Structures of the target receptors that were not in complex with a chemokine were kept.

Multimer model generation settings Calculations for chemokine – CKR complexes were performed using LIT-AlphaFold (github.com/LIT-CCM-lab/LIT-AlphaFold), a modified input generation and structure prediction pipeline based on AlphaPulldown[41] and Colabfold[38], integrating many features developed for multistate modelling.

We generated models using different combinations of MSA editing, template selection, and inference parameters to determine their effects on the prediction quality of chemokine – CKR interactions, as well as the activation state of the receptor model. Structure templates were selected with LIT-AlphaFold based on their activation state, according to their GPCRdb annotation[42, 43]. For the MSA, we used the option of removing the unpaired MSA in combination with state-specific templates aiming to bias the receptor structure towards a desired activation state. Calculations were run one and five times for each chemokine – CKR pair, with and without dropout, to determine the effect of extended sampling on the predictions. Combining these different options provided nine different protocols to generate chemokine – CKR complex models (Table 6.2.1). All calculations were performed using AlphaFold v2.3.0. models' weights.

Generation of receptor monomer structures The monomer structures of the nine studied receptors were determined using the same input files (all templates, active state templates, inactive templates) used for multimers. Different prediction protocols were used: AlphaFold v2.0, AlphaFold v2.0 without templates, AlphaFold-multistate[35], AF2_GPCR_Kinase[33], AlphaFold v2.0 with state-specific templates, and AlphaFold v2.0 using state-specific templates and without MSA. AlphaFold v2.0 is formed by five models

with similar architecture but different weights. Each model is used to generate one structure of the target protein, the structure with the highest confidence score is selected as the output of AlphaFold. Of these five models only two use template information at inference time, therefore for prediction protocols including template selection in their pipeline, calculations were performed using only the two models. Prediction protocols not based on specific template information used all five models to predict the protein structure.

Preparation of reference structures

The reference structure files were prepared from the publicly available structures in the PDB according to the following steps: chains not containing the receptor or chemokine were removed. Non-receptor or non-chemokine residues in the remaining chains were removed, which included auxiliary proteins like T4-Lysozyme or rubredoxin, and solvent and membrane atoms. When required, the chemokine and the receptor were separated into distinct chains. Residues were renumbered in accordance with UniProt residue numbering. For chemokines, we used residue numbering starting after the signal peptide.

Model evaluation

The chemokine – CKR complex models generated for benchmarking were compared to their corresponding reference structures. The DockQ score was computed for the top-ranking models that were predicted using nine different protocols (Table 6.2.1). In the case of protocols using state-specific templates, DockQ was calculated only for models built using templates in the same activation state as the reference. The analysis of the activation state was performed on all structures generated after repeating calculations five times with and without dropout, using different templates (all templates, active state, inactive state), and MSA information (entire MSA, only paired MSA) (protocols 1,2,5-8, Table 6.2.1).

DockQ score DockQ is a score used to evaluate the quality of protein-protein docking predictions against reference structures[44]. It is based on three measurements: the fraction of native contacts (Fnat), the Ligand Root Mean Square deviation (LRMS) and the interface Root Mean Square deviation (iRMS). DockQ combines these into a continuous score between 0 and 1, facilitating the benchmarking of docking algorithms and development of scoring functions.

Analysis of the receptor activation state We classified the activation state of the predicted structures based on the distances between TM6 and TM2 (positions 6.37 and 2.46), and TM7 and TM3 (positions 7.53 and 3.46) (positions denoted by the Ballesteros-Weinstein numbering scheme[45]). According to activation-state labels of experimental structures of CKRs in the GPCRdb, we defined that a structure is active if it presented a TM6-TM2 distance between 16.66 Å and 19.01 Å, and a TM7-TM3 distance between 8.42 Å and 10.40 Å. A structure was defined as inactive if it presented a TM6-TM2 distance between 11.18 Å and 12.99 Å, and a TM7-TM3 distance between 13.57 Å and 16.79 Å. Structures with distances outside the original threshold but within 1 Å were classified as active-like or inactive-like. Structures not fitting into any of these four categories were labelled as intermediate.

Table 6.2.2: Dataset of reference chemokine – CKR structures

Subfamily	Receptor	Ligand	Technique	Resolution (Å)	PDB ID	Date	Receptor template	PCA	Reference
CC	CCR1	CCL15	cryo-EM	2.6	7VL9	1.3.2022	×		[13]
	CCR2	CCL2	cryo-EM	2.9	7XA3	15.5.2022	×		[14]
	CCR5	CCL3	cryo-EM	2.9	7F1Q	6.7.2021	×		[18]
	CCR5	CCL5	cryo-EM	3.0	7F1R	6.7.2021	×		[18]
	CCR6	CCL20	cryo-EM	3.3	6WWZ	15.6.2020			[16]
	CCR5	[5P7]-CCL5[a]	X-ray	2.2	5UIW	20.6.2017	×	×	[19]
	CCR5	[6P4]-CCL5[a]	cryo-EM	3.1	7O7F	16.6.2021	×	×	[8]
CXC	CXCR2	CXCL8	cryo-EM	3.4	6LFO	1.9.2020	×		[9]
	CXCR4	vCCL2b	X-ray	3.1	4RWS	6.3.2015	×		[12]
CX3C	CX3CR1	CX3CL1	cryo-EM	3.4	7XBX	29.6.2022	×	×	[10]
ACKR	ACRK3	CXCL12	cryo-EM	3.8	7SK3	13.7.2022	×		[17]
	ACRK3	$CXCL12^{a}_{LRHQ}$	cryo-EM	3.3	7SK4	13.7.2022	×		[17]
vCKR	US28[b]	CX3CL1	X-ray	2.9	4XT1	6.3.2015	×	×	[7]
	US28[b]	CX3CL1.35[a]	X-ray	3.5	5WB2	8.6.2018	×		[11]

[a]Engineered chemokine. [b]Viral chemokine or CKR. Receptor template: A structure of the receptor that is not in complex with a chemokine is available in the PDB. PCA: pyroglutamic acid is the N-terminal residue of the chemokine. Bolded dates signify structures released before the training-set cutoff date of AlphaFold v2.3.0.

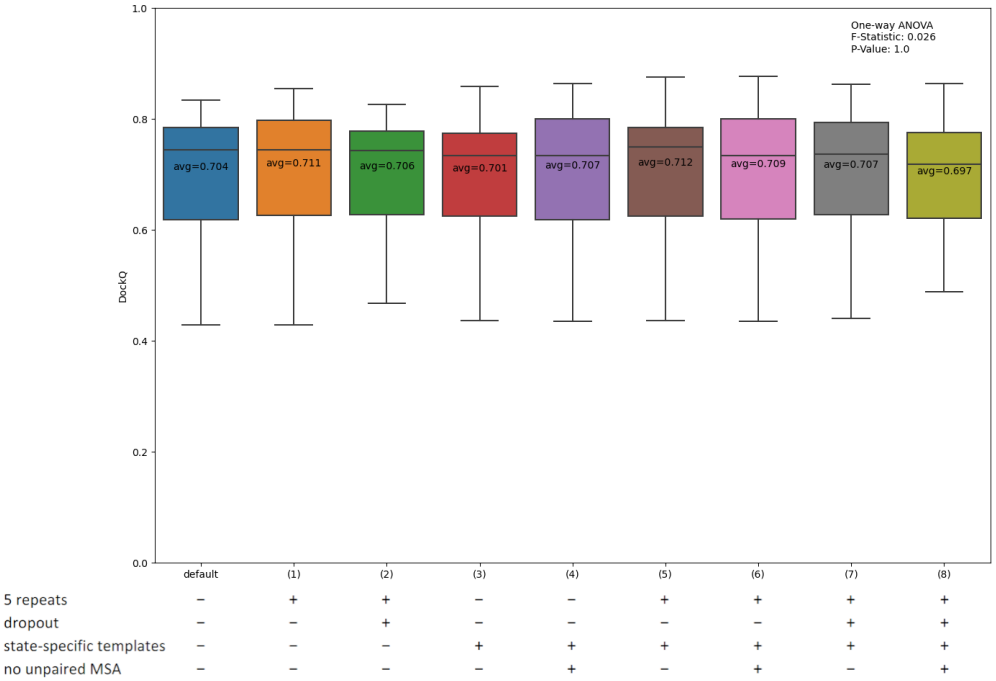
6.2.3 Results

The reference dataset of chemokine – chemokine receptor complexes

As the goal of the study was to predict the structures of experimentally solved chemokine – CKR complexes, we curated a reference dataset comprising structures of 14 unique chemokine – CKR pairs (Table 6.2.2). These complexes included CKRs and chemokines from various subfamilies, as well as atypical and viral CKRs, and engineered and viral chemokines. The reference dataset of chemokine – CKR complexes curated for this study allowed us to address multiple issues concerning the prediction of protein complexes' structures, considering both the built-in limits of AlphaFold, and its applicability towards targets with limited availability of data. The complexes in the dataset can be divided in endogenous complexes (CCL15-CCR1, CCL2-CCR2, CCL3-CCR5, CCL5-CCR5, CCL20-CCR6, CXCL8-CXCR2, CX3CL1-CX3CR1 and CXCL12-ACK3) and non-native complexes, this second group containing complexes with engineered components ([5P7]CCL5-CCR5, [6P4]CCL5-CCR5, and CXCL12LRHQ-ACKR3) and complexes between human and viral proteins (vCCL2-CXCR4, CX3CL1-US28, and CX3CL1.35-US28). As AlphaFold utilizes coevolutionary data to predict protein-protein interactions, its accuracy could be reduced for protein pairs that don't have a clear shared evolutionary history. This is the case for proteins from two different species and partly for engineered proteins with mutated residues.

Considering instead the availability of data for each complex from AlphaFold's point of view, different scenarios are present in the dataset: nine of the fourteen structures were released before the cutoff date of 30.09.2021, therefore they were included in the training set used for AlphaFold-Multimer (AlphaFold v2.3.0). This is something we could not modify, unlike the structures provided as templates. For the templates, eight out of nine receptors in our dataset have structures available in apo-form or bound to non-chemokine ligands, which were included in our protocol as structure templates. CCR6 had no such structures in the PDB.

Figure 6.2.2: Comparison of nine multimer model generation protocols, based on DockQ scores of chemokine - CKR complexes.



Model quality

Using LIT-AlphaFold, we generated structural predictions for the 14 chemokine – CKR complexes using nine different protocols (Table 6.2.1) and compared them to the reference structures. We used the DockQ scores of the top-ranking structures of each complex to compare the different model generation settings. All protocols generated structures with an average DockQ score of 0.70 or 0.71, with 13 of 14 models (93%) being of medium or high quality (Figure 6.2.2). The average DockQ score and its distribution over the fourteen models showed no significant difference in the quality of the protein-protein interfaces predicted using the different methods. Based on these results, we carried out further analysis on the structures with the highest confidence score generated after repeating calculations five times with default settings.

Chemokine N-Terminus Binding Mode

Across all models, the chemokine N-terminus was correctly predicted to occupy the receptor transmembrane cavity. However, a high degree of accuracy is important for the binding mode prediction to be informative about key interactions regarding binding and activation. To evaluate the precision of the predicted binding modes, we measured the all-atom RMSDs of the chemokine N-termini with the surrounding binding site side chains aligned.

The resulting all-atom RMSD values ranged between 1.03 and 6.15 Å, with an average of 3.18 Å among all complexes and 2.88 Å among complexes without viral components (Figure 6.2.3a). Upon visual analysis, the models included instances of well replicated binding modes (CXCL12-ACKR3:1.88 Å, CCL2-CCR2:2.81 Å), as well as unhelpful, inaccurate binding modes (CX3CL1.35-US28:6.15 Å, CCL5-CCR5:5.24 Å) (Figure 6.2.3b). In this context, we

aimed to determine if AlphaFold's internal confidence score pLDDT, calculated for each residue, could reliably indicate prediction accuracy. After calculating the mean pLDDT among the chemokine N-terminus residues for each complex, a moderate negative correlation of -0.57 was found between the mean pLDDT values and the RMSD values (Figure 6.2.3a). With two exceptions, a mean pLDDT score higher than 80 indicated an RMSD less than 3 Å.

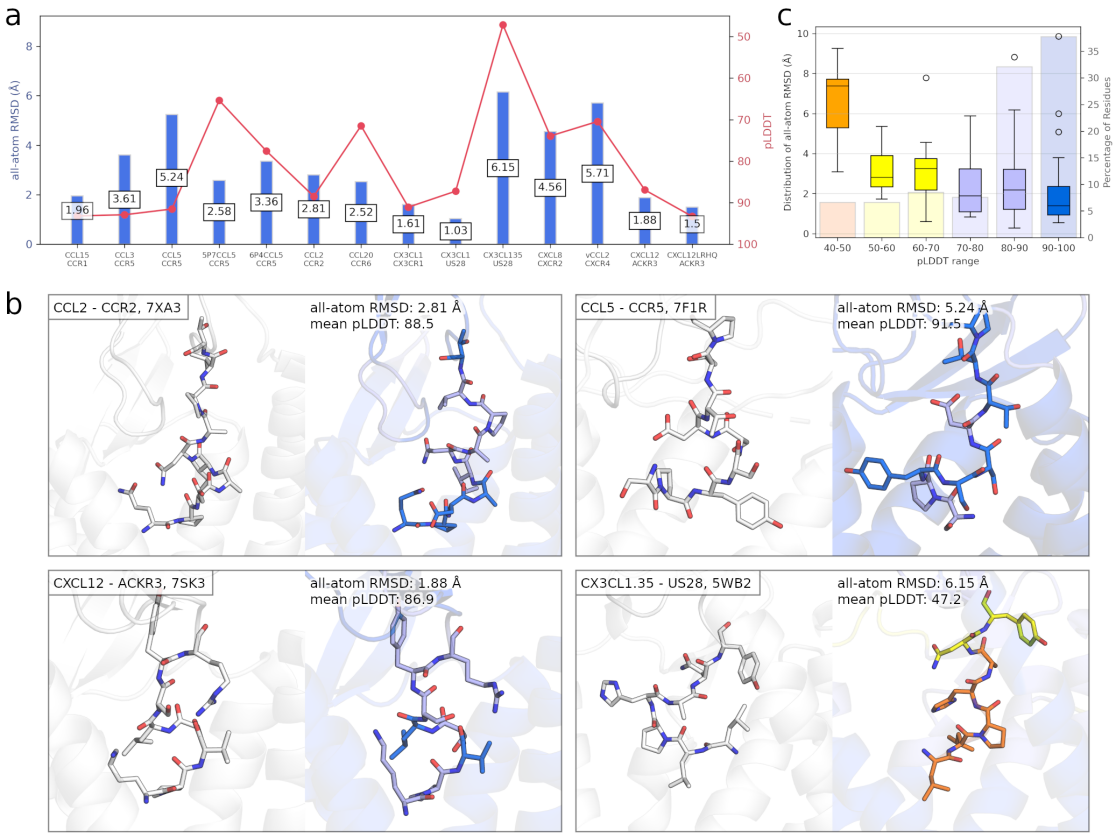
We then compared the per-residue pLDDT scores to the per-residue RMSD values (Figure 4c). Out of all individual N-terminus residues, 39% exhibited a pLDDT value of 90 or above. This subset displayed a mean RMSD of 1.99 Å but also included six outlier residues (out of 41) above 3 Å. For residues with pLDDT values greater or equal to 80 and below 90 (constituting 31% of all residues), the average RMSD remained reasonably modest at 2.65 Å. However, as an increased number of 11 residues (out of 33) showed an RMSD above 3 Å within the 80–90 pLDDT range, the reliability of these predictions diminishes. On the other end of the spectrum, residues with pLDDT values between 40 and 50 had a mean RMSD of 6.53 Å. These results suggest that individual residues with pLDDT scores above 90 are likely to be informative of physiological binding modes of the chemokine N-terminus. Also, low pLDDT scores seem to indicate an unreliable binding mode.

An important aspect to consider when using AlphaFold to model chemokine N-terminus binding modes is the presence of a nonstandard amino acid, pyroglutamic acid (PCA), as the first residue of the N-terminus in some chemokines (Table 6.2.2, column "PCA"). Since AlphaFold does not consider nonstandard amino acids, these were replaced in the input sequence with glutamine, based on UniProt annotation. For the RMSD measurements of the chemokine N-terminus, the glutamine residues standing in place for PCA residues were ignored. The two amino acids have differing steric volumes and physicochemical properties, which might influence the ability of AlphaFold in correctly placing the N-terminus inside the binding cavity. This was the case for the complexes with CX3CL1 but not as much for [5P7]CCL5 or [6P4]CCL5 (Figure S3).

Activation state of predicted structures

GPCRs are characterized by their ability to assume two main conformational states: an active state and an inactive state. We classified the structures generated by different prediction protocols using 2D reference distance zones describing the inactive and active states. The thresholds were determined by measuring from all available experimental structures of CKRs the outward tilt of TM6 and the inward tilt of TM7, which are the two main motions characterizing class A GPCR activation. The default AlphaFold pipeline generated structures for the free receptors in three conformational states: five structures were predicted in an inactive or inactive-like state (CCR2, CCR5, CCR6, CXCR2, CXCR4), three in an active or active-like state (ACKR3, US28, CCR1), and one (CX3CR1) in an intermediate state. Using both state specific templates and modified MSAs, it was possible to generate both active or active-like, and inactive or inactive-like conformations for all targets. AlphaFold-Multimer generated different activation states for CKRs that were in complex with chemokines depending on the prediction protocol (Table S6, Figures S5-S18). The default prediction pipeline, with and without dropout, generated only active or active-like state receptor structures for four receptors, only in the inactive state for CXCR4 (Figure 6.2.4a), and in multiple activation states for four receptors. If multiple receptor conformations were predicted, at least one structure in an active or active-like state was always generated. Using

Figure 6.2.3: Analysis of AlphaFold-predicted chemokine N-terminus binding modes. a) Bar plot showing the all-atom RMSD values for the predicted chemokine N-terminus binding modes. Superimposed on this plot are the average pLDDT values as red dots for the N-terminus residues, offering a glance into the correlation between prediction confidence and accuracy. b) Comparison of reference (in white sticks) and model chemokine N-terminal binding modes in four chemokine–CKR complexes. The sticks of each residue in the model are individually color-coded based on their per-residue pLDDT scores. This color scheme is consistent with the boxes in panel c. c) In front, a boxplot showing the distribution of per-residue all-atom RMSD values within various pLDDT confidence ranges. Behind, a bar plot indicating the percentage of residues falling within each pLDDT range.



state specific templates, we were able to generate structures in both activation states for five receptors, while for the remaining four we obtained only one of the two main conformational states. Combining template selection and MSA editing, by removing the unpaired region of the multimer MSA, increased the number of receptors modelled in both activation states to six (Activation plot of CXCR4 in Figure 6.2.4c). Combining template selection with dropout generated structures in both activation states for seven receptors. Combining all three methods did not improve the results obtained using template selection and dropout. None of the tested prediction protocols could generate an inactive or inactive-like conformation for CX3CR1 and US28 (Activation plot of CX3CR1 in Figure 6.2.4d).

6.2.4 Discussion

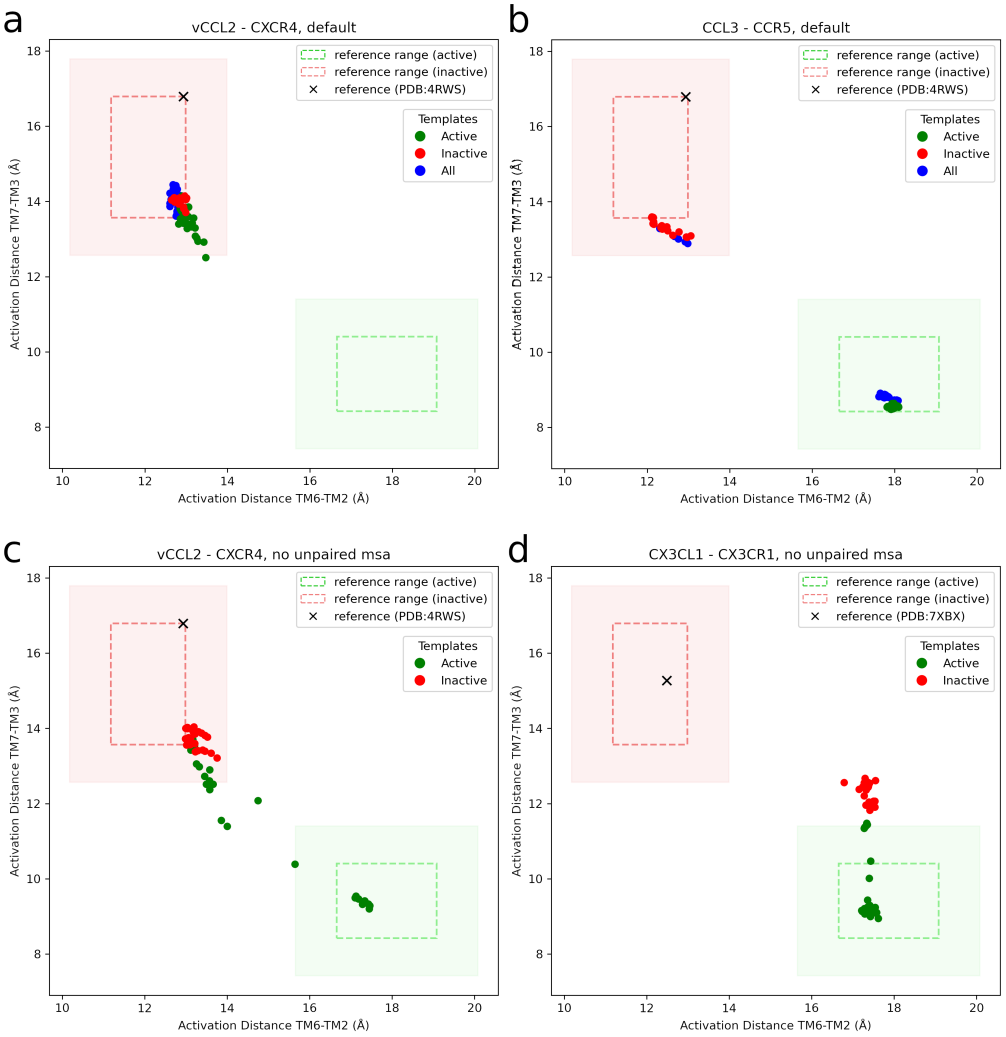
In this study, we evaluated the ability of AlphaFold-Multimer to predict the structures of chemokine – CKR complexes by employing global measurements as well as local measurements at different interaction sites. On average, all the different prediction protocols we applied yielded complexes with a similar quality at the protein-protein interface. The mean DockQ score of 0.71 indicates a relatively good quality of prediction. Comparing to numbers cited in other studies (ranging between 0.5 and 0.7 using different methods and datasets[28, 31, 46, 47] suggests that chemokine – CKR complexes are relatively easy targets for AlphaFold-multimer.

Regarding the modeling of unsolved chemokine–CKR complexes and their precise binding modes on the atomic level, we would recommend careful consideration of the pLDDT confidence scores provided for each residue. Our findings aligned with AlphaFold's general guidelines. If AlphaFold-Multimer can generate a prediction with pLDDT values exceeding 90 for a segment of interest, such as the chemokine N-terminus, the prediction is most likely accurate. For some complexes, the pLDDT scores remained modest even when the calculations were repeated multiple times, and in those cases, the models should be approached with caution. For the prediction of atomic-level interactions, AlphaFold-Multimer serves best as a tool for generating valuable starting hypotheses, which can be further validated using experimental methods such as site-directed mutagenesis.

A question of interest in predicting structures of GPCRs is their activation state. CKRs behave similarly to other class A GPCRs, with most of the monomer structures being generated in an inactive or inactive-like conformation. This bias has been explained by the imbalance between active state and inactive state structures in the AlphaFold training set. However, when predicting chemokine – CKRs structures the active state was prevalent. AlphaFold-Multimer was trained on protein-protein complex structures, and according to GPCRdb among the 171 structures of class A GPCRs released before the training set cutoff date and bound to another biomolecule, 140 are in an active state conformation (10/16 for CKRs). The training set of AlphaFold-Multimer is therefore biased towards active state conformations for class A GPCRs and CKRs, explaining the higher number of generated active state structures.

As expected, monomer structures of the receptor could be generated in both the active and inactive state by using state specific templates and by subsampling or masking the MSA. For chemokine – CKR complexes, selecting state specific templates using the automatic LIT-AlphaFold preparation pipeline was sufficient to obtain a specific conformation of the receptor for five of the nine targets. Forced sampling with enabled dropout, using the modified inputs, was successful in generating structures in different activation states for more difficult

Figure 6.2.4: Activation state analysis of receptors in chemokine-CKR complexes. a) Activation states of vCCL2-CXCR4 complex models. b) Activation states of CCL3-CCR5 complex models. c) Activation states of vCCL2-CXCR4 complex models generated without unpaired MSA. d) Activation states of CX3CL1-CX3CR1 complex models generated without unpaired MSA.



targets. Only two receptors could not be modelled in both activation states with any of the investigated methods, the viral chemokine receptor US28, and CX3CR1.

Modelling the different activation states of CX3CR1 highlights one of the limits of AlphaFold in modelling specific conformational states for unknown targets. CX3CR1 has a non-canonical active state not observed in other CKR structures. The intracellular binding cavity used by G-protein is not formed by an outward movement of TM6, but rather of TM7. AlphaFold could generate monomer structures resembling a canonical active state, or a canonical inactive state, while AlphaFold-Multimer generated mostly structures in a canonical active state. Neither of the two methods was able to reproduce the non-canonical active state observed in the reference structure.

6.2.5 Conclusions

Our analysis showed that AlphaFold-Multimer is capable of predicting the structural elements that interact with each other and the general orientation of the complex in native chemokine–CKR complexes. On the level of side chain conformations and specific interactions, AlphaFold-Multimer cannot replace experimental determination of structures but works well in generating very good initial structural hypotheses, which can be further strengthened by site directed mutagenesis or cross-linking experiments. We suggest carefully considering the confidence scores when generating models of unsolved complexes. With our data set, the binding of chemokine N-terminus residues in the orthosteric binding pocket of CKRs was predicted well for 85% of residues with confidence scores of 90 or above. In our analysis of receptor activation states, the generated receptor structures were mostly in an active state conformation. Using state specific structural templates was generally sufficient to stir predictions toward the desired conformation. Finally, we generated a structure of the unsolved CXCL12-CXCR4 complex which had high confidence scores and could explain experimental results that were not used to determine the complex.

Data and Software Availability

LIT-AlphaFold can be found at `https://github.com/LIT-CCM-lab/LIT-AlphaFold`. The code of AlphaFold and AlphaFold-Multimer for the v2.3.1, as well as the related models' weights, genetic databases, and template database can be found at `https://github.com/google-deepmind/alphafold`. The FASTA files containing the sequences of the chemokines and CKRs, the structure files generated by AlphaFold-Multimer used for benchmarking, and the structure of the CXCL12 – CXCR4 with the highest confidence score are available in the supporting information. All AlphaFold output, reference structures and data analysis scripts are available on Zenodo at `https://zenodo.org/records/10948412`.

Author Information

Corresponding Author

*Esther Kellenberger – Laboratoire d'Innovation Thérapeutique, UMR 7200 CNRS, Université de Strasbourg, 67400, Illkirch, France; Email:ekellen@unistra.fr

Author Contributions

Model benchmarking analysis was carried out by L.U. Software development and model generation conducted by L.C. AlphaFold installation and setup, software engineering, and technical assistance by G.B. Model evaluation protocols proposed by L.U., L.C., C.J., and E.K. Writing—original draft preparation by L.U., L.C., and E.K. Writing—review and editing by L.U., L.C., G.B., C.J., and E.K. All authors have given approval to the final version of the manuscript. L.U. and L.C. contributed equally

Acknowledgements

This research was supported by the French National Research Agency (Project ANR-21-CE44-0040), Centre National de la Recherche Scientifique (CNRS), and Université de Strasbourg. This work of the Interdisciplinary Thematic Institute IMS, the drug discovery and development institute, as part of the ITI 2021-2028 program of the University of Strasbourg, CNRS and Inserm, was supported by IdEx Unistra (ANR-10-IDEX-0002), and by SFRI-STRAT'US project (ANR-20-SFRI-0012) under the framework of the French Investments for the Future Program.

Abbreviations

7TM, seven-transmembrane; ACKR, atypical chemokine receptor; CKR, chemokine receptor; CRS, chemokine recognition site; CRS1, chemokine recognition site 1; CRS1.5, chemokine recognition site 1.5; CRS2, chemokine recognition site 2; CRS3, chemokine recognition site 3; cryo-EM, cryo-electron microscopy; ECL, extracellular loop; GPRC, G protein-coupled receptor; ICL, intracellular loop; MSA, multiple sequence alignment; PCA, pyroglutamic acid; PDB, protein data bank; pLDDT, predicted local distance difference test; RMSD, root mean square deviation; TM, transmembrane domain.

References

[1] Francoise Bachelerie, Adit Ben-Baruch, Amanda M. Burkhardt, Christophe Combadiere, Joshua M. Farber, Gerard J. Graham, Richard Horuk, Alexander Hovard Sparre-Ulrich, Massimo Locati, Andrew D. Luster, Alberto Mantovani, Kouji Matsushima, Philip M. Murphy, Robert Nibbs, Hisayuki Nomiyama, Christine A. Power, Amanda E. I. Proudfoot, Mette M. Rosenkilde, Antal Rot, Silvano Sozzani, Marcus Thelen, Osamu Yoshie, and Albert Zlotnik. International Union of Basic and Clinical Pharmacology. LXXXIX. Update on the Extended Family of Chemokine Receptors and Introducing a New Nomenclature for Atypical Chemokine Receptors. *Pharmacological Reviews*, 66(1):1–79, January 2014.

[2] Catherine E. Hughes and Robert J. B. Nibbs. A guide to chemokines and their receptors. *The FEBS Journal*, 285(16):2944–2971, 2018.

[3] Nisha Nagarsheth, Max S. Wicha, and Weiping Zou. Chemokines in the cancer microenvironment and their relevance in cancer immunotherapy. *Nature Reviews Immunology*, 17(9):559–572, September 2017.

[4] Amanda E. I. Proudfoot. Chemokine receptors: Multifaceted therapeutic targets. *Nature Reviews Immunology*, 2(2):106–115, February 2002.

[5] Thomas J. Schall and Amanda E. I. Proudfoot. Overcoming hurdles in developing successful drugs targeting chemokine receptors. *Nature Reviews Immunology*, 11(5):355–363, May 2011.

[6] Antonella Viola and Andrew D. Luster. Chemokines and Their Receptors: Drug Targets in Immunity and Inflammation. *Annual Review of Pharmacology and Toxicology*, 48(Volume 48, 2008):171–197, February 2008.

[7] John S. Burg, Jessica R. Ingram, A. J. Venkatakrishnan, Kevin M. Jude, Abhiram Dukkipati, Evan N. Feinberg, Alessandro Angelini, Deepa Waghray, Ron O. Dror, Hidde L. Ploegh, and K. Christopher Garcia. Structural basis for chemokine recognition and activation of a viral G protein–coupled receptor. *Science*, 347(6226):1113–1117, March 2015.

[8] Polina Isaikina, Ching-Ju Tsai, Nikolaus Dietz, Filip Pamula, Anne Grahl, Kenneth N. Goldie, Ramon Guixà-González, Camila Branco, Marianne Paolini-Bertrand, Nicolas Calo, Fabrice Cerini, Gebhard F. X. Schertler, Oliver Hartley, Henning Stahlberg, Timm Maier, Xavier Deupi, and Stephan Grzesiek. Structural basis of the activation of the CC chemokine receptor 5 by a chemokine agonist. *Science Advances*, 7(25):eabg8685, June 2021.

[9] Kaiwen Liu, Lijie Wu, Shuguang Yuan, Meng Wu, Yueming Xu, Qianqian Sun, Shu Li, Suwen Zhao, Tian Hua, and Zhi-Jie Liu. Structural basis of CXC chemokine receptor 2 activation and signalling. *Nature*, 585(7823):135–140, September 2020.

[10] Minmin Lu, Wenli Zhao, Shuo Han, Xiaowen Lin, Tingyu Xu, Qiuxiang Tan, Mu Wang, Cuiying Yi, Xiaojing Chu, Weibo Yang, Ya Zhu, Beili Wu, and Qiang Zhao. Activation of the human chemokine receptor CX3CR1 regulated by cholesterol. *Science Advances*, 8(26):eabn8048, July 2022.

[11] Timothy F Miles, Katja Spiess, Kevin M Jude, Naotaka Tsutsumi, John S Burg, Jessica R Ingram, Deepa Waghray, Gertrud M Hjorto, Olav Larsen, Hidde L Ploegh, Mette M Rosenkilde, and K Christopher Garcia. Viral GPCR US28 can signal in response to chemokine agonists of nearly unlimited structural degeneracy. *eLife*, 7:e35850, June 2018.

[12] Ling Qin, Irina Kufareva, Lauren G. Holden, Chong Wang, Yi Zheng, Chunxia Zhao, Gustavo Fenalti, Huixian Wu, Gye Won Han, Vadim Cherezov, Ruben Abagyan, Raymond C. Stevens, and Tracy M. Handel. Crystal structure of the chemokine receptor CXCR4 in complex with a viral chemokine. *Science*, 347(6226):1117–1122, March 2015.

[13] Zhehua Shao, Qingya Shen, Bingpeng Yao, Chunyou Mao, Li-Nan Chen, Huibing Zhang, Dan-Dan Shen, Chao Zhang, Weijie Li, Xufei Du, Fei Li, Honglei Ma, Zhi-Hua Chen, H. Eric Xu, Songmin Ying, Yan Zhang, and Huahao Shen. Identification and mechanism of G protein-biased ligands for chemokine receptor CCR1. *Nature Chemical Biology*, 18(3):264–271, March 2022.

[14] Zhehua Shao, Yangxia Tan, Qingya Shen, Li Hou, Bingpeng Yao, Jiao Qin, Peiyu Xu, Chunyou Mao, Li-Nan Chen, Huibing Zhang, Dan-Dan Shen, Chao Zhang, Weijie Li, Xufei Du, Fei Li, Zhi-Hua Chen, Yi Jiang, H. Eric Xu, Songmin Ying, Honglei Ma, Yan Zhang, and Huahao Shen. Molecular insights into ligand recognition and activation of chemokine receptors CCR2 and CCR3. *Cell Discovery*, 8(1):44, May 2022.

[15] Naotaka Tsutsumi, Shoji Maeda, Qianhui Qu, Martin Vögele, Kevin M. Jude, Carl-Mikael Suomivuori, Ouliana Panova, Deepa Waghray, Hideaki E. Kato, Andrew Velasco, Ron O. Dror, Georgios Skiniotis, Brian K. Kobilka, and K. Christopher Garcia. Atypical structural snapshots of human cytomegalovirus GPCR interactions with host G proteins. *Science Advances*, 8(3):eabl5442, January 2022.

[16] David Jonathan Wasilko, Zachary Lee Johnson, Mark Ammirati, Ye Che, Matthew C. Griffor, Seungil Han, and Huixian Wu. Structural basis for chemokine receptor CCR6 activation by the endogenous protein ligand CCL20. *Nature Communications*, 11(1):3031, June 2020.

[17] Yu-Chen Yen, Christopher T. Schafer, Martin Gustavsson, Stefanie A. Eberle, Pawel K. Dominik, Dawid Deneka, Penglie Zhang, Thomas J. Schall, Anthony A. Kossiakoff, John J. G. Tesmer, and Tracy M. Handel. Structures of atypical chemokine receptor 3 reveal the basis for its promiscuity and signaling bias. *Science Advances*, 8(28):eabn8063, July 2022.

[18] Hui Zhang, Kun Chen, Qiuxiang Tan, Qiang Shao, Shuo Han, Chenhui Zhang, Cuiying Yi, Xiaojing Chu, Ya Zhu, Yechun Xu, Qiang Zhao, and Beili Wu. Structural basis for chemokine recognition and receptor activation of chemokine receptor CCR5. *Nature Communications*, 12(1):4151, July 2021.

[19] Yi Zheng, Gye Won Han, Ruben Abagyan, Beili Wu, Raymond C. Stevens, Vadim Cherezov, Irina Kufareva, and Tracy M. Handel. Structure of CC Chemokine Receptor 5 with a Potent Chemokine Antagonist Reveals Mechanisms of Chemokine Recognition and Molecular Mimicry by HIV. *Immunity*, 46(6):1005–1017.e5, June 2017.

[20] Martin Gustavsson. New insights into the structure and function of chemokine receptor:chemokine complexes from an experimental perspective. *Journal of Leukocyte Biology*, 107(6):1115–1122, June 2020.

[21] Marta Arimont, Shan-Liang Sun, Rob Leurs, Martine Smit, Iwan J. P. de Esch, and Chris de Graaf. Structural Analysis of Chemokine Receptor–Ligand Interactions. *Journal of Medicinal Chemistry*, 60(12):4735–4779, June 2017.

[22] Marta Arimont, Carsten Hoffmann, Chris de Graaf, and Rob Leurs. Chemokine Receptor Crystal Structures: What Can Be Learned from Them? *Molecular Pharmacology*, 96(6):765–777, December 2019.

[23] Ram Prasad Bhusal, Simon R. Foster, and Martin J. Stone. Structural basis of chemokine and receptor interactions: Key regulators of leukocyte recruitment in inflammatory responses. *Protein Science*, 29(2):420–432, 2020.

[24] Archana Sonawani, Shalmali Kharche, Debjani Dasgupta, and Durba Sengupta. Insights into the dynamic interactions at chemokine-receptor interfaces and mechanistic models of chemokine binding. *Journal of Structural Biology*, 214(3):107877, September 2022.

[25] Lauri Urvas and Esther Kellenberger. Structural Insights into Molecular Recognition and Receptor Activation in Chemokine–Chemokine Receptor Complexes. *Journal of Medicinal Chemistry*, 66(11):7070–7085, June 2023.

[26] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, August 2021.

[27] N. Bouatta, P. Sorger, and M. AlQuraishi. Protein structure prediction by AlphaFold2: Are attention and symmetries all you need? *Acta Crystallographica Section D: Structural Biology*, 77(8):982–991, August 2021.

[28] Richard Evans, Michael O'Neill, Alexander Pritzel, Natasha Antropova, Andrew Senior, Tim Green, Augustin Žídek, Russ Bates, Sam Blackwell, Jason Yim, Olaf Ronneberger, Sebastian Bodenstein, Michal Zielinski, Alex Bridgland, Anna Potapenko, Andrew Cowie, Kathryn Tunyasuvunakool, Rishub Jain, Ellen Clancy, Pushmeet Kohli, John Jumper, and Demis Hassabis. Protein complex prediction with AlphaFold-Multimer, March 2022.

[29] Wensi Zhu, Aditi Shenoy, Petras Kundrotas, and Arne Elofsson. Evaluation of AlphaFold-Multimer prediction on multi-chain protein complexes. *Bioinformatics*, 39(7):btad424, July 2023.

[30] Marc F. Lensink, Guillaume Brysbaert, Nessim Raouraoua, Paul A. Bates, Marco Giulini, Rodrigo V. Honorato, Charlotte van Noort, Joao M. C. Teixeira, Alexandre M. J. J. Bonvin, Ren Kong, Hang Shi, Xufeng Lu, Shan Chang, Jian Liu, Zhiye Guo, Xiao Chen, Alex Morehead, Raj S. Roy, Tianqi Wu, Nabin Giri, Farhan Quadir, Chen Chen, Jianlin Cheng, Carlos A. Del Carpio, Eichiro Ichiishi, Luis A. Rodriguez-Lumbreras, Juan Fernandez-Recio, Ameya Harmalkar, Lee-Shin Chu, Sam Canner, Rituparna Smanta, Jeffrey J. Gray, Hao Li, Peicong Lin, Jiahua He, Huanyu Tao, Sheng-You Huang, Jorge Roel-Touris, Brian Jimenez-Garcia, Charles W. Christoffer, Anika J. Jain, Yuki Kagaya, Harini Kannan, Tsukasa Nakamura, Genki Terashi, Jacob C. Verburgt, Yuanyuan Zhang, Zicong Zhang, Hayato Fujuta, Masakazu Sekijima, Daisuke Kihara, Omeir Khan, Sergei Kotelnikov, Usman Ghani, Dzmitry Padhorny, Dmitri Beglov, Sandor Vajda, Dima Kozakov, Surendra S. Negi, Tiziana Ricciardelli, Didier Barradas-Bautista, Zhen Cao, Mohit Chawla, Luigi Cavallo, Romina Oliva, Rui Yin, Melyssa Cheung, Johnathan D. Guest, Jessica Lee, Brian G. Pierce, Ben Shor, Tomer Cohen, Matan Halfon, Dina Schneidman-Duhovny, Shaowen Zhu, Rujie Yin, Yuanfei Sun, Yang Shen, Martyna Maszota-Zieleniak, Krzysztof K. Bojarski, Emilia A. Lubecka,

Mateusz Marcisz, Annemarie Danielsson, Lukasz Dziadek, Margrethe Gaardlos, Artur Gieldon, Adam Liwo, Sergey A. Samsonov, Rafal Slusarz, Karolina Zieba, Adam K. Sieradzan, Cezary Czaplewski, Shinpei Kobayashi, Yuta Miyakawa, Yasuomi Kiyota, Mayuko Takeda-Shitaka, Kliment Olechnovic, Lukas Valancauskas, Justas Dapkunas, Ceslovas Venclovas, Bjorn Wallner, Lin Yang, Chengyu Hou, Xiaodong He, Shuai Guo, Shenda Jiang, Xiaoliang Ma, Rui Duan, Liming Qui, Xianjin Xu, Xiaoqin Zou, Sameer Velankar, and Shoshana J. Wodak. Impact of AlphaFold on structure prediction of protein complexes: The CASP15-CAPRI experiment. *Proteins: Structure, Function, and Bioinformatics*, n/a(n/a).

[31] Isak Johansson-Åkhe and Björn Wallner. Improving peptide-protein docking with AlphaFold-Multimer using forced sampling. *Frontiers in Bioinformatics*, 2, 2022.

[32] Diego del Alamo, Davide Sala, Hassane S Mchaourab, and Jens Meiler. Sampling alternative conformational states of transporters and receptors with AlphaFold2. *eLife*, 11:e75751, March 2022.

[33] Davide Sala, Peter W. Hildebrand, and Jens Meiler. Biasing AlphaFold2 to predict GPCRs and kinases with user-defined functional or structural properties. *Frontiers in Molecular Biosciences*, 10, 2023.

[34] Richard A. Stein and Hassane S. Mchaourab. SPEACH_AF: Sampling protein ensembles and conformational heterogeneity with Alphafold2. *PLOS Computational Biology*, 18(8):e1010483, August 2022.

[35] Lim Heo and Michael Feig. Multi-state modeling of G-protein coupled receptors at experimental accuracy. *Proteins: Structure, Function, and Bioinformatics*, 90(11):1873–1885, 2022.

[36] D. Sala, F. Engelberger, H. S. Mchaourab, and J. Meiler. Modeling conformational states of proteins with AlphaFold. *Current Opinion in Structural Biology*, 81:102645, August 2023.

[37] Hannah K. Wayment-Steele, Adedolapo Ojoawo, Renee Otten, Julia M. Apitz, Warintra Pitsawong, Marc Hömberger, Sergey Ovchinnikov, Lucy Colwell, and Dorothee Kern. Predicting multiple conformations via sequence clustering and AlphaFold2. *Nature*, 625(7996):832–839, January 2024.

[38] Milot Mirdita, Konstantin Schütze, Yoshitaka Moriwaki, Lim Heo, Sergey Ovchinnikov, and Martin Steinegger. ColabFold: Making protein folding accessible to all. *Nature Methods*, 19(6):679–682, June 2022.

[39] Björn Wallner. AFsample: Improving multimer prediction with AlphaFold using massive sampling. *Bioinformatics*, 39(9):btad573, September 2023.

[40] The UniProt Consortium. UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, 51(D1):D523–D531, January 2023.

[41] Dingquan Yu, Grzegorz Chojnowski, Maria Rosenthal, and Jan Kosinski. AlphaPulldown—a python package for protein–protein interaction screens using AlphaFold-Multimer. *Bioinformatics*, 39(1):btac749, January 2023.

[42] Vignir Isberg, Bas Vroling, Rob van der Kant, Kang Li, Gert Vriend, and David Gloriam. GPCRDB: An information system for G protein-coupled receptors. *Nucleic Acids Research*, 42(D1):D422–D425, January 2014.

[43] Albert J Kooistra, Stefan Mordalski, Gáspár Pándy-Szekeres, Mauricio Esguerra, Alibek Mamyrbekov, Christian Munk, György M Keserű, and David E Gloriam. GPCRdb in 2021: Integrating GPCR sequence, structure and function. *Nucleic Acids Research*, 49(D1):D335–D343, January 2021.

[44] Sankar Basu and Björn Wallner. DockQ: A Quality Measure for Protein-Protein Docking Models. *PLOS ONE*, 11(8):e0161879, August 2016.

[45] Juan A. Ballesteros and Harel Weinstein. [19] Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. In Stuart C. Sealfon, editor, *Methods in Neurosciences*, volume 25 of *Receptor Molecular Biology*, pages 366–428. Academic Press, January 1995.

[46] Björn Wallner. Improved multimer prediction using massive sampling with AlphaFold in CASP15. *Proteins: Structure, Function, and Bioinformatics*, 91(12):1734–1746, 2023.

[47] Mu Gao, Davi Nakajima An, Jerry M. Parks, and Jeffrey Skolnick. AF2Complex predicts direct physical interactions in multimeric proteins with deep learning. *Nature Communications*, 13(1):1744, April 2022.

6.3 Critical analysis and perspectives

AlphaFold and other deep learning methods for protein structure prediction have become key tools for structural biology, drug discovery, and biomolecular modelling. Modelling of different protein conformational states however remains a complex task, that general purpose state of the art methods like AlphaFold3 are still unable to solve[1]. Thanks to multiple contributions from the research community the predictive power of AlphaFold2[2] could be applied to multistate modelling. LIT-AlphaFold aims to be a continuation of this effort, allowing an easier access to multistate modelling tools for prospective users and developers, and extending their applicability to multimers. Despite being a target agnostic pipeline, LIT-AlphaFold was mostly developed to tackle the problem of GPCRs multistate modelling. Template selection based on structural annotations is currently only supported for GPCRdb[3]. Other structural databases like KLIFS[4] have already been used to perform automatic structure selection[5], and their inclusion in the pipeline will expand its application in high throughput contexts to other targets.

The benchmark study on chemokine-chemokine receptor complexes proved that AlphaFold-Multimer can be a valuable tool for the study of chemokine receptors. The generated structural models are at their worst a good starting hypothesis for the modelling of unknown chemokine – chemokine receptor complexes, while at their best they faithfully reproduced some of the most important features observed in experimental structures, as highlighted by the CXCL12-CXCR4 complex described in Section 4.1.2. Furthermore, using the chemokine-CKRs benchmark dataset it was possible to generate multimer structures with one protomer in a specific conformational state, tackling one of the issues for multistate modelling of multichain targets. The use of state specific templates was sufficient to model most chemokine receptors in both the active and inactive state, while for more difficult targets enabling dropout during inference was sufficient to sample both the active and inactive state. Despite the limited size of the benchmark dataset, we can conclude that multistate modelling of single chains in multimeric complexes does not require editing or subsampling of the MSA unlike what was observed for monomers.

Future endeavours should expand on the possibility of modelling GPCRs in a specific activation state in multimeric complexes. One possible application is the modelling of GPCR dimers[6] since alternative protein-protein interfaces might be observed depending on the activation state of the protomers. It is still an open question whether AlphaFold-Multimer could identify different biologically relevant protein-protein interfaces. Some of the multistate modelling tools implemented in LIT-AlphaFold, like in silico mutations or MSA clustering, could facilitate the exploration of the protein-protein interfaces conformational landscape.

References

[1] Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, Sebastian W. Bodenstein, David A. Evans, Chia-Chun Hung, Michael O'Neill, David Reiman, Kathryn Tunyasuvunakool, Zachary Wu, Akvilė Žemgulytė, Eirini Arvaniti, Charles Beattie, Ottavia Bertolli, Alex Bridgland, Alexey Cherepanov, Miles Congreve, Alexander I. Cowen-Rivers, Andrew Cowie, Michael Figurnov, Fabian B. Fuchs, Hannah Gladman, Rishub Jain, Yousuf A. Khan, Caroline M. R. Low, Kuba Perlin, Anna Potapenko, Pascal Savy, Sukhdeep Singh, Adrian Stecula, Ashok Thillaisundaram, Catherine Tong, Sergei

Yakneen, Ellen D. Zhong, Michal Zielinski, Augustin Žídek, Victor Bapst, Pushmeet Kohli, Max Jaderberg, Demis Hassabis, and John M. Jumper. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, pages 1–3, May 2024.

[2] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, August 2021.

[3] Gáspár Pándy-Szekeres, Jimmy Caroli, Alibek Mamyrbekov, Ali A Kermani, György M Keserű, Albert J Kooistra, and David E Gloriam. GPCRdb in 2023: State-specific structure models using AlphaFold2 and new ligand resources. *Nucleic Acids Research*, page gkac1013, November 2022.

[4] Georgi K Kanev, Chris de Graaf, Bart A Westerman, Iwan J P de Esch, and Albert J Kooistra. KLIFS: An overhaul after the first 5 years of supporting kinase research. *Nucleic Acids Research*, 49(D1):D562–D569, January 2021.

[5] Davide Sala, Peter W. Hildebrand, and Jens Meiler. Biasing AlphaFold2 to predict GPCRs and kinases with user-defined functional or structural properties. *Frontiers in Molecular Biosciences*, 10, 2023.

[6] Jun Jin, Fanny Momboisse, Gaelle Boncompain, Florian Koensgen, Zhicheng Zhou, Nelia Cordeiro, Fernando Arenzana-Seisdedos, Franck Perez, Bernard Lagane, Esther Kellenberger, and Anne Brelot. CCR5 adopts three homodimeric conformations that control cell surface delivery. *Science Signaling*, 11(529):eaal2869, May 2018.

Chapter 7

General conclusions

The work described in this thesis started from the need to design of a small molecule agonist of the C-X-C chemokine receptor type 4 (CXCR4) to use as a chemical probe to efficiently characterize the signalling regulation of the receptor. This initial practical issue led to the study of three problems of structure-based drug design, with a particular focus on the discovery of ligands of the G-protein coupled receptor family (GPCRs): How to selectively identify new molecules with a desired efficacy profile? How can data from molecular dynamics simulations be used in virtual screening? How to efficiently and accurately model a relevant conformation of a protein target?

The design of ligands with a desired efficacy profile has been approached using both ligand-based and structure-based techniques. For structure-based methods the success of virtual screening depended on using a relevant binding site conformation for docking, and by considering the presence of relevant protein-ligand interactions associated to a specific efficacy. These observations were initially identified by an analysis of the literature, comparing the results of different virtual screening campaigns, and further confirmed by benchmarking different computational tools on a panel of different GPCR targets, confirming their utility as general guidelines for future virtual screening campaigns.

Selecting a relevant receptor conformation, and efficacy specific protein-ligand interactions, requires to have access to multiple experimental structures of the receptor bound to ligands with different efficacies. Despite the exponential increase in experimental structures, multiple relevant target still lack sufficient structural information. To maximize the amount of information extracted from a single structure, being experimental or in silico, molecular dynamics could be used to obtain a better insight. To facilitate the use of molecular dynamics information in virtual screening a new method, called InteractionOCSVM, was developed. A machine learning model is trained on the ensemble of binding modes extracted from a simulation, and used to evaluate the binding modes generated by docking. Since InteractionOCSVM was trained to identify relevant binding modes, it was further tested and developed as a tool for efficacy-driven virtual screening targeting GPCRs. Retrospective tests using the developed method showed its potentialities, since the trained models are robust towards imprecisions in the initial reference binding mode.

Protein modelling has been heavily changed in the last years by the release of AlphaFold and AlphaFold-Multimer. The existing pipelines were adapted to incorporate the state of the art methods for the modelling of alternative protein conformations, in order to facilitate their use for the modelling of GPCRs. The developed pipeline was tested for the modelling of both monomeric GPCR structures, and to model chemokine - chemokine receptor structures. The

developed input preparation LIT-AlphaFold was able to model both active and inactive state conformations of the receptor bound to a chemokine, representing one of the first reported cases of multistate modelling for protein-protein complexes.

The different methods and techniques developed and studied in this thesis were applied to the discovery of a small molecule agonist of CXCR4. While the first attempt was unsuccessful, the obtained data could be used in the future to identify a relevant ligand. Additionally the same techniques were used to model and characterize the duration of action of agonists of the β2-adrenergic receptor, including other relevant pharmacological properties in the virtual screening pipeline.

To conclude, in this thesis different computational tools and techniques were benchmarked and developed for structure-based drug discovery, covering multiple aspect from protein modelling, to docking pose evaluation. While many topics were covered in this thesis, an equal if not greater number of questions was further raised, proposing multiple suggestions for future research projects. Most of the developed methods are already or will be soon made public to the research community, hoping that they could help to further develop in silico drug discovery.

Appendix A

Chapter 1.1.3 supporting information

A.1 Section 2.1 supporting information

Table A.1.1: Experimental GPCR structures used in the study (5HT2A, AA2AR, ACM2)

Receptor	PDB ID	Conformation	Ligand
5HT2A	6WHA	Active	25-CN-NBOH
5HT2A	7RAN	Active	(R)-69
5HT2A	7WC9	Inactive-Agonist	93F
5HT2A	7WC4	Inactive-Agonist	5-hydroxytryptamine
5HT2A	7WC8	Inactive-Agonist	LUMATEPERONE
5HT2A	7WC6	Inactive-Agonist	Lysergide
5HT2A	6WGT	Inactive-Agonist	Lysergide
5HT2A	7VOE	Inactive-Agonist	aripiprazole
5HT2A	7VOD	Inactive-Agonist	cariprazine
5HT2A	7WC7	Inactive-Agonist	lisuride
5HT2A	7WC5	Inactive-Agonist	psilocin
5HT2A	6WH4	Inactive-Antagonist	CHEMBL428892
5HT2A	6A93	Inactive-Antagonist	risperidone
5HT2A	6A94	Inactive-Antagonist	zotepine
AA2AR	4UHR	Active	CGS 21680
AA2AR	6GDG	Active	NECA
AA2AR	5G53	Active	NECA
AA2AR	2YDV	Active	NECA
AA2AR	7EZC	Active	UK-432,097
AA2AR	5WF5	Active	UK-432,097
AA2AR	5WF6	Active	UK-432,097
AA2AR	3QAK	Active	UK-432,097
AA2AR	7ARO	Inactive-Agonist	CHEMBL124345
AA2AR	8DU3	Inactive-Antagonist	(4M)-6-bromo-4-(furan-2-yl)quinazolin-2-amine
AA2AR	5OLZ	Inactive-Antagonist	CHEMBL2024114
AA2AR	3UZC	Inactive-Antagonist	CHEMBL2024114
AA2AR	7T32	Inactive-Antagonist	ZM-241385
AA2AR	6S0L	Inactive-Antagonist	ZM-241385
AA2AR	6MH8	Inactive-Antagonist	ZM-241385
AA2AR	5UVI	Inactive-Antagonist	ZM-241385
AA2AR	3VGA	Inactive-Antagonist	ZM-241385
AA2AR	3PWH	Inactive-Antagonist	ZM-241385
AA2AR	3RFM	Inactive-Antagonist	caffeine
AA2AR	8GNG	Inactive-Antagonist	istradefylline
AA2AR	5OLO	Inactive-Antagonist	tozadenant
AA2AR	3REY	Inactive-Antagonist	xanthine amine congener
ACM2	7T8X	Active	acetylcholine
ACM2	4MQS	Active	iperoxo
ACM2	5ZKB	Inactive-Antagonist	AF-DX 384
ACM2	5ZK3	Inactive-Antagonist	CHEMBL558910
ACM2	3UON	Inactive-Antagonist	CHEMBL558910
ACM2	5ZK8	Inactive-Antagonist	N-methyl scopolamine
ACM2	5ZKC	Inactive-Antagonist	N-methyl scopolamine
ACM2	5YC8	Inactive-Antagonist	N-methyl scopolamine

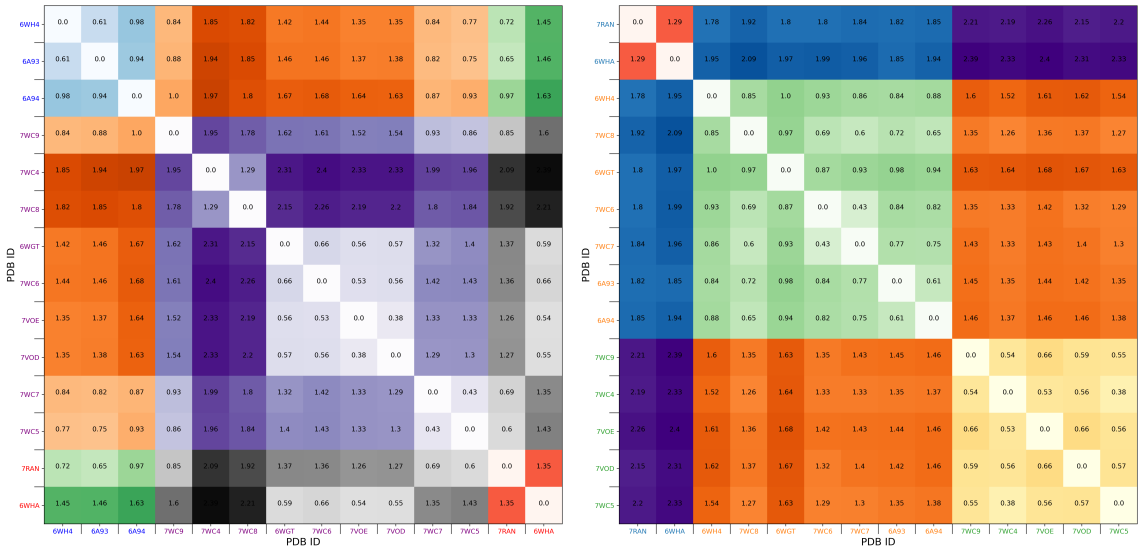
Table A.1.2: Experimental GPCR structures used in the study (ADRB2, CNR1, DRD2, OPRK)

Receptor	PDB ID	Conformation	Ligand
ADRB2	4LDO	Active	(-)-adrenaline
ADRB2	6NI3	Active	CHEMBL1615159
ADRB2	4LDE	Active	CHEMBL1615159
ADRB2	3SN6	Active	CHEMBL1615159
ADRB2	3P0G	Active	CHEMBL1615159
ADRB2	7DHR	Active	LEVISOPRENALINE
ADRB2	7BZ2	Active	arformoterol
ADRB2	4LDL	Active	hydroxybenzylisoproterenol
ADRB2	7DHI	Active	levosalbutamol
ADRB2	6MXT	Active	salmeterol
ADRB2	6PS2	Inactive-Antagonist	CHEMBL1160734
ADRB2	6PRZ	Inactive-Antagonist	CHEMBL1160734
ADRB2	3NYA	Inactive-Antagonist	CHEMBL1160734
ADRB2	3NY9	Inactive-Antagonist	CHEMBL1233771
ADRB2	6PS3	Inactive-Antagonist	CHEMBL3799125
ADRB2	5D6L	Inactive-Antagonist	Carazolol
ADRB2	5JQH	Inactive-Antagonist	Carazolol
ADRB2	5D5A	Inactive-Antagonist	Carazolol
ADRB2	5D5B	Inactive-Antagonist	Carazolol
ADRB2	4GBR	Inactive-Antagonist	Carazolol
ADRB2	2RH1	Inactive-Antagonist	Carazolol
ADRB2	6PS4	Inactive-Antagonist	ICI 118551
ADRB2	3NY8	Inactive-Antagonist	ICI 118551
ADRB2	6PS0	Inactive-Antagonist	carazolol
ADRB2	6PS5	Inactive-Antagonist	propranolol
ADRB2	6PS6	Inactive-Antagonist	timolol
ADRB2	6PS1	Inactive-Antagonist	timolol
ADRB2	3D4S	Inactive-Antagonist	timolol
CNR1	5XRA	Active	AM11542
CNR1	6KPG	Active	AM841
CNR1	5XR8	Active	AM841
CNR1	7V3Z	Active	CP55940
CNR1	6N4B	Active	MDMB-Fubinaca
CNR1	5TGZ	Inactive-Antagonist	AM6538
CNR1	5U09	Inactive-Antagonist	taranabant
DRD2	7JVR	Active	bromocriptine
DRD2	6VMS	Active	bromocriptine
DRD2	6LUQ	Inactive-Antagonist	haloperidol
DRD2	7DFP	Inactive-Antagonist	spiperone
OPRK	6B73	Active	MP1104
OPRK	6VI4	Inactive-Antagonist	JDTic

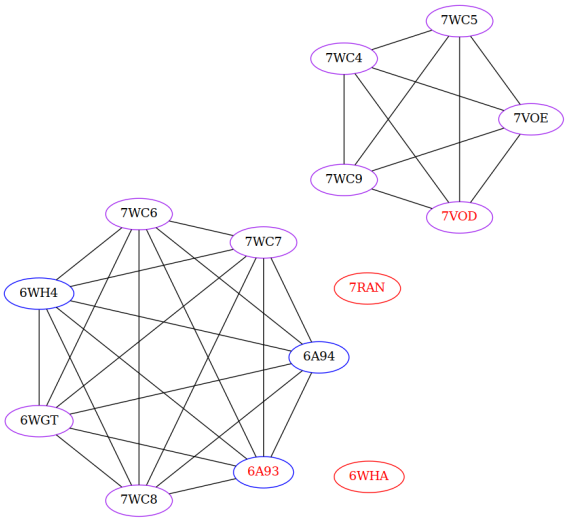
Table A.1.3: Mean, maximum, and minimum ROCAUC values obtained by ranking known agonists and antagonists by molecular similarity to a reference. The positive class was defined as the efficacy of the reference molecule.

