UNIVERSITÉ DE STRASBOURG

École Doctorale des Sciences de la Vie et de la Santé - ED 414

THÈSE

présentée par :

Ana MILINSKI

soutenue le 17 décembre 2024

pour obtenir le grade de **Docteur de l'université de Strasbourg**
Spécialité : Biophysique et Biologie structurale

Dynamique structurale du récepteur nucléaire PPARgamma

THÈSE dirigée par :

Mr STOTE Roland	Dr, CNRS UMR 7104, Inserm U 1258, Université de Strasbourg
Mme DEJAEGERE Annick	Pr, Université de Strasbourg

RAPPORTEURS :

Mme BAUD Stéphanie	Pr, Université de Reims
Mr ZOETE Vincent	Pr, Université de Lausanne, Suisse

EXAMINATEURS :

Mme LE MAIRE Albane	Dr, CNRS UMR 5048, Inserm U 1054
Mr SPICHTY Martin	Dr, UMR 7042 Université de Haute Alsace

Avertissement au lecteur / Warning to the reader

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition des membres de la communauté universitaire. Il est soumis à la propriété intellectuelle de l'auteur. Cela implique une obligation de citation et de référencement lors de l'utilisation de ce document. D'autre part, toute contrefaçon, plagiat, reproduction ou représentation illicite encourt une poursuite pénale.
This document is the result of a long process approved by the jury and made available to members of the university community. It is subject to the intellectual property rights of its author. This implies an obligation to quote and reference when using this document. Furthermore, any infringement, plagiarism, unlawful reproduction or representation will be prosecuted.

Code de la Propriété Intellectuelle

Article L122-4 :
Toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite. Il en est de même pour la traduction, l'adaptation ou la transformation, l'arrangement ou la reproduction par un art ou un procédé quelconque.
Any representation or reproduction in whole or in part without the consent of the author or his successors in title or assigns is unlawful. The same applies to translation, adaptation or transformation, arrangement or reproduction by any art or process whatsoever.

Articles L335-1 à L335-9. : Dispositions pénales / Penal provisions.

Licence attribuée par l'auteur / Licence attributed by the author

[*Veillez à ne conserver que la licence qui vous convient* / Make sure you only keep the licence that suits you best]



https://creativecommons.org/licenses/?lang=fr-FR

Acknowledgements

My first acknowledgments go to the members of my thesis jury: Prof. Vincent Zoete, Prof. Stéphanie Baud, Dr. Martin Spichty, and Dr. Albane Le Maire. Thank you for accepting to review my work and for taking the time to participate in this jury. I look forward to having interesting discussions during the defense.
I also express my gratitude to Prof. Manuel Dauchez, Prof. Petra Hellwig and Prof. Xabier Lopez for being part of the Thesis Follow-up Committee and for your guidance throughout these years.

I am deeply thankful to my team and supervisors. First and foremost, I thank Prof. Annick Dejaegere for welcoming me to the team as an undergraduate student, allowing me to begin working on molecular modelling during the second year of my Master's. Annick, from the very first Molecular Modelling class, I was captivated by this field, and that is entirely thanks to you. You have been an incredible inspiration and role model, particularly for girls and women aspiring to pursue careers in science and beyond. Thank you for being my supervisor over these four years.

My greatest thanks go to my supervisor, Dr. Roland H. Stote, without whom this work would not have been possible.
Roland, thank you for everything - for your guidance, patience, and all that you've taught me. What I know today is all thanks to you. Working alongside you, discussing science (and more) with you, over these four years has been a privilege.

Special thanks go to Valentin Loux and Meilin An, my colleagues and now friends. Thank you for joining the team and making me feel less alone. Thank you for being there to talk about science and life, for sharing desserts in the lab, and for our walks throughout the streets of Strasbourg (and Europa Park (with Nicolas of course!)). I wish you both the best in your future careers and lives.

I extend my gratitude to other team members: Dr. Laurent Bianchetti, whom I wish all the best in work and beyond - thank you for all the insightful discussions we had in the lab. Thanks also to Dr. Natacha Rochel, Carole Peluso Illtis, and especially Dr. Judit Osz, for your contributions to the collaborative study, for the joint meetings and discussions, and for the best cookies. Aurora Silvestri, the wonderful Italian in our friend group, thank you for always bringing something precious and fun. I wish you all success and happiness.

I also wish to officially thank the organizations that made this work possible: the Agence Nationale de la Recherche (ANR) for funding three years of my thesis under grant ANR-20-CE29-0013, Title: FRIDaY, and the University of Strasbourg for giving me the opportunity to teach and fund the fourth year of my studies. My gratitude also goes to the IGBMC staff for their unwavering organization and support - it was greatly appreciated.

I am honoured to be part of two research groups in Reims: The MEDyC lab at the University of Reims and the company Apmonia Therapeutics, where I currently work.

To my Master's friends and colleagues - Fadwa, Delphine, Christelle, Victor, Franck and Georges - thank you for sharing the good and the bad during these 6 years together. Fad, Del, Chris, thank you for being such amazing girls, and for sharing all the wonderful (and less wonderful) memories, but here we are - we made it! Thank you for being there, I am so grateful

to have you as my friends. I also thank Abir, friend with a special place in our group, for being so cheerful after all the hardships, and being an example of resilience and beauty. As an extension, I would like to express my gratitude to Kretz - Krempp family, and Louis, for their help in a time of need, for encouragement, and for the moments we shared. I will always cherish these memories and be grateful for them.

I am also thankful for meeting so many wonderful people, especially Mau, Susi, and Ivi (and Hung, of course). Ivi and Susi, thank you for your constant support throughout these years. I love you both and can't wait to see you soon! Clément, merci pour tout - merci d'avoir été à mes côtés, pour ton soutien et encouragement (parfois, c'était la seule chose qui me donnait la force de continuer). I would also like to thank my childhood friends, Jovana, Jelena, Ogi, and their families. Our friendship has gone through a lot of challenges and has only grown stronger. Jocko, Jecke, thank you for always being there, for supporting me, and for sharing more than 20 years of friendship. Here's to many more adventures together!

Lastly, I want to thank my family. First, my mother, whom everyone knows as being an extraordinary woman - but even that doesn't begin to describe her. She has instilled in me countless values that I carry with me wherever I go. I am here today, writing this because of her. Maki, hvala za sve.
Thank you to my grandmother, in my home village of Bakionica, Serbia, who raised me.
I owe so much to my sisters (and by extension their families), Milica (and Djordje and Megi) and Sara (and Neno and Maksim) - your families have been a constant source of support, and I thank you all. Seki Mici, Seki Saki, you are the two best older sisters a girl can have.
Volim vas najviše.

This thesis is dedicated to my family.
Posvećeno mojoj porodici.

~ From planets to atoms ~

I have always found beauty in the duality of the moment:
As Man prepared to take first steps on the Moon, he also prepared the first force fields to describe unseen motions of molecules.

Table of Contents

Acknowledgements .....

List of Figures *4*

List of tables *6*

List of Abbreviations *7*

Chapter I - Introduction **8**

1. Biological context - nuclear receptor proteins *9*

2. Roles and classification of NRs *9*

3. Structural organisation *12*
3.1. A/B domain 13
3.2. C domain 14
3.3. D domain 17
3.4. E domain 17

4. Mechanisms of transcription regulation *19*
4.1. Transcriptional activation 19
4.2. Transcriptional repression 22
4.3. Coregulator proteins 24
4.3.1. Coactivators 25
4.3.2. Corepressors 26
4.4. Dimerisation 26

5. Peroxisome proliferator - activated receptors (PPARs) *27*
5.1. Peroxisome proliferator - activated receptor gamma (PPARγ) 28

6. Objectives of the thesis *33*
6.1 Experimental and Computational study of protein dynamics 33

Chapter II - Methodology **38**

1. Molecular modelling *39*

2. Molecular mechanics *39*
2.1. Potential energy function: All - Atom Additive Force Field 40
2.1.1. Bonded terms 41
2.1.2. Non - bonded terms 43
2.2. Potential energy function: All - Atom Polarizable Force Field 43
2.2.1. Atomic polarizability 44
2.2.1.1. Drude particle 44
2.2.1.2. Anisotropic polarizability 46
2.3. Long - range interactions 47
2.4. Solvent representation 48
2.5. Energy minimization 49
2.5.1. Steepest descent method 50
2.5.2. Conjugate gradient 51
2.5.3. Adopted-Basis Newton-Raphson 52
2.5. Normal Mode Analysis 53
2.5.1. Analysis of normal mode calculations 55
2.5.1.1. Fluctuations 55
2.5.1.2. Correlations 56
2.5.1.3. Infra - red spectra 56

3. Molecular dynamics *57*
3.1. Protocol 57
3.2. Integrating algorithm - All - Atom Additive FF 58

3.3. Integrating algorithm - All - Atom Polarizable FF....60
3.4. Periodic boundary conditions....61
3.5. Analysis....63
3.5.1. RMS deviation and fluctuation....63
3.5.2. Correlations....63
3.5.3. Secondary structure element analysis....64

Chapter III - Ensemble averaged Normal modes - a prelude....65

1. Introduction....66

2. Methods....68
2.1. Structure preparation....68
2.2. Analysis of the trajectories....69
2.3. Normal Mode Analysis....69

3. Results and discussion....71
3.1. RMSD, RGYR and RMSF....71
3.2. Normal mode calculations....74
3.2.1. Fluctuations....74
3.2.2. Correlated motions....76
3.2.3 Far-IR spectra....80

4. Conclusions....82

Chapter IV - Application of Ensemble-averaged Normal modes to PPARγ - a combined MD/Far-IR study....84

1. Introduction....85

2. Methods....90
2.1. Model preparation....90
2.2 Determination of histidine protonation states....91
2.3. Molecular dynamics simulation parameters....94

3. Results and discussion....95
3.1. RMSD and RMSF....95
3.2. RMSD - Radius of gyration based free-energy landscapes....100
3.3. Normal Mode Analysis....107
3.3.1. Computed fluctuations....107
3.3.2. Computed Infrared spectra....110
3.3.3. Secondary structure analysis....115
3.3.4. Correlated motions....120
3.3.5. H-bond analysis....124
3.3.5.1. Backbone H-bond analysis....124
3.3.5.2. Sidechain H-bond analysis....125
3.3.6. Community network analysis (CNA)....130

4. Conclusions....138

5. Annex: Experimental methodology....140
5.1. Mid-Infrared (Mid-IR) Measurements....140
5.2. Far-Infrared (Far-IR) Measurements....140
5.3. Determination of Protein Secondary Structure....141

Chapter V - Impact of force field polarization on the collective motions of PPARγ....142

1. Introduction....143

2. Methods and Analysis....147

3. Results and Discussion....149
3.1. Structural dynamics of PPARγ – RMSD and RMSF....149
3.2. Electric dipole moments....154
3.3. Correlated motions....156
3.3.1. Community Network Analysis....158

3.3.2. Shortest Path Map (SPM) 166
4. Conclusions *170*
Chapter VI - Conclusions and Perspectives **172**
Résumé **178**
References **193**

List of Figures

FIGURE 1. REPRESENTATION OF NUCLEAR RECEPTOR MODULAR ORGANISATION....12
FIGURE 2. DNA BINDING DOMAIN....15
FIGURE 3. REPRESENTATION OF NR - DNA BINDING....16
FIGURE 4. LIGAND BINDING DOMAIN....18
FIGURE 5. LIGAND BINDING POCKETS OF THREE NUCLEAR RECEPTORS....19
FIGURE 6. CONFORMATIONAL CHANGE OF THE LBD HELIX 12 UPON LIGAND AND COACTIVATOR PEPTIDE BINDING....20
FIGURE 7. SCHEMATIC OF NUCLEAR RECEPTOR ACTIVATION BY SMALL MOLECULE AGONIST....23
FIGURE 8. CRYSTALLOGRAPHIC STRUCTURE OF THE FULL-LENGTH PPARGAMMA AND RXRALPHA HETERODIMER BOUND TO DNA DR1....32
FIGURE 9. SCHEMATIC REPRESENTATION OF THE PPARGAMMA LIGAND BINDING DOMAIN....33
FIGURE 10. DIFFERENT CONTRIBUTIONS OF THE POTENTIAL ENERGY FUNCTION....42
FIGURE 11. REPRESENTATION OF THE DRUDE OSCILLATOR MODEL....44
FIGURE 12. EXPLICIT SOLVENT MODELS: TIP3P AND SWM4 NDP....48
FIGURE 13. REPRESENTATION OF DIFFERENT ENERGY MINIMIZATION ALGORITHMS....51
FIGURE 14. SCHEMATIC OF PERIODIC BOUNDARY CONDITIONS....62
FIGURE 15. PPARGAMMA LIGAND BINDING DOMAIN....67
FIGURE 16. RMSD OF THE CA ATOMS OF THE COMPLETE BACKBONE....71
FIGURE 17. RADIUS OF GYRATION OF THE CA ATOMS OF THE COMPLETE BACKBONE....71
FIGURE 18. BY-RESIDUE AVERAGES BACKBONE RMS FLUCTUATIONS....72
FIGURE 19. THE EFFECTIVE FREE ENERGY LANDSCAPE BUILT FROM THE RMSD AND THE RGYR....73
FIGURE 20. ATOMIC FLUCTUATIONS OF BACKBONE ATOMS FROM THE MOLECULAR DYNAMICS SIMULATIONS....75
FIGURE 21. CORRELATED MOTIONS FROM THREE INDIVIDUAL STRUCTURES....76
FIGURE 22. CORRELATED MOTIONS CALCULATED FROM THE NMA....77
FIGURE 23. CORRELATED MOTIONS CALCULATED FROM NORMAL MODES....79
FIGURE 24. CALCULATED FAR-IR SPECTRA....81
FIGURE 25. CONVERGENCE OF THE SPECTRUM....82
FIGURE 26. THE CHEMICAL STRUCTURE OF THE PPARGAMMA AGONIST GW1929....88
FIGURE 27. STRUCTURES USED IN THE STUDY....89
FIGURE 28. THE THREE DIFFERENT PROTONATION STATE OF HISTIDINE....92
FIGURE 29. THE CRYSTALLOGRAPHIC STRUCTURE OF THE HOLO WT PPARGAMMA BOUND TO GW1929....94
FIGURE 30. RMSD TIMESERIES GRAPHS OF 200 NS SIMULATIONS OF PPARGAMMA LBD....97
FIGURE 31. RMS FLUCTUATIONS OF PPARGAMMA LBD, FOR FOUR BIOLOGICAL SYSTEMS....99
FIGURE 32. RMSD – RGYR FREE ENERGY LANDSCAPES FOR INDIVIDUAL REPLICAS....102
FIGURE 33. RMSD – RGYR FREE ENERGY LANDSCAPES FOR INDIVIDUAL REPLICAS....103
FIGURE 34. RMSD – RGYR FREE ENERGY LANDSCAPES FOR INDIVIDUAL REPLICAS....104
FIGURE 35. RMSD – RGYR FREE ENERGY LANDSCAPES FOR INDIVIDUAL REPLICAS....105
FIGURE 36. RMS FLUCTUATIONS CALCULATED FROM MD TRAJECTORIES, NM CALCULATIONS AND CRYSTALLOGRAPHIC B FACTORS....108
FIGURE 37. THE EXPERIMENTAL MID- AND FAR- IR ABSORBANCE SPECTRA....112
FIGURE 38. COMPUTED IR SPECTRA, FROM NORMAL MODE CALCULATIONS....114
FIGURE 40. SECONDARY STRUCTURE CONTENT AS A FUNCTION OF TIME POINT....119
FIGURE 41. CORRELATED MOTIONS CALCULATED FROM NORMAL MODES....122

FIGURE 42. CORRELATED MOTIONS CALCULATED FROM NORMAL MODES.....123
FIGURE 43. AVERAGE NUMBER OF HYDROGEN BONDS 125
FIGURE 44. H-BONDS MAPPED ONTO THE AVERAGE STRUCTURE.....127
FIGURE 45. H-BONDS MAPPED ONTO THE AVERAGE STRUCTURE.....129
FIGURE 46. CNA OF THE APO WT SYSTEM.....132
FIGURE 47. CNA OF THE HOLO-WT WITH GW1929 SYSTEM.....133
FIGURE 48. CNA OF THE T475M SYSTEM 136
FIGURE 49. CNA OF THE F310S SYSTEM. 137
FIGURE 50. PPARΓ LIGAND BINDING DOMAIN. 145
FIGURE 51. RMSD OF PPARGAMMA LBD.....150
FIGURE 52. RMSF OF PPARΓ LBD 152
FIGURE 53. RADIALLY AVERAGED ATOMIC FLUCTUATIONS 153
FIGURE 54. PROTEIN DIPOLE MOMENT TIMESERIES.....154
FIGURE 55. BY-RESIDUE DIPOLE MOMENTS OF PPARGAMMA. 155
FIGURE 56. CORRELATED MOTIONS CALCULATED FROM THE SIMULATION OF THE PPARΓ LBD APO AND COREPRESSOR BOUND FORM 157
FIGURE 57. COMMUNITY NETWORK ANALYSIS OF THE PPARGAMMA LBD.....162
FIGURE 58. COMMUNITY NETWORK ANALYSIS OF PPARGAMMA WITH COREPRESSOR PEPTIDE.....164
FIGURE 59. SHORTEST PATH METHOD OF PPARGAMMA LBD APO FORM.. 168
FIGURE 60. SHORTEST PATH METHOD OF PPARGAMMA LBD WITH COREPRESSOR PEPTIDE BOUND169

List of tables

TABLE 1. NUCLEAR RECEPTOR CLASSIFICATION. 11
TABLE 2. THE PROPKA SUGGESTIONS FOR HISTIDINE PROTONATION STATES. 93
TABLE 3. VALUES OF RMSD AND RADIUS OF GYRATION 106
TABLE 4. NUMBER OF ATOMS AND NUMBER OF NORMAL MODES OF EACH SYSTEM 107
TABLE 5. SECONDARY STRUCTURE CONTENT AS DETERMINED BY MID-IR SPECTROSCOPY FROM ANALYSIS OF THE AMIDE I AND AMIDE VI BANDS. 116
TABLE 6. PERCENT OF SECONDARY STRUCTURE CALCULATED FROM THE CRYSTAL STRUCTURES. 117
TABLE 7. SECONDARY STRUCTURE CONTENT CALCULATED FROM THE MOLECULAR DYNAMICS SIMULATIONS. 118
TABLE 8. COMPOSITION OF THE NODES FROM THE COMMUNITY NETWORK ANALYSIS. 160

List of Abbreviations

NR: Nuclear Receptor protein
DBD: DNA Binding Domain
LBD: Ligand Binding Domain
LBP: Ligand Binding Pocket
DNA: Deoxyribonucleic acid
RNA: Ribonucleic acid
RNA Pol II: RNA Polymerase II
PGC1α: Peroxisome proliferator-activated receptor gamma coactivator 1-alpha
NMR: Nuclear Magnetic Resonance
PDB: Protein Data Bank
SD: Steepest descent
CG: Conjugate Gradient
ABNR: Adopted-Basis Newton Raphson
NMA: Normal Mode Analysis
RMS: Root Mean Square
RMSD: Root Mean Square Deviation
RMSF: Root Mean Square Fluctuations
RGYR: Radius of gyration
PBC: Periodic Boundary Conditions
WT: Wild Type
TF: Transcription factors
ID: intrinsically disordered
FA: fatty acids
MUFA: mono-unsaturated fatty acids
PUFA: polyunsaturated
GTF: general transcription factor
AF-1: Activating Function 1
AF-2: Activating Function 2
SMRT: Silencing mediator of retinoid and thyroid hormone receptors
NCoR: Nuclear receptor corepressor
PES: potential energy surface

Chapter I - Introduction

1. Biological context - nuclear receptor proteins

The proper and efficient functioning of living cells relies on the intricate regulation of gene expression (Cramer 2019). This process is vital for maintaining cellular homeostasis in response to constant changes in cell environment. In eucaryotic cells, protein gene transcription is carried out sequentially, comprising three main stages: initiation, elongation and termination. Each of these steps is regulated by a variety of proteins, namely general transcription factors (GTFs), activators and co-activators (Maston, Evans, and Green 2006). For the transcription to begin, various activators and coactivators must bind to specific upstream regions of the DNA, and their function is to regulate and facilitate chromatin remodelling, as well as to assemble additional proteins necessary for the initiation of transcription. Activators, also named transcription factors (TFs), are sequence-specific DNA-binding proteins that are classified based on their DNA-binding domain structural organization. One superclass of transcription factors named "Zinc-coordinating DNA-binding domains", has either cysteines, histidines, or both, coordinating zinc ions. Belonging to this superclass, along with nine others, is the class of proteins called nuclear receptors (NRs) (Wingender, Schoeps, and Dönitz 2013).

2. Roles and classification of NRs

Nuclear receptor proteins are found in all metazoans classes, from sponges to vertebrates (Miglioli et al. 2021). They govern the transcription of a large variety of genes necessary for driving key biological processes, including development, cell proliferation and apoptosis (J. P. Renaud and Moras 2000a). They are also critical determinants of everyday health via their roles in metabolism and circadian rhythms (Ray 2022).

In humans, there are 48 nuclear receptor proteins. They modulate transcription by selectively binding small-molecule lipophilic ligands, thus providing a direct link between signalling molecules and gene transcription (Rastinejad et al. 2013). Their natural ligands are small hydrophobic molecules, such as hormones, vitamins, sterols, bile- and fatty acids. NRs that have no identified natural ligands are called nuclear orphan receptors (19 out of 48 human

NRs), but they also exert active regulation of numerous genes (Tao et al. 2020). The question remains open whether this is due to their intrinsic activation activity, making them ligand-independent ("constitutively active"), or due to unknown metabolites acting as their ligands (Tao et al. 2020). Due to their large number of proteins, their different functions and ligands, there are several classifications of NRs based on different criteria. Here we adopt the classification based on sequence alignment and phylogenetic tree analysis of NRs shown in Table 1 (Germain et al. 2006; Owen and Zelent 2000; Weikum, Liu, and Ortlund 2018a).

Table 1. Nuclear receptor classification. Adapted from (Weikum et al., 2018).

Family	Name	Abbreviation	Gene name	Ligand
0B	Dosage-sensitive sex reversal-adrenal hypoplasia congenital critical region on the X chromosome, Gene 1	DAX1	*NROB1*	Orphan
	Short heterodimeric partner SHP NROB2 Orphan	SHP	*NR0B2*	Orphan
1A	Thyroid hormone receptor-α	TRα	*THRA*	Thyroid hormones
	Thyroid hormone receptor-β	TRβ	*THRB*	Thyroid hormones
1B	Retinoic acid receptor-α	RARα	*RARA*	Retinoic acids
	Retinoic acid receptor-β	RARβ	*RARB*	Retinoic acids
	Retinoic acid receptor-γ	RARγ	*RARG*	Retinoic acids
1C	Peroxisome proliferator-activated receptor-α	PPARα	*PPARA*	Fatty acids
	Peroxisome proliferator-activated receptor-β	PPARβ	*PPARD*	Fatty acids
	Peroxisome proliferator-activated receptor-γ	PPARγ	*PPARG*	Fatty acids
1D	Reverse-Erb-α REV-ERBα	REV-ERBα	*NR1D1*	Heme
	Reverse-Erb-β REV-ERBβ	REV-ERBβ	*NR1D2*	Heme
1F	Retinoic acid-related orphan-α	RORα	*RORA*	Sterols
	Retinoic acid-related orphan-β	RORβ	*RORB*	Sterols
	Retinoic acid-related orphan-γ	RORγ	*RORC*	Sterols
1H	Farnesoid X receptor	FXRα	*NR1H4*	Bile Acids
	Farnesoid X receptor-β	FXRβ	*NR1H5P*	Orphan
	Liver X receptor-α	LXRα	*NR1H3*	Oxysterols
	Liver X receptor-β	LXRβ	*NR1H2*	Oxysterols
1I	Vitamin D receptor	VDR	*VDR*	1α,25-dihydroxyvitamin D3
	Pregnane X receptor PXR NR1I2	PXR	*NR1I2*	Endobiotics and xenobiotics
	Constitutive androstane receptor		*NR1I3*	Xenobiotics
2A	Hepatocyte nuclear Factor-4-α	HNF4α	*HNF4A*	Fatty acids
	Hepatocyte nuclear Factor-4-γ	HNF4γ	*HNF4G*	Fatty acids
2B	Retinoid X receptor-α	RXRα	*RXRA*	9-Cis retinoic acid
	Retinoid X receptor-β	RXRβ	*RXRB*	9-Cis retinoic acid
	Retinoid X receptor-γ	RXRγ	*RXRG*	9-Cis retinoic acid
2C	Testicular Receptor 2	TR2	*NR2C1*	Orphan
	Testicular Receptor 4	TR4	*NR2C2*	Orphan
2E	Tailless homolog orphan receptor	TLX	*NR2E1*	Orphan
	Photoreceptor-cell-specific nuclear receptor	PNR	*NR2E3*	Orphan
2F	Chicken ovalbumin upstream promoter-transcription factor α	COUP-TFα	*NR2F1*	Orphan
	Chicken ovalbumin upstream promoter-transcription factor β	COUP-TFβ	*NR2F2*	Orphan
	Chicken ovalbumin upstream promoter-transcription factor γ	COUP-TFγ	*NR2F6*	Orphan
3A	Estrogen receptor-α	ERα	*ESR1*	Estrogens
	Estrogen receptor-β	Erβ	*ESR2*	Estrogens
3B	Estrogen-related receptor-α	ERRα	*ESRRA*	Orphan
	Estrogen-related receptor-β	ERRβ	*ESRRB*	Orphan
	Estrogen-related receptor-γ	ERRγ	*ESRRC*	Orphan
3C	Androgen receptor	AR	*AR*	Androgens
	Glucocorticoid receptor	GR	*NR3C1*	Glucocorticoids
	Mineralocorticoid receptor	MR	*NR3C2*	Mineralocorticoids and glucocorticoids
	Progesterone receptor	PR	*PGR*	Progesterone
4A	Nerve growth Factor	NGF1-B	*NR4A1*	Orphan
	Nurr-related Factor 1	NURR1	*NR4A2*	Unsaturated fatty acids
	Neuron-derived orphan Receptor 1	NOR-1	*NR4A3*	Orphan
5A	Steroidogenic Factor 1	SF-1	*NR5A1*	Phospholipids
	Liver receptor Homolog-1	LRH-1	*NR5A2*	Phospholipids
6A	Germ cell nuclear factor	GCNF	*NR6A1*	Orphan

3. Structural organisation

Nuclear receptors are multi-domain proteins that have a shared modular structure where different regions corresponding to autonomous domains execute different functions. This canonical structural and functional organisation is represented in Fig. 1A, where the canonical organization starts with the A/B N terminal domain also called the Activating Function 1 (AF-1) domain; followed by the C, or DNA binding domain (DBD), the hinge region D and terminating with the E or ligand binding domain (LBD), also called the Activating Function 2 (AF-2) domain. The F domain is not present in all nuclear receptor proteins. Nuclear receptors generally function as homo- or hetero-dimers. In Fig. 1B, the domain structures of several nuclear receptor proteins are represented to illustrate the diversity of specific domains, in particular the length of the A/B domain.

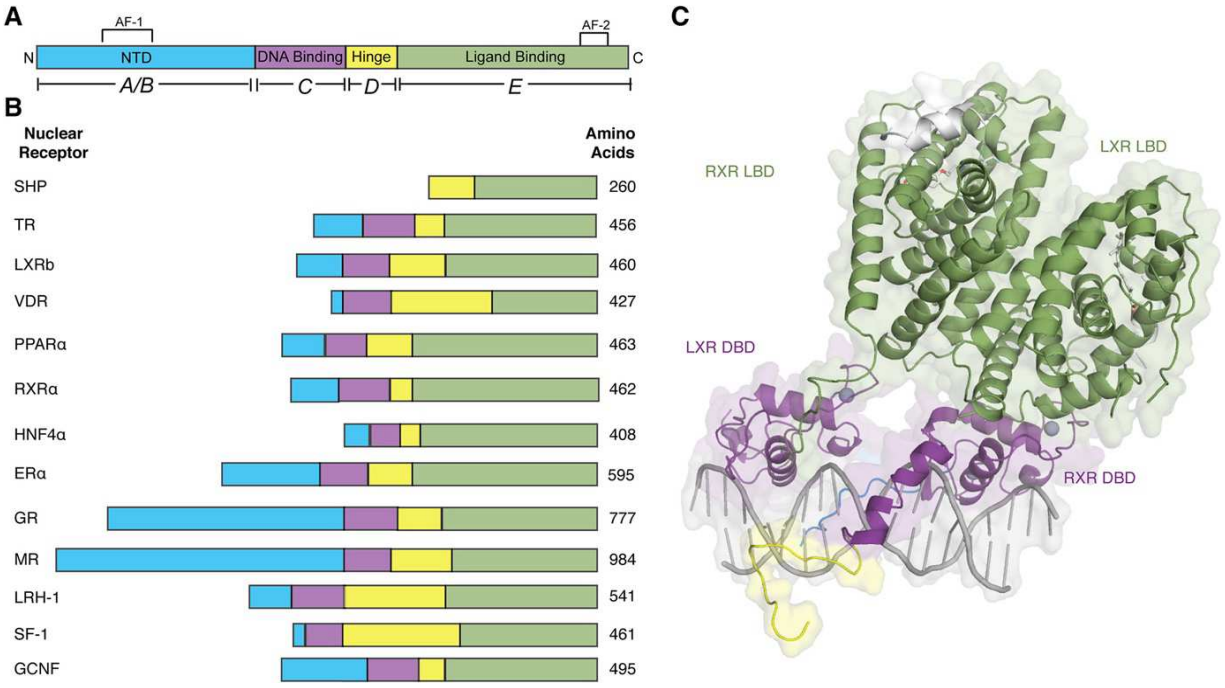


Figure 1. Representation of nuclear receptor modular organisation. A) Sequential order of modular domains. (The F domain is not represented since it is not common to all receptors) B) Examples of nuclear receptors, their sequence versatility and length. C) Full-length NR structure of LXR-RXR heterodimer (PDB: 4NQA). The colours of the structure match the colours of domains in the schematic representation. Figure from (Weikum et al., 2018)

3.1. A/B domain

The N-terminal A/B domain of nuclear receptors is an intrinsically disordered, sequence variable domain with lengths ranging from 8 to 600 residues, depending on the NR protein (Fig. 1A, B) (Wärnmark et al. 2003). Many nuclear receptors have several isoforms, which are derived from alternative splicing of a single gene. Differences between isoforms are mainly found in the A/B domain. It is revealed that the ordering of this domain is not substantially changed upon ligand binding (Goswami et al. 2014). The A/B domain contains the autonomous AF-1 activation function that mediates the recruitment of multiple transcriptional coregulatory proteins in gene regulation (Shamilov and Aneskievich 2019). NRs recruit coregulator proteins mediated by the AF-1 domain in ligand-independent manner. The AF-1 surface is known to bind coregulators which enable cooperative function between AF-1 and AF-2 regions and thus cooperatively enhance transactivation (Bugge et al. 2009; Pawlak, Lefebvre, and Staels 2012). In addition, coregulator-linked interactions with the N-terminal and C-terminal domains were found for AR, ER and PR (Table 1) (Wärnmark et al. 2003).

The A/B domains are frequently targeted by phosphorylation and other post-translational, covalent modifications, such as SUMOylation, which confer distinct functional properties of nuclear receptors. In the case of ligand-activated receptors, AF-1 modifications generally have a tissue-specific modulatory effect on their transcriptional properties (Gronemeyer, Gustafsson, and Laudet 2004). Studies show that receptors like RAR can be phosphorylated by cyclin-dependent kinases, a process that is important for both ligand-dependent and ligand-independent transactivation (Bour et al. 2005a; Gaillard et al. 2006). For example, phosphorylation of the A/B domain of GR by p38 MAPK was shown to induce stable tertiary structure formation in this domain, hence favouring its interaction with coregulatory proteins (Huppunen, Wohlfahrt, and Aarnisalo 2004; Nader et al. 2010). Due to it being an intrinsically disordered (ID) domain, no crystallographic structures of the N-terminal domain have been resolved. However, one experimental study, by cryo-electron microscopy (cryo-EM) shows full-length androgen receptor (AR) homodimer bound to DNA and two coactivator proteins. The N-terminal domains are experimentally characterised as being wrapped around

the LBDs of the dimer, providing a platform for coactivator binding (Yu et al. 2020). Kumar and Thompson proposed that, since the A/B domain is structured when involved in transcriptional activation, it can rapidly and reversibly adopt various configurations that are available for binding by coregulator proteins (R. Kumar and Thompson 2012).

3.2. C domain

The DNA-binding domain (DBD), or the C domain, enables specific recognition of, and binding to target DNA sequences (Claessens and Gewirth 2004). It is, structurally and functionally, a highly conserved domain consisting of 60-70 amino acids that are organised as a globular domain of two α-helices with short anti-parallel β-strands, see Fig. 2 (Pohl and Tomlinson 2020). It contains two zinc finger motifs, in which zinc ions are coordinated by four highly conserved cysteine residues (Fig. 2A). These two zinc finger motives are responsible for DNA sequence recognition (Freedman et al. 1988). The C domain contains several sequence elements, named P-, D-, T- and A-boxes (Novac and Heinzel 2004). The P-box consist of key residues that enable the accurate identification of the major groove of specific DNA sequences (Fig. 2B). D-box is responsible for DNA-dependent dimerization and, more precisely, for the half-site spacing. Additional T- and A-boxes, which are at the C-terminal extension (CTE) of this domain (or sometimes even in the D domain), contain amino acid residues that are essential for the creation of functional dimers. Studies have shown that the dimerization of DNA-binding domains takes place simultaneously with their interaction with DNA, rendering DNA an allosteric effector (Claessens and Gewirth 2004).

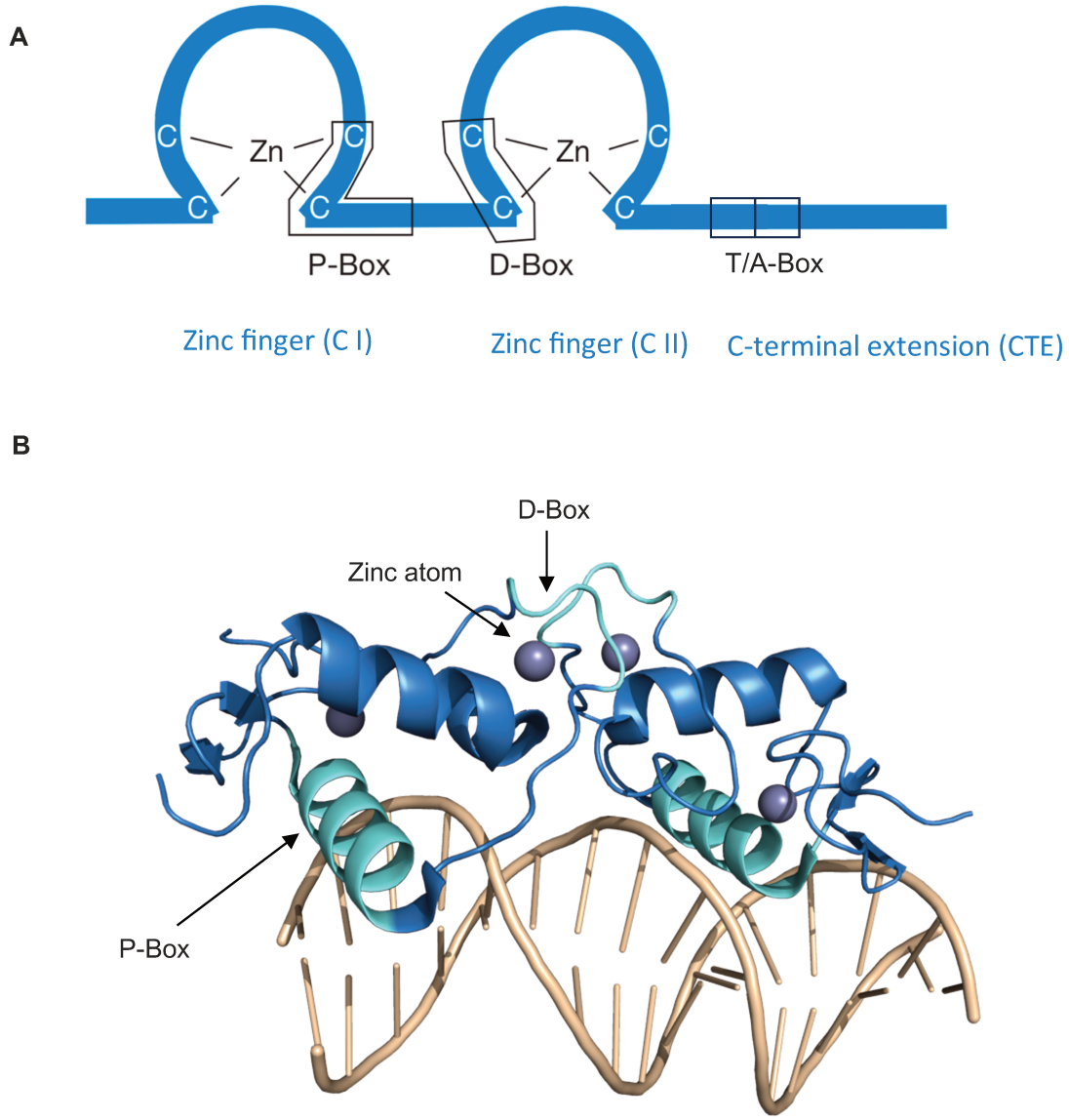


Figure 2. DNA binding domain. A) Schematic representation of the DNA-binding domain with two zinc fingers, and B) the structure of steroid receptor GR (glucocorticoid receptor) homodimer DBD, bound to DNA double helix, through interactions in major groove. P- and D-boxes are coloured in light blue. The rest of GR domain is coloured in dark blue, and the DNA in light orange. C) One-letter code sequence of the P- and D-boxes for the GR receptor. Figure A) adapted from Claessens & Gewirth, 2004.

In order to regulate transcription, nuclear receptors bind DNA by recognising specific sequences called hormone response elements (HREs). HREs are organised in 'half-sites', and can take on different patterns of organisation of the consensus recognition motif. This recognition motif contains nucleotide sequences 5'-AGGTCA-3' or 5'-AGAACA-3'. The specificity of NR binding to a particular promoter region is related to the placement of these repeats (Evans and Mangelsdorf 2014). They are organised as one (for monomeric receptors) or two (for dimeric receptors) repeats. Two repeats can be direct or inverted, see Fig. 3. For example, the steroid hormone receptors bind to direct or inverted repeats, while other heterodimeric nuclear receptors bind direct repeats. For the NRs that bind direct repeats, the space between the core recognition motifs dictates the binding specificity. The spacing between repeats can be 1 - 5 base pairs, denominated ID1 - ID5, or DR1 - DR5. If there is no spacing (no nucleotides) between direct or inverted repeats, they are called tandem and palindrome repeats, respectively. The discrimination between these hexamers is mainly due to differences in the structure of the response elements, allowing the intercalation of water molecules between the DNA and DBD residues, thus destabilising the complex when there is an incorrect match (Gewirth and Sigler 1995). One classification of nuclear receptors is based on the organisation of half-sites to which they bind. The class I and class II receptors bind to two half-sites organised as inverted repeats (IRs) of different sequences. Class III receptors bind to direct repeats (DRs), while class IV receptors typically bind to unique half-sites as monomers.