Target	Mean MFP	Max MFP	Min MFP	Mean Pharma	Max Pharma	Min Pharma	Mean Gobbi	Max Gobbi	Min Gobbi
5HT2A	0.62	0.88	0.31	0.55	0.88	0.22	0.58	0.82	0.35
AA2AR	0.64	0.99	0.13	0.64	0.92	0.12	0.73	0.96	0.12
ACM2	0.84	0.99	0.33	0.8	0.98	0.33	0.85	0.98	0.2
ADRB2	0.77	0.92	0.17	0.84	0.97	$8.44 \cdot 10^{-2}$	0.79	0.96	0.24
CNR1	0.87	0.98	0.14	0.69	0.88	0.33	0.87	1	0.14
DRD2	0.65	0.81	0.34	0.67	0.89	0.32	0.65	0.82	0.37
OPRK	0.56	0.77	0.33	0.54	0.85	0.18	0.54	0.87	0.27

Figure A.1.1: (A) Pairwise RMSD between the binding sites from crystallographic structures of 5HT2A. Blue labels indicate antagonist bound inactive state structures, purple labels indicate agonist bound inactive state structures, and red labels indicate active state structures. The same colors in the plot indicate the RMSD between structures in the same conformation. Regions in orange indicate the similairty between agonist and antagonist binding modes from inactive state structures, grey between agonists bound to active and inactive state structures, and green between agonists and antagonists. (B) Results of structures clustering based on binding site. (C) Graph obtained by analysis of the given structures, each structure is identified by its PDB ID, IDs in red highlight the representative conformations from each graph.

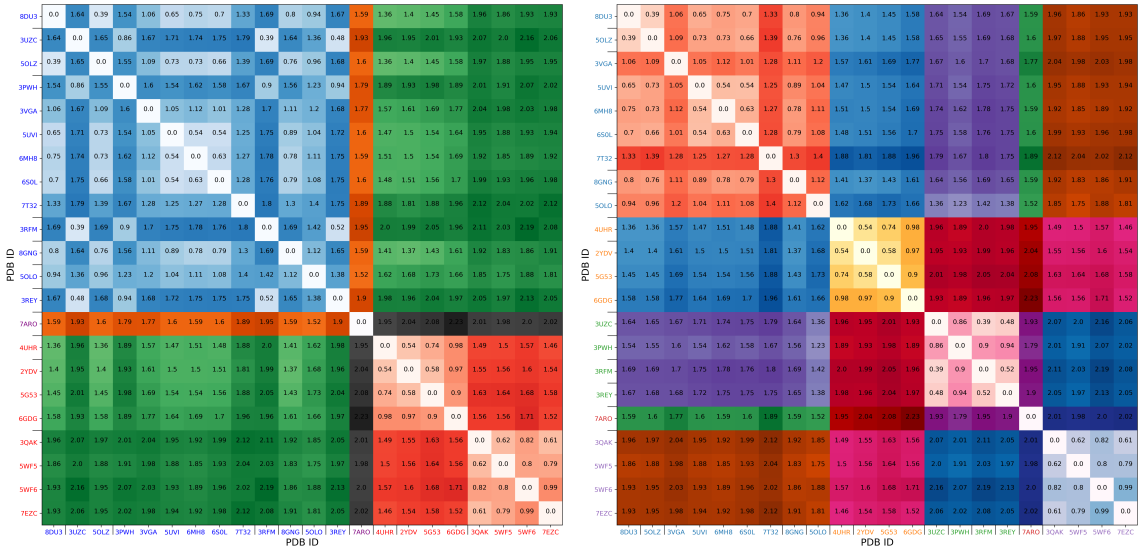


(A) **(B)**

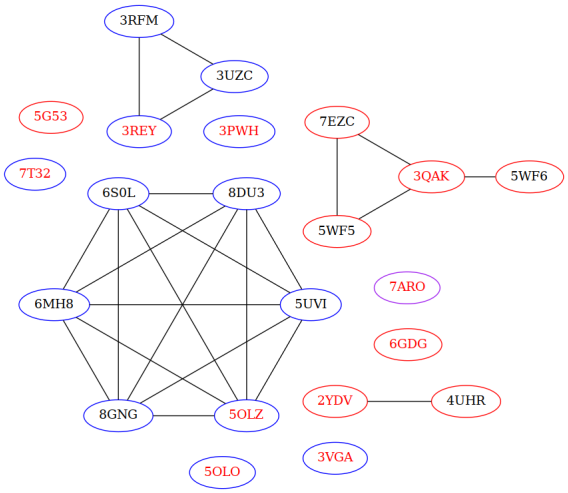


(C)

Figure A.1.2: (A) Pairwise RMSD between the binding sites from crystallographic structures of AA2AR. Blue labels indicate antagonist bound inactive state structures, purple labels indicate agonist bound inactive state structures, and red labels indicate active state structures. The same colors in the plot indicate the RMSD between structures in the same conformation. Regions in orange indicate the similairty between agonist and antagonist binding modes from inactive state structures, grey between agonists bound to active and inactive state structures, and green between agonists and antagonists. (B) Results of structures clustering based on binding site. (C) Graph obtained by analysis of the given structures, each structure is identified by its PDB ID, IDs in red highlight the representative conformations from each graph.



(A) **(B)**



(C)

Figure A.1.3: (A) Pairwise RMSD between the binding sites from crystallographic structures of ACM2. Blue labels indicate antagonist bound inactive state structures, and red labels indicate active state structures. The same colors in the plot indicate the RMSD between structures in the same conformation. Regions in green indicate RMSD values between agonists and antagonists. (B) Results of structures clustering based on binding site. (C) Graph obtained by analysis of the given structures, each structure is identified by its PDB ID, IDs in red highlight the representative conformations from each graph.

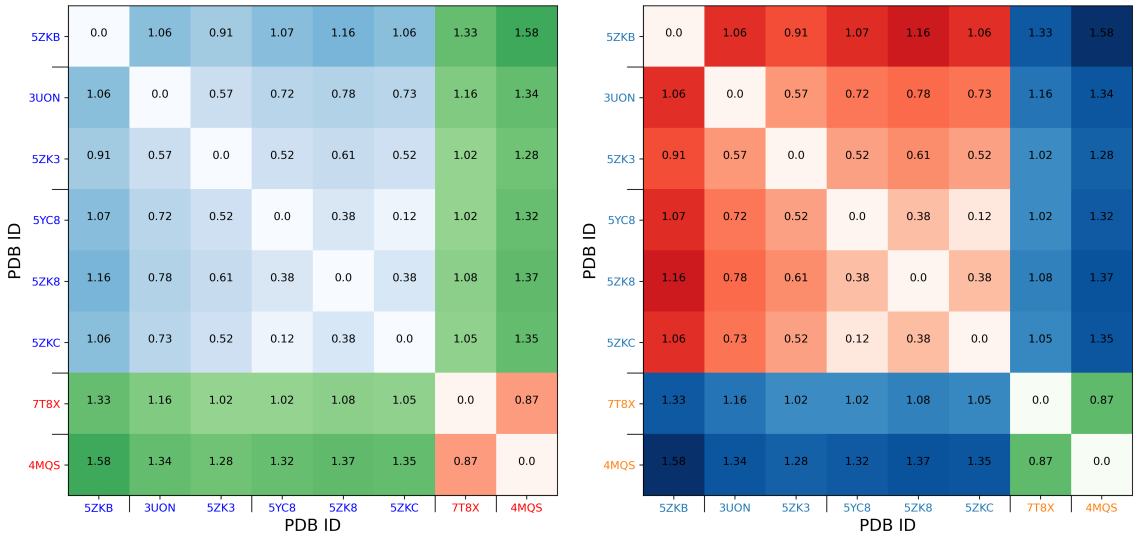


Figure A.1.4: (A) Pairwise RMSD between the binding sites from crystallographic structures of ADRB2. Blue labels indicate antagonist bound inactive state structures, and red labels indicate active state structures. The same colors in the plot indicate the RMSD between structures in the same conformation. Regions in green indicate RMSD values between agonists and antagonists. (B) Results of structures clustering based on binding site. (C) Graph obtained by analysis of the given structures, each structure is identified by its PDB ID, IDs in red highlight the representative conformations from each graph.

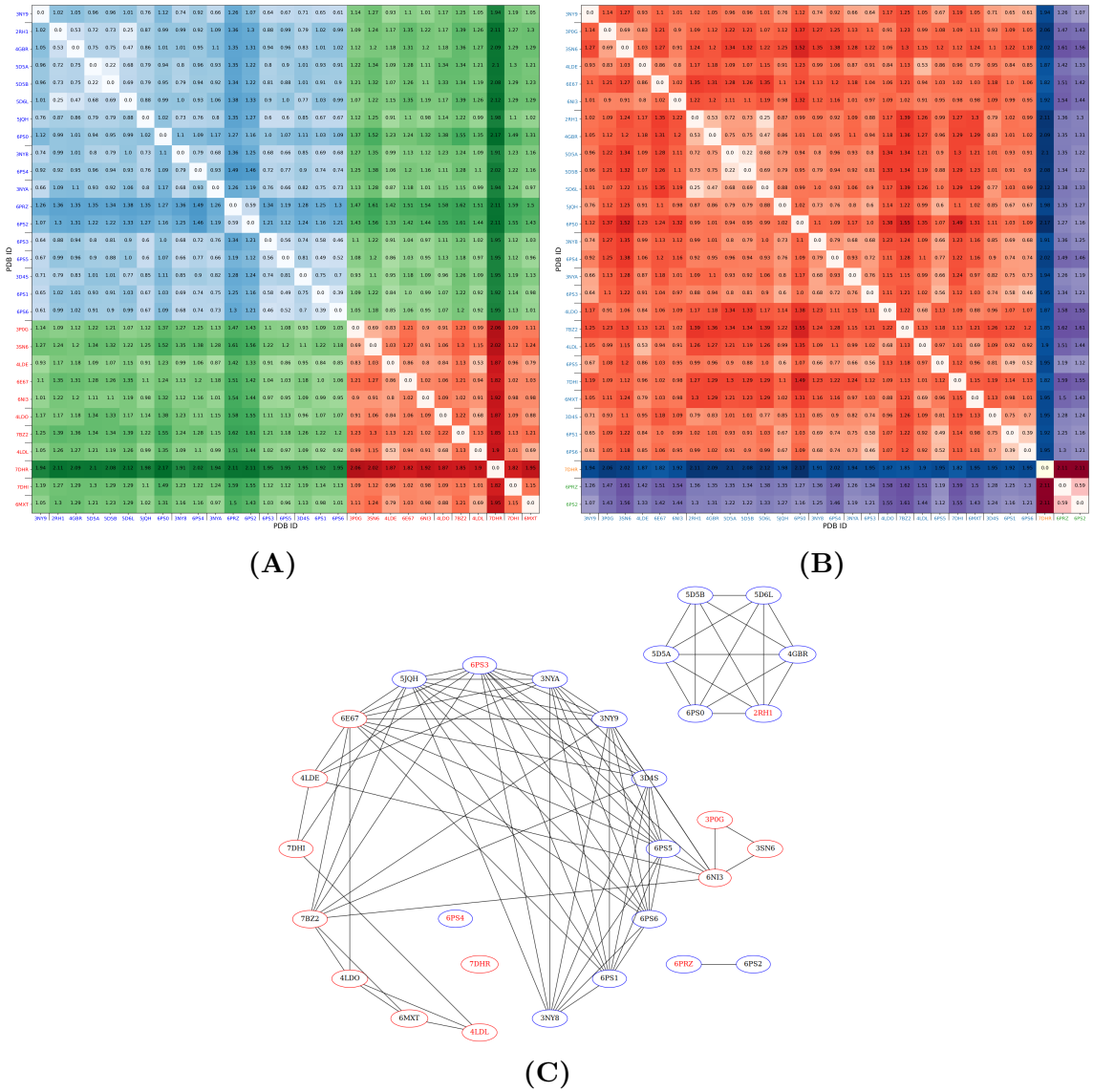


Figure A.1.5: (A) Pairwise RMSD between the binding sites from crystallographic structures of CNR1. Blue labels indicate antagonist bound inactive state structures, and red labels indicate active state structures. The same colors in the plot indicate the RMSD between structures in the same conformation. Regions in green indicate RMSD values between agonists and antagonists. (B) Results of structures clustering based on binding site. (C) Graph obtained by analysis of the given structures, each structure is identified by its PDB ID, IDs in red highlight the representative conformations from each graph.

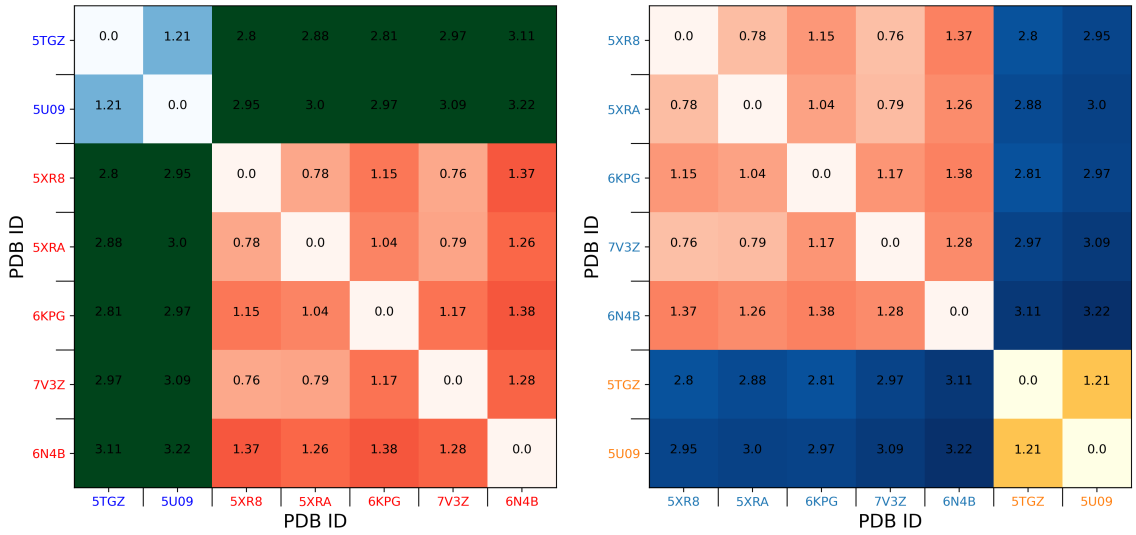
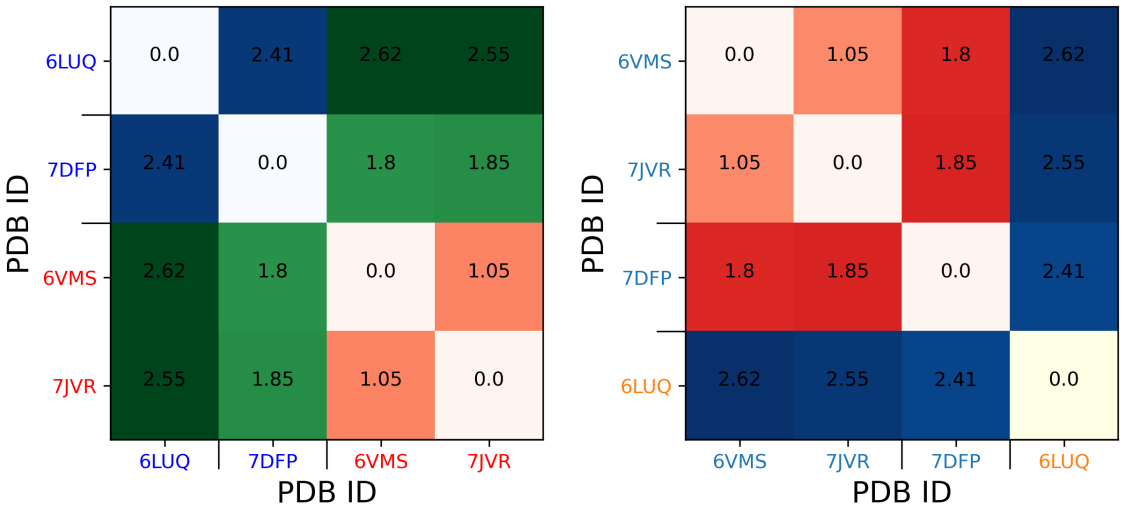
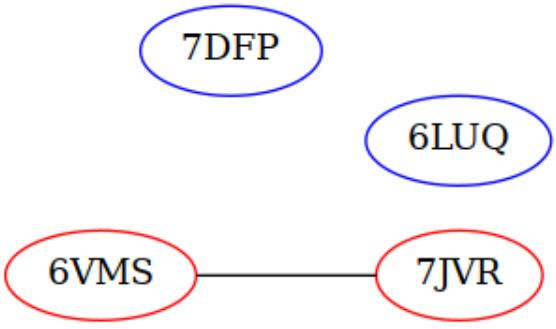


Figure A.1.6: (A) Pairwise RMSD between the binding sites from crystallographic structures of DRD2. Blue labels indicate antagonist bound inactive state structures, and red labels indicate active state structures. The same colors in the plot indicate the RMSD between structures in the same conformation. Regions in green indicate RMSD values between agonists and antagonists. (B) Results of structures clustering based on binding site. (C) Graph obtained by analysis of the given structures, each structure is identified by its PDB ID, IDs in red highlight the representative conformations from each graph.

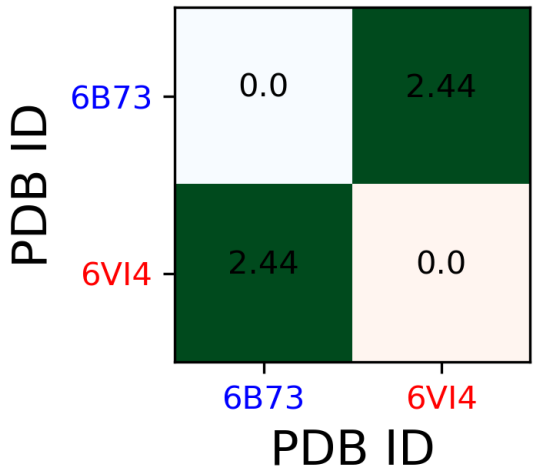


(A) **(B)**



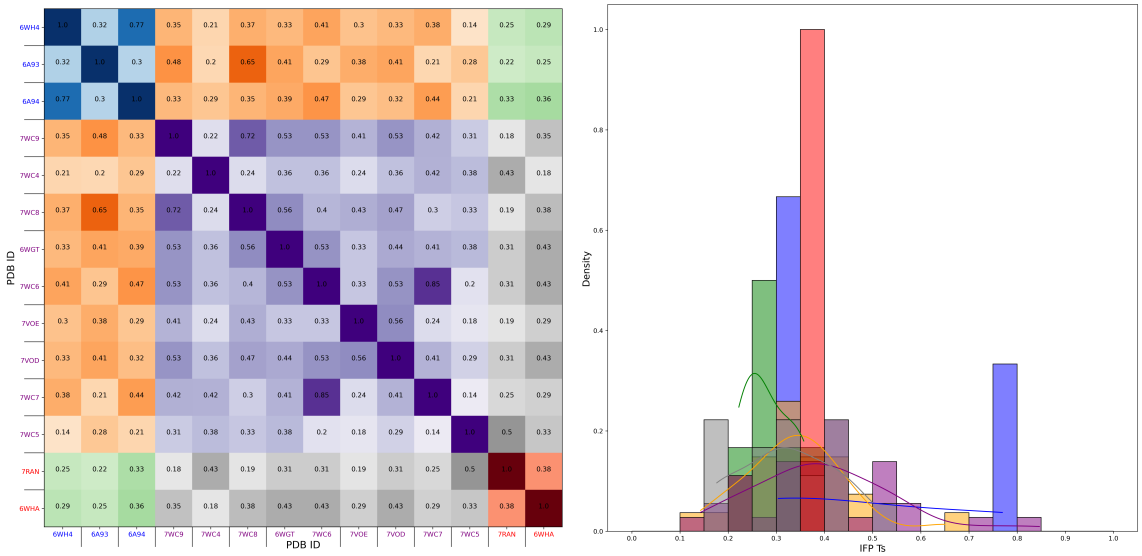
(C)

Figure A.1.7: (A) Pairwise RMSD between the binding sites from crystallographic structures of OPRK. Blue labels indicate antagonist bound inactive state structures, and red labels indicate active state structures. The same colors in the plot indicate the RMSD between structures in the same conformation. Regions in green indicate RMSD values between agonists and antagonists. (B) Results of structures clustering based on binding site. (C) Graph obtained by analysis of the given structures, each structure is identified by its PDB ID, IDs in red highlight the representative conformations from each graph.

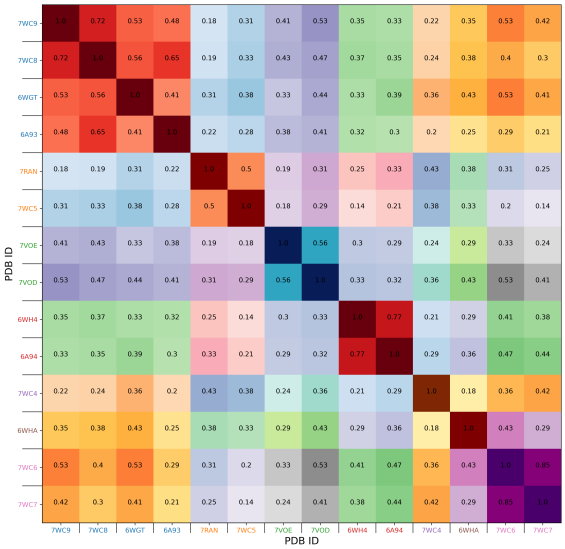


(A)

Figure A.1.8: (A) Pairwise Tanimoto similarity between IFPs extracted from crystallographic structures of 5HT2A. Blue labels indicate antagonist bound inactive state structures, purple labels indicate agonist bound inactive state structures, and red labels indicate active state structures. The same colors in the plot inddicate interaction similarity between structures in the same conformation. Regions in orange indicate the similairty between agonist and antagonist binding modes from inactive state structures, grey between agonists bound to active and inactive state structures, and green between agonists and antagonists. (B) Normalized distribution of pairwise similarities between different binding modes. The color scheme is the same as for plot A. (C) Results of structures clustering based on IFP similarity.

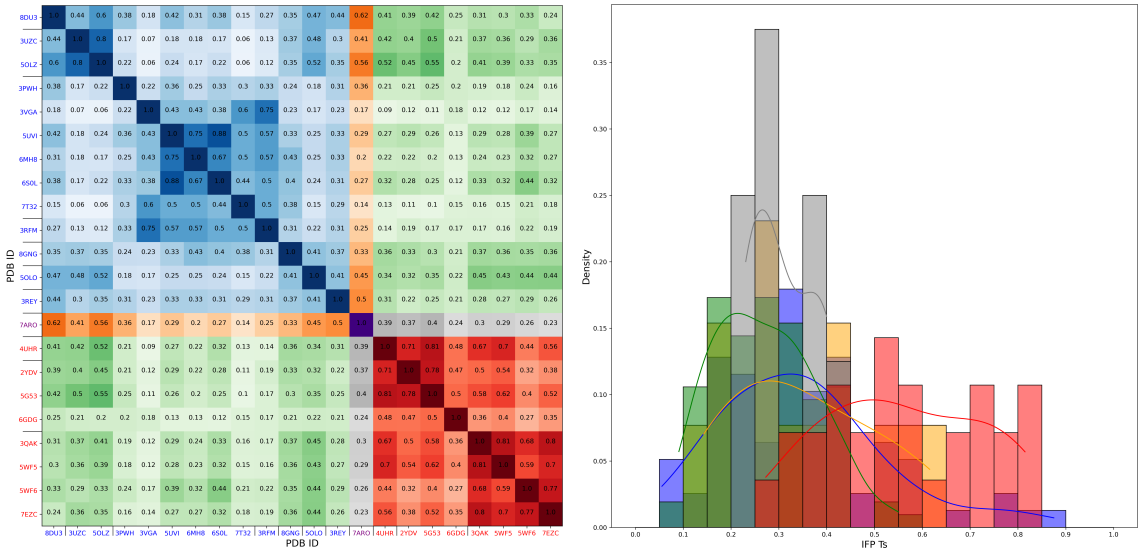


(A) **(B)**

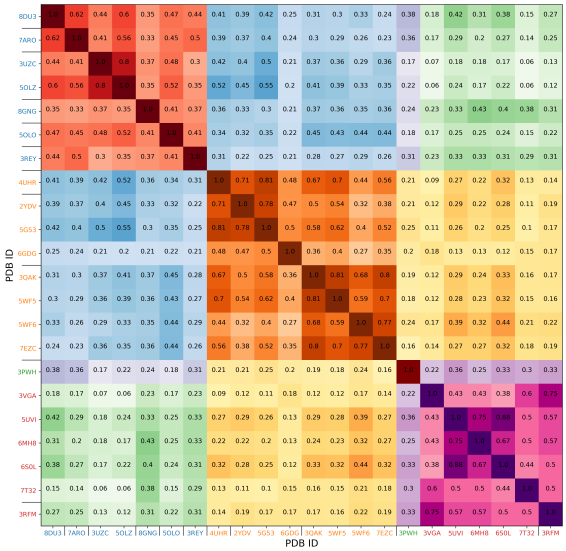


(C)

Figure A.1.9: (A) Pairwise Tanimoto similarity between IFPs extracted from crystallographic structures of AA2AR. Blue labels indicate antagonist bound inactive state structures, purple labels indicate agonist bound inactive state structures, and red labels indicate active state structures. The same colors in the plot inddicate interaction similarity between structures in the same conformation. Regions in orange indicate the similairty between agonist and antagonist binding modes from inactive state structures, grey between agonists bound to active and inactive state structures, and green between agonists and antagonists. (B) Normalized distribution of pairwise similarities between different binding modes. The color scheme is the same as for plot A. (C) Results of structures clustering based on IFP similarity.



(A) **(B)**



(C)

Figure A.1.10: (A) Pairwise Tanimoto similarity between IFPs extracted from crystallographic structures of ACM2. Blue labels indicate inactive state structures, and red labels indicate active state structures. The same colors in the plot indicate interaction similarity between structures in the same conformation. The region in green indicates interaction similarity between agonists and antagonists. (B) Normalized distribution of pairwise similarities between different binding modes. The color scheme is the same as for plot A. (C) Results of structures clustering based on IFP similarity.

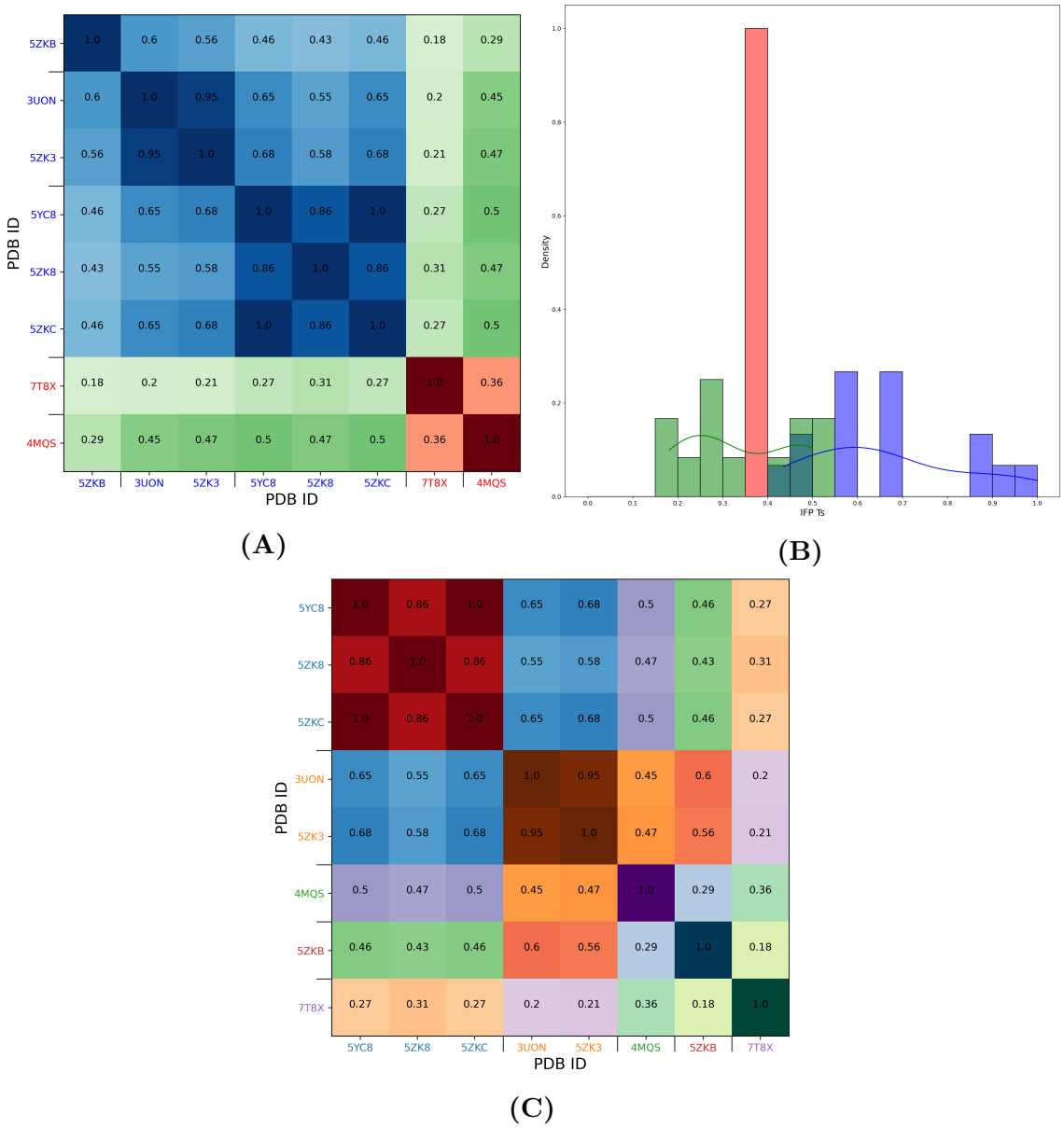
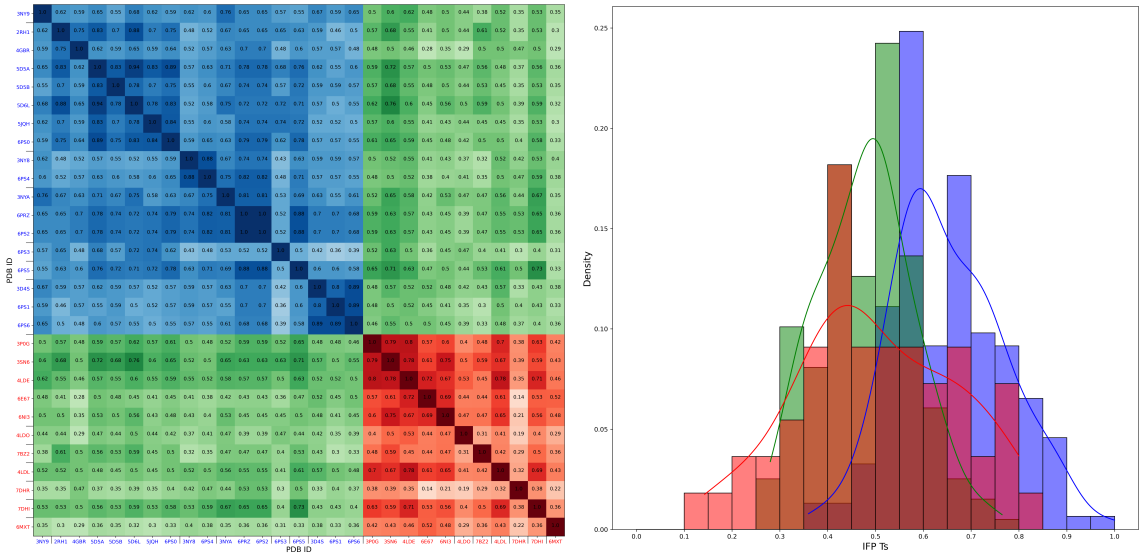
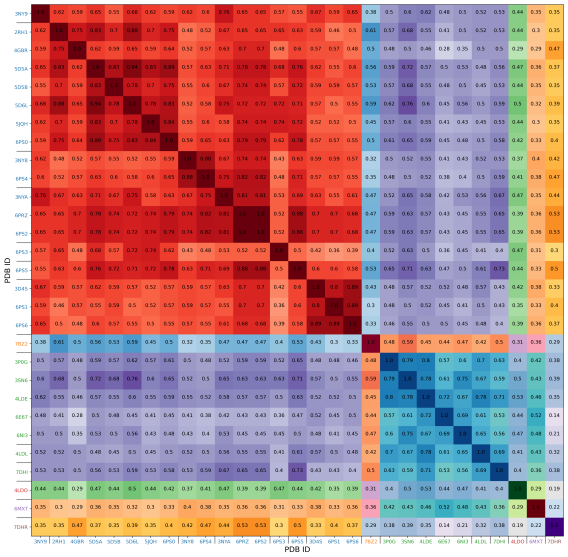


Figure A.1.11: (A) Pairwise Tanimoto similarity between IFPs extracted from crystallographic structures of ADRB2. Blue labels indicate inactive state structures, and red labels indicate active state structures. The same colors in the plot indicate interaction similarity between structures in the same conformation. The region in green indicates interaction similarity between agonists and antagonists. (B) Normalized distribution of pairwise similarities between different binding modes. The color scheme is the same as for plot A. (C) Results of structures clustering based on IFP similarity.

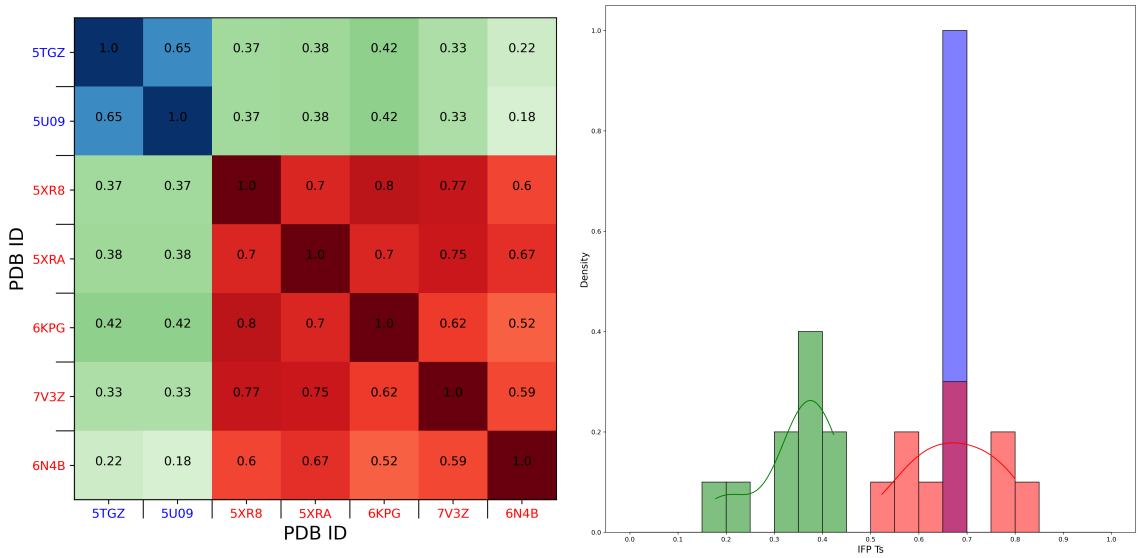


(A) **(B)**

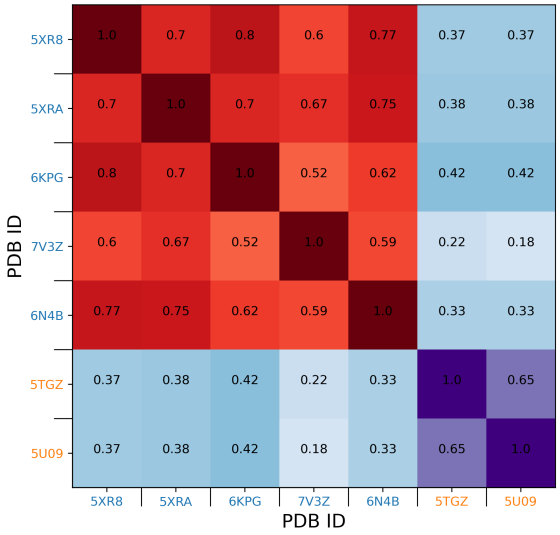


(C)

Figure A.1.12: (A) Pairwise Tanimoto similarity between IFPs extracted from crystallographic structures of CNR1. Blue labels indicate inactive state structures, and red labels indicate active state structures. The same colors in the plot indicate interaction similarity between structures in the same conformation. The region in green indicates interaction similarity between agonists and antagonists. (B) Normalized distribution of pairwise similarities between different binding modes. The color scheme is the same as for plot A. (C) Results of structures clustering based on IFP similarity.

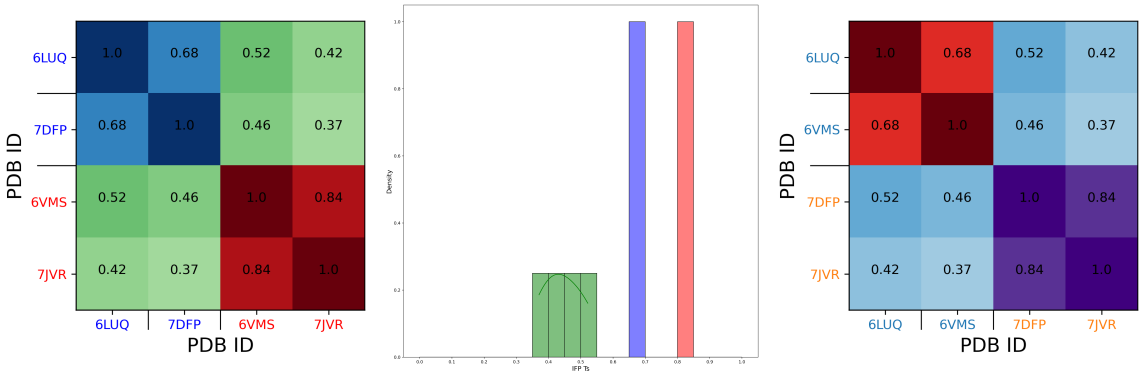


(A) **(B)**



(C)

Figure A.1.13: (A) Pairwise Tanimoto similarity between IFPs extracted from crystallographic structures of DRD2. Blue labels indicate inactive state structures, and red labels indicate active state structures. The same colors in the plot indicate interaction similarity between structures in the same conformation. The region in green indicates interaction similarity between agonists and antagonists. (B) Normalized distribution of pairwise similarities between different binding modes. The color scheme is the same as for plot A. (C) Results of structures clustering based on IFP similarity.



(A) (B) (C)

Figure A.1.14: Pairwise Tanimoto similarity between IFPs extracted from crystallographic structures of OPRK. Blue labels indicate inactive state structures, and red labels indicate active state structures. The same colors in the plot indicate interaction similarity between structures in the same conformation. The region in green indicates interaction similarity between agonists and antagonists.

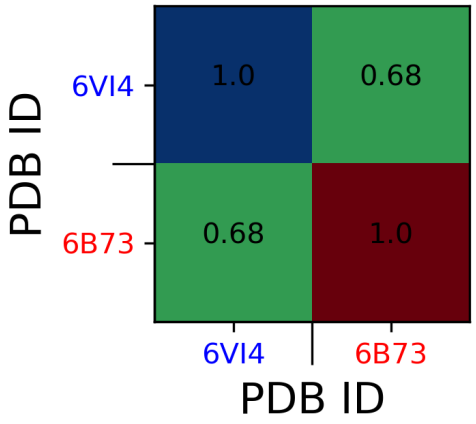
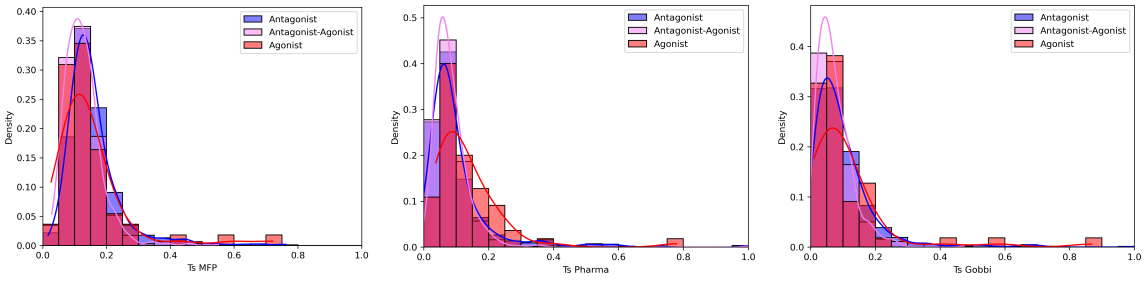
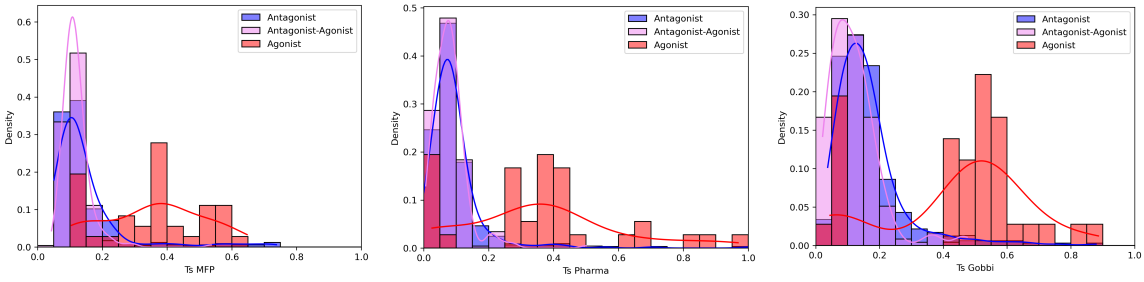


Figure A.1.15: Distribution of pairwise similarity between agonists, antagonists, and agonists and antagonists of 5HT2A using different fingerprint representation: (A) Morgan fingerprints (r=2) (B) Pharmacophoric fingerprints (C) Gobbi pharmacophoric fingerprints.



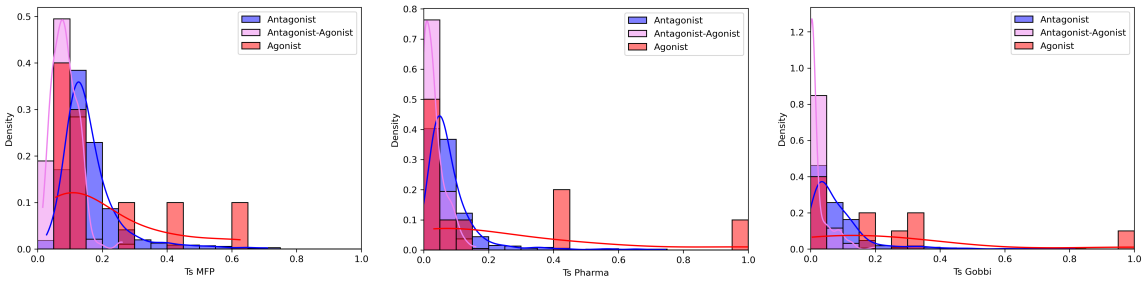
(A) **(B)** **(C)**

Figure A.1.16: Distribution of pairwise similarity between agonists, antagonists, and agonists and antagonists of AA2AR using different fingerprint representation: (A) Morgan fingerprints (r=2) (B) Pharmacophoric fingerprints (C) Gobbi pharmacophoric fingerprints



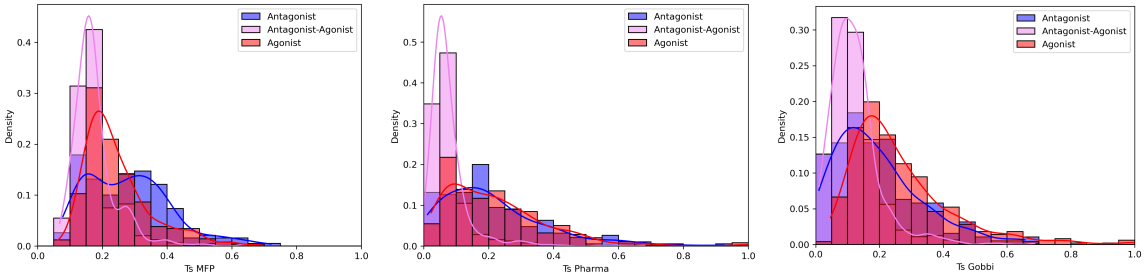
(A) **(B)** **(C)**

Figure A.1.17: Distribution of pairwise similarity between agonists, antagonists, and agonists and antagonists of ACM2 using different fingerprint representation: (A) Morgan fingerprints (r=2) (B) Pharmacophoric fingerprints (C) Gobbi pharmacophoric fingerprints



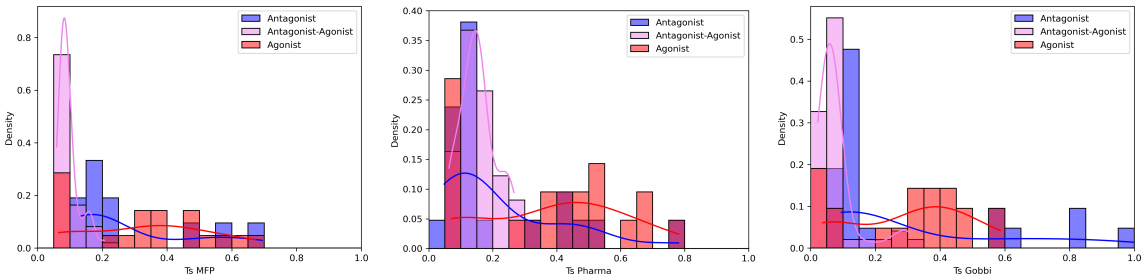
(A) **(B)** **(C)**

Figure A.1.18: Distribution of pairwise similarity between agonists, antagonists, and agonists and antagonists of ADRB2 using different fingerprint representation: (A) Morgan fingerprints (r=2) (B) Pharmacophoric fingerprints (C) Gobbi pharmacophoric fingerprints



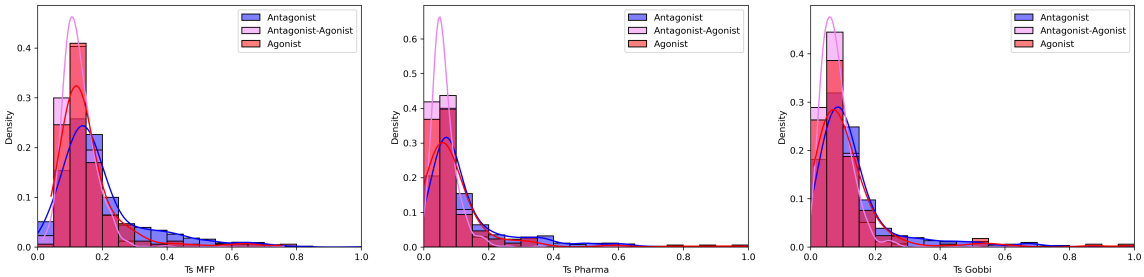
(A) **(B)** **(C)**

Figure A.1.19: Distribution of pairwise similarity between agonists, antagonists, and agonists and antagonists of CNR1 using different fingerprint representation: (A) Morgan fingerprints (r=2) (B) Pharmacophoric fingerprints (C) Gobbi pharmacophoric fingerprints



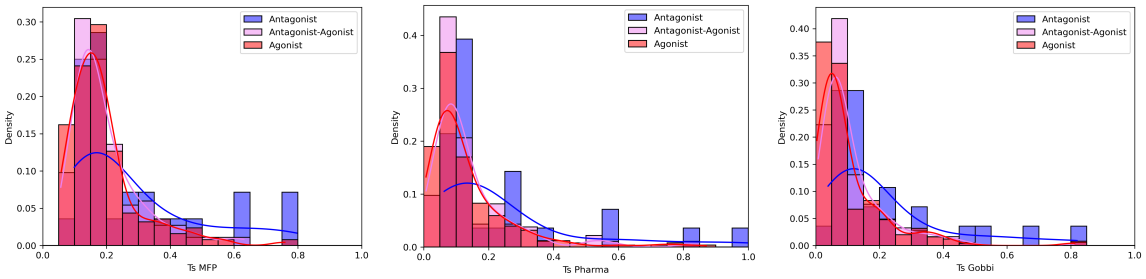
(A) **(B)** **(C)**

Figure A.1.20: Distribution of pairwise similarity between agonists, antagonists, and agonists and antagonists of DRD2 using different fingerprint representation: (A) Morgan fingerprints (r=2) (B) Pharmacophoric fingerprints (C) Gobbi pharmacophoric fingerprints



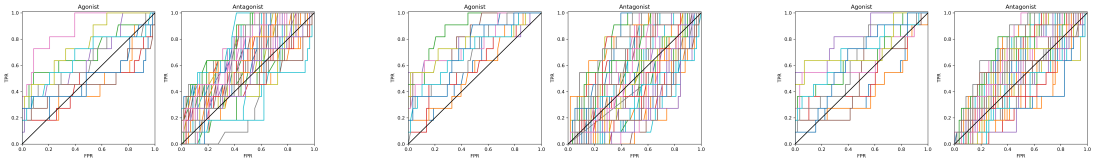
(A) **(B)** **(C)**

Figure A.1.21: Distribution of pairwise similarity between agonists, antagonists, and agonists and antagonists of OPRK using different fingerprint representation: (A) Morgan fingerprints (r=2) (B) Pharmacophoric fingerprints (C) Gobbi pharmacophoric fingerprints



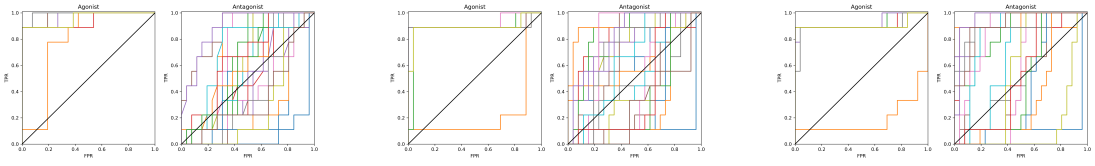
(A) **(B)** **(C)**

Figure A.1.22: ROCs obtained by molecular or pharmacophoric similarity using as reference an agonist or antagonist ligand of 5HT2A. Molecules with the same efficacy as the reference were selected as the positive class.



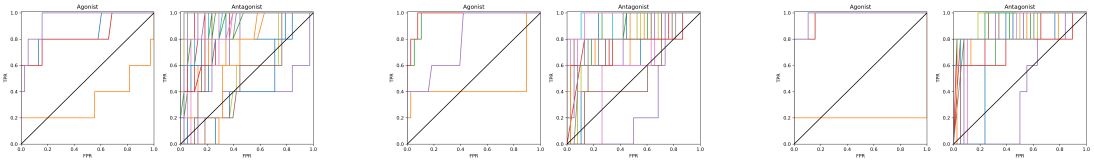
(A) MFP **(B)** Pharma **(C)** Gobbi

Figure A.1.23: ROCs obtained by molecular or pharmacophoric similarity using as reference an agonist or antagonist ligand of AA2AR. Molecules with the same efficacy as the reference were selected as the positive class.



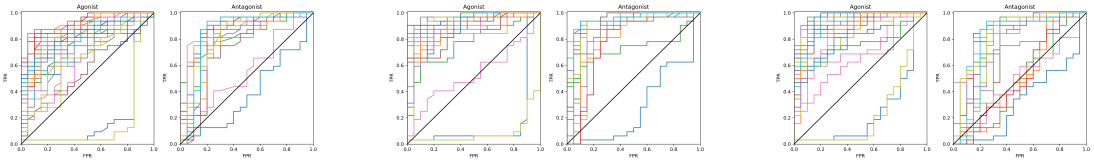
(A) MFP **(B)** Pharma **(C)** Gobbi

Figure A.1.24: ROCs obtained by molecular or pharmacophoric similarity using as reference an agonist or antagonist ligand of ACM2. Molecules with the same efficacy as the reference were selected as the positive class.



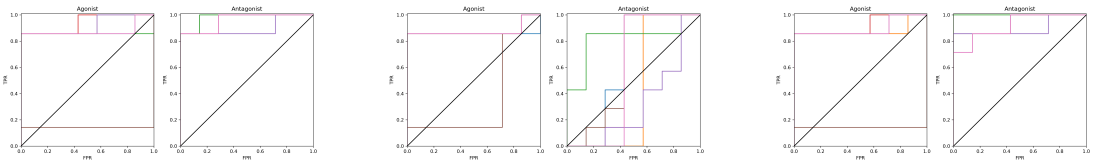
(A) MFP **(B)** Pharma **(C)** Gobbi

Figure A.1.25: ROCs obtained by molecular or pharmacophoric similarity using as reference an agonist or antagonist ligand of ADRB2. Molecules with the same efficacy as the reference were selected as the positive class.



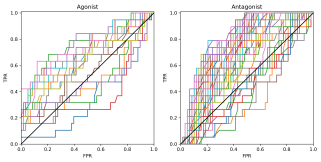
(A) MFP **(B)** Pharma **(C)** Gobbi

Figure A.1.26: ROCs obtained by molecular or pharmacophoric similarity using as reference an agonist or antagonist ligand of CNR1. Molecules with the same efficacy as the reference were selected as the positive class.

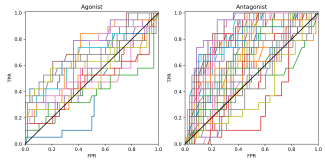


(A) MFP **(B)** Pharma **(C)** Gobbi

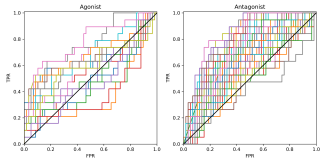
Figure A.1.27: ROCs obtained by molecular or pharmacophoric similarity using as reference an agonist or antagonist ligand of DRD2. Molecules with the same efficacy as the reference were selected as the positive class.



(A) MFP2

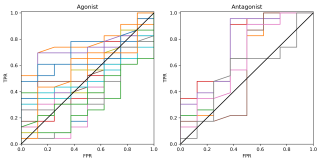


(B) Pharma

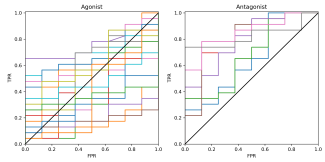


(C) Gobbi

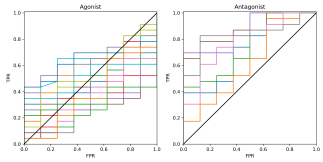
Figure A.1.28: ROCs obtained by molecular or pharmacophoric similarity using as reference an agonist or antagonist ligand of OPRK. Molecules with the same efficacy as the reference were selected as the positive class.



(A) MFP2

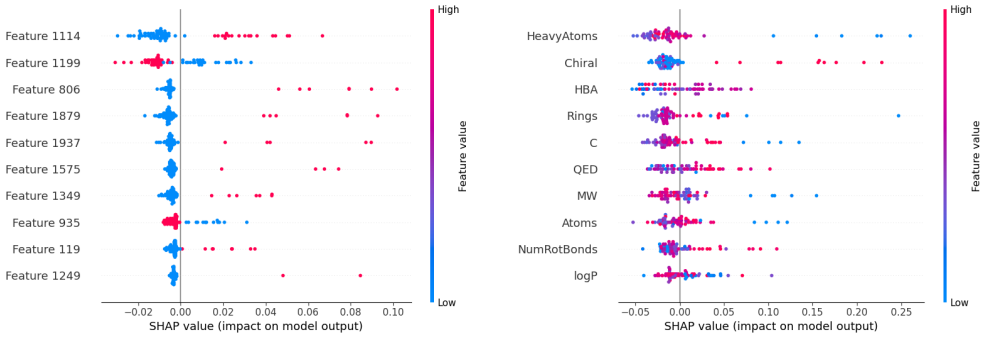


(B) Pharma

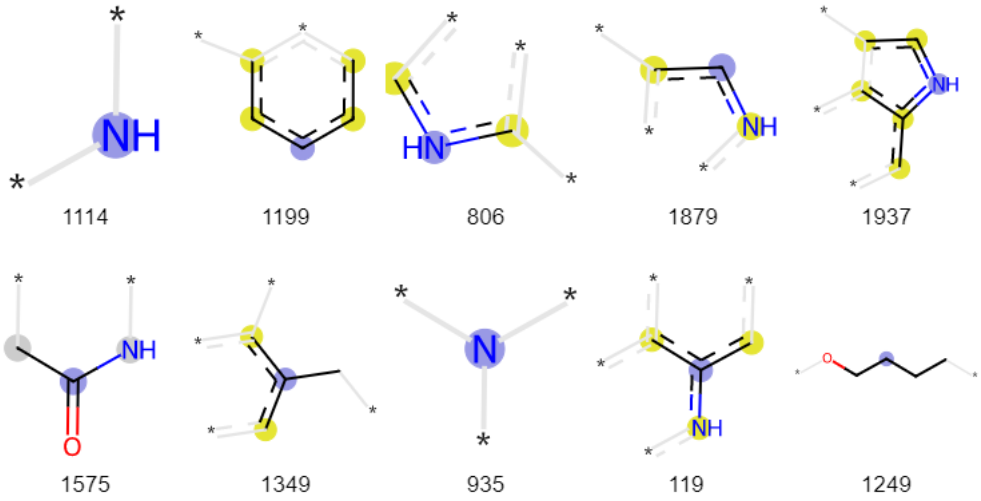


(C) Gobbi

Figure A.1.29: Analysis of the efficacy-classification models trained on 5HT2A ligands. (A) Results from SHAP analysis performed in LOO-CV for the random forest rained of Morgan fingerprints (radius=2). (B) Results from SHAP analysis performed on the random forest model trained on physicochemical properties. (C)Important features for efficacy classification identified by SHAP analysis.

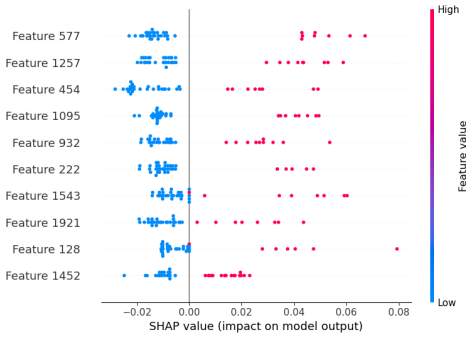


(A) Morgan fingerprints r=2 **(B)** Physicochemical properties

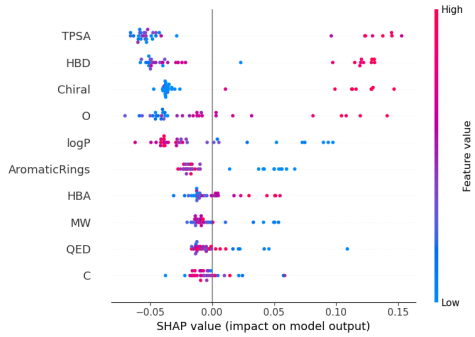


(C) MFP2 features

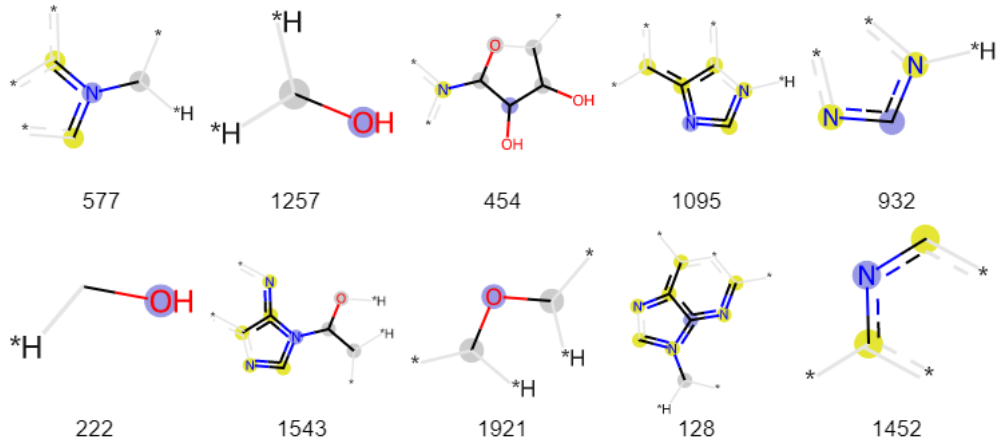
Figure A.1.30: Analysis of the efficacy-classification models trained on AA2AR ligands. (A) Results from SHAP analysis performed in LOO-CV for the random forest rained of Morgan fingerprints (radius=2). (B) Results from SHAP analysis performed on the random forest model trained on physicochemical properties. (C)Important features for efficacy classification identified by SHAP analysis.



(A) Morgan fingerprints r=2

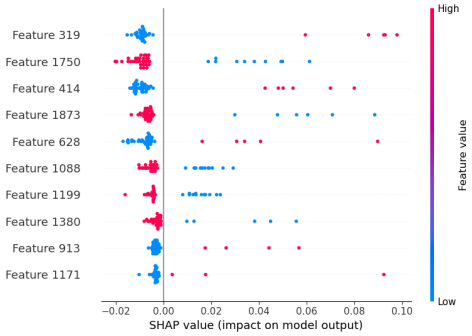


(B) Physicochemical properties

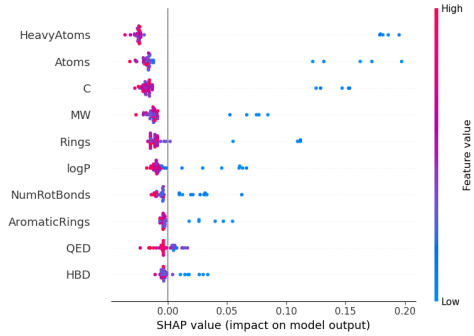


(C) MFP2 features

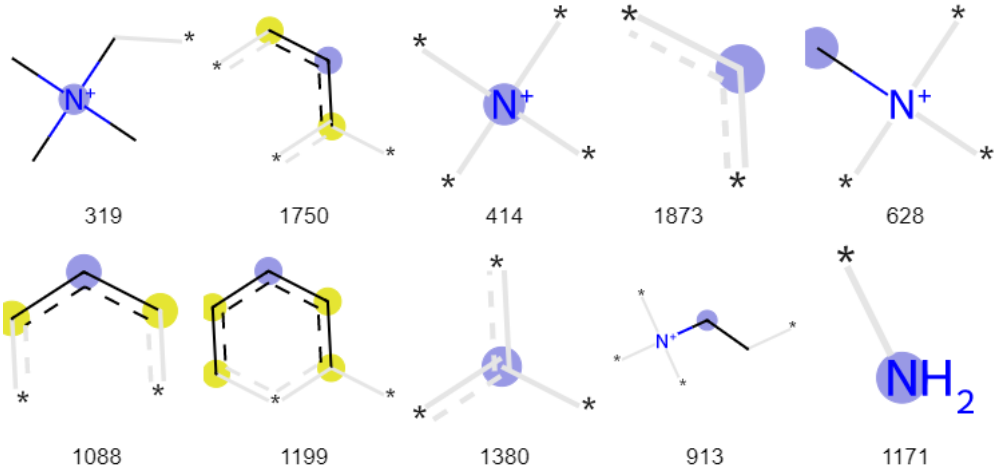
Figure A.1.31: Analysis of the efficacy-classification models trained on ACM2 ligands. (A) Results from SHAP analysis performed in LOO-CV for the random forest rained of Morgan fingerprints (radius=2). (B) Results from SHAP analysis performed on the random forest model trained on physicochemical properties. (C)Important features for efficacy classification identified by SHAP analysis.



(A) Morgan fingerprints r=2

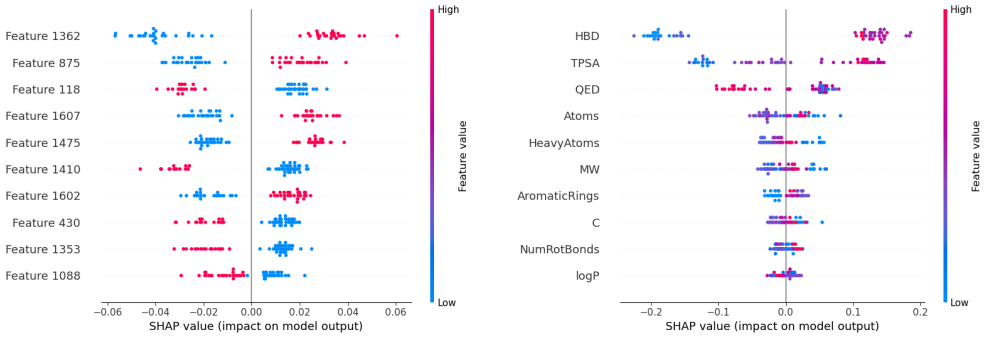


(B) Physicochemical properties

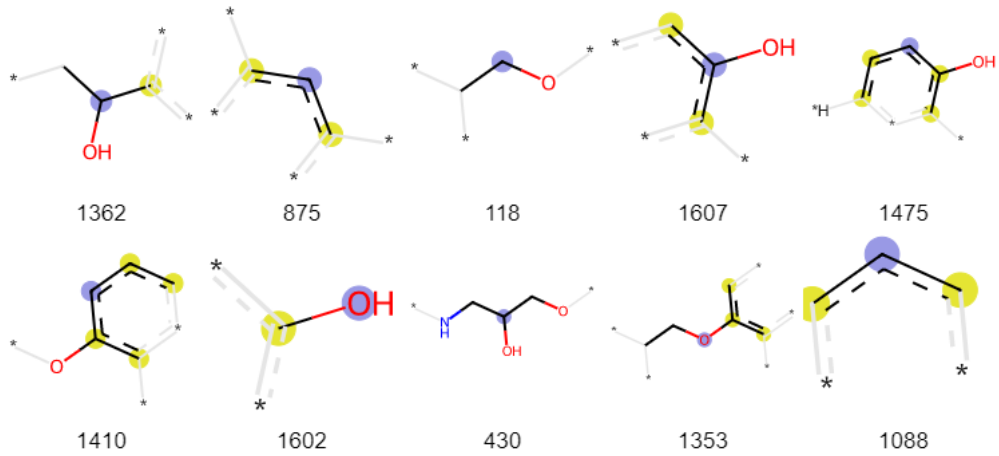


(C) MFP2 features

Figure A.1.32: Analysis of the efficacy-classification models trained on ADRB2 ligands. (A) Results from SHAP analysis performed in LOO-CV for the random forest rained of Morgan fingerprints (radius=2). (B) Results from SHAP analysis performed on the random forest model trained on physicochemical properties. (C)Important features for efficacy classification identified by SHAP analysis.

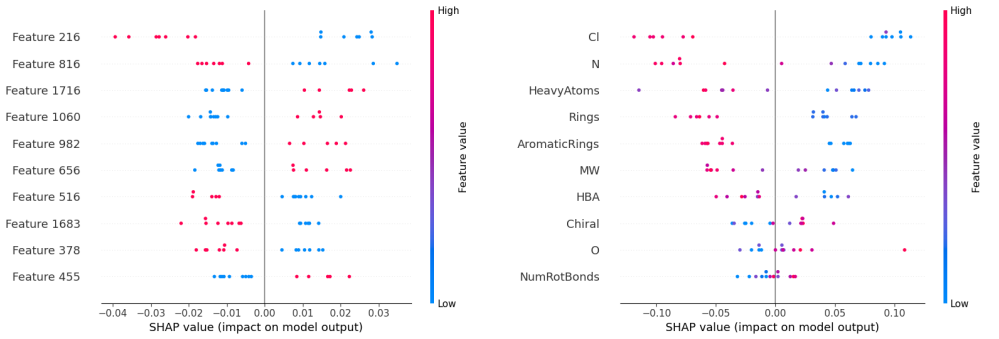


(A) Morgan fingerprints r=2 **(B)** Physicochemical properties

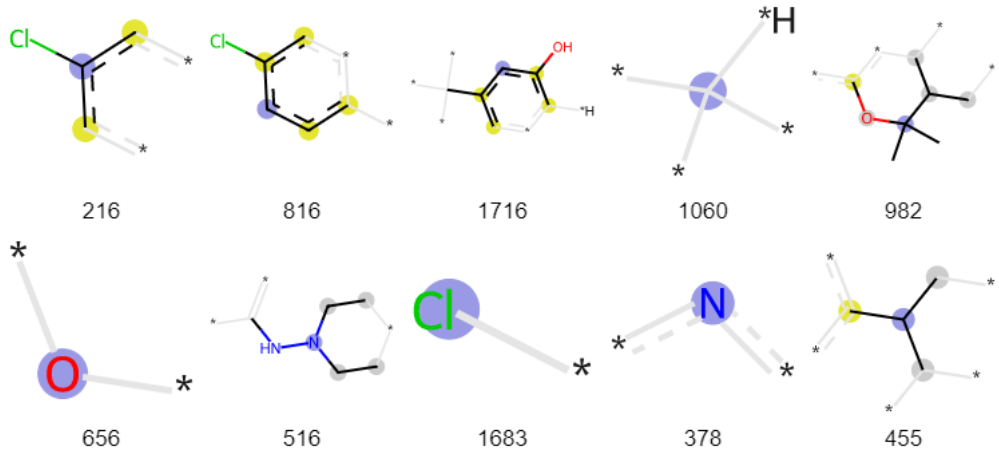


(C) MFP2 features

Figure A.1.33: Analysis of the efficacy-classification models trained on CNR1 ligands. (A) Results from SHAP analysis performed in LOO-CV for the random forest rained of Morgan fingerprints (radius=2). (B) Results from SHAP analysis performed on the random forest model trained on physicochemical properties. (C)Important features for efficacy classification identified by SHAP analysis.

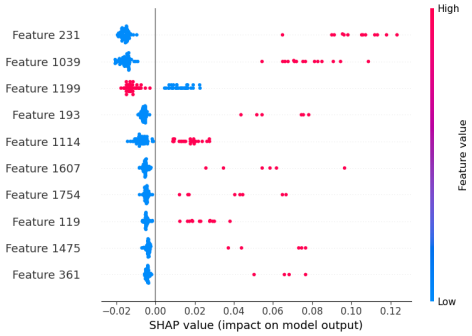


(A) Morgan fingerprints r=2 **(B)** Physicochemical properties

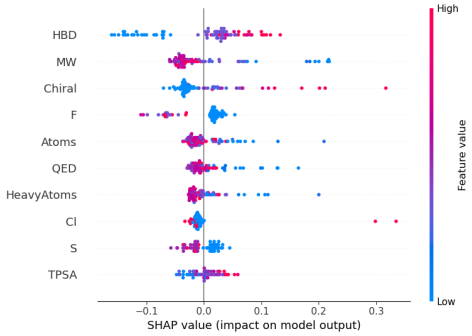


(C) MFP2 features

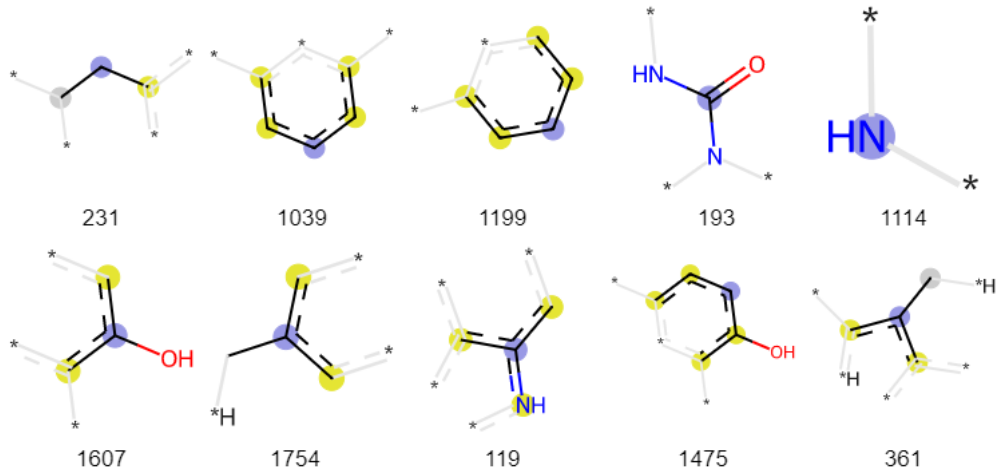
Figure A.1.34: Analysis of the efficacy-classification models trained on DRD2 ligands. (A) Results from SHAP analysis performed in LOO-CV for the random forest rained of Morgan fingerprints (radius=2). (B) Results from SHAP analysis performed on the random forest model trained on physicochemical properties. (C)Important features for efficacy classification identified by SHAP analysis.



(A) Morgan fingerprints r=2

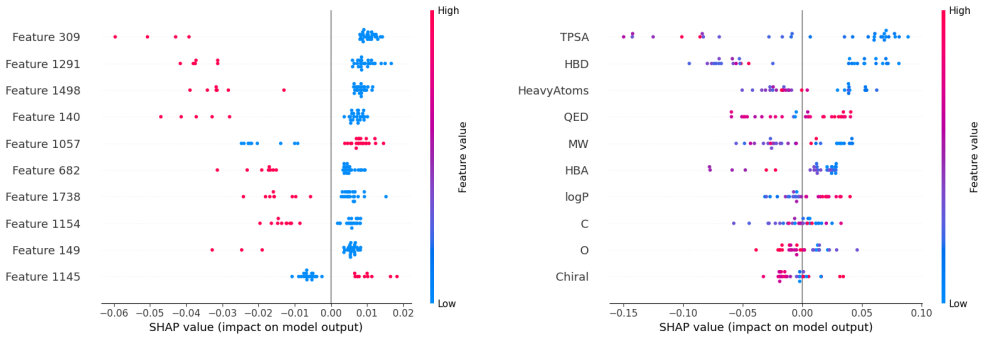


(B) Physicochemical properties

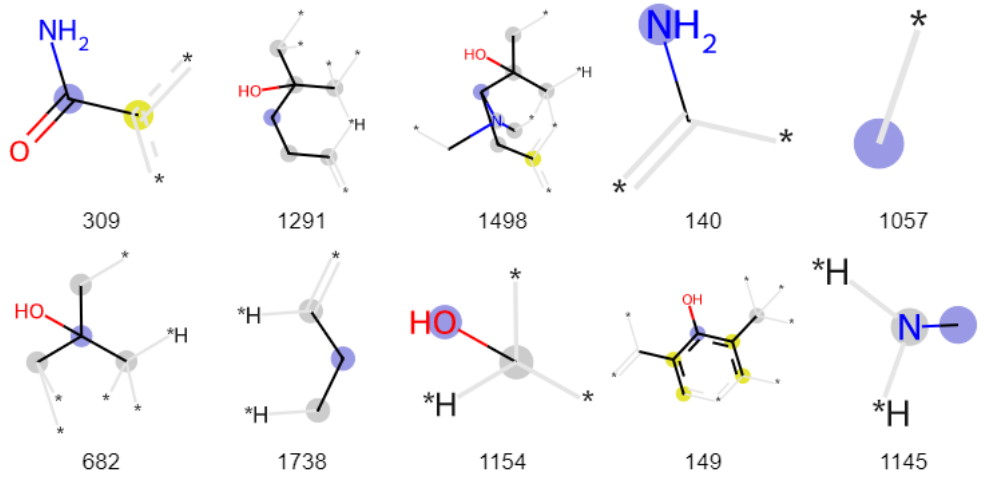


(C) MFP2 features

Figure A.1.35: Analysis of the efficacy-classification models trained on OPRK ligands. (A) Results from SHAP analysis performed in LOO-CV for the random forest rained of Morgan fingerprints (radius=2). (B) Results from SHAP analysis performed on the random forest model trained on physicochemical properties. (C)Important features for efficacy classification identified by SHAP analysis.

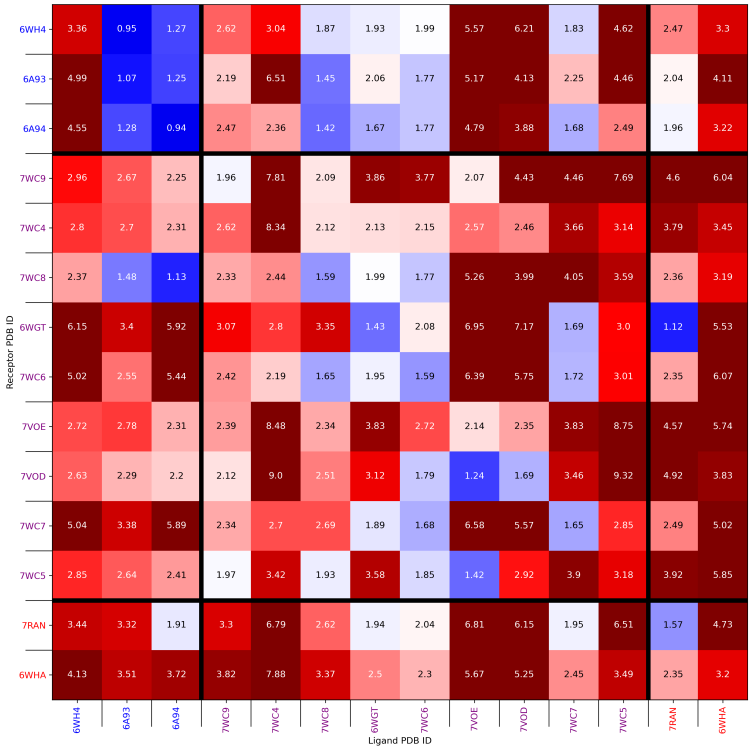


(A) Morgan fingerprints r=2 **(B)** Physicochemical properties

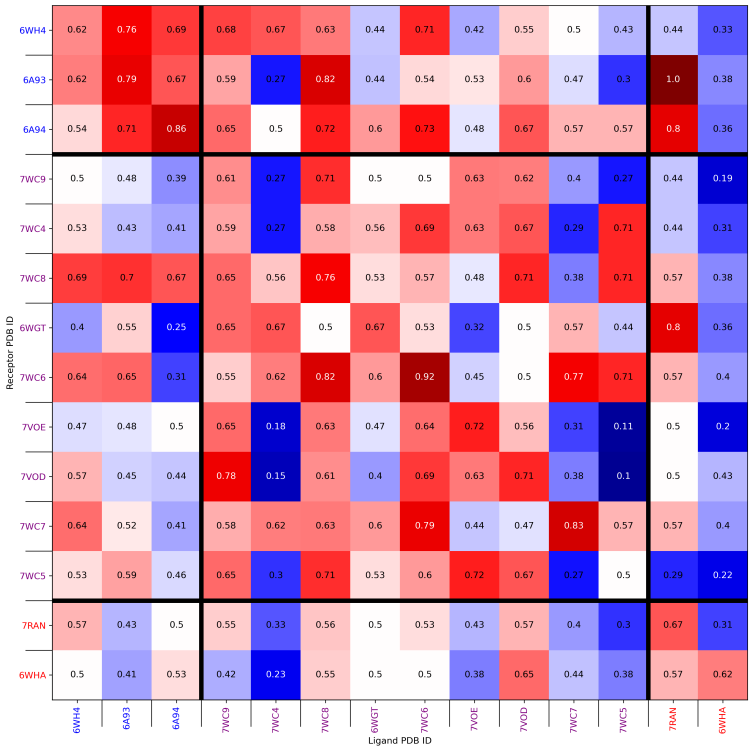


(C) MFP2 features

Figure A.1.36: Crossdocking on 5HT2A structures RMSD(Å) and IFP similarity for the best performing pose based on the calculated quality measure. Blue label indicates antagonist-bound inactive state structures, purple agonist-bound inactive state structures, red agonist-bound active state structures. Black lines are used to separate ensembles of structures in different conformational states.

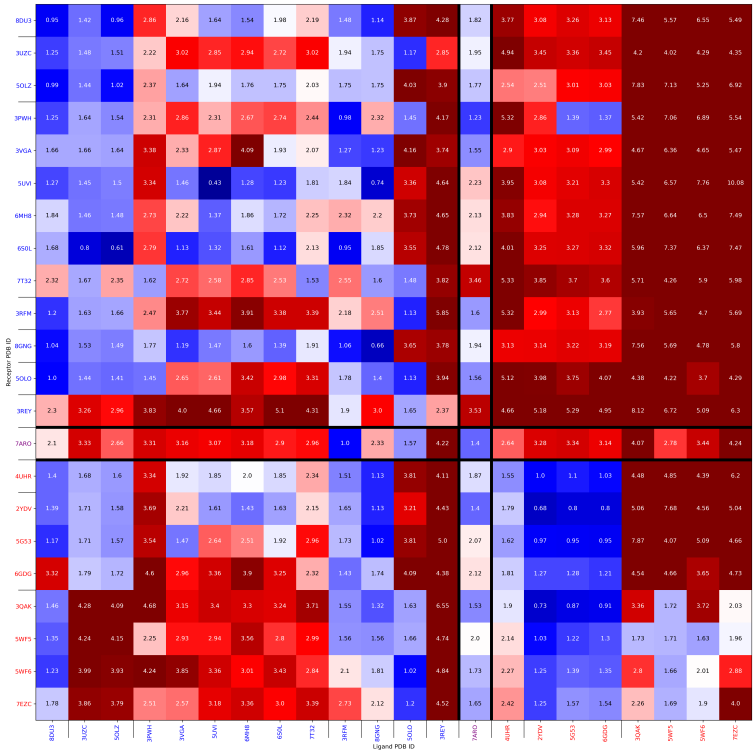


(A) Cross-docking RMSD

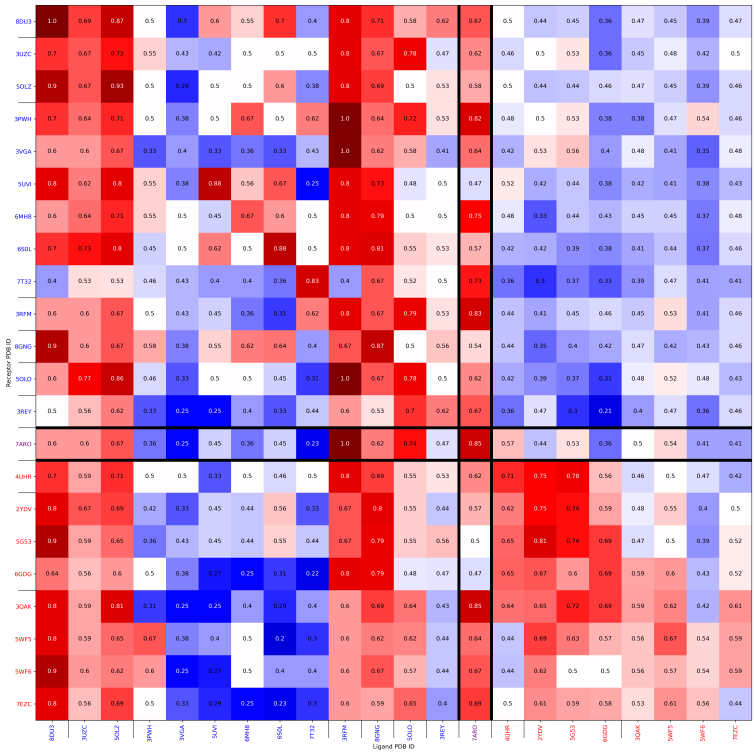


(B) Cross-docking IFP similarity

Figure A.1.37: Crossdocking on AA2AR structures RMSD(Å) and IFP similarity for the best performing pose based on the calculated quality measure. Blue label indicates antagonist-bound inactive state structures, purple agonist-bound inactive state structures, red agonist-bound active state structures. Black lines are used to separate ensembles of structures in different conformational states.

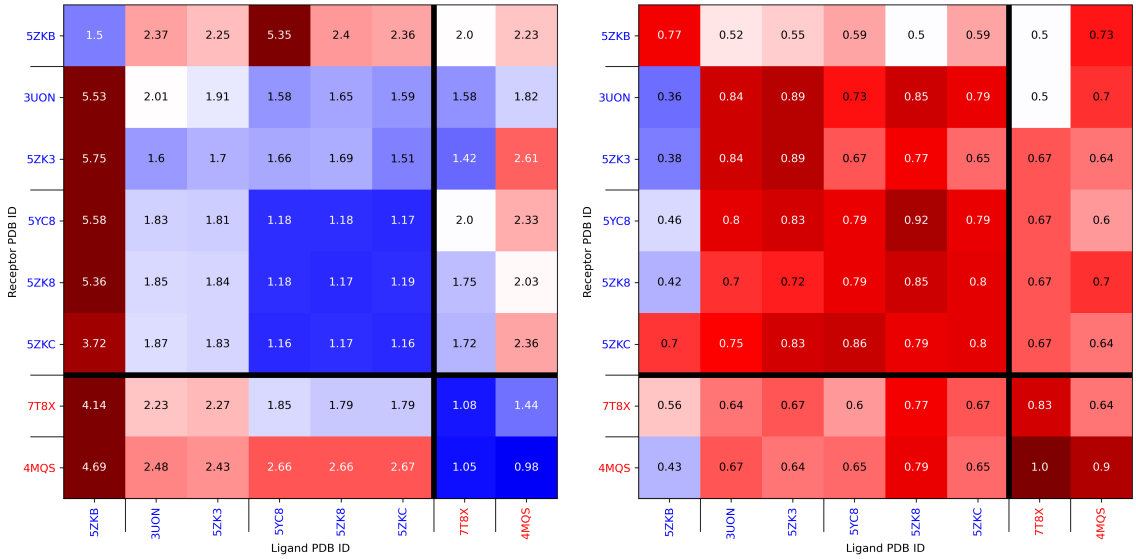


(A) Cross-docking RMSD



(B) Cross-docking IFP similarity

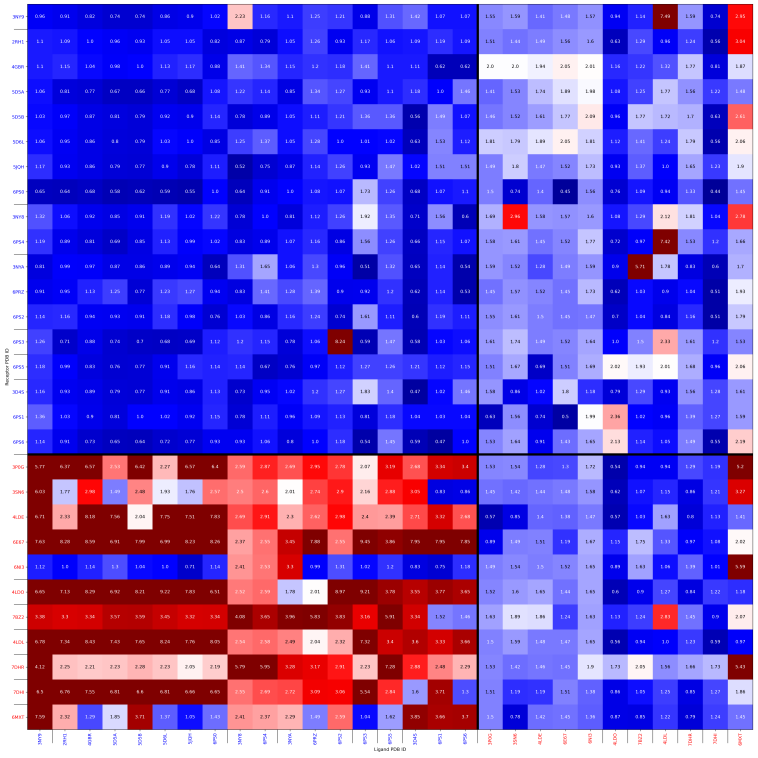
Figure A.1.38: Crossdocking on ACM2 structures RMSD(Å) and IFP similarity for the best performing pose based on the calculated quality measure. Blue label indicates antagonist-bound inactive state structures, purple agonist-bound inactive state structures, red agonist-bound active state structures. Black lines are used to separate ensembles of structures in different conformational states.



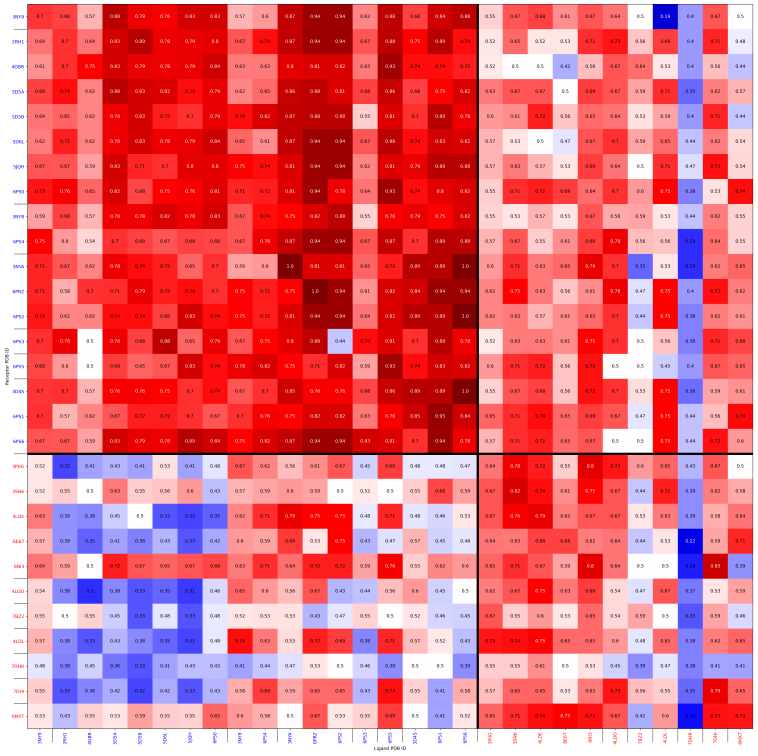
(A) Cross-docking RMSD

(B) Cross-docking IFP similarity

Figure A.1.39: Crossdocking on ADRB2 structures RMSD(Å) and IFP similarity for the best performing pose based on the calculated quality measure. Blue label indicates antagonist-bound inactive state structures, purple agonist-bound inactive state structures, red agonist-bound active state structures. Black lines are used to separate ensembles of structures in different conformational states.

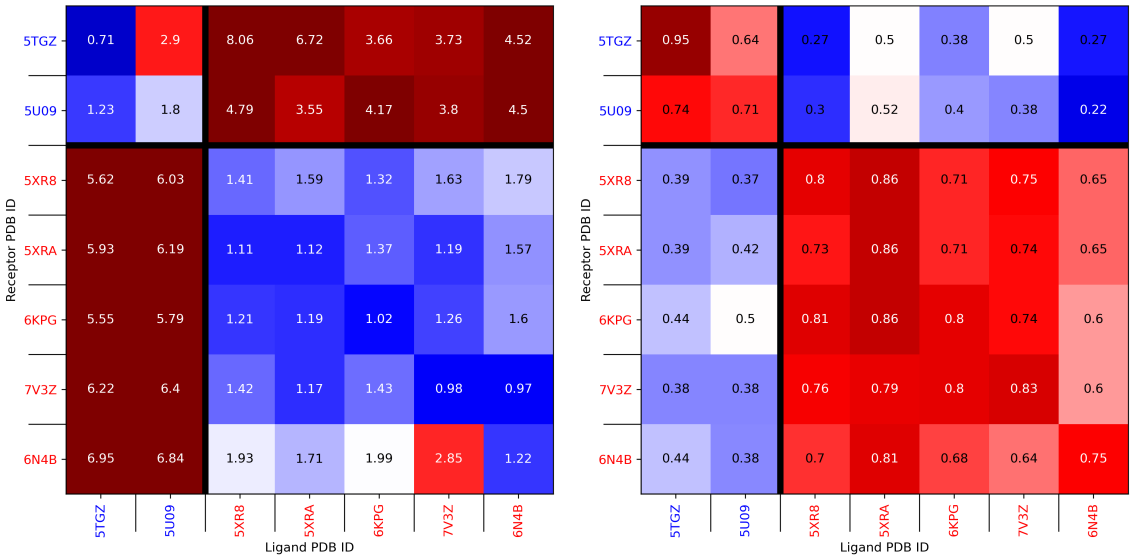


(A) Cross-docking RMSD



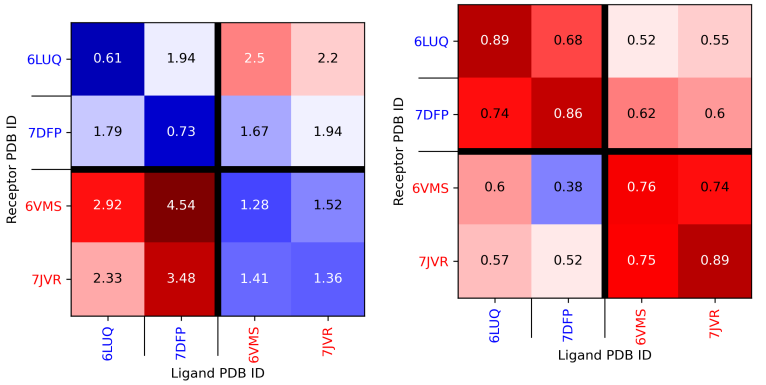
(B) Cross-docking IFP similarity

Figure A.1.40: Crossdocking on CNR1 structures RMSD(Å) and IFP similarity for the best performing pose based on the calculated quality measure. Blue label indicates antagonist-bound inactive state structures, purple agonist-bound inactive state structures, red agonist-bound active state structures. Black lines are used to separate ensembles of structures in different conformational states.



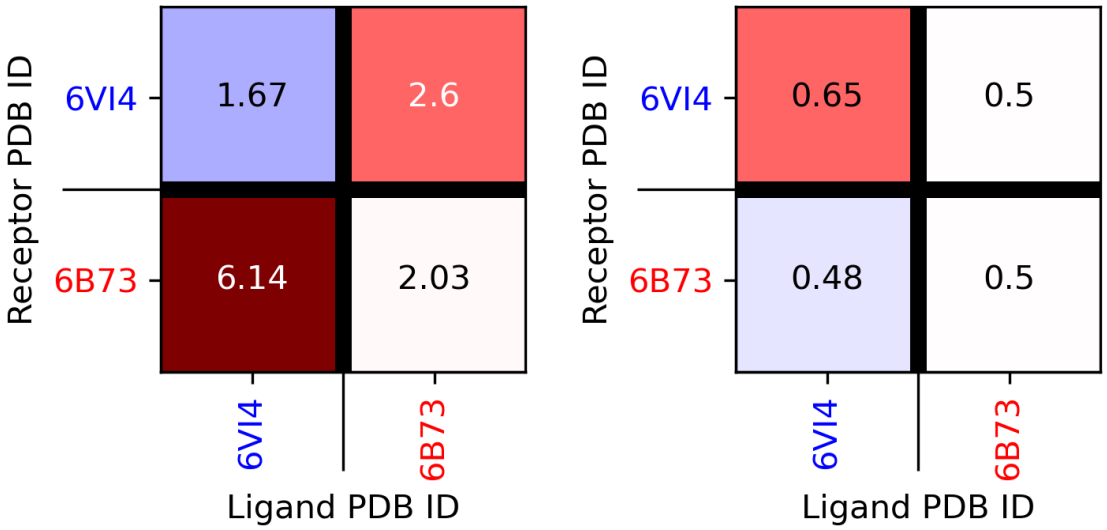
(A) Cross-docking RMSD **(B)** Cross-docking IFP similarity

Figure A.1.41: Crossdocking on DRD2 structures RMSD(Å) and IFP similarity for the best performing pose based on the calculated quality measure. Blue label indicates antagonist-bound inactive state structures, purple agonist-bound inactive state structures, red agonist-bound active state structures. Black lines are used to separate ensembles of structures in different conformational states.



(A) Cross-docking RMSD **(B)** Cross-docking IFP similarity

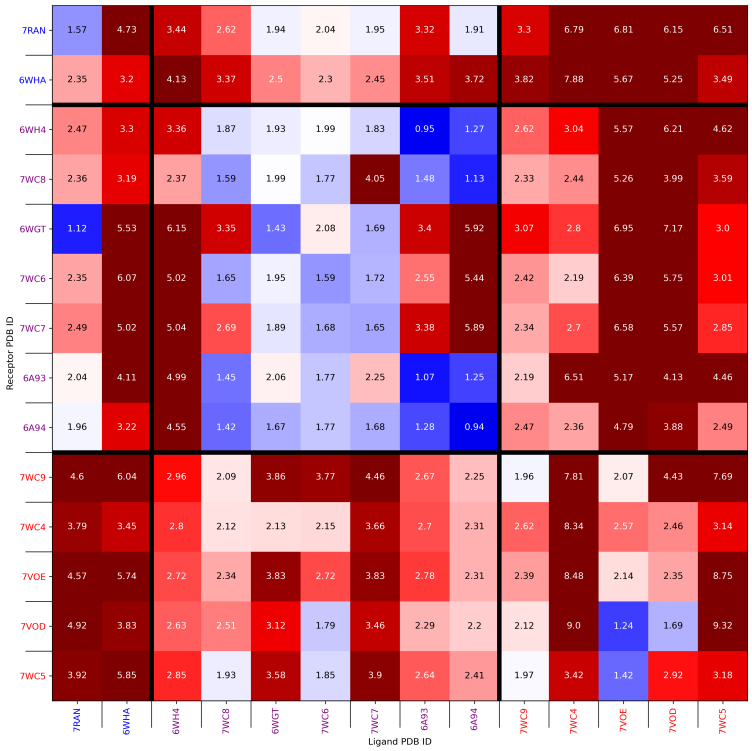
Figure A.1.42: Crossdocking on OPRK structures RMSD(Å) and IFP similarity for the best performing pose based on the calculated quality measure. Blue label indicates antagonist-bound inactive state structures, purple agonist-bound inactive state structures, red agonist-bound active state structures. Black lines are used to separate ensembles of structures in different conformational states.



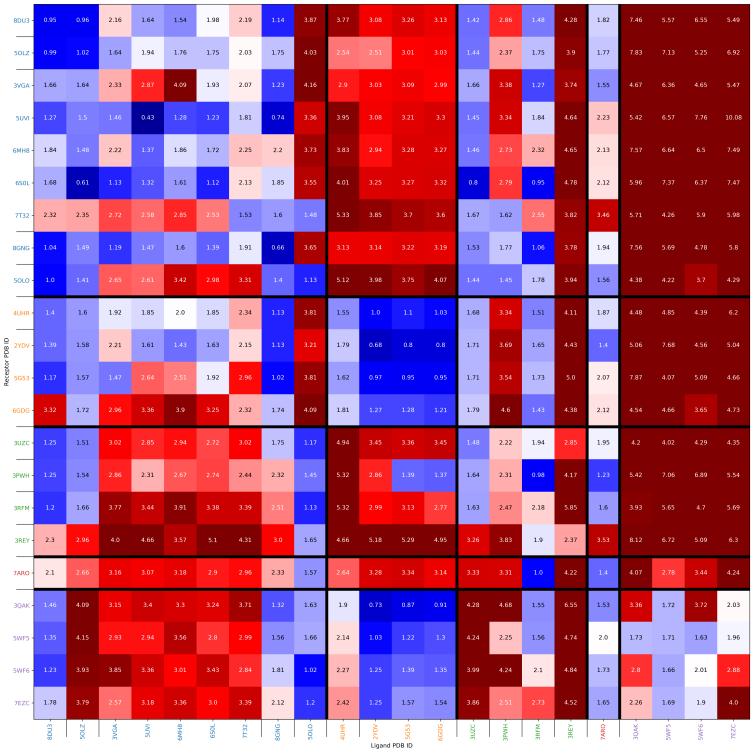
(A) Cross-docking RMSD

(B) Cross-docking IFP similarity

Figure A.1.43: RMSD(Å) from crossdocking on 5HT2A and AA2AR structures clustered based on binding site clustering. Black lines are used to separate ensembles of structures in different conformational states.



(A) 5HT2A



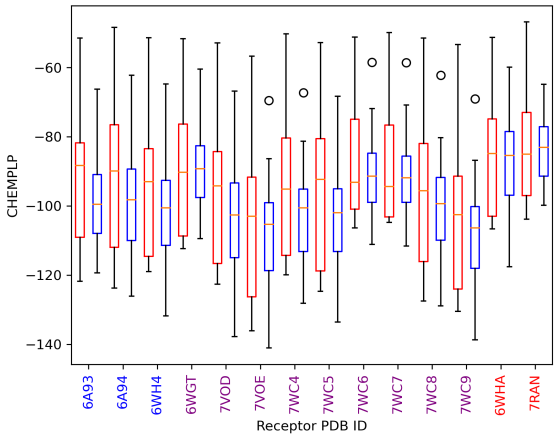
(B) AA2AR

A.2 Section 2.2 supporting information

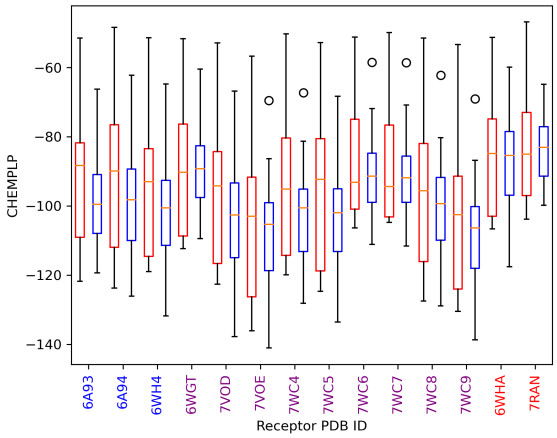
Table A.2.1: Quality of the decoys generated by DeepCoy according to the generated quality values

Target	AUC ROC - 1NN	AUC ROC - RF	DOE score	LADS score	Doppelganger score mean	Doppelganger score max
5HT2A	0.54	0.74	$3.95 \cdot 10^{-2}$	0.45	0.26	0.34
AA2AR	0.55	0.56	$4.6 \cdot 10^{-2}$	0.33	0.22	0.34
ACM2	0.54	0.65	$3.82 \cdot 10^{-2}$	0.46	0.25	0.35
ADRB2	0.53	0.55	$3.17 \cdot 10^{-2}$	0.48	0.28	0.35
CNR1	0.49	0.25	$4.06 \cdot 10^{-2}$	0.48	0.23	0.26
DRD2	0.57	0.79	$4.48 \cdot 10^{-2}$	0.44	0.27	0.4
OPRK	0.65	0.7	0.12	0.51	0.24	0.35

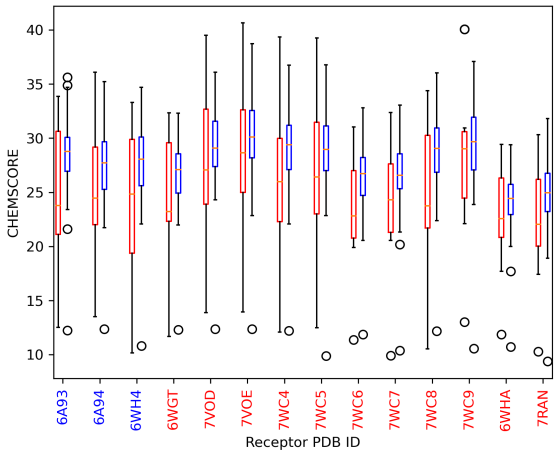
Figure A.2.1: Distribution of docking scores of 5HT2A agonists and antagonists on different receptor structures. PDB ID labels in blue indicate antagonist-bound inactive state structures, purple indicate agonist-bound inactive state structures, red indicate active state structures receptor. For CHEMPLP, NORM CHEMPLP lower score values indicate a higher affinity of the molecule, while for GOLDSCORE, CHEMSCORE and GOLD PLP higher scores are associated to a higher affinity.



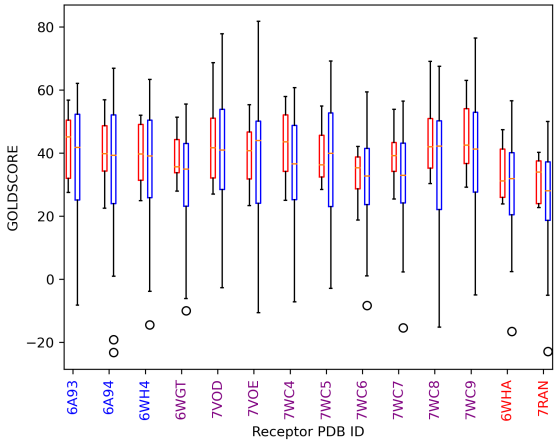
(A) CHEMPLP



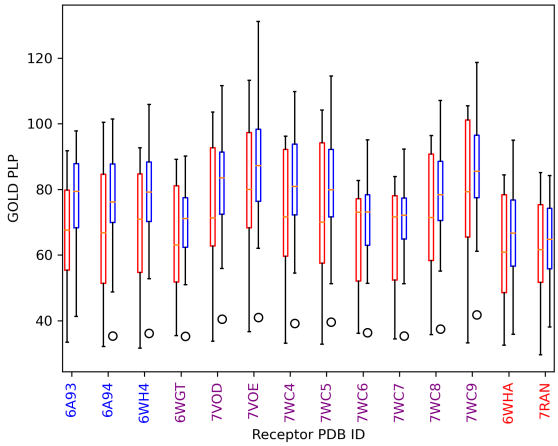
(B) NORMCHEMPLP



(C) CHEMSCORE

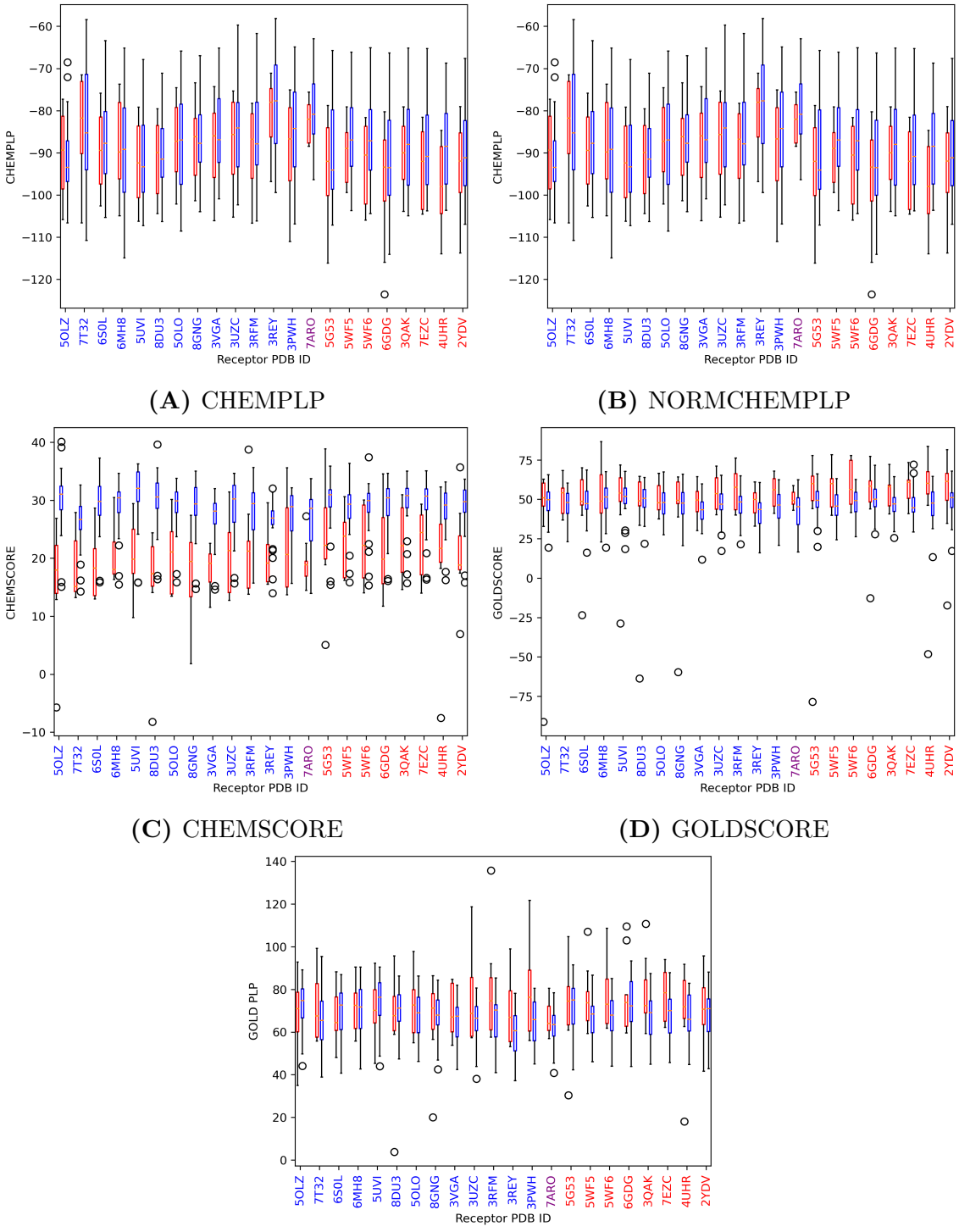


(D) GOLDSCORE



(E) GOLD PLP

Figure A.2.2: Distribution of docking scores of AA2AR agonists and antagonists on different receptor structures. PDB ID labels in blue indicate antagonist-bound inactive state structures, purple indicate agonist-bound inactive state structures, red indicate active state structures receptor. For CHEMPLP, NORM CHEMPLP lower score values indicate a higher affinity of the molecule, while for GOLDSCORE, CHEMSCORE and GOLD PLP higher scores are associated to a higher affinity.



(A) CHEMPLP

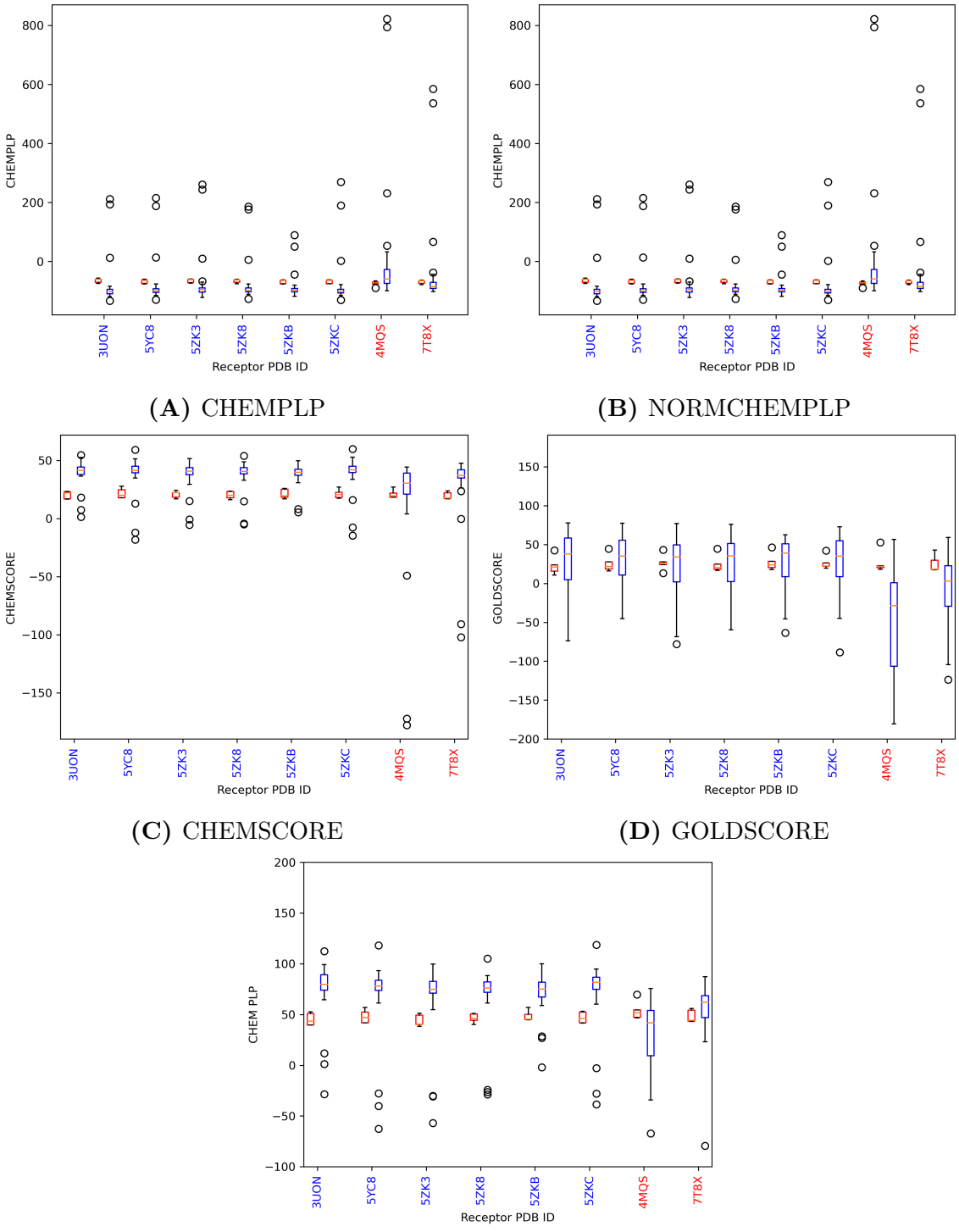
(B) NORMCHEMPLP

(C) CHEMSCORE

(D) GOLDSCORE

(E) GOLD PLP

Figure A.2.3: Distribution of docking scores of ACM2 agonists and antagonists on different receptor structures. PDB ID labels in blue indicate antagonist-bound inactive state structures, purple indicate agonist-bound inactive state structures, red indicate active state structures receptor. For CHEMPLP, NORM CHEMPLP lower score values indicate a higher affinity of the molecule, while for GOLDSCORE, CHEMSCORE and GOLD PLP higher scores are associated to a higher affinity.