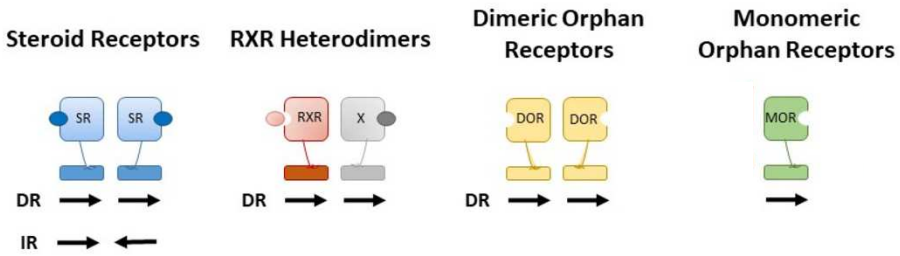


Figure 3. Representation of NR - DNA binding. Schematic representation of NR binding to DNA sequences called repose elements. DR: Direct repeats; IR: Inverted repeats. Examples of nuclear receptors: SR: steroid receptors. RXR: Retinoid X receptor, in heterodimer, where X is another NR. DOR: orphan receptors without known ligands. Monomeric orphan receptor is for example NGF1-B. Figure adapted from Pastori, V., Pozzi, S., Labedz, A., Ahmed, S., & Ronchi, A. E. (2022).

3.3. D domain

The D domain, also referred to as the hinge region, connects the DNA and ligand-binding domains, facilitating the receptor's structural flexibility (Novac and Heinzel 2004). This region had been initially seen as a flexible region with no particular function, however several nuclear receptor studies have revealed different functions. The D domain contains the nuclear receptor localisation signal (NLS). It also provides the ability to bind to DNA through the T- and A-boxes, which are found in the conserved N-terminal of this domain. Even though these two box elements are in the hinge region, they are ascribed the function of the DNA-binding domain. The T- and A-boxes are involved in NR dimerisation and half-site recognition (DNA minor groove), respectively. The hinge region of thyroid receptor is seen to accommodate corepressor proteins that inactivate the nuclear receptor. The hinge region is also known to accommodate some residue mutations that impede ligand binding when bound to the DNA, and, therefore, by extension prevent corepressor protein release (Safer et al. 1998).

3.4. E domain

The ligand-binding domain (LBD) is a multifunction domain comprised of around 250 residues and is structurally highly conserved within and between species (Mitsis et al. 2019). Nuclear receptors share a similar overall conformation of LBDs, consisting of 11 - 13 alpha-helices, and a small beta sheet, ordered in a 'three-layered sandwich' (Moras and Gronemeyer 1998). The canonical structure of this domain has been resolved by crystallographic experiments many times (Fig. 4). Variations of the canonical structure involve certain receptors, such that they present an additional helix between H1 and H3. The 'sandwich'-like structure has the first layer made of helices H1 - H3, the second layer is composed of helices H4, H5, H6, H8 and H9, and the final layer composed of helices H7, H10 and H11. The H12 is the C-terminal flexible helix, also known as the AF-2, contributes to the NR functionality by its flexibility. The structure is assembled around a hydrophobic core, or pocket, which can accommodate various lipophilic ligands (Rastinejad et al. 2013).

The ligand-binding domains, although similar in their architecture, have a varying sequence in the ligand-binding pocket (LBP), which makes them highly variable in size and specificity (Fig. 5). This contributes to the specificity of each nuclear receptor for its endogenous ligands.

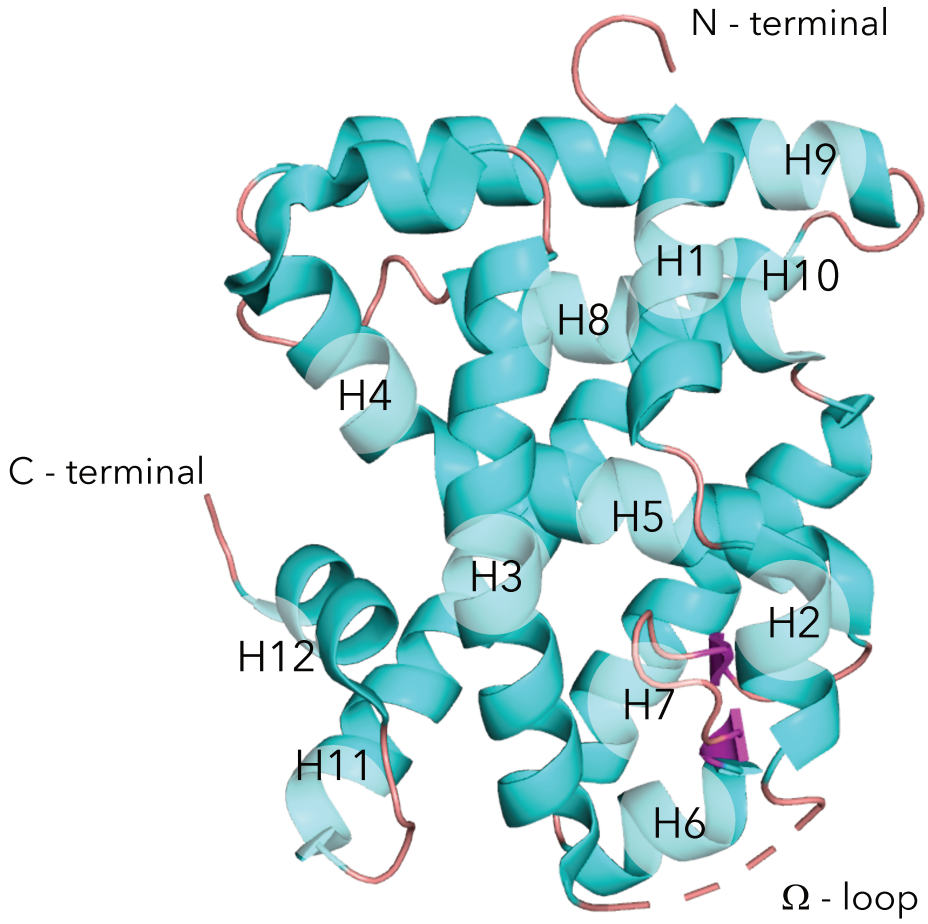


Figure 4. Ligand binding domain. Ligand-binding domain of structure of the nuclear receptor retinoid X receptor (PDB ID 1MZN). Structure contains 12 alpha helical elements (in cyan), one beta sheet (in pink). The Omega-loop is flexible.

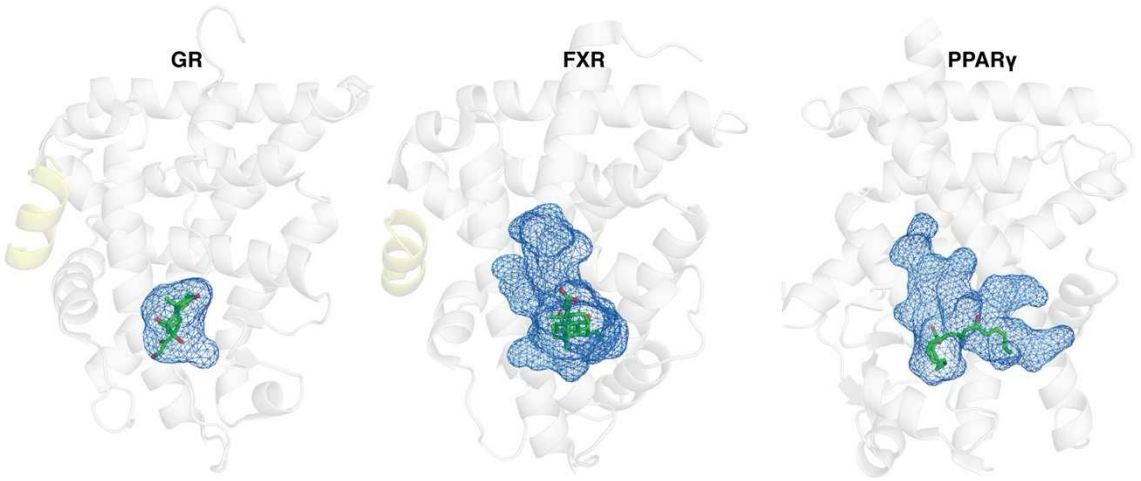


Figure 5. Ligand binding pockets of three nuclear receptors. LBPs of (GR, FXR, and PPARgamma are in blue, represented for comparison of their size. Figure from (Weikum et al., 2018)

4. Mechanisms of transcription regulation

4.1. Transcriptional activation

The LBD is implicated in multiple functions - ligand binding, protein dimerization and the co-regulatory protein binding, through the ligand-dependent transactivation function, AF-2 (Germain and Bourguet 2013). The structural changes seen in the LBD upon ligand binding concern the increase in the compactness of the domain and a specific change in the position of the C-terminal helix H12. The helix adopts a distinct transcriptionally 'active conformation', characterised by numerous crystallographic structures (Fig. 6). The active conformation of H12 makes it more stable, positioned adjacent to helices H3, H4 and H11. The interface made of these helices constitutes a platform for coactivator protein binding, after the release of corepressor protein. This model of receptor activation in which H12 closes on the ligand-binding site in response to ligand binding, rendering the receptor active and ready to accommodate coactivator protein is termed 'the mousetrap model' (J.-P. Renaud et al. 1995).

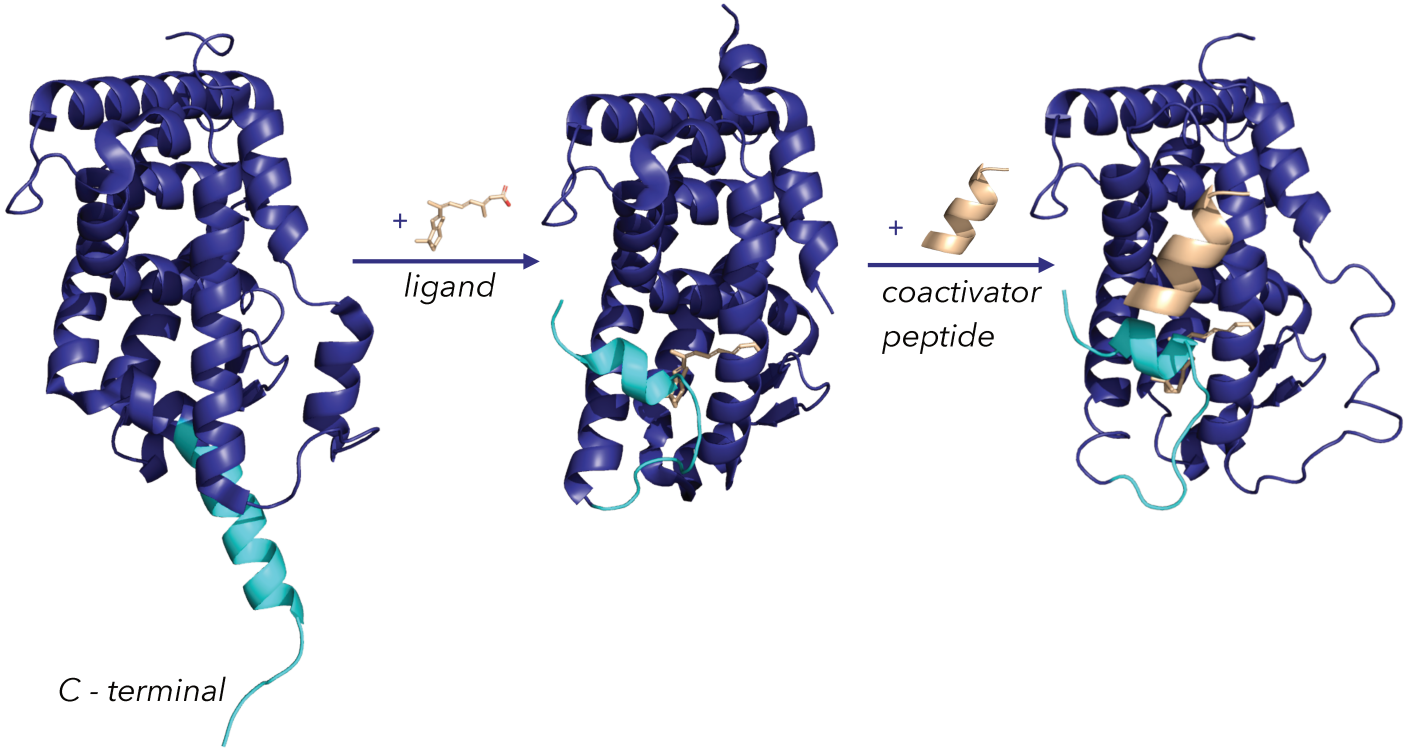


Figure 6. Conformational change of the LBD helix 12 upon ligand and coactivator peptide binding. Crystallographic structures PDB ID (from left to right): 6HN6, 1FBY, 1FM9. LBD is in dark blue, H12 residues in cyan, ligand (9-cis-retinoic acid) and coactivator peptide (steroid receptor coactivator SRC1) in beige.

Today, this model is considered to be an oversimplification of the receptor activation. The first structural studies of NRs rendered a static image of the function of individual domains, and while this has been an incredible achievement, many complementary experimental studies, including those of full-length nuclear receptor complexes suggest a more complex view of NR regulation (Orlov et al. 2012; Chandra et al. 2008). This is the case, for example, for NRs that are activated by different ligands acting in a tissue-specific fashion that impacts their function, comparing to NRs with only one specific ligand (Holzer, Markov, and Laudet 2017).

It is clear that not all understanding can be inferred from 3D crystal structures. Indeed, there is a growing awareness that the local conformational changes revealed by structural studies are just one piece of the regulatory puzzle. Ligand binding has a profound effect on the structural dynamics of NRs (de Vera et al. 2017; Hughes et al. 2012a; B A Johnson et al. 2000;

Kallenberger et al. 2003; le Maire et al. 2010) and there are examples where the binding of dissimilar ligands results in similar structures of the LBD/ligand complex as determined by X-ray crystallography, but the regulatory consequences are distinct. This suggests a contribution of structural dynamics to selective targeting of different regulatory pathways (Nwachukwu et al. 2016; Choi et al. 2010; 2011; Chrisman et al. 2018; Nettles et al. 2008; Zheng et al. 2018).

So, adding to the foundations of the mouse trap model is a more dynamical view of the activation. Dynamical models describe nuclear receptors as a dynamic ensemble of conformations, where ligand binding shifts the population of these conformations, influencing receptor activity (Khan et al. 2022). This model captures the receptor's intrinsic flexibility and accommodates observations that different ligands stabilize distinct conformations with varying transcriptional outcomes.

Furthermore, the allosteric modulation of NRs is a topic of interest where novel mechanisms trigger an allosteric response at the level of the LBD or spanning across domains, to a transcriptionally active receptor. These include (non)canonical ligand binding (Cossins and Lawson 2015; Meijer et al. 2019), interactions with different proteins (Fernandez 2018) or DNA (A. K. M. Patel et al. 2023), and post-translational modifications (PTMs). One example of PTM triggered allostery is found in the RAR nuclear receptor, whose primary regulator is retinoic acid. Phosphorylation of the ligand binding domain has been shown to modulate downstream phosphorylation of the regulatory A/B domain and thus nuclear signalling (Gaillard et al. 2006; Bour et al. 2005b; Samarut et al. 2011a). Molecular dynamics simulations showed that phosphorylation of the RARγ (and RARα) receptors of this family leads to changes in the dynamic properties of the protein without producing significant conformational rearrangements (Chebaro et al. 2013; 2017). Along with these studies, the whole genome sequencing helped understand how epigenetics can dictate NR binding sites, which are not only found in gene promoter regions, but in the regions between genes, acting as enhancers, and regulate the transcription of target genes (D. X. Zhang and Glass 2013).

4.2. Transcriptional repression

The repressive conformation of nuclear receptors is characterised by specific structural features that distinguish it from the active state. The most prominent, and the most studied, structural change happens at the LBD level and involves the region carrying the activation function 2 (AF-2) region. In most cases, in the absence of ligand, the AF-2 H12 helix adopts a position different from the transcriptionally active conformation, where it is against the LBD making a hydrophobic platform with helices H3 and H4 for coactivator protein binding. Instead, the AF-2 favours interactions with corepressor proteins. In order to ensure a stable repressive configuration, the helix 12 can either adopt a conformation where it is stabilized against LBD, while creating a surface that is highly compatible with corepressors, or experience a higher degree of dynamics, resulting in movements where it can take multiple conformations, including being extended in solution. This is most often translated as the absence of electron density for this region, or a secondary structure that is stabilised by crystal contacts. One example of distinct active and repressive conformations is the retinoid X receptor alpha (RXRα) (Fig. 6).

While the detailed mechanism of molecular switching between an active and repressed form is not known in detail, and coupled with the fact that there are relatively few structures of nuclear receptors in their inactive form, crystallographic structures show a distinction between the "unligated", or "apo" form, and the “ligand-bound” or “holo” form (Torchia, Glass, and Rosenfeld 1998a). Furthermore, the mechanism of NR repression is not necessarily universal for different NRs, unlike NR ligand-dependent activation. The most straightforward mechanism, which was first discovered, is that in the presence of activating ligand, NRs are associated with coactivators, and in the absence of ligands, they are associated with corepressors (Fig. 7) (McKenna, Lanz, and O’Malley 1999; Nagy 2004). A significant characteristic of these interactions is that both corepressors and coactivators bind to overlapping surfaces of LBDs, rendering their binding mutually exclusive. This way, both states are seen as structured conformations actively repressing or activating transcription. The majority of nuclear receptors are in the cytoplasm in their unbounded form, precluded from

interactions with chromatin, while some NRs, such as the thyroid hormone receptor (TR) and retinoic acid receptor (RAR), still bind DNA in their repressive states, but their interactions with corepressor proteins actively inhibit transcription.

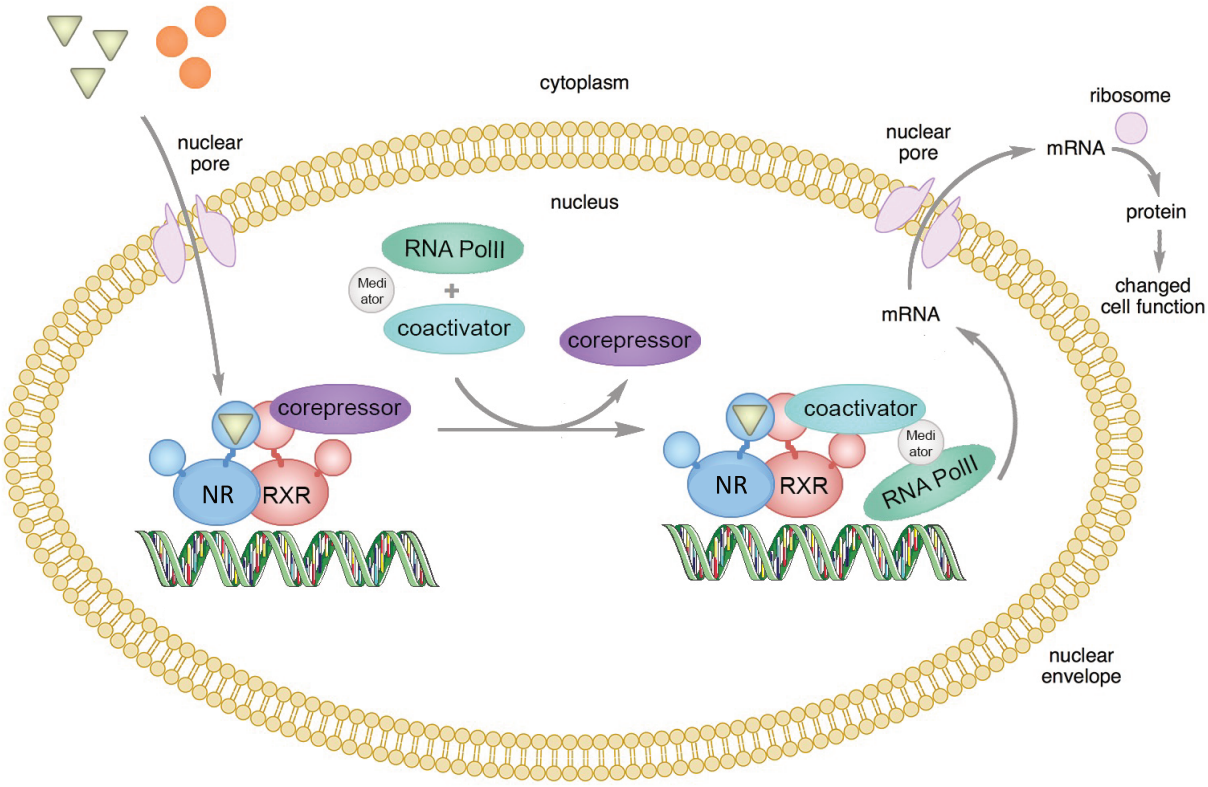


Figure 7. Schematic of nuclear receptor activation by small molecule agonist. *Adapted from Wikipedia.*

Over the years, cases of repression emerged that do not follow this pattern (Santos, Fairall, and Schwabe 2011). Mechanisms have been proposed to explain cases of gene repression by NRs in their ligand-bound form. In the case of glucocorticoid receptor (GR) and thyroid hormone receptor (TR), their binding to DNA is solely to prevent other NRs from binding and competing for their place on promoters (Subramaniam, Cairns, and Okret 1998). In addition, it has been found that coregulators can reverse roles and a corepressor complex can activate transcription, depending on promoter binding (Tagami, Park, and Jameson 1999). The need for diverse mechanisms of NR activity regulation, and especially repression, becomes

evident when considering the number and importance of genes whose transcription they modulate. More specifically, given the absence of a recognised category of endogenous antagonists, it is not unexpected that the control of repression encompasses a large variety of mechanisms (Nagy 2004). Additionally, post-translational modifications (PTMs), such as phosphorylation or ubiquitination, can modulate the stability and dynamics of both active and the repressive states (Rosenfeld, Lunyak, and Glass 2006). In conclusion, nuclear receptors exhibit conformations with varying affinity for corepressors and DNA.

4.3. Coregulator proteins

Since nuclear receptor proteins have no intrinsic enzymatic activity, they rely on other proteins to carry out their function and facilitate the transcriptional regulation of target genes. Coregulator proteins, or cofactors, are essential for enabling nuclear receptors (NRs) to modulate transcription. They significantly contribute to the stabilization of large complexes of NRs along with the basal transcription machinery at the promoter region through a series of molecular interactions (Raj Kumar, Johnson, and Thompson 2004a; Millard et al. 2013). Coregulators are large proteins that can count more than 2500 residues, and whose structures are often characterised by intrinsically disordered (ID) regions. This supports their role in flexible and dynamic interaction with different partners and participation in signalling complexes. The ID regions, which are subjected to covalent PTMs, primarily phosphorylation, acetylation, and ubiquitination, probably allow coregulators to adopt different conformations when binding to different nuclear receptors or other coregulators. Forming transient, dynamic interactions with many interactants would be an essential feature for rapid assembly and disassembly of large complexes. More than 300 coregulator proteins are known ("dkNET | dkNET Data Archive: NURSA (Nuclear Receptor Signaling Atlas)" 2024), data from 2020). Most coregulators belong to the group responsible for covalently modifying histones or transcriptional machinery through PTMs, through ATP-dependent activity, ubiquitination and SUMOylation. The second group physically bridges nuclear receptors with other essential components of the transcriptional machinery (Kishimoto et al. 2006; Lonard and O'Malley

2007; Rosenfeld, Lunyak, and Glass 2006; Torchia, Glass, and Rosenfeld 1998b). These coregulators primarily function in two distinct capacities, either as coactivators or as corepressors of the transcriptional process.

4.3.1. Coactivators

Coactivators are proteins that enhance the transcription process. They contain a characteristic leucine-rich recognition motif defined by a consensus amino acid sequence of LXXLL, which is responsible for the interaction with nuclear receptors. This motif is a part of a longer helix that carries residues implicated in a charge clamp, where two charged amino acids of the coactivator interact with two charged amino acids from helices H3 and H12 (Weikum, Liu, and Ortlund 2018a). Their enzymatic activity encompasses chromatin and nucleosome remodelling. For example, the acetylation of histones tails by histone acetylases (HATs) on sites H3K9 and H4K20, the methylation on site H3K4, and phosphorylation of linker histones. These modifications weaken the electrostatic interaction between the positively charged histone tails and the negatively charged backbone of the DNA, resulting in chromatin decondensation, which opens the way for transcriptional activation. Examples of coactivators include protein complexes such as the cAMP response element binding protein (CBP/p300), as well as the CBP association factor known as p/CAF, both of which exhibit intrinsic HAT activity. Other important coactivators encompass members of various families, including the p160s/Steroid Receptor Coactivator (SRC-1) complex, coactivator associated arginine methyltransferase (CARM), and protein arginine methyltransferase (PRMT) families. The binding of coactivator complexes to NRs allows the association of the pre-initiation complex (PIC) in the correct position and its stabilization. This complex contains RNA polymerase II general transcription factors (GTFs), necessary for transcription (J. P. Renaud and Moras 2000b).

4.3.2. Corepressors

In contrast, NR corepressor proteins, which repress transcription, are characterized by the motif LXX I/H IXXX I/L, alternatively referred to as the Cornr-box (X. Hu and Lazar 1999). The two most extensively studied nuclear receptor corepressors are the Silencing mediator of retinoid and thyroid hormone receptors (SMRT) and the Nuclear receptor corepressor (NCoR), which are large homologous proteins that function as a structural platform, facilitating the binding of multiple cofactors. These hub proteins engage in complex assembly with histone deacetylase enzymes (HDACs). The activity of HDACs closes the chromatin around nucleosomes, physically preventing gene transcription. The interaction between a nuclear receptor and a corepressor can be enhanced by the addition of an inverse agonist ligand, which is disrupted when an agonist ligand binds to the receptor, underscoring the dynamic nature of these molecular interactions. The majority of LBD crystal structures in complex with corepressor peptides have synthetic antagonists bound, in order to enhance their interaction or stabilise the H12 in a position suitable for corepressor binding (H. Zhang et al. 2011).

4.4. Dimerisation

Nuclear receptor dimerization is a critical mechanism that enhances the transcriptional regulation of genes. NRs can assemble in homo- or hetero-dimers. The receptor RXR plays an important role since it can both homodimerize and act as a common dimerisation partner for other NRs. Dimerisation at the LBD level essentially helps stabilize the dimer, while dimerization at the DBD helps DNA specific sequence recognition. LBD dimerisation interfaces have been mainly determined through crystallographic studies, showing that the topologically conserved dimerization surface is comprised of helices H7, H9 and H10-11, including the loops H8-9 and L9-10 (Germain and Bourguet 2013). The sequence specificity determines the binding specificity of different receptors. The group of steroid receptors (AR, PR, GR, and MR) represents an exception to this established dimerization interface, but

bioinformatic tools can help decipher their alternative homodimer binding surfaces (Bianchetti et al. 2018).

5. Peroxisome proliferator - activated receptors (PPARs)

Peroxisome proliferator - activated receptors, commonly referred to as PPARs belong to subgroup 1 of the NR superfamily, along with receptors TR (thyroid hormone receptors), RAR (retinoic acid receptors), REV-ERB (reverse-Erb), ROR (retinoic acid related orphan receptor), FXR (farnesoid X receptor), LXR (liver X receptor) and VDR (vitamin D receptor) (Weikum, Liu, and Ortlund 2018b). Within the PPAR subfamily, there exist three distinct isotypes (subtypes): PPARα (NR1C1), PPARβ/δ (NR1C2) and PPARγ (NR1C3), encoded by different genes. Even though they play a central role in lipid metabolism, they are named after their ability to bind peroxisome proliferators, in the first study from 1990 (Issemann and Green 1990). Peroxisome proliferators (PPs) are a class of structurally diverse endogenous substances and exogenous chemicals that increase the number and size of peroxisomes (M. Jiang and Yang 2014). Exogenic PPs are proven to have cancerogenic effect in rodent models and are used for different purposes, such as herbicides, plasticizers, and industrial solvents (Abdelmalak, Yang, and Ray 2024). Peroxisomes are single - membrane cytoplasmic organelles, which assure several metabolic functions such as β-oxidation of fatty acids, ether phospholipid and bile acid synthesis, and glyoxylate detoxification (Wanders et al. 2023).

Due to the considerable size of their ligand binding pockets, PPARs can accommodate a large number of various ligands. The endogenous ligands of PPARs are long chain saturated, mono- (MUFA) and polyunsaturated (PUFA) fatty acids (FA), as well as and their lipophilic derivatives, called eicosanoids, highlighting the connection between these receptors and lipid metabolism (Gervois et al. 2000). Three PPAR isotypes differ in tissue expression, together controlling the expression of genes involved in lipid and glucose metabolism, in development, and inflammatory response (Berger and Moller 2002). PPARα is primarily expressed in tissues having high FA oxidation rate, such as liver, heart, skeletal muscle, white adipose tissue and kidney. It promotes FA uptake in these tissues, through mitochondrial and peroxisomal β-

oxidation, producing acetyl-CoA and ATP the latter affecting the indirect modulation of glucose metabolism. High-affinity PPARα agonists, called fibrates have been used for treating dyslipidaemia (unhealthy levels of one or more kinds of lipid in blood). The PPARβ/δ is the least studied isotype, found predominantly in skeletal and cardiac muscles, where it controls the metabolic switch from glucose to FA utilization, and decreases lipid accumulation. Trials on treatments for dyslipidaemia and type 2 diabetes mellitus targeting PPARβ/δ have been discontinued due to safety concerns, warranting further studies (Lamas Bervejillo and Ferreira 2019).

5.1. Peroxisome proliferator - activated receptor gamma (PPARγ)

Peroxisome proliferator - activated receptor gamma (PPARγ) was initially discovered based on its similarity to PPARα. PPARγ protein is coded by the *PPARG* gene, found on chromosome 3. It has two separate promoters and different 5' exons resulting in at least four mRNAs. The isoform PPARγ2, which has 505 residues, is considered to be the full-length protein. It has molecular mass of 57.62 kDa (UniProt 2024). PPARγ is primarily expressed in brown and white adipose tissue where it plays a crucial role in regulating adipogenesis and glucose metabolism by promoting the differentiation of pre-adipocytes into mature adipocytes. It stimulates glucose uptake by regulating the secretion of adipocytokines - the mediators of insulin action (Janani and Ranjitha Kumari 2015). PPARγ is present in macrophages, dendritic- and T-cells, where it acts as an immune-modulator, specifically as a repressor of inflammation. Furthermore, it has a dual role in cancer - it can act as a tumor suppressor and as an initiator, depending on cancer type (Hernandez-Quiles, Broekema, and Kalkhoven 2021a). Therefore, PPARγ represents a target for treatment of type 2 diabetes mellitus, cancer, inflammation and hypertension (Berger and Moller 2002).

With a ligand binding pocket of about 1300 $Å^3$, PPARγ can accommodate a variety of lipophilic ligands - polyunsaturated fatty acids (PUFAs) and their oxidised derivatives called eicosanoids. Fatty acids, such as linoleic, docosahexaenoic, and eicosatetraenoic acids, bind

PPARγ at μM levels. PPARγ governs hormone- and nutrient-mediated responses, by accommodating different ligands: eicosanoids (ex. prostalglandins), nitrated fatty acids, flavonoids, nutrients such as glutamine and arginine (Marion-Letellier, Savoye, and Ghosh 2016a). Man-made compounds with an agonistic effect on PPARγ are phtalates and bisphenols. PPARγ is a cognate receptor for thiazolidinediones (such as rosiglitazone and pioglitazone), which are a class of anti-hyperglycaemic drugs developed for the treatment of type 2 diabetes. However, their use has been challenged due to serious side - effects (Dubois et al. 2020).

PPARγ is localized both in the cytoplasm and in the cell nucleus, where it exerts its function of nuclear repressor. Nuclear - cytoplasmic shuttling of PPARγ in the cell is largely determined by ligand concentration (Umemoto and Fujiki 2012). In absence of ligand, PPARγ interacts with corepressor complexes, such as NCoR or SMRT, which recruit chromatin-modifying enzymes (HDACs) and actively repress transcription. After binding an agonist ligand, PPARγ forms a heterodimer with the receptor retinoid X receptor (RXR)(Fig. 8). The heterodimer can recruit coactivators such as PPARγ coactivator 1-α (PGC-1α) or E1A binding protein p300 (EP300), which carry HAT enzymatic activity. DNA binding occurs on the peroxisome proliferator response element (PPRE) gene promoter, which are usually direct hexamer repeats spaced by 1 (DR1) or 2 (DR2) nucleotides, leading to regulation of gene transcription (Schoonjans, Staels, and Auwerx 1996). Binding profiles have indicated that PPARγ binds to thousands of PPRE sites in the genome (many of which are located far from proximal promoters) and that the PPARγ binding is cell-dependent (binding differs between cell types and adipocytes from different anatomical locations) (Nielsen et al. 2008; Siersbæk et al. 2012). Additional level of PPARγ regulation of activity is at the epigenetic level.
Covalent post-translational modifications, such as phosphorylation or SUMOylation can modulate the activity of PPARγ. One particular PTM, the over-phosphorylation of Ser273 by cyclin-dependent kinase 5 (CDK5) leads to suppression of genes that promote insulin sensitivity (Choi et al. 2010). This turn of events results in insulin resistance, increasing the risk of type 2 diabetes and cardiovascular diseases.

The PPARγ ligand binding domain has the canonical NR structure, composed of 12 α helices, one 2 or 3-strand β sheet, and a small additional α helix H2` (Fig. 9). To date, there are over 300 resolved structures in the PDB, of the ligand binding domain (residues 230 - 505), with or without RXR receptor, and with the resolution as high as 1.42 Å. In addition, we find 3 crystallographic structures of the full-length receptor (residues 102 - 5005 minus the N-terminal domain). Most of the LBD structures depict the active state of the protein, in the presence of natural or synthetic ligands, or through the stabilisation of the H12 by crystal contacts. In some of these structures, the omega loop structure is not resolved, indicating its flexibility. The inactive conformation of the PPARγ remains controversial. One structure of PPARγ LBD bound to an antagonist and SMRT corepressor peptide does not have H12 electron density (PDB ID: 7SQA). The structure which binds covalent synthetic inverse agonist (PDB ID: 6ONI) can be seen as having the helix H12 inside the ligand binding pocket. We will not discuss the physiological importance of such structures, but we will note that the true "apo" form of PPARγ LBD, which would represent an inactive conformation, are scarce (PDB IDs: 7WOX chain B, and 2PRG chain B). The structures of “apo” PPARγ in the Protein Data Bank generally have H12 stabilised in a position that is not considered transcriptionally active (Fig. 9). Concerning the full-length repressive conformation of PPARγ, studies showed that the AF-1 of N-terminal domain inhibits PPARγ activity, through the (MAP) kinase-mediated phosphorylation (E. Hu et al. 1996), and a recent study (from BioRxiv) found that this inhibition happens through the interaction of NTD with the β sheet and H12 of the LBD, which could compete for coactivator binding (Mosure et al. 2024).

The conformational dynamics of PPARγ was first studied using NMR experiments, which showed that more than half of the pics were missing in "apo" structure compared to an agonist - bound form. This further suggested that adding the agonist (here rosiglitazone) led to the stabilisation of the domain, and that the domain activation of this is a result of a population shift of a dynamic ensemble of conformations, rather than a switch from an inactive to an active conformation (of the helix H12) (Bruce A. Johnson et al. 2000). Further studies of hydrogen-deuterium exchange (HDX) coupled mass-spectrometry (MS), NMR and fluorescence

anisotropy, revealed high flexibility of H12, the last portion of helix H11and the loop H11-H12, which were stabilized solely in response to ligand binding (Bruning et al. 2007; Hughes et al. 2012b). Sampling of PPARγ apo state was also done using molecular dynamics simulations. Studies by Fratev and colleagues (Fratev 2016; Fratev et al. 2015) by accelerated MD (aMD) and metadynamics (metaD), of ~12μs, indicate that the C-terminal H12 can adopt an antagonist conformation in mainly 2 clusters, starting from active conformations, and one cluster shows H12 conformation similar to the one found in two 'apo' crystallographic structures. Another combined experimental - MD study by (Chrisman et al. 2018a) showed the that helix 12 and the coregulator-binding surface are a dynamic structural ensemble, and reveal several clusters of H12 conformational state, in apo and corepressor - bound form, among others.

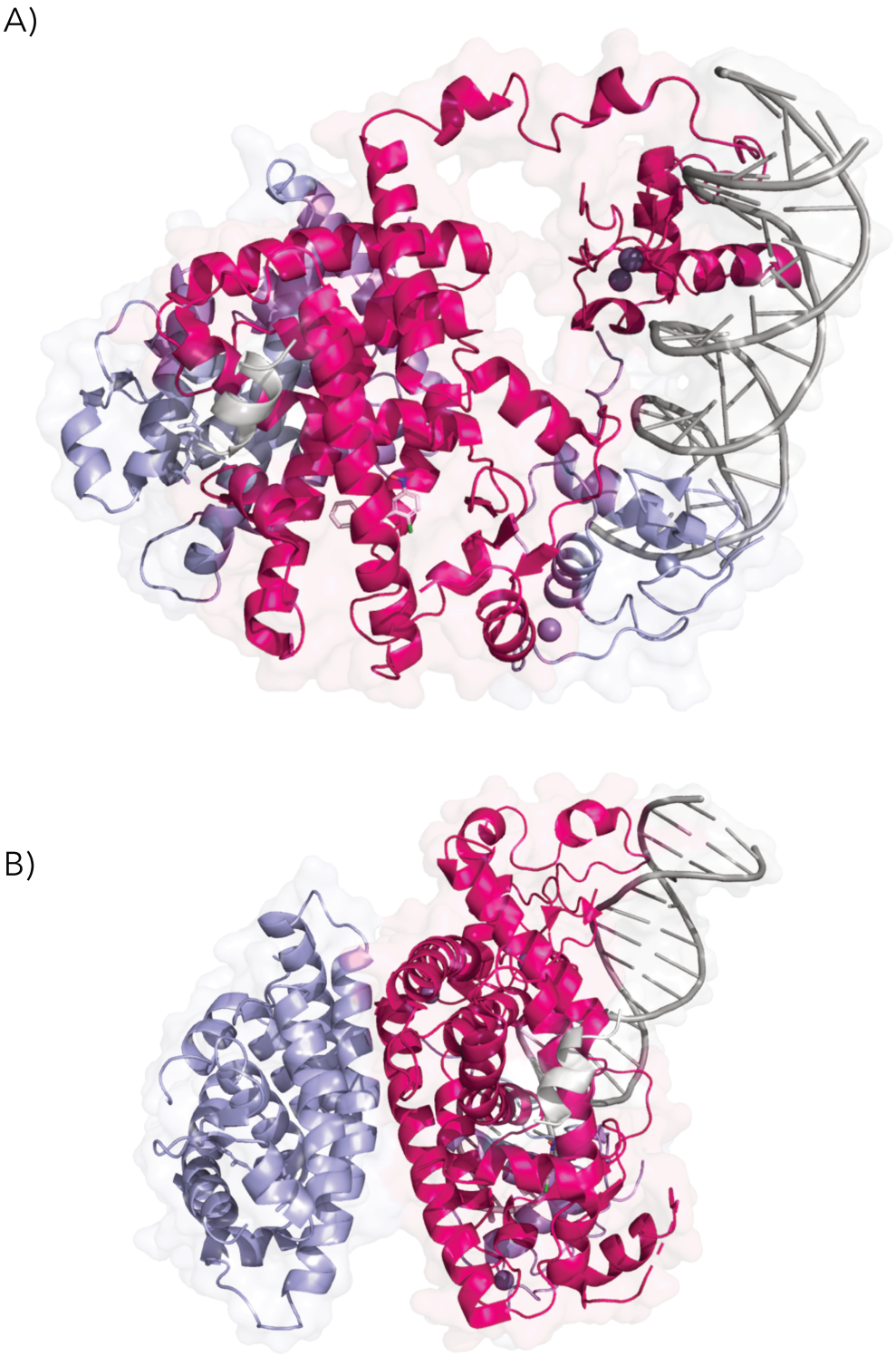


Figure 8. Crystallographic structure of the full-length PPARgamma and RXRalpha heterodimer bound to DNA DR1. A) PPARg bounds roziglitazone, and (RXRa 9-cis)-retinoic acid. PPARg in pink, RXRa in violet, DNA in gray. Ligand and coactivator peptide for PPARg are in white. B) Dimerization interface between PPARg and RXRa. Structure PDB ID: 1FM9.

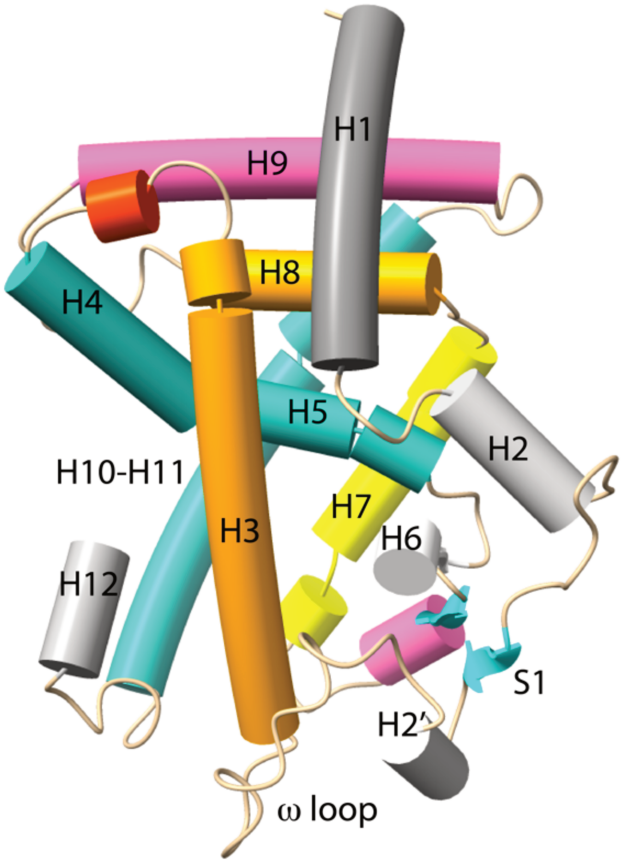


Figure 9. Schematic representation of the PPARgamma ligand binding domain. PPARs have an additional H2' helix. PDB ID: 7WOX.

6. Objectives of the thesis

6.1 Experimental and Computational study of protein dynamics

Advances in both experimental techniques and computational methods have been contributing to an evolution in the thought about the relation between protein structure and function. The traditional view of protein function is strongly associated with specific structures presumably at the global energy minimum. But methods such as NMR, SAXS measurements and molecular dynamics simulations force the reflection away from single structures to the role of structural ensembles in biological function. Changes in protein conformational ensembles are implicated

in many different physiological processes, as well as disease (Tompa 2016). For example, recurrent mutations of PPARγ were found to activate the PPARγ/RXRα pathway in luminal bladder cancer (Natacha Rochel et al. 2019). Mutations were found throughout the protein - including N-terminal, DNA-binding and ligand-binding domains, and most of them enhance protein activity. Structure-function and molecular dynamics studies of some of PPARγ variants with mutations in the ligand-binding domain allowed for the identification of structural dynamic elements that underpin the gain-of-function of PPARG mutants that lead to pro-tumorigenic PPARγ/RXRα pathway activation in luminal bladder tumours.

The atomic level understanding of molecular mechanisms of biological activity, both physiological and pathological, has been largely dominated by the structural analysis of implicated biomolecules. However, there is a growing realization that understanding can be significantly improved by including consideration of the underlying structural dynamics, particularly those of collective motions (Ponzoni and Bahar 2018; Bahar et al. 2015; Seo et al. 2014; Shukla, Shukla, and Tripathi 2018). These trends underscore a shift towards a more dynamic understanding of biological structures, emphasizing the importance of movement, flexibility, and temporal changes in biomolecular function. Advances in technology and computational methods are key drivers of these developments, providing unprecedented detail and insight into the complexity of life at the molecular level.

It remains a major scientific challenge to quantify changes in structural dynamics and understand how these changes are coupled to different physiological mechanisms. The most used method for studying structural dynamics in proteins is solution NMR (Salvi, Abyzov, and Blackledge 2017; Walinda, Morimoto, and Sugase 2018). However, this approach is often limited to proteins under 50 kDa, although new labelling techniques have been pushing the size limit upwards. Spectroscopic techniques, such as far-IR and terahertz (THz) spectroscopies (both time dependent and absorption) are emerging as attractive for studying biophysical processes (Khoury and Hellwig 2017), but their application has remained limited to isolated proteins and peptides, as discussed in the next section.

The objective of this thesis is to develop new approaches to measure physical properties directly related to changes in low frequency collective structural dynamics of proteins. Toward

this end, we plan to develop an integrated far infrared absorption spectroscopy – molecular dynamics simulation approach. Far infrared absorption spectroscopies probe the low frequency region of the vibrational spectrum and reveal the collective vibrational modes in the spectral region from 0.06 THz (2 cm^{-1}) to 10 THz (333 cm^{-1}), a frequency region that is accessible to vibrational analysis based on molecular mechanics and dynamics calculations. Our hypothesis is that spectroscopic techniques such as far-IR absorption spectroscopy, when combined with molecular modelling and structural and biophysical analysis, can be used to advance our understanding of the structural dynamic response of proteins to ligand binding, even when the structural changes are minimal or non-existent (Cooper and Dryden 1984b). Not limited by protein size, the establishment and integration of far-IR spectroscopy with other biophysical approaches will provide an innovative means to quantify a low-frequency vibrational fingerprint of the protein as well as changes to this fingerprint as a function of ligand binding and complexation. We will be able to study a much wider range of proteins and their complexes than current approaches (i.e. NMR) permit.

This project is founded upon work studying ligand binding and allostery in the MAGI1 PDZ domain (Cote et al. 2017a). In that work, far-IR absorption spectroscopy and molecular dynamics simulations were combined to study the structural dynamic response of this PDZ domain to the binding of a small peptide ligand. These first results demonstrated the potential of combining far-IR experiments and molecular dynamics (MD) simulations for the study of ligand binding by proteins. We believe that, through the development of far-IR/modelling approaches, we will be able to uniquely characterize low-frequency motions and exploit this information in novel ways to target proteins which can be considered in drug development projects. The first work also made clear certain developments were in order, including moving towards more complex systems and the consideration of ensembles of conformations rather than single conformations and improvements in simulation technology for interpreting spectroscopic data.

Concerning the computational approach, normal mode (NM) analysis is a technique that continues to contribute to new technologies. NMs explore the collective motions within a molecule by examining its vibrational modes. These modes represent the natural, low-

frequency movements that the molecule can undergo without significant changes in its structure.

Some of the notable techniques where normal mode analysis has been used includes cryo-electron microscopy (cryo-EM). Cryo-EM has become a critical tool for determining the structures of large macromolecular complexes at near-atomic resolutions and more recently has been providing insights in the dynamics of biological assemblies through the elucidation of the subject molecules in different conformations. An essential computational tool underpinning these applications is normal mode analysis (Harastani et al. 2022; Grudinin, Laine, and Hoffmann 2020). Single molecule fluorescence, smFRET, enables the study of dynamic processes at the single-molecule level, providing detailed information about conformational changes and interactions in real-time. Normal mode analysis has been employed in the interpretation of the experiments (Gabba et al. 2014).

Normal modes can play a significant role when they are employed in conjunction with molecular dynamics simulations, for example being an integral part of enhanced sampling algorithms that improve the exploration of conformational space. One such approach is the Molecular Dynamics with Excited Normal Modes (MDeNM) method (Costa et al. 2023). MDeNM is an enhanced sampling method that combines molecular dynamics simulations with normal mode analysis (NMA) to accelerate sampling. It uses a few low-frequency normal modes to guide the dynamics and has been shown to explore protein conformational space more effectively than standard MD. With these methods, one is able to study large scale conformational changes, such as domain movements, binding events, allosteric regulation, or folding processes.

In the next chapter, we present the methods used in the course of this thesis, followed by chapter III, which presents our ensemble approach to calculating properties from normal mode analysis. Chapter III serves as a prelude to the applications presented in the subsequent chapters. Chapter IV presents the results obtained in a large study of PPARγ in apo WT, holo WT and two mutant forms, using the combined far-IR/molecular dynamics approach. The objective was to investigate the effects of different perturbations to the underlying low frequency collective motions, which are often attributed to important physiological function.

Chapter V presents results for a theoretical study on the effects of polarization on the collective motions of proteins. Polarization was introduced via the Drude polarizable force field (Lopes, Huang, Shim, Luo, Li, Roux, and Mackerell 2013). The final chapter presents conclusions and perspectives.

Chapter II - Methodology

1. Molecular modelling

According to A. Leach. « molecular modelling represents a simplified description of a system or a process (...) devised to facilitate calculations and prediction of behaviour of molecules an molecular systems. » (Leach 2001a). A system is the portion or the subset of the physical world that we are modelling. It is described by its constituents and the interactions between them (Haile 1997). In order to observe, and further analyse them, we must first model them - meaning we have to assign numerical values to the constituents of our system. The art of modelling lies in the ability to make a system simpler than reality, to exclude the constituents that have little to no impact on the results. The system's behaviour will remain consistent with physical reality, in the scope of a restricted set of input conditions. One set of methods used for describing properties of molecular systems is molecular mechanics (Allinger 1982).

2. Molecular mechanics

Molecular mechanics (MM) is a computational method which describes molecular systems using classical mechanics principles. Its advantage lies in the simplified and, therefore, less computationally expensive, calculations, compared to quantum mechanics (QM). Despite the advancement in the computing capacities, it is still largely used for quick calculations, applications to large systems and for simulating longer molecular dynamics. In all-atom MM, atoms are represented as *particles* (no separation of nuclei and electrons), connected to other atoms via *springs* (Hooke's law), which undergo vibrational motion. The specific parameters describing atoms and bonds are derived from experimental data, empirical or ab initio (quantum mechanical) calculations, and will be described in the next paragraph. Molecular mechanics calculations are used, primarily, for calculating the potential energy of molecular systems, energy minimization, molecular dynamics simulations, or even in molecular docking studies. We will describe it in terms of potential energy, energy minimization and normal mode analysis.

2.1. Potential energy function: All - Atom Additive Force Field

The basis of molecular mechanics that characterizes the dynamical behaviour of molecules of biological interest is the potential energy surface. The potential energy surface (PES) represents the energy of a system as a function of atomic positions, i.e. how the energy changes for different molecular conformations. The first information we get from the PES are relative stabilities of different conformations of our system. The most accurate PES of a molecular system can be obtained by quantum calculations. However, historically, the capacities for doing these complex calculations were limited by computer power, so the first models that would represent the PES were created as a way of simplifying and approximating the systems that were studied. The model that is applied for calculating the potential surface is given in the form of potential energy function (U), also called *force field* (FF). A force field is a sum of individual bonded and non – bonded terms, as a function of atomic coordinates, describing inter- and intramolecular interactions of the atoms in a system (Eq.1).

$$U = E_{bonded} + E_{non-bonded} \quad (1)$$

Since the development of the approach, different force fields have been developed with some common features such as the use of harmonic potentials with a force constant that restores the equilibrium positions for bonds and valence angles (an angle between three atoms). Here, we will consider the all atom force field developed for the CHARMM program, called ALLATOM36 (A. D. MacKerell, Jr, et al. 1998), in its integral form (Eq. 2):

$$U = \sum_{bonds} k_r\,(r - r_0)^2 + \sum_{angles} k_\theta\,(\theta - \theta_0)^2 + \sum_{Urey-Bradley} k_{UB}\,(S - S_0)^2 + \sum_{dihedrals} k_\phi\left(1 + cos(n\phi - \delta)\right) + \sum_{impropers} k_\omega\,(\omega - \omega_0)^2 + \sum_{residues} U_{CMAP}\,(\varphi, \psi) + \sum_{non-bonded\ pairs} \left\{4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon r_{ij}}\right\} \quad (2)$$

2.1.1. Bonded terms

The reference to bonded terms implicates groups of atoms separated by one, two or three bonds (Fig. 10, Eq. 3).

$$U = \sum_{bonds} k_r\,(r - r_0)^2 + \sum_{angles} k_\theta\,(\theta - \theta_0)^2 + \sum_{Urey-Bradley} k_{UB}\,(S - S_0)^2 + \sum_{dihedrals} k_\phi\left(1 + cos(n\phi - \delta)\right) + \sum_{impropers} k_\omega\,(\omega - \omega_0)^2 + \sum_{residues} U_{CMAP}\,(\varphi, \psi) \quad (3)$$

The terms modelled by a harmonic potential include atomic bonds (r), valence angles (θ), Urey-Bradley (S), and improper angle (ω), torsional dihedral angles (ϕ) are modelled by a periodic function, and the torsional correction for backbone atoms (CMAP φ,ψ) is an energy contribution rather than a continuous function. Variables with subscript represent the respective equilibrium values. Each term has a force constant associated: k_r, k_θ, k_{UB}, k_ϕ, *and* k_ω. In this way, the energy of individual terms increases with the deviation of the values from their respective equilibriums. The only term not represented by a harmonic potential is the dihedral angle, which has sinusoidal expression, where n represents the periodicity of the dihedral angle, and δ is the phase shift. The improper angle is maintaining the planarity of atoms, and it is described as an angle between an atom an imaginary plane composed of three adjacent atoms. For atoms 1-2-3, the Urey – Bradley term describes quadratic function of the distance between

atoms 1 and 3, separated by two covalent bonds. The protein main chain has an additional term called CMAP correction, which keeps the backbone dihedral values of φ and ψ consistent with the allowed values from Ramachandran plots. The values for these terms were determined using structural and vibrational data of model compounds, and further refined by iterative process of adjusting according to interaction parameters, the structures, vibrational spectra, and energy surfaces of the model compounds (A. D. Jr. MacKerell et al. 1998).

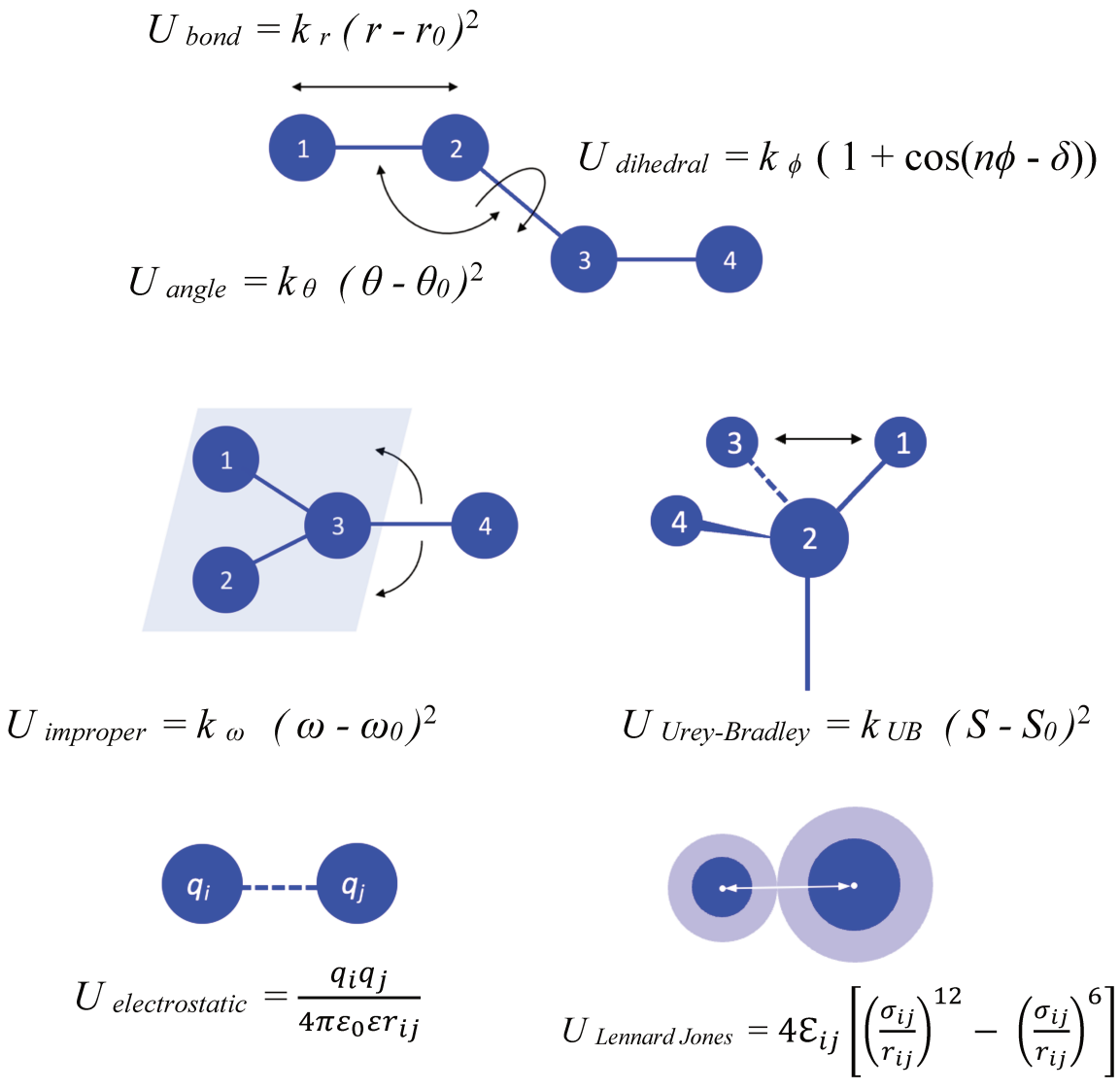


Figure 10. Different contributions of the potential energy function. The atoms are represented as balls and bonds as springs. The bonds are approximated by a harmonic oscillator (all except the dihedral bond), but for the sake of simplicity, they are represented as straight lines.

2.1.2. Non - bonded terms

Non-bonded terms include two terms, the Coulombic interactions between the point charges (q_i and q_j) and the Lennard-Jones (LJ) 6-12 potential term, which defines the repulsion and the attractive van der Waals dispersion (Fig. 10, Eq. 4).

$$U = \sum_{non-bonded\ pairs} \underbrace{\left\{4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]\right.}_{Lennard\ Jones\ potential} + \underbrace{\left.\frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon r_{ij}}\right\}}_{electrostatic\ term} \qquad (4)$$

The non-bonded calculations are calculated over all atom pairs, within a set cut-off distance, except for the covalently bound atoms (1-2 interactions), and pairs separated by two bonds (1-3 interactions). The electrostatic interaction can be attractive or repulsive, depending on the charges q_i and q_j. It approaches 0 when the r_{ij}, the interatomic distance, increases. The term ε is the relative dielectric constant (set to 1 in explicit solvent simulations), and ε_0 is the permittivity of vacuum. The Lennard – Jones term is also an expression of interaction energy as a function of interatomic distance and has both an attractive interaction (r^6) and a repulsive interaction (r^{12}). The ε_{ij} is the depth of the energy well, and r_{ij} is the distance between atoms *i* and *j*. The σ_{ij} represents the atomic distance r_{ij} where the LJ value is zero.

2.2. Potential energy function: All - Atom Polarizable Force Field

The accuracy of force fields governs the quality of the results of describing a molecular system and running simulations. The additive FF has a number of implicit approximations in its formulation and any inaccuracies do not necessarily stem from insufficient parameters, but they rather highlight the inherent limitation of their functional form. For that reason, there has been a considerable effort in improving the functional form of empirical energy FFs. One such improvement was introduced was to try and account for polarization, which is missing from the standard additive force field because the partial charges are fixed. One formulation for

including polarization is in the form of a Drude polarizable force field, which is built upon the additive FF previously described. This polarizable force field introduces a term based on Drude oscillator, or "charge-on-springs", and accounts for electronic polarization, which is the way atomic or molecular dipoles adjust in response to their surroundings (Lopes, Huang, Shim, Luo, Li, Roux, and MacKerell 2013; Vanommeslaeghe and MacKerell 2015a). In addition, the Drude FF includes virtual "lone pair" particles that allow a better representation of hydrogen bond acceptors (Fig. 11).

2.2.1. Atomic polarizability

2.2.1.1. Drude particle

In this FF, there is a "Drude particle" carrying negative charge $q_{D,i,}$ connected to the core of the polarizable atom, *i,* via harmonic potential (or a spring) with force constant k_D. The following term is added to the classical CHARMM36m force field formulation:

$$U_{drude} = \frac{1}{4\pi D}\left(\sum_{i<j}\frac{q_{D,i}q_j}{\|r_{D,i}-r_j\|}+\sum_{i<j}\frac{q_{D,i}q_{D,j}}{\|r_{D,i}-r_{D,j}\|}\right)+\frac{1}{2}\sum_{\substack{polarizable\\atoms}}k_D\left\|r_{D,i}-r_i\right\|^2 \quad (5)$$

The position $r_{D,i}$ of the Drude particle is able to move freely depending on the electrostatic potential environment, emulating the deformation of the atom's electron cloud in response to

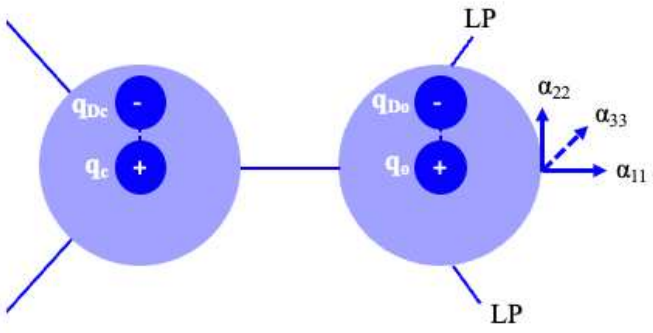


Figure 11. Representation of the Drude oscillator model. Two atoms of carbon and oxygen are represented in light blue, and their charges (qc and qo), in addition to their respective Drude particles qDc and qDo are represented. 'LP' - lone pair. Three arrows and values α_{11}, α_{22}, and α_{33} are the tensor components of the anisotropic polarizability along three axes.

that environment. The isotropic atomic polarizability, α (Å^3), is a constant that depends on how easily its electron cloud can be distorted by an external field. The isotropic atomic polarizability of a given atom is achieved by distributing the atomic charge q between the core atom (q_A) and its Drude particle (q_D), such that the $q = q_A + q_D$. The atomic polarizability is calculated through the following expression:

$$\alpha = \frac{q_D^2}{k_D} \tag{6}$$

The value of the force constant k_D is set to a constant value of 1000 kcal.mol^{-1}.Å^{-2} for all Drude–atom pairs. The atomic dipole moment $\boldsymbol{\mu}$, in response to the electric field $\boldsymbol{E}$, can be calculated using:

$$\boldsymbol{\mu} = \frac{q_D^2 \boldsymbol{E}}{k_D} \tag{7}$$

The first generation Drude FF had a problem of unrealistic over polarization, which arises when two atoms are physically close and the induced dipole moment being a linear function of the electric field. This can cause the Drude particle of one atom to become "trapped" in the potential well of a nearby atom. Furthermore, the effect on polarization is greater along the axis formed by two adjacent atoms than along the perpendicular axis. In order to suppress these effects, a specific constraint was introduced to the first generation Drude potential energy function. This was the "HardWall' modification, which places a "hard limit" on how close the Drude particles can approach, thus preventing the distance between charges from going to zero and having infinite electrostatic forces that would therefore arise. If a Drude particle moves beyond the HardWall limit set to 0.2 Å, the relative velocity of the particle is scaled down, even though it still continues to move outwards. The velocity of the core atom is adjusted to maintain overall momentum during MD simulation integration (Vanommeslaeghe and MacKerell 2015b).

A second adjustment of the Drude FF is the implementation of the explicit dipole - dipole interaction for atoms within three bonds. Non-bonded interactions between 1-2 (one bond) and 1-3 (two bonds) atom pairs were only implicitly represented by bond-stretching and

angle-bending terms in the additive FF. The interactions of atom pairs 1- 4 and beyond are represented by classical Coulomb interactions. The interactions of the Drude oscillators, i.e. no core atom - core atom or core atom - Drude for 1-2, 1-3, corresponding to 1−2 and 1−3 pairs are represented through scaling of electrostatic terms by a Thole-like screening function S_{ij}, which reduces dipole-dipole interactions as distances decrease (Thole 1981). The function S_{ij} depends on the distance r_{ij} between Drude particles i and j in the following manner:

$$S_{ij}\left(r_{ij}\right) = 1 - \left(1 + \frac{\left(t_i + t_j\right) r_{ij}}{2\left(\alpha_i \alpha_j\right)^{1/6}}\right) e^{\frac{-\left(t_i + t_j\right) r_{ij}}{\left(\alpha_i \alpha_j\right)^{1/6}}} \quad (8)$$

where r_{ij} is the distance between atoms i and j, α_i *and* α_j are respective atomic polarizabilities. The atomistic Thole factor, t_i (or t_j), is a damping constant associated with atom i (j) and is defined individually for each atom.

2.2.1.2. Anisotropic polarizability

To include anisotropy (until now it was all isotropy), the Drude model includes two more extensions that further improve the representation of hydrogen bond acceptors. One concerns the anisotropy of the charge distribution and the second the polarizability. First term, the anisotropy of the charge distribution, describes the uneven spatial distribution of charges in a molecule. It is represented by lone pairs (LP), massless, negatively charged point charges that are exclusively added to electronegative atoms. Essentially, LPs represent virtual sites carrying negative charge. (Harder et al. 2006)

The second term accounts for anisotropy of the polarizability, which is the variation in how easily the electron cloud can distort in different directions in response to an external electric field. To explain anisotropy, recall that isotropic polarizability is described by a harmonic self-polarization term, U_{self}, that is calculated on the basis of the harmonic bond energy between the Drude oscillators and their core atoms. While the U_{self} term was represented by a displacement $\boldsymbol{d}$, and an isotropic scalar force constant k_D,

$$U_{self}^{isotropic}(\boldsymbol{d}) = \frac{1}{2} k_D \boldsymbol{d}^2 \tag{9}$$

anisotropic polarizability, which varies with direction, is represented by a tensor (matrix), K_D, with off-diagonal elements set to zero. (Fig 11., tensor components)

$$U_{self}^{anisotropic} = \frac{1}{2}\left(\left[\boldsymbol{K}_{11}^{(D)}\right]{\alpha_1}^2 + \left[\boldsymbol{K}_{22}^{(D)}\right]{\alpha_2}^2 + \left[\boldsymbol{K}_{33}^{(D)}\right]{\alpha_3}^2\right) \tag{10}$$

The combination of lone pairs and anisotropic polarization improves the accuracy of describing hydrogen bonding.

2.3. Long - range interactions

The long-range non-bonded interactions are the most computationally demanding, since their number scales with N^2, as opposed to the number of bonded interactions (involving bonds, angles, and dihedrals) which is proportional to the number of atoms (N). To accelerate these calculations, we generally disregard interactions between atoms separated by the distance greater than a predefined *cutoff* distance. Two most common methods used for implementing the cutoff in the calculations of energy, while maintaining energy smoothness, are called shift and switch truncation functions which, by their algorithmic construction « turn off » the interaction in a continuous way so that the interaction is zero at the cutoff distance (Bernard R. Brooks et al. 1983).

The difficulty here is that terminating interactions at a certain cutoff distance results in the neglect of long-range interactions arise from electrostatics, in particular. Numerous studies have shown the importance of long-range electrostatic interactions in biomolecular systems (Zuegg and Gready 1999; Ahsan, Pindi, and Senapati 2020).

The inclusion of the long-range electrostatic interactions in molecular dynamics simulations employs, for the most part, the use of an algorithm based on the Ewald summation method. The most currently used implementation is called Particle Mesh Ewald (York,

Darden, and Pedersen 1993). The method splits the electrostatic interaction into two parts, a short-range contribution and a long-range contribution. For this, we use the implementation of the Particle Mesh algorithm.

2.4. Solvent representation

For the sake of a better representation of biological systems, it is preferable to study protein systems in aqueous solution, instead of using the gas-phase simulations. The representation of water molecules is of great importance, since it plays an important role in the structure and dynamics of biological molecules. Two common ways of modelling solvent are explicit solvent and implicit solvent models. In our simulations, we use the explicit solvent, which represents water molecules as individual particles.

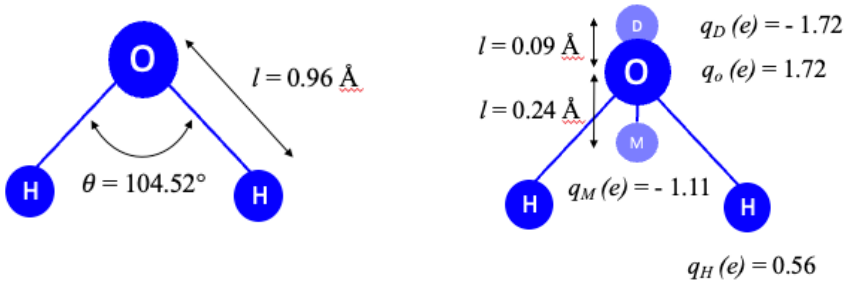


Figure 12. Explicit solvent models: TIP3P and SWM4 NDP. Both models have the same geometry. TIP3P is the model on the left, and SWMP4 NDP (Drude model) is on the right. The Drude model has 2 additional terms: the Drude particle (D), and the 'M' site (M).

In the additive force field simulations, we are using the TIP3P ("Transferable Intermolecular Potential functions" 3-point) water model, meaning it represents water molecules using three interaction sites corresponding to the positions of the oxygen atom and the two hydrogen atoms (Fig 3)(Jorgensen 1981). The corresponding Drude model for water molecules is the 4-point

SWM4 - NDP ("simple water model with negative Drude polarization") (Lamoureux et al. 2006), which is essentially a rigid configuration of HOH particles, with two additional elements. Firstly, the model includes a negatively charged 'M site' of the oxygen, placed along the axis of symmetry of the molecule. Secondly, the Drude particle is attached to the oxygen atom, carrying a charge q_D = -1.72, that is equal in magnitude, but opposite in sign, to the charge of the oxygen atom (q_O = 1.72). The total charge of the oxygen atom (between oxygen and its Drude particle) is zero, which is an important aspect of the model's design, and together these two terms contribute to the accurate representation of dipole and quadrupole moments of water (Lemkul, Huang, Roux, and Alexander D. MacKerell 2016). The polarizability is represented by the Drude particle, and the permanent dipole by the M-site charge (q_M = -1.11) and 2 hydrogen partial charges (2 x q_H = 0.56).

2.5. Energy minimization

Energy minimization is a method employed to find a minimum-energy conformation of a molecular system. Energy minimization is used to optimize the molecular coordinates, or geometry. For most applications, it is used to relieve strain in conformations that are generally present in experimental structures, resolving structural conflicts like bond angles or torsion strain and steric clashes. Considering the complexity of a potential energy surface, finding a minimum-energy structure is not an easy task. The potential energy surface describes the way potential energy varies with respect to coordinates. The potential energy is a multidimensional function of atomic coordinates; *3N - 6* internal or *3N* Cartesian coordinates for an *N* atom system. The nonlinear nature of the potential energy function results in potential energy surface with multiple *minima*, most that are called local minima and one global energy minimum, characterized by having the lowest energy of the system. Energy minimization algorithms can determine local minima. These algorithms are in general characterized as first-derivative algorithms (i.e., algorithms that use only the first derivative of the potential), second derivative algorithms (i.e., algorithms that use both the first and second derivative of the potential). The first derivative of the potential energy function (the gradient) describes the rate of change of

energy with respect to atomic positions, or the slope. At both the global minimum and all local minima, as well as at any other stationary points such as saddle points or maxima, the first derivative of the energy with respect to the coordinates is zero. It also related to the force, since the force is the negative gradient of the potential (Eq. 11) (McCammon and Stephen 1987). The second derivative matrix, representing the curvature of the potential energy surface, has all positive values, at a minimum point.

$$\boldsymbol{F} = -\boldsymbol{\nabla} U \tag{11}$$

2.5.1. Steepest descent method

Steepest descent (SD) method is a first-derivative minimization algorithm that is generally used in the initial stages of preparation to relieve bad contacts that may be present in an experimental structure (Wiberg 1965). At each step, the gradient of the potential (the vector of the first derivative) is calculated.

$$\boldsymbol{s}_k = -\frac{\boldsymbol{g}_k}{|\boldsymbol{g}_k|} \tag{12}$$

Here, $\boldsymbol{s}_k$ is the $3N$-dimensional unit vector, and $\boldsymbol{g}_k$ is the potential energy gradient vector component of the coordinate k. Next, a change is added to all the coordinates, in a direction opposite to the gradient. Starting from a point $\boldsymbol{r}_k$, the displacement is added using:

$$\boldsymbol{r}_{k+1} = \boldsymbol{r}_k + \lambda_k \boldsymbol{s}_k \tag{13}$$

where the λ_k is the displacement parameter. The size of the displacement, or how far to move along the gradient, is determined through different approaches. The algorithm gradually changes atomic coordinates as the system moves closer to the minimum point. Both the gradients and the directions of successive steps are orthogonal. Despite having a poor convergence and accuracy, coming from oscillating around a minimum, this algorithm is computationally fast, and useful as a first step in minimizing a structure.

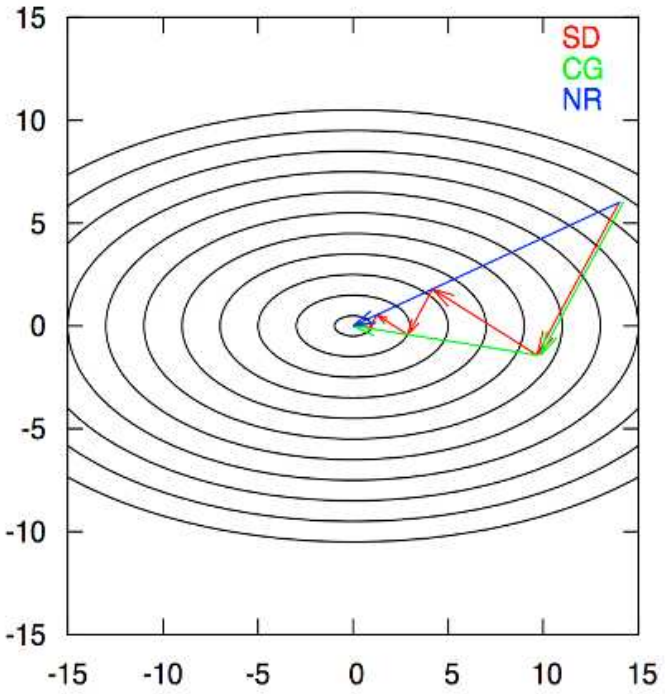


Figure 13. Representation of different energy minimization algorithms. Steepest descent (SD) requires the most steps, while conjugate gradient (CG) requires fewer steps, but is more computationally costly. Newton Raphson (NR) can minimize quadratic function in one step, but it is more most costly computationally. Image adapted from Leach, 2001.

2.5.2. Conjugate gradient

Conjugate gradient (CD) algorithm (Williams and Schleyer 1968) also uses the first-derivative information. It uses the gradient from the previous minimization step, in addition to the current gradient, to determine the next displacement direction $\boldsymbol{v}_k$, as their linear combination:

$$\boldsymbol{v}_k = \boldsymbol{g}_k + \gamma_k \boldsymbol{v}_{k-1} \tag{14}$$

where γ_k can be calculated using:

$$\gamma_k = \frac{\boldsymbol{g}_k \cdot \boldsymbol{g}_k}{\boldsymbol{g}_{k-1} \cdot \boldsymbol{g}_{k-1}} \tag{15}$$

The following position is obtained through:

$$\boldsymbol{r}_{k+1} = \boldsymbol{r}_k + \lambda_k \boldsymbol{v}_k \tag{16}$$

For an *N*-dimensional quadratic surface, the algorithm reaches the minimum in *N* steps. This algorithm is more computationally demanding than the steepest descent, but it converges more quickly and produces a lower energy once it has converged.

2.5.3. Adopted-Basis Newton-Raphson

The Adopted-Basis Newton-Raphson (ABNR) is a variant of the Newton-Raphson algorithm that uses the second-derivative of the potential energy function (Bernard R. Brooks et al. 1983). The Newton-Raphson algorithm, starting from a point $\boldsymbol{r}_k$, has the next step $\boldsymbol{r}_{k+1}$ calculated by the following expression:

$$\boldsymbol{r}_{k+1} = \boldsymbol{r}_k - \frac{\boldsymbol{U}'_{pot}(\boldsymbol{r}_k)}{\boldsymbol{U}''_{pot}(\boldsymbol{r}_k)} \tag{17}$$

Calculating the second derivative entails the construction and inversion of the second derivative (Hessian) matrix, which can be computationally costly. The ABNR method consists of calculating only a subset of the matrix (of basis vectors) at each step and only taking into account the vectors for which the systems has made the most progress in the previous steps. This approach converges extremely fast, but it is more computationally demanding than the first-derivative methods.

In general, the Steepest Descent and Conjugate Gradient methods are used to quickly minimize structures and eliminate significant high-energy interactions. Afterwards, the second derivative methods can be used to produce lower energy conformations.

2.5. Normal Mode Analysis

The idea of exploring vibrational motions of proteins came hand in hand with the advancement of computational methods in the field of chemistry. The first study of the vibrational dynamics of an alpha-helix was done in 1979, by Karplus and Levy (Levy and Karplus 1979). This idea was developed further through the Normal Mode Analysis (NMA), and shortly after, the first protein systems were studied - glucagon and bovine pancreatic trypsin inhibitor (BPTI) (Tasumi et al. 1982; B. Brooks and Karplus 1983). Since then, normal mode analysis has been applied to study many protein systems and has been used to elucidate how protein conformational dynamics are associated to their function (Skjaerven, Hollup, and Reuter 2009; Mahajan and Sanejouand 2015). Normal modes describe the vibrational motion of molecular systems, and so the normal mode analysis represents a powerful tool for studying structural and dynamic properties of molecular systems. By definition, NMA involves the analytical study of harmonic potential wells, beginning with a stable conformation of the system that corresponds to a minimum on the potential energy surface, and then describing harmonic motions of small amplitude within the potential well (where “small” means “small enough that the harmonic approximations hold”). A stable conformational state is defined as a point on potential energy surface where the first derivative of the potential energy is zero, or sufficiently close to zero, and the second derivative of the potential energy is positive, which indicates a local minimum. Applying the formulation of normal mode analysis, the motions in these wells can be described as vibrational motions, called normal modes, and they represent a specific pattern of motion, each associated with a specific frequency. The technique has been widely used for small molecules in the study of their vibrational motions. For a system with N particles, there are $3N$ number of degrees of freedom for of a non-linear system. Six degrees of freedom represent rigid-body motions such as translation and rotation of a molecule, around three axes,

with no energy. The remaining *3N-6* degrees of freedom are called vibrational normal modes, or non-trivial modes, and are grouped by their frequencies, ω. The typical protein vibrational frequencies span from couple of wave-numbers (cm^{-1}) the tera-hertz (THz) region, for the entire protein, up to approximately 3280 cm^{-1} for individual bond vibration (stretching of C-N bond) (Foggia et al., n.d.). Protein dynamics can be described as a linear combination of all normal modes, and the modes are orthogonal - the motion of one mode does not interfere with another, allowing for clear separation of different vibrational patterns. The lowest frequency modes, also named soft modes, are believed to be functionally relevant (Skjaerven, Hollup, and Reuter 2009). Low frequency modes have the highest amplitude and the lowest energetic cost. They encompass all, or most of, the atoms in a system, and describe the largest movements in a protein. The vibrational modes in which all atoms are moving are called collective motions, and they exclude localised motions associated with higher frequencies.