(A) CHEMPLP

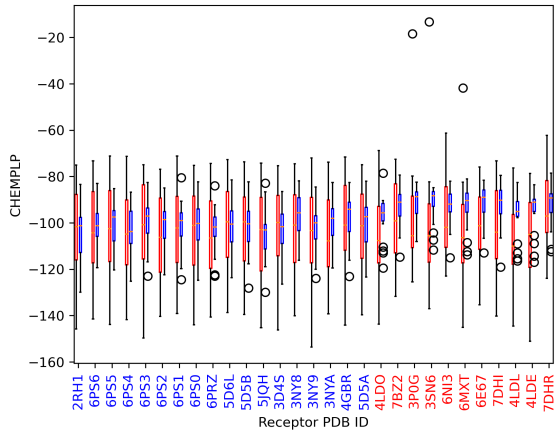
(B) NORMCHEMPLP

(C) CHEMSCORE

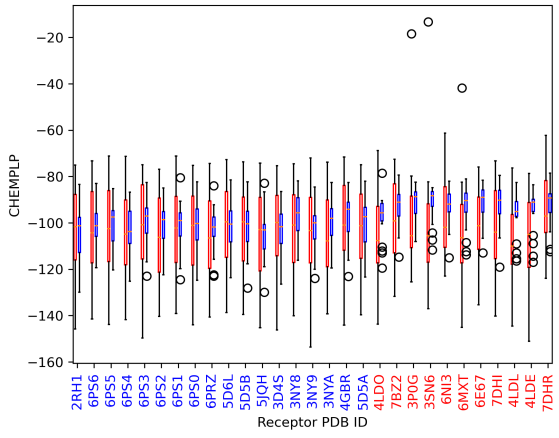
(D) GOLDSCORE

(E) GOLD PLP

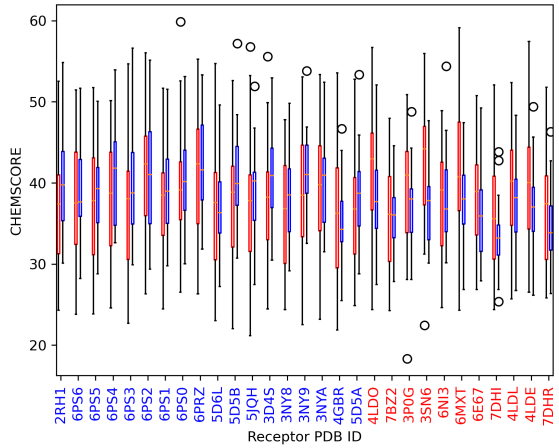
Figure A.2.4: Distribution of docking scores of ADRB2 agonists and antagonists on different receptor structures. PDB ID labels in blue indicate antagonist-bound inactive state structures, purple indicate agonist-bound inactive state structures, red indicate active state structures receptor. For CHEMPLP, NORM CHEMPLP lower score values indicate a higher affinity of the molecule, while for GOLDSCORE, CHEMSCORE and GOLD PLP higher scores are associated to a higher affinity.



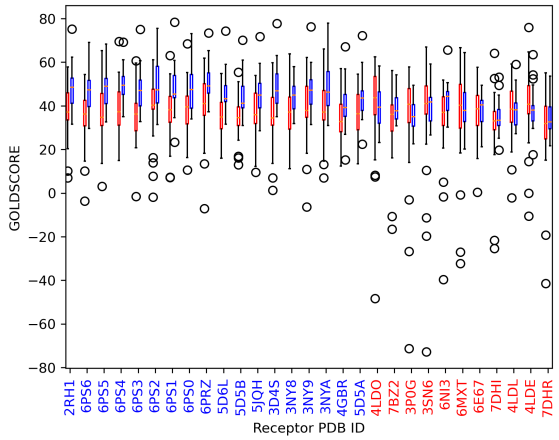
(A) CHEMPLP



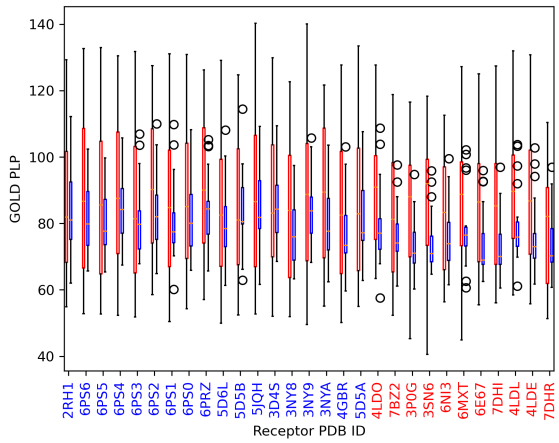
(B) NORMCHEMPLP



(C) CHEMSCORE

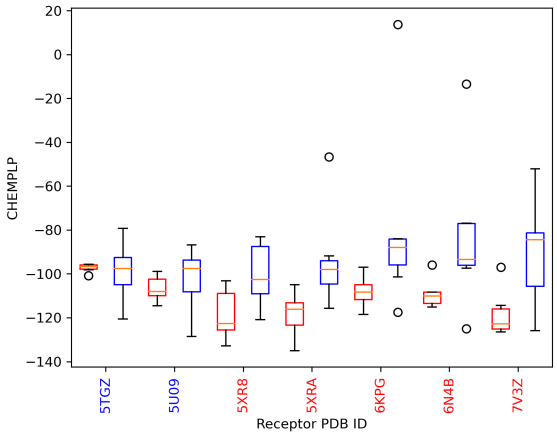


(D) GOLDSCORE

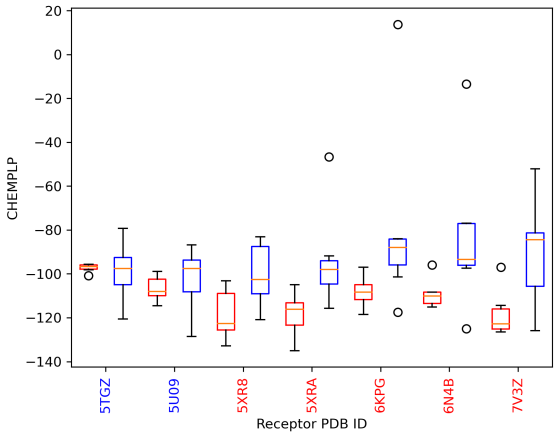


(E) GOLD PLP

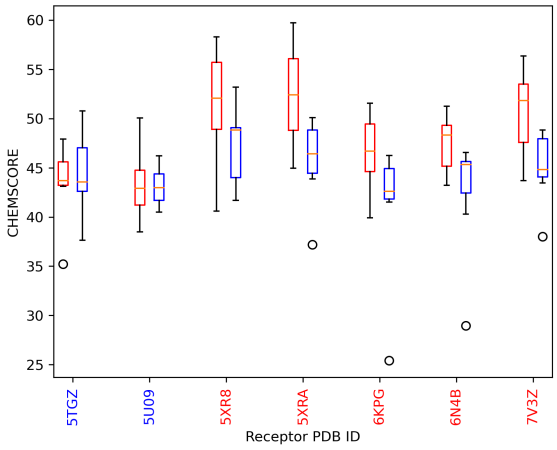
Figure A.2.5: Distribution of docking scores of CNR1 agonists and antagonists on different receptor structures. PDB ID labels in blue indicate antagonist-bound inactive state structures, purple indicate agonist-bound inactive state structures, red indicate active state structures receptor. For CHEMPLP, NORM CHEMPLP lower score values indicate a higher affinity of the molecule, while for GOLDSCORE, CHEMSCORE and GOLD PLP higher scores are associated to a higher affinity.



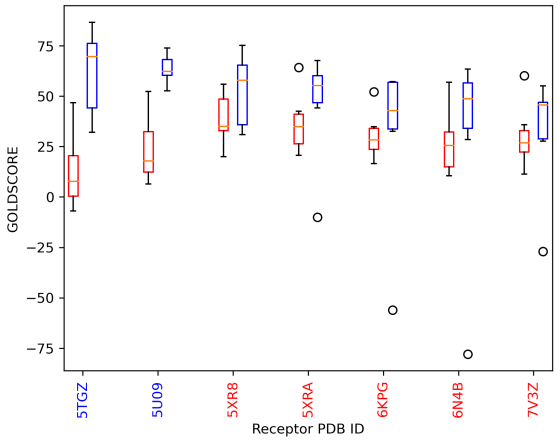
(A) CHEMPLP



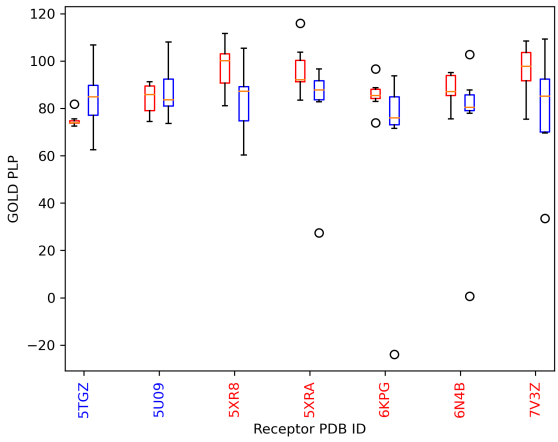
(B) NORMCHEMPLP



(C) CHEMSCORE

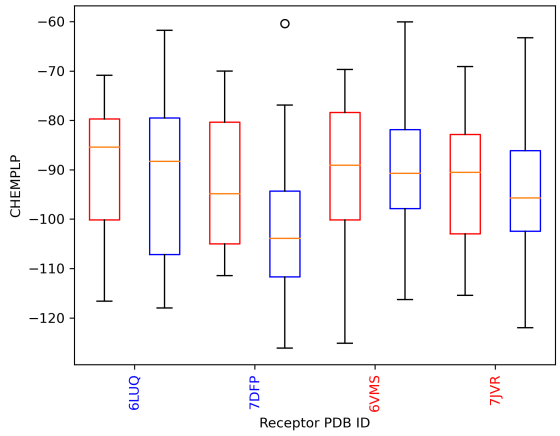


(D) GOLDSCORE

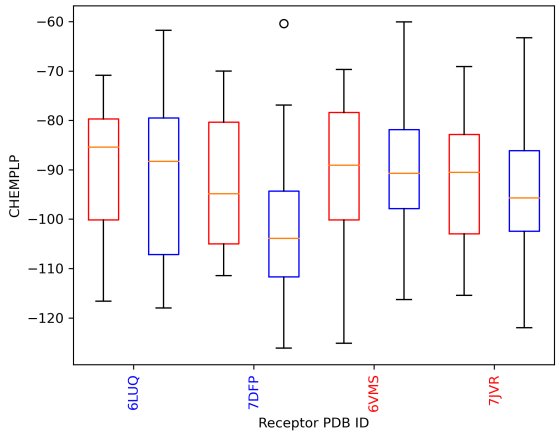


(E) GOLD PLP

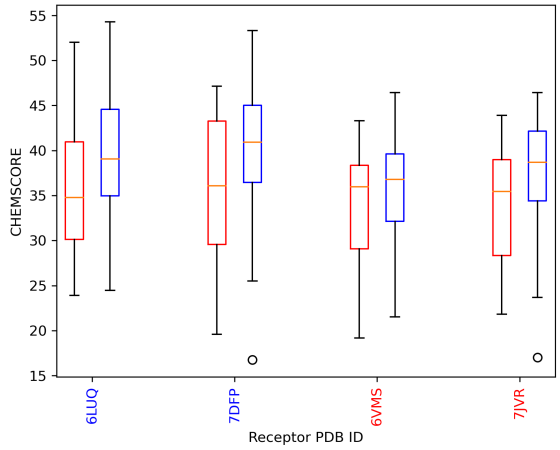
Figure A.2.6: Distribution of docking scores of DRD2 agonists and antagonists on different receptor structures. PDB ID labels in blue indicate antagonist-bound inactive state structures, purple indicate agonist-bound inactive state structures, red indicate active state structures receptor. For CHEMPLP, NORM CHEMPLP lower score values indicate a higher affinity of the molecule, while for GOLDSCORE, CHEMSCORE and GOLD PLP higher scores are associated to a higher affinity.



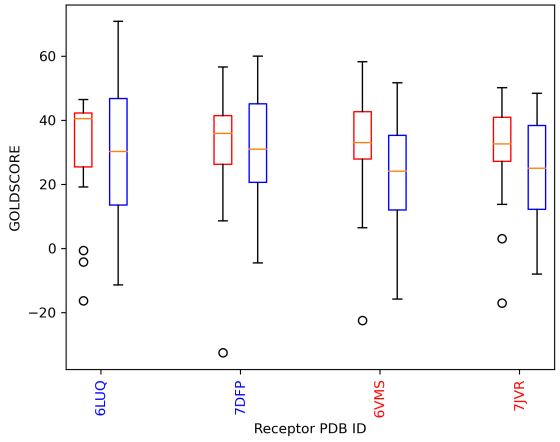
(A) CHEMPLP



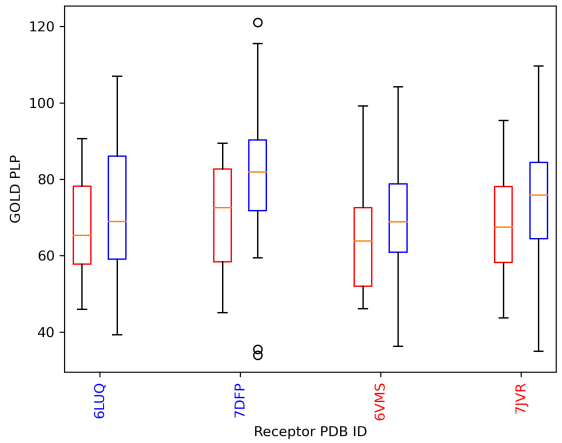
(B) NORMCHEMPLP



(C) CHEMSCORE

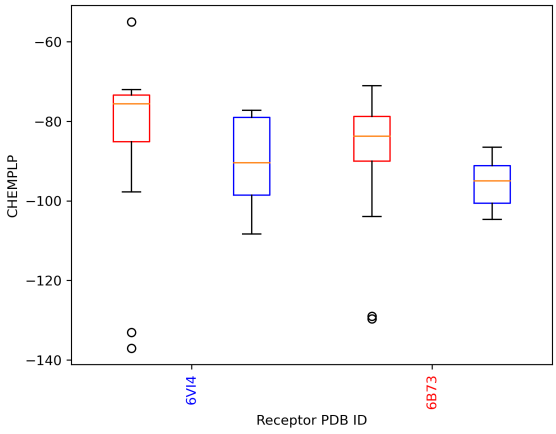


(D) GOLDSCORE

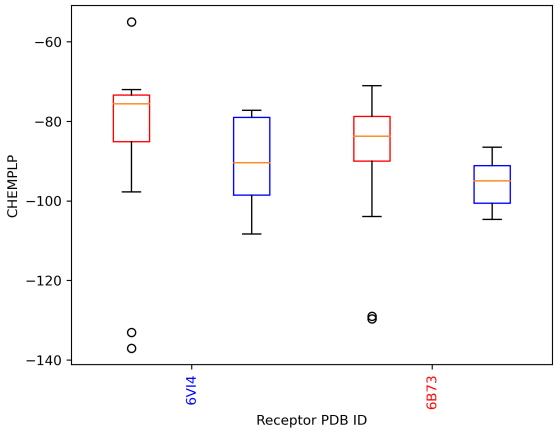


(E) GOLD PLP

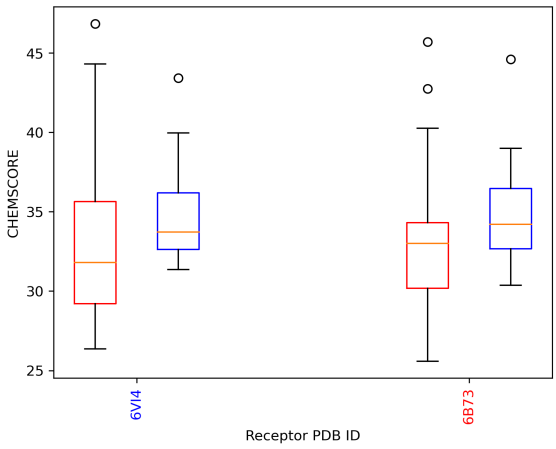
Figure A.2.7: Distribution of docking scores of OPRK agonists and antagonists on different receptor structures. PDB ID labels in blue indicate antagonist-bound inactive state structures, purple indicate agonist-bound inactive state structures, red indicate active state structures receptor. For CHEMPLP, NORM CHEMPLP lower score values indicate a higher affinity of the molecule, while for GOLDSCORE, CHEMSCORE and GOLD PLP higher scores are associated to a higher affinity.



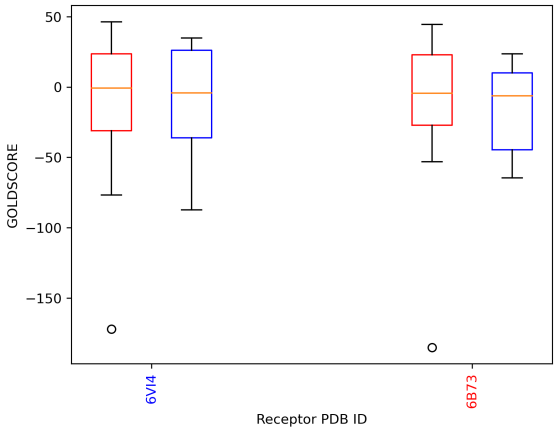
(A) CHEMPLP



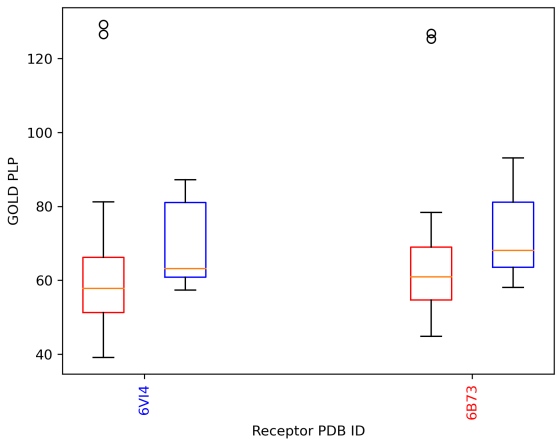
(B) NORMCHEMPLP



(C) CHEMSCORE

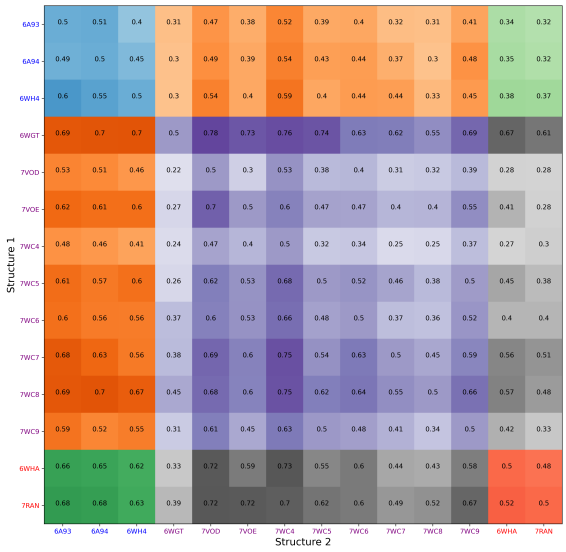


(D) GOLDSCORE

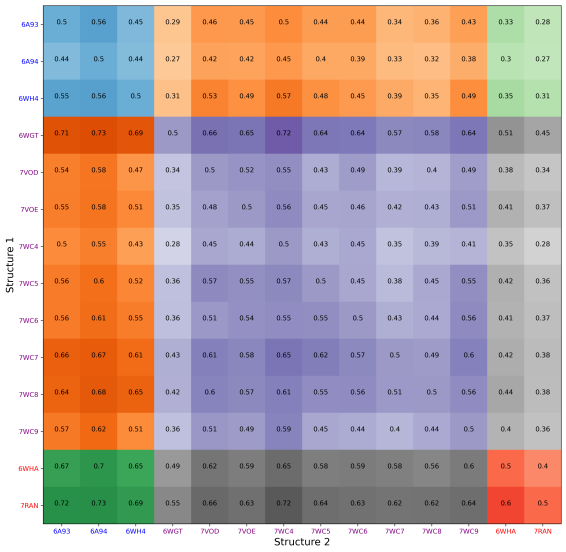


(E) GOLD PLP

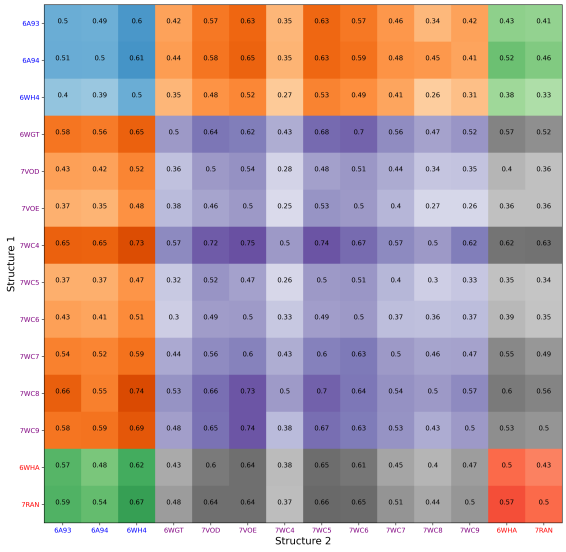
Figure A.2.8: ROCAUC values obtained by ranking 5HT2A agonists and antagonists based on the difference in docking score. For the CHEMPLP scoring function the difference was calculated as Structure 1 - Structure 2, since the score is negative. For the other scoring functions the difference was computed as Structure 2 - Structure 1. Compounds were ranked in descending order by difference in docking score.



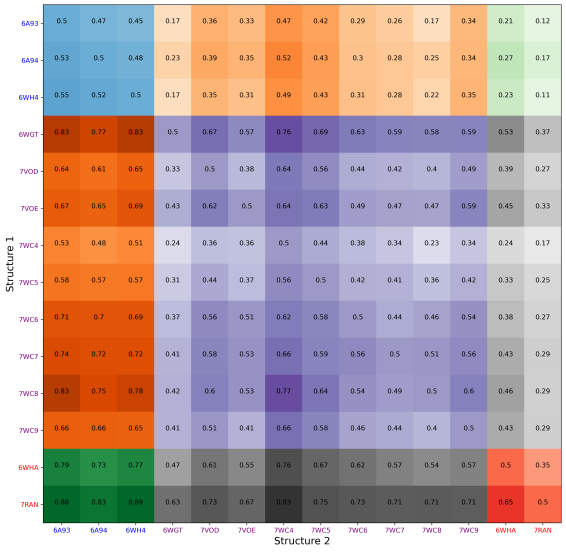
(A) CHEMPLP



(B) CHEMSCORE

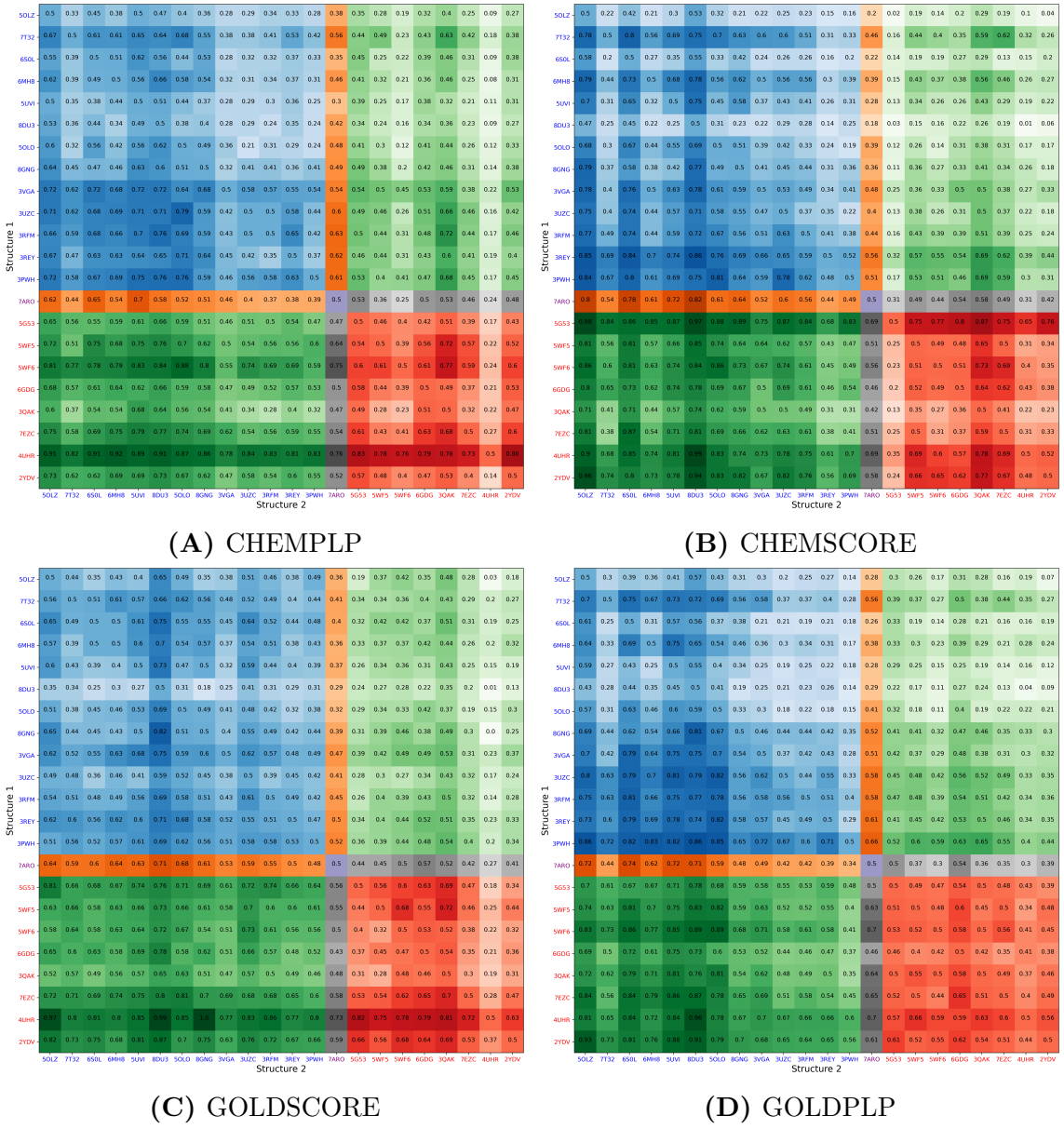


(C) GOLDSCORE



(D) GOLDPLP

Figure A.2.9: ROCAUC values obtained by ranking AA2AR agonists and antagonists based on the difference in docking score. For the CHEMPLP scoring function the difference was calculated as Structure 1 - Structure 2, since the score is negative. For the other scoring functions the difference was computed as Structure 2 - Structure 1. Compounds were ranked in descending order by difference in docking score.



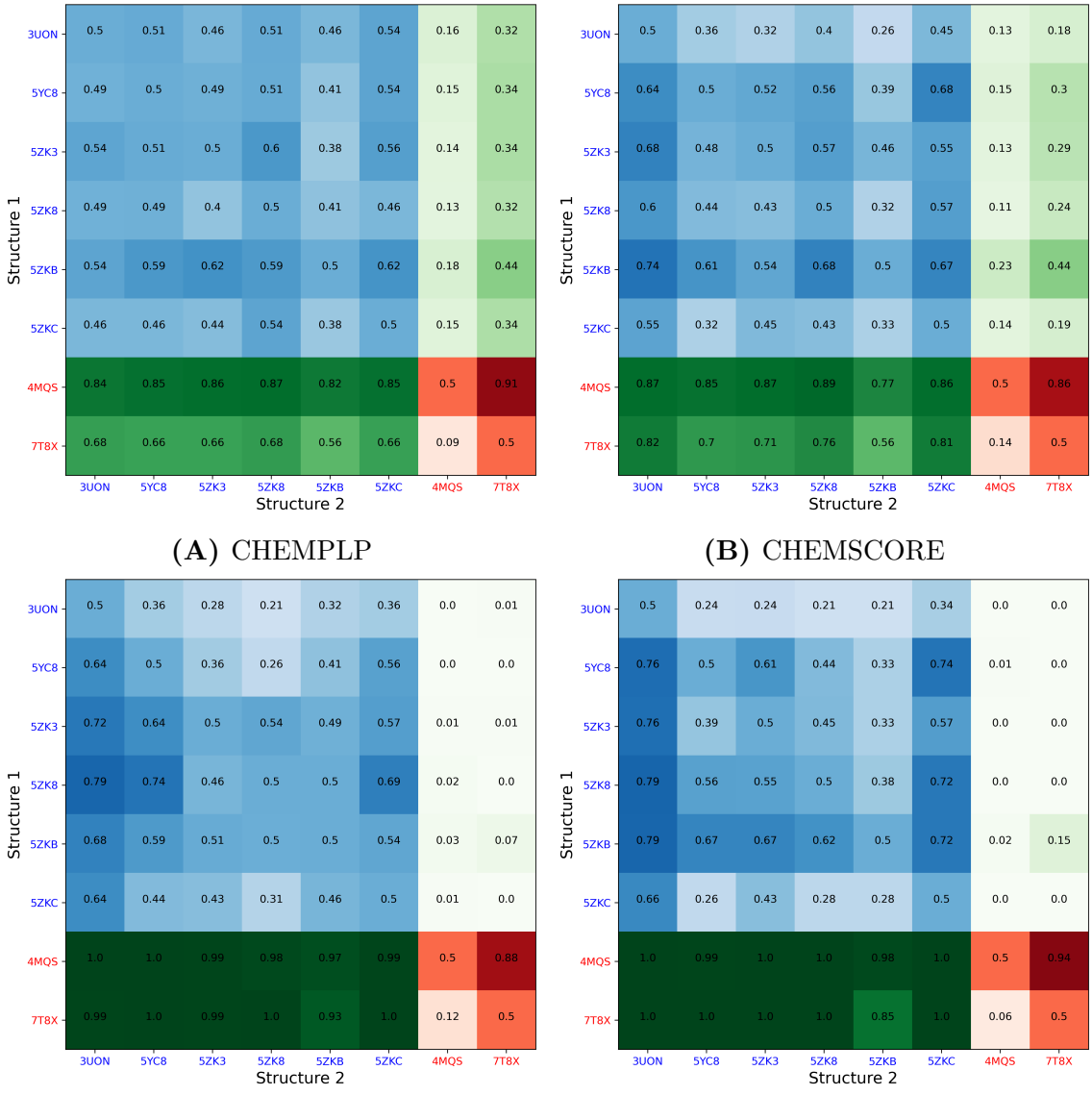
(A) CHEMPLP

(B) CHEMSCORE

(C) GOLDSCORE

(D) GOLDPLP

Figure A.2.10: ROCAUC values obtained by ranking ACM2 agonists and antagonists based on the difference in docking score. For the CHEMPLP scoring function the difference was calculated as Structure 1 - Structure 2, since the score is negative. For the other scoring functions the difference was computed as Structure 2 - Structure 1. Compounds were ranked in descending order by difference in docking score.



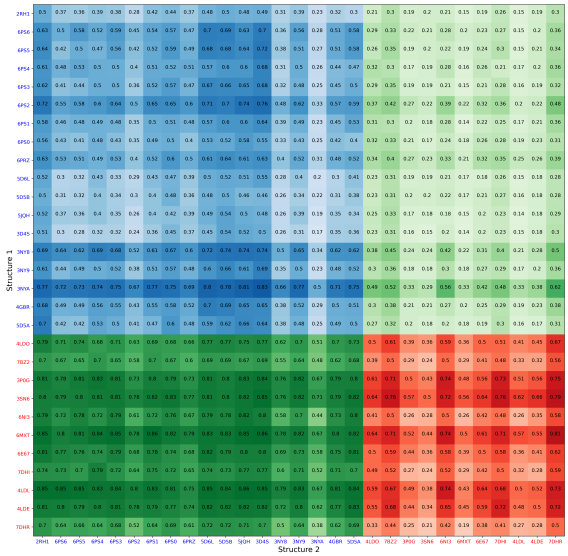
(A) CHEMPLP

(B) CHEMSCORE

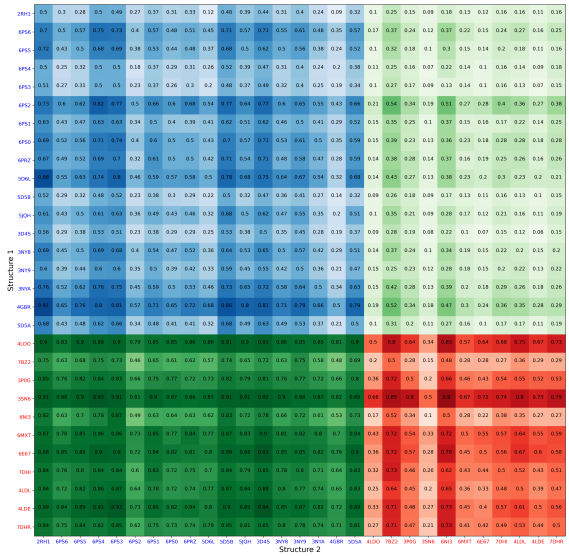
(C) GOLDSCORE

(D) GOLDPLP

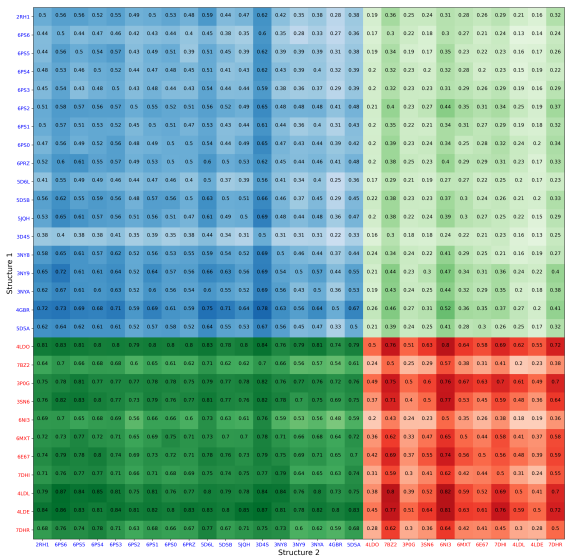
Figure A.2.11: ROCAUC values obtained by ranking ADRB2 agonists and antagonists based on the difference in docking score. For the CHEMPLP scoring function the difference was calculated as Structure 1 - Structure 2, since the score is negative. For the other scoring functions the difference was computed as Structure 2 - Structure 1. Compounds were ranked in descending order by difference in docking score.



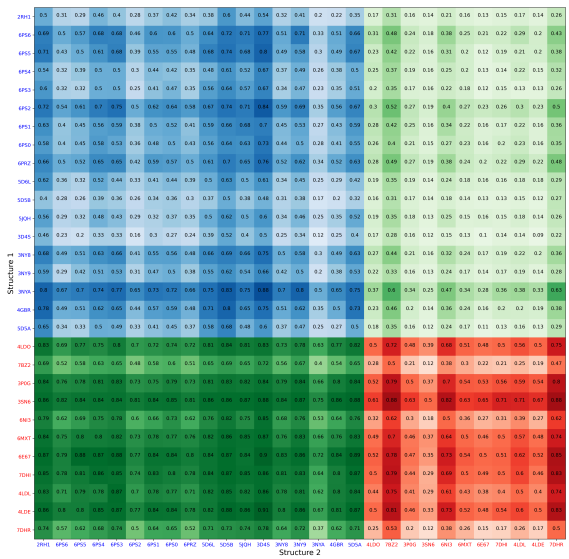
(A) CHEMPLP



(B) CHEMSCORE

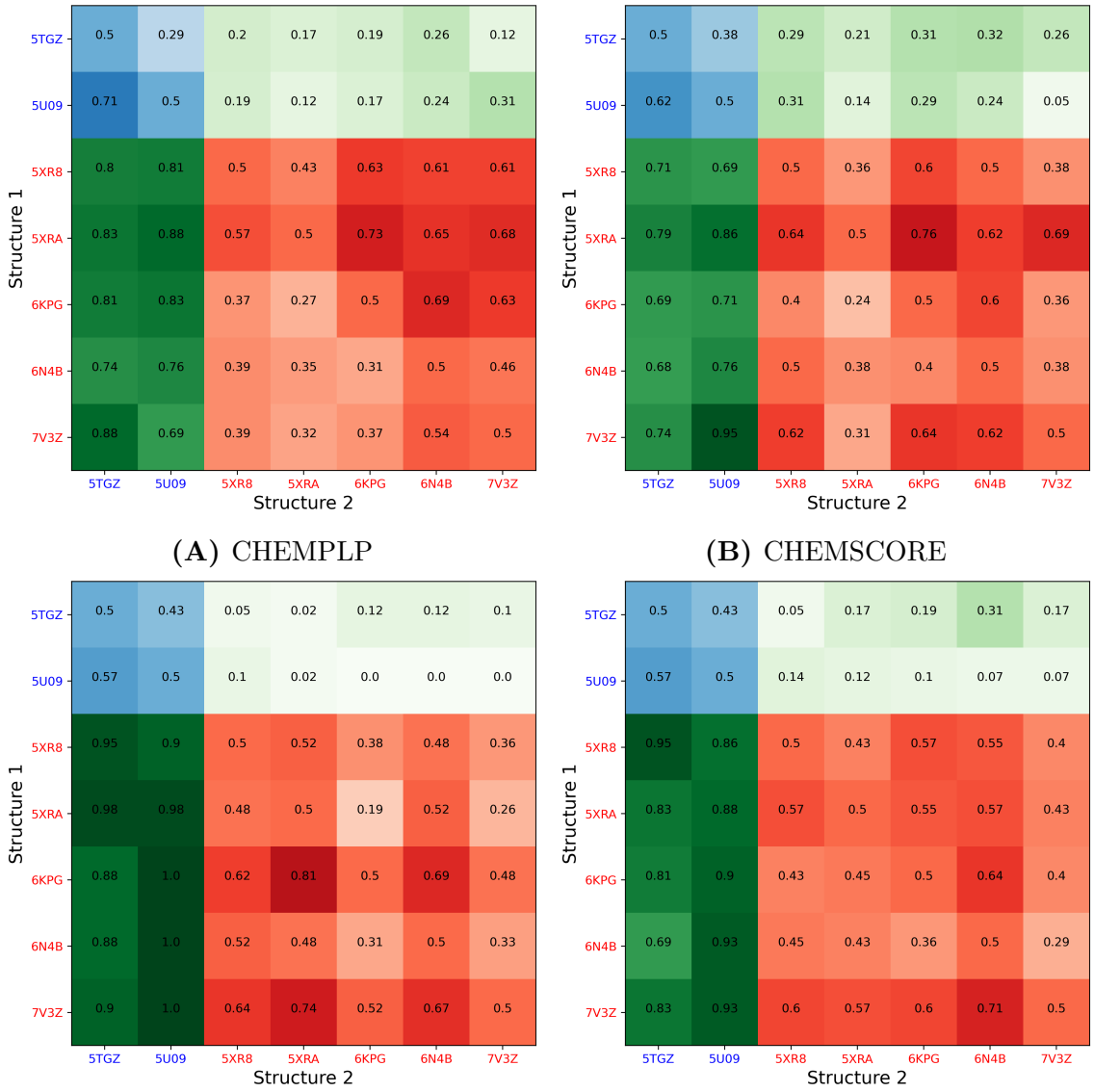


(C) GOLDSCORE



(D) GOLDPLP

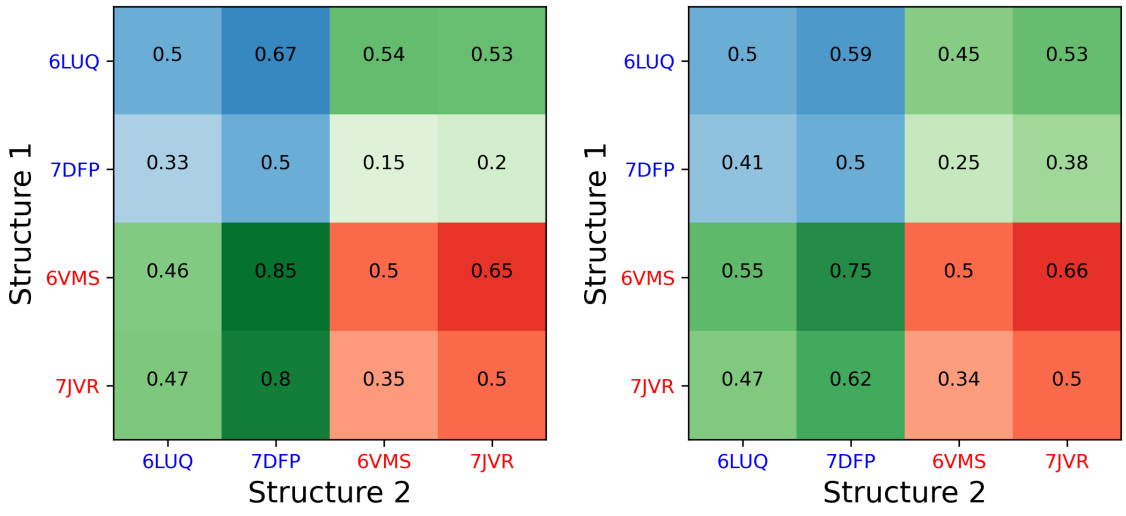
Figure A.2.12: ROCAUC values obtained by ranking CNR1 agonists and antagonists based on the difference in docking score. For the CHEMPLP scoring function the difference was calculated as Structure 1 - Structure 2, since the score is negative. For the other scoring functions the difference was computed as Structure 2 - Structure 1. Compounds were ranked in descending order by difference in docking score.



(A) CHEMPLP **(B)** CHEMSCORE

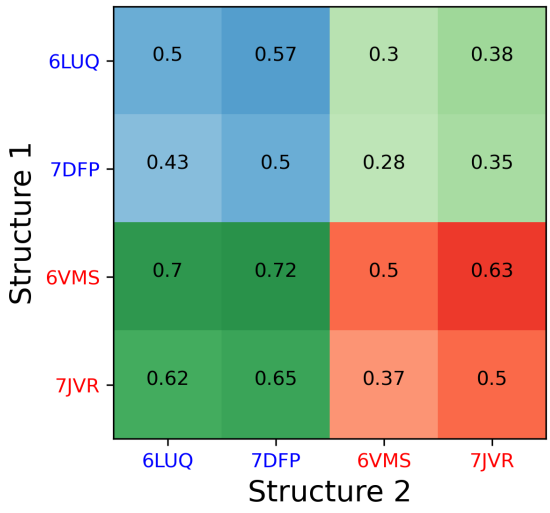
(C) GOLDSCORE **(D)** GOLDPLP

Figure A.2.13: ROCAUC values obtained by ranking DRD2 agonists and antagonists based on the difference in docking score. For the CHEMPLP scoring function the difference was calculated as Structure 1 - Structure 2, since the score is negative. For the other scoring functions the difference was computed as Structure 2 - Structure 1. Compounds were ranked in descending order by difference in docking score.

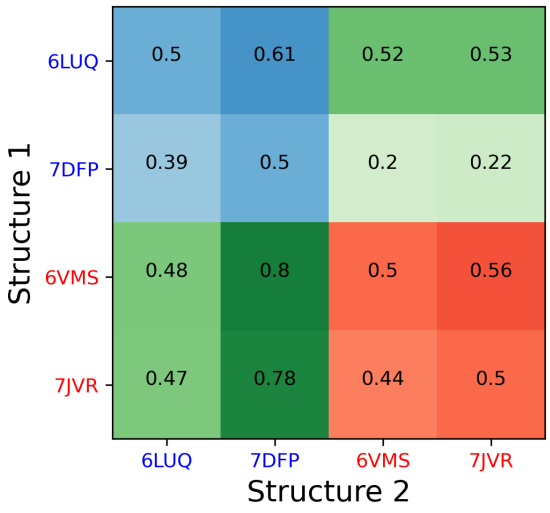


(A) CHEMPLP

(B) CHEMSCORE

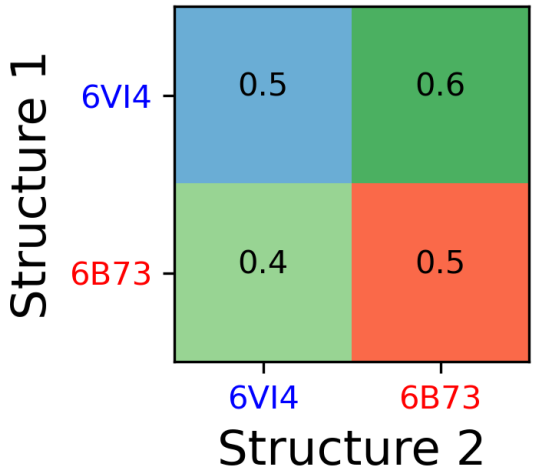


(C) GOLDSCORE

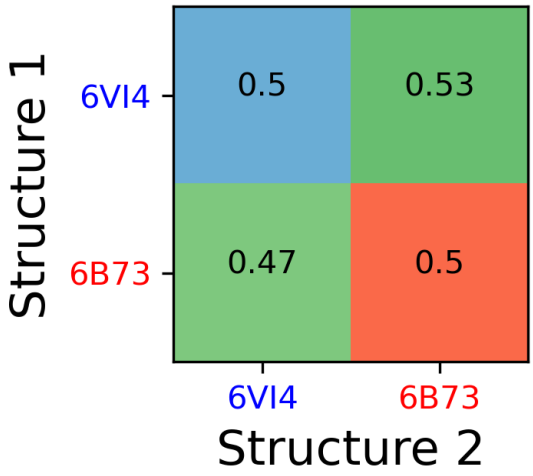


(D) GOLDPLP

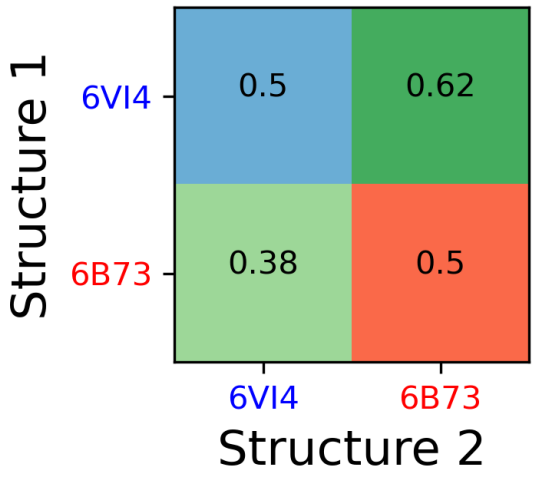
Figure A.2.14: ROCAUC values obtained by ranking OPRK agonists and antagonists based on the difference in docking score. For the CHEMPLP scoring function the difference was calculated as Structure 1 - Structure 2, since the score is negative. For the other scoring functions the difference was computed as Structure 2 - Structure 1. Compounds were ranked in descending order by difference in docking score.



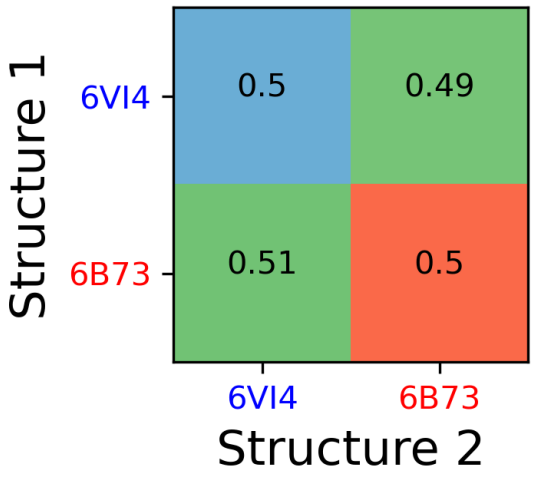
(A) CHEMPLP



(B) CHEMSCORE

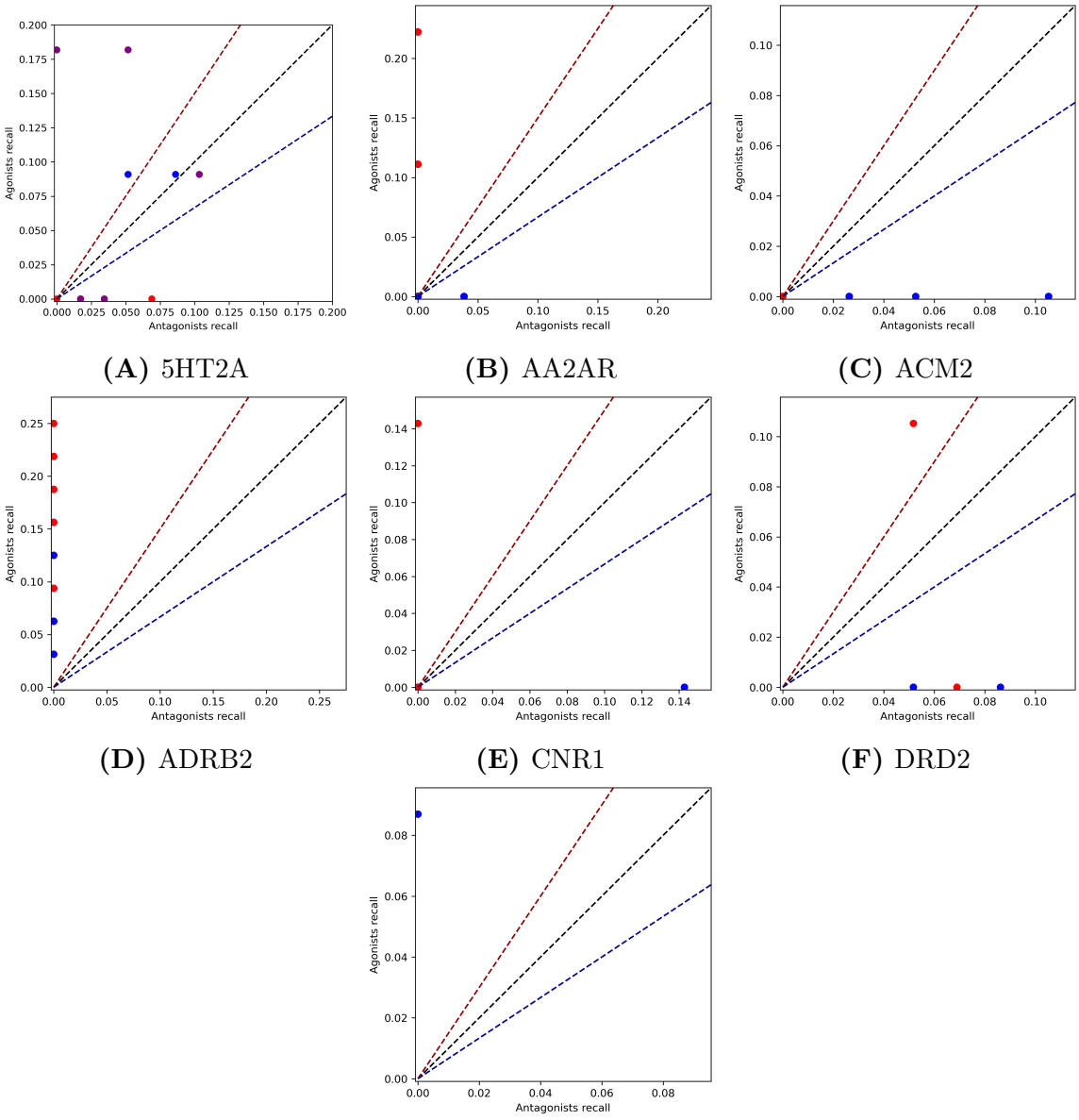


(C) GOLDSCORE



(D) GOLDPLP

Figure A.2.15: Fraction of agonists and fraction of antagonists with a better docking score than the top 1% of scored decoys. Each point represent a different receptor structure used for docking: blue for antagonist-bound inactive state structures, purple for agonist-bound inactive state structures, red for active state structures. The blue dotted line indicates a 1.5 higher ratio the fraction of selected antagonists than agonists. The red dotted line indicates a 1.5 higher ratio the fraction of selected agonists than antagonists.

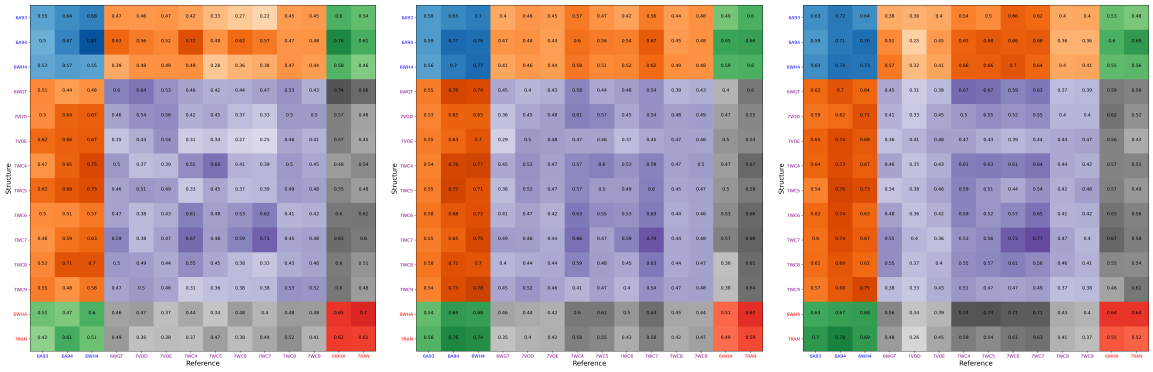


(A) 5HT2A **(B)** AA2AR **(C)** ACM2

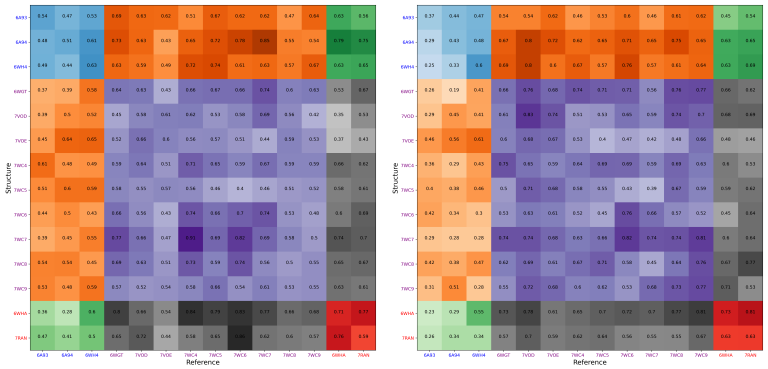
(D) ADRB2 **(E)** CNR1 **(F)** DRD2

(G) OPRK

Figure A.2.16: ROCAUC values based on ranking of known 5HT2A agonists and antagonists using interaction similarity to crystallographic references. The positive class used to calculate ROCAUC correspond to the efficacy of the reference ligand. (A) Tanimoto IFP similarity. (B) Graph kernel similarity on Hyd IPA. (C) GRIM score on Newhyd IPA. (D) Graph kernel similarity on Hyd IPA. (E) GRIM score on Newhyd IPA.

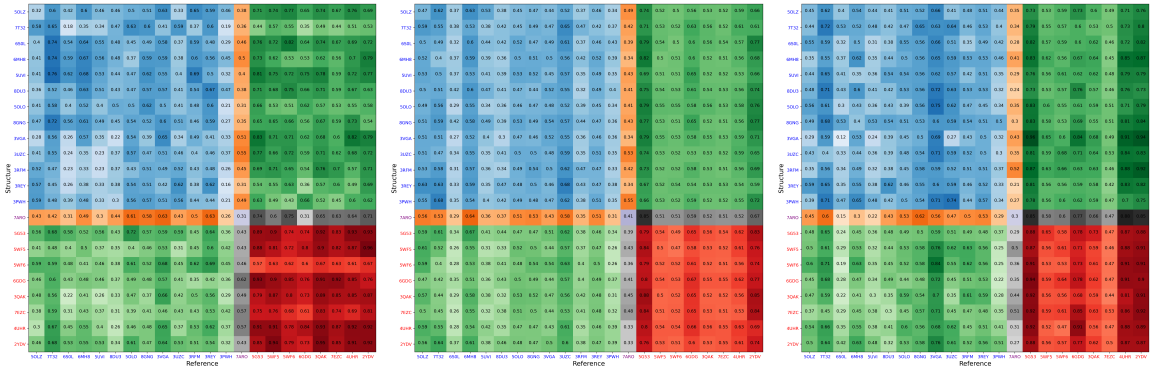


(A) IFP **(B)** Kernel Hyd **(C)** Kernel Newhyd

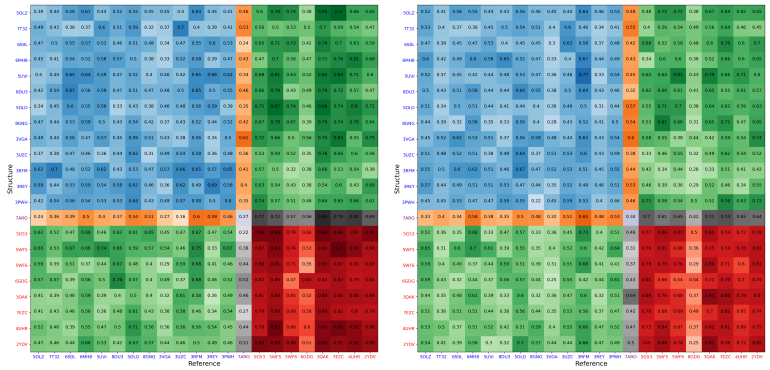


(D) GRIM Hyd **(E)** GRIM Newhyd

Figure A.2.17: ROCAUC values based on ranking of known AA2AR agonists and antagonists using interaction similarity to crystallographic references. The positive class used to calculate ROCAUC correspond to the efficacy of the reference ligand. (A) Tanimoto IFP similarity. (B) Graph kernel similarity on Hyd IPA. (C) GRIM score on Newhyd IPA. (D) Graph kernel similarity on Hyd IPA. (E) GRIM score on Newhyd IPA.

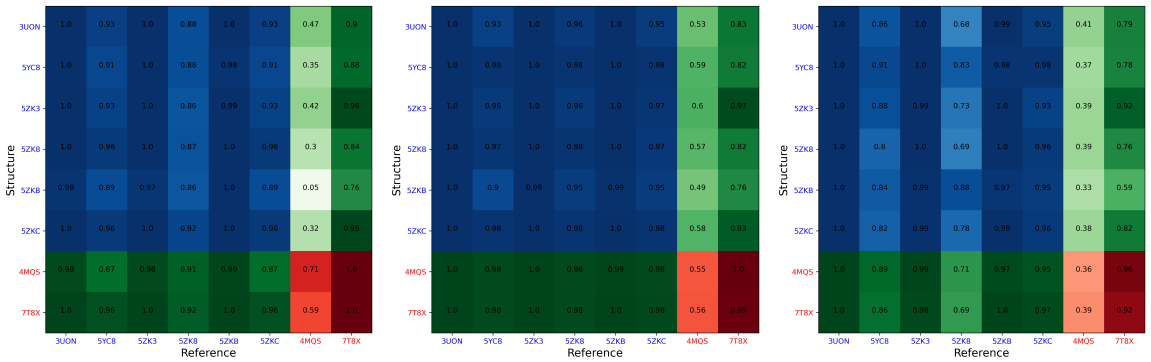


(A) IFP **(B)** Kernel Hyd **(C)** Kernel Newhyd

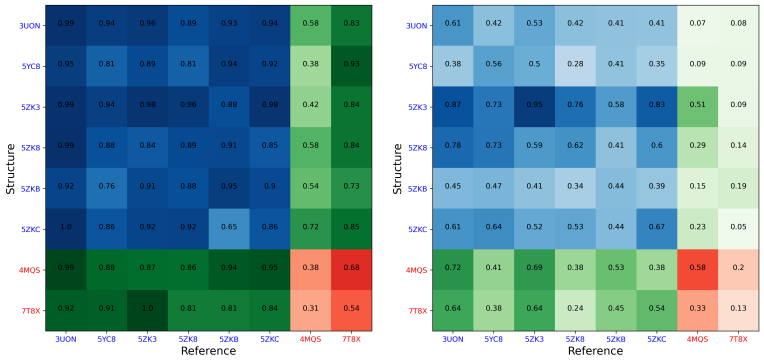


(D) GRIM Hyd **(E)** GRIM Newhyd

Figure A.2.18: ROCAUC values based on ranking of known ACM2 agonists and antagonists using interaction similarity to crystallographic references. The positive class used to calculate ROCAUC correspond to the efficacy of the reference ligand. (A) Tanimoto IFP similarity. (B) Graph kernel similarity on Hyd IPA. (C) GRIM score on Newhyd IPA. (D) Graph kernel similarity on Hyd IPA. (E) GRIM score on Newhyd IPA.

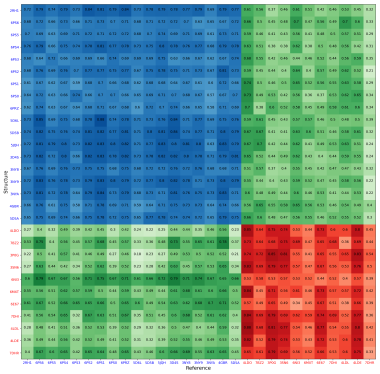


(A) IFP **(B)** Kernel Hyd **(C)** Kernel Newhyd

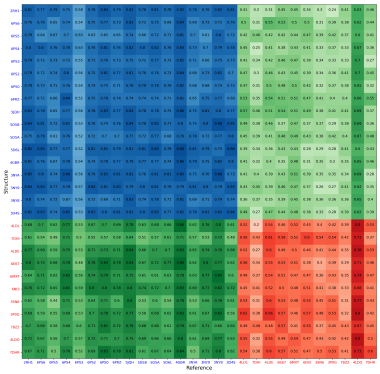


(D) GRIM Hyd **(E)** GRIM Newhyd

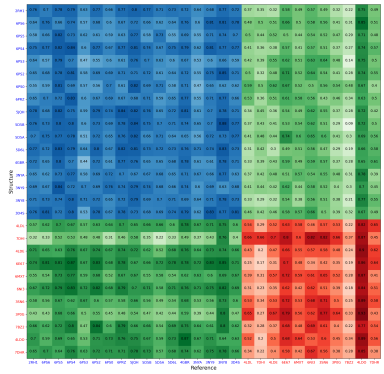
Figure A.2.19: ROCAUC values based on ranking of known ADRB2 agonists and antagonists using interaction similarity to crystallographic references. The positive class used to calculate ROCAUC correspond to the efficacy of the reference ligand. (A) Tanimoto IFP similarity. (B) Graph kernel similarity on Hyd IPA. (C) GRIM score on Newhyd IPA. (D) Graph kernel similarity on Hyd IPA. (E) GRIM score on Newhyd IPA.



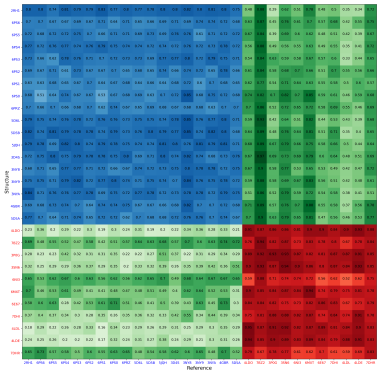
(A) IFP



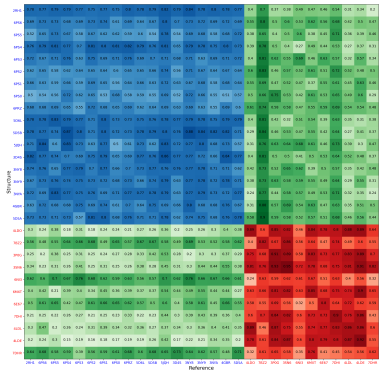
(B) Kernel Hyd



(C) Kernel Newhyd

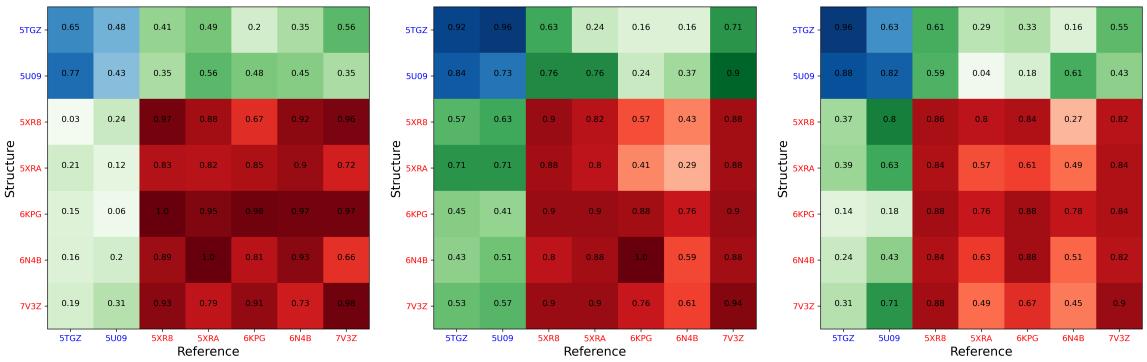


(D) GRIM Hyd

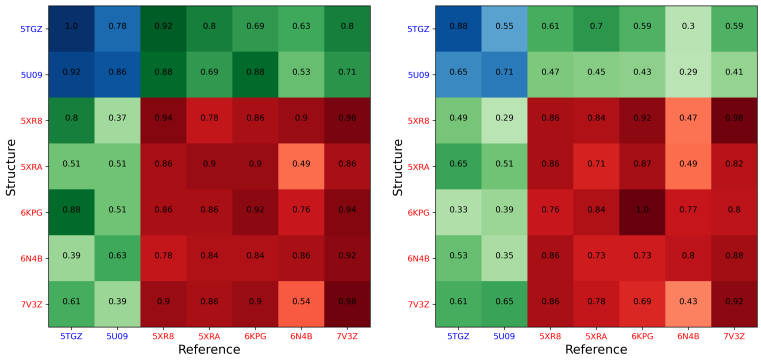


(E) GRIM Newhyd

Figure A.2.20: ROCAUC values based on ranking of known CNR1 agonists and antagonists using interaction similarity to crystallographic references. The positive class used to calculate ROCAUC correspond to the efficacy of the reference ligand. (A) Tanimoto IFP similarity. (B) Graph kernel similarity on Hyd IPA. (C) GRIM score on Newhyd IPA. (D) Graph kernel similarity on Hyd IPA. (E) GRIM score on Newhyd IPA.



(A) IFP **(B)** Kernel Hyd **(C)** Kernel Newhyd



(D) GRIM Hyd **(E)** GRIM Newhyd

Figure A.2.21: ROCAUC values based on ranking of known DRD2 agonists and antagonists using interaction similarity to crystallographic references. The positive class used to calculate ROCAUC correspond to the efficacy of the reference ligand. (A) Tanimoto IFP similarity. (B) Graph kernel similarity on Hyd IPA. (C) GRIM score on Newhyd IPA. (D) Graph kernel similarity on Hyd IPA. (E) GRIM score on Newhyd IPA.

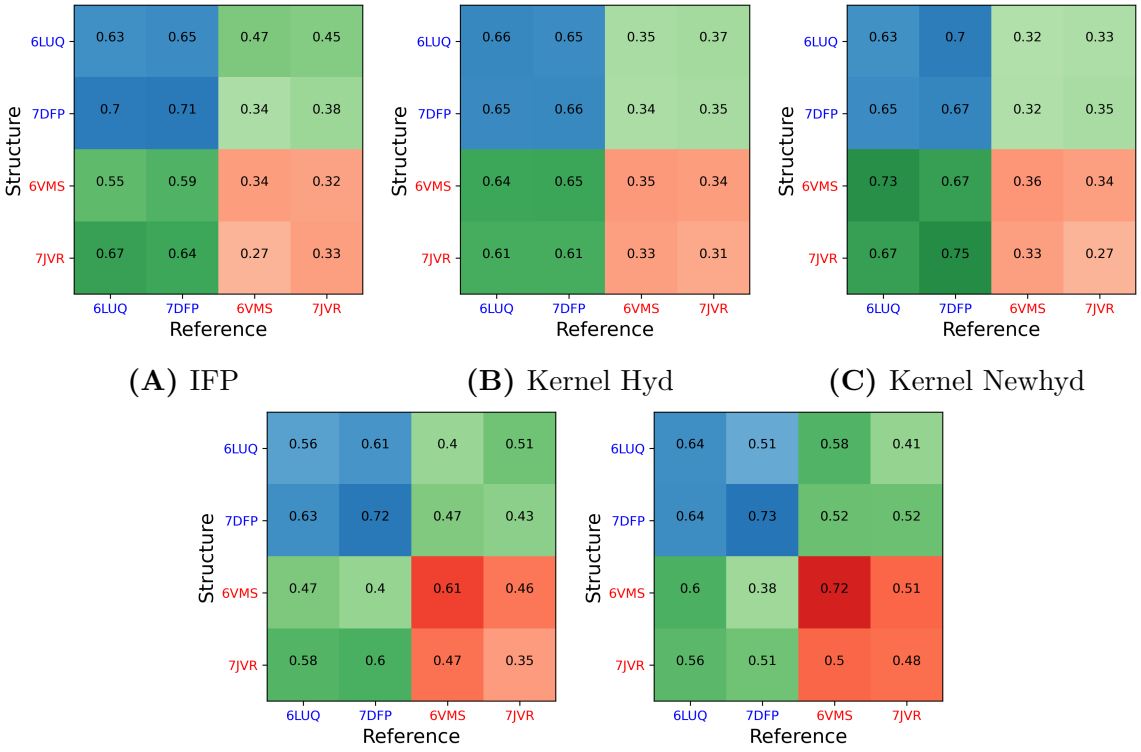
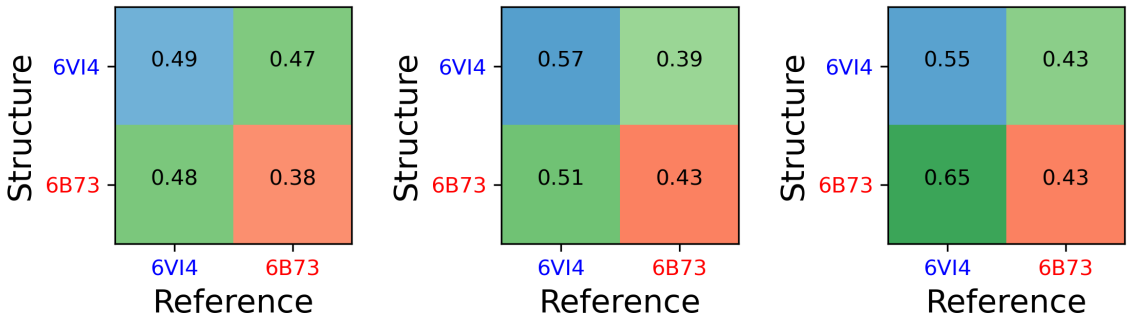
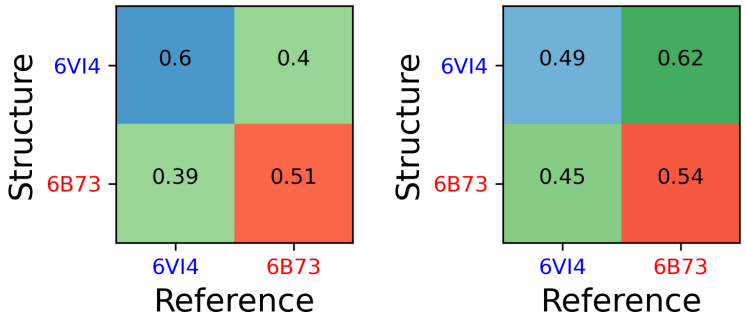


Figure A.2.22: ROCAUC values based on ranking of known OPRK agonists and antagonists using interaction similarity to crystallographic references. The positive class used to calculate ROCAUC correspond to the efficacy of the reference ligand. (A) Tanimoto IFP similarity. (B) Graph kernel similarity on Hyd IPA. (C) GRIM score on Newhyd IPA. (D) Graph kernel similarity on Hyd IPA. (E) GRIM score on Newhyd IPA.

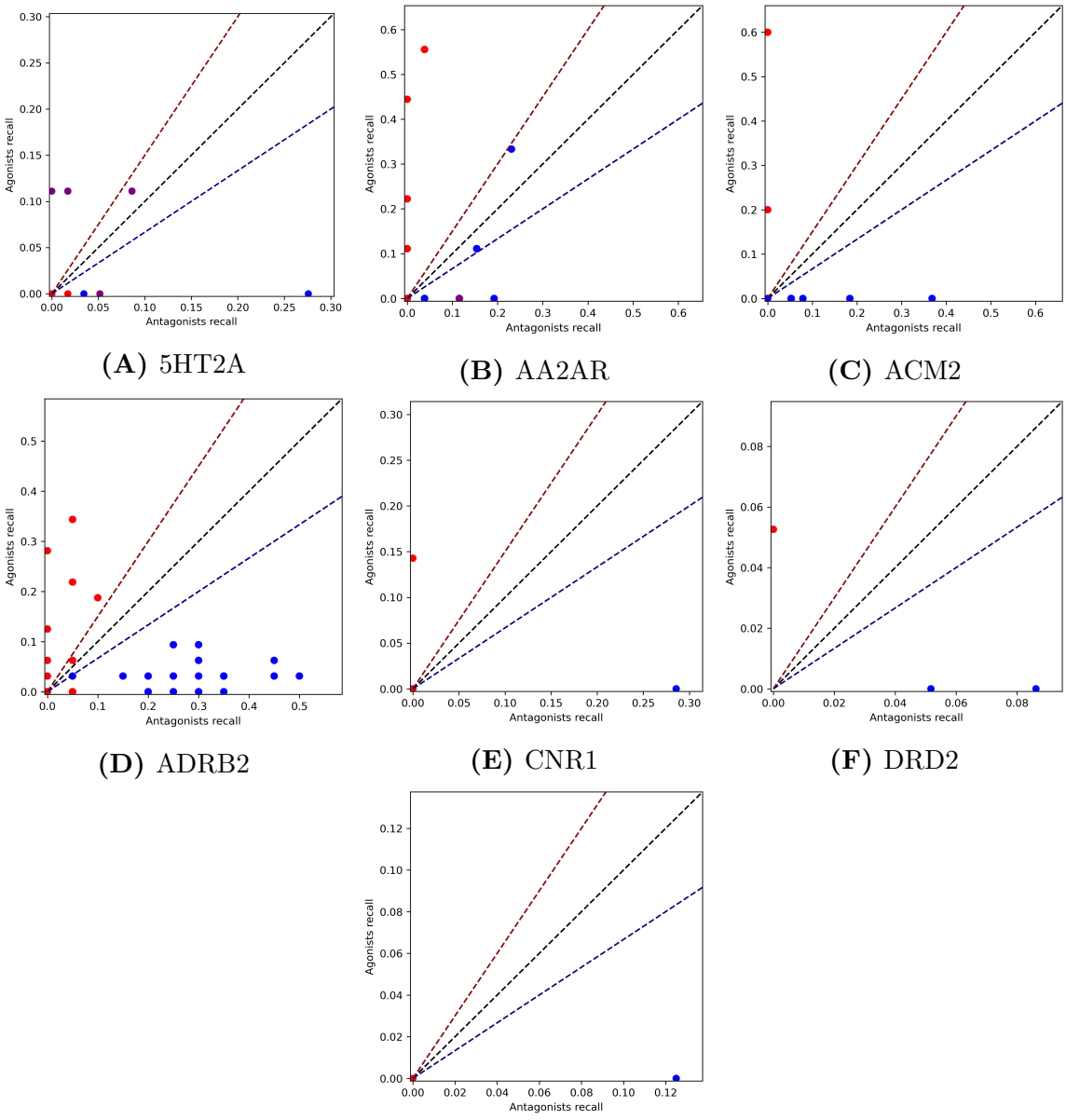


(A) IFP **(B)** Kernel Hyd **(C)** Kernel Newhyd



(D) GRIM Hyd **(E)** GRIM Newhyd

Figure A.2.23: Fraction of agonists and fraction of antagonists with a higher IFP similarity than the top 1% of scored decoys. The binding mode of the ligand co-crystallized with the structure used for docking was selected as reference. Each point represent a different receptor structure used for docking: blue for antagonist-bound inactive state structures, purple for agonist-bound inactive state structures, red for active state structures. The blue dotted line indicates a 1.5 higher ratio the fraction of selected antagonists than agonists. The red dotted line indicates a 1.5 higher ratio the fraction of selected agonists than antagonists.

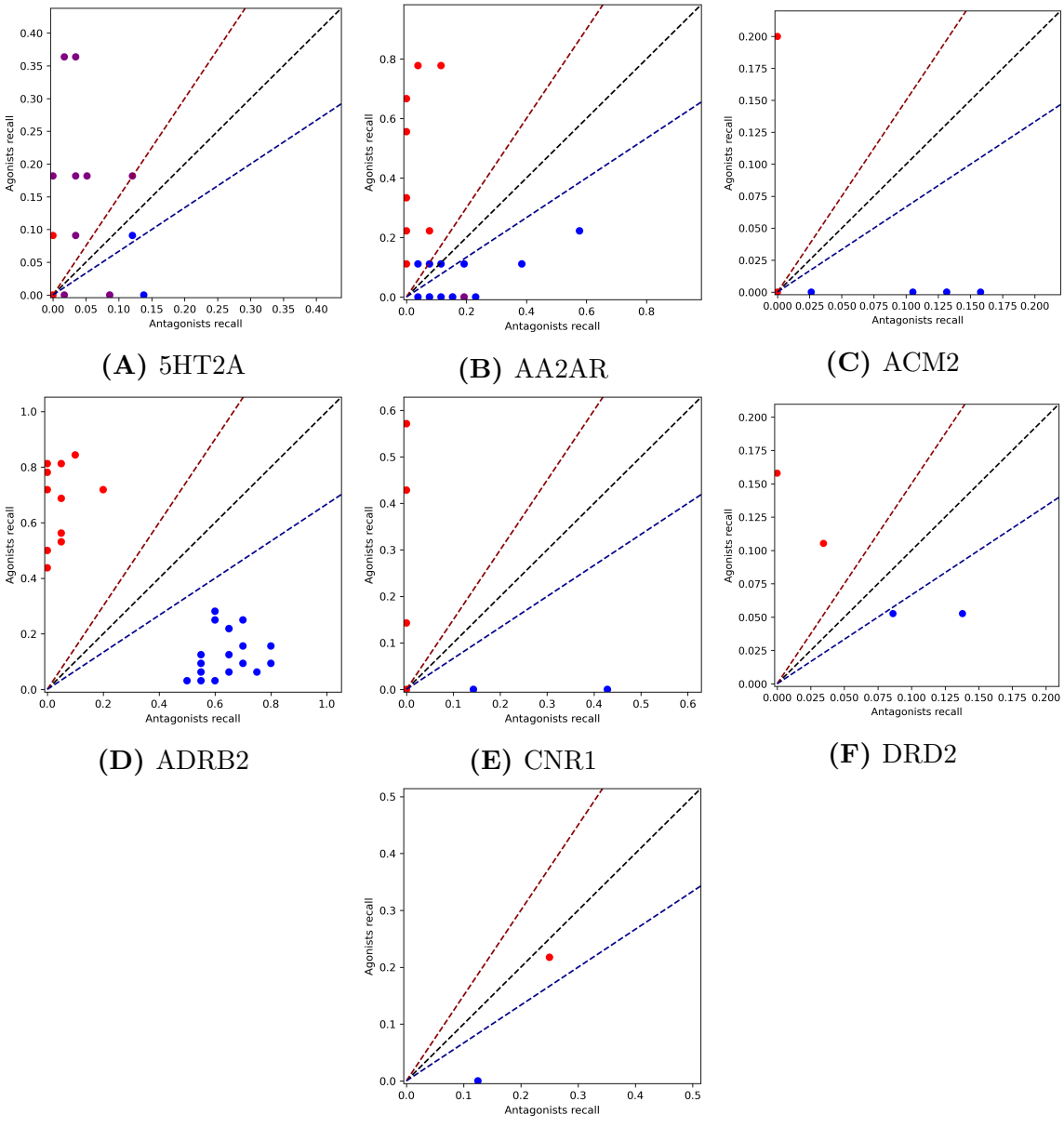


(A) 5HT2A **(B)** AA2AR **(C)** ACM2

(D) ADRB2 **(E)** CNR1 **(F)** DRD2

(G) OPRK

Figure A.2.24: Fraction of agonists and fraction of antagonists with a higher GRIM score than the top 1% of scored decoys. The Hyd definition of hydrophobic contacts was used for interaction detection. The binding mode of the ligand co-crystallized with the structure used for docking was selected as reference. Each point represent a different receptor structure used for docking: blue for antagonist-bound inactive state structures, purple for agonist-bound inactive state structures, red for active state structures. The blue dotted line indicates a 1.5 higher ratio the fraction of selected antagonists than agonists. The red dotted line indicates a 1.5 higher ratio the fraction of selected agonists than antagonists.

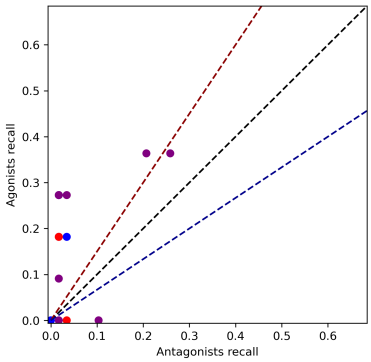


(A) 5HT2A **(B)** AA2AR **(C)** ACM2

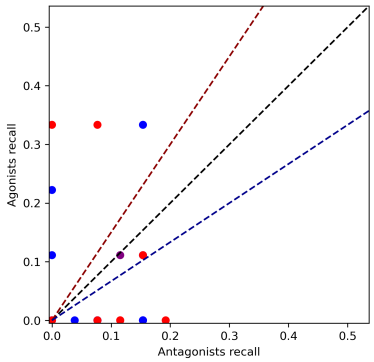
(D) ADRB2 **(E)** CNR1 **(F)** DRD2

(G) OPRK

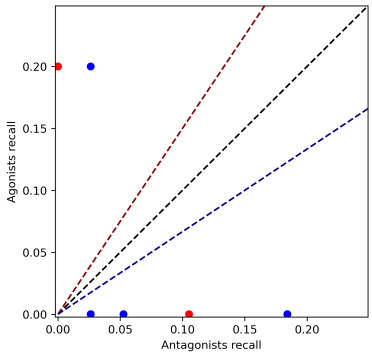
Figure A.2.25: Fraction of agonists and fraction of antagonists with a higher GRIM score than the top 1% of scored decoys. The Newhyd definition of hydrophobic contacts was used for interaction detection. The binding mode of the ligand co-crystallized with the structure used for docking was selected as reference. Each point represent a different receptor structure used for docking: blue for antagonist-bound inactive state structures, purple for agonist-bound inactive state structures, red for active state structures. The blue dotted line indicates a 1.5 higher ratio the fraction of selected antagonists than agonists. The red dotted line indicates a 1.5 higher ratio the fraction of selected agonists than antagonists.



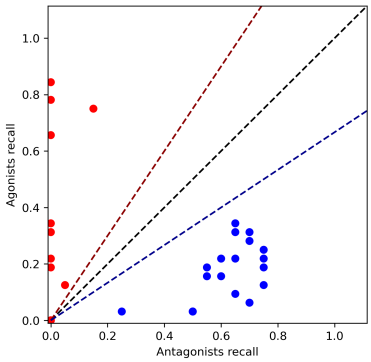
(A) 5HT2A



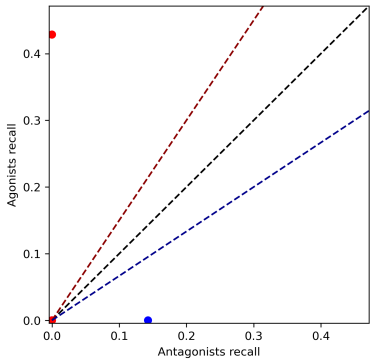
(B) AA2AR



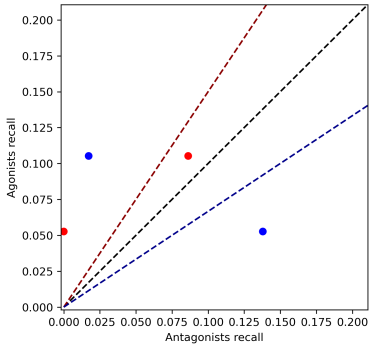
(C) ACM2



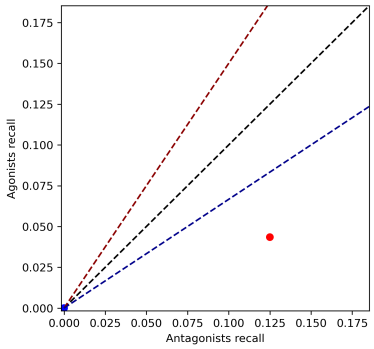
(D) ADRB2



(E) CNR1

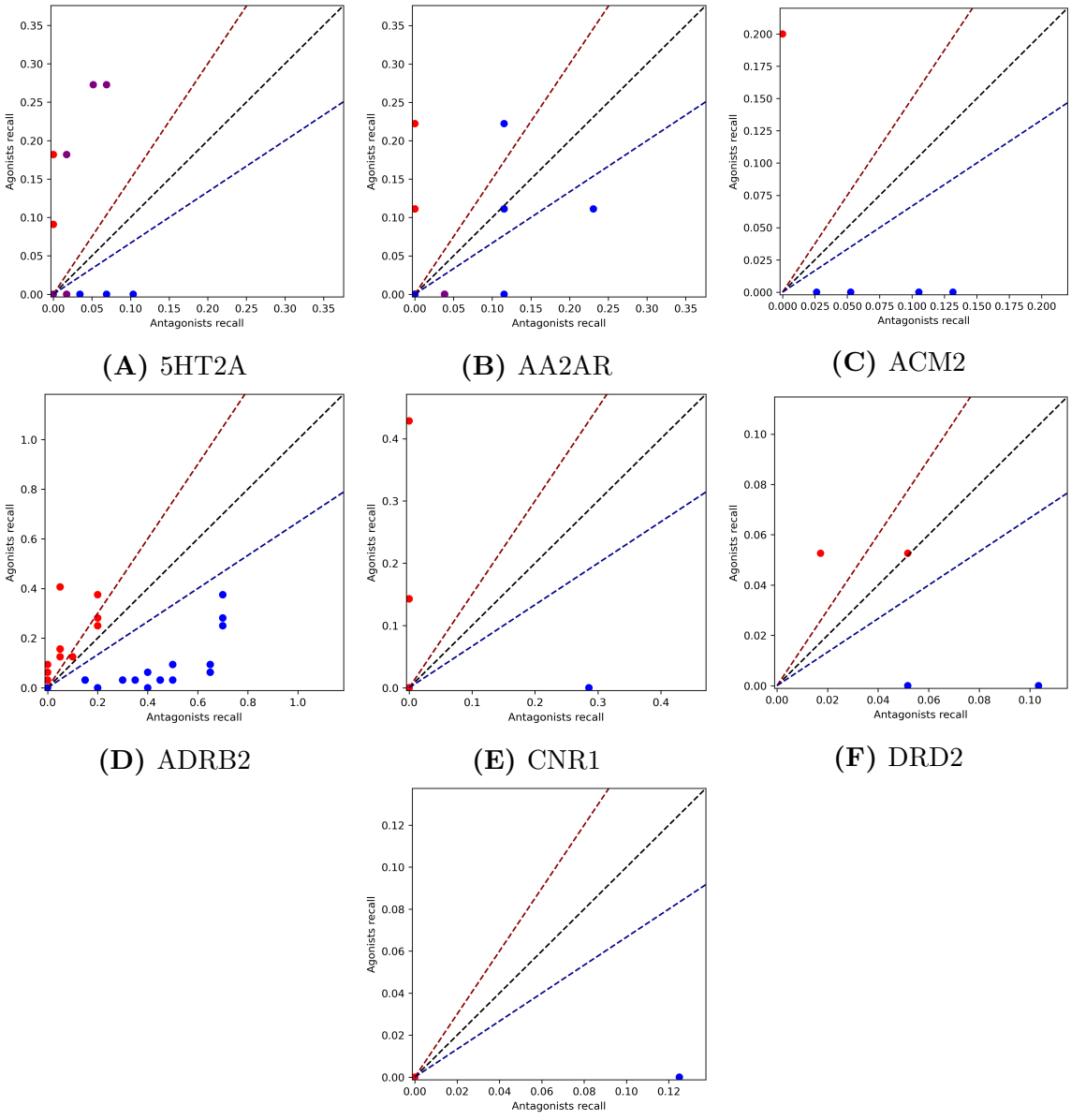


(F) DRD2



(G) OPRK

Figure A.2.26: Fraction of agonists and fraction of antagonists with a higher kernel similarity score than the top 1% of scored decoys. The Hyd definition of hydrophobic contacts was used for interaction detection. The binding mode of the ligand co-crystallized with the structure used for docking was selected as reference. Each point represent a different receptor structure used for docking: blue for antagonist-bound inactive state structures, purple for agonist-bound inactive state structures, red for active state structures. The blue dotted line indicates a 1.5 higher ratio the fraction of selected antagonists than agonists. The red dotted line indicates a 1.5 higher ratio the fraction of selected agonists than antagonists.

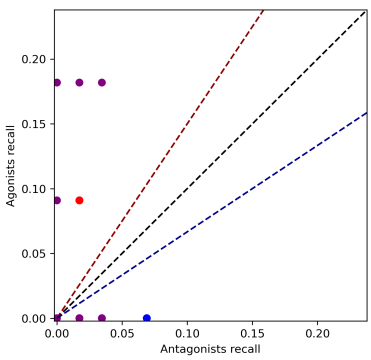


(A) 5HT2A **(B)** AA2AR **(C)** ACM2

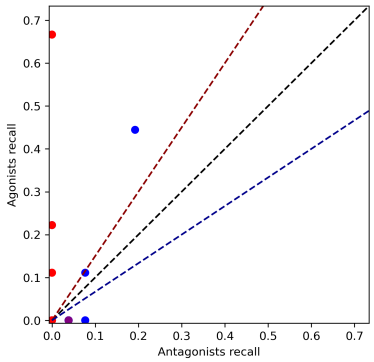
(D) ADRB2 **(E)** CNR1 **(F)** DRD2

(G) OPRK

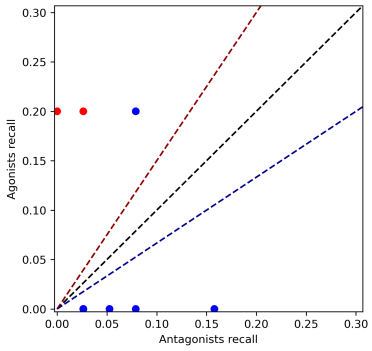
Figure A.2.27: Fraction of agonists and fraction of antagonists with a higher kernel similarity score than the top 1% of scored decoys. The Newhyd definition of hydrophobic contacts was used for interaction detection. The binding mode of the ligand co-crystallized with the structure used for docking was selected as reference. Each point represent a different receptor structure used for docking: blue for antagonist-bound inactive state structures, purple for agonist-bound inactive state structures, red for active state structures. The blue dotted line indicates a 1.5 higher ratio the fraction of selected antagonists than agonists. The red dotted line indicates a 1.5 higher ratio the fraction of selected agonists than antagonists.



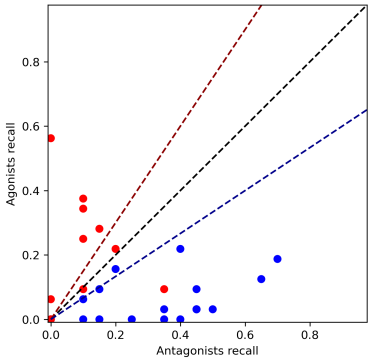
(A) 5HT2A



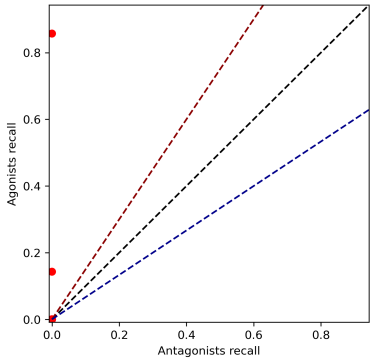
(B) AA2AR



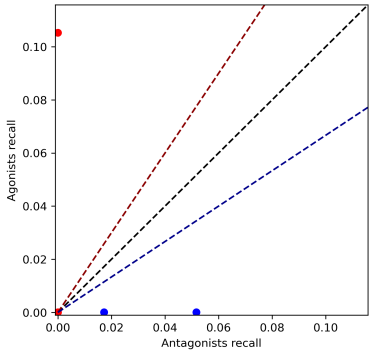
(C) ACM2



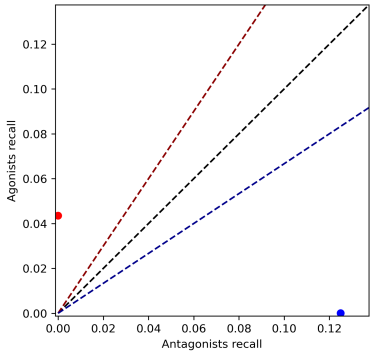
(D) ADRB2



(E) CNR1

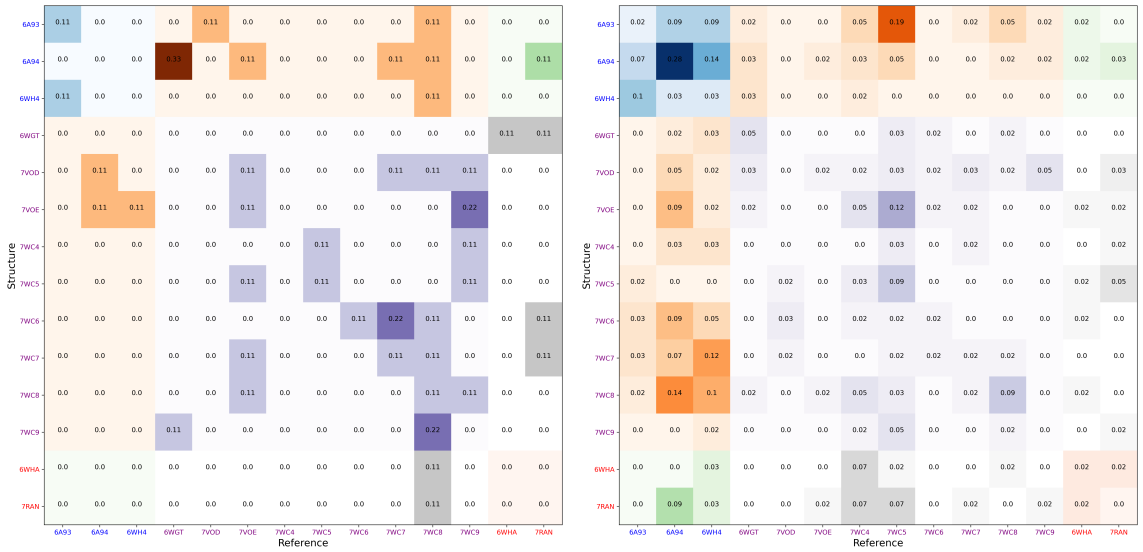


(F) DRD2



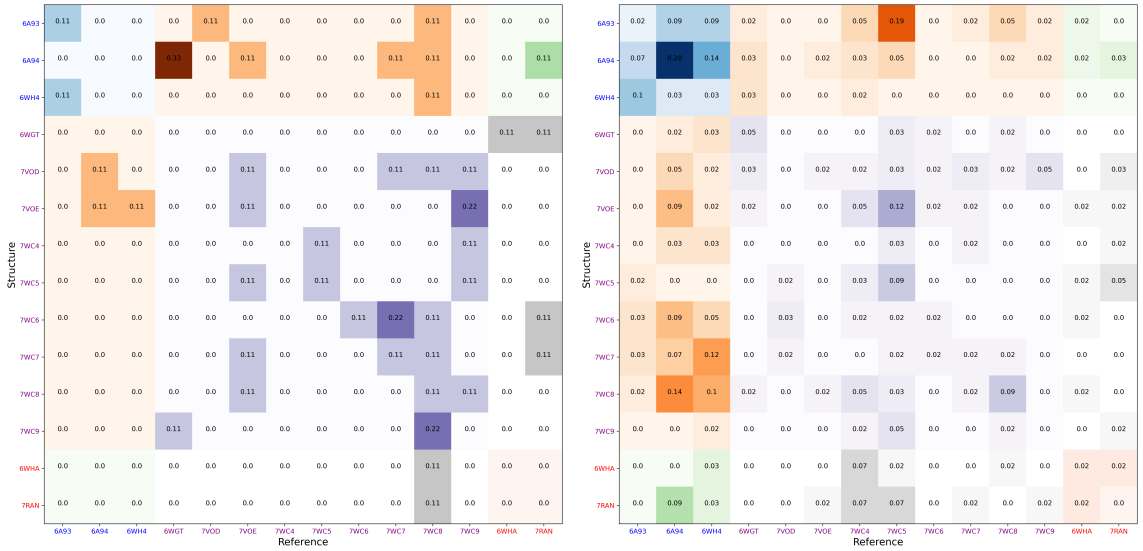
(G) OPRK

Figure A.2.28: Fraction of agonist (A) and antagonist (B) with a IFP Tanimoto similarity score better than the best 1% of evaluated decoys.



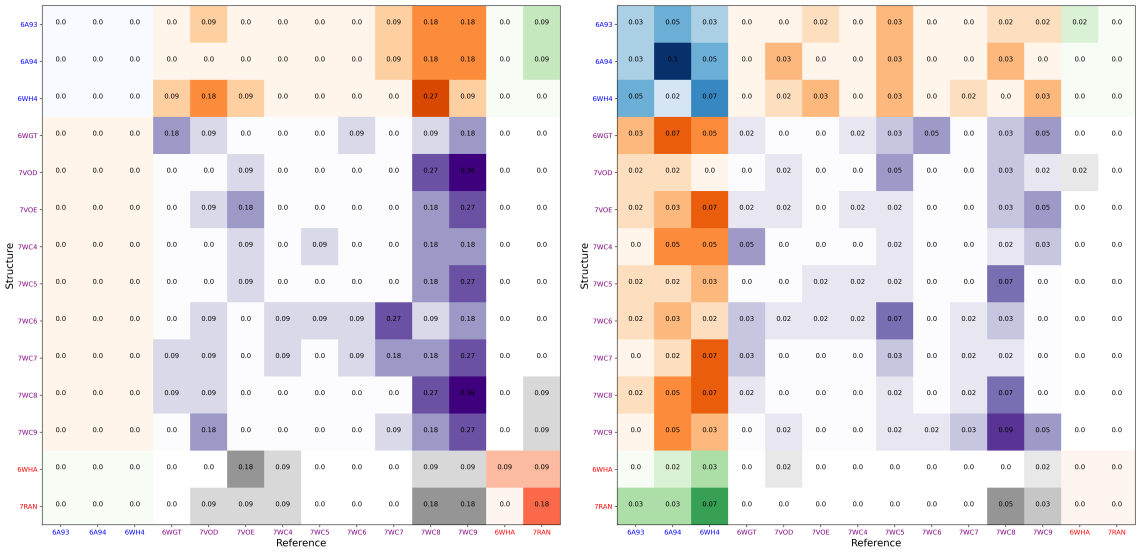
(A) Agonist recall **(B)** Antagonist recall

Figure A.2.29: Fraction of agonist (A) and antagonist (B) of 5HT2A with a IFP Tanimoto similarity score better than the best 1% of evaluated decoys.



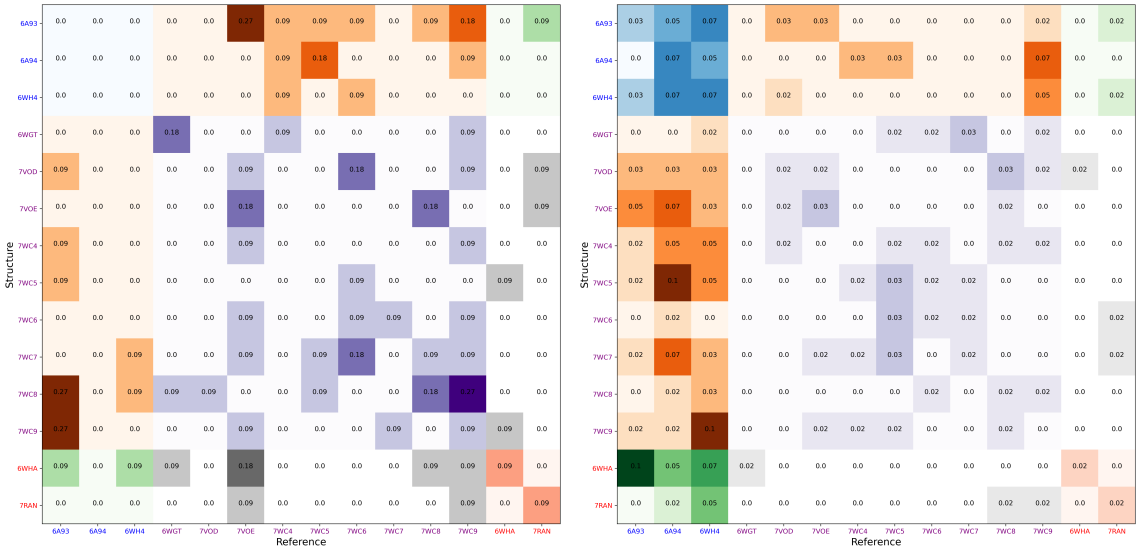
(A) Agonist recall **(B)** Antagonist recall

Figure A.2.30: Fraction of agonist (A) and antagonist (B) of 5HT2A with a kernel on Hyd IPA score better than the best 1% of evaluated decoys.



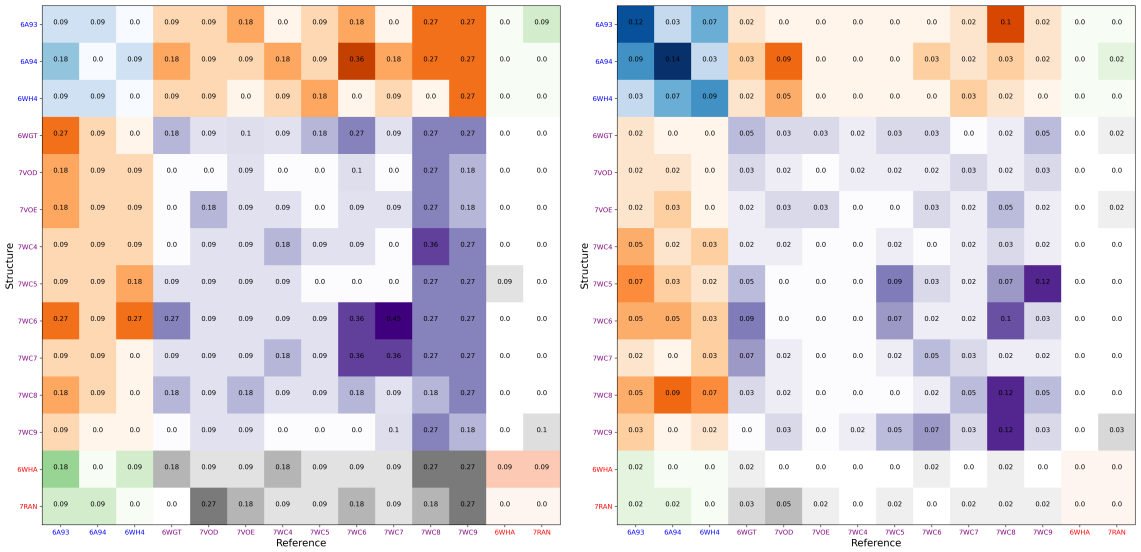
(A) Agonist recall **(B)** Antagonist recall

Figure A.2.31: Fraction of agonist (A) and antagonist (B) of 5HT2A with a kernel on Newhyd IPA score better than the best 1% of evaluated decoys.



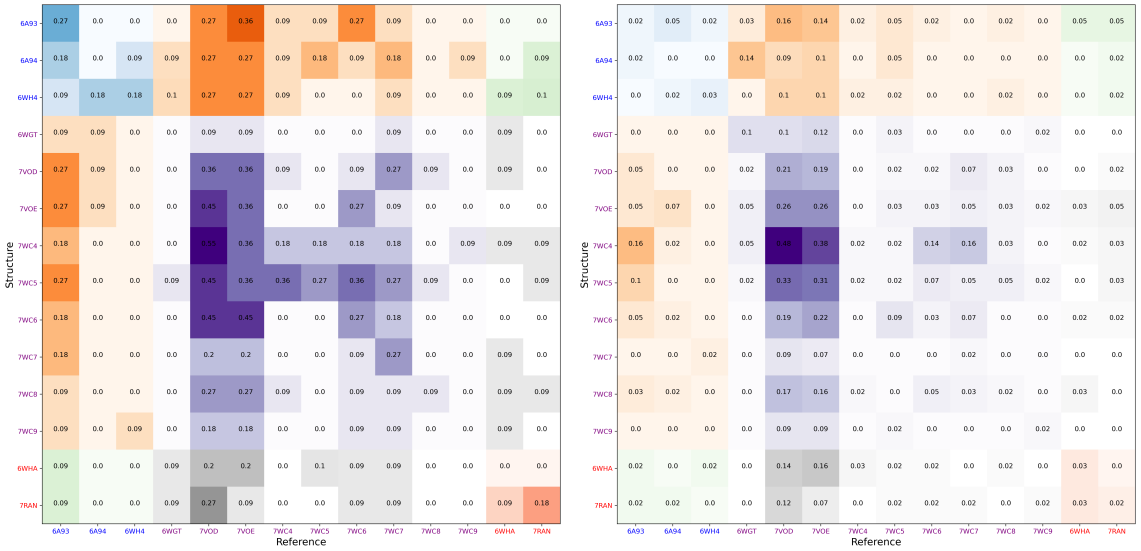
(A) Agonist recall **(B)** Antagonist recall

Figure A.2.32: Fraction of agonist (A) and antagonist (B) of 5HT2A with a GRIM on Hyd IPA score better than the best 1% of evaluated decoys.



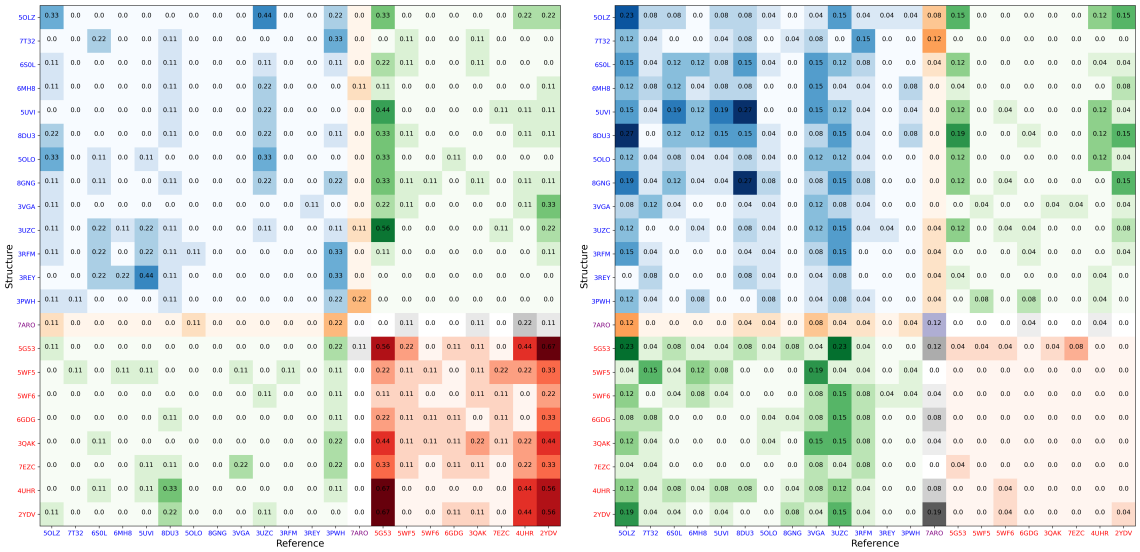
(A) Agonist recall **(B)** Antagonist recall

Figure A.2.33: Fraction of agonist (A) and antagonist (B) of 5HT2A with a GRIM on Newhyd IPA score better than the best 1% of evaluated decoys.



(A) Agonist recall **(B)** Antagonist recall

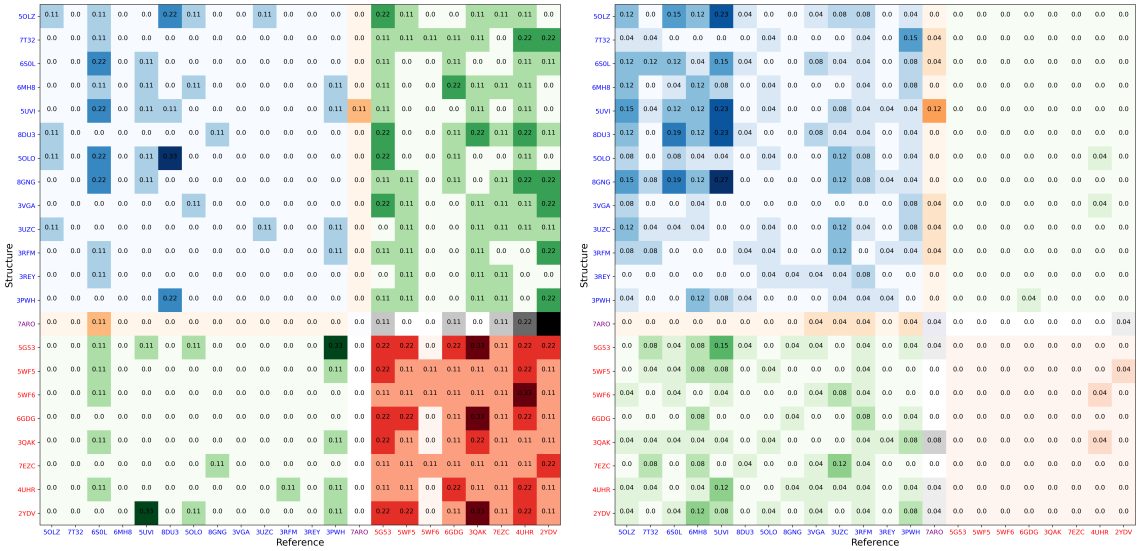
Figure A.2.34: Fraction of agonist (A) and antagonist (B) of AA2AR with a IFP Tanimoto similarity score better than the best 1% of evaluated decoys.



(A) Agonist recall

(B) Antagonist recall

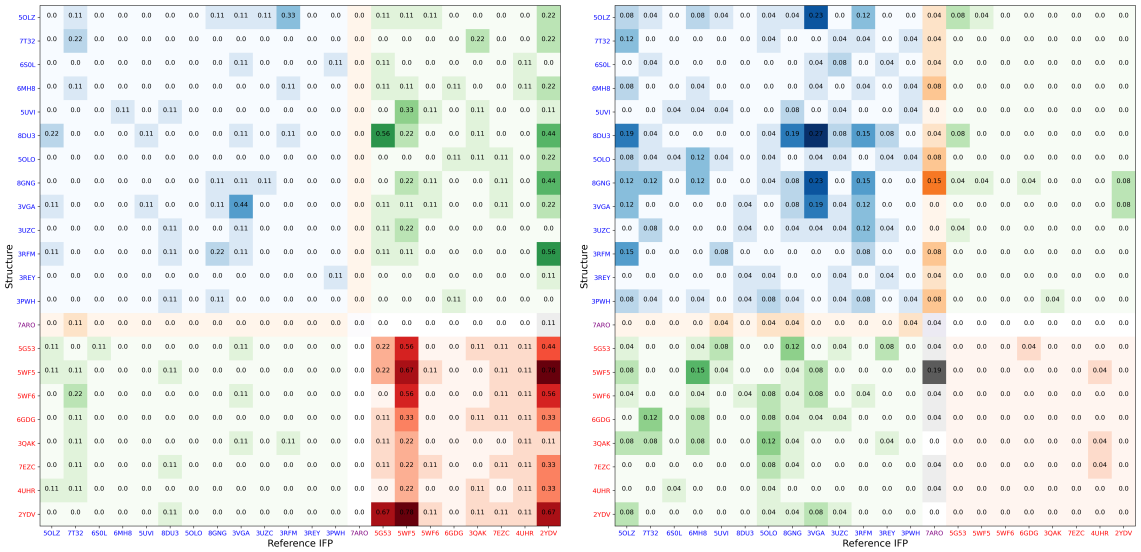
Figure A.2.35: Fraction of agonist (A) and antagonist (B) of AA2AR with a kernel on Hyd IPA score better than the best 1% of evaluated decoys.



(A) Agonist recall

(B) Antagonist recall

Figure A.2.36: Fraction of agonist (A) and antagonist (B) of AA2AR with a kernel on Newhyd IPA score better than the best 1% of evaluated decoys.



(A) Agonist recall

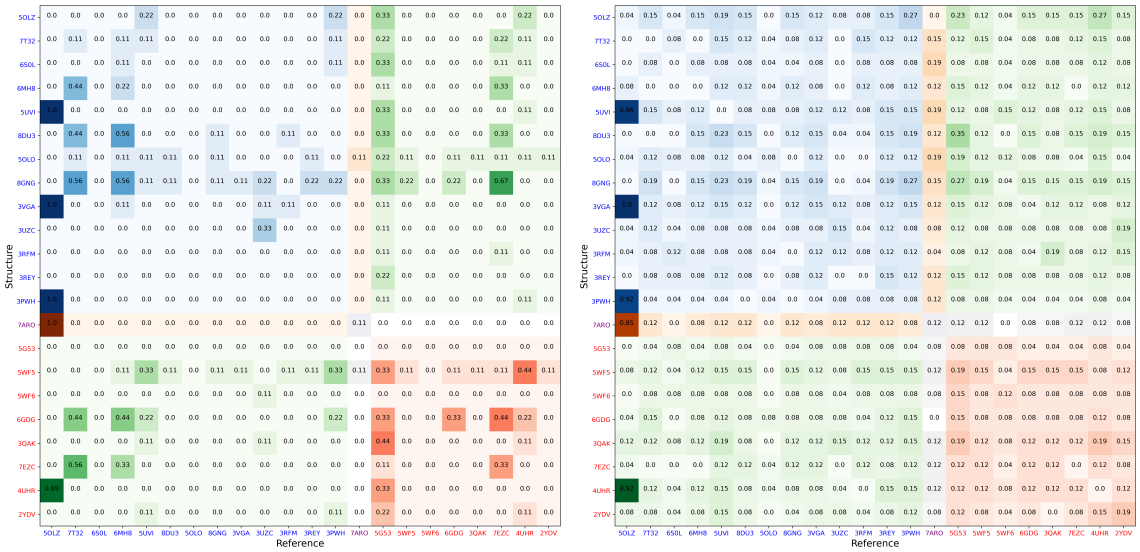
(B) Antagonist recall

Figure A.2.37: Fraction of agonist (A) and antagonist (B) of AA2AR with a GRIM on Hyd IPA score better than the best 1% of evaluated decoys.