Normal mode frequencies and the displacements of atoms are obtained by calculating the second derivative of the potential energy (described by molecular mechanics), also called the Hessian matrix (**F**). After minimizing a structure, and calculating the Hessian matrix, the values are weighted by atomic mass according to the equation:

$$\boldsymbol{F} = \boldsymbol{M}^{-\frac{1}{2}}\boldsymbol{\nabla}^2 U(\boldsymbol{r}^N)\boldsymbol{M}^{-\frac{1}{2}} \tag{18}$$

where **M** is the diagonal matrix of *3N* x *3N*, containing atomic masses only along the diagonal. $\mathbf{M}^{-\frac{1}{2}}$ represents the matrix with the inverse square root of atomic masses and weighing them is required since forces act differently with respect to atomic mass. The $\boldsymbol{\nabla}^2$ is the second derivative of the potential energy U of the system of the atoms r. By calculating eigenvalues (λ_i), and eigenvectors (a_i) of our matrix we can diagonalize it:

$$|\boldsymbol{\nabla}^2 U(r^N) - \lambda \boldsymbol{M}| = 0 \tag{19}$$

Each normal mode is described by two components: eigenvector $\boldsymbol{A}_i$, which represents displacements of each atom moves in a vibrational pattern, scaled by eigenvalue λ_i, corresponding to vibrational frequencies, or how fast the atoms are oscillating:

$$\boldsymbol{F} \times \boldsymbol{A}_l = \lambda_i \times \boldsymbol{A}_l \tag{20}$$

In summary, NMA uses mathematical tools to study how atoms in a molecule move slightly around a stable point, helping us understand their vibrations and motions.

2.5.1. Analysis of normal mode calculations

2.5.1.1. Fluctuations

From the results of a normal mode calculation, we can calculate properties such as root mean square fluctuations (RMSF). Atomic fluctuations for atom *i* at temperature T, can be calculated by Eq. 21, which uses normal mode (eigenvector) $\boldsymbol{k}$ and the corresponding eigenvalue:

$$< (\Delta r_i)^2 >_{\boldsymbol{k}} = k_B T \sum_{\boldsymbol{\alpha}=x,y,z} \frac{a_{i\alpha,k}^2}{\omega_k^2} \tag{21}$$

The Boltzmann constant is represented by k_B. ω_k is the frequency of mode k, Δr_i is the displacement of atom i from its minimum energy position, the scalar $a_{i\alpha,k}$ of mode k is the component of the eigenvector, $\boldsymbol{\alpha}$ = *(x,y,z)* that describes the displacement the atom *i*.

$$< (\Delta r_i)^2 > = \sum_{k=1}^{n} < (\Delta r_i)^2 >_k \tag{22}$$

Eq. 22 describes the contribution of all normal modes to the fluctuation of an atom *i*. These fluctuations can be compared to the fluctuations derived from the crystallographic *B* factors, calculated using the following equation:

$$B = \frac{3}{8}\pi^2(\Delta r_i)^2 \tag{23}$$

2.5.1.2. Correlations

The second analysis that we can carry out using the normal modes is to calculate the correlation coefficients for each protein atom pairs. This will give us information about the correlated motions exhibited by residues, and in more general view, the correlations between different secondary structure elements during the course of a trajectory. We can obtain the C_{ij} correlation coefficients using the following expression:

$$C_{ij} = \frac{< \Delta r_i \Delta r_j >}{\left(< \Delta r_i >< \Delta r_j >\right)^{\frac{1}{2}}} \tag{24}$$

where the Δr_i is the displacement from the average position of an atom i.

2.5.1.3. Infra - red spectra

Calculating the Infrared spectrum (IR) from normal modes is achieved through the extraction of frequencies and dipoles associated to a structure and, more precisely, the derivative of a dipole moment. The IR spectrum is represented as function of the integrated intensity, Γ_k, which has units of the molar absorptivity, calculated from the dipole derivative values:

$$\Gamma_k = \frac{N_0\pi^2}{3c^2\varepsilon_0\omega_k}\left(\frac{d\mu}{dQ_k}\right)^2 \tag{25}$$

where N_0 is Avogadro's number, ε_0 is the permittivity of vacuum, c is the speed of light, μ is the dipole moment, Q_k is the normal coordinate, and ω_k is the respective frequency. The IR

spectra were then smoothed by multiplying the intensity of each peak by a Gaussian function with a full width at half-maximum of 10 cm^{-1}.

We have described the calculation of normal modes, starting from a minimized structure of a molecular system. The protocol we develop in chapter III consists of extracting multiple representative structures from the most populated well of a free energy landscape calculated from a molecular dynamics simulation, using these for NMA and averaging the results. Since the normal mode analysis of a single structure is highly sensitive to that structure, we take this ensemble approach and average over hundreds of structures. This will be described in a separate chapter.

3. Molecular dynamics

3.1. Protocol

Molecular Dynamics (MD) simulation is a computational method that allows us to describe the microscopic world of molecules using numerical models. MD simulations calculate the dynamical evolution of a modelled system through time using classical dynamics. By carrying out MD simulations, we can analyse the properties of atomic assemblies in terms of structure, microscopic interactions and overall dynamics (Michael P. Allen 2004). This approach consists of integrating equations of motion by breaking down the calculation into discrete time steps. For classical MD simulations using an all-atom force field, the time step, *dt,* is in the order of 1 - 2 femtoseconds. The forces acting on atoms are calculated at each step and combined with the positions of the atoms at a time *t*, to generate positions at time t_{+1}. The new atomic positions are then determined, and a new set of forces is calculated. The product of a simulation is a trajectory, representing how atomic positions change with time.

MD simulations start with the preparation of the system followed by the production of a dynamics trajectory. The preparation involves modelling the molecule of interest by an empirical energy force field; in general, the initial molecular structure is determined experimentally usually by X-ray crystallography or NMR spectroscopy, or modelled using

different approaches, such as homology modelling or other structure prediction algorithms such as AlphaFold2. To include environmental effects, the molecule is most often modelled in an explicit water environment, along with additional elements used to emulate the physiological and physicochemical properties of biological systems (for example, ions). One necessary step before running an MD simulation is energy minimization, as discussed above – a well minimized structure is crucial for stable MD simulations. The molecular system is brought to a desired temperature in a series of steps designed to heat the system gradually by modulating the atomic velocities. Multiple heating and minimization steps are usually combined to ensure a proper preparation of the system. The system is then equilibrated by molecular dynamics simulations for some given time during which the temperature is monitored and readjusted accordingly. Different parameters, for example total energy and/or RMSD are monitored as a measure of stability. The final step, called the production, phrase will generate a trajectory of the molecular motions over time that is then used for analysis.

3.2. Integrating algorithm - All - Atom Additive FF

MD simulations calculate atomic trajectories by solving Newton's equations of motion for each of the atoms in the system. Atomic trajectories are a sequence of atomic positions in space as a function of time. The Newton's second law of motion stipulates that the force acting on particle i is equal to its mass m_i times acceleration $\boldsymbol{a}_i$ (the rate at which velocity v changes over time t, being the second derivative of the position $\boldsymbol{r}_i$):

$$\boldsymbol{F}_i = m_i \cdot \boldsymbol{a}_i \tag{26}$$

$$\boldsymbol{F}_i = m_i \cdot \frac{d\boldsymbol{v}_i}{dt} = m_i \cdot \frac{d^2\boldsymbol{r}_i}{dt^2} \tag{27}$$

The forces acting on the particles come from interactions with other particles, and they can be derived from potential energy function, where the total force on a particle, $\boldsymbol{F}_i$, is the negative gradient of the potential energy with respect to its position:

$$\boldsymbol{F}_i = -\boldsymbol{\nabla}_i U \tag{28}$$

Starting from initial atomic position $\boldsymbol{r}_i$, and the potential energy function U, the force is known. We can calculate the acceleration for every particle of the system and then update the atomic positions and velocities of particles using numerical integration methods, which are given by different algorithms. An analytical solution is not achievable, since the motions of the N particles are coupled. Initial velocities, $\boldsymbol{v}_i$, are assigned randomly from the Maxwell-Boltzmann distribution at a specified temperature (Leach 2001a).

The NAMD program (Phillips et al. 2020a), which is the most used in the present work to run the MD simulations, uses the Velocity Verlet algorithm for calculating atomic velocities and positions (Allen and Tildesley 1989; Swope et al. 1982). This algorithm calculates positions $\boldsymbol{r}_i$ and velocities $\boldsymbol{v}_i$ at the same time. The positions are calculated in the following manner:

$$\boldsymbol{r}_i(t + dt) = \boldsymbol{r}_i(t) + \boldsymbol{v}_i(t) \cdot dt + \frac{1}{2}\boldsymbol{a}_i(t) \cdot (dt)^2 \tag{29}$$

and are updated based on the current velocities and acceleration. Forces are then updated for $t + dt$ the based on new positions (and new potential energy U), and allow for the calculation of acceleration for $t + dt$:

$$\boldsymbol{a}_i\ (t\ +\ dt) = \frac{\boldsymbol{F}_i(t\ +\ dt)}{m_i} \tag{30}$$

Velocities $\boldsymbol{v}_i$ are updated based on:

$$\boldsymbol{v}_i\,(t\,+\,dt) = \boldsymbol{v}_i\,(t) + \frac{1}{2}\,[\,\boldsymbol{a}_i\,(t) + \boldsymbol{a}_i\,(t\,+\,dt)]\,\cdot dt \tag{31}$$

The velocity is updated using both the old and new accelerations, which ensures that the velocity is correctly synchronized with the positions.

3.3. Integrating algorithm - All - Atom Polarizable FF

The Drude oscillator model requires a different integration algorithm for including the auxiliary particles attached to core atoms. Extended Lagrangian dynamics is implemented in the NAMD program, instead of the computationally expensive, albeit accurate, self-consistent field (SCF) regime. (W. Jiang et al. 2011; Lamoureux and Roux 2003) This ensures lower computational cost since the energy is not minimized at each step of the dynamics. In extended Lagrangian dynamics, Drude particles are not restricted to their energetic minima, but instead carry kinetic energy that is kept at low values using a designated thermostat. Dual stochastic Langevin thermostats modify the equations of motion by adding both a frictional force and a random force to each particle's motion. The motions of atom-Drude pairs with coordinates r_i and $r_{D,i}$ are treated separately. We now have two different terms accounting for motions. The motion of the center-of-mass $\boldsymbol{R_i}$ of the atom-Drude pair and the relative internal motion of the oscillator $d_i = r_{D,i} - r_i$. The atom-Drude pair has a total mass m_i, and the oscillator has a reduced mass $m_i' = m_D\,(1 - m_D/m_i)$. The resulting pair of equations of motions for atom-Drude pairs are the following Langevin equations Eq. 32 and 33:

$$m_i \frac{d^2}{dt^2}\boldsymbol{R}_i \;=\; \boldsymbol{F}_{R,i} \;-\; \gamma \frac{d^2}{dt}\boldsymbol{R}_i \;+\boldsymbol{f}_i \tag{32}$$

$$m_i' \frac{d^2}{dt^2}\boldsymbol{d}_i \;=\; \boldsymbol{F}_{d,i} \;-\; \gamma' \frac{d^2}{dt}\boldsymbol{d}_i \;+\boldsymbol{f}_i' \tag{33}$$

The $\boldsymbol{F}_{R,i}$ and $\boldsymbol{F}_{d,i}$ are the forces acting on the centre-of-mass and on the oscillator reduced mass, respectively, γ and γ' are the internal and external Langevin friction coefficients, $\boldsymbol{f}_i$ and $\boldsymbol{f}_i'$ are two random forces whose values randomly change over time, according to the stochastic (random) force of Langevin dynamics, and with respect to two Langevin friction coefficients respectively. The forces acting on the centers-of-mass and on the displacements are represented as actual forces acting on particles:

$$\boldsymbol{F}_{R,i} = -\frac{dU}{dr_i} - \frac{dU}{dr_{D,i}} \tag{34}$$

and:

$$\boldsymbol{F}_{d,i} = -\left(1 - \frac{m_D}{m_i}\right)\frac{dU}{dr_{D,i}} + \left(\frac{m_D}{m_i}\right)\frac{dU}{dr_i} \tag{35}$$

The Brünger−Brooks−Karplus (BBK) method is implemented in NAMD for integrating the Langevin equation (Brünger, Brooks, and Karplus 1984). The integration of the centers-of-mass and the displacements are identical to the integration of the individual atoms. NAMD treats the entire system as standard atomic coordinates. After an initial equilibration using 0.5 femtosecond timestep, integration time steps of 1 or 2 femtoseconds are appropriate for generating accurate molecular dynamics trajectories.

3.4. Periodic boundary conditions

The most commonly used method for studying dynamical properties of bulk solvent and solvated systems is by molecular dynamics simulations using periodic boundary conditions (PBC) (Leach 2001b; Rapaport 2004). PBC allow for a relatively large number of particles to be simulated as if they were experiencing interactions in a bulk fluid (or other states of matter, not applicable in our studies). This approach takes our system, which is essentially a protein submerged in a water box, as a central cell and it replicates it in all spatial directions, called spatial images (Fig. 14). The box has no physical walls and, as a consequence, if a particle leaves the box from one side, it is reintroduced from the opposite side of the central box (in the

3 dimensions of space). By applying PBC, the system is treated as infinitely large and, when coupled to different algorithms, allows one to carry out simulations at thermodynamic equilibrium in different ensembles, for example, the microcanonical ensemble - NVE (constant number of particles (N), constant volume (V) and constant energy (E)), or the isothermal-isobaric ensemble - NPT (constant number of particles (N), pressure (P) and temperature (T)). Simulations in these ensembles are used alongside PBC, by coupling the central cell of atoms to a constant-temperature bath or a constant-pressure piston.

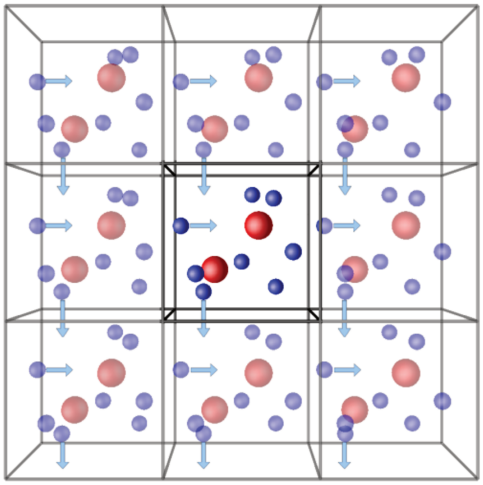


Figure 14. Schematic of periodic boundary conditions. Taken from https://people.se.cmich.edu/petko1vg/isaacs/manual/page22_ct.html#fpbc

3.5. Analysis

3.5.1. RMS deviation and fluctuation

Because the positions of the atoms of a dynamical system change over time, several properties can be calculated. We can calculate the average structural deviation (RSMD) from a reference structure, given by Eq. 36:

$$RMSD = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(r_j - r_j^{reference}\right)^2} \tag{36}$$

where j indexes a particular atom, N is the number of equivalent atoms, r_j is the actual position and $r_j^{reference}$ is the position of the same atom in the reference structure.

If the reference structure is the average structure calculated from the dynamics simulation, we can get a measure of the flexibility about that average structure for a particular atom. The RMSF (root mean square fluctuation) represents a measure of the deviation between the position of particle *i* and a reference position, when the reference position is the average structure over time, and it can be calculated using Eq. 37:

$$RMS_i^{fluct} = \sqrt{\frac{1}{T}\sum_{t_j=1}^{T}\left(r_i(t_j) - r_i^{average}\right)^2} \tag{37}$$

3.5.2. Correlations

The calculation of correlated motions is described in detail in section 2.5.1.2., in the context of Normal Mode Analysis, but correlations C_{ij} can be calculated directly from an MD trajectory, using the same equation:

$$C_{ij} = \frac{< \Delta r_i \Delta r_j >}{\left(< \Delta r_i >< \Delta r_j >\right)^{\frac{1}{2}}} \tag{38}$$

where the Δr_i is the displacement from the average position of a residue *i*.

3.5.3. Secondary structure element analysis

Secondary structure elements are well characterized protein regions that are stabilized by hydrogen bonds between atoms in the polypeptide backbone, providing a scaffold for the tertiary structure of the protein. The fluctuations in the secondary structure content can influence the protein's structure and function. In this work, we sought to quantify the secondary structure content of PPARγ systems for the purpose of comparing them with experimental IR measurements (chapters III and IV). We used a combination of the software STRIDE (Heinig and Frishman 2004a) and the python package MDAnalysis (R. Gowers et al. 2016). STRIDE software uses a knowledge-based approach to determine hydrogen bonding patterns and backbone geometry (notably the geometry of torsion angles). Weighted product of hydrogen bond energy and the probabilities of torsional angles determine the starting and ending position of secondary structure elements, with respect to empirically optimized thresholds. We developed a python script to pilot these two programs to efficiently analyse our dynamics trajectories.

Chapter III - Ensemble averaged Normal modes - a prelude

1. Introduction

Normal mode analysis (NMA) is well suited to study the conformational dynamics of proteins because of its relative simplicity and rapidity in terms of calculation time. Normal mode analysis has been especially useful for understanding the contributions of low-frequency collective vibrational modes to conformational changes. However, a shortcoming of normal mode analysis is that it is generally performed for a single energy minimized structure, obtained either from a crystal structure or from a molecular dynamics simulation. It is well appreciated today that a representative experimental "structure" actually corresponds to an ensemble of structures, so the calculation of properties from a single structure can lead to questions on the robustness of the results. Several approaches to address the issue of an ensemble of representative structures has been developed. In the work by van Vlijmen and Karplus (van Vlijmen and Karplus 1999), normal modes (NMs) were calculated for several individual energy-minimized protein structures sampled from 100ps MD simulations and properties calculated and averaged from the individual NM analyses gave a better agreement with experiment. Kitao and co-workers developed the ''jumping among minima'' (JAM) model (Kitao, Hayward, and Go 1998). This approach accounts for multiple structures by assuming the system can transition between local minima that are separated by barriers that can be crossed on the timescale of MD simulations. Batista and coworkers developed a 'consensus modes' approach where the harmonic information present in multiple minima of the potential energy surface of a protein was exploited (Batista et al. 2010).

As simulations have become much longer, there is the realization that crystal structures represent a restrictive view of the conformational ensemble sampled by a given protein. So, much more extensive conformational sampling is considered the norm. For example, in the method of Karplus and Hermann, the simulations were 100ps in length; today, simulations are generally 100s nano- to microsecond in length.

We present here a method that builds on the use of multiple structures for normal mode analysis (NMA) and that is better adapted to the analysis of longer simulations. Our ensemble approach involves combining data from multiple structures from a molecular dynamics

simulation extracted from the most populated well of a free energy surface. The free energy surface is generated from RMSD and radius of gyration, as in Cote, et al. (Cote et al. 2017a), the structures are reoriented to the same reference frame and an NMA is performed on each structure to obtain the normal modes. We calculate the normal modes for the extracted structures and subsequently average the results. From the normal modes, we calculate the average RMSF, correlated motions and the far-IR spectra that we will compare to experiment in later chapters. The combination of spectra from multiple structures allows us to obtain an ensemble averaged IR spectra. Overall, this approach is more robust and less dependent on any individual structure. We will present this ensemble averaged approach through an application to the ligand binding domain of PPARγ, shown in Fig. 15.

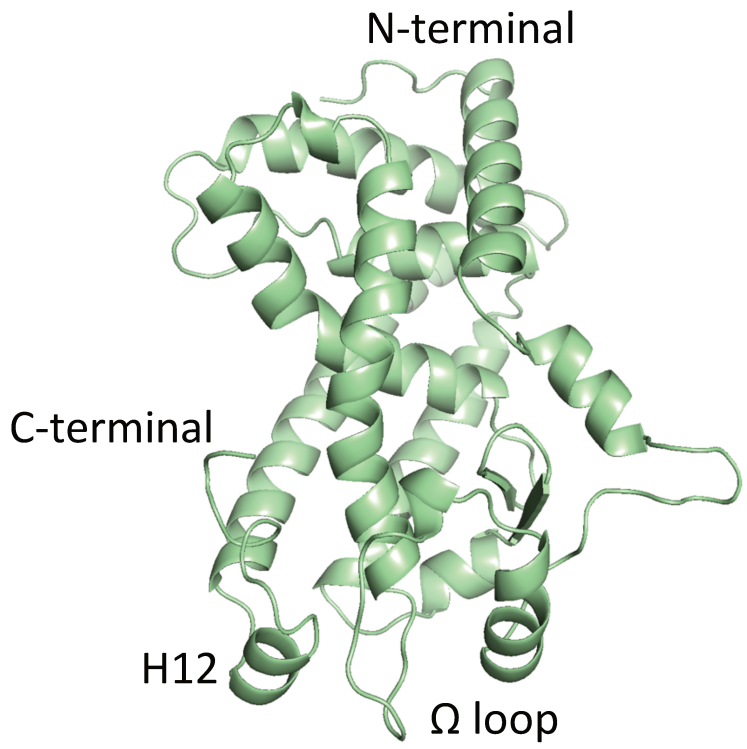


Figure 15. PPARgamma ligand binding domain. Residues 230 - 505, from the 3.2Å crystal structure PDB ID: 7WOX chain B, in apo form. Key structural elements are labelled.

2. Methods

2.1. Structure preparation

The PPARγ LBD (residues 230 - 505) of wild type (WT) apo structure was prepared for this analysis using a crystallographic structure available in the Protein Data Bank (PDB) (Burley et al. 2023). The WT apo model of 276 residues was based on the structure of chain B from PDBID 7WOX (Yoshizawa et al. 2022a), which has a 3.20 Å resolution. Chain A in this PDB entry is bound to the antagonist MMT-160, but chain B was considered to be in apo form because it did not show any electron density, representing a ligand, in the binding pocket.

The initial coordinate and protein structure files (PSF) needed for the molecular dynamics simulations were prepared using the PDB Reader & Manipulator option of the CHARMM-GUI web interface (Jo et al. 2008) using default parameters. The interface was also used to build missing protein residues. The protonation states of the histidine residues of our systems were determined using PROPKA program (Olsson et al. 2011; Søndergaard et al. 2011) via the webserver https://server.poissonboltzmann.org/pdb2pqr, and verified manually.

The following protocol was used for the molecular dynamics simulation using the CHARMM all-atom force field, version 36m (A. D. MacKerell, Bashford, et al. 1998). The molecular dynamics simulation was carried out for 100 ns using NAMD program (Phillips et al. 2020b) with the time step of 1 fs under NPT conditions. System preparation was done in four stages following solvation of the protein in a 100Å cubic water box.. First stage consisted of 1000 steps of minimization with a fixed protein chain and non-fixed water-ion box. Next, the system was heated up to 600K during 23000 steps. The system was, once again, minimized for 1000 steps, and then heated up to 296.5 K. Then, with the constraints on the protein chain removed, the system was minimized for 2000 steps. The entire system was heated up to 296.5 K, for 15000 steps. This was followed by an equilibration of 85 000 steps of dynamics, followed by the production phase. The production simulation was carried out for a duration of 100 ns.

2.2. Analysis of the trajectories

From the molecular dynamics simulation, we calculated the root-mean-square coordinate difference (RMSD) and the radius of gyration (RGYR) considering only the Cα backbone atoms. Using the results for RMSD and RGYR, a free energy surface was constructed and the most populated well was identified. From this well, 100 structures were extracted for further analysis. These 100 structures compose our structural ensemble. We further calculated the backbone atomic root-mean-square fluctuations (RMSFs) averaged by residue from the simulations.

2.3. Normal Mode Analysis

Normal mode analysis was carried out using the VIBRAN module of the CHARMM program, see Methods section 2.5 for a description of the methodology. All the modes were calculated in this analysis (*3N* atoms) corresponding to 13464 Cartesian displacement modes for the APO form of PPARγ. The first six modes correspond to the global translational and rotational modes and were removed from the analysis.

From the results of the normal mode analysis of the ensemble of structures, the intrinsic dynamics were characterized by the atomic root-mean-square fluctuations (RMSF), which were calculated using Eqs. 21 and 22. The cross-correlation coefficients for the ensemble of structures were also calculated from the NMA using Eq. 24 of Methods. From the C_{ij} correlation coefficients, which are organized as a matrix, a cross-correlation map was calculated using a color-coded 2D representation. In this representation, C_{ij} = 1 identifies correlated motions and C_{ij} = -1 anti-correlated motions. These values give us information concerning the global collective motions.

Using structures in the ensemble, which are representative of highly probable structures through the analysis of the free energy surface, and the results of their NMA, the IR spectra were computed (Cote et al. 2017a). Each IR spectrum was represented as function of the

integrated intensity, Γ_k, which has units of the molar absorptivity, calculated from the dipole derivative values, using Eq. 25 of Methods.

3. Results and discussion

3.1. RMSD, RGYR and RMSF

The RMSD and RGYR time series were calculated from the 100ns simulation of PPARγ apo-WT. The time series results are shown in Fig. 16 and 17. We see that in both cases, the time series level off at plateau values even though the RMSD time series shows some variation.

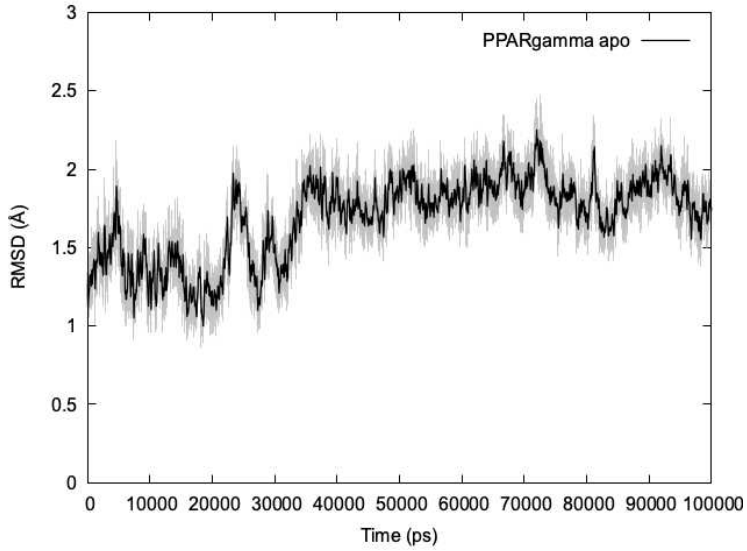


Figure 16. RMSD of the Cα atoms of the complete backbone. In black is the running average over 500 timeframes and in grey are the values at each time point.

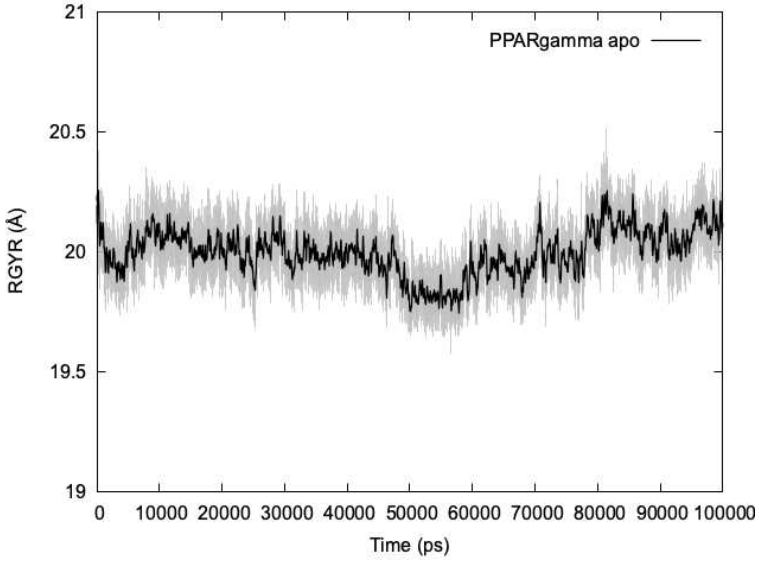


Figure 17. Radius of gyration of the Cα atoms of the complete backbone. In black is the running average over 500 timeframes and in grey are the values at each time point.

We additionally calculated the RMSF from the molecular dynamics simulation and averaged the results by-residue for the backbone atoms. The results are shown in Fig. 18, which also shows the secondary structure of the protein as a function of sequence. We see that, for the most part, secondary structure elements, such as alpha helices (green rectangles), are more stable than the terminal ends and loops between the secondary structure elements, but being an apo structure, the LBD displays a certain degree of flexibility.

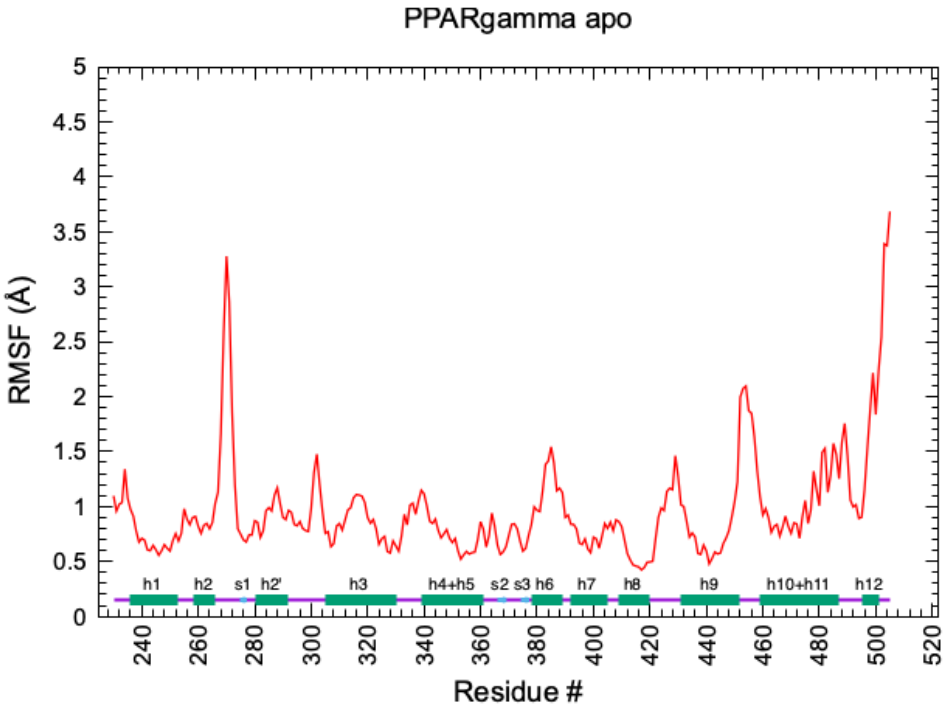


Figure 18. By-residue averages backbone RMS fluctuations from the molecular dynamics simulations of PPARγ apo form.

To assess the conformational landscape of the protein, an effective 2D free-energy landscape (FEL) based on the RMSD and RGYR was constructed from molecular dynamics simulations. The 2D FEL was based on the values of the RMSD and the RGYR computed from the protein structures extracted from the simulations using the relationship as follows:

$$FEL(RMSD, RGYR) = -k_B T ln P(RMSD, RGYR) \quad (39)$$

where k_B and T are the Boltzmann constant and the temperature, respectively, and P is the joint probability of a structure having the values of RMSD and RGYR. The landscape result is shown in Fig. 19, where we see that, in this simulation, there is one principal well between RMSD values of, roughly 1.65Å and 1.95Å and an RGYR of 19.0Å and 20Å. A minor well is observed at smaller RMSD values and slightly higher RGYR values.

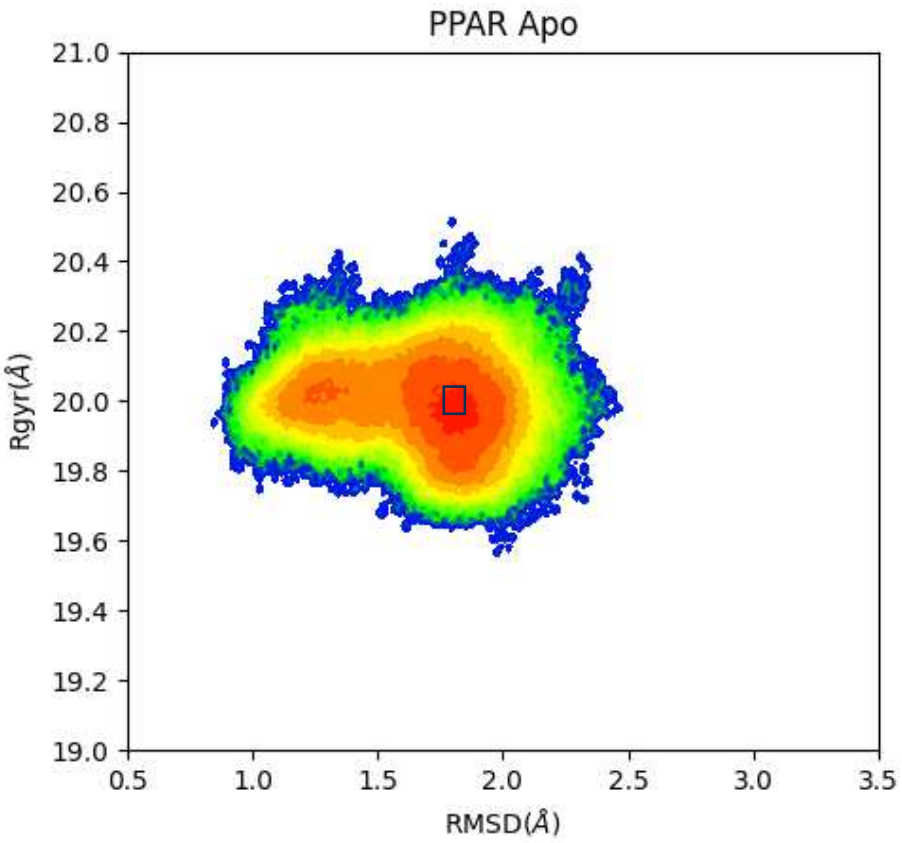


Figure 19. The effective free energy landscape built from the RMSD and the RGYR values computed from the MD simulation the APO form of PPARγ. The more red the region, the deeper is the well. Indicated in the square is the region from which structures were extracted for further analysis.

3.2. Normal mode calculations

3.2.1. Fluctuations

From the deepest minimum on the surface, that being for RMSD between 1.85 and 1.95Å and a radius of gyration between 19.9 and 19.95, denoted by the square on the FES plot, Fig. 19, we generated a dynamics trajectory of structures that fell within those ranges. Nearly 10000 structures were extracted (9754) from this region of the free energy landscape. From this cluster, we randomly extracted 100 structures for normal mode analysis (NMA). Each structure was subjected to enough steps of minimization using the adapted basis Newton–Raphson (ABNR) algorithm to reach an RMS gradient of 10^{-7} kcal.mol^{-1}.Å^{-1} or less. The treatment of the nonbond interactions in the normal mode calculations was the same as that used for the energy minimization. This ensured that each structure was at the local minimum of the potential energy surface, a necessary condition for NMA. A switching function was used for the van der Waals non-bonded interactions and a shift function with the distance-dependent dielectric, $\varepsilon = 4r$, was used for the electrostatic interactions. An atom-based 15Å cutoff was used.

The atomic fluctuations were calculated from the normal modes using equation Eq. 21. In Fig. 20, the fluctuations from normal modes are compared to the fluctuations calculated from the molecular dynamics simulations. We see that the trends between the simulation results and the normal mode results follow well. As expected, the fluctuations are generally greater in loop regions for both the MD simulations and the normal mode results. Fluctuations calculated from NMA are generally smaller than corresponding fluctuations calculated from molecular dynamics simulations.

In Figure 21, we show the RMSF calculated from several different structures extracted from the lowest free energy well. The structures themselves are all roughly 1.9Å from the initial crystal structure. We see that the detailed results, in this case, of RMSF can show a certain degree of variation between structures, in particular, for the more flexible loop regions. This is a manifestation of the fact that even slight differences in energy minimized structures can lead to different results, raising the question when comparing results from different

systems, are the differences due to the inherent differences between the systems or could the conclusions change if different, but perfectly acceptable minimized structures are used.

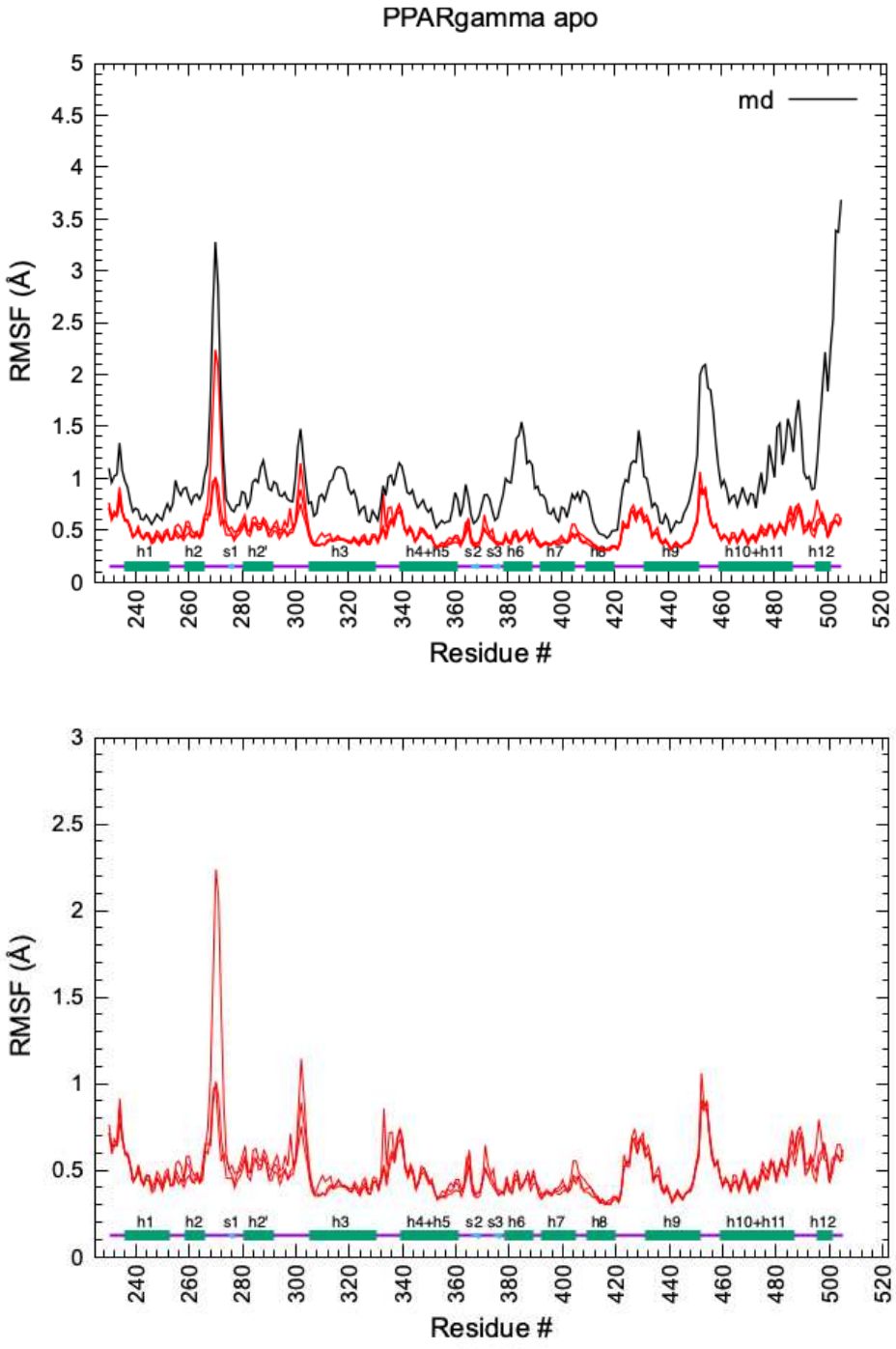


Figure 20. Atomic fluctuations of backbone atoms from the molecular dynamics simulations (black) and the atomic fluctuations of backbone atoms from normal mode calculations (red). Examples of several results from normal modes are shown. In the bottom panel, we zoom in on the fluctuations from normal modes.

3.2.2. Correlated motions

From the normal modes for several individual structures, we calculated the correlated motions using Eq. 24. We compare the correlated motions calculated from a single structure to the correlated motions calculated and averaged from the full ensemble extracted from the molecular dynamics trajectory. We see in Fig. 21, that, for the most part, the general characteristics concerning secondary structure elements are similar, reflecting the topology of the structure.