(A) Agonist recall

(B) Antagonist recall

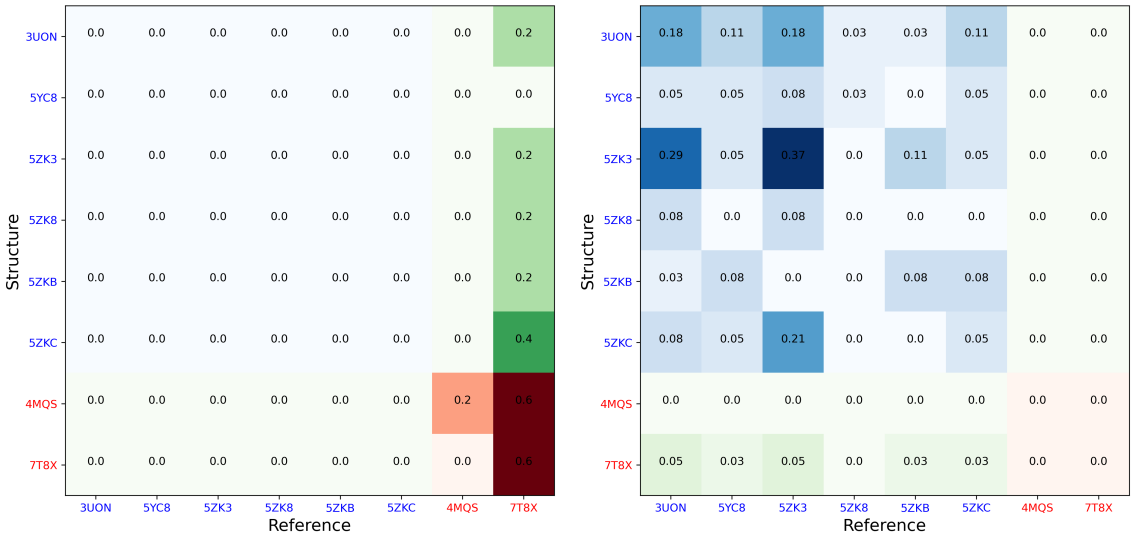
Figure A.2.38: Fraction of agonist (A) and antagonist (B) of AA2AR with a GRIM on Newhyd IPA score better than the best 1% of evaluated decoys.



(A) Agonist recall

(B) Antagonist recall

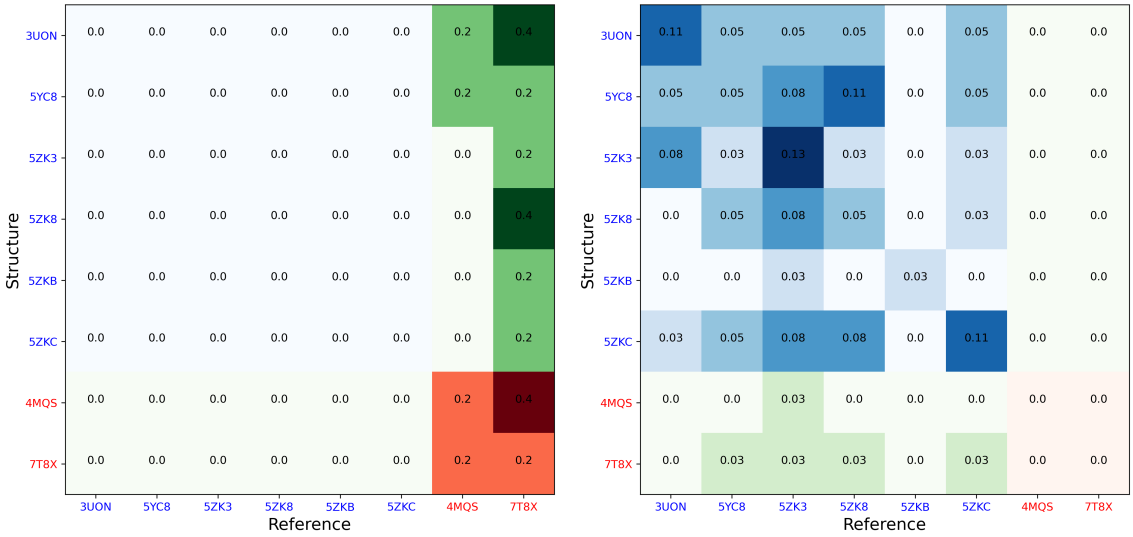
Figure A.2.39: Fraction of agonist (A) and antagonist (B) of ACM2 with a IFP Tanimoto similarity score better than the best 1% of evaluated decoys.



(A) Agonist recall

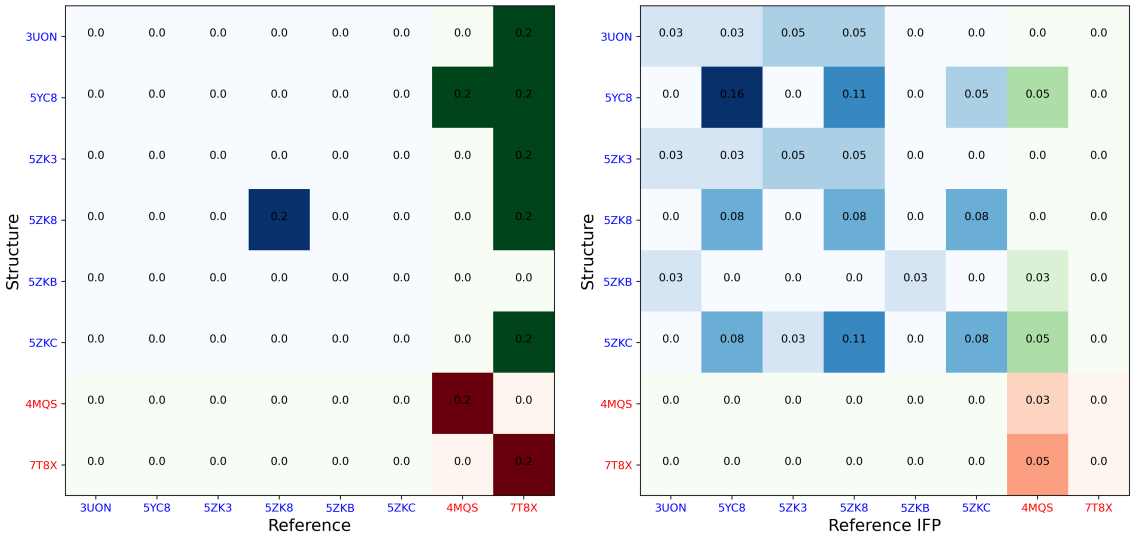
(B) Antagonist recall

Figure A.2.40: Fraction of agonist (A) and antagonist (B) of ACM2 with a kernel on Hyd IPA score better than the best 1% of evaluated decoys.



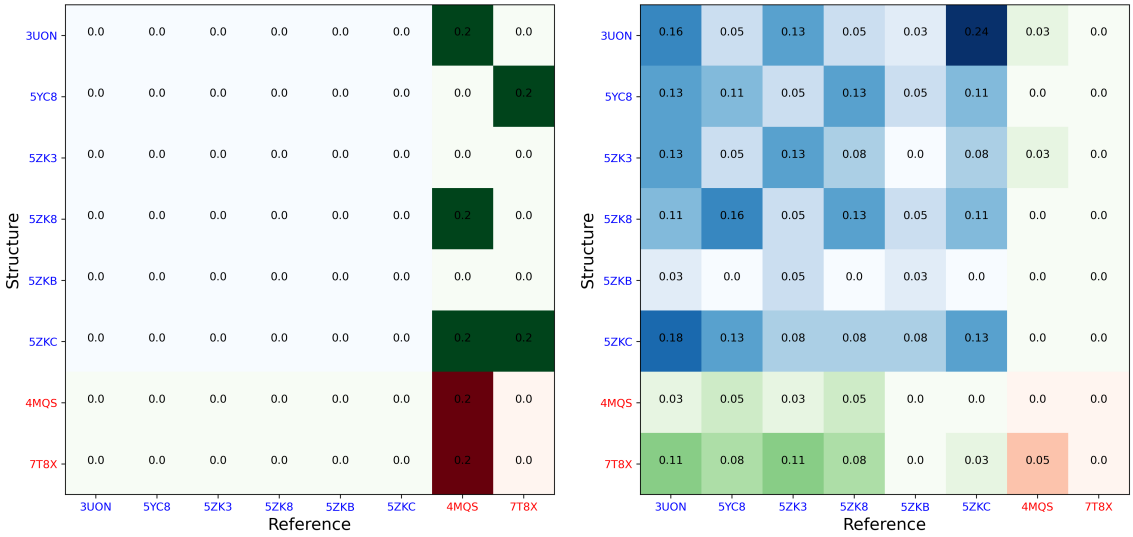
(A) Agonist recall **(B)** Antagonist recall

Figure A.2.41: Fraction of agonist (A) and antagonist (B) of ACM2 with a kernel on Newhyd IPA score better than the best 1% of evaluated decoys.



(A) Agonist recall **(B)** Antagonist recall

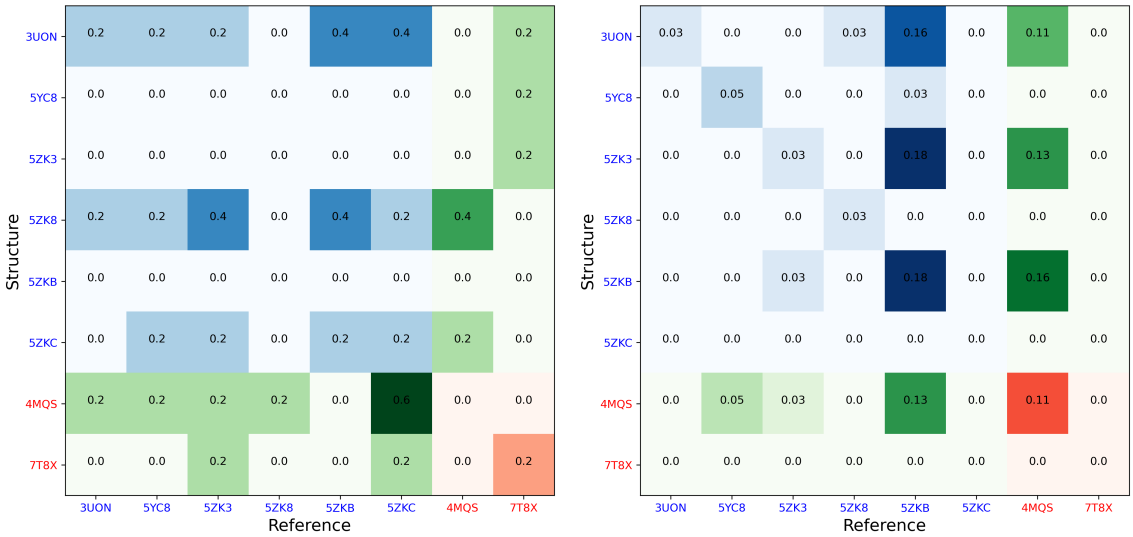
Figure A.2.42: Fraction of agonist (A) and antagonist (B) of ACM2 with a GRIM on Hyd IPA score better than the best 1% of evaluated decoys.



(A) Agonist recall

(B) Antagonist recall

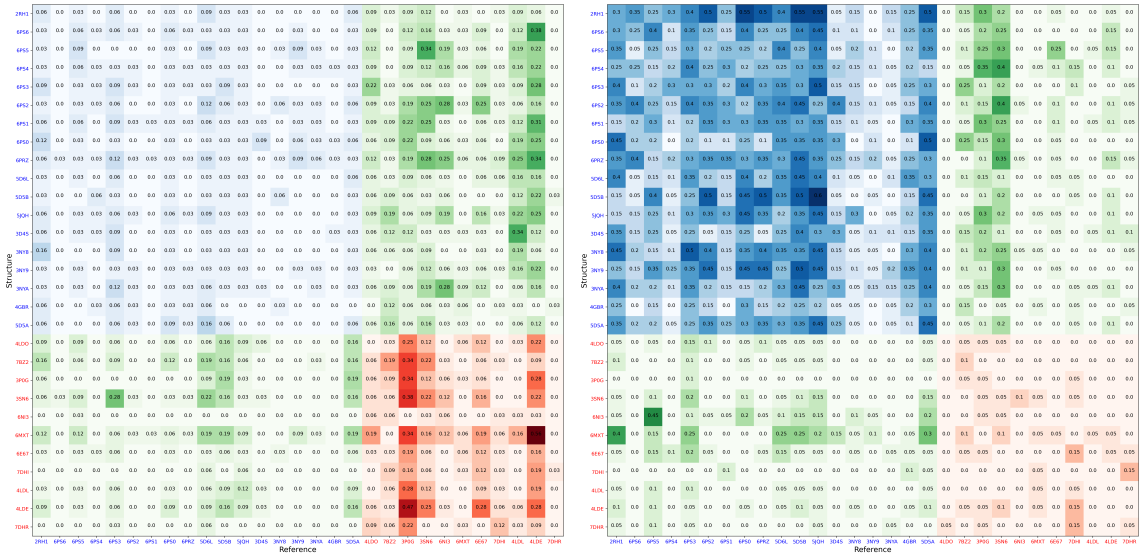
Figure A.2.43: Fraction of agonist (A) and antagonist (B) of ACM2 with a GRIM on Newhyd IPA score better than the best 1% of evaluated decoys.



(A) Agonist recall

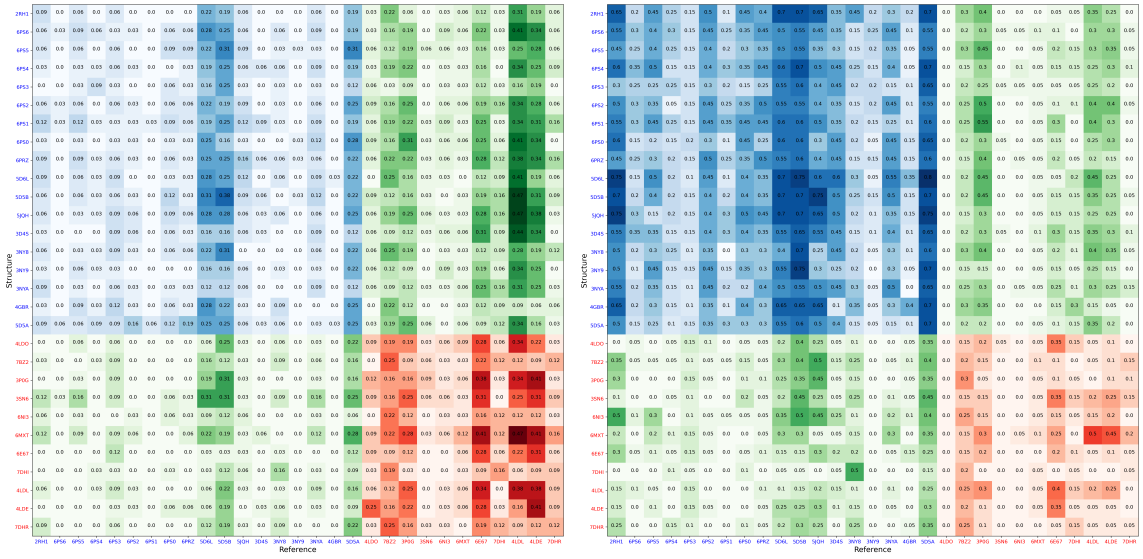
(B) Antagonist recall

Figure A.2.44: Fraction of agonist (A) and antagonist (B) of ADRB2 with a IFP Tanimoto similarity score better than the best 1% of evaluated decoys.



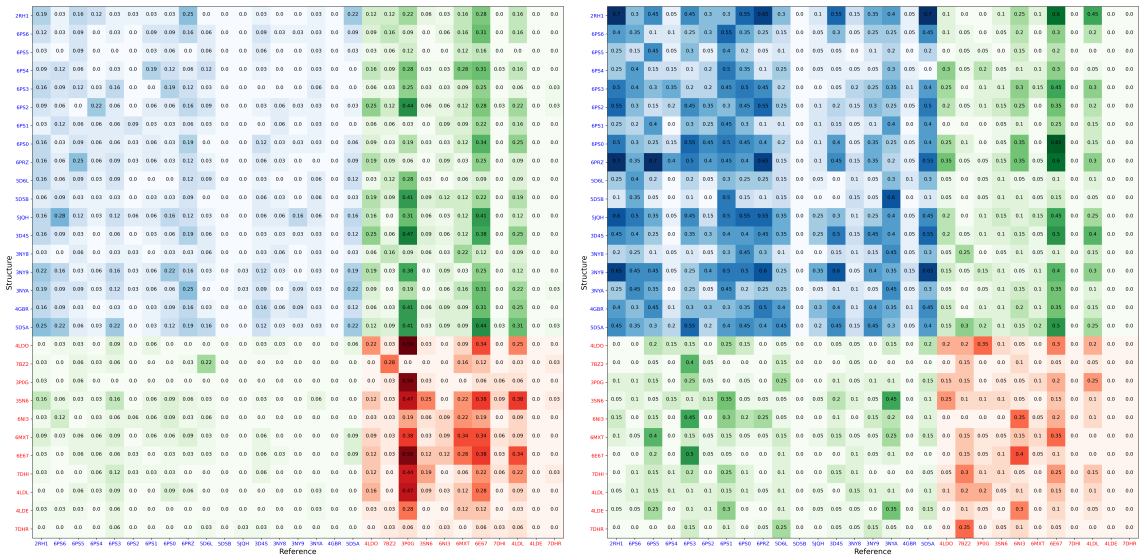
(A) Agonist recall **(B)** Antagonist recall

Figure A.2.45: Fraction of agonist (A) and antagonist (B) of ADRB2 with a kernel on Hyd IPA score better than the best 1% of evaluated decoys.



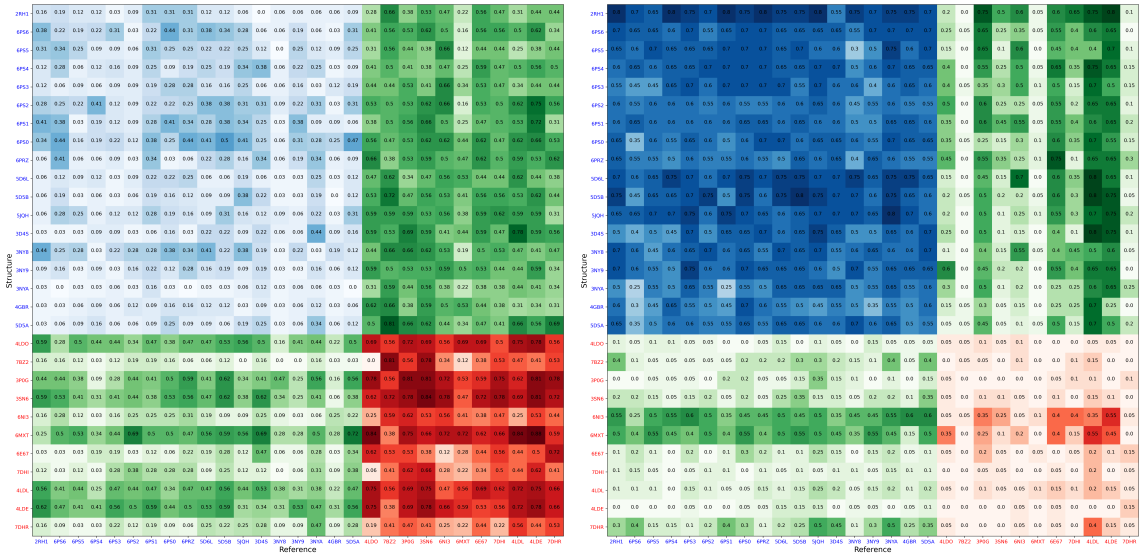
(A) Agonist recall **(B)** Antagonist recall

Figure A.2.46: Fraction of agonist (A) and antagonist (B) of ADRB2 with a kernel on Newhyd IPA score better than the best 1% of evaluated decoys.



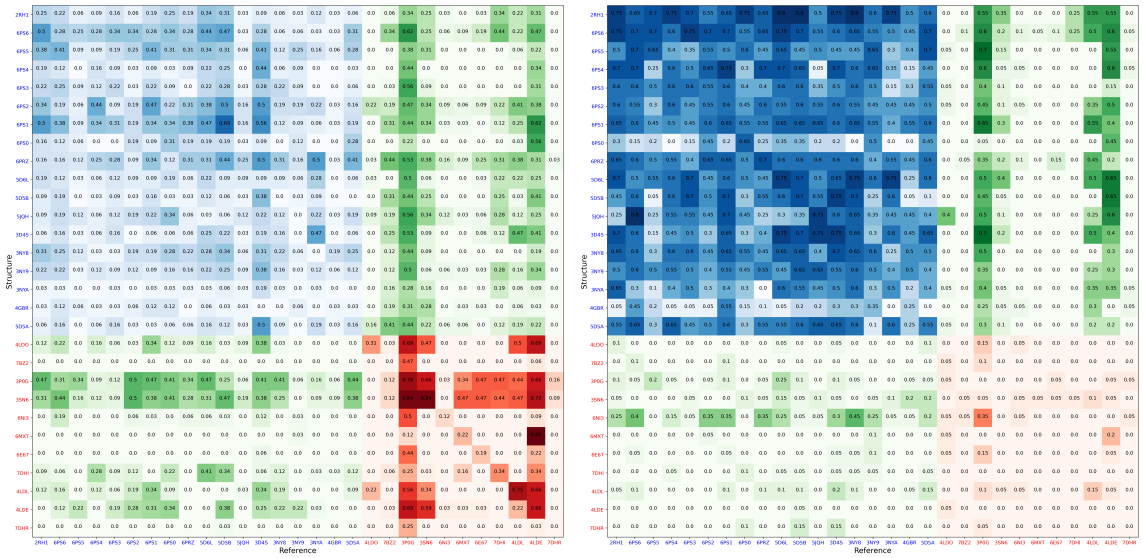
(A) Agonist recall **(B)** Antagonist recall

Figure A.2.47: Fraction of agonist (A) and antagonist (B) of ADRB2 with a GRIM on Hyd IPA score better than the best 1% of evaluated decoys.



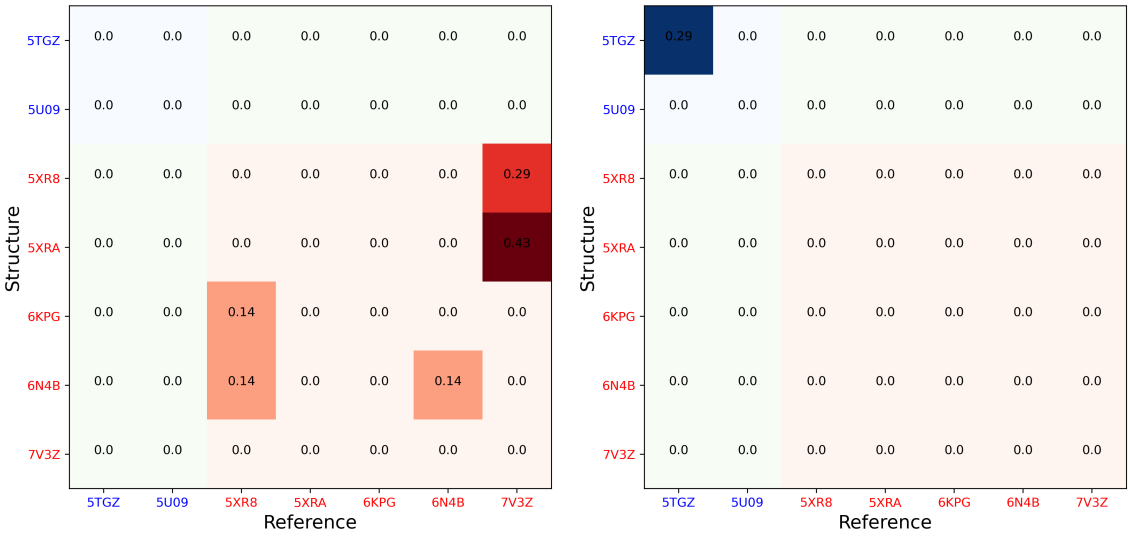
(A) Agonist recall **(B)** Antagonist recall

Figure A.2.48: Fraction of agonist (A) and antagonist (B) of ADRB2 with a GRIM on Newhyd IPA score better than the best 1% of evaluated decoys.



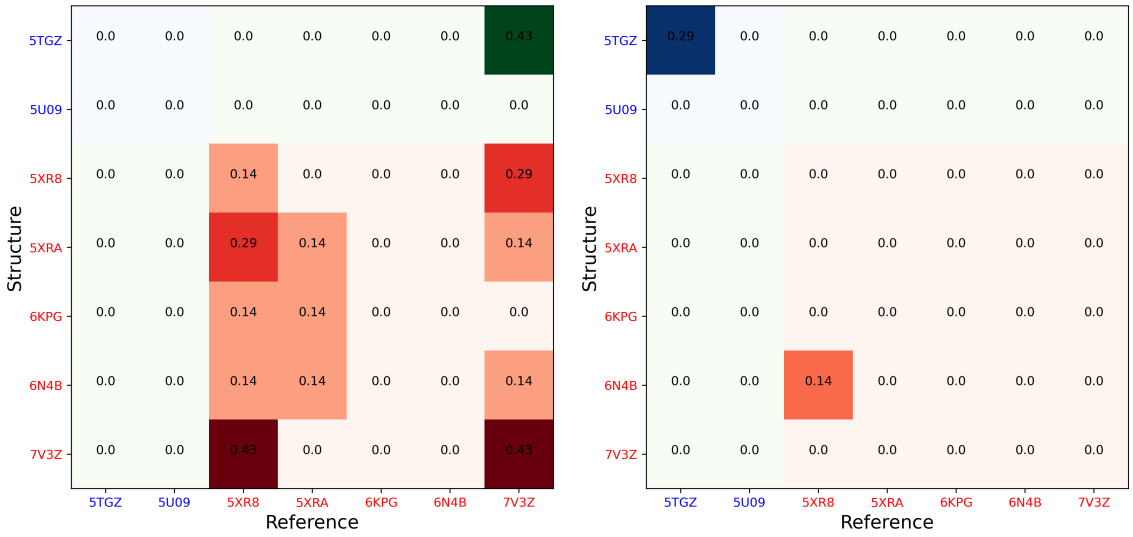
(A) Agonist recall **(B)** Antagonist recall

Figure A.2.49: Fraction of agonist (A) and antagonist (B) of CNR1 with a IFP Tanimoto similarity score better than the best 1% of evaluated decoys.



(A) Agonist recall **(B)** Antagonist recall

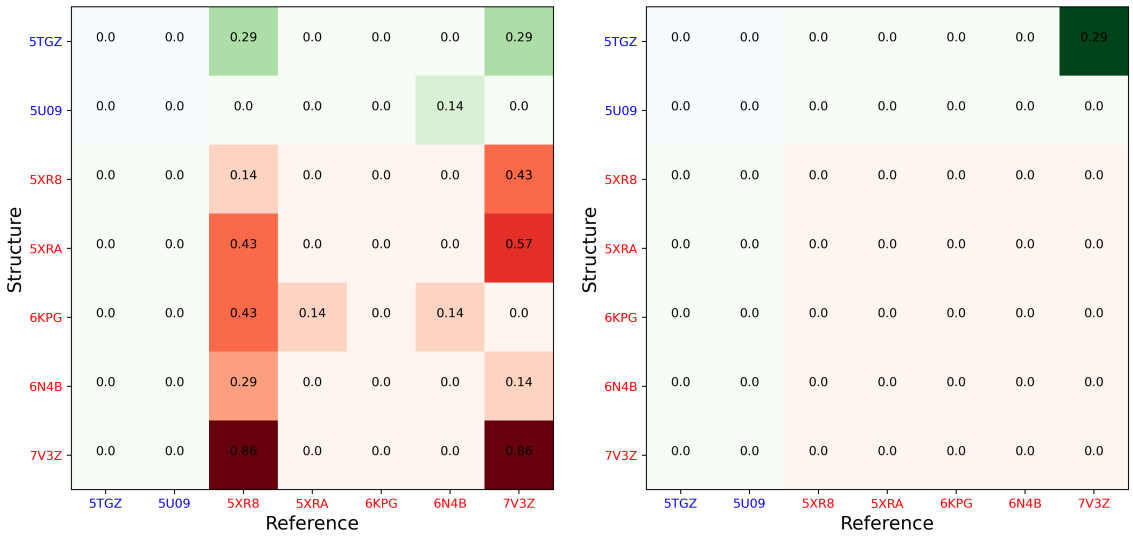
Figure A.2.50: Fraction of agonist (A) and antagonist (B) of CNR1 with a kernel on Hyd IPA score better than the best 1% of evaluated decoys.



(A) Agonist recall

(B) Antagonist recall

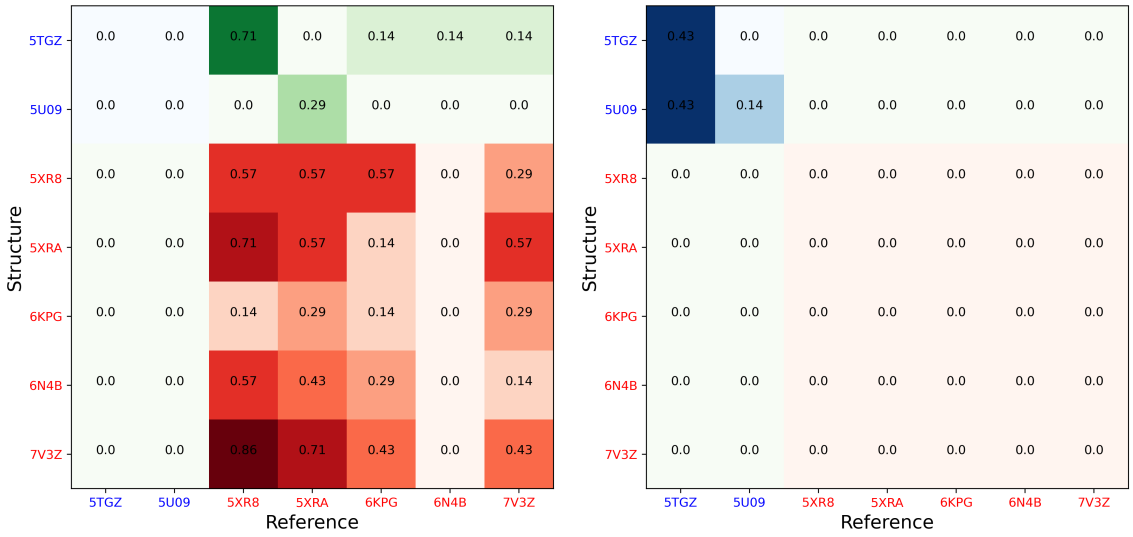
Figure A.2.51: Fraction of agonist (A) and antagonist (B) of CNR1 with a kernel on Newhyd IPA score better than the best 1% of evaluated decoys.



(A) Agonist recall

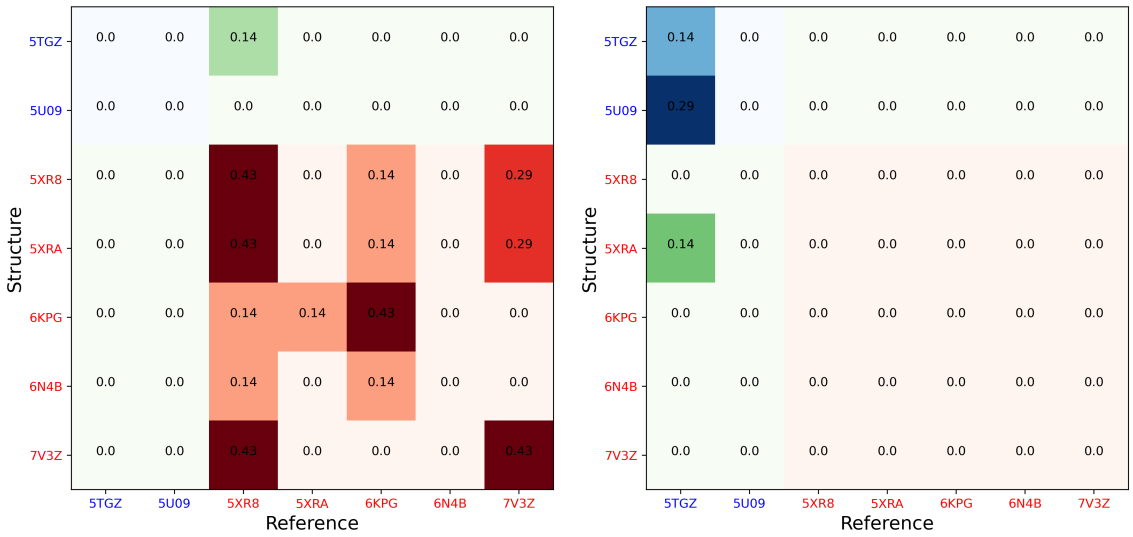
(B) Antagonist recall

Figure A.2.52: Fraction of agonist (A) and antagonist (B) of CNR1 with a GRIM on Hyd IPA score better than the best 1% of evaluated decoys.



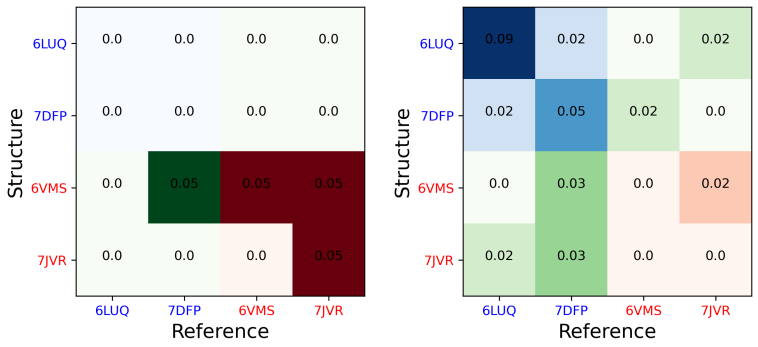
(A) Agonist recall **(B)** Antagonist recall

Figure A.2.53: Fraction of agonist (A) and antagonist (B) of CNR1 with a GRIM on Newhyd IPA score better than the best 1% of evaluated decoys.



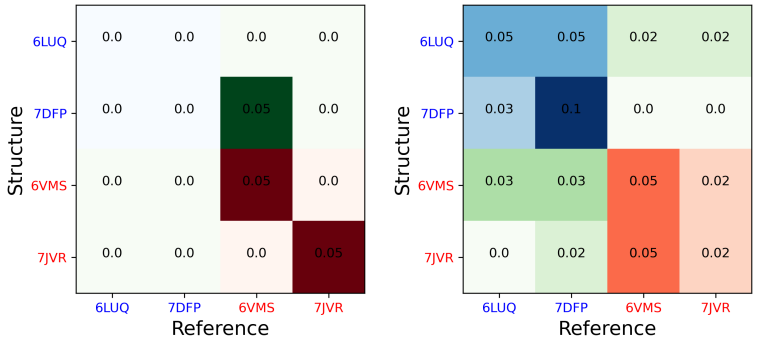
(A) Agonist recall **(B)** Antagonist recall

Figure A.2.54: Fraction of agonist (A) and antagonist (B) of DRD2 with a IFP Tanimoto similarity score better than the best 1% of evaluated decoys.



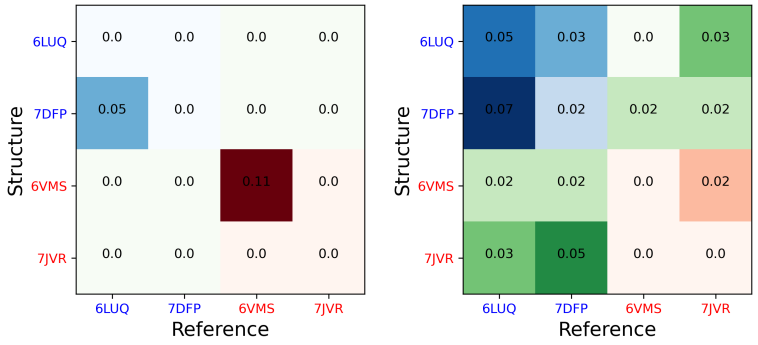
(A) Agonist recall **(B)** Antagonist recall

Figure A.2.55: Fraction of agonist (A) and antagonist (B) of DRD2 with a kernel on Hyd IPA score better than the best 1% of evaluated decoys.



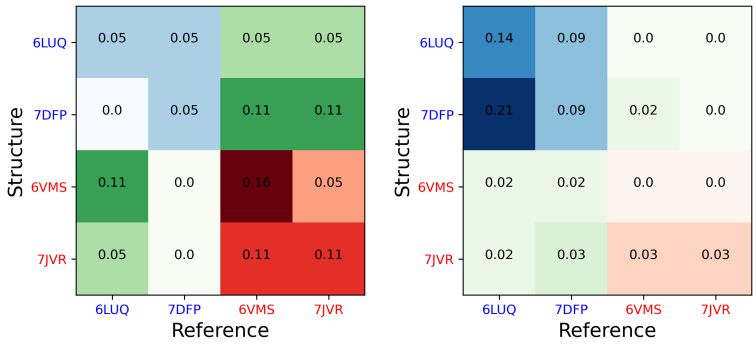
(A) Agonist recall **(B)** Antagonist recall

Figure A.2.56: Fraction of agonist (A) and antagonist (B) of DRD2 with a kernel on Newhyd IPA score better than the best 1% of evaluated decoys.



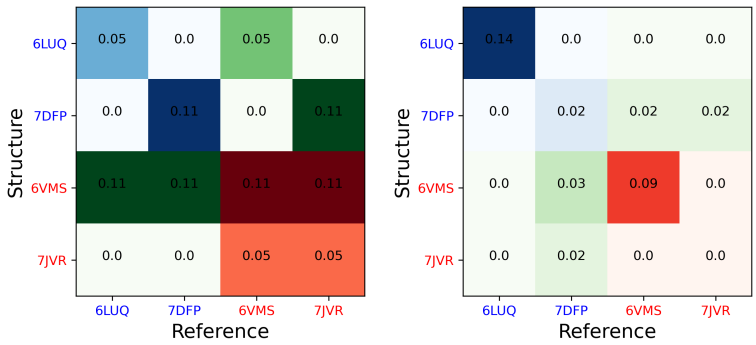
(A) Agonist recall **(B)** Antagonist recall

Figure A.2.57: Fraction of agonist (A) and antagonist (B) of DRD2 with a GRIM on Hyd IPA score better than the best 1% of evaluated decoys.



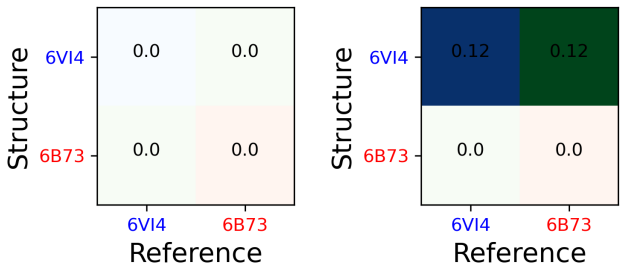
(A) Agonist recall **(B)** Antagonist recall

Figure A.2.58: Fraction of agonist (A) and antagonist (B) of DRD2 with a GRIM on Newhyd IPA score better than the best 1% of evaluated decoys.



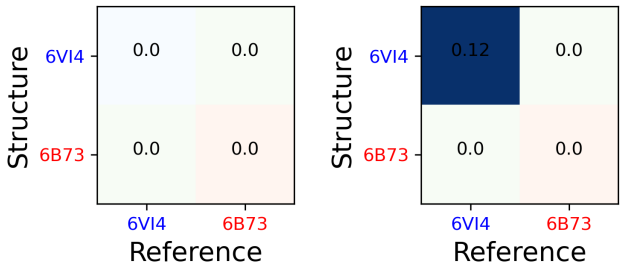
(A) Agonist recall **(B)** Antagonist recall

Figure A.2.59: Fraction of agonist (A) and antagonist (B) of OPRK with a IFP Tanimoto similarity score better than the best 1% of evaluated decoys.



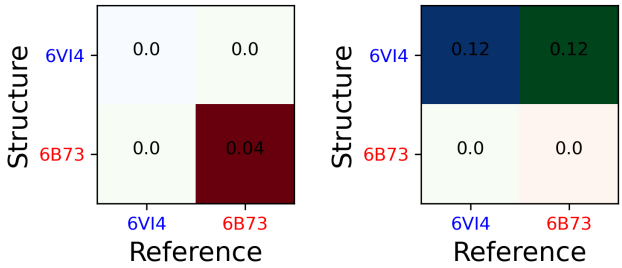
(A) Agonist recall **(B)** Antagonist recall

Figure A.2.60: Fraction of agonist (A) and antagonist (B) of OPRK with a kernel on Hyd IPA score better than the best 1% of evaluated decoys.



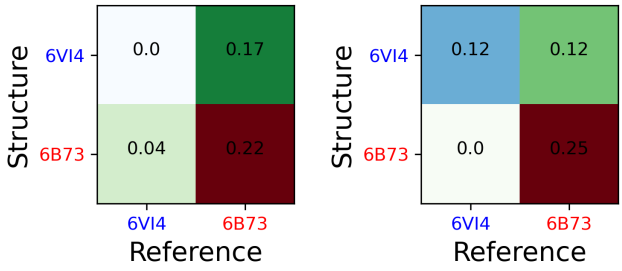
(A) Agonist recall **(B)** Antagonist recall

Figure A.2.61: Fraction of agonist (A) and antagonist (B) of OPRK with a kernel on Newhyd IPA score better than the best 1% of evaluated decoys.



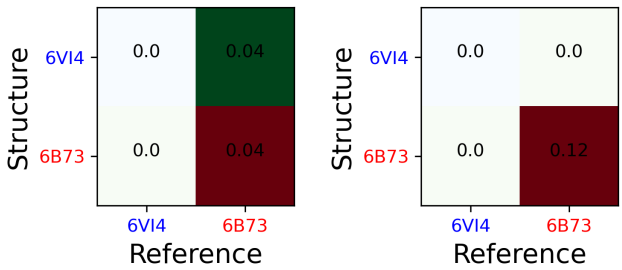
(A) Agonist recall **(B)** Antagonist recall

Figure A.2.62: Fraction of agonist (A) and antagonist (B) of OPRK with a GRIM on Hyd IPA score better than the best 1% of evaluated decoys.



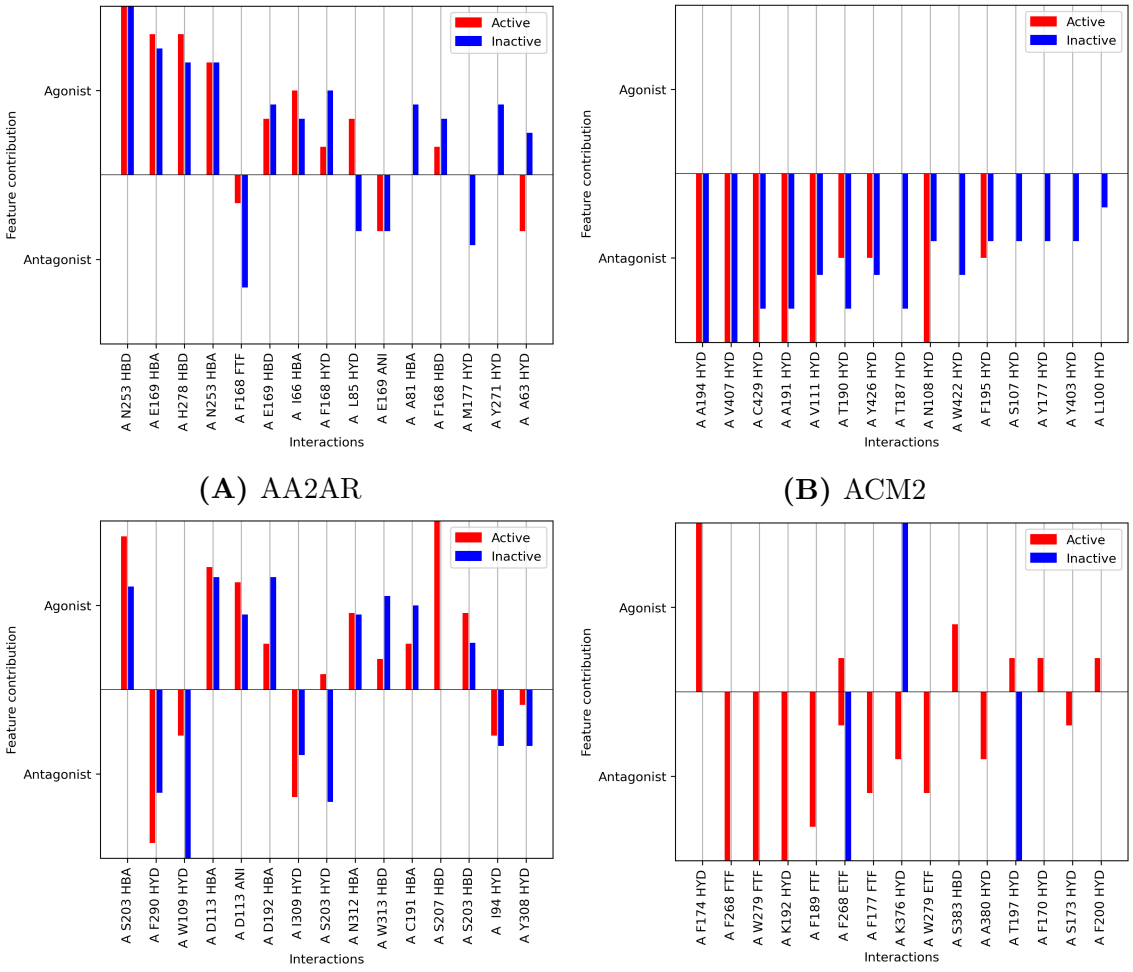
(A) Agonist recall **(B)** Antagonist recall

Figure A.2.63: Fraction of agonist (A) and antagonist (B) of OPRK with a GRIM on Newhyd IPA score better than the best 1% of evaluated decoys.



(A) Agonist recall **(B)** Antagonist recall

Figure A.2.64: Frequency of an interaction being between the ten most important features for efficacy classification for models trained using all generated docking poses. Frequencies are relative to the activation state of the structure used for docking (active or inactive), or by using two structures in different activation states. Results are indicated for the four best performing targets: (A)AA2AR, (B)ACM2, (C)ADRB2, (D)CNR1



(A) AA2AR

(B) ACM2

(C) ADRB2

(D) CNR1

Appendix B

Chapter 3 supporting information

B.1 Section 3.2 supporting information

Table B.1.1: Fraction of agonist and fraction of antagonist with a better docking score than the best 1% best ranked decoys. Reference results are referred to the rankings obtained by docking on the experimental structures.

Receptor	PDB ID	Conformation	Agonist recall	Antagonist recall	Agonist ref	Antagonist ref
5HT2A	6A94	Inactive	0.27	0.31	0.09	0.05
AA2AR	3REY	Inactive	0.22	0.19	0.00	0.00
AA2AR	3UZC	Inactive	0.00	0.00	0.11	0.00
AA2AR	5OLO	Inactive	0.00	0.00	0.00	0.04
AA2AR	5OLZ	Inactive	0.00	0.00	0.00	0.00
AA2AR	5UVI	Inactive	0.00	0.00	0.00	0.00
AA2AR	6MH8	Inactive	0.00	0.00	0.00	0.04
AA2AR	4UHR	Active	0.11	0.00	0.11	0.00
AA2AR	5G53	Active	0.00	0.00	0.22	0.00
ACM2	3UON	Inactive	0.00	0.03	0.00	0.11
ACM2	5ZK3	Inactive	0.00	0.03	0.00	0.11
ACM2	5ZKB	Inactive	0.00	0.05	0.00	0.03
ACM2	4MQS	Active	0.00	0.61	0.00	0.00
ADRB2	3D4S	Inactive	0.25	0.15	0.09	0.00
ADRB2	3NY8	Inactive	0.13	0.05	0.03	0.00
ADRB2	3NY9	Inactive	0.06	0.05	0.09	0.00
ADRB2	3NYA	Inactive	0.09	0.00	0.16	0.00
ADRB2	4GBR	Inactive	0.25	0.10	0.06	0.00
ADRB2	5D6L	Inactive	0.06	0.00	0.03	0.00
ADRB2	5JQH	Inactive	0.00	0.00	0.06	0.00
ADRB2	6PRZ	Inactive	0.19	0.00	0.19	0.00
ADRB2	6PS0	Inactive	0.19	0.00	0.16	0.00
ADRB2	6PS1	Inactive	0.19	0.05	0.13	0.00
ADRB2	6PS2	Inactive	0.09	0.00	0.16	0.00
ADRB2	6PS3	Inactive	0.19	0.05	0.06	0.00
ADRB2	6PS4	Inactive	0.09	0.00	0.06	0.00
ADRB2	6PS6	Inactive	0.13	0.00	0.19	0.00
ADRB2	4LDL	Active	0.16	0.00	0.13	0.00
CNR1	5U09	Inactive	0.00	0.00	0.00	0.14
CNR1	5XRA	Active	0.14	0.00	0.14	0.00
DRD2	6LUQ	Inactive	0.21	0.33	0.00	0.05

Table B.1.2: Fraction of agonist and fraction of antagonist with a better docking score than the best 1% best ranked decoys, after performing pose filtering using the IFP-OCSVM models. Reference results are referred to the rankings obtained by docking on the experimental structures, and rescoring based on IFP similarity to the co-crystallized ligand.

Receptor	PDB ID	Conformation	Agonist recall QMS2	Antagonist recall QMS2	Agonist recall MAD	Antagonist recall MAD	Agonist ref	Antagonist ref
5HT2A	6A94	Inactive	0.09	0.45	0.18	0.55	0.00	0.26
AA2AR	3REY	Inactive	0.22	0.12	0.22	0.12	0.00	0.00
AA2AR	3UZC	Inactive	0.11	0.00	0.00	0.00	0.11	0.15
AA2AR	5OLO	Inactive	0.00	0.00	0.11	0.00	0.00	0.04
AA2AR	5OLZ	Inactive	0.00	0.12	0.00	0.15	0.33	0.23
AA2AR	5UVI	Inactive	0.00	0.00	0.00	0.04	0.00	0.19
AA2AR	6MH8	Inactive	0.11	0.04	0.00	0.00	0.00	0.04
AA2AR	4UHR	Active	0.22	0.08	0.22	0.00	0.44	0.00
AA2AR	5G53	Active	0.11	0.00	0.22	0.04	0.56	0.04
ACM2	3UON	Inactive	0.00	0.21	0.00	0.50	0.00	0.18
ACM2	5ZK3	Inactive	0.00	0.08	0.00	0.21	0.00	0.37
ACM2	5ZKB	Inactive	0.00	0.03	0.00	0.03	0.00	0.08
ACM2	4MQS	Active	0.20	0.08	0.40	0.39	0.20	0.00
ADRB2	3D4S	Inactive	0.09	0.55	0.19	0.70	0.00	0.30
ADRB2	3NY8	Inactive	0.06	0.30	0.28	0.55	0.00	0.05
ADRB2	3NY9	Inactive	0.00	0.10	0.00	0.10	0.03	0.05
ADRB2	3NYA	Inactive	0.00	0.35	0.03	0.15	0.03	0.15
ADRB2	4GBR	Inactive	0.00	0.10	0.00	0.40	0.00	0.20
ADRB2	5D6L	Inactive	0.03	0.30	0.03	0.35	0.03	0.35
ADRB2	5JQH	Inactive	0.03	0.40	0.03	0.35	0.03	0.45
ADRB2	6PRZ	Inactive	0.00	0.20	0.03	0.05	0.03	0.35
ADRB2	6PS0	Inactive	0.03	0.35	0.03	0.55	0.03	0.25
ADRB2	6PS1	Inactive	0.00	0.10	0.00	0.25	0.03	0.30
ADRB2	6PS2	Inactive	0.00	0.10	0.00	0.20	0.00	0.35
ADRB2	6PS3	Inactive	0.00	0.05	0.00	0.15	0.09	0.30
ADRB2	6PS4	Inactive	0.06	0.15	0.09	0.05	0.03	0.20
ADRB2	6PS6	Inactive	0.00	0.20	0.00	0.10	0.00	0.25
ADRB2	4LDL	Active	0.19	0.00	0.25	0.05	0.00	0.00
CNR1	5U09	Inactive	0.00	0.14	0.00	0.14	0.00	0.00
CNR1	5XRA	Active	0.00	0.00	0.00	0.00	0.00	0.00
DRD2	6LUQ	Inactive	0.05	0.10	0.11	0.28	0.00	0.09

Table B.1.3: Fraction of agonist and fraction of antagonist with a better docking score than the best 1% best ranked decoys, after performing pose filtering using the Hyd IG-OCSVM models. Reference results are referred to the rankings obtained by docking on the experimental structures, and rescoring based on the interaction similarity to the co-crystallized ligand measured using a graph kernel on Hyd IPAs.

Receptor	PDB ID	Conformation	Agonist recall QMS2	Antagonist recall QMS2	Agonist recall MAD	Antagonist recall MAD	Agonist ref	Antagonist ref
5HT2A	6A94	Inactive	0.09	0.28	0.09	0.21	0.00	0.26
AA2AR	3REY	Inactive	0.22	0.08	0.22	0.08	0.00	0.00
AA2AR	3UZC	Inactive	0.00	0.04	0.00	0.04	0.11	0.15
AA2AR	5OLO	Inactive	0.00	0.00	0.00	0.00	0.00	0.04
AA2AR	5OLZ	Inactive	0.00	0.08	0.00	0.08	0.33	0.23
AA2AR	5UVI	Inactive	0.00	0.00	0.00	0.00	0.00	0.19
AA2AR	6MH8	Inactive	0.00	0.04	0.00	0.00	0.00	0.04
AA2AR	4UHR	Active	0.22	0.00	0.22	0.00	0.44	0.00
AA2AR	5G53	Active	0.11	0.04	0.00	0.00	0.56	0.04
ACM2	3UON	Inactive	0.00	0.05	0.00	0.08	0.00	0.18
ACM2	5ZK3	Inactive	0.00	0.24	0.00	0.24	0.00	0.37
ACM2	5ZKB	Inactive	0.00	0.00	0.00	0.00	0.00	0.08
ACM2	4MQS	Active	0.00	0.08	0.00	0.13	0.20	0.00
ADRB2	3D4S	Inactive	0.06	0.40	0.09	0.35	0.00	0.30
ADRB2	3NY8	Inactive	0.13	0.20	0.13	0.35	0.00	0.05
ADRB2	3NY9	Inactive	0.00	0.10	0.03	0.10	0.03	0.05
ADRB2	3NYA	Inactive	0.09	0.35	0.19	0.45	0.03	0.15
ADRB2	4GBR	Inactive	0.13	0.50	0.16	0.50	0.00	0.20
ADRB2	5D6L	Inactive	0.03	0.20	0.03	0.15	0.03	0.35
ADRB2	5JQH	Inactive	0.06	0.10	0.09	0.15	0.03	0.45
ADRB2	6PRZ	Inactive	0.22	0.15	0.22	0.15	0.03	0.35
ADRB2	6PS0	Inactive	0.13	0.20	0.03	0.15	0.03	0.25
ADRB2	6PS1	Inactive	0.03	0.25	0.06	0.25	0.03	0.30
ADRB2	6PS2	Inactive	0.09	0.10	0.06	0.15	0.00	0.35
ADRB2	6PS3	Inactive	0.16	0.10	0.03	0.10	0.09	0.30
ADRB2	6PS4	Inactive	0.06	0.10	0.13	0.00	0.03	0.20
ADRB2	6PS6	Inactive	0.09	0.05	0.09	0.30	0.00	0.25
ADRB2	4LDL	Active	0.31	0.15	0.34	0.10	0.00	0.00
CNR1	5U09	Inactive	0.00	0.00	0.00	0.00	0.00	0.00
CNR1	5XRA	Active	0.00	0.00	0.00	0.00	0.00	0.00
DRD2	6LUQ	Inactive	0.11	0.16	0.11	0.31	0.00	0.09

Table B.1.4: Fraction of agonist and fraction of antagonist with a better docking score than the best 1% best ranked decoys, after performing pose filtering using the Hyd IG-OCSVM models. Reference results are referred to the rankings obtained by docking on the experimental structures, and rescoring based on the interaction similarity to the co-crystallized ligand measured using a graph kernel on Hyd IPAs.

Receptor	PDB ID	Conformation	Agonist recall QMS2	Antagonist recall QMS2	Agonist recall MAD	Antagonist recall MAD	Agonist ref	Antagonist ref
5HT2A	6A94	Inactive	0.09	0.19	0.09	0.16	0.00	0.26
AA2AR	3REY	Inactive	0.22	0.12	0.22	0.15	0.00	0.00
AA2AR	3UZC	Inactive	0.00	0.00	0.00	0.04	0.11	0.15
AA2AR	5OLO	Inactive	0.00	0.00	0.00	0.00	0.00	0.04
AA2AR	5OLZ	Inactive	0.00	0.08	0.11	0.12	0.33	0.23
AA2AR	5UVI	Inactive	0.00	0.00	0.00	0.00	0.00	0.19
AA2AR	6MH8	Inactive	0.00	0.00	0.00	0.00	0.00	0.04
AA2AR	4UHR	Active	0.22	0.00	0.22	0.04	0.44	0.00
AA2AR	5G53	Active	0.11	0.00	0.22	0.00	0.56	0.04
ACM2	3UON	Inactive	0.00	0.03	0.00	0.03	0.00	0.18
ACM2	5ZK3	Inactive	0.00	0.00	0.00	0.05	0.00	0.37
ACM2	5ZKB	Inactive	0.00	0.00	0.00	0.00	0.00	0.08
ACM2	4MQS	Active	0.00	0.47	0.00	0.34	0.20	0.00
ADRB2	3D4S	Inactive	0.22	0.30	0.19	0.25	0.00	0.30
ADRB2	3NY8	Inactive	0.09	0.15	0.16	0.30	0.00	0.05
ADRB2	3NY9	Inactive	0.06	0.10	0.13	0.10	0.03	0.05
ADRB2	3NYA	Inactive	0.06	0.20	0.09	0.20	0.03	0.15
ADRB2	4GBR	Inactive	0.16	0.55	0.41	0.40	0.00	0.20
ADRB2	5D6L	Inactive	0.03	0.20	0.03	0.15	0.03	0.35
ADRB2	5JQH	Inactive	0.13	0.20	0.13	0.20	0.03	0.45
ADRB2	6PRZ	Inactive	0.16	0.15	0.22	0.25	0.03	0.35
ADRB2	6PS0	Inactive	0.09	0.30	0.06	0.15	0.03	0.25
ADRB2	6PS1	Inactive	0.22	0.25	0.19	0.30	0.03	0.30
ADRB2	6PS2	Inactive	0.00	0.15	0.06	0.15	0.00	0.35
ADRB2	6PS3	Inactive	0.22	0.20	0.13	0.25	0.09	0.30
ADRB2	6PS4	Inactive	0.06	0.00	0.13	0.00	0.03	0.20
ADRB2	6PS6	Inactive	0.06	0.05	0.03	0.10	0.00	0.25
ADRB2	4LDL	Active	0.38	0.10	0.44	0.05	0.00	0.00
CNR1	5U09	Inactive	0.00	0.14	0.00	0.14	0.00	0.00
CNR1	5XRA	Active	0.00	0.00	0.00	0.00	0.00	0.00
DRD2	6LUQ	Inactive	0.16	0.36	0.16	0.36	0.00	0.09

Table B.1.5: Fraction of agonist and fraction of antagonist with a better docking score than the best 1% best ranked decoys, after performing pose filtering using the Hyd IG-OCSVM+IFP-OCSVM ensemble models. Reference results are referred to the best rankings obtained by docking on the experimental structures, and rescoring based on IFP similarity or the interaction similarity to the co-crystallized ligand measured using a graph kernel on Hyd IPAs.

Receptor	PDB ID	Conformation	Agonist recall QMS2	Antagonist recall QMS2	Agonist recall MAD	Antagonist recall MAD	Agonist ref	Antagonist ref
5HT2A	6A94	Inactive	0.09	0.31	0.09	0.34	0.00	0.26
AA2AR	3REY	Inactive	0.11	0.15	0.33	0.27	0.00	0.00
AA2AR	3UZC	Inactive	0.67	0.46	0.67	0.50	0.11	0.15
AA2AR	5OLO	Inactive	0.11	0.19	0.22	0.23	0.00	0.04
AA2AR	5OLZ	Inactive	0.00	0.08	0.00	0.08	0.33	0.23
AA2AR	5UVI	Inactive	0.22	0.12	0.33	0.42	0.11	0.23
AA2AR	6MH8	Inactive	0.00	0.19	0.44	0.54	0.00	0.12
AA2AR	4UHR	Active	0.22	0.04	0.22	0.12	0.44	0.00
AA2AR	5G53	Active	0.11	0.00	0.11	0.00	0.56	0.04
ACM2	3UON	Inactive	0.00	0.18	0.00	0.32	0.00	0.18
ACM2	5ZK3	Inactive	0.00	0.08	0.00	0.13	0.00	0.37
ACM2	5ZKB	Inactive	0.00	0.00	0.00	0.00	0.00	0.08
ACM2	4MQS	Active	0.20	0.03	0.40	0.08	0.20	0.00
ADRB2	3D4S	Inactive	0.03	0.40	0.09	0.55	0.03	0.45
ADRB2	3NY8	Inactive	0.00	0.10	0.03	0.20	0.00	0.20
ADRB2	3NY9	Inactive	0.00	0.05	0.00	0.05	0.03	0.05
ADRB2	3NYA	Inactive	0.00	0.35	0.03	0.55	0.09	0.50
ADRB2	4GBR	Inactive	0.00	0.05	0.00	0.25	0.00	0.40
ADRB2	5D6L	Inactive	0.03	0.10	0.03	0.20	0.28	0.70
ADRB2	5JQH	Inactive	0.03	0.20	0.03	0.25	0.06	0.65
ADRB2	6PRZ	Inactive	0.03	0.40	0.06	0.55	0.03	0.50
ADRB2	6PS0	Inactive	0.03	0.05	0.03	0.10	0.03	0.45
ADRB2	6PS1	Inactive	0.00	0.35	0.03	0.50	0.03	0.35
ADRB2	6PS2	Inactive	0.00	0.10	0.00	0.25	0.03	0.45
ADRB2	6PS3	Inactive	0.00	0.05	0.00	0.10	0.09	0.30
ADRB2	6PS4	Inactive	0.06	0.30	0.19	0.65	0.03	0.20
ADRB2	6PS6	Inactive	0.00	0.15	0.03	0.20	0.03	0.30
ADRB2	4LDL	Active	0.13	0.00	0.41	0.00	0.38	0.20
CNR1	5U09	Inactive	0.00	0.14	0.00	0.14	0.00	0.00
CNR1	5XRA	Active	0.43	0.00	0.43	0.14	0.14	0.00
DRD2	6LUQ	Inactive	0.05	0.17	0.21	0.43	0.00	0.09

Table B.1.6: Fraction of agonist and fraction of antagonist with a better docking score than the best 1% best ranked decoys, after performing pose filtering using the Newhyd IG-OCSVM+IFP-OCSVM ensemble models. Reference results are referred to the best rankings obtained by docking on the experimental structures, and rescoring based on IFP similarity or the interaction similarity to the co-crystallized ligand measured using a graph kernel on Newyd IPAs.

Receptor	PDB ID	Conformation	Agonist recall QMS2	Antagonist recall QMS2	Agonist recall MAD	Antagonist recall MAD	Agonist ref	Antagonist ref
5HT2A	6A94	Inactive	0.09	0.33	0.09	0.50	0.00	0.26
AA2AR	3REY	Inactive	0.11	0.19	0.67	0.62	0.00	0.04
AA2AR	3UZC	Inactive	0.56	0.46	0.89	0.77	0.11	0.15
AA2AR	5OLO	Inactive	0.11	0.31	0.33	0.50	0.00	0.04
AA2AR	5OLZ	Inactive	0.00	0.12	0.00	0.12	0.33	0.23
AA2AR	5UVI	Inactive	0.33	0.42	0.56	0.73	0.00	0.19
AA2AR	6MH8	Inactive	0.11	0.31	0.44	0.65	0.00	0.04
AA2AR	4UHR	Active	0.22	0.08	0.33	0.23	0.44	0.00
AA2AR	5G53	Active	0.00	0.00	0.22	0.00	0.56	0.04
ACM2	3UON	Inactive	0.00	0.18	0.00	0.45	0.00	0.18
ACM2	5ZK3	Inactive	0.00	0.08	0.00	0.11	0.00	0.37
ACM2	5ZKB	Inactive	0.00	0.00	0.00	0.00	0.00	0.08
ACM2	4MQS	Active	0.20	0.08	0.40	0.32	0.20	0.03
ADRB2	3D4S	Inactive	0.09	0.45	0.28	0.70	0.03	0.50
ADRB2	3NY8	Inactive	0.00	0.15	0.28	0.50	0.06	0.10
ADRB2	3NY9	Inactive	0.00	0.05	0.00	0.25	0.03	0.40
ADRB2	3NYA	Inactive	0.00	0.45	0.09	0.60	0.03	0.45
ADRB2	4GBR	Inactive	0.00	0.05	0.00	0.25	0.00	0.20
ADRB2	5D6L	Inactive	0.03	0.10	0.03	0.35	0.09	0.35
ADRB2	5JQH	Inactive	0.03	0.20	0.03	0.35	0.03	0.45
ADRB2	6PRZ	Inactive	0.03	0.30	0.06	0.60	0.13	0.65
ADRB2	6PS0	Inactive	0.03	0.10	0.03	0.05	0.03	0.45
ADRB2	6PS1	Inactive	0.06	0.30	0.13	0.55	0.03	0.45
ADRB2	6PS2	Inactive	0.00	0.10	0.00	0.30	0.00	0.35
ADRB2	6PS3	Inactive	0.00	0.05	0.00	0.15	0.16	0.30
ADRB2	6PS4	Inactive	0.34	0.60	0.53	0.90	0.03	0.20
ADRB2	6PS6	Inactive	0.00	0.20	0.06	0.30	0.03	0.35
ADRB2	4LDL	Active	0.16	0.00	0.38	0.00	0.09	0.10
CNR1	5U09	Inactive	0.29	0.14	0.29	0.14	0.00	0.00
CNR1	5XRA	Active	0.29	0.00	0.57	0.00	0.00	0.00
DRD2	6LUQ	Inactive	0.05	0.19	0.53	0.81	0.00	0.09

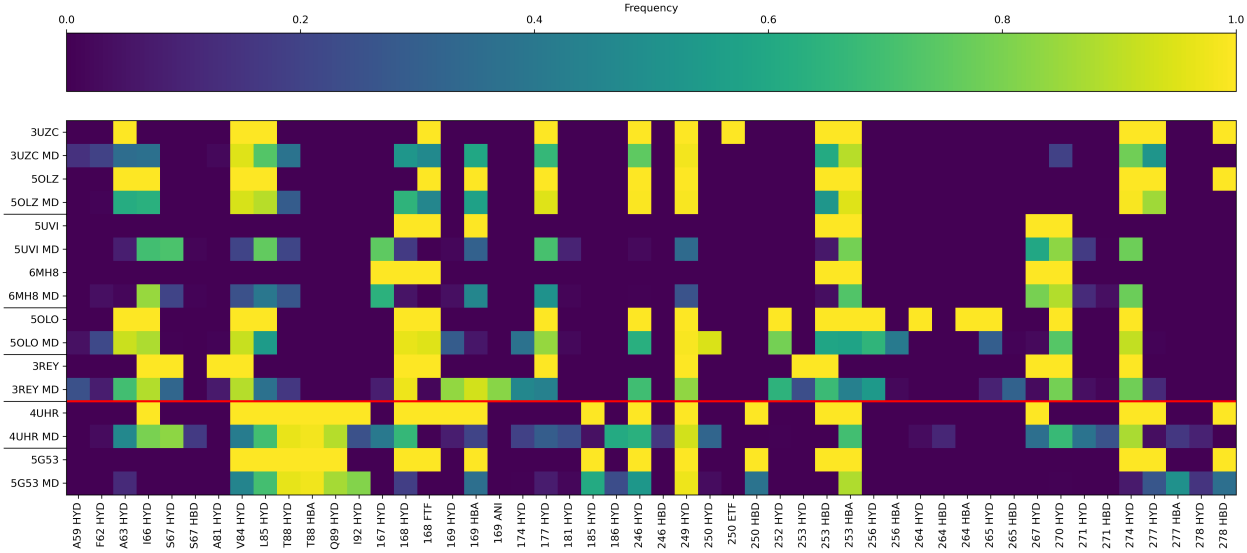
Table B.1.7: Fraction of agonist and fraction of antagonist with a better docking score than the best 1% best ranked decoys, after performing pose filtering using the Hyd+Newhyd IG-OCSVM ensemble models. Reference results are referred to the best rankings obtained by docking on the experimental structures, and rescoring based on interaction similarity to the co-crystallized ligand measured using a graph kernel on Hyd IPAs or Newhyd IPAs.

Receptor	PDB ID	Conformation	Agonist recall QMS2	Antagonist recall QMS2	Agonist recall MAD	Antagonist recall MAD	Agonist ref	Antagonist ref
5HT2A	6A94	Inactive	0.09	0.40	0.09	0.40	0.00	0.10
AA2AR	3REY	Inactive	0.22	0.38	0.33	0.35	0.00	0.04
AA2AR	3UZC	Inactive	0.67	0.58	0.67	0.58	0.11	0.12
AA2AR	5OLO	Inactive	0.33	0.35	0.33	0.35	0.00	0.04
AA2AR	5OLZ	Inactive	0.33	0.31	0.22	0.31	0.11	0.12
AA2AR	5UVI	Inactive	0.44	0.65	0.44	0.54	0.11	0.23
AA2AR	6MH8	Inactive	0.67	0.65	0.78	0.65	0.00	0.12
AA2AR	4UHR	Active	0.22	0.23	0.22	0.23	0.22	0.00
AA2AR	5G53	Active	0.56	0.15	0.56	0.19	0.22	0.00
ACM2	3UON	Inactive	0.00	0.42	0.00	0.32	0.00	0.11
ACM2	5ZK3	Inactive	0.00	0.24	0.00	0.18	0.00	0.13
ACM2	5ZKB	Inactive	0.00	0.00	0.00	0.00	0.00	0.03
ACM2	4MQS	Active	0.40	0.08	0.40	0.11	0.20	0.03
ADRB2	3D4S	Inactive	0.16	0.65	0.25	0.65	0.03	0.50
ADRB2	3NY8	Inactive	0.06	0.20	0.13	0.25	0.06	0.20
ADRB2	3NY9	Inactive	0.09	0.35	0.06	0.30	0.00	0.40
ADRB2	3NYA	Inactive	0.09	0.55	0.22	0.60	0.09	0.50
ADRB2	4GBR	Inactive	0.13	0.55	0.25	0.60	0.00	0.40
ADRB2	5D6L	Inactive	0.13	0.20	0.16	0.30	0.28	0.70
ADRB2	5JQH	Inactive	0.06	0.35	0.06	0.35	0.06	0.65
ADRB2	6PRZ	Inactive	0.41	0.65	0.44	0.60	0.13	0.65
ADRB2	6PS0	Inactive	0.13	0.50	0.06	0.50	0.03	0.45
ADRB2	6PS1	Inactive	0.16	0.75	0.06	0.75	0.03	0.45
ADRB2	6PS2	Inactive	0.03	0.45	0.09	0.50	0.03	0.45
ADRB2	6PS3	Inactive	0.03	0.10	0.00	0.10	0.16	0.20
ADRB2	6PS4	Inactive	0.09	0.45	0.31	0.80	0.03	0.15
ADRB2	6PS6	Inactive	0.25	0.70	0.25	0.75	0.03	0.35
ADRB2	4LDL	Active	0.34	0.05	0.50	0.05	0.38	0.20
CNR1	5U09	Inactive	0.29	0.29	0.43	0.43	0.00	0.00
CNR1	5XRA	Active	0.43	0.14	0.43	0.14	0.14	0.00
DRD2	6LUQ	Inactive	0.16	0.26	0.37	0.60	0.00	0.05

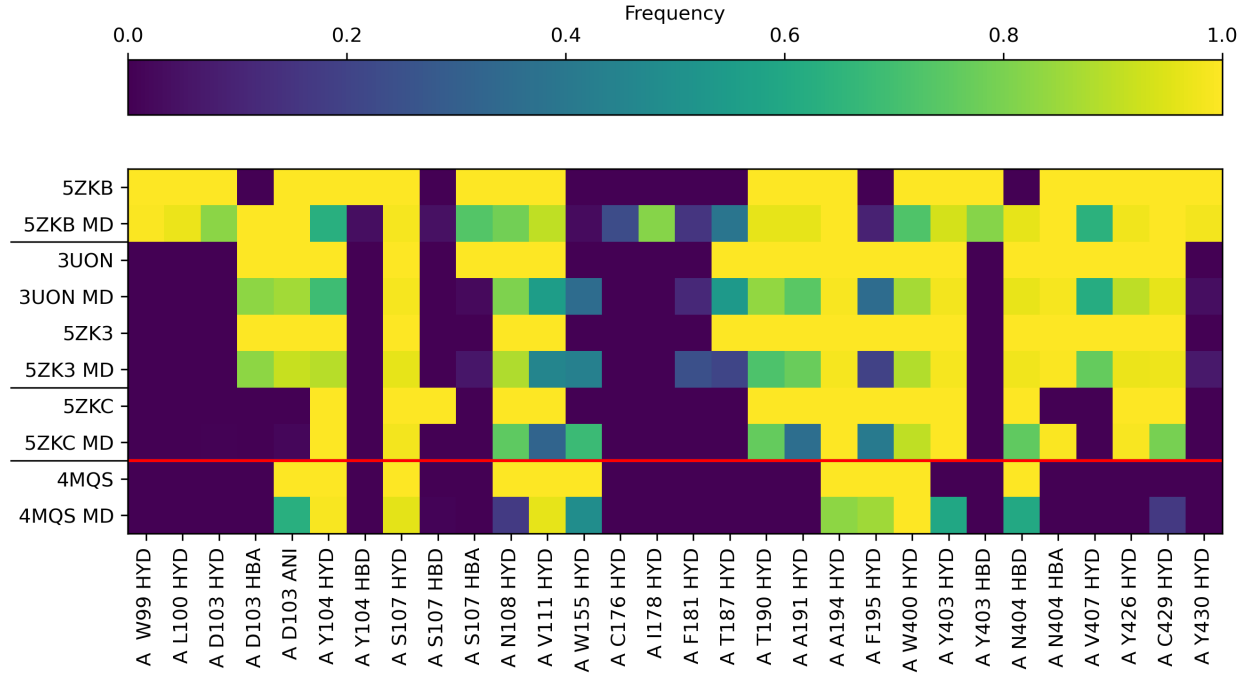
Table B.1.8: Fraction of agonist and fraction of antagonist with a better docking score than the best 1% best ranked decoys, after performing pose filtering using the the ensembles of the three models. Reference results are referred to the best rankings obtained by docking on the experimental structures, and rescoring based on IFP similarity or interaction similarity to the co-crystallized ligand measured using a graph kernel on Hyd IPAs or Newhyd IPAs.

Receptor	PDB ID	Conformation	Agonist recall QMS2	Antagonist recall QMS2	Agonist recall MAD	Antagonist recall MAD	Agonist ref	Antagonist ref
5HT2A	6A94	Inactive	0.09	0.26	0.09	0.29	0.00	0.26
AA2AR	3REY	Inactive	0.11	0.15	0.33	0.27	0.00	0.04
AA2AR	3UZC	Inactive	0.56	0.46	0.67	0.50	0.11	0.15
AA2AR	5OLO	Inactive	0.11	0.19	0.22	0.19	0.00	0.04
AA2AR	5OLZ	Inactive	0.00	0.08	0.00	0.08	0.33	0.23
AA2AR	5UVI	Inactive	0.22	0.12	0.33	0.42	0.11	0.23
AA2AR	6MH8	Inactive	0.00	0.19	0.44	0.50	0.00	0.12
AA2AR	4UHR	Active	0.22	0.04	0.22	0.12	0.44	0.00
AA2AR	5G53	Active	0.00	0.00	0.00	0.00	0.56	0.04
ACM2	3UON	Inactive	0.00	0.18	0.00	0.26	0.00	0.18
ACM2	5ZK3	Inactive	0.00	0.08	0.00	0.08	0.00	0.37
ACM2	5ZKB	Inactive	0.00	0.00	0.00	0.00	0.00	0.08
ACM2	4MQS	Active	0.20	0.03	0.40	0.05	0.20	0.03
ADRB2	3D4S	Inactive	0.03	0.40	0.06	0.55	0.03	0.50
ADRB2	3NY8	Inactive	0.00	0.10	0.03	0.15	0.06	0.20
ADRB2	3NY9	Inactive	0.00	0.05	0.00	0.05	0.03	0.40
ADRB2	3NYA	Inactive	0.00	0.30	0.00	0.55	0.09	0.50
ADRB2	4GBR	Inactive	0.00	0.05	0.00	0.25	0.00	0.40
ADRB2	5D6L	Inactive	0.03	0.10	0.03	0.15	0.28	0.70
ADRB2	5JQH	Inactive	0.03	0.20	0.03	0.20	0.06	0.65
ADRB2	6PRZ	Inactive	0.03	0.30	0.06	0.50	0.13	0.65
ADRB2	6PS0	Inactive	0.03	0.00	0.03	0.05	0.03	0.45
ADRB2	6PS1	Inactive	0.00	0.25	0.03	0.40	0.03	0.45
ADRB2	6PS2	Inactive	0.00	0.10	0.00	0.25	0.03	0.45
ADRB2	6PS3	Inactive	0.00	0.05	0.00	0.10	0.16	0.30
ADRB2	6PS4	Inactive	0.03	0.30	0.19	0.65	0.03	0.20
ADRB2	6PS6	Inactive	0.00	0.15	0.03	0.20	0.03	0.35
ADRB2	4LDL	Active	0.09	0.00	0.34	0.00	0.38	0.20
CNR1	5U09	Inactive	0.00	0.14	0.00	0.14	0.00	0.00
CNR1	5XRA	Active	0.29	0.00	0.29	0.00	0.14	0.00
DRD2	6LUQ	Inactive	0.05	0.16	0.21	0.36	0.00	0.09

Figure B.1.1: Average IFPs extracted from a MD simulation, and crystallographic IFPs of AA2AR, and ACM2



(A) AA2AR



(B) ACM2

Figure B.1.2: Average IFPs extracted from a MD simulation, and crystallographic IFPs of ADRB2

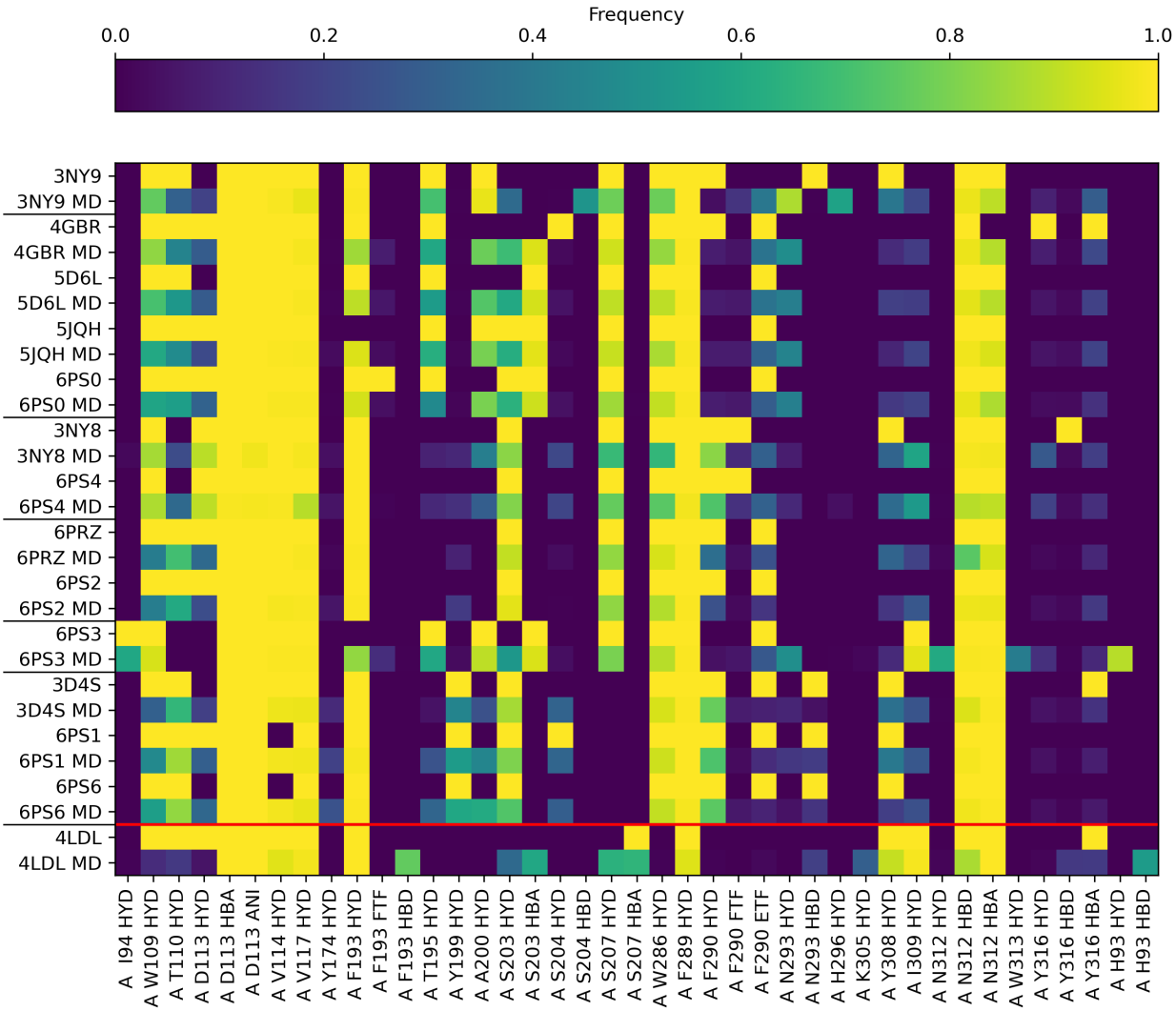
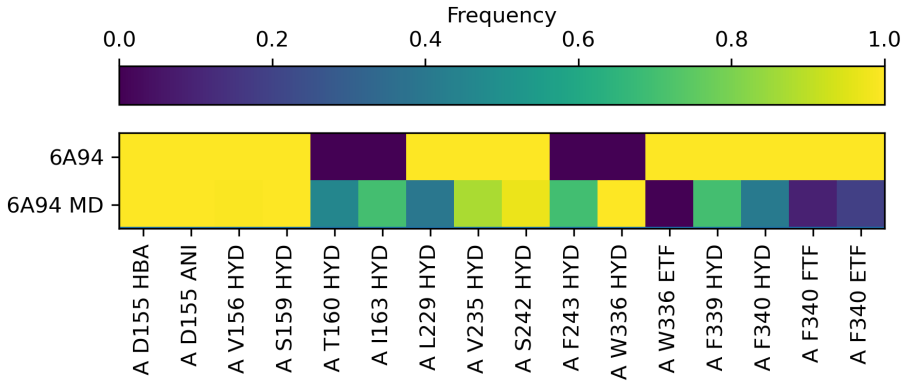
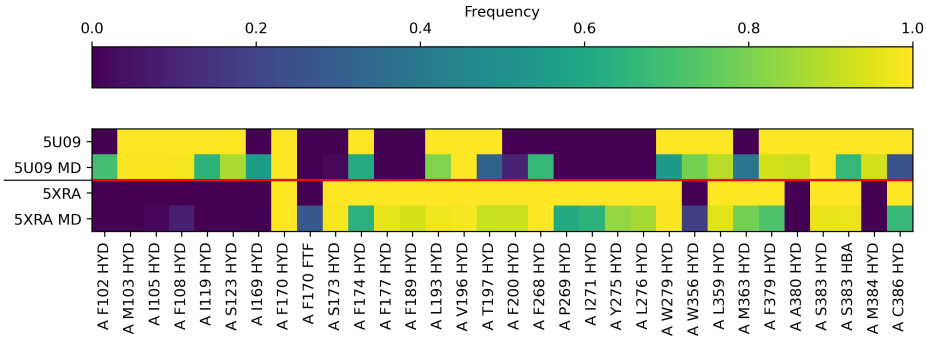


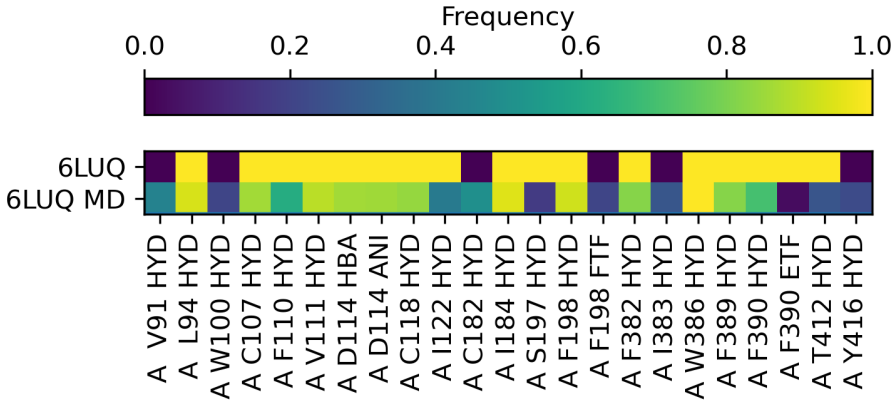
Figure B.1.3: Average IFPs extracted from a MD simulation, and crystallographic IFPs of 5HT2A, CNR1, and DRD2



(A) 5HT2A

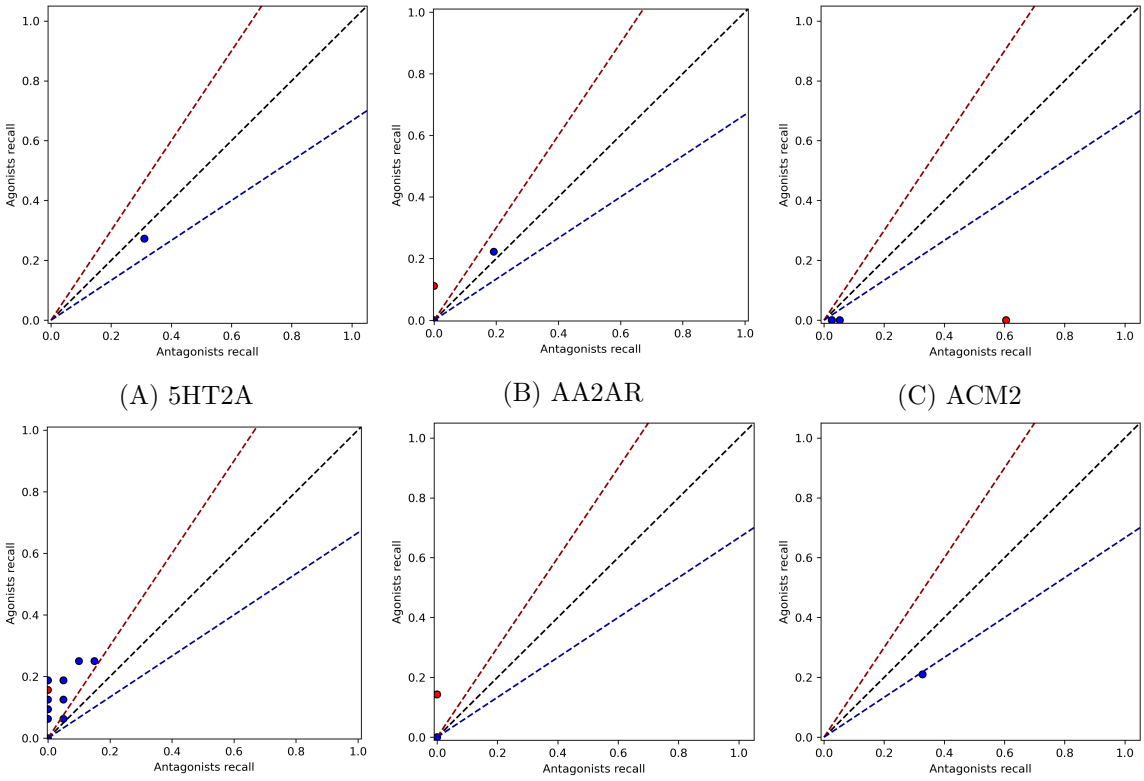


(B) CNR1



(C) DRD2

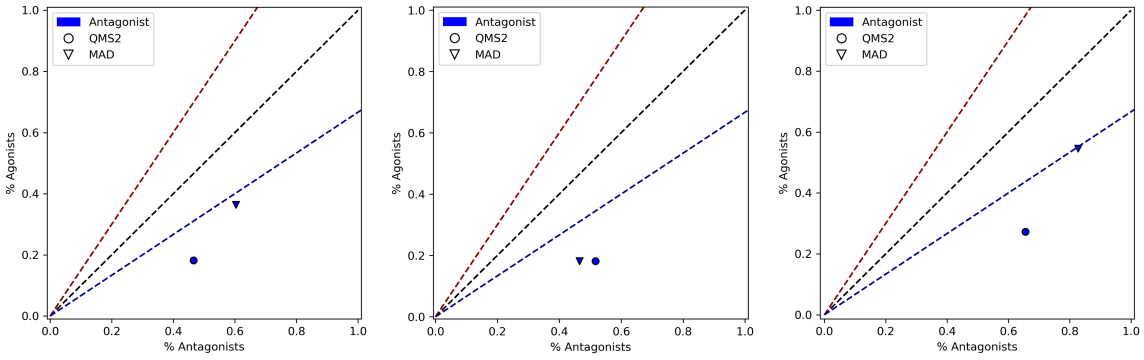
Figure B.1.4: Fraction of recovered agonists and fraction of recovered antagonists with a better docking score than the 1% of decoys with the best docking scores.



(A) 5HT2A (B) AA2AR (C) ACM2

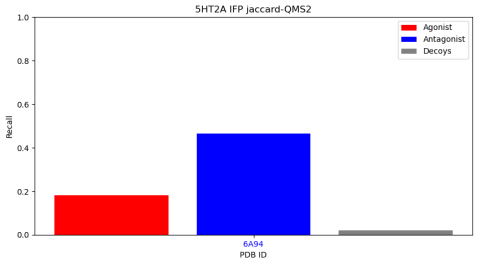
(D) ADRB2 (E) CNR1 (F) DRD2

Figure B.1.5: Fraction of agonist and fraction of antagonists of 5HT2A selected by InteractionOCSVM models.

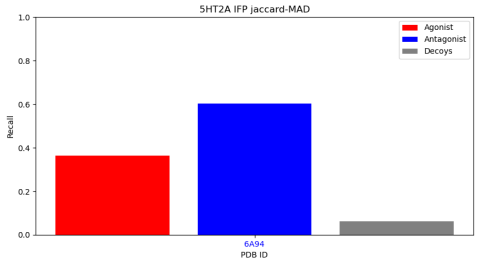


(A) IFP-OCSVM (B) Hyd IG-OCSVM (C) Newhyd IG-OCSVM

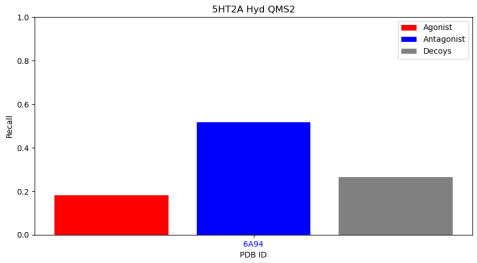
Figure B.1.6: Fraction of agonists, fraction of antagonists, and fraction of decoys selected by the corresponding 5HT2A InteractionOCSVM models. The color of the PDB ID label indicates the efficacy of the reference ligand, red for agonist, and blue for antagonist.



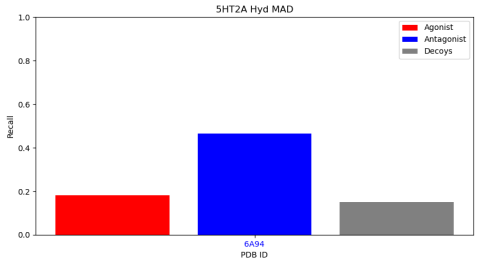
(A) IFP-OCSVM QMS2



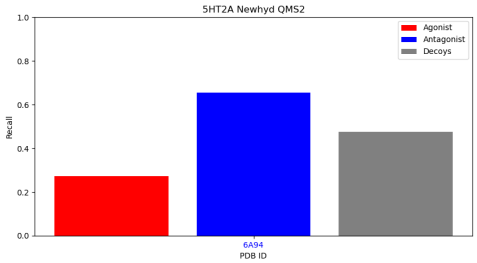
(B) IFP-OCSVM MAD



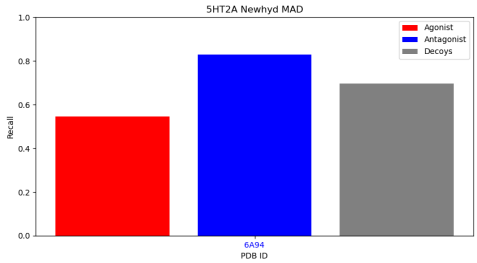
(C) Hyd IG-OCSVM QMS2



(D) Hyd IG-OCSVM MAD

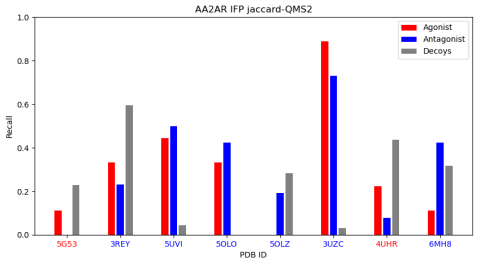


(E) Newhyd IG-OCSVM QMS2

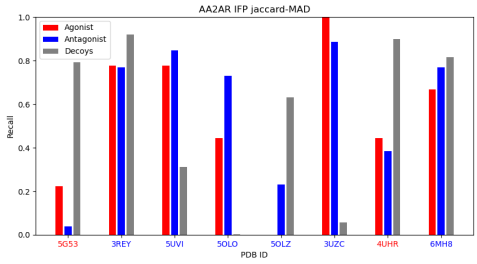


(F) Newhyd IG-OCSVM MAD

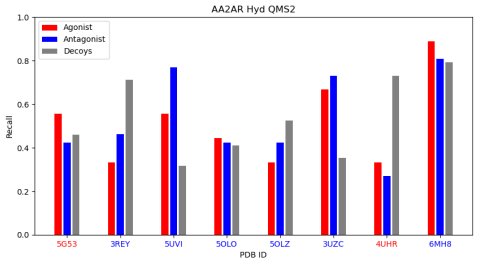
Figure B.1.7: Fraction of agonists, fraction of antagonists, and fraction of decoys selected by the corresponding AA2AR InteractionOCSVM models. The color of the PDB ID label indicates the efficacy of the reference ligand, red for agonist, and blue for antagonist.



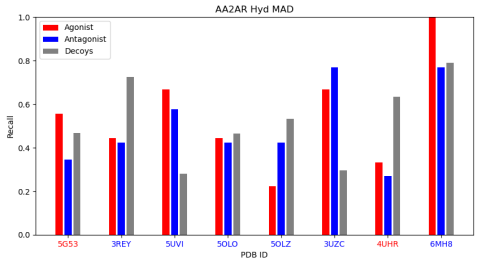
(A) IFP-OCSVM QMS2



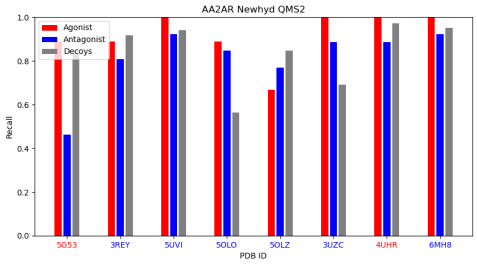
(B) IFP-OCSVM MAD



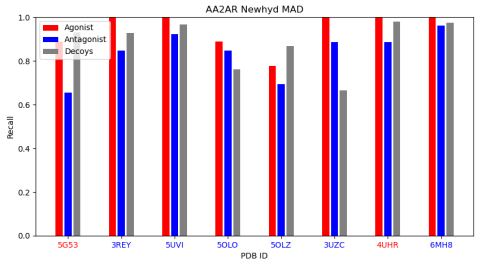
(C) Hyd IG-OCSVM QMS2



(D) Hyd IG-OCSVM MAD

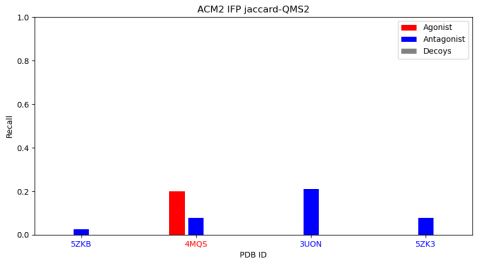


(E) Newhyd IG-OCSVM QMS2

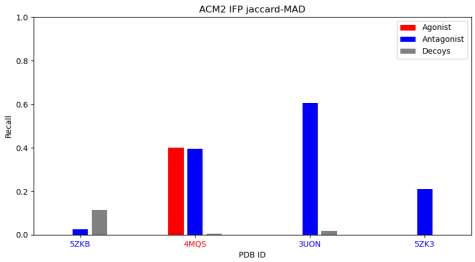


(F) Newhyd IG-OCSVM MAD

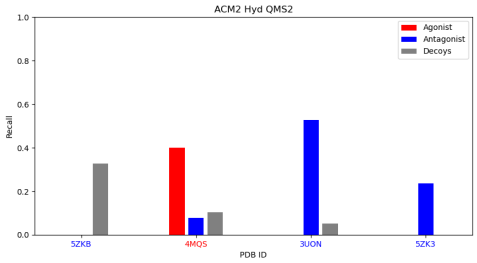
Figure B.1.8: Fraction of agonists, fraction of antagonists, and fraction of decoys selected by the corresponding ACM2 InteractionOCSVM models. The color of the PDB ID label indicates the efficacy of the reference ligand, red for agonist, and blue for antagonist.



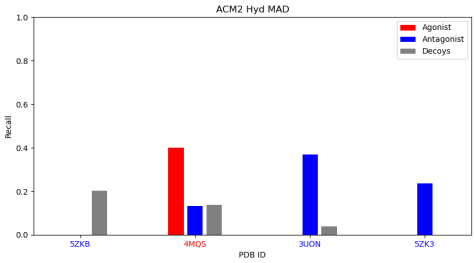
(A) IFP-OCSVM QMS2



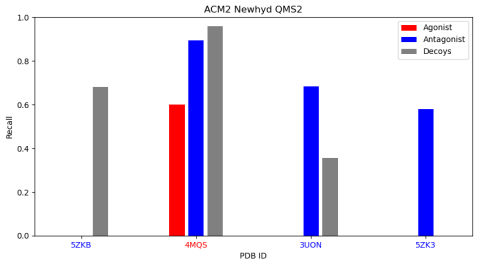
(B) IFP-OCSVM MAD



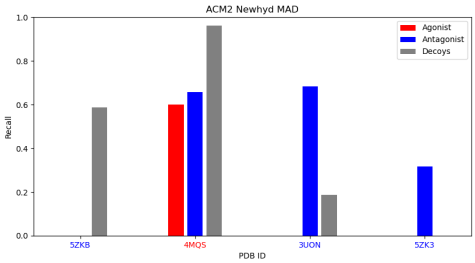
(C) Hyd IG-OCSVM QMS2



(D) Hyd IG-OCSVM MAD

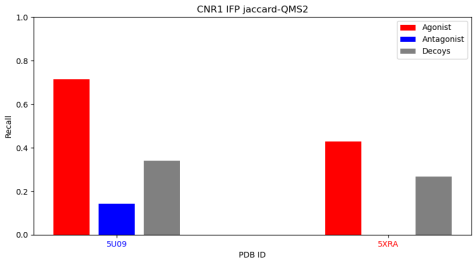


(E) Newhyd IG-OCSVM QMS2

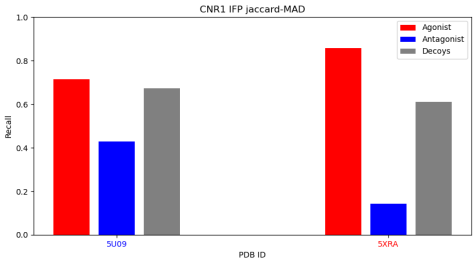


(F) Newhyd IG-OCSVM MAD

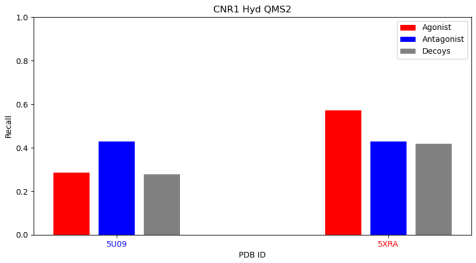
Figure B.1.9: Fraction of agonists, fraction of antagonists, and fraction of decoys selected by the corresponding CNR1 InteractionOCSVM models. The color of the PDB ID label indicates the efficacy of the reference ligand, red for agonist, and blue for antagonist.



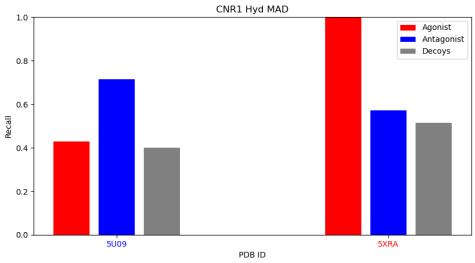
(A) IFP-OCSVM QMS2



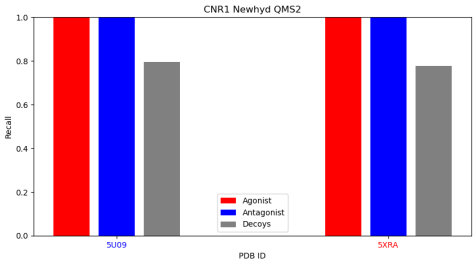
(B) IFP-OCSVM MAD



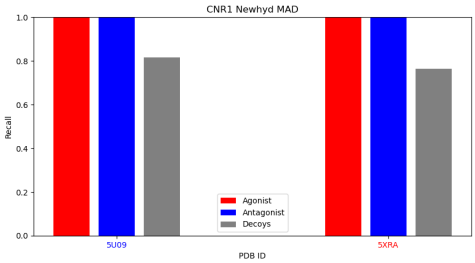
(C) Hyd IG-OCSVM QMS2



(D) Hyd IG-OCSVM MAD

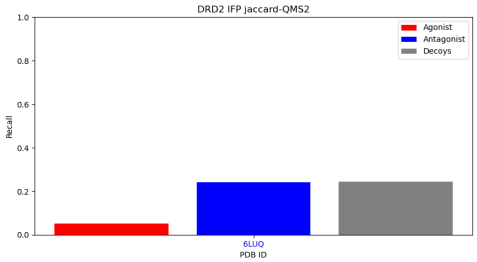


(E) Newhyd IG-OCSVM QMS2

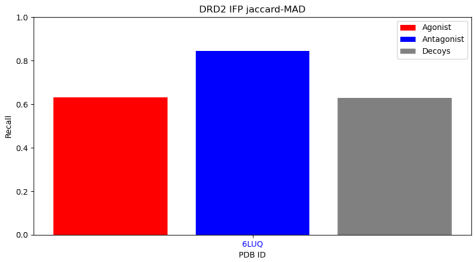


(F) Newhyd IG-OCSVM MAD

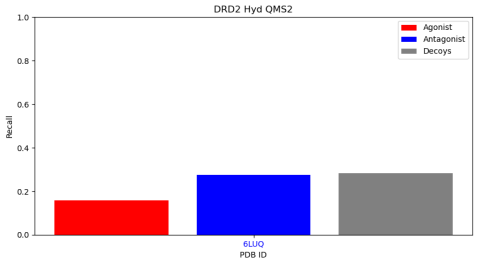
Figure B.1.10: Fraction of agonists, fraction of antagonists, and fraction of decoys selected by the corresponding DRD2 InteractionOCSVM models. The color of the PDB ID label indicates the efficacy of the reference ligand, red for agonist, and blue for antagonist.



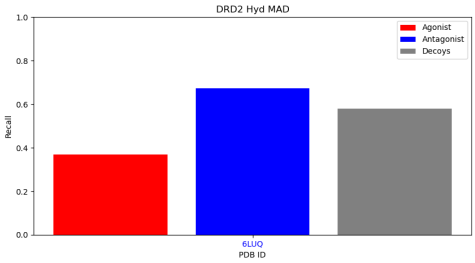
(A) IFP-OCSVM QMS2



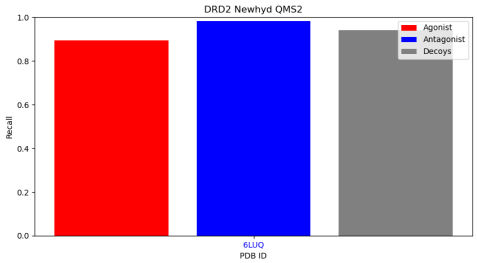
(B) IFP-OCSVM MAD



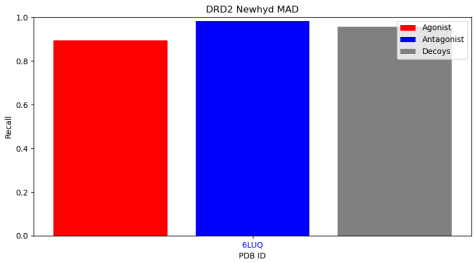
(C) Hyd IG-OCSVM QMS2



(D) Hyd IG-OCSVM MAD

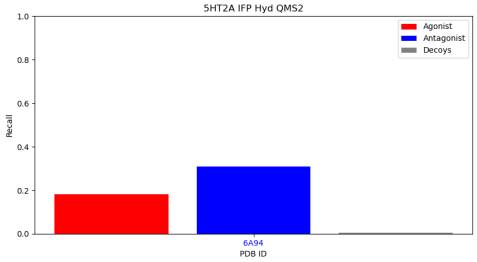


(E) Newhyd IG-OCSVM QMS2

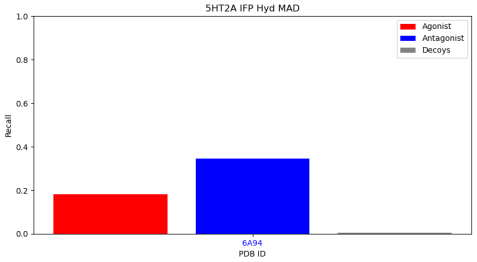


(F) Newhyd IG-OCSVM MAD

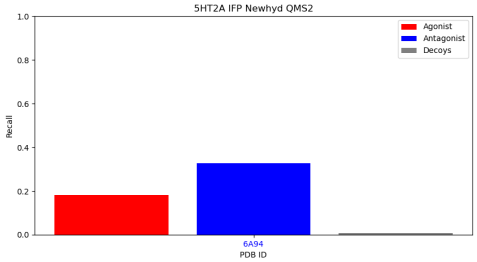
Figure B.1.11: Fraction of agonists, fraction of antagonists, and fraction of decoys selected by the corresponding 5HT2A ensemble InteractionOCSVM models. The color of the PDB ID label indicates the efficacy of the reference ligand, red for agonist, and blue for antagonist.



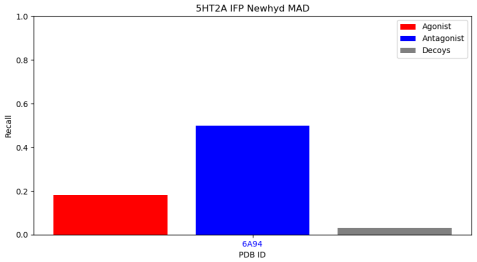
(A) IFP+Hyd IG-OCSVM QMS2



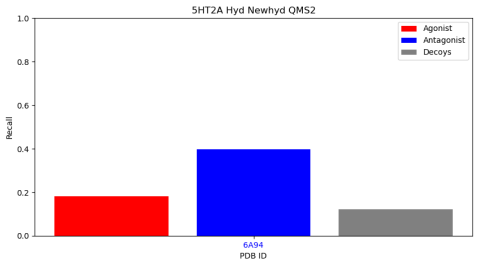
(B) IFP+Hyd IG-OCSVM MAD



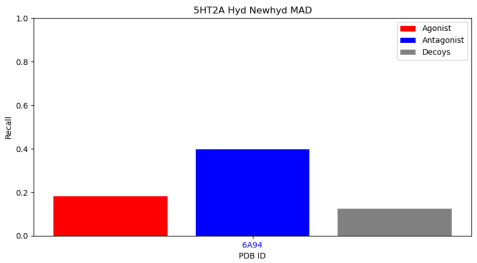
(C) IFP+Newhyd IG-OCSVM QMS2



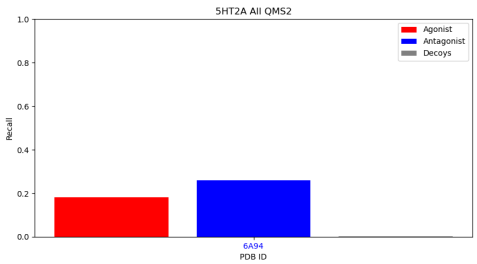
(D) IFP+Newhyd IG-OCSVM MAD



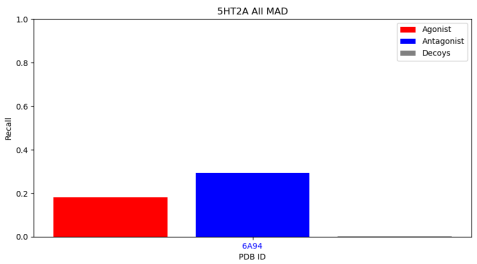
(E) Hyd+Newhyd IG-OCSVM QMS2



(F) Hyd+Newhyd IG-OCSVM MAD

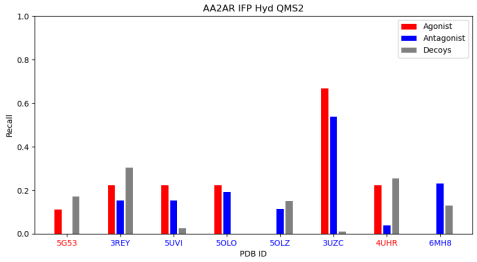


(G) All InteractionOCSVM QMS2

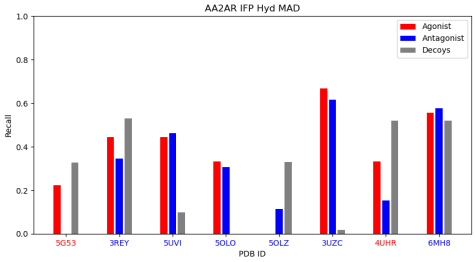


(H) All InteractionOCSVM MAD

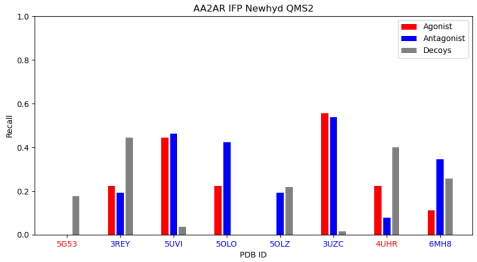
Figure B.1.12: Fraction of agonists, fraction of antagonists, and fraction of decoys selected by the corresponding AA2AR ensemble InteractionOCSVM models. The color of the PDB ID label indicates the efficacy of the reference ligand, red for agonist, and blue for antagonist.



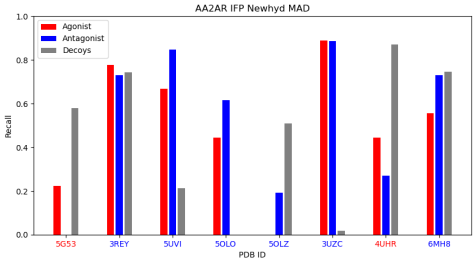
(A) IFP-OCSVM+Hyd IG-OCSVM QMS2



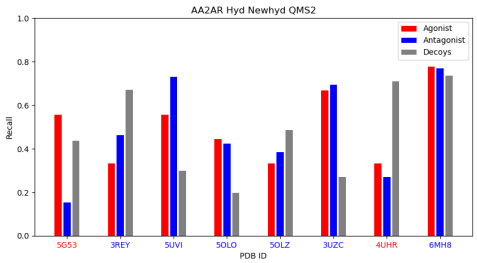
(B) IFP+Hyd IG-OCSVM MAD



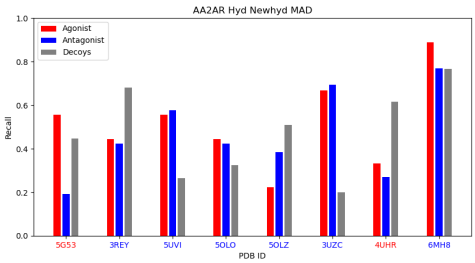
(C) IFP+Newhyd IG-OCSVM QMS2



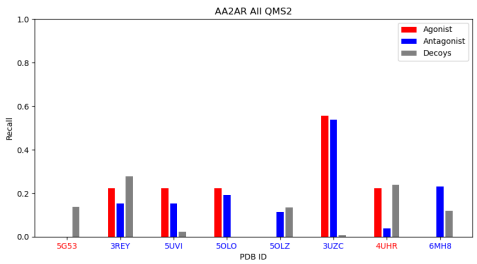
(D) IFP+Newhyd IG-OCSVM MAD



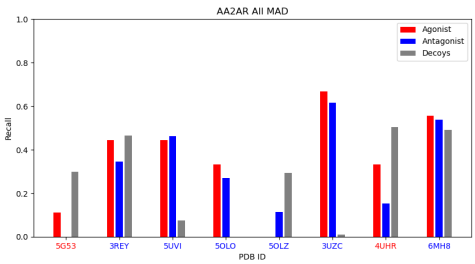
(E) Hyd+Newhyd IG-OCSVM QMS2



(F) Hyd+Newhyd IG-OCSVM MAD

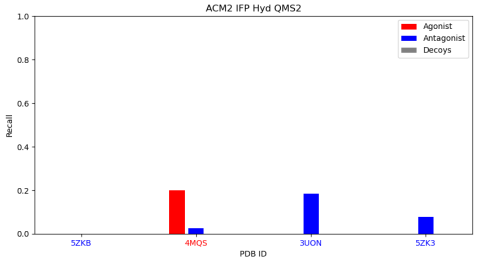


(G) All InteractionOCSVM QMS2

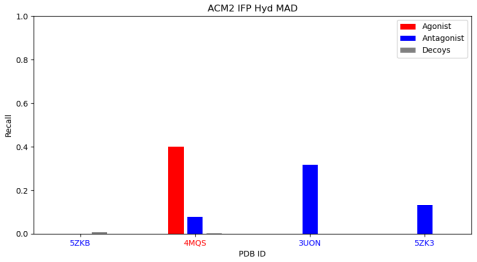


(H) All InteractionOCSVM MAD

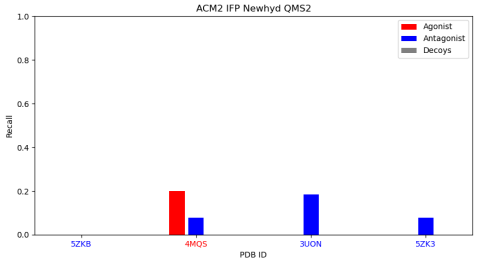
Figure B.1.13: Fraction of agonists, fraction of antagonists, and fraction of decoys selected by the corresponding ACM2 ensemble InteractionOCSVM models. The color of the PDB ID label indicates the efficacy of the reference ligand, red for agonist, and blue for antagonist.