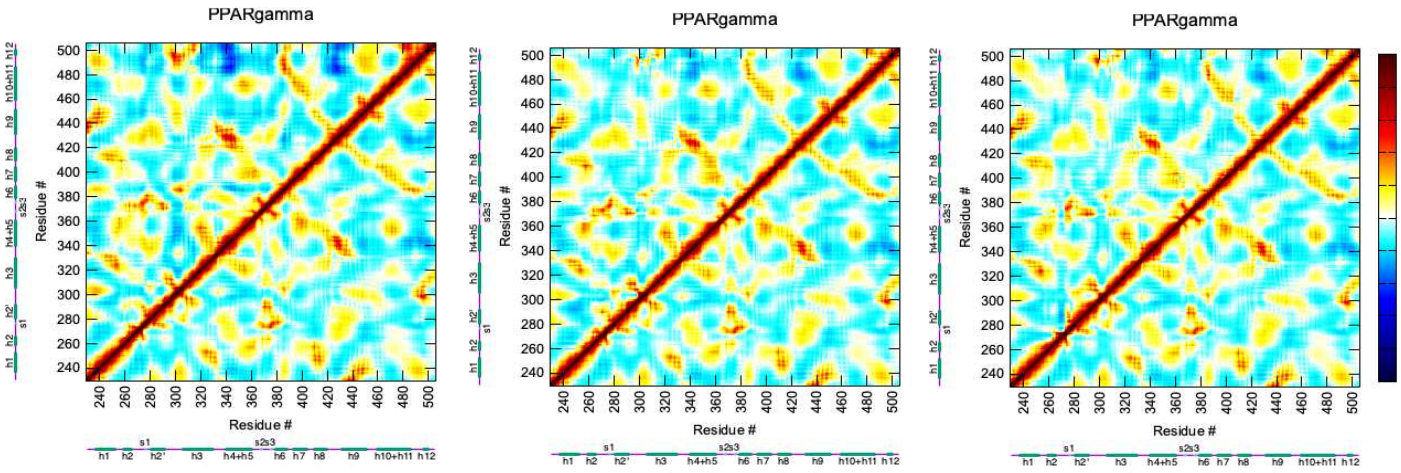


Figure 21. Comparison of correlated motions from three individual structures (upper triangle) to the correlated motions calculated and averaged from the ensemble of structures extracted from the molecular dynamics simulations (lower triangle).

But we see differences between regions that do not belong to the same secondary structure elements, for example, the anticorrelated motions implicating helix 4-5 and helix 12 differ between conformations and from the correlated motions calculated from the full ensemble of structures. Other regions where anticorrelated motions differ are between helix 9 and helix 12. Or the correlated motions between the beta sheets and the Ω loop region including helix 2.

By averaging over the ensemble of structures, we get a more robust representation of the correlated motions. In Fig. 22, we show the comparison of the correlated motions averaging of 10, 50, 75 and 95 structures to the correlations from the full ensemble of 100 structures.

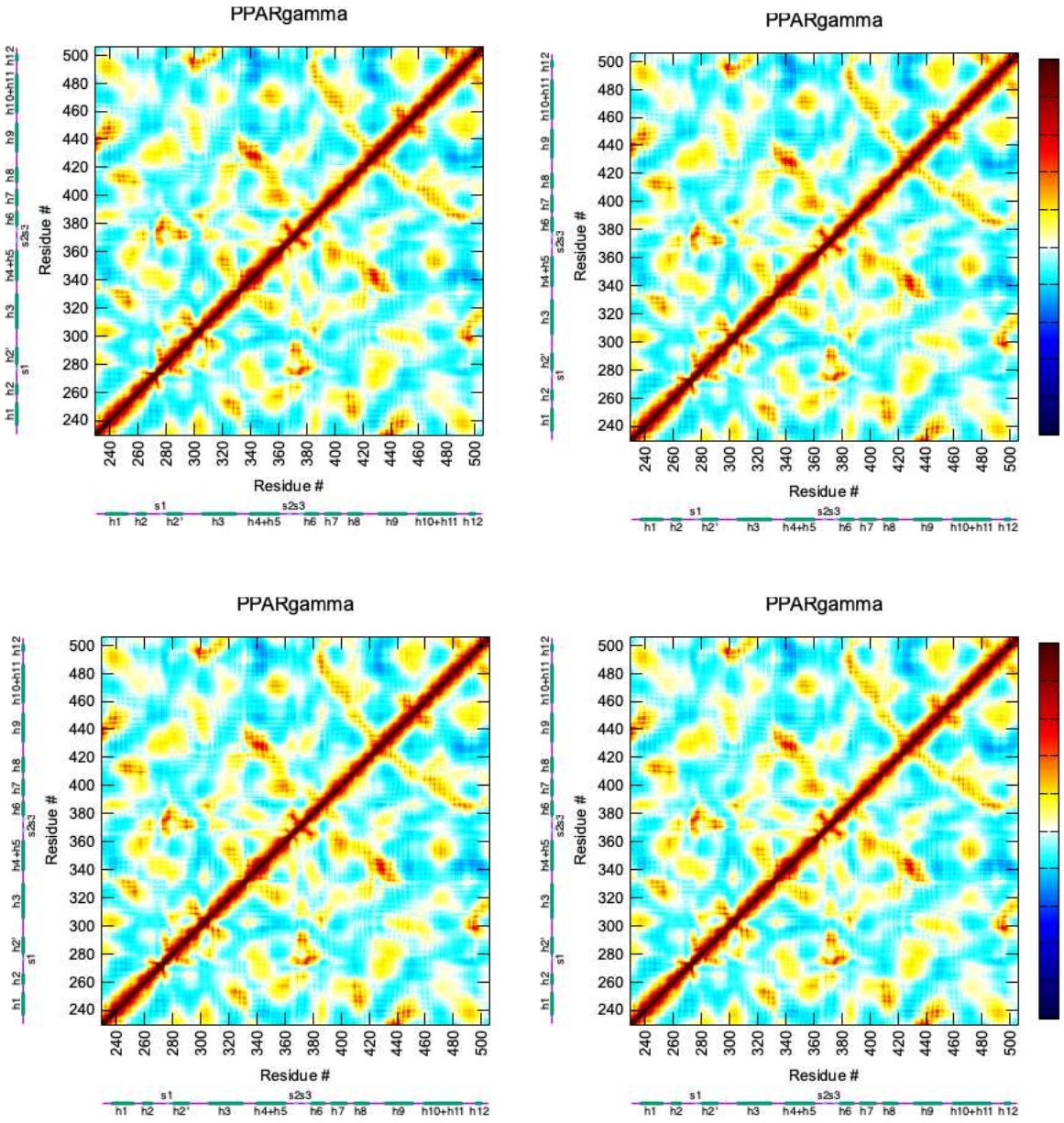


Figure 22. Correlated motions calculated from the NMA. The top graphs represent 10 and 50 structures, and bottom graphs 75 and 95 structures, from left to right. Upper left and bottom right triangles of a given individual graph are identical.

Averaging over 50 structures seems sufficient for good convergence to the full ensemble result based on a visual inspection.

To demonstrate that the correlated motions are well reproduced by the low frequency vibrational modes, we calculated the correlated motions maps for PPARγ for a single structure, taking a progressively larger number of modes determined by normal mode analysis and Eq. 24. We compared the calculation of the correlated motions using the full set of normal modes to the correlated motions calculated from the first 15 modes which range from 1.2 to 4.5 cm^{-1}, the first 150 modes ranging from 1.2 to 22.2 cm^{-1} and the first 1500 mode range from 1.2 to 158.6 cm^{-1} modes. The corelated motions maps are shown in Fig. 23, where in the lower triangle, the correlated motions are calculated from the full set of normal modes and in the upper triangle, for the first 15, 150 and 1500 modes, respectively. We see that, even just the first few modes, up to 4.5 cm^{-1} include enough information to characterize the topological features of the secondary structural elements. The same conclusion is made when including modes up to 22 cm^{-1}. However, the dynamic properties are exaggerated in the correlated motions calculated by just the lowest frequency modes. Higher frequencies are required to get convergence with the correlated motions calculated from the full set of modes. While the lowest frequency modes are crucial for the underlying dynamics of the protein, the higher frequency modes in the far-IR region also contribute to the larger-scale collective motions, and by extension, more accurate calculation of dynamical properties.

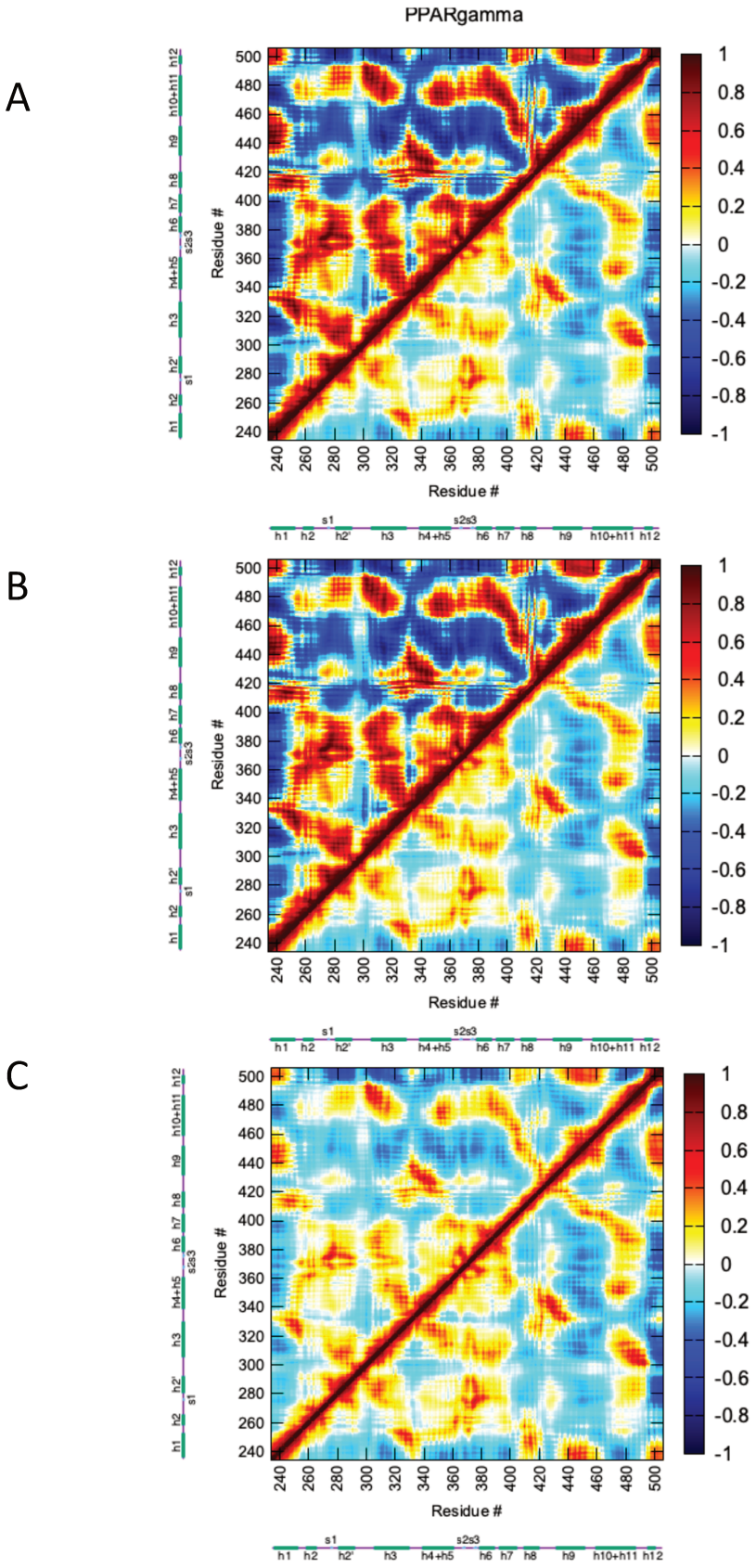


Figure 23. Correlated motions calculated from normal modes comparing correlations calculated from the full set of normal modes to correlations calculated from a subset of normal modes. A. modes up to 4.5cm^{-1}, B. modes up to 22.2cm^{-1}, C. modes up to 158.6cm^{-1}.

3.2.3 Far-IR spectra

We calculated the far-infrared (far-IR) spectra from normal mode analysis. The normal mode analysis was performed on each of the structures extracted from the lowest energy well. Each structure was energy optimized. The NMA provided frequencies, corresponding to normal modes that represent collective atomic displacements, and the change in the dipole moment with respect to the atomic displacements. Far-IR spectra are derived from these low-frequency vibrational modes, typically in the range of 10 to 400 cm^{-1}. The intensity of each mode in the far-IR region depends on the change in dipole moment with respect to atomic displacements. The spectrum is then generated by summing up the contributions from all modes, with each mode's intensity and frequency contributing to a broadened peak. The result is a spectrum that reflects the interactions and motions in the low-frequency region, often related to collective motions such as vibrations or other large-scale rearrangements within the molecular structure.

As with the other properties calculated from the normal modes, we show that the detailed results vary with the choice of specific structure. In Fig. 24, we see the far-IR spectra from two sets of three structures taken from two different PPARγ systems. We see that the variation in maximum peak position and features of individual spectra vary significantly. As with the other properties, this makes it difficult to decide if differences between two systems, when interpreted at a detailed level, could change if another valid structure was used.

In our ensemble averaged approach, we average the results over the ensemble of structures taken from the minimum of the free energy surface. In Fig. 25, we see the results comparing two systems, after averaging structures represented in Fig. 24.

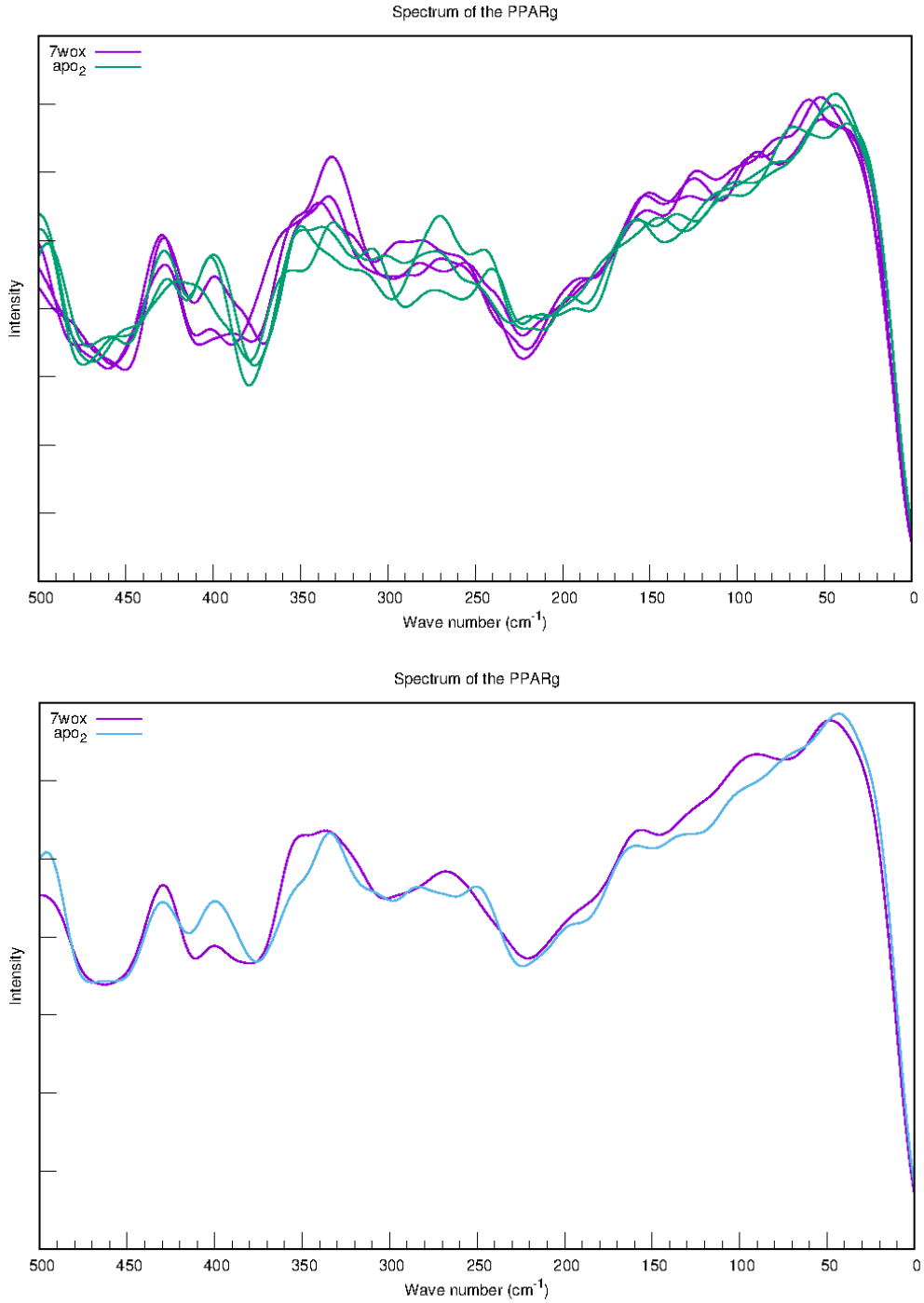


Figure 24. Calculated Far-IR spectra. Upper panel: Far-IR spectrum obtained from NMA of individual structures. Three structures were taken from each system. Systems are annotated '7wox' and 'apo2'. Bottom panel: Average Far-IR spectrum obtained from averaging the results from NMA of the ensemble of structures.

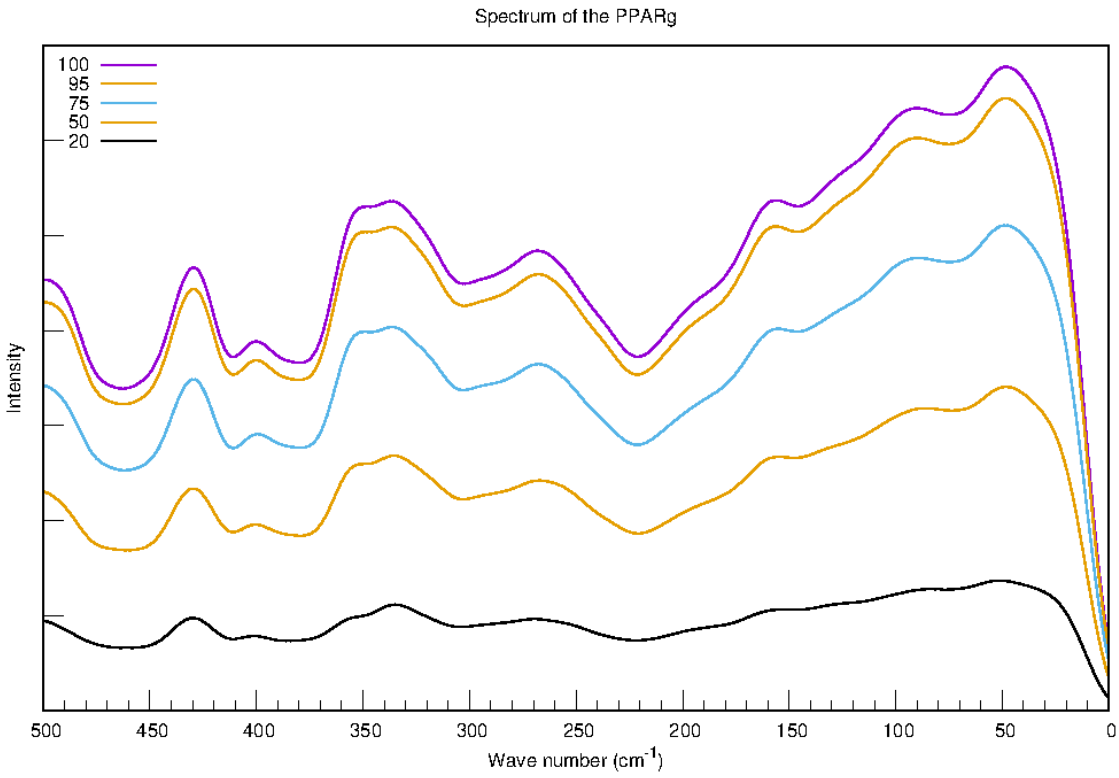


Figure 25. Convergence of the spectrum as a function of the number of structures included in the averaging.

Averaging results until convergence ensures robustness in the reported results. In Fig. 25, we see that peak positions and curve features begin to converge after averaging over about 75 structures from the ensemble and with 95 structures there is little variation in the calculated results. Throughout this work in this thesis, we therefore limit our ensembles to 100 structures.

4. Conclusions

Here we introduced a new approach for the calculation of properties from an ensemble of conformations obtained from molecular dynamics simulations. In contrast to previous works, the ensemble of structures was extracted from the deepest well on a pseudo free energy surface constructed from RMSD and RGYR values of the structures sampling in the molecular dynamics trajectory.

In most normal-mode calculations, a single energy minimized structure is used. Here, we averaged the results from 100 structures. We showed that for single structures and using all the modes in the calculation of the properties, there was significant variation in structural dynamic properties, such as the RMSF by residue. We also showed that, using single structures, the calculation of the correlated motions and the far-IR spectrum displayed significant variation.

Normal-mode calculations based on a single minimized structure can vary significantly from one structure to another. The averaging of normal mode results over an ensemble of subspace conformations leads to better converged results. Experimental observations reflect an average over different native subspace conformations of the protein. However, from a computational perspective, averaging seems to defy the purpose of normal-mode calculations, which is in part to avoid the calculation of computationally expensive MD trajectories. To average, one needs an ensemble of native subspace conformations. In conclusion, the utility of normal-mode calculations for analysing structural dynamic properties of proteins is increased by calculating them for several structures obtained from MD simulations and averaging the results. This is expected to provide more meaningful values and makes it possible to determine error bars.

Chapter IV - Application of Ensemble-averaged Normal modes to PPARγ - a combined MD/Far-IR study

1. Introduction

Nuclear receptor (NR) proteins form the largest family of transcription factors that regulate the transcription of genes in metazoans. They are implicated in the transcriptional cascade underlying many physiological phenomena related to cell cycle, differentiation, apoptosis, development, reproduction and homeostasis (Laudet and Gronemeyer 2001). An important feature of NRs is that their regulation of gene expression is dependent on the fixation of small ligands, thus acting as a hub that translates a chemical signal into a biological response. The transmission of this information implicates conformational changes in the protein. This ligand dependent activity of NRs makes them central targets for drug development in many diseases including diabetes, arteriosclerosis, inflammatory diseases and cancer (Xu 2015).

While much is known about the structure of nuclear receptor protein, in particular, the ligand binding domain, increasing attention is being focused on the role of protein structural dynamics in crucial cellular signaling pathways. Large scale structural dynamics play a critical role in the ability of nuclear receptor proteins to function properly (J. P. Renaud and Moras 2000c). Structural dynamics without large-scale conformational changes likely allow nuclear receptor proteins to bind to a variety of ligands, both natural and synthetic, with varying affinity and specificity allowing nuclear receptors to respond to a wide range of signaling molecules (Ekins et al. 2009; Kroker and Bruning 2015). Structural dynamics in nuclear receptor proteins is necessary for dimerization either at the level of the ligand binding or the DNA binding domains. In full length nuclear receptor proteins, flexibility of the linker region between the DBD and the LBD is necessary for dimerization and accommodation of varying DNA sequences allowed a more expansive range of DNA target sequences and subsequent gene regulation (A. K. M. Patel et al. 2023). Large-scale structural dynamics are implicated in co-regulator binding, in particular when agonist binding leads to the stabilization of H12 in a conformation where the recruitment of coactivator proteins is promoted (J. P. Renaud and Moras 2000c). The binding of antagonists prevents H12 from populating this transcriptionally active conformation and the binding of an inverse agonist promotes the binding of corepressor proteins. This strong dependence of function on conformation has led to the development of

selective modulators for a variety of nuclear receptor proteins, which exploit this dependence to modulate nuclear receptor activity, often in a tissue-selective context (Holzer, Markov, and Laudet 2017). Several clinically approved modulators exist to treat a variety of conditions, including cancers, metabolic disorders, osteoporosis, and hormone-related diseases.

Besides modulating conformations in nuclear receptor proteins, ligand binding is believed to influence the underlying collective dynamics (Fidelak et al. 2010a). Collective dynamics results from the concerted motions of a large number of atoms throughout the protein and have been suggested to play a significant role in controlling functional dynamical mechanisms, such as enzyme catalysis (Agarwal et al. 2002), ligand binding (Meireles et al. 2011) and allosteric signaling (Smith et al. 2016). Modulating these collective structural dynamics is becoming an important avenue of exploitation for the discovery of new therapeutic compounds, in particular allosteric drugs (Tee and Berezovsky 2024).

Collective motions occur in the low-frequency region of vibrational spectra and have been measured by techniques such as neutron scattering (Hong et al. 2016), vibrational spectroscopy (Rischel et al. 1998) and NMR spectroscopy (Lewandowski et al. 2011). Previously, far-IR spectroscopy and molecular dynamics simulations were combined to study the response of a PDZ domain to the binding of a small peptide ligand, elucidating the mechanism of allostery in this protein domain (Cote et al. 2017a). This integrated approach was used to quantify changes in low-frequency collective motions even for proteins without substantial conformational change upon ligand binding. Infrared (IR) spectroscopy, encompassing both mid-infrared (Mid-IR) and far-infrared (Far-IR) regions, is a powerful analytical technique for studying protein structure and dynamics. Mid-IR spectroscopy is particularly useful for examining protein secondary structures, as it probes vibrational modes that are sensitive to the backbone conformation of proteins, such as the amide I and II bands. The amide I band, arising mainly from the C=O stretching vibration, is located in the spectral range of 1700 to 1600 cm^{-1} and provides detailed information about the secondary structure elements like alpha-helices, beta-sheets, and random coils. Far-IR spectroscopy explores the low-frequency vibrational modes below 400 cm^{-1}, which are often associated with collective motions and large-scale conformational changes within the protein. These motions are crucial

for understanding dynamic allostery and the functional conformational changes that occur upon ligand binding. Ligand binding can influence the underlying collective motions in this frequency range. In the work presented here, we extend the development and application of an integrated approach to characterize the structural dynamics of a nuclear receptor (NR) protein. We apply our ensemble approach to normal mode analysis to all the systems studied. We focus on Peroxisome Proliferator-Activated Receptor gamma (PPARγ), a nuclear receptor that is a particularly important target for the development of therapeutic compounds for diseases such as diabetes and cancer. In collaboration with the team of Professor Petra Hellwig (UMR 7140, University of Strasbourg), both mid-IR and far-IR spectroscopies were employed to analyze the LBDs of PPARγ in both wild-type (apo and holo) and two mutant forms, T475M and F310S. The proteins were provided by Dr. Natacha Rochel and Dr. Judit Osz of our team at the IGBMC. The work presented in this chapter are the molecular dynamics simulations and associated developments that contributed to this collaborative study through integration with the experimental data and the subsequent analysis done to extract information on the collective motions of the protein. Studying the protein by this combined experimental/computational approach in the presence and absence of ligands, as well as a function of mutation, we gained insights into how ligand binding and mutations influence the structural dynamics of the protein at the level of allosteric regulation. This dual approach allowed us to capture a comprehensive picture of the conformational landscape and the underlying mechanisms of PPARγ's functional modulation. Such information can contribute to the development of new therapeutic strategies.

In this chapter, we present the molecular dynamics simulations of PPARγ in WT apo and holo forms, where PPARγ was complexed to the ligand GW1929. GW1929 is a non-thiazolidinedione selective PPARγ agonist (Heppner et al. 2004) and its chemical structure is shown in Fig. 26. In addition to the apo and holo WT forms, we also studied two mutant forms - T475M, which is mutant form that is known to have increased transcriptional activity even in apo form (Natacha Rochel et al. 2019) and F310S, which is known to be a loss-of-function mutation (Coutos-Thévenot, Beji, Neyret-Kahn, Pippo, Fontugne, Osz, Krucker, Groeneveld, et al. 2019). T475M and F310S mutations are implicated in luminal and basal bladder cancer, respectively. The structures of the 4 PPARγ systems are shown in Fig. 27 . Also shown for the

mutant structures are the positions of the point mutations. The mutation T475M is located on the helix called H10-11, while the F310S mutation is located on H3 (for the detailed numbering of helices in PPARγ, see Fig. 9 in the introduction).

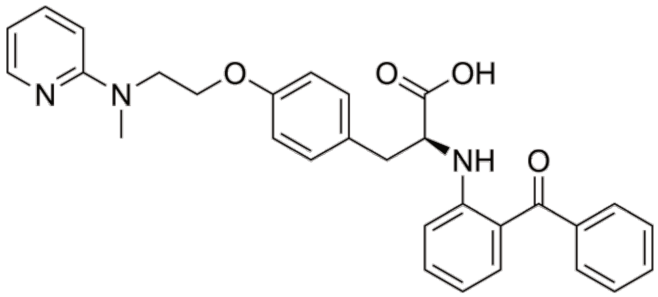


Figure 26. The chemical structure of the PPARγ agonist GW1929 used in this study.

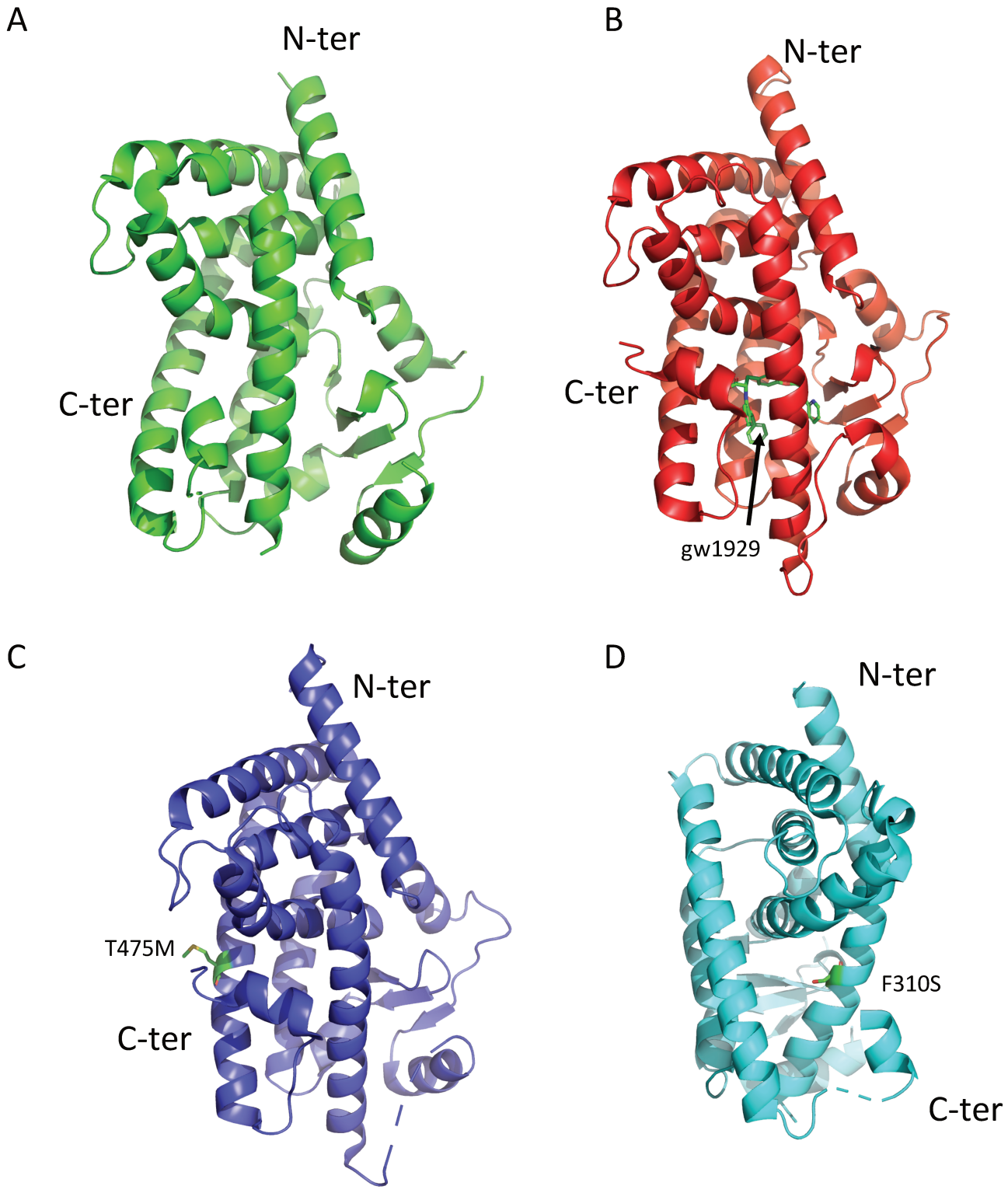


Figure 27. Structures used in the study. A) Apo WT (PDB ID: 7WOX, chain B). B) WT complexed to GW1929 (PDB ID: 6D8X). C) Apo T575M mutant (6FYZ). D) Apo F310S mutant constructed on 7WOX template.

2. Methods

2.1. Model preparation

Four different PPARγ LBD (residues 230 - 505) models, including the WT apo structure, WT holo with ligand GW1929, mutants T475M and F310S in their apo forms, were prepared for numerical simulation based on crystallographic structures available in the Protein Data Bank (PDB) (www.rcsb.org) (Berman et al. 2000; Burley et al. 2023). In this study, the numbering of the LBD sequence corresponds to PPARγ2 isoform (residues 230 - 505).

The WT apo model of 276 residues was based on the structure of chain B from PDBID 7WOX (Yoshizawa et al. 2022b), which has a 3.20 Å resolution. Chain A in this PDB entry is bound to the antagonist MMT-160, but chain B was considered to be in apo form because it did not show any electron density representing a ligand in the binding pocket. The initial coordinate and protein structure files (PSF) needed for simulations were prepared using the PDB Reader & Manipulator option of the CHARMM-GUI web interface (Jo et al. 2008), using default parameters. The interface was also used to build missing protein residues. The protein chain was energy minimized using 700 steps of SD algorithm. The protein chain, along with crystallographic waters and neutralizing ions (90 Na^+ ions, 87 Cl^- ions, ~150mM concentration) was placed into a 100$Å^3$ explicit TIP3P model water box.

The mutant F310S model is based on the crystal apo structure 7WOX. The protein chain B was extracted from the PDB file and residue F310 was modified to a serine using the CHARMM program (Bernard R. Brooks et al. 1983). The protein chain was energy minimized using 700 steps of SD algorithm. The protein chain, along with crystallographic waters and neutralizing ions (90 Na^+ ions, 87 Cl^- ions, ~150mM concentration) was placed into a 100$Å^3$ explicit TIP3P model water box.

The PPARγ bound to the agonist GW1929 (see Fig. 26) was modeled using the structure from PDBID: 6D8X (PDB DOI: https://doi.org/10.2210/pdb6D8X/pdb, to be published), chain A, which has a resolution of 1.90 Å. The 276 residue LBD coordinate and protein structure file were prepared using CHARMM-GUI interface and CHARMM program as described previously. Based on our observation of a ‘water – network’ representing a hydrogen bond

network bridging ligand and protein residues that seemingly contributes to stabilizing the ligand binding pocket, the crystallographic water molecules were kept in the preparation. Initial parameters for the ligand GW1929 were obtained from the ParamChem webserver (Vanommeslaeghe and MacKerell 2012) and used without modification as no parameter penalty exceeded the recommended limit. The protein chain was energy minimized using 700 steps of SD algorithm. The system was further neutralized with 90 ions of Na^+ and 85 ions of Cl^-, in a 100 $Å^3$ water box.

The structure of mutant T475M used for the simulations was based on the structure of chain A from, PDB ID: 6FZY, which has a resolution of 3.10 Å (Natacha Rochel et al. 2019). This structure has 279 residues (residues 227 – 505 of the LBD). Seven residues are missing from the Ω - loop region in the crystal structure, reflecting local flexibility. These missing residues were built using the tools available in the CHARMM-GUI interface. The protein chain was energy minimized using 700 steps of SD algorithm, then placed in the 100 $Å^3$ water box, along with 90 Na^+ and 87 Cl^- neutralizing ions.

2.2 Determination of histidine protonation states

Prior to setting up molecular dynamics simulations of a protein, it is necessary to determine the protonation states of the titratable residues. While most residues are taken in their standard protonation states because their pKa values are relatively far from physiological pH. Because of its imidazole group side chain, histidine can readily exist in different protonation states depending on the local environment. For the isolated His amino acid, the three protonation states of His are shown in Fig 28. In an acidic environment (low pH), histidine can exist in a fully protonated form where both Nδ and Nε are protonated giving the residue a +1 charge (HSP). In an environment that is more pH neutral, histidine exists in a neutral (deprotonated form) where it can be protonated on either Nδ or Nε atoms, depending on the local environment. However, in a protein, depending on the local environment, the protonated HSP form can often be stabilized at neutral pH.

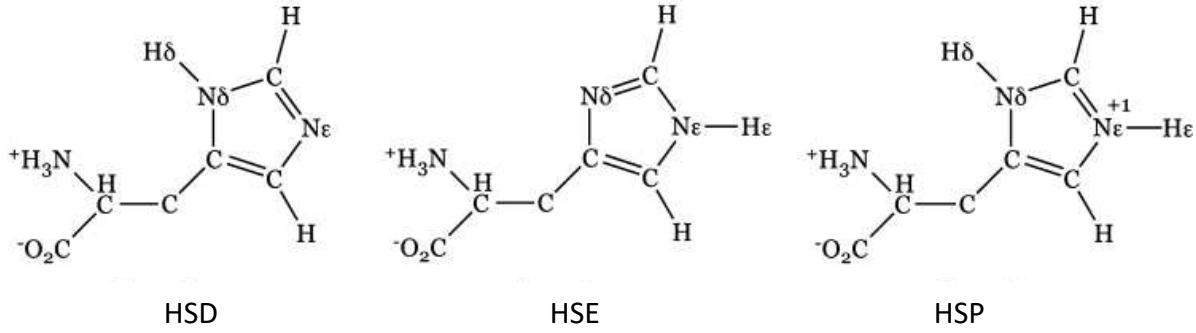


Figure 28. The three different protonation state of histidine. The names are those used in the CHARMM force field.

How the protonation state is set in the simulation can have an important impact on the final results. To determine the protonation states of the histidine residues of our systems, we used the PROPKA program (Olsson et al. 2011; Søndergaard et al. 2011) via the webserver https://server.poissonboltzmann.org/pdb2pqr. PROPKA takes an empirical approach to estimate titration states using parameters derived from experimental data and pKa values. Based on the local structural, it estimates pKa values for the titratable residues in the protein. The results for PPARγ are presented in Table 2. The protonation states given in Table 2 were taken for all apo structures (WT and two mutants). For the holo-WT complex (PDBID 6D8X), the protonation states were the same as in Table 2, except for His245, taken as HSD, His453, taken as HSD and His477, taken as HSD.

Table 2. The PROPKA suggestions for histidine protonation states of the apo protein. The HSD, HSE and HSP annotations represent histidines carrying hydrogens on atoms Nδ, Nε and on both nitrogen atoms, respectively. The mutant T475M has one additional histidine on N-terminal end of the protein (first row in the table).

histidine residue n°	protonation state
(229)	(HSD)
245	HSE
294	HSD
351	HSP
453	HSP
477	HSE
494	HSE

His351 is located on H4-H5 and extends into the ligand binding pocket. In the holo-WT structure, PDBID 6D8X, the PROPKA server first suggested this His as being neutral and protonated on the Nδ of the imidazole ring (HSD). However, the PROPKA server does not take into account the presence of ligands, and the structure PDBID 6D8X is of PPARγ bound to the ligand GW1929. Visual inspection of the binding pocket showed that, in the vicinity of His351, there is a water molecule that can stabilize protonation on the Nδ, thus justifying the assignment made by PROPKA, but in the presence of the ligand, protonation can also occur on the Nε, which points towards the carboxyl group of the ligand, see Fig. 29. Given that there is no crystallographic water molecule in the vicinity acting as H bond donor for the ligand and that a unprotonated Nε– ligand interaction would be unfavorable, His351 was finally reassigned as HSP, or doubly protonated for both apo and holo forms.

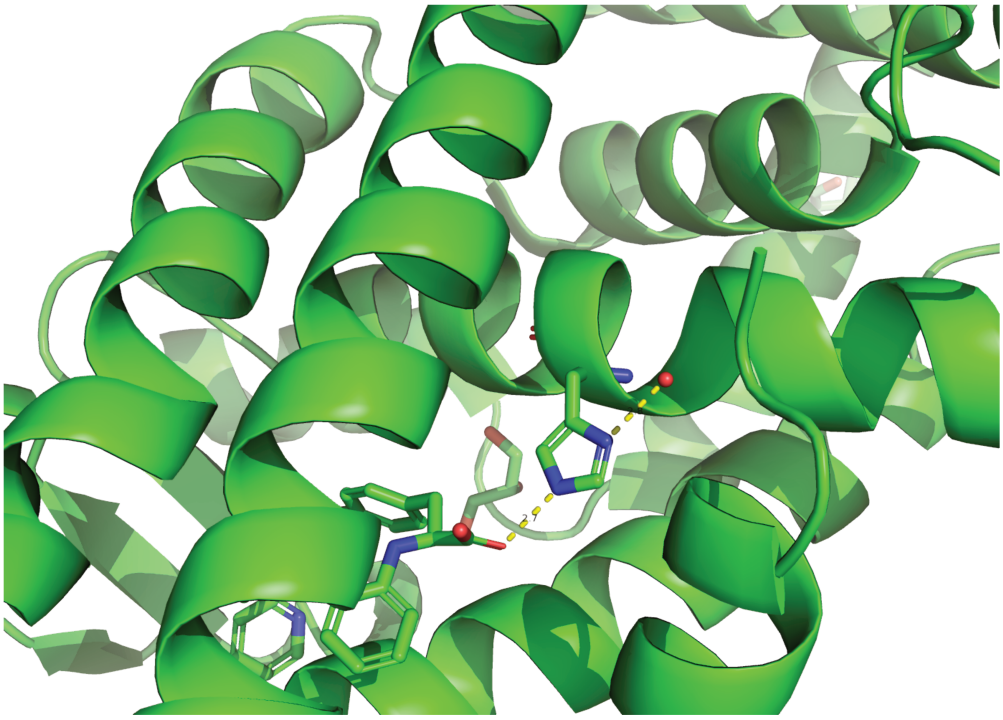


Figure 29. Shown is the crystallographic structure of the holo WT PPARg bound to GW1929 (PDB ID: 6D8X) Shown in stick representation are the ligand, GW1929 and His351.

Even though the PROPKA is generally considered to be a validated approach for assigning protonation states, visual inspection provides an additional degree of validation, as witnessed by the case of His351 in the structure PDBID 6D8X.

2.3. Molecular dynamics simulation parameters

The CHARMM program (B. R. Brooks et al. 2009) was used to prepare the solvated proteins. The protein chain was first energy minimized for 700 steps using the steepest descent algorithm (SD). A simulated-annealing type protocol was carried out to prepare the system for molecular dynamics simulation. In the first step, the protein and, if applicable, ligand, were fixed in their position and 1000 steps of energy minimization was done to allow the water and ions to adjust position in response to the presence of the protein. Next, the system was heated up to 600K, during 23000 steps, again with the protein/ligand fixed. This was followed by another energy

minimization for 1000 steps. This was followed by a heating to 296.5 K. The constraints on the protein/ligand were removed and the entire system was energy minimized for of 2000 steps. The entire system was then heated up to 296.5 K over 15000 steps, followed by an equilibration run of 85 000 steps of dynamics that was followed by the production phase. A time step of 2 fs was used. The duration of each simulation was 100 stages of 1 x 10^6 timesteps, which resulted in 200 ns - long simulations. Structures were saved every 250 steps, resulting in 400000 structures per trajectory. Three simulation replicas were carried out for all four PPARγ LBD systems yielding a total of 2.4 μs of simulation.

3. Results and discussion

3.1. RMSD and RMSF

The RMSD time series provides a measure of the change in conformation of the protein as a function of time and is used to assess the stability of a molecular dynamics simulation. For each of the three replicas of the four different PPARγ systems being studied here, the RMSD time series were calculated from the MD. The results from each replica of each system were averaged and the average time series was plotted for the 200 ns of trajectory, along with the variation (highest and lowest values) at each time point. Referring to Fig 30, all four systems show relatively stable trajectories given that the RMSD time series reach plateau values. Upon visual inspection of final structures, there is no significant loss of secondary structure integrity. The RMSD mean value of PPARγ apo WT system has the average RMSD of 2.7 Å (st. dev. 0.21 Å). After 150ns, a slight increase in RMSD value is observed and there is an increase of variability between replicas (variation in pink of the highest and lowest values). That apo form of the WT shows increased flexibility, especially in comparison to the holo form, is not unexpected given that experimental measurements suggest this behaviour (B A Johnson et al. 2000; Chalmers et al. 2011). The WT – GW1929 complex has an average RMSD of 3.0 Å (st. dev. 0.11 Å) and, although this is a higher number compared to the WT RMSD, we see that the trajectory is more stable and reaches a plateau after 60 ns of trajectory. We calculated the RMSD of the agonist ligand separately (average value: 2.3 Å, st. dev. 0.1 Å). After some

conformational adjustment in the first 100 ns of trajectory, it reaches a plateau and becomes more stable at around an RMSD of 2.5 Å. Both mutant structures present high RMSD values, with the RMSD of 3.6 Å (st. dev. 0.09 Å) for the mutant T575M and 3.6 Å (st. dev. 0.07 Å) for F310S. Both simulations are stable, with F310S displaying less variability between replicas. The F310S model was based on the WT apo structure, but it presents a higher RMSD value much sooner in the trajectory.

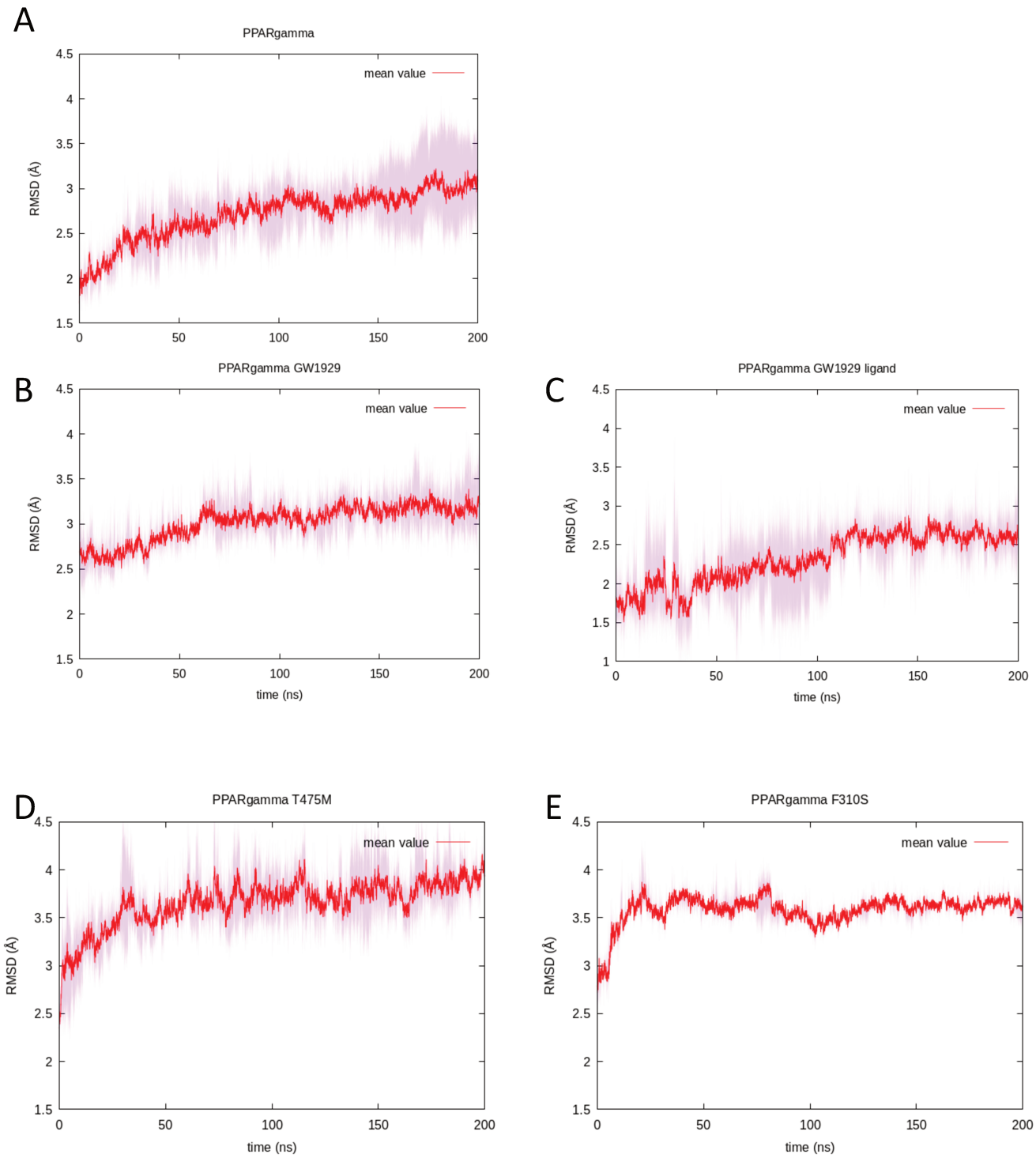


Figure 30. RMSD timeseries graphs of 200 ns simulations of PPARgamma LBD, averaged over 3 replicas. Four different systems: A) WT apo, B) WT - GW1929, C) for the GW1292 ligand alone, and mutants D) T475M and E. F310S. The mean value of 3 replicas is represented as a red line and the variation at each time point by the shaded pink.

The root mean square fluctuation (RMSF) values provide a quantification of the degree of flexibility of the protein around an average structure. Here, the RMSF values averaged over the three replicas are plotted by residue number, from residues 230 (226 for the mutant T475M) to 505. To facilitate the understanding of these results, the secondary structure elements are represented on a horizontal line, with alpha helices represented as green, and beta strands as blue rectangles. The mean values are plotted in red and the highest/lowest values as pink area. In general manner, all four systems present values that are characteristic for the PPARγ LBD. Lower flexibility is observed for residues of secondary structure elements and higher flexibility for loops and terminal end regions. Still, differences between systems are noticeable. The PPARγ apo WT form has stable regions, including helices H1, H3, H5, H8, H9, H10. The most flexible regions are the loop H2 – S1, with RMSF values up to 4 Å and the C – terminal residues with RMSF values up to 4.5 Å (Fig. 31A). As expected, all loops are more flexible than the secondary structure regions. The highest variability between replicas is noticeable in the helix H12 and the C – terminal region.

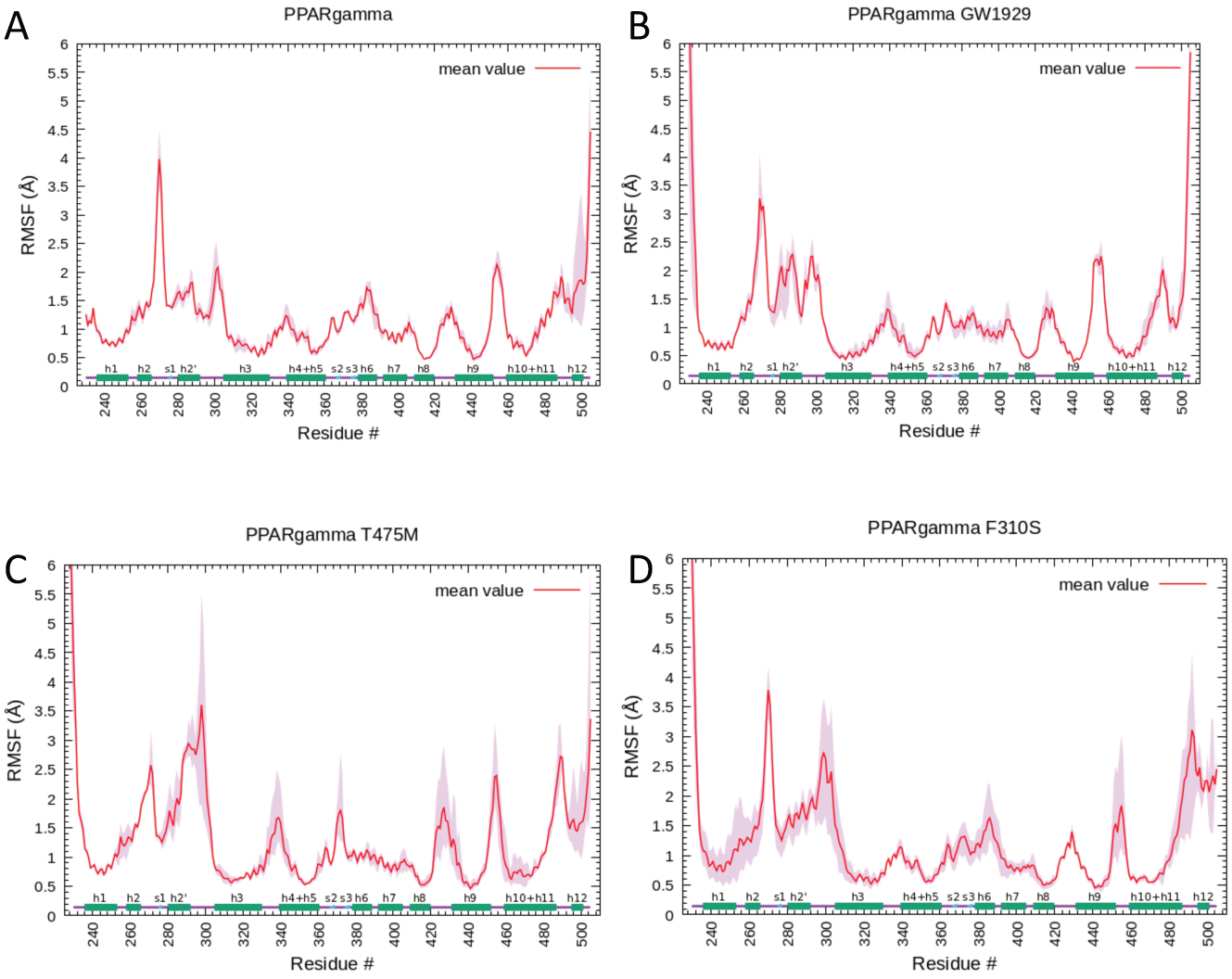


Figure 31. RMS fluctuations of PPARgamma LBD, for four biological systems. The mean value of 3 replicas is represented as a red line. Secondary structure elements are shown on the x axis: alpha helices (h1 – h12) as green, and beta strands (s1 – s3) as blue rectangles. A)WT apo, B) WT with GW1929, C) the T475M mutant and D) the F310S mutant.

The RMSF values for PPARγ – GW1929 also follow the characteristic pattern (Fig. 31B). The N-terminal region is not of particular interest here since it is the continuation of the hinge region connecting the DBD and LBD domains, and it is expected to be flexible. In the PPARγ – GW1929 system, the RMSF values are slightly higher than those in the WT apo form, this is particularly noticeable in the omega loop region. The RMSF increases for the loop residues. Helices H1, H3, H6, H10 and H12 appear more stable comparing to the WT apo system. The C – terminal residues fluctuate up to 6 Å, but the residues of the helix H12 exhibit

low fluctuations of around 2 Å. The two PPARγ mutated system, T475M and F310S (Fig. 31C and D, respectively), also follow the classical pattern of PPARγ fluctuations. However, they present higher variability of RMSF values among three replicas, represented by a larger pink area.

The T475M system has the most significant fluctuation values of up to 3.5 Å in the Ω - loop region. It is the only system where the Ω - loop region presents higher values of RMSF than the H2 – S1 loop. Another area of high fluctuation is helix H12 and, in particular, the very C – terminal residues, which have RMSF values of around 3.7 Å. The fluctuations of the H3 – H4 loop have the mean values of 2.7 Å, the highest of all four systems, which also displays the highest variability among replicas. The same is observed for the H8 – H9 loop, which surpasses other PPARγ systems with the average value of about 1.8 Å. The helix H6, loop H6 – H7 and the helix H7 have somewhat uniform fluctuations that are slightly lower than the WT apo and F310S systems, and similar to the PPARγ – GW1929 system.

The F310S mutant system displays high variability of RMSF values when comparing three replicas (Fig. 33D). The most variability is seen in the region around helices H1, H2, Ω - loop, H8 – H9 loop, H9 – H10 loops and helix H12. Recall that F310S is a loss-of-function mutant, while T475M is a gain-of-function mutant (Natacha Rochel et al. 2019)

3.2. RMSD - Radius of gyration based free-energy landscapes

In the subsequent analysis of the simulations, we will follow the protocol outlined in the previous chapter, where we developed our Ensemble normal mode approach for calculating dynamical properties. In the first step of applying this approach here, we calculate the free energy landscape from RMSD and radius of gyration (RGYR).

To assess the conformational landscape of PPARγ in the different systems studied here, an effective free energy landscape (FEL) was constructed from the molecular dynamics simulations. Based on the vales of the backbone RMSD and the RGYR, the FEL was computed

using Eq. 39, in chapter III. The FEL was calculated for each replica simulation; the results are presented in Figs. 32 - 35.

For each system, we show three individual plots, one for each molecular dynamics replica simulation. For each replica simulation, we calculated the backbone RMSD and RGYR in order to construct the landscapes through binning of the discrete values. Each simulation saved 400 thousand structures for analysis. The ‘pseudo’ potential energy surface maps constructed from these trajectories are used to visually cluster the large number of structures that we have for each simulation as they serve as an indicator of different clusters or ‘wells’ that are populated during the simulations. In Fig 32-35, the more red the area, the more populated are those regions of conformational space, meaning that the majority of structures in a given well have similar values of RMSD and RGYR. As per our ensemble approach, structures were extracted from the most populated well for further analysis.

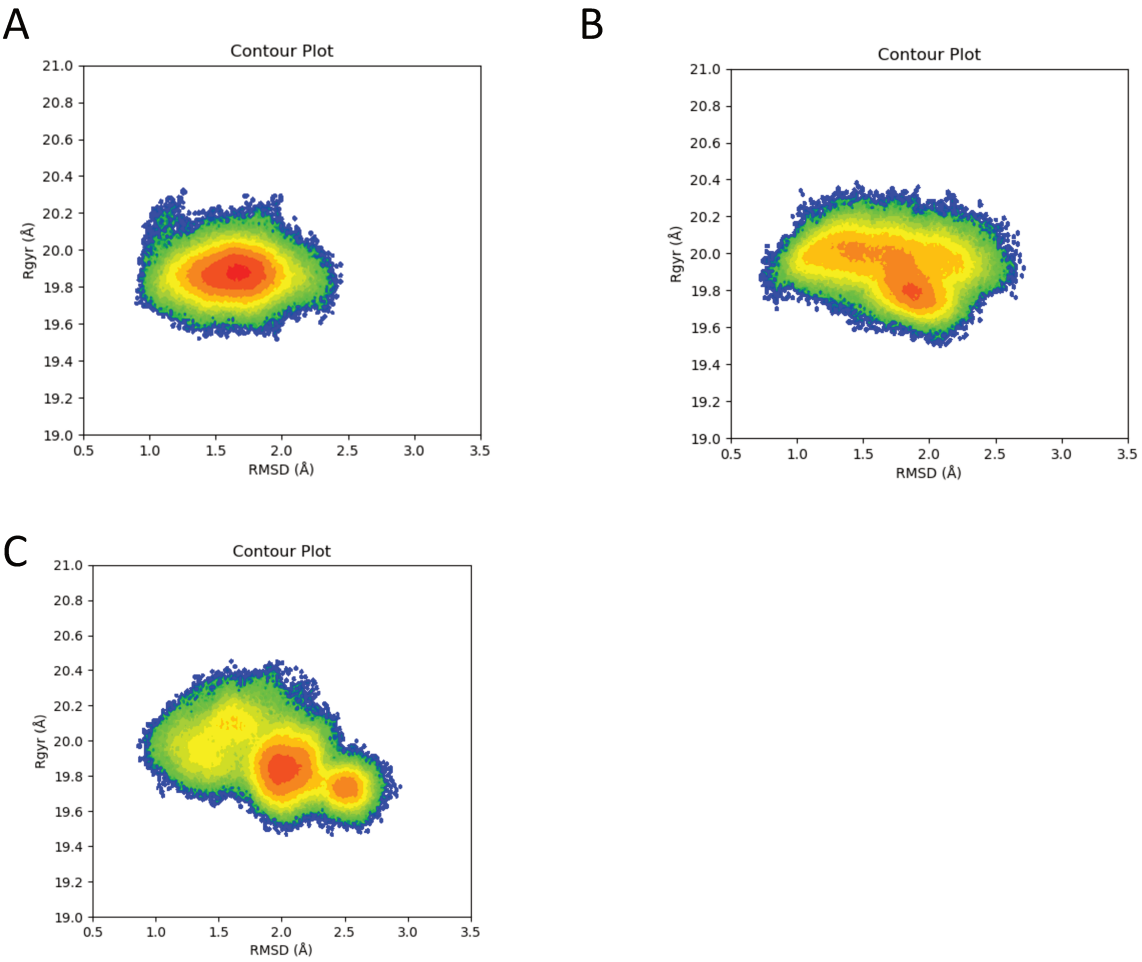


Figure 32. RMSD – RGYR free energy landscapes for individual replicas for the WT APO system. A) replica 1, B) replica 2, C) replica 3.

Referring to Fig. 32, the WT apo system has the scale of 19 – 20 Å for the radius of gyration, and 0.5 – 3.5 Å scale for the RMSD values. The three FEL plots display different aspects. For replica 1 (Fig. 32A) the FEL suggests that the trajectory is more restrained in sampling than replicas 2 and 3 (Fig. 32B, C respectively) samples. Replica 3 shows structures that are less compact and with values of RMSD up to 2.9 Å, which are higher than for the others. That the simulation of the apo WT structures show larger variability is not unexpected.

The FELs for the PPARγ – GW1929 system are shown in Fig. 34. The plots have scales ranging from 19.2 – 20.3 Å for the radius of gyration and 1.2 – 4.2 Å for the RMSD. The first replica (Fig. 33A) displays a plot with two wells, one more populated than the other. Other two replicas (replica2, Fig. 33B and replica3, Fig. 33C) sampled less conformational space. This is probably due to the fact that, in this set of simulations, PPARγ is complexed to the strong agonist, GW1929, which stabilizes the LBD structure. Even though GW1929 is a strong agonist and stabilizes the LBD, it can also stabilize the conformation in a well that can display calculated results that differ more significantly from the other replicas. We note that one of the replicas (33A) displays two minima, which highlights the importance of extensive conformational sampling, a consideration that underpins our ensemble sampling approach.

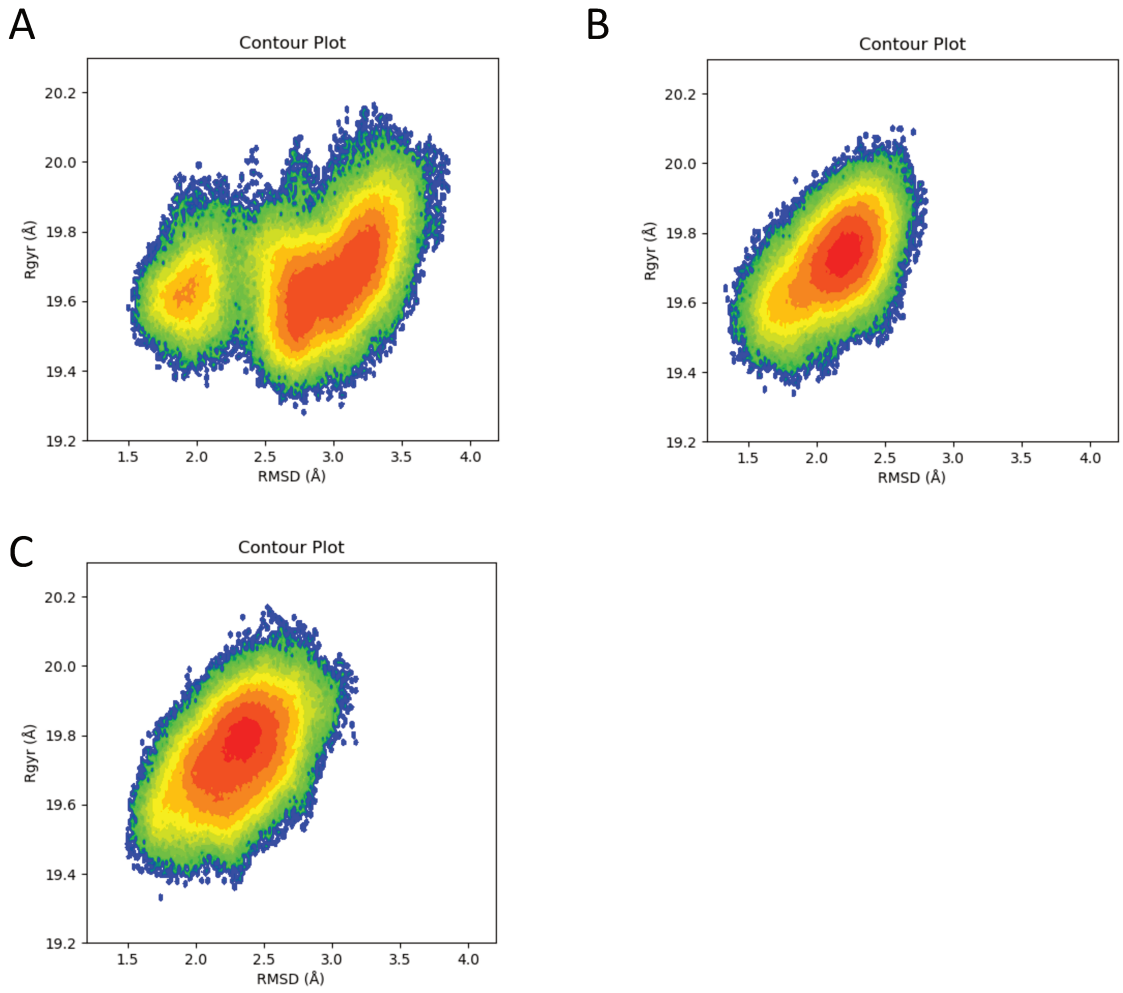


Figure 33. RMSD – RGYR plots free energy landscape for individual replicas for the WT with GW1929 system. A) replica 1, B) replica 2, C) replica 3.

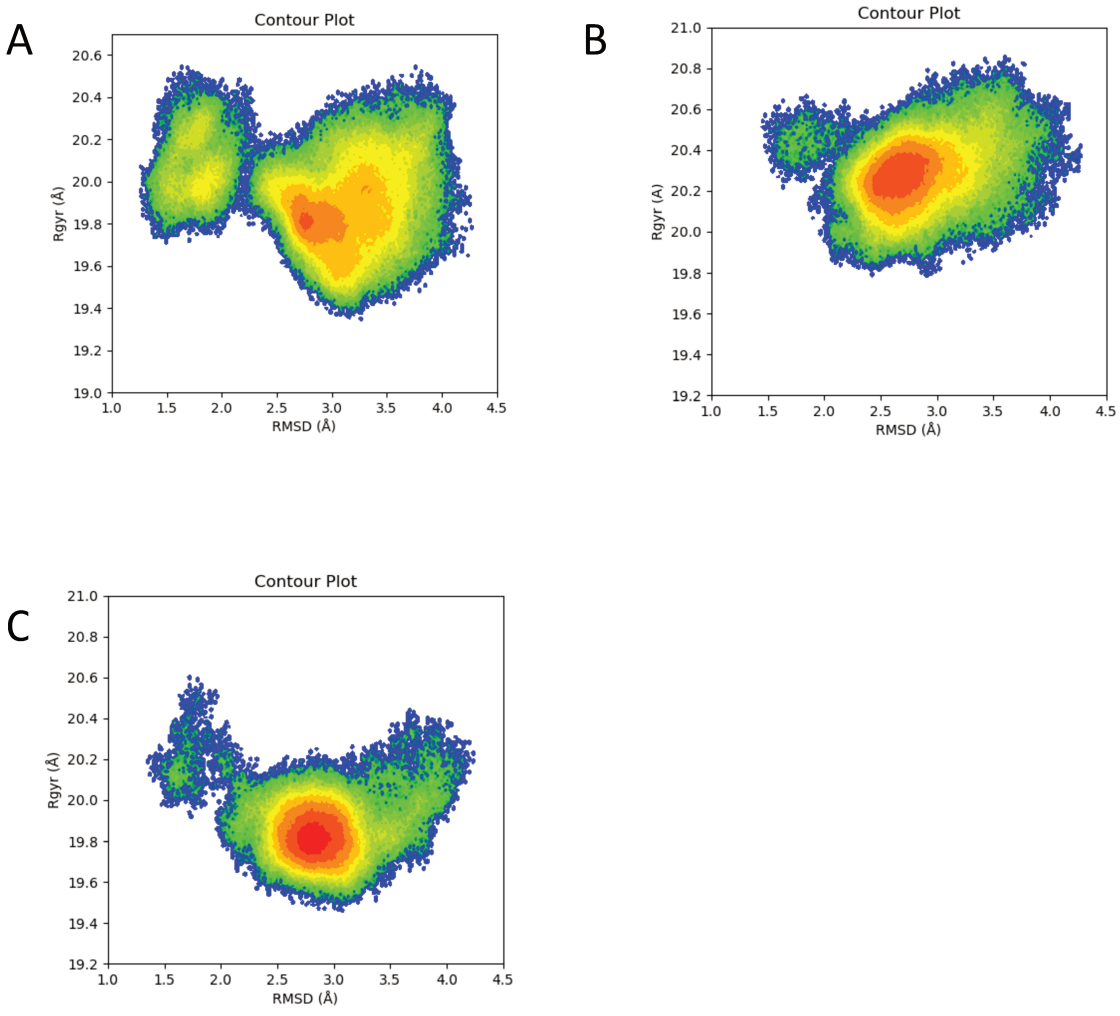


Figure 34. RMSD – RGYR plots free energy landscape for individual replicas for the T475M system. A) replica 1, B) replica 2, C) replica 3.

The mutant T475M (Fig. 34) is plotted on a scale of 19.2 – 21.0 Å for the radius of gyration and 1.0–4.5 Å for the RMSD. All three replicas show larger conformational sampling of trajectories with respect to the WT protein, with irregularly shaped plots.

The mutant F310S (Fig. 35) is plotted on scales 19.0 – 20.4 Å for the radius of gyration and 1.5 – 4.0 Å for the RMSD. All three plots suggest a more flexible LBD, with the highest RMSD values going higher than 4 Å.

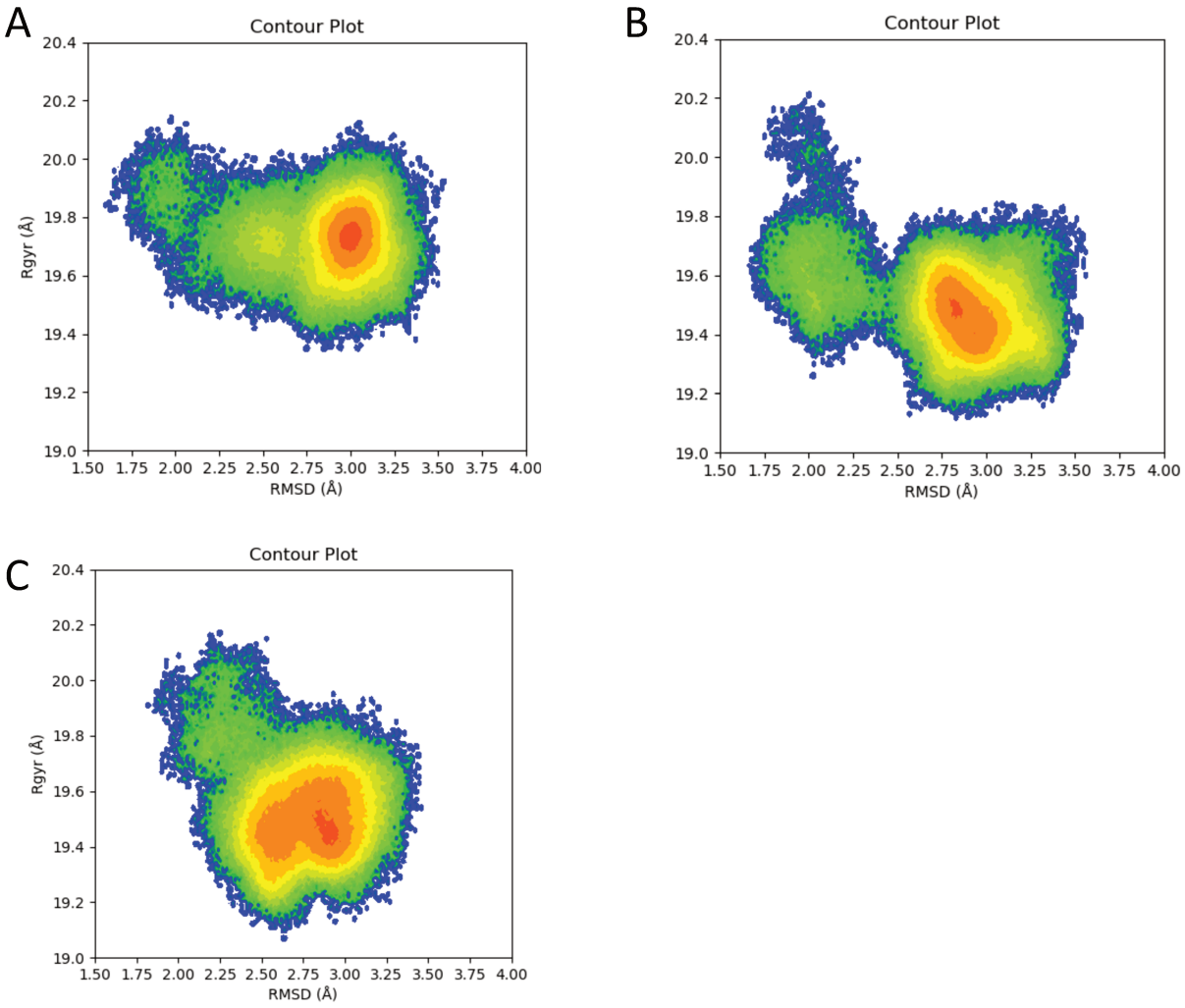


Figure 35. RMSD – RGYR plots free energy landscape for individual replicas for the F310S system. A) replica 1, B) replica 2, C) replica 3.

The comparison of these different plots reveal that the mutant forms explore more conformational space even though their most populated wells are relative constrained. The WT apo shows more conformational exploration that the WT complexed to the agonist GW1929, which demonstrates the stabilizing effect of the ligand on the LBD.

As per our ensemble approach, for subsequent analysis, structures were extracted from the most populated wells. The areas shaded in deep red/deep red correspond to those regions of the FEL that were most populated. Defining limits in terms of both RMSD and RGYR, (not shown on plots) a CHARMM script was used to extract structures from the most populated wells. The values for extracting structures for individual replicas are summarized in Table 3.

For example, for first replica of the WT apo system, we extracted structures that had RMSD values between 1.6 and 1.75 Å, and a RGYR between 19.8 and 19.9 Å. For each replica, the number of structures which satisfied these conditions is reported in the Table 3. From each group of representative structures (n° of extracted structures, Table 3), one hundred structures were extracted for further analysis. As explained earlier, a dynamics trajectory of all the structures that satisfied the RMSD and RGYR criteria were selected. A dynamics trajectory was constructed, and 100 structures were extracted from this trajectory.

Table 3. Values of RMSD and Radius of gyration used to define the most populated wells, and the total number of structures extracted for each system.

		replica 1	**replica 2**	**replica 3**
WT apo	**RMSD (Å)**	1.6 – 1.75	1.8 – 1.95	1.9 – 2.15
	R_{gyr} (Å)	19.8 – 19.9	19.77 – 19.82	19.75 – 19.9
	n° of structures extracted	48 897	14 888	78 811
WT - GW1929	**RMSD (Å)**	2.67 – 3.0	2.12 – 2.3	2.22 – 2.47
	R_{gyr} (Å)	19.56 – 19.65	19.68 – 19.8	19.7 – 19.85
	n° of structures extracted	5 100	84 714	90 671
F310S	**RMSD (Å)**	2.97 – 3.0	2.81 – 2.87	2.85 – 2.98
	R_{gyr} (Å)	19.7 – 19.76	19.4 – 19.5	19.4 – 19.5
	n° of structures extracted	11 970	23 611	30 885
T475M	**RMSD (Å)**	2.7 – 2.84	2.5 – 2.8	2.69 – 2.97
	R_{gyr} (Å)	19.76 – 19.84	20.2 – 20.34	19.7 – 19.88
	n° of structures extracted	14 972	77 056	123 965

3.3. Normal Mode Analysis

Normal Mode Analysis (NMA) was done for each set of 100 structures extracted from the individual replicas. Each replica is treated separately. The NMA calculation was carried out using the *Vibran* module of the CHARMM program. Each structure was subjected to an energy minimization using the adopted basis Newton Raphson (ABNR) minimizer in the CHARMM program until the GRMS, a measure of the closeness to the local energy minimum, was less than the tolerance of 1 x 10^{-6} kcal/(mol.Å). The minimized structure was then used for the NMA calculation. The output file of the *vibran* calculation contains *3N* number of frequencies, where *N* is the number of atoms, each associated to a normal mode (Table 4). To check if the calculation was executed properly, we verified the values of first 6 modes, which, in the CHARMM output, correspond to translational and rotational degrees of freedom. Their values should be close to zero. If this condition was satisfied, the results from the normal mode analysis frequencies associated with normal modes are used for further calculation of atomic fluctuations, IR spectra, and correlation plots.

Table 4. Number of atoms and number of normal modes of each system.

	n° of atoms	***3N* normal modes**
WT apo	4 488	13 464
WT - GW1929	4 559	13 677
F310S	4 479	13 437
T475M	4 543	13 629

3.3.1. Computed fluctuations

From the normal modes calculated for each set of 100 structures from each replica of each system under study, we calculate the RMS fluctuations using Eq.21 in Methods. The low frequency normal modes capture the collective motions, which are often associated with functionally important movement. Calculating the RMS fluctuations provides a verification of

the correctness of the calculated result. Comparison to the fluctuations calculated from molecular dynamics simulations should be favorable.