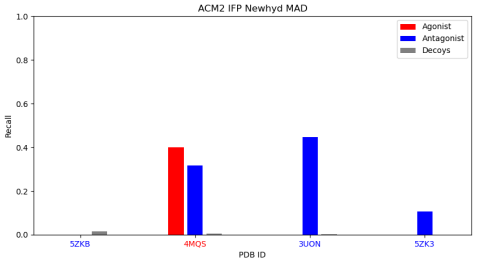
(A) IFP+Hyd IG-OCSVM QMS2



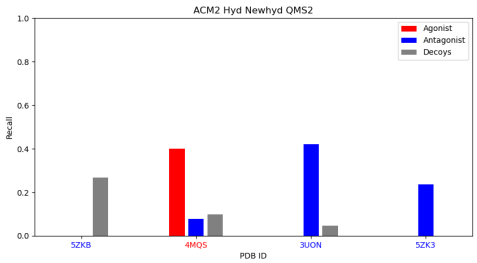
(B) IFP+Hyd IG-OCSVM MAD



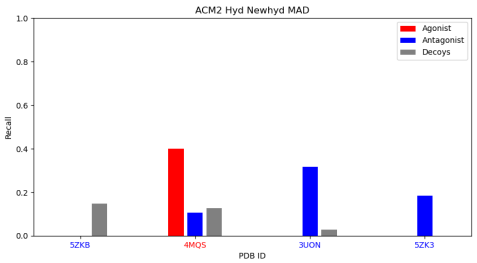
(C) IFP+Newhyd IG-OCSVM QMS2



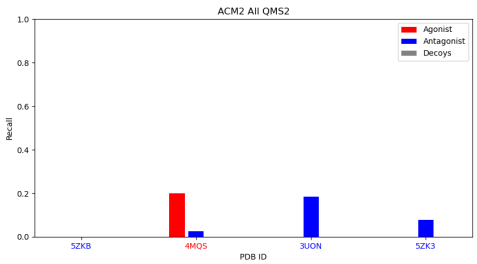
(D) IFP+Newhyd IG-OCSVM MAD



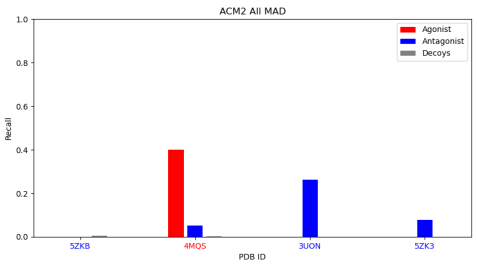
(E) Hyd+Newhyd IG-OCSVM QMS2



(F) Hyd+Newhyd IG-OCSVM MAD

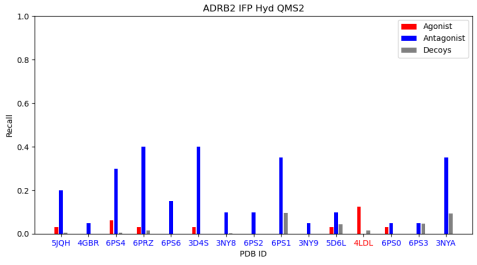


(G) All InteractionOCSVM QMS2

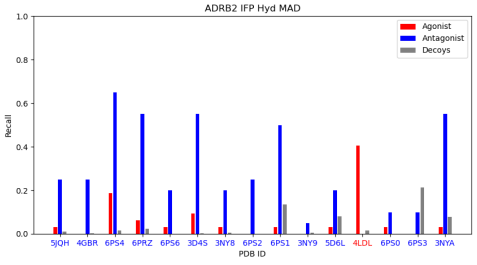


(H) All InteractionOCSVM MAD

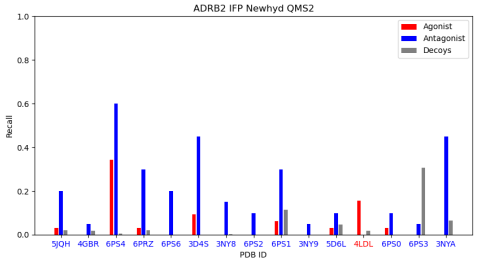
Figure B.1.14: Fraction of agonists, fraction of antagonists, and fraction of decoys selected by the corresponding ADRB2 ensemble InteractionOCSVM models. The color of the PDB ID label indicates the efficacy of the reference ligand, red for agonist, and blue for antagonist.



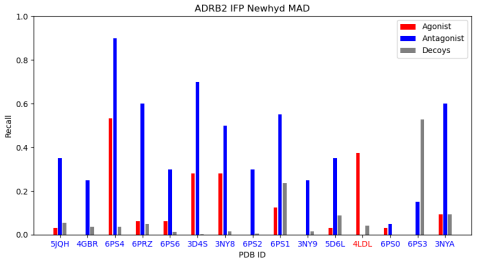
(A) IFP+Hyd IG-OCSVM QMS2



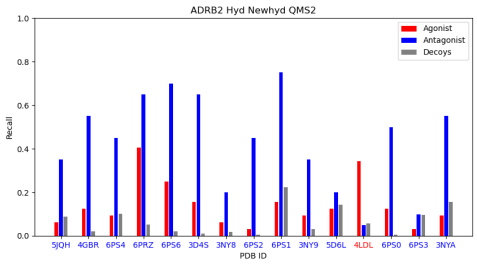
(B) IFP+Hyd IG-OCSVM MAD



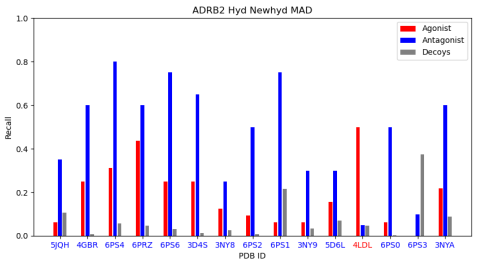
(C) IFP+Newhyd IG-OCSVM QMS2



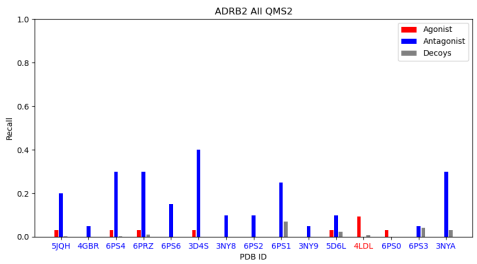
(D) IFP+Newhyd IG-OCSVM MAD



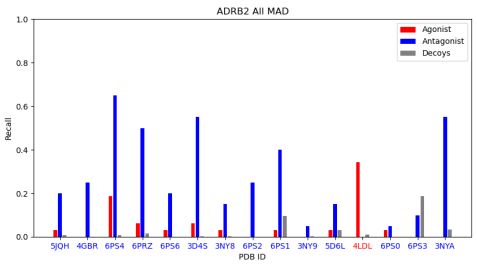
(E) Hyd+Newhyd IG-OCSVM QMS2



(F) Hyd+Newhyd IG-OCSVM MAD

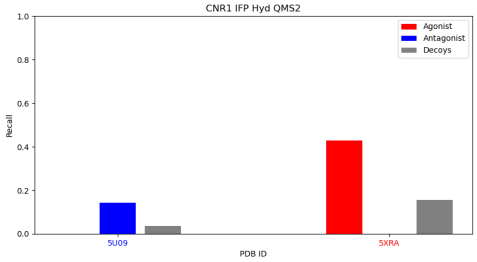


(G) All InteractionOCSVM QMS2

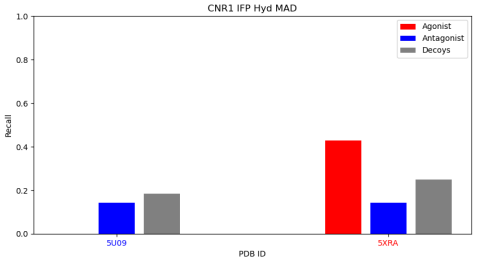


(H) All InteractionOCSVM MAD

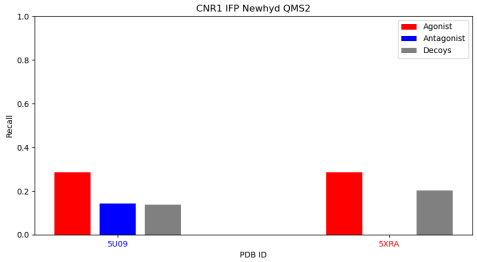
Figure B.1.15: Fraction of agonists, fraction of antagonists, and fraction of decoys selected by the corresponding CNR1 ensemble InteractionOCSVM models. The color of the PDB ID label indicates the efficacy of the reference ligand, red for agonist, and blue for antagonist.



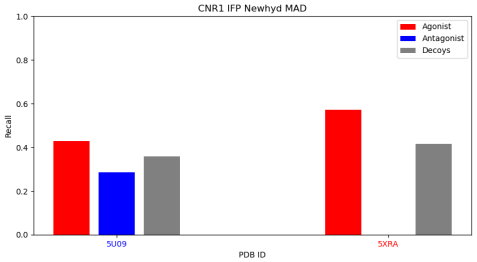
(A) IFP+Hyd IG-OCSVM QMS2



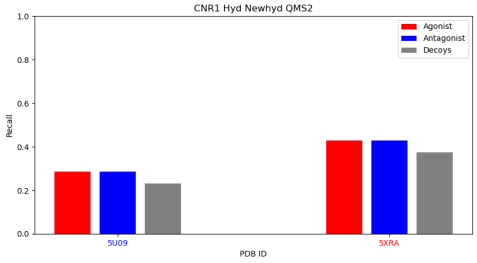
(B) IFP+Hyd IG-OCSVM MAD



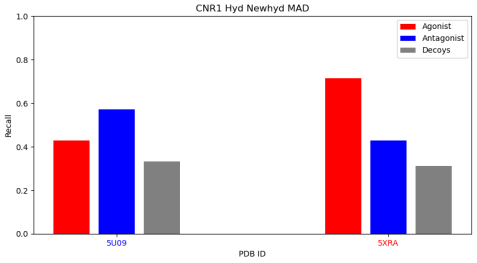
(C) IFP+Newhyd IG-OCSVM QMS2



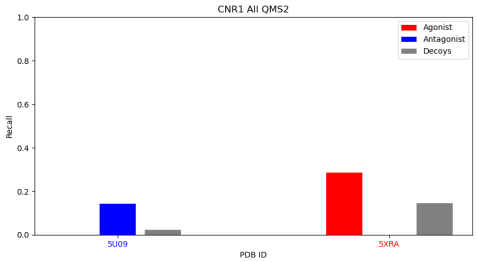
(D) IFP+Newhyd IG-OCSVM MAD



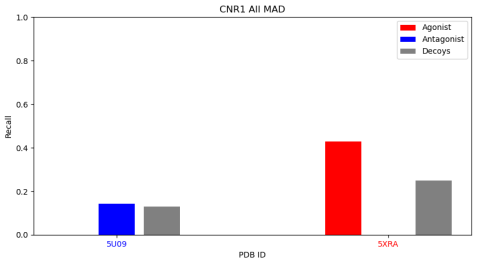
(E) Hyd+Newhyd IG-OCSVM QMS2



(F) Hyd+Newhyd IG-OCSVM MAD

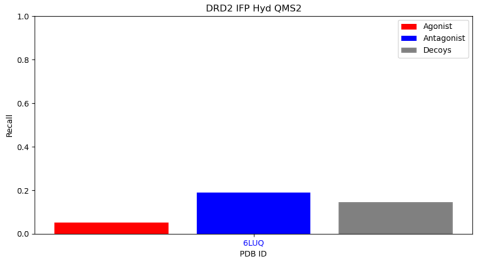


(G) All InteractionOCSVM QMS2

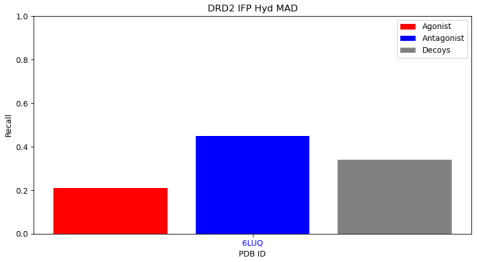


(H) All InteractionOCSVM MAD

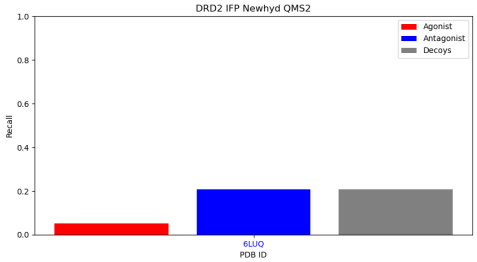
Figure B.1.16: Fraction of agonists, fraction of antagonists, and fraction of decoys selected by the corresponding DRD2 ensemble InteractionOCSVM models. The color of the PDB ID label indicates the efficacy of the reference ligand, red for agonist, and blue for antagonist.



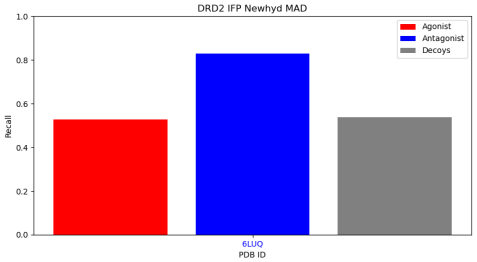
(A) IFP+Hyd IG-OCSVM QMS2



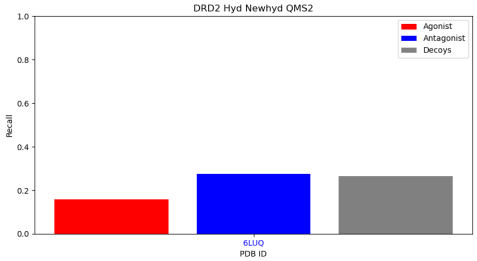
(B) IFP+Hyd IG-OCSVM MAD



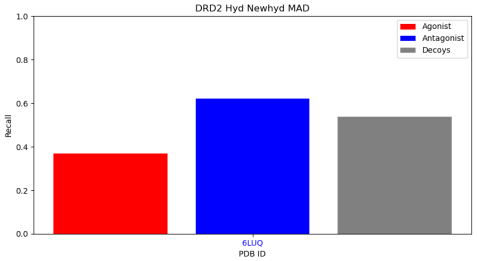
(C) IFP+Newhyd IG-OCSVM QMS2



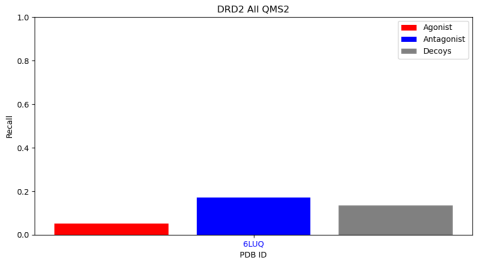
(D) IFP+Newhyd IG-OCSVM MAD



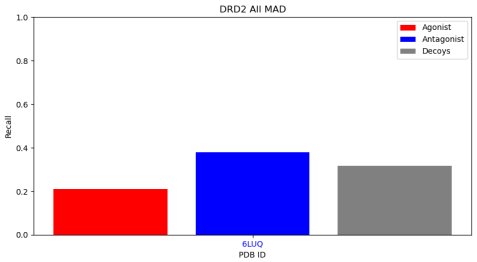
(E) Hyd+Newhyd IG-OCSVM QMS2



(F) Hyd+Newhyd IG-OCSVM MAD

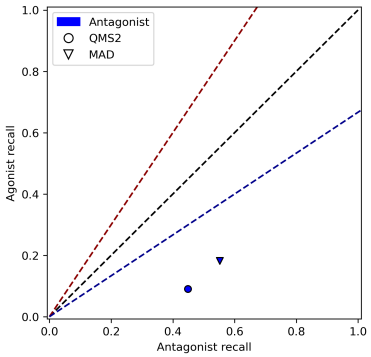


(G) All InteractionOCSVM QMS2

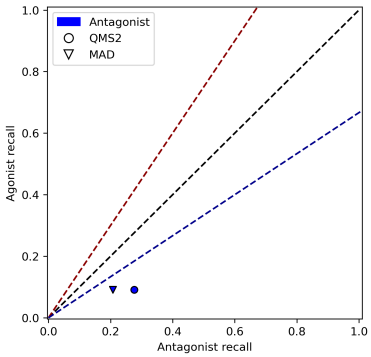


(H) All InteractionOCSVM MAD

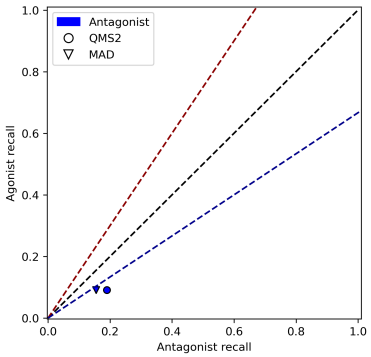
Figure B.1.17: Fraction of agonist and fraction of antagonists of 5HT2A selected by InteractionOCSVM models, and presenting a better docking score than the 1% of the best decoys after pose filtering using InteractionOCSVM.



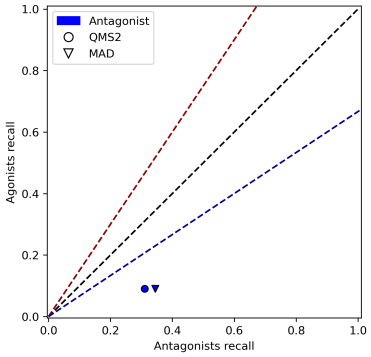
(A) IFP-OCSVM



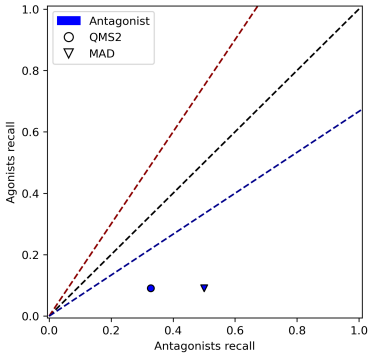
(B) Hyd IG-OCSVM



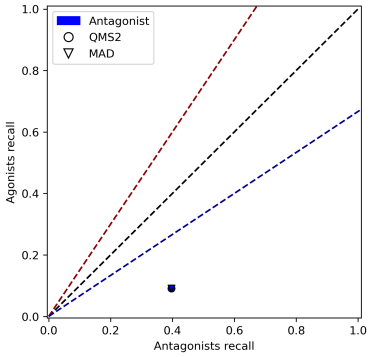
(C) Newhyd IG-OCSVM



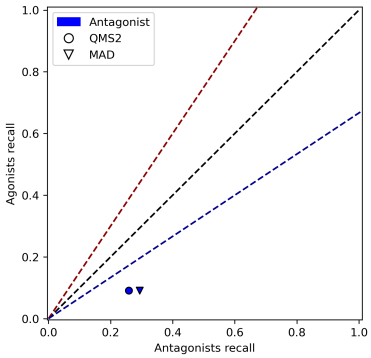
(D) IFP+Hyd IG-OCSVM



(E) IFP+Newhyd IG-OCSVM

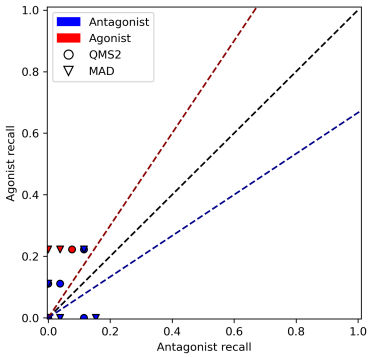


(F) Hyd+Newhyd IG-OCSVM

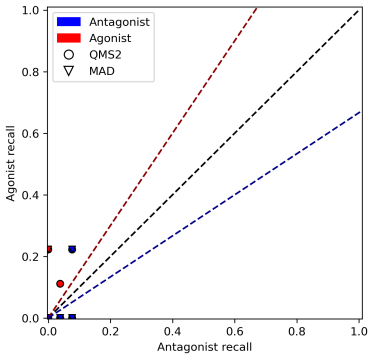


(G) All - OCSVM

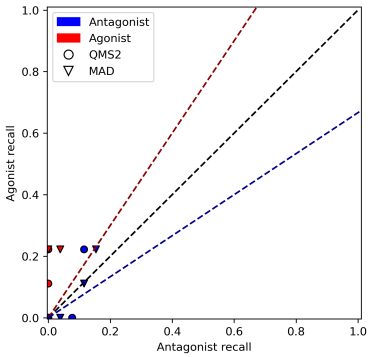
Figure B.1.18: Fraction of agonist and fraction of antagonists of AA2AR selected by InteractionOCSVM models, and presenting a better docking score than the 1% of the best decoys after pose filtering using InteractionOCSVM.



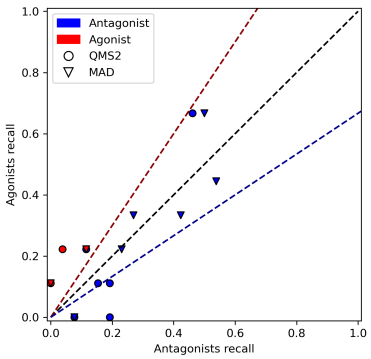
(A) IFP-OCSVM



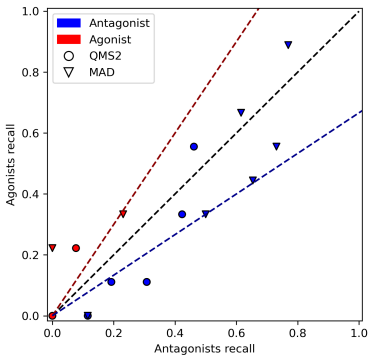
(B) Hyd IG-OCSVM



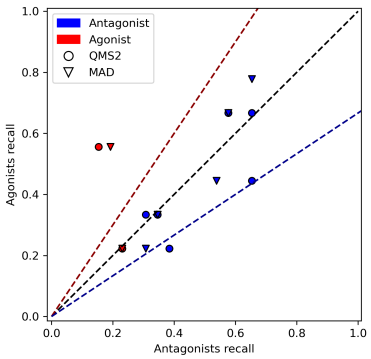
(C) Newhyd IG-OCSVM



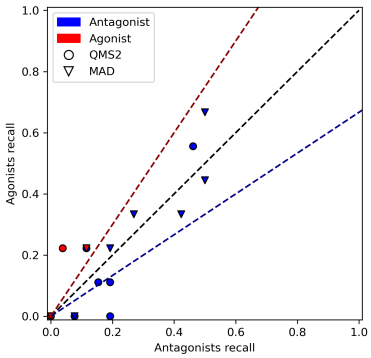
(D) IFP+Hyd IG-OCSVM



(E) IFP+Newhyd IG-OCSVM

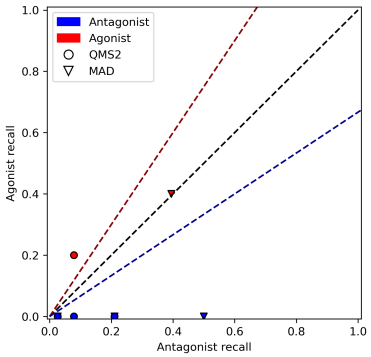


(F) Hyd+Newhyd IG-OCSVM

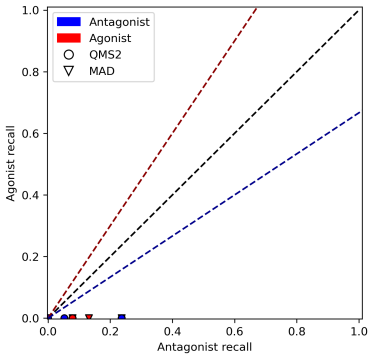


(G) All - OCSVM

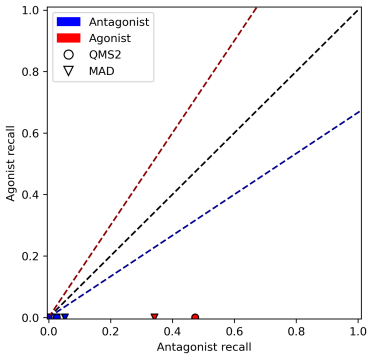
Figure B.1.19: Fraction of agonist and fraction of antagonists of ACM2 selected by InteractionOCSVM models, and presenting a better docking score than the 1% of the best decoys after pose filtering using InteractionOCSVM.



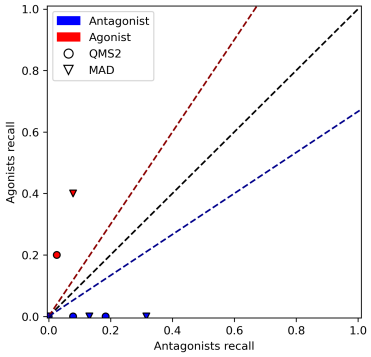
(A) IFP-OCSVM



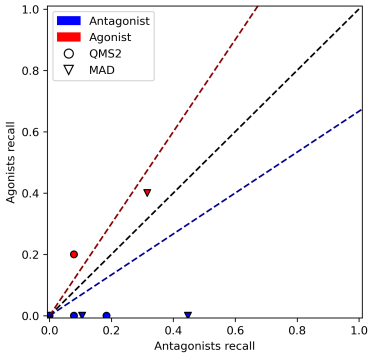
(B) Hyd IG-OCSVM



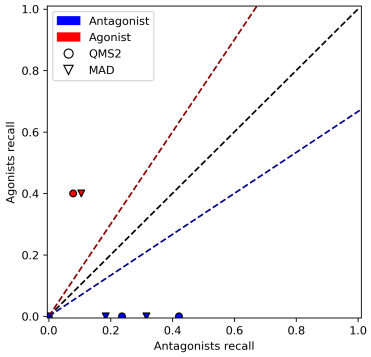
(C) Newhyd IG-OCSVM



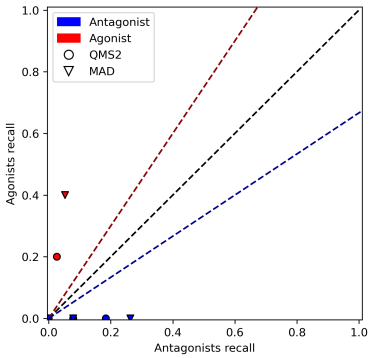
(D) IFP+Hyd IG-OCSVM



(E) IFP+Newhyd IG-OCSVM

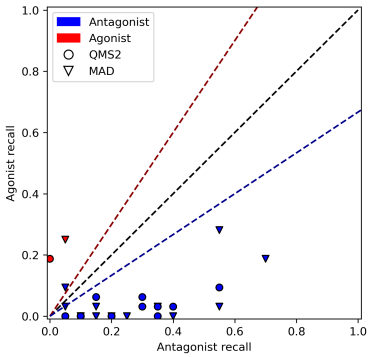


(F) Hyd+Newhyd IG-OCSVM

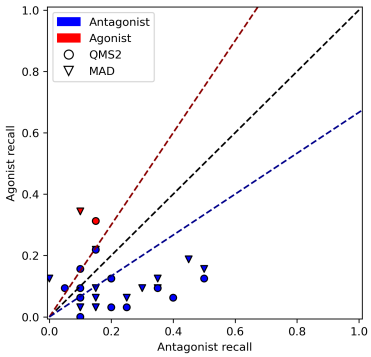


(G) All - OCSVM

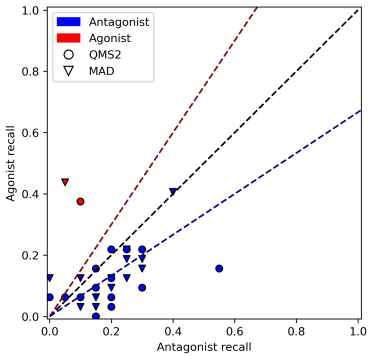
Figure B.1.20: Fraction of agonist and fraction of antagonists of ADRB2 selected by InteractionOCSVM models.



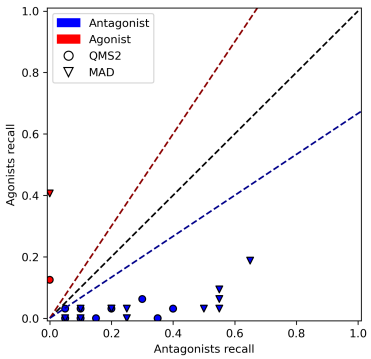
(A) IFP-OCSVM



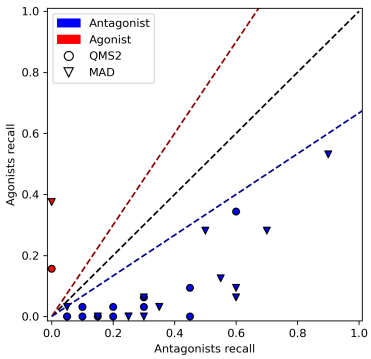
(B) Hyd IG-OCSVM



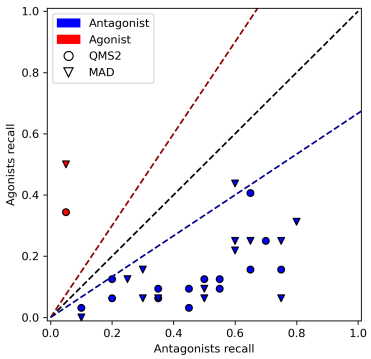
(C) Newhyd IG-OCSVM



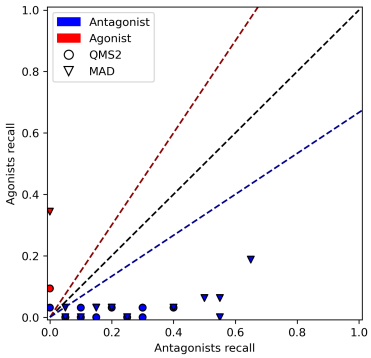
(D) IFP+Hyd IG-OCSVM



(E) IFP+Newhyd IG-OCSVM

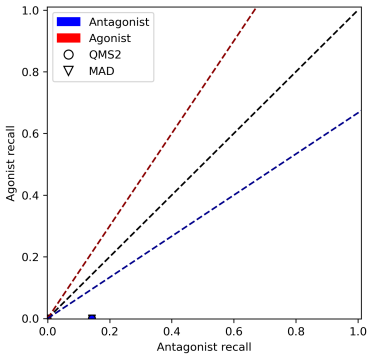


(F) Hyd+Newhyd IG-OCSVM

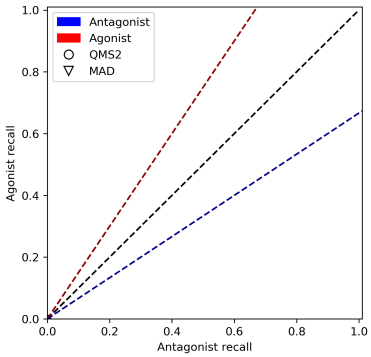


(G) All - OCSVM

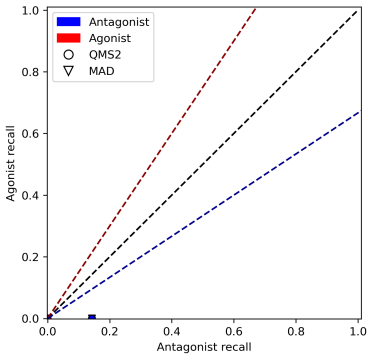
Figure B.1.21: Fraction of agonist and fraction of antagonists of CNR1 selected by InteractionOCSVM models.



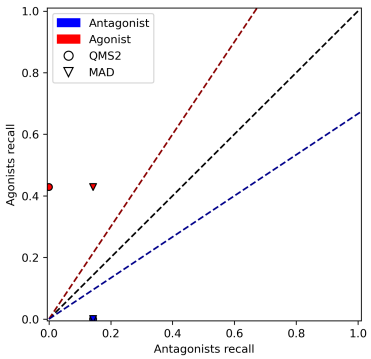
(A) IFP-OCSVM



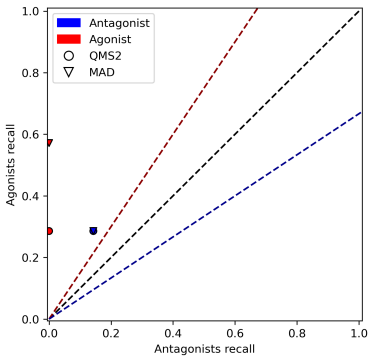
(B) Hyd IG-OCSVM



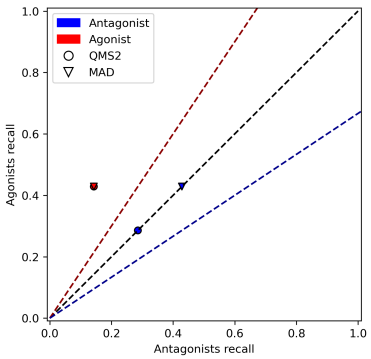
(C) Newhyd IG-OCSVM



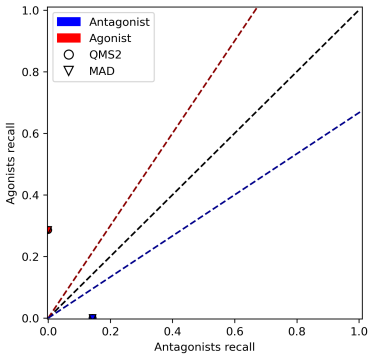
(D) IFP+Hyd IG-OCSVM



(E) IFP+Newhyd IG-OCSVM

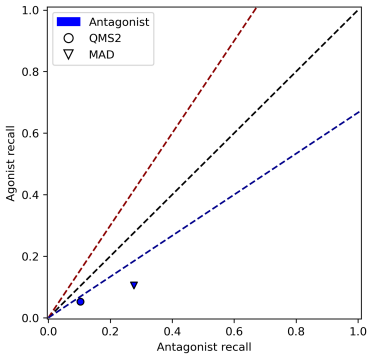


(F) Hyd+Newhyd IG-OCSVM

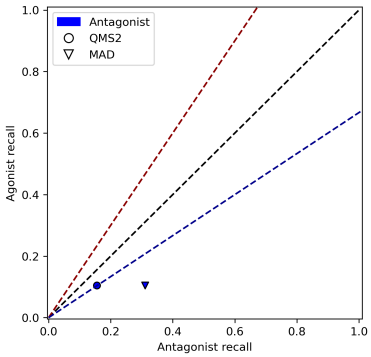


(G) All - OCSVM

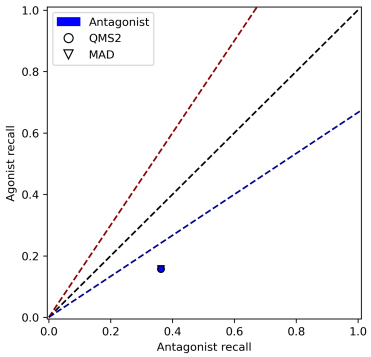
Figure B.1.22: Fraction of agonist and fraction of antagonists of DRD2 selected by InteractionOCSVM models, and presenting a better docking score than the 1% of the best decoys after pose filtering using InteractionOCSVM.



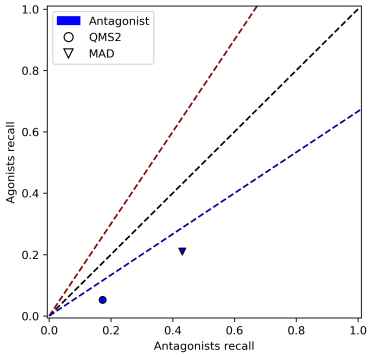
(A) IFP-OCSVM



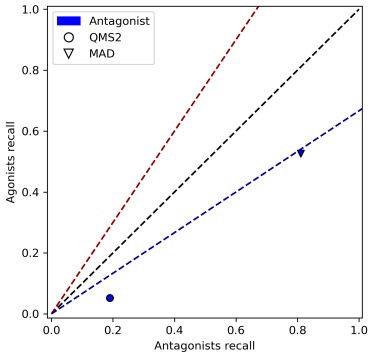
(B) Hyd IG-OCSVM



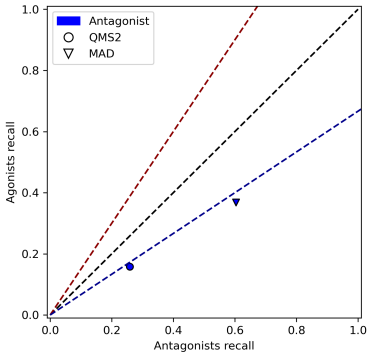
(C) Newhyd IG-OCSVM



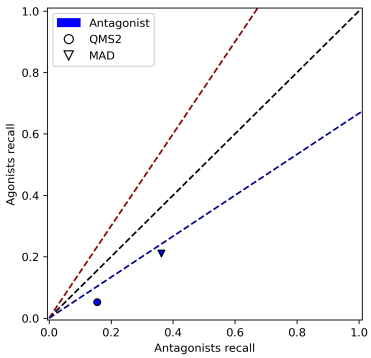
(D) IFP+Hyd IG-OCSVM



(E) IFP+Newhyd IG-OCSVM



(F) Hyd+Newhyd IG-OCSVM



(G) All - OCSVM

Appendix C

Chapter 4 supporting information

C.1 Section 4.2 supporting information

Figure C.1.1: RMSD of IT1t after structural alignment on the rigid domains of the receptor.

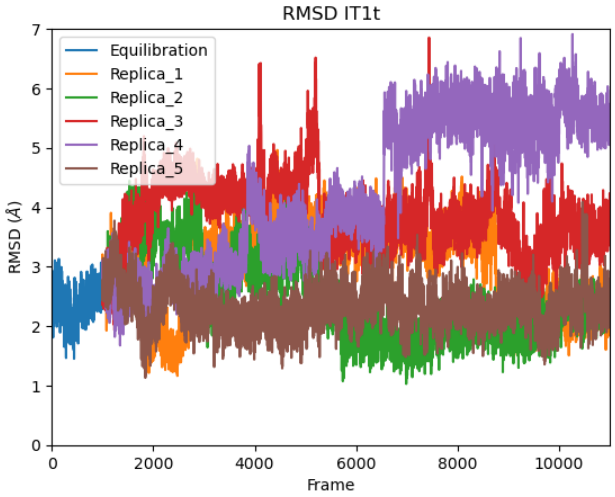


Figure C.1.2: Average interaction fingerprint describing the frequency of interaction between IT1t and CXCR4

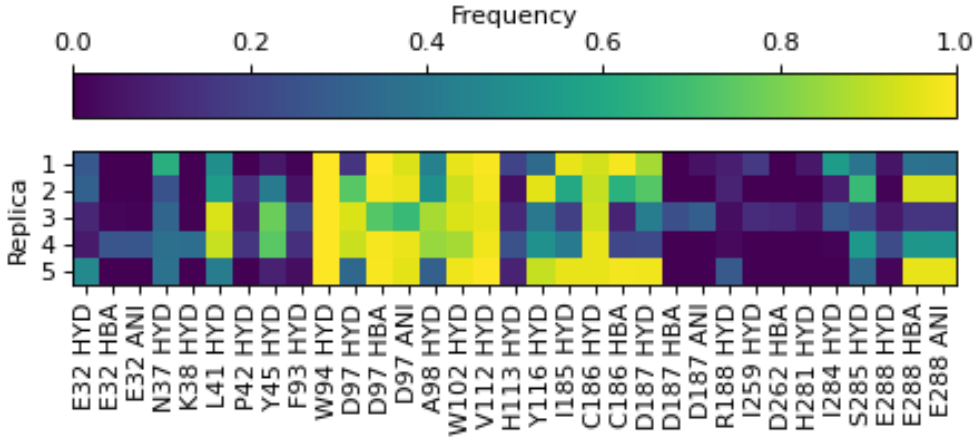
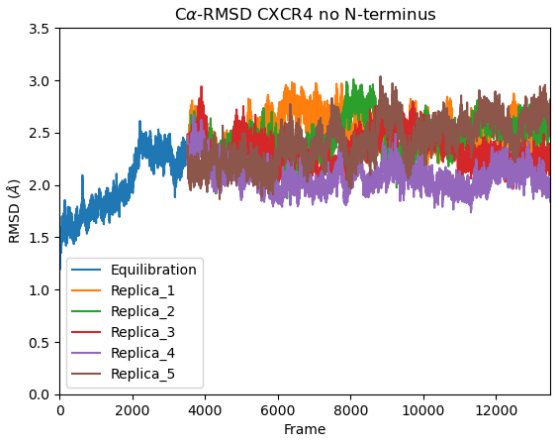
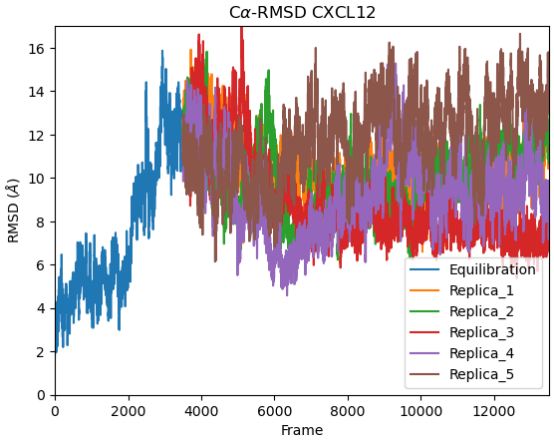


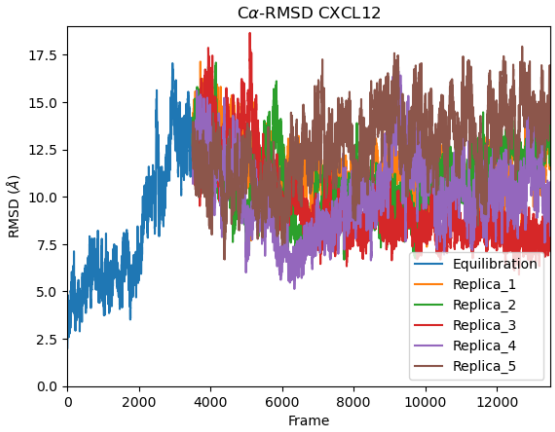
Figure C.1.3: RMSD of CXCL12 and CXCR4 after alignment of the structure based on the rigid domains of the receptor. (A) Receptor structure excluding the N-terminus (B) Full length CXCL12 (B) Rigid core (B) N-terminus (E) Distal three residues (K1-P2-V3)



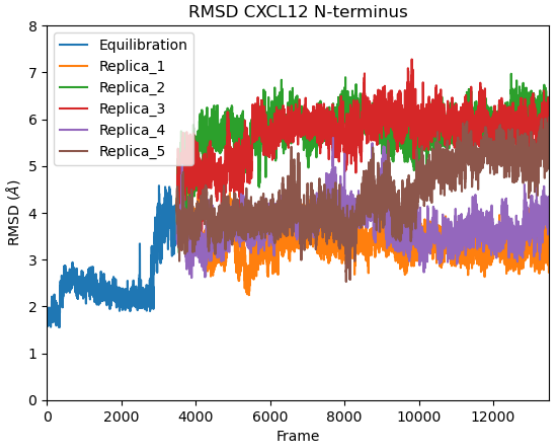
(A) CXCR4



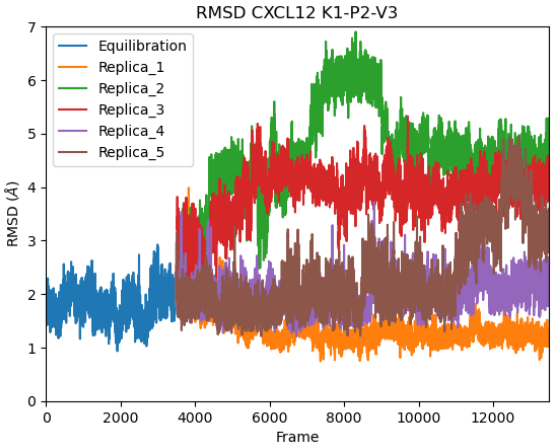
(B) CXCL12



(C) Core



(D) N-terminus



(E) K1-P2-V3

Figure C.1.4: Average interaction fingerprint describing the frequency of interaction between the three distal residues of CXCL12 and CXCR4

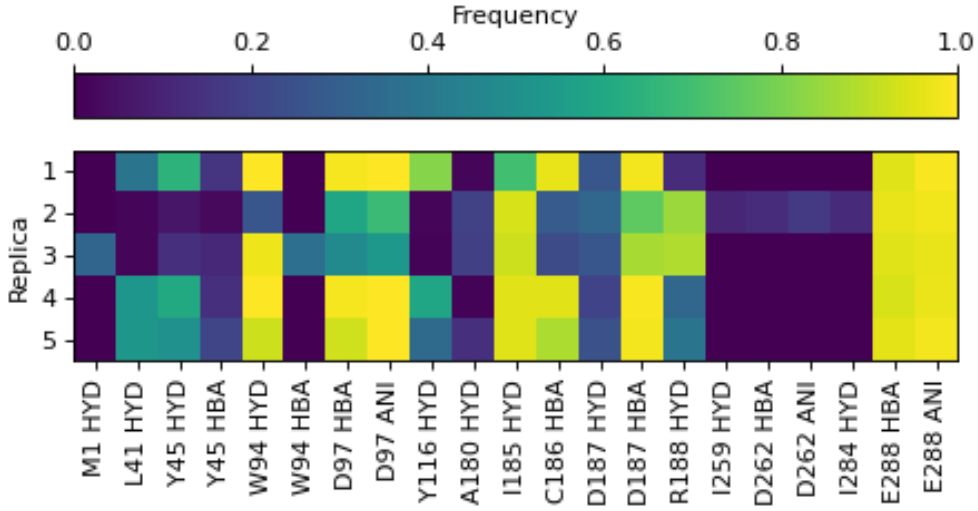


Figure C.1.5: RMSF of the chemokine after aligning the structure based on the receptor rigid transmembrane domain. Highlighted are the three distal residues of the chemokine (KPV), and the flexible N-terminus of the chemokine.

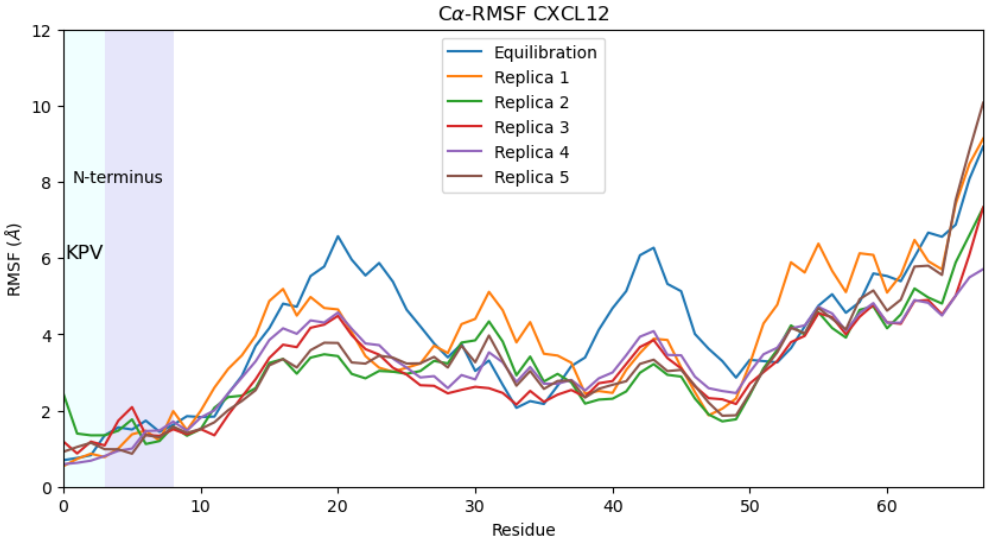


Figure C.1.6: Distribution of χ_1 angle values for Tyr45, Trp94, Tyr116, and χ_2 of Trp94 during the simulation of the IT1t-CXCR4 complex

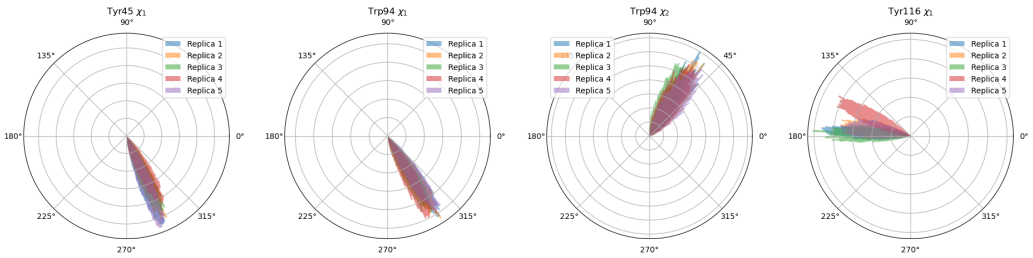


Figure C.1.7: Distribution of χ_1 angle values for Tyr45, Trp94, Tyr116, and χ_2 of Trp94 during the simulation of the CXCL12-CXCR4 complex

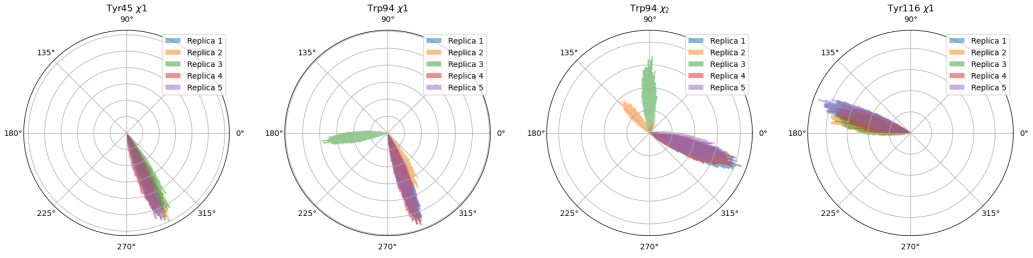
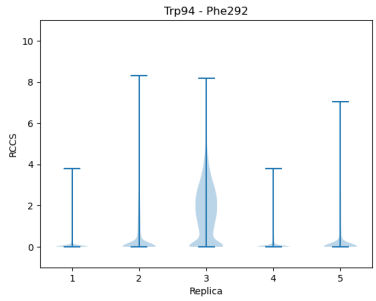
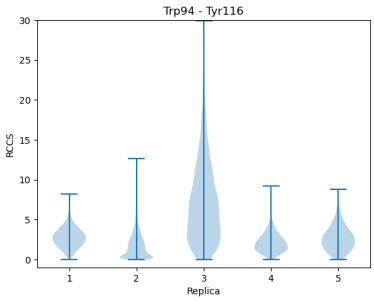


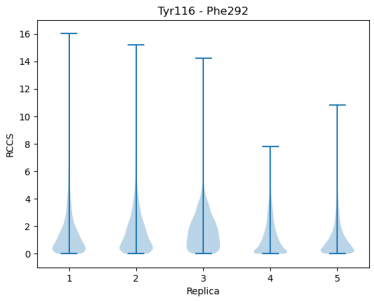
Figure C.1.8: Distribution of RRCS between Trp94, Tyr116, Phe292. Distributions are separated between different replicas. (A)(B)(C) Contacts observed in the CXCL12-CXCR4 complex. (D)(E)(F) Contacts observed in the IT1t-CXCR4 complex.



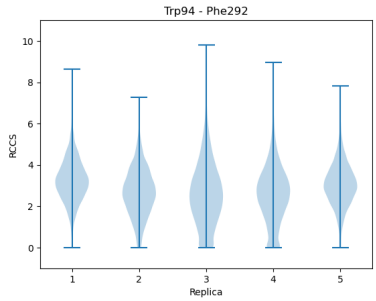
(A) Trp94-Phe292(CXCL12)



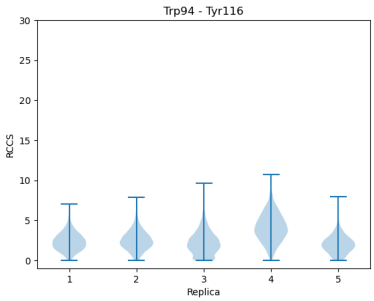
(B) Trp94-Tyr116(CXCL12)



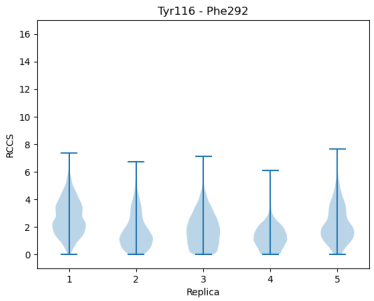
(C) Tyr116-Phe292(CXCL12)



(D) Trp94-Phe292(IT1t)

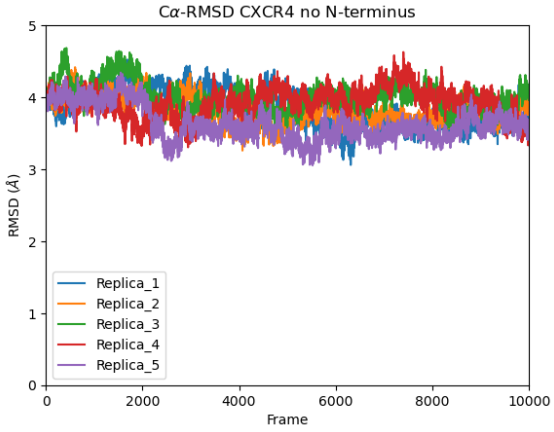


(E) Trp94-Tyr116(IT1t)

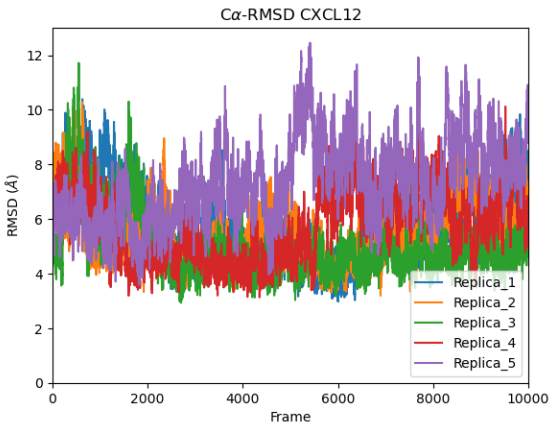


(F) Tyr116-Phe292(IT1t)

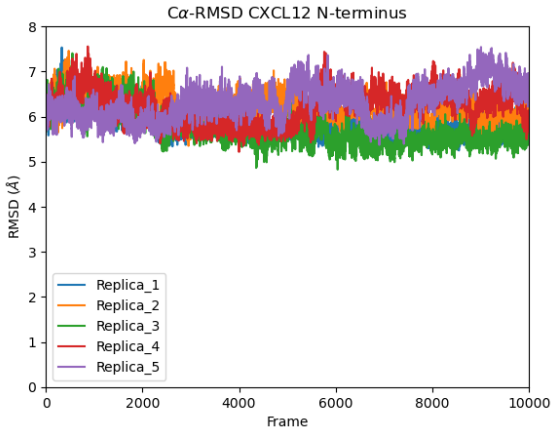
Figure C.1.9: RMSD of CXCL12 and CXCR4 after alignment of the structure based on the rigid domains of the receptor on the experimental reference. (A) Receptor structure excluding the N-terminus (B) Full length CXCL12 (C) N-terminus (D) Distal three residues (K1-P2-V3)



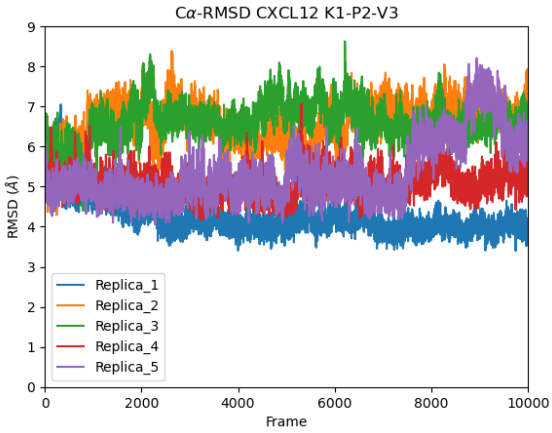
(A) CXCR4



(B) CXCL12



(C) N-terminus



(D) K1-P2-V3

Luca CHIESA

Development of artificial intelligence methods to help the design of new ligands of G-protein coupled receptors

Résumé

Les récepteurs couplés aux protéines G représentent l'une des familles de protéines les plus importantes pour la découverte de médicaments. Les techniques de conception de médicaments assistée par ordinateur ont été largement appliquées pour cibler les membres de cette famille, conduisant à la découverte de multiples molécules bioactives. Cette thèse décrit le développement, le test et l'application de différents outils informatiques et d'apprentissage automatique pour aider à la découverte de nouveaux composés présentant un profil pharmacologique souhaité. Divers aspects du processus de découverte de médicaments *in silico* ont été abordés dans ce travail, depuis la modélisation d'une protéine cible jusqu'à l'évaluation systématique des molécules identifiées par criblage virtuel.

Mots-clés : RCPG, découverte de médicaments, apprentissage automatique, récepteurs des chimiokines, basé sur la structure, chémoinformatique, criblage virtuel, similarité d'interaction

Résumé en anglais

G-protein coupled receptor represent one of the most important protein families for drug discovery. Computer aided drug design techniques have been extensively applied to target members of this family, leading to the discovery of multiple bioactive molecules. This thesis describes the development, testing, and application of different computational and machine learning tools to assist in the discovery of new compounds with a desired pharmacological profile. Different aspects of the *in silico* drug discovery pipeline were covered in this work, from the modelling of a target protein, to the systematic evaluation of molecules identified by virtual screening.

Keywords: GPCR, drug discovery, machine learning, chemokine receptors, structure-based, Cheminformatics, virtual screening, interaction similarity