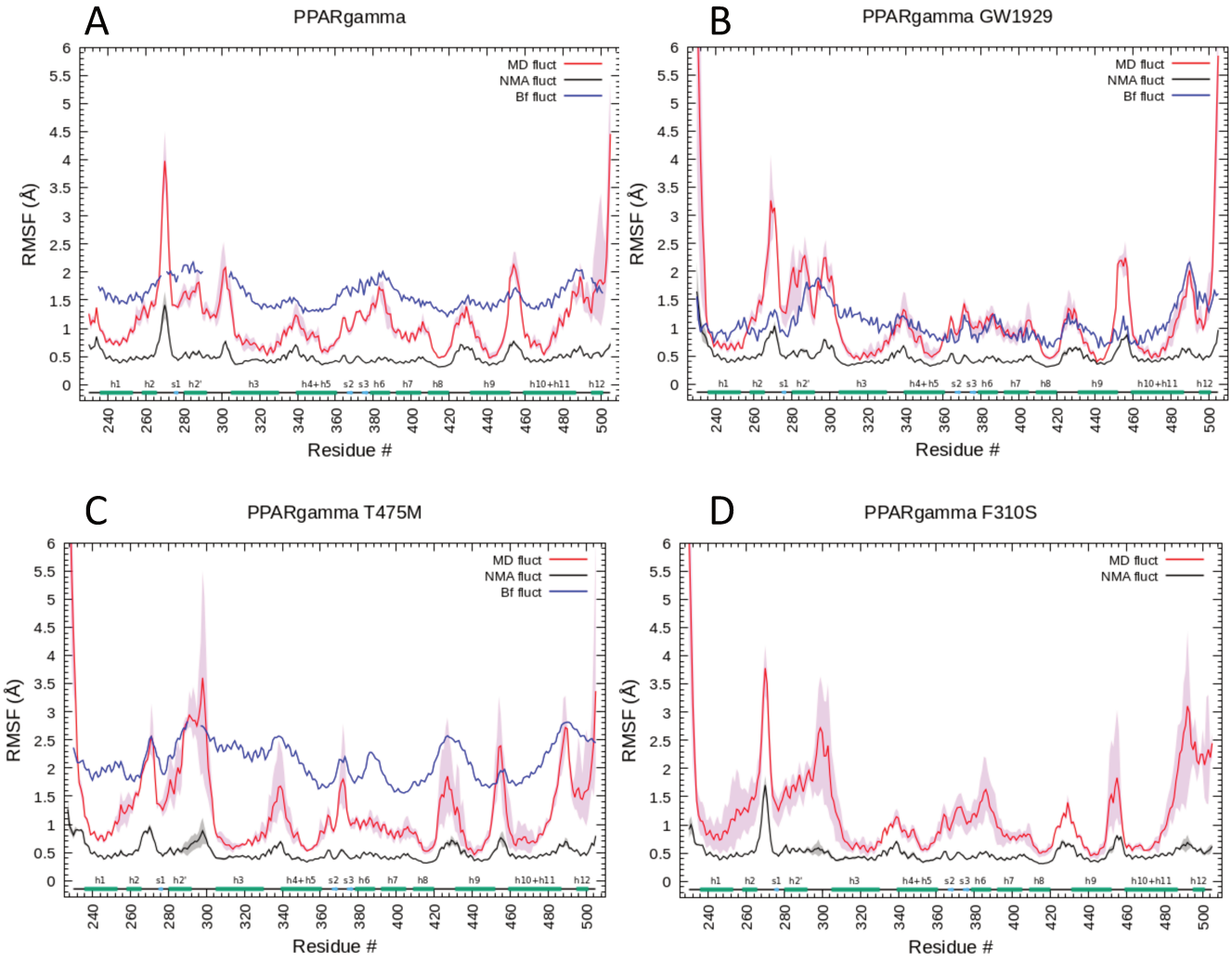


Figure 36. RMS fluctuations calculated from MD trajectories, NM calculations and crystallographic B factors. A) WT apo; B) WT with GW1929; C) T475M; D) F310S.

To compare different results, three different calculations of RMS fluctuations are represented on one plot: the RMSF calculated from MD simulations, from normal mode analysis and from the crystallographic B- factors (Fig. 36). Crystallographic B-factors, also known as temperature factors or Debye -Waller factors, represent atomic displacements from X- ray crystallography data (C. L. Brooks, Karplus, and Montgomery Pettitt 1988). They provide information on positional flexibility and dynamics of atoms in crystallographic structures of proteins. The

information they give should not be over-interpreted, since their accuracy is affected by various factors, such as crystal defects, diffraction data quality or experimental conditions (Na, Hinsen, and Song 2021; Mlynek, Djinović-Carugo, and Carugo 2024). Nevertheless, they can still be useful to assess the internal mobility or rigidity of proteins, which can help understand protein function.

For each system, a plot of all three RMSF (Å) calculations as a function of residue is provided in Fig. 36. The fluctuations are calculated from different sources, from MD simulations, from NMA and from crystallographic B-factors. Their comparison shows how they can be similar - or different - for a particular protein system. Note that, in Fig. 36D for the mutant F310S, there are no no fluctuations from experimental B – factors, as this mutant does not have a corresponding crystallographic structure, rather it is a model constructed based on the structure of the apo WT.

The MD and NMA fluctuation are represented as an average of 3 replicas, with the shaded area of the same color around the curve being the standard deviation value. The first thing we notice is the scale (Å) at which these fluctuations take place. The highest values in fluctuations are seen in MD simulations, going up to 6 Å in the case of the N-terminal end of PPARγ in the cases of the mutant structures and the structure complexed with GW1929. Next, the fluctuations from B factors ranging from 0.8 - 3 Å, and finally the NMA fluctuations with the smallest values, of around 0.5 Å for the stable secondary structure elements. It comes as no surprise that the values for these fluctuations fall in this order. The crystallographic physical and chemical environment of individual proteins restricts their movement, accounting for the smaller values of atomic displacements with respect to the simulation results. In the cases where a protein region moves significantly, there will be no electron density, and consequently no B- factors. This is depicted as a gap in the curve on the plot in the case of PPARγ WT apo and mutant T475M. Regarding the fluctuations calculated from normal modes, recall that these motions are small harmonic displacements around an atomic equilibrium, so it is expected that they will not have high values. But, from these figures we can see a consistency in the pattern of motions. On each figure, we see that the three curves follow the same trends, with lower fluctuation values for more stable regions of secondary structure elements, and higher values

for more flexible protein sections. This provides confirmation of the correctness of our simulations and suggests that the trajectories and structures extracted from them can be used for further analysis.

3.3.2. Computed Infrared spectra

From the results of the normal mode calculations, we calculated the mid- and far-IR spectra for each replica of each system under study using Eq. 25. This calculation used the results from the normal mode calculation presented above. The full set of normal modes was used. Following the protocol of ensemble normal modes presented in chapter III, one infra – red spectrum was calculated for each structure extracted from the MD simulations of each replica, meaning 100 spectra for each replica. From the output file, we extracted frequencies and dipole derivatives associated to each vibrational normal mode. The frequencies and intensities were then binned into 4000 bins which allowed us to merge the different spectra. In order to simulate an experimental IR spectrum, we applied a convolution factor. A value of ten for the smoothing factor was chosen since it gave the best qualitative result and corresponded to the value used in the experiments.

The experimental results were generated by Professor Petra Hellwig and Dr. Filipa Seica. With their permission, the data are reproduced here for comparison to the calculated results. The experimental far-IR spectra of PPARγ in apo WT form, holo form bound to the agonist GW1929, and two apo mutant forms, one with the activating mutant T475M and the inactivating mutant F310S were recorded by following the protocol described in Annex section at the end of this chapter IV. The results are shown in the Fig. 37, where one observes that the measured far-IR spectra all have a similar shape. One also notices that the calculated peaks are generally broader. There is one very intense and broad peak that appears in the frequency range [100–250] cm^{-1} and a second peak that is present in the frequency range [500–600] cm^{-1}. A smaller peak appears in the frequency range [400–450] cm^{-1}. The signal at 140 cm^{-1} observed in the apo WT form shifts to higher frequencies [162 cm^{-1}] in the holo form spectrum. This blue shift in the far-IR is likely to be a consequence of the interaction between the ligand and

the protein. With respect to the apo WT protein, the first peak for both mutants also shows a blue shift (the shift towards higher frequencies). The same trends are observed in the higher frequency peaks in the ranges of range [500–600] cm^{-1} and [400–450] cm^{-1}.

Mid-IR range from [1500-1800] cm^{-1} shows very little effect of changes due either to ligand binding or point mutations. For the holo form in complex with GW1929, the region [1100-1200] cm^{-1} shows significant peak shifts, and the apo mutant forms show slightly less shifts. But overall, ligand binding seems to affect the low-frequency region more than the mid- and amide regions of the spectrum.

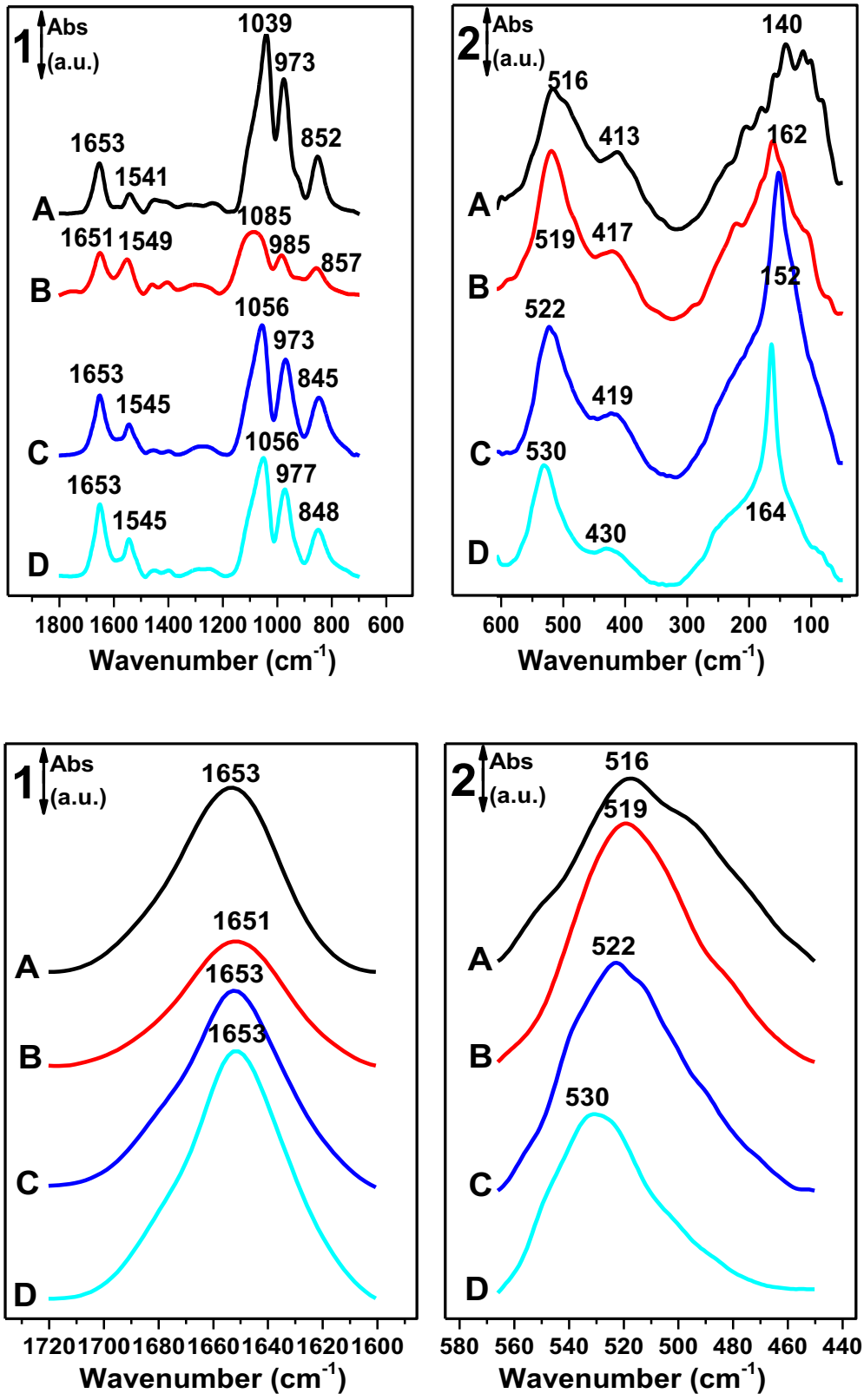


Figure 37. Shown here are the experimental mid- and far- IR absorbance spectra (top) for WT APO (black), WT+GW1929 (red), and two mutant forms T475M (blue) and F310S (cyan). The bottom figure shows the absorbance spectra for the amide I and amide VI bands. Reproduced with the permission of Professor Petra Hellwig and Dr. Filipa Seica.

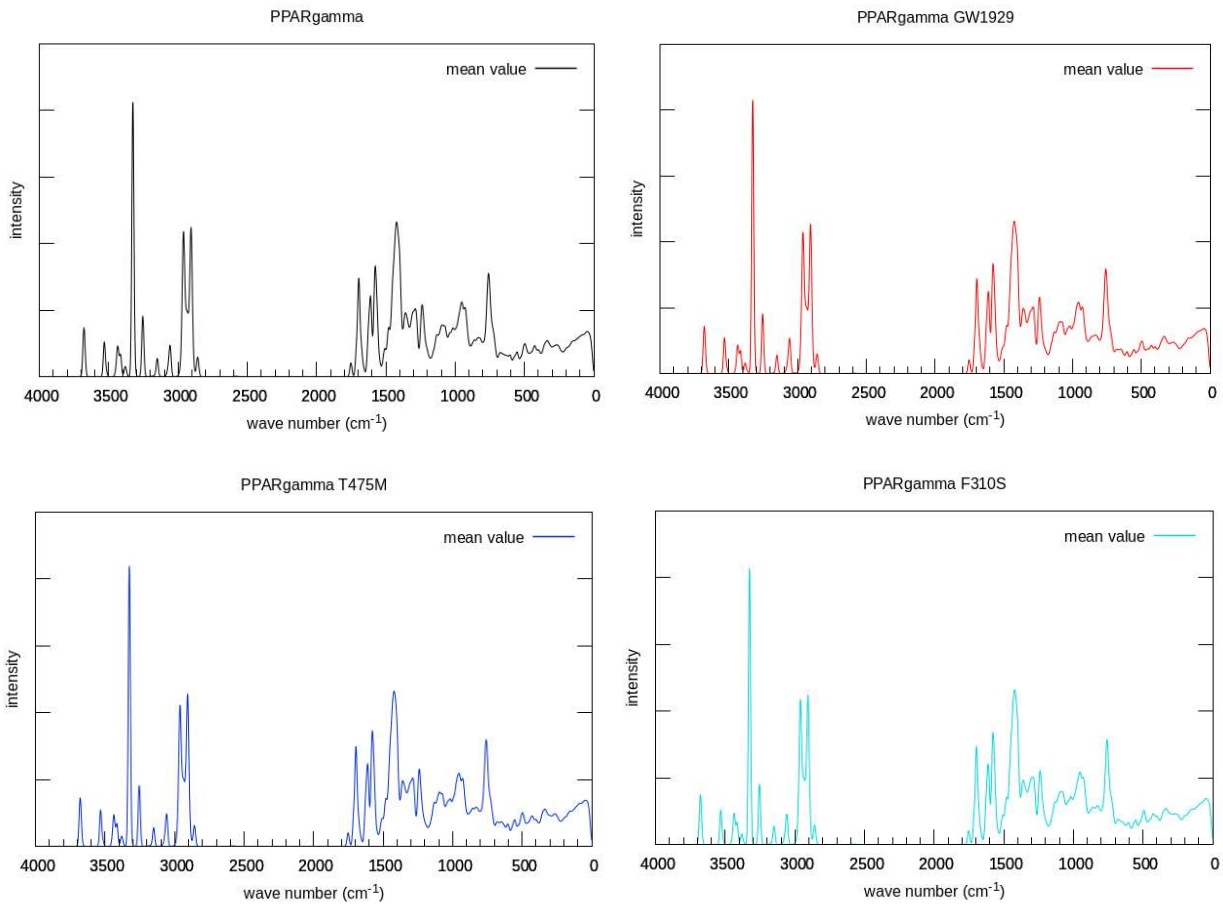


Figure 38. Computed IR spectra, from normal mode calculations, of four biological systems: WT apo (black), WT – GW1929 (red), mutant T475M (blue) and mutant F310S (cyan).

Figure 38 shows the full IR spectra calculated for all four systems. Plotted are wave numbers (cm^{-1}) ranging from 0 to 4000 cm^{-1} and spectral intensities (from right to left). We distinguish three ranges of wave numbers, the far – IR, with $\omega < 700$ cm^{-1}, the mid – IR (from 1200 to 1700 cm^{-1}) and near – infra red (from 3000 to ~ 3300 cm^{-1}). The 0 – 4000 cm^{-1} range plots are given as a qualitative assessment of the validity of the results, ensuring that the entire spectrum is calculated. The different regions of the spectrum correspond to different molecular motions corresponding to different vibrational frequencies ranging from entire collective protein motions (low frequencies) to more localized bond vibrational motions (high frequencies). The mid-IR reflects motions of secondary structural elements. We note that the high frequency

vibrations (near IR range, high wavenumbers) are nearly identical across all proteins. Our focus here is on the low - frequency motions that are unique for each system.

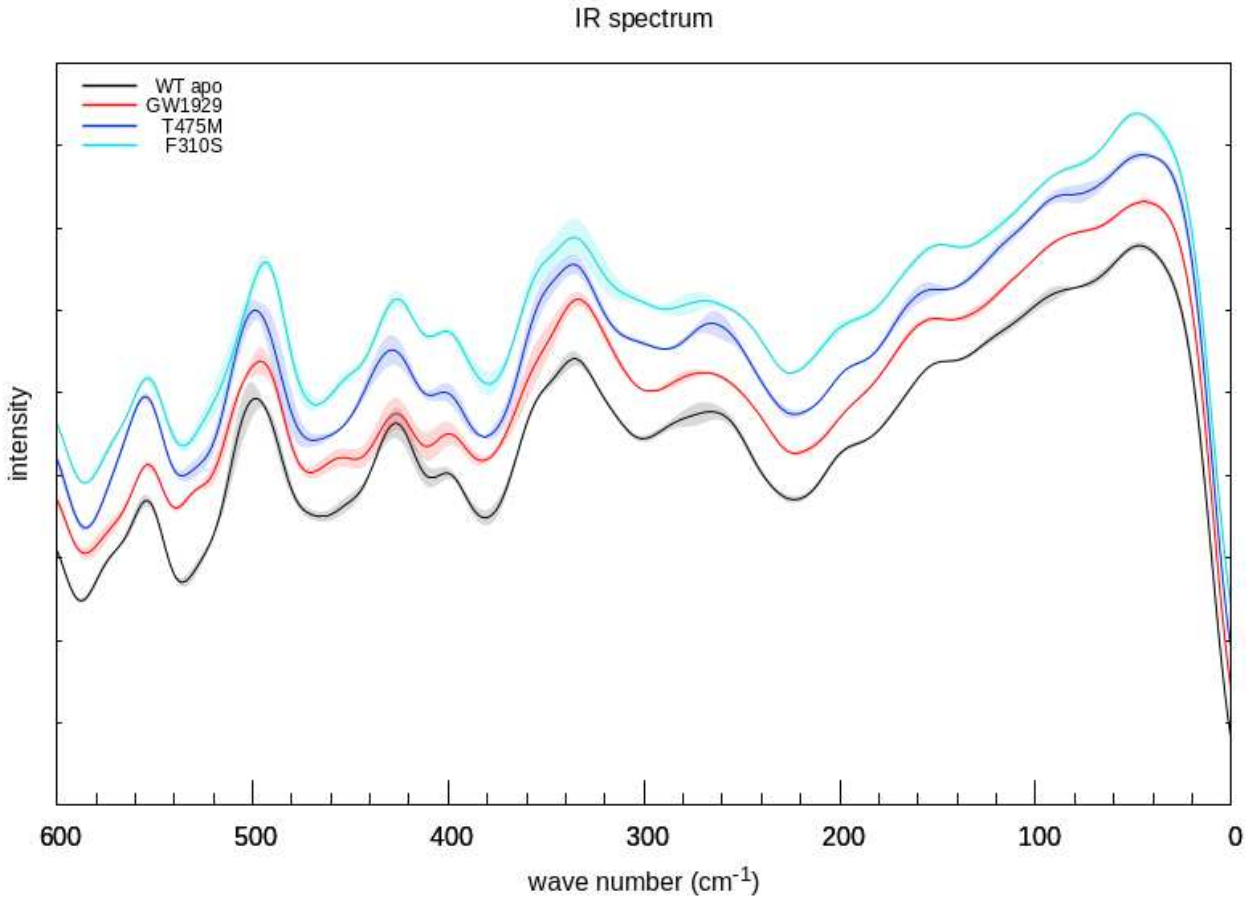


Figure 39. Computed IR spectra, ω < 600 cm^{-1}, from normal mode calculations, of four biological systems.

As the primary interest in this work are the collective motions of the different protein systems, the low frequency regions were plotted in one plot to facilitate comparison; the low-frequency region lies between $\omega = 0$ - 600 cm^{-1} (Fig. 39). The intensity values are not indicated as they are not of primary interest here, only peak positions, shifts and shapes. Averaging over the replicas results in the mean values of different spectra being plotted as a solid line and the standard deviations being represented as shaded areas around curves in the same color. The relative position of curves is displayed with an increment along the y-axis to facilitate the visual comparison. Focusing on the lowest frequency peak of the spectra, we find they are centered around ω = 47 cm^{-1} (WT apo), ω = 45 cm^{-1} (WT - GW1929), ω = 45 cm^{-1} (mutant T475M), ω

= 48 cm^{-1} (mutant F310S). Small shifts in these lowest frequency peaks are also observed experimentally. Moving to higher frequencies, we see similar trends between the experimental and calculated spectra. We see that the descending slope after the first peak (after 100cm^{-1}) shows more variation in the apo WT than in the holo WT with the ligand GW1929. Experimentally, the descending slope of the apo WT has more features than the holo WT. Similarly, the next set of peaks in the calculated spectra find correspondence in the experimental peaks with similar shifts.

3.3.3. Secondary structure analysis

By analyzing the amide I and amide VI infra-red bands, which are sensitive to secondary structure motifs, infrared (IR) spectroscopy provides a direct experimental measurement of protein secondary structure content. Following the methodology described in the Annex of this chapter, the secondary structure content of PPARγ in the different forms studied here, that is WT apo and holo forms, and two mutant structures, T475M and F310S, was determined by the experimental team of Professor Petra Hellwig. The results are presented in Table 5.

Table 5. Secondary structure content as determined by Mid-IR spectroscopy from analysis of the amide I and amide VI bands. Shown are the percentages for each type of secondary structure and, in parentheses, the wave numbers of the peaks analyzed.

Amide I	**WT [cm^{-1}]**	**WT_GW1929 [cm^{-1}]**	**T475M [cm^{-1}]**	**F310S [cm^{-1}]**
% α-helix	48 (1654)	57 (1652)	69 (1653)	66 (1652)
% antiparallel β-sheet	27 (1638, 1687)	24 (1634, 1686)	18 (1631, 1689)	21 (1631, 1689)
% β-turns	22 (1671)	16 (1671)	11 (1677)	11 (1677)
% intermolecular aggregate % β-sheet	3 (1622)	3 (1616)	2 (1615)	2 (1615)

Amide VI	**WT [cm^{-1}]**	**WT_GW1929 [cm^{-1}]**	**T475M [cm^{-1}]**	**F310S [cm^{-1}]**
% α-helix	59 (518)	69 (502, 521)	54 (510, 525)	70 (502, 518 532)
% β-sheet	10 (539)	22 (538)	17 (540)	18 (548)
% β-turns	5 (552)	4 (558, 567)	6 (555)	6 (564)
% other contributions	26 (472, 478, 493)	6 (481)	19 (477, 493)	6 (485)

Here we focus only on α-helical content, given that α-helices are the predominant structural motif in the NR LBD. The analysis of the experimental data (Table 5) indicates first that the analysis of the two bands, amide I and amide VI, yields different absolute values of percent content. However, for the most part, the analysis of the two bands shows the same general trend from one form to the other. Only T475M shows a deviation from the pattern. In the analysis of

the amide I band, T475M shows a greater α-helical content with respect to apo and holo WT forms, while in the analysis of the amide VI band, it shows lower α-helical content than apo and holo WT forms. Comparing to the values calculated from the crystal structures used in this study (Table 6), the percent of α-helical content determined from the IR experiments are generally lower, while the percent β-sheet are considerably higher. The reasons for these discrepancies are not clear.

Table 6. Percent of secondary structure calculated from the crystal structures used in this study. There is no experimental structure for the F310S mutant.

Crystal Structure	α helix (%)	% β sheet (%)
Apo WT (7WOX)	69.5	4.7
WT_GW1929 (6D8X)	76.8	4.7
T475M (6FYZ)	74.5	4.7
F310S	-	-

The secondary structure content of PPARγ was calculated from the molecular dynamics simulations using the Stride program (Heinig and Frishman 2004b) piloted by MDAnalysis (R. J. Gowers et al. 2016) and an in-house python script developed during the course of this thesis. For each replica of each system, the percentages of α-helices and random coils, and β-sheets were calculated. For each system, the results were averaged over the three replicas (last column). The results are shown in Table 7.

Table 7. Secondary structure content calculated from the molecular dynamics simulations. The background colors are associated to the colors of the far-IR spectra shown in Fig. 40.

		rep1	rep2	rep3	mean value
WT apo	α helix (%)	73.4	72.2	74.09	73.2 ±1.0
	β sheet (%)	3.9	4	4	4.0 ±0.0
GW1929	α helix (%)	75.8	77	76	76.3 ±0.6
	β sheet (%)	4.2	3.7	4.2	4.0 ±0.3
T475M	α helix (%)	77	78	77	77.3 ±0.6
	β sheet (%)	4.6	4	4	4.0 ±0.0
F310S	α helix (%)	72.7	72	71.7	71.9 ±0.2
	β sheet (%)	4.3	4.3	4.1	4.2 ±0.1

In general, we see that the percentages calculated from the simulations are more in line with the values calculated from the experimental crystal structures than from the mid-IR experiments. However, for the α helices, we see that, for the most part, the trends observed in the IR results are followed by the simulation results. In particular, the percentage of α helix increases when going from the wt-APO to the wt-HOLO, which could reflect stabilization of H12, as well as other regions around the ligand binding pocket. A further increase in α-helical content is observed in the T475M mutant. An increase is also observed in the results from the mid-IR measurements but is not observed in the crystal structures. This increase in α-helical content in the T475M mutant could reflect further stabilization of H12 due to interaction with the methionine 475. As seen in the original simulations of T475M, and in the subsequent X-ray crystal structure, the C-terminal tyrosine (Y505) interacts with the sidechain of 475M, leading to stabilization of the H12 helix (Natacha Rochel et al. 2019). Concerning the F310S mutant, the mid-IR results show that the α-helical content in this mutant is larger than in the WT apo protein, but the simulations show that the content is essentially the same. This could

reflect the fact that this mutant structure was constructed using the WT apo structure as a template.

In Fig. 40, we show the percent secondary structure at three points throughout the simulation- beginning, middle and end. This shows that over the course of the simulations, the secondary structure content does not change significantly, although there is a slight increase in α-helix content for the WT holo and the T475M proteins. The percent content of β-sheet remains low, consistent with the values calculated from the crystal structures, but less than the values determined by the mid-IR measurements.

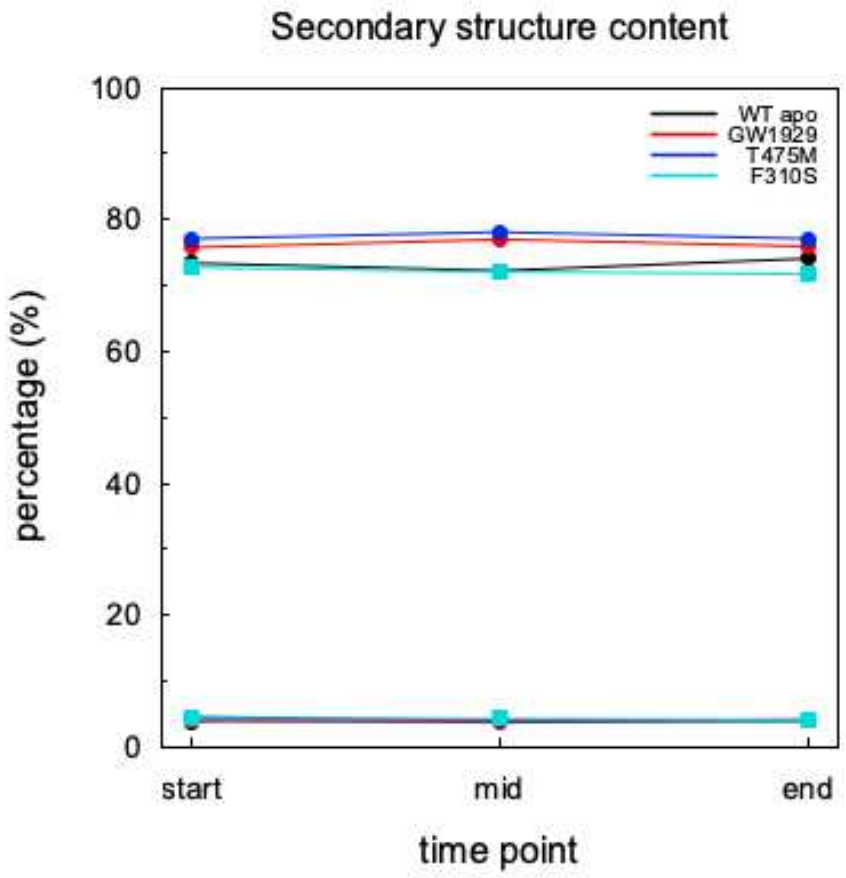


Figure 40. Secondary structure content as a function of time point (values between 70 - 80 % represent α helix, and around 4% are β sheets). The results are averaged over the three replicas. Multiple structures are taken from 3 time points: beginning, middle and at the end of each simulation.

When compared to experimental IR spectroscopy data, these results further validate the molecular dynamics simulations. The discrepancies between calculated and experimental data can arise from factors such as differences in hydration and environmental conditions between the experiments and the molecular dynamics simulations.

3.3.4. Correlated motions

Through the comparison of the measured mid- and far-IR spectra to those calculated from our molecular dynamics simulations, we established that the atomic motions in our simulations represent well the low-frequency dynamics of the different forms of PPARγ studied here. This then allows us to examine the collective motions and elucidate the effects ligand binding and point mutations. Following our ensemble NMA approach (chapter III), we computed the correlated motions of the Cα atoms for each of the 100 representative structures extracted for each replica of each system. For these calculations, we included all *3N* normal modes, but in chapter III, we showed in Fig. 23 that the correlated motions themselves are well represent by modes up to 200 cm^{-1}. The results are shown in Figs. 41 for the WT proteins and Fig. 42 for the mutant proteins. The general aspect of the correlation maps for all four systems is quite similar. We notice differences mostly in the intensities of the correlations. Along the diagonal, we have completely correlated motions, and the rest of the plot represents correlated (red) or anti-correlated motions (blue). On all four plots, extending perpendicularly from the diagonal, are correlated regions representing β sheet, while α helices run along the diagonal.

In the correlated motions of the apo WT (Fig. 41 top) calculated and averaged from the NMA, we see correlations between different secondary structure elements, in particular H12 and H3, as well as the ω loop region following H2'. One also sees correlated motions between H4-5 and H8-loop-H9. The strands of the β-sheets show significant correlation between them, as well as with other regions of the protein. The PPARγ WT presents other prominent islets of red. These represent correlations of β1, β2, and β3 strands in the β - sheet, H1 and the C-terminal end of H3, H1 and H9, and between H5 and H10-11. These correlations can be explained by the spatial proximity of these residues and their immediate interaction. Another

correlated region is found between residues of H12 and the Ω - loop. Highly anticorrelated motions are seen less across the plot. Still, smaller anticorrelations are concentrated in the regions of H1 and the region between H2 - H2', H1 and H4, H1 and the loop H8 - H9, the end of H2 and the end of H9, Ω - loop and H5, H4 and end of H9. The region between H11 and C - terminal is anticorrelated with helix H4 and with the region of H8 - H9 loop and N - terminal of H9.

The correlations of PPARγ - GW1929 (Fig. 41 bottom), have a highly similar aspect, but with significant fading of anticorrelations. For example, the anticorrelated motions depicted in darker blue shade, between H1 and H8 - H9 loop, are completely softened. H12 shows significant correlation with H3. Putting H12 in a different conformation in the holo-WT simulation has a significant effect on the correlated motions of the ligand binding domain.

In a previous study (Natacha Rochel et al. 2019), it was shown that the T475M mutant, even in apo form, could activate gene transcription to a degree similar to the WT holo form. Concerning the correlated motions of the T475M mutant structures (Fig. 42 top), we see, in fact, strong similarities to the holo WT structure, especially if we look at the anti-correlated regions. For example, the mutant T475M displays anti-correlated patterns and intensities that are similar to the WT - GW1929 system. The T475M mutant shows some additional correlated motions not seen in the WT holo, for example, H4-5 and H12. The correlated motion of the mutant F310S is similar to those of the WT apo form (Fig. 42 bottom).

While it was previously shown that ligand binding alone can affect the correlated motions (Fidelak et al. 2010a), these molecular dynamics simulations have identified both sequence- and structure-specific differences in correlated motions involving different secondary structure elements. We see that differences in the local conformation and sequence can have profound effects on the structural dynamics, interaction networks, and long-range correlations in this protein domain.

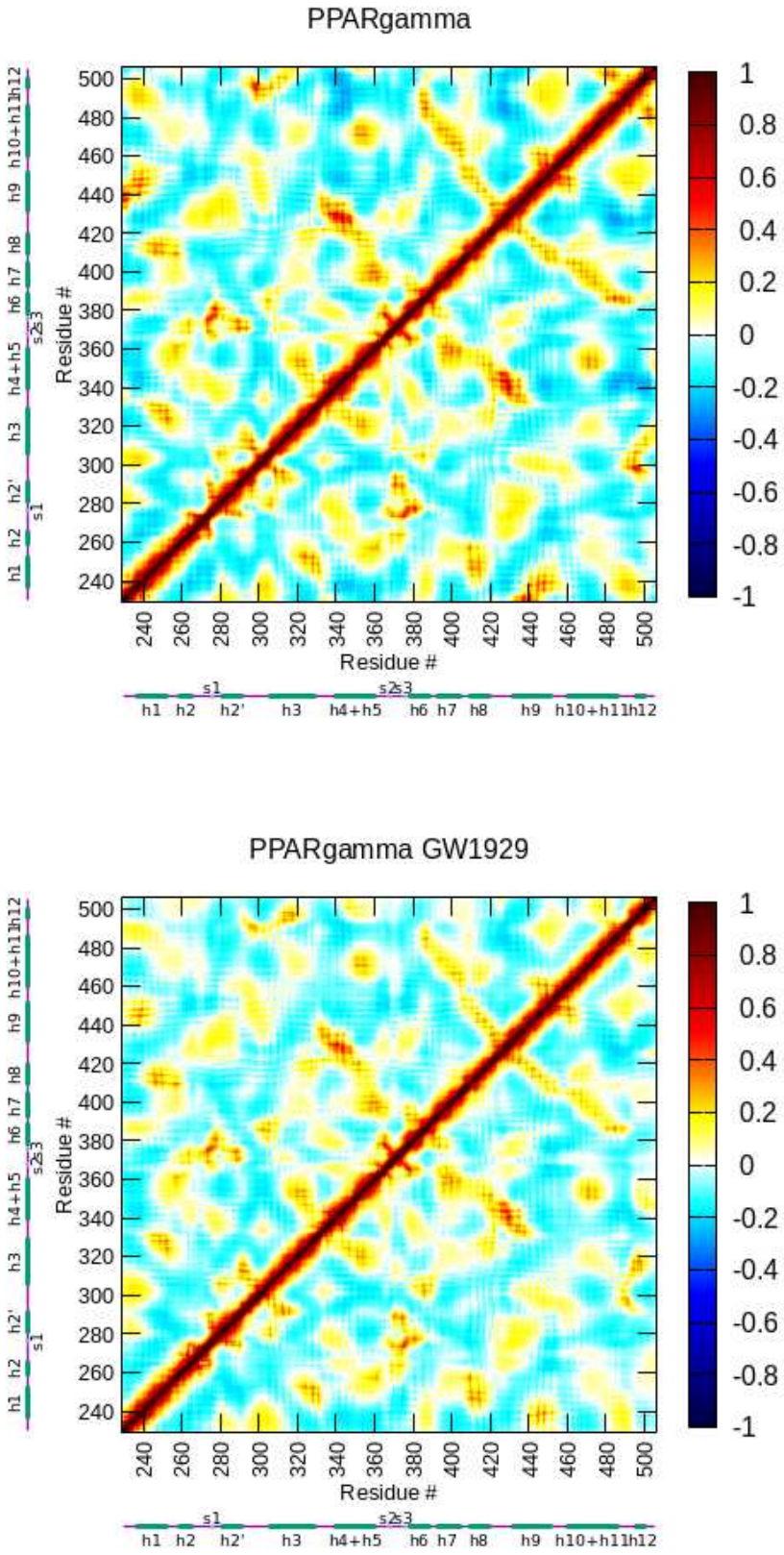


Figure 41. Correlated motions calculated from normal modes for the WT apo (top) and WT holo systems (bottom).

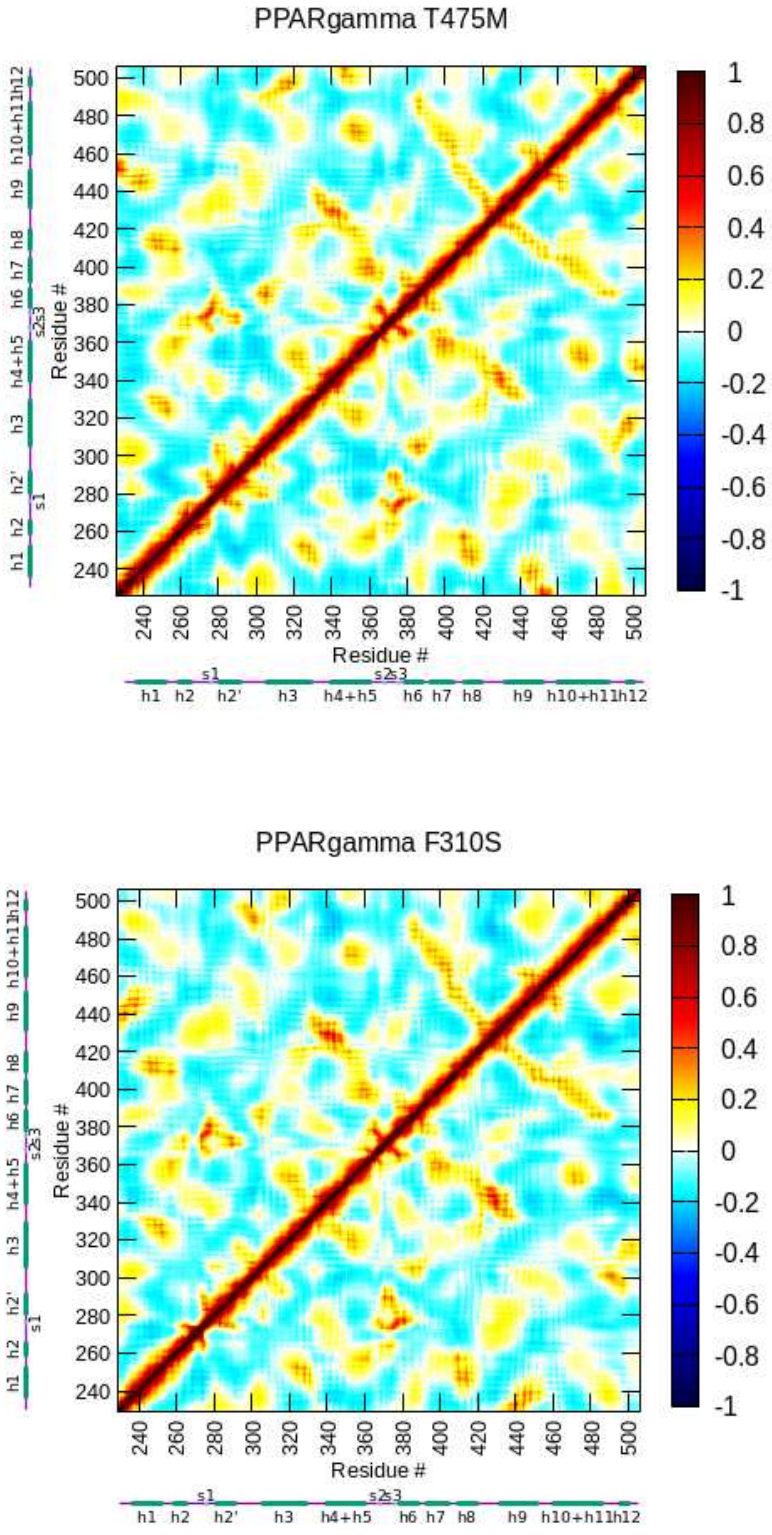


Figure 42. Correlated motions calculated from normal modes for the mutant T475M (top) and F310S systems (bottom).

3.3.5. H-bond analysis

Shifts in the mid- and far-IR spectra can be attributed, in part, to changes in hydrogen bonding. In the mid-IR, these are changes to C=O vibrations in secondary structure. Here we carried out an analysis of the hydrogen bonds present in ensembles of structures used for the NMA. All hydrogen bond pairs were chosen, including both backbone and sidechain hydrogen bonds. Analysis of the hydrogen bonds can provide insight into protein stability and dynamics.

To identify the hydrogen bonds populated by the structures used for the normal mode calculations, we took the 100 structures of each replica of each system and generated a dynamics trajectory that was subsequently analyzed using the CHARMM *coor hbond* module. The selection was made for the entire protein chain, which gave results for backbone-backbone and hydrogen bonds involving sidechains. Backbone hydrogen bonds are essential for the stability of secondary structure elements, such as α helices, β sheets and β turns. These hydrogen bonds form between the carbonyl oxygen of one amino acid and the amide hydrogen of another. The analysis also identified hydrogen bonds between side chains and between side chains and backbone atoms. These later hydrogen bonds can contribute to the stabilization of the overall 3D fold of the protein. Water molecules were not included in this analysis. The hydrogen bonds were calculated for each replica using a distance cutoff of 2.4 Å and no angle limitation, as defined by the default parameters in the CHARMM program. Changes in the hydrogen bonds of the protein could manifest itself in shifts of the low frequency IR peaks.

3.3.5.1. Backbone H-bond analysis

From the analysis of the 100 structures per replica per system used in the ensemble normal mode analysis, we calculated the average number of backbone hydrogen bonds (Fig. 44). We see that the WT apo system has the fewest number of backbone H-bonds, while the other systems have larger numbers. The WT holo system has, on average, the highest number of

hydrogen bonds in the ensemble of structures extracted from the molecular dynamics simulations.

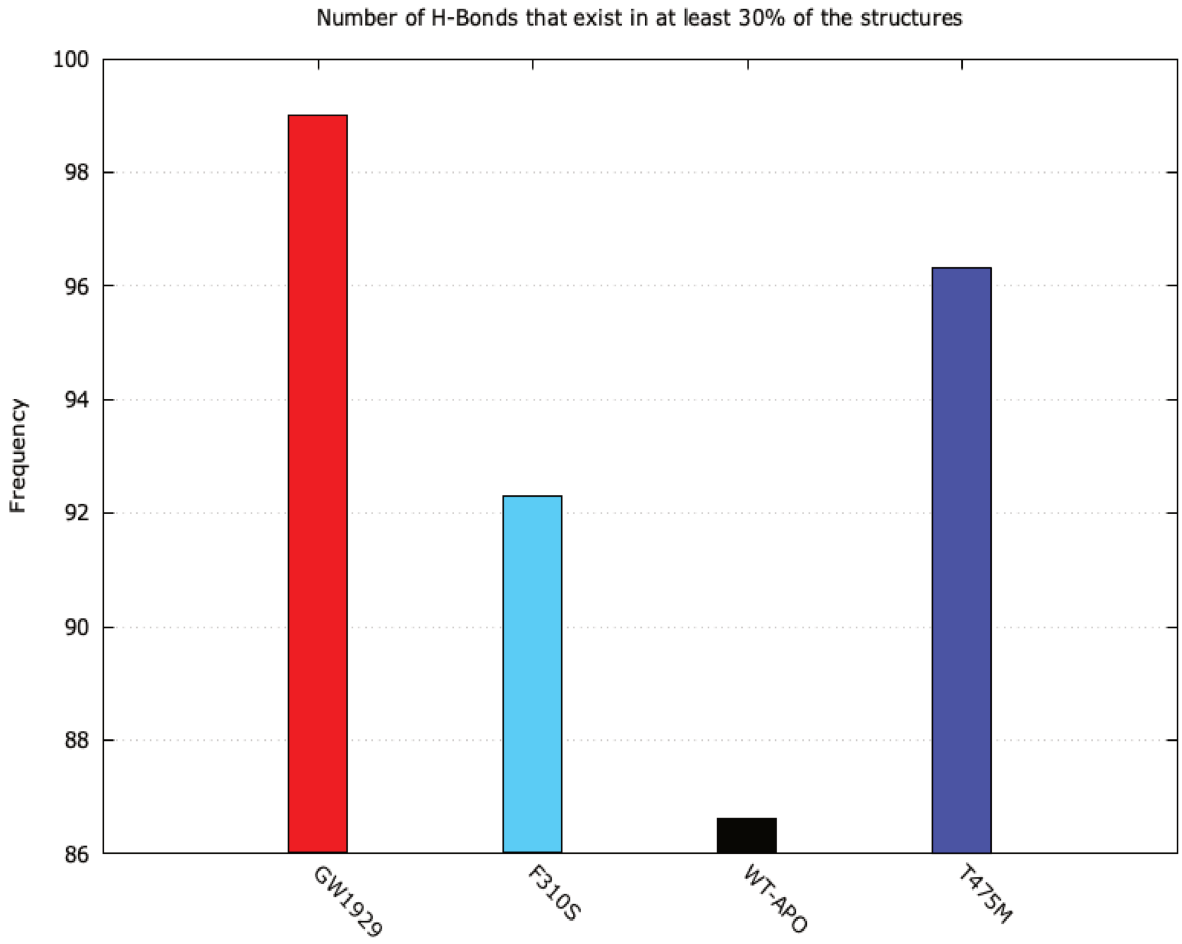


Figure 43. Average number of backbone hydrogen bonds found in at least 30% of the structures of each ensemble of structures extracted from the molecular dynamics simulation. Averaged over 3 replicas.

3.3.5.2. Sidechain H-bond analysis

Through a combination of perl and python scripts, we extracted the hydrogen bonds formed outside of secondary structure elements and displayed them on the average structure of the protein calculated over the 100 structures. H-bonds for each replica are displayed on the same structure.

In Fig. 44 top, we show the hydrogen bonds calculated from the structure ensembles used for the normal mode calculations of the WT apo system. Only the backbone structure is shown,

and the hydrogen bonds are shown as thin lines. Some of the lines appear to be very long for hydrogen bonds, which is a consequence of putting the bonds on the average structure as opposed to individual structures. Indeed, the long lines represent regions of high conformational flexibility. In this present case of the WT apo structure, this suggests that the C-terminal helix is somewhat flexible. One does not see any coupling by H-bonds between H3 and the Ω loop region.

In Fig. 45 bottom, we see the results of the H-bond analysis for the WT complexed to the agonist GW1929. The H-bonds clustered around the C-terminal end are relatively short, which suggests a less flexible C-terminal H12 helix. We see, in addition, H-bonds that couple the H11-H12 loop to H3 and to the Ω loop region. So, there is more significant coupling than what is observed the apo form. In addition, we see that His494 at the N-terminal end of H12 is involved in H-bond formation with Gln314, Lys293, His294 and Ser 492. His 494 is a residue of therapeutic interest since its mutation to Tyr has been associated to inactivation of PPARγ in baldder cancer (Coutos-Thévenot, Beji, Neyret-Kahn, Pippo, Fontugne, Osz, Krucker, Dos, et al. 2019).

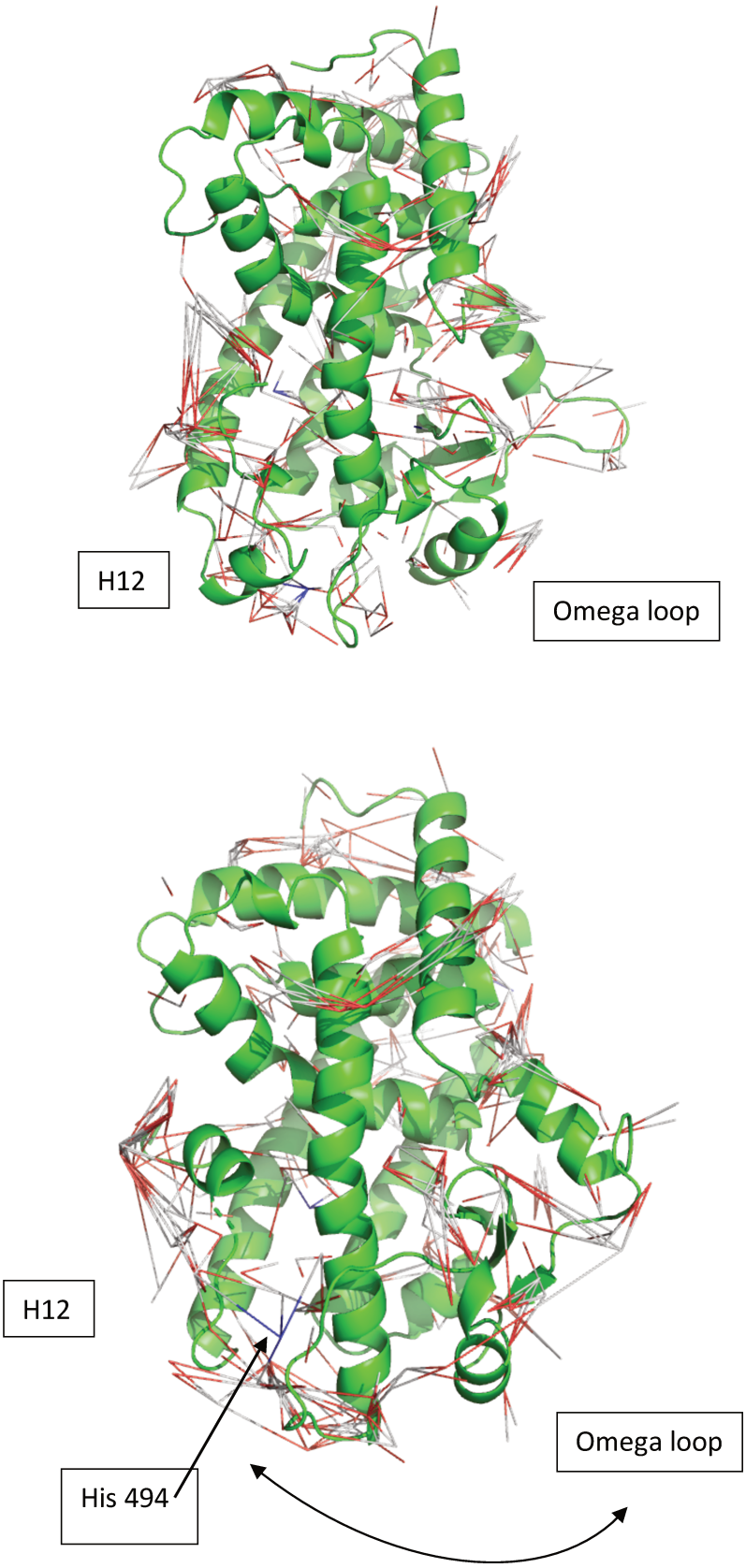


Figure 44. H-bonds mapped onto the average structure over the 100 extracted for normal mode analysis. Systems: WT apo (top), and WT structure complexed to GW1929 (bottom). In the ligand bound system, the arrow highlights a wider range coupling of the omega loop region to H12.

The T475M mutation has been identified as an activating point mutation in luminal bladder cancer (Natacha Rochel et al. 2019). From the analysis of the H-bonds in this structure, we see there is significant coupling between the H11-H12 loop and the N-terminal end of H3 (Fig.45 top). There is also some coupling with the Ω loop, as seen in the WT protein complexed with the agonist GW1929, but we see that, given the length of the hydrogen bonds, the region remains quite flexible. While there are various couplings via H-bonds of the C-terminal end of H12, the cluster is less populated than that observed in the WT complexed to the agonist GW1929 (see Fig. 44). His494 is involved in hydrogen bonds with Gln311, Leu496, and Gln498. Even in its apo form, the mutant T475M has shown some transcriptional activity (Natacha Rochel et al. 2019).

The F310S has been identified as an inactivating point mutation in basal bladder cancer. In Fig. 45 (bottom), we see the H-bonds calculated from these structures. The F310S mutation show a significant amount of hydrogen bonding around H12, but interestingly, there is almost no H-bond coupling to the Ω loop region. Given the length of the illustrated H-bonds, the region remains quite flexible. Compared to the WT apo form, F310S shows a much more local H-bonding between H12 and the N-terminal end of H3, which suggest that the mutation leads to a stabilization of this region in its inactive conformation. This H-bonding could impede the conformational changes necessary to populate the active form. This could contribute to the inactivating effect of the mutation.

These results suggest that the shift in the IR spectra might be partially due to changes in hydrogen bond numbers.

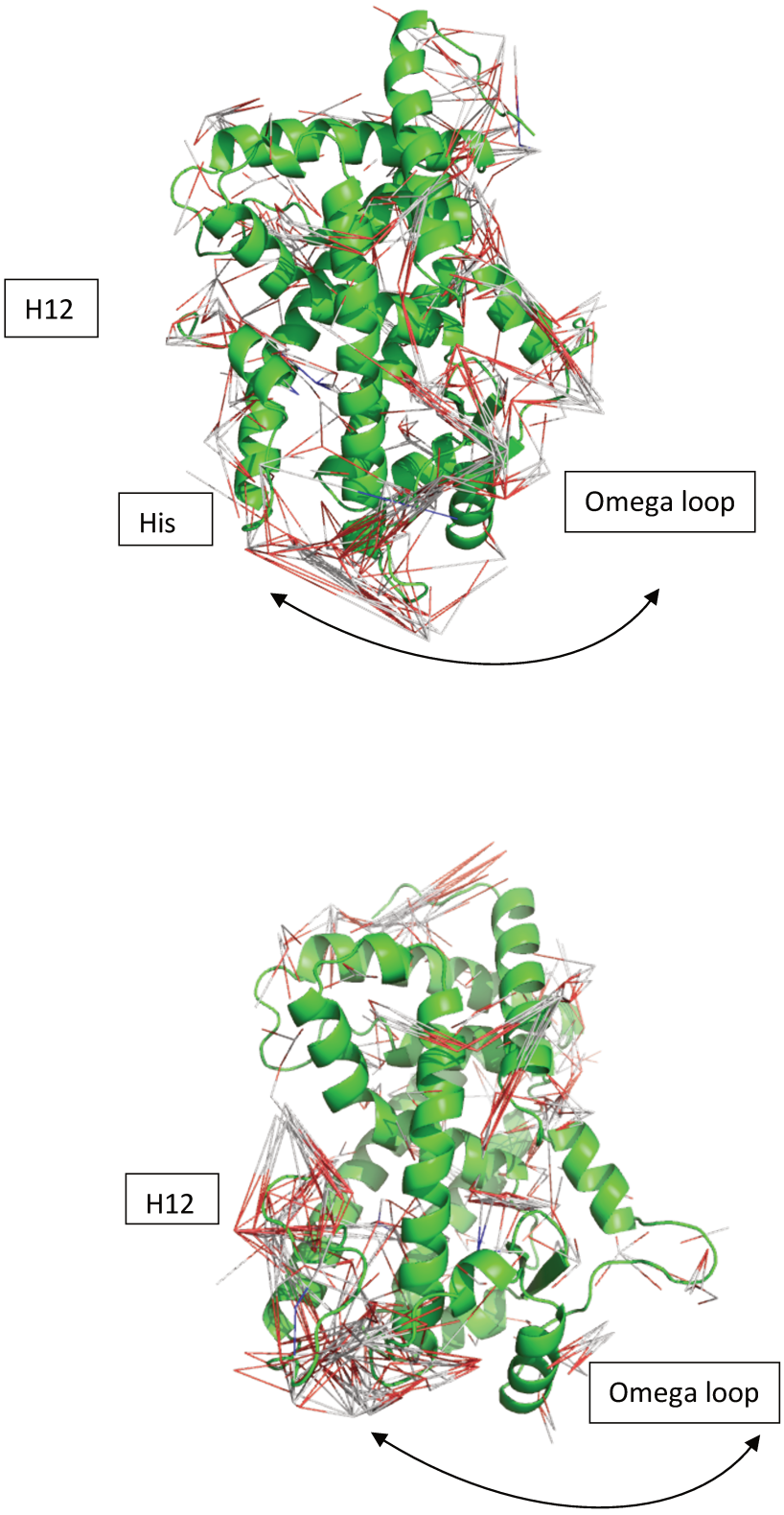


Figure 45. H-bonds mapped onto the average structure over the 100 extracted for normal mode analysis. Systems: apo T475M structure (top), and the apo F310S structure (bottom).

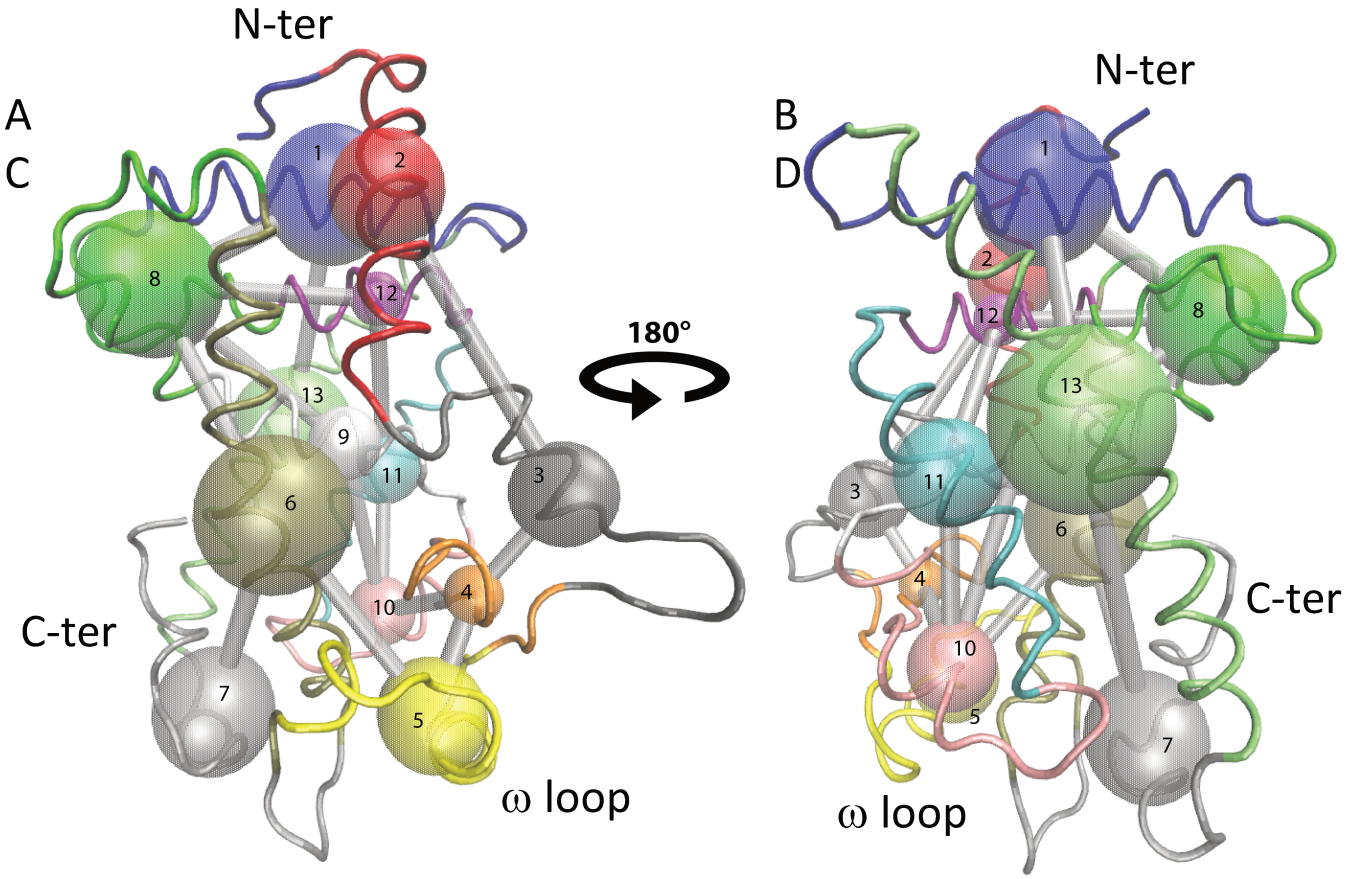
3.3.6. Community network analysis (CNA)

Community network analysis was performed using the Bio3D software (Grant, Skjærven, and Yao 2021). Contact maps were produced using an atom-atom distance (cut-off of ≤10 Å) and the correlated motions obtained from the MD trajectories. The CNA uses graph-based network approach that is based on the edge-betweenness centrality measure, where the edge-betweenness centrality of an individual residue is defined as the number of the shortest paths connecting other residue pairs that pass through it, thus providing an estimate of the influence of this residue on communication, or modularity. Communities of residues are characterized by high modularity values, that is, residues in the same community share dense connections, whereas residues of different communities have sparse or no connections at all. The size of the nodes is related to the size of a community and larger spheres depict higher number of residues. The edges connect coupled communities, where thicker edges correspond to higher degree of correlation. The correlation threshold for edge detection (c_{ij} *cutoff*) was 0.5. The community map analysis results are depicted using coloured spheres mapped on the average 3D structure in tube representation.

The Community Network Analysis (CAN) of the PPARγ systems was done using the correlated motions calculated and averaged from the 3 replicas simulations for each PPARγ variant. The results (Figs. 49 to 52) show community nodes placed on the average backbone structure for each system. Two orientations are shown, one rotated about 180° with respect to the other (A and B). Also shown are the detailed compositions of each node (C), as well as the network 2D map (D). Many nodes correspond to entire secondary structure elements, such as helices, however, some nodes encompass spatially adjacent residues belonging to different structural elements. Coupling between nodes is indicated by an edge, drawn as a rod connecting two nodes. The thickness of the rod indicates the degree of coupling.

For the WT apo protein, the results of the CNA are shown in Fig. 46. 13 nodes were detected by the analysis. Interestingly, we see nodes formed on H3 (node 6), H12 (node 7), the Ω loop region (node 5). The β sheet region forms a single node in the WT apo protein (node 4). The loop between H8 and H9 forms a sizable node (node 8) that is highly connected

to the rest of the LBD (Fig. 46A). H3 shows connections to H12 and the ω loop region. H10-H11 (Fig. 46B) forms a single node (node 13) that is connected to H9 and the N-terminus (node 1), as well as to loop H8-H9 (node 8).



WT APO

id	size		members
1	1	29	c(1:4, 202:226)
2	2	21	5:25
3	3	19	26:44
4	4	11	c(45:47, 139:146)
5	5	20	48:67
6	6	29	c(68:69, 76:102)
7	7	25	c(70:75, 258:276)
8	8	29	c(103:119, 190:201)
9	9	16	120:135
10	10	16	c(136:138, 147:159)
11	11	19	160:178
12	12	11	179:189
13	13	31	227:257

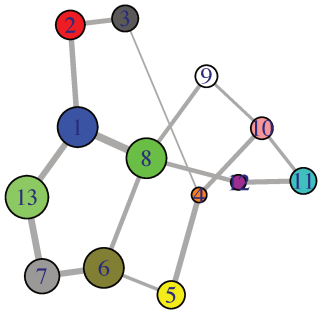
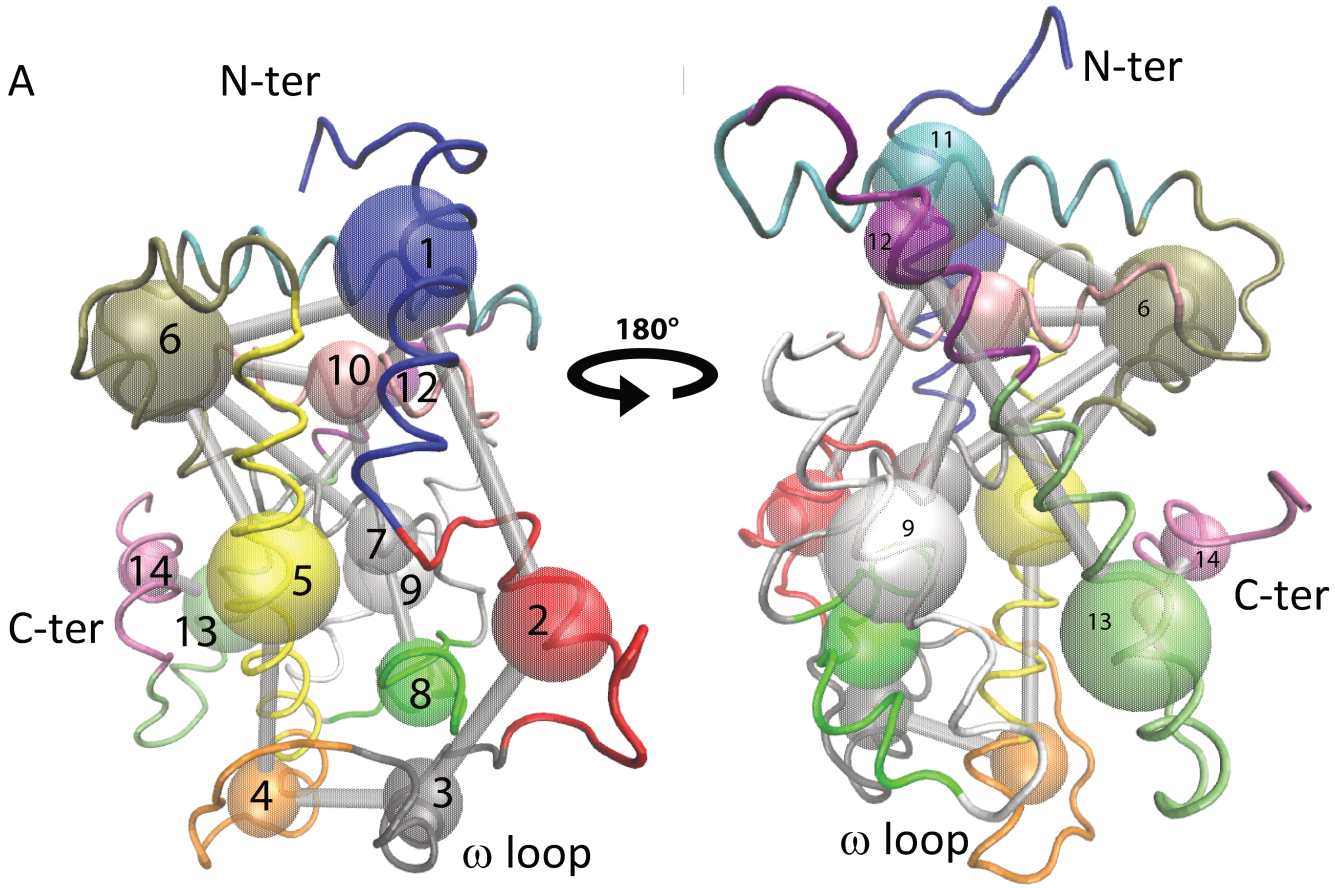


Figure 46. CNA of the apo WT system. A) LBD, B) LBD rotated by 180°, C) detailed list of node members (add 229 for isoform 2 numbering), D) network graph.



C

GW1929

id	size	members
1	25	1:25
2	20	26:45
3	16	46:61
4	16	62:77
5	25	78:102
6	27	c(103:120, 196:203, 205)
7	16	121:136
8	19	137:155
9	23	156:178
10	17	179:195
11	22	c(204, 206:226)
12	13	227:239
13	24	240:263
14	13	264:276

D

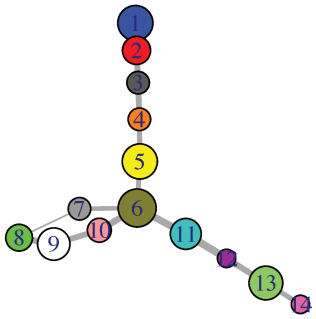


Figure 47. CNA of the holo-WT with GW1929 system. A) LBD, B) LBD rotated by 180°, C) detailed list of node members (add 229 for isoform 2 numbering), D) network image.

Adding the agonist GW1929 to the WT system results in a radical modification of the node organization (Fig. 47). The first community node regroups the N-terminal residues of H1, but unlike the apo protein, the first node does not include any amino acids of H9. The effect of this is a decoupling of H1 (node 1) and the residues of loop 8-9 (node 6). As in the apo system, the node encompassing loop 8-9 (node 6) is the most connected node in the LDB. In the presence of the ligand, we see that the β sheet (node 8) no longer contains all three strands, but just the two closest to the ligand binding pocket. We also see that H3 (node 5) decouples from H12 (node 14). While in the WT-apo protein, H10-H11 composes an entire node, in WT-holo, this long helix is broken into 2 nodes. This illustrates that ligand binding affects the motions of H10-H11, which is the principal helix found at the dimer interface.

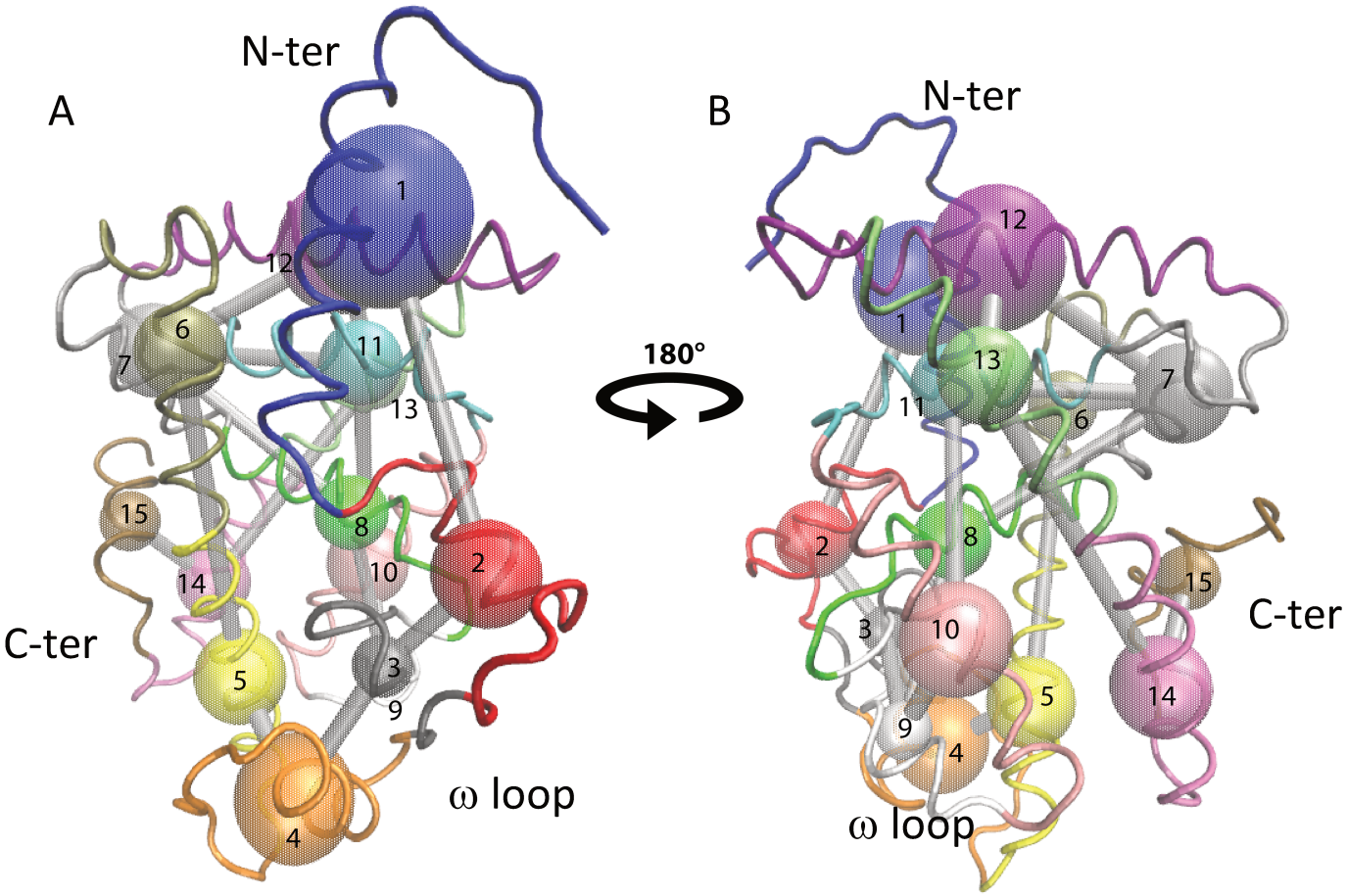
The apo mutant protein, T475M, has features that resemble both the WT-apo and the WT-holo protein. Recall that the mutation T475M is an activating mutation, in that, even in apo form, it displays transcriptional activity (N Rochel et al. 2019). In the T475M mutant, the β sheet region contains all three strands (node 3), but H3 (node 5) is decoupled from H12 (node 15). As in both the WT- apo and holo proteins, loop 8-9 (node 7) is the best-connected node. And as in the WT-holo, H10-H11 is represented as 2 nodes with the transition from one node (node 13) to the other (node 14) occurs just before the point mutation (Fig. 48). T475M mutation affects the size of the nodes attributed to H10-11. In the holo form (Fig. 47), a large node encompassing much of H10 and H1 was determined. Introduction of the point mutation T475M, located in H10-H11 leads to a decrease in the size of this major node.

Interestingly, the WT-apo system has a strong edge between H3 (node 6) and H12 (node 7) (Fig. 46), but introduction of the ligand or the T475M mutation leads to a loss of this edge and an increased coupling of H12 with H10-H11 (nodes 12 and 13), which is the helix found at the dimer interface. This could reflect a mechanism to transfer information to the dimer partner, receptor RXR.

The apo mutant F310S, which is an inactivating mutation, displays effects from the point mutation (Fig. 49). Modelled on the structure of the apo chain from the PDB structure 7WOX, which was also used for the wild-type apo simulations, the point mutation is located in helix H3 and breaks the coupling between H3 (node 4) and H12 (node 13), shifting the

couplings towards the β sheet region of the protein. As in the other protein systems, the loop between H8 and H9 (node 5) is well connected to the rest of the protein.

Loop 8-9 of the LBD is known to interact with cyclin H in the nuclear receptor RARα (Samarut et al. 2011b). The PPARγ LBD is known to interact with cyclin D in the context of regulating adipogenesis (Sarruf et al. 2005). We saw that in all cases, the node encompassing loop 8-9 is the most connected in the networks, which underscores its importance in LBD physiology. Besides the loop between H8-H9, the loop between H9-H10 is also well connected to the rest of the protein. This loop is known to be important in the allostery related to phosphorylation in other receptors (Chebaro et al. 2017; 2013; Samarut et al. 2011b).



C

T475M

id	size	members
1	29	1:29
2	18	30:47
3	11	c(48:50, 142:149)
4	23	51:73
5	20	74:93
6	17	94:110
7	21	c(111:123, 196:203)
8	16	124:139
9	11	c(140:141, 150:158)
10	20	159:178
11	17	179:195
12	27	204:230
13	16	231:246
14	19	247:265
15	14	266:279

D

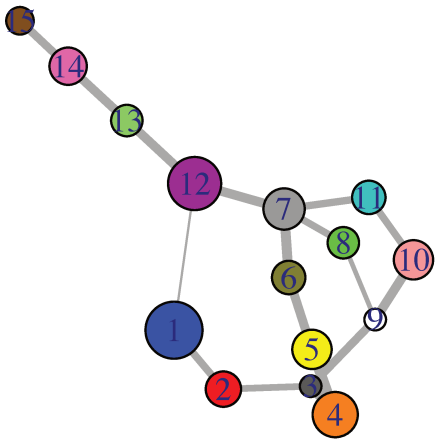
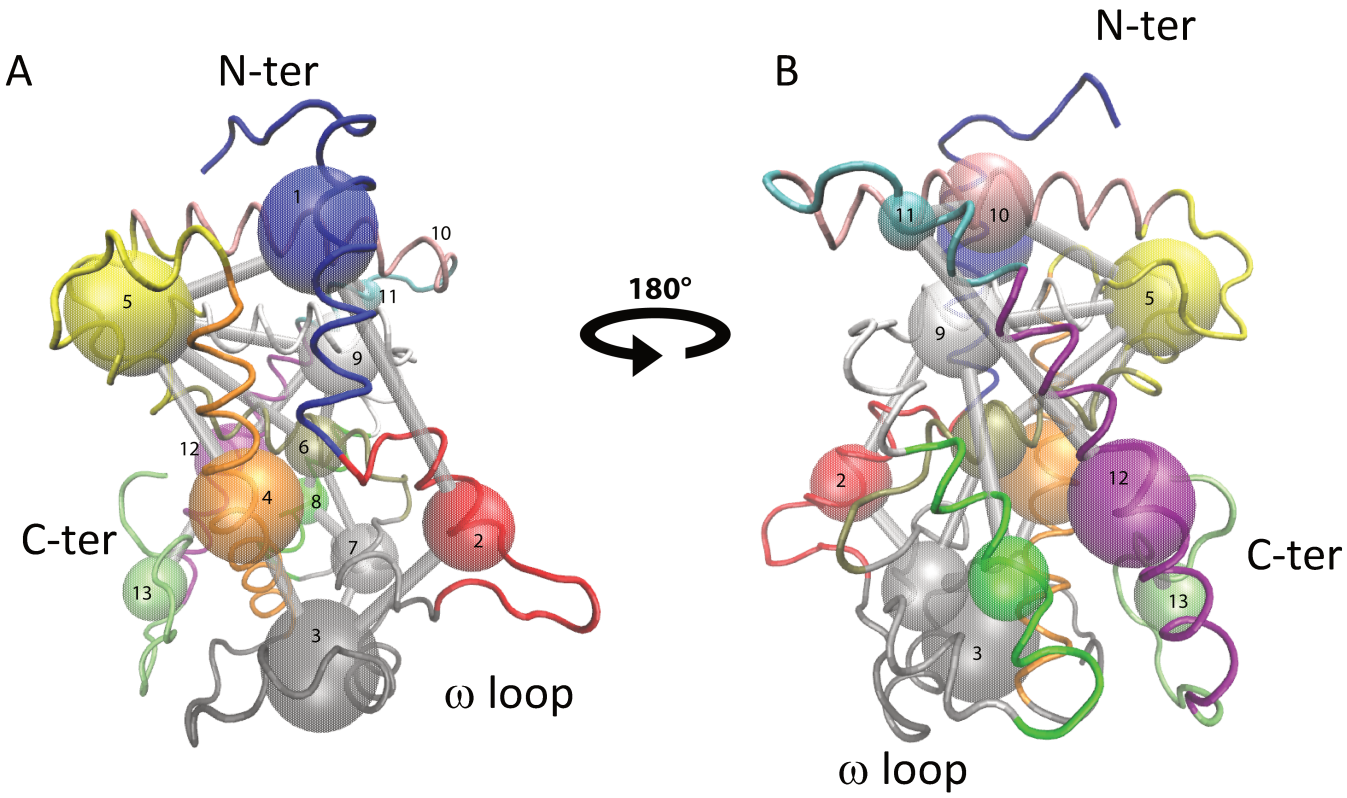


Figure 48. CNA of the T475M system. A) LBD, B) LBD rotated by 180°, C) detailed list of node members (add 226 for isoform 2 numbering), D) network image.



C

F310S

id	size	members
1	25	1:25
2	20	26:45
3	30	46:75
4	27	76:102
5	29	c(103:119, 193:203, 205)
6	17	120:136
7	19	137:155
8	15	156:170
9	22	171:192
10	21	c(204, 206:225)
11	10	226:235
12	24	236:259
13	17	260:276

D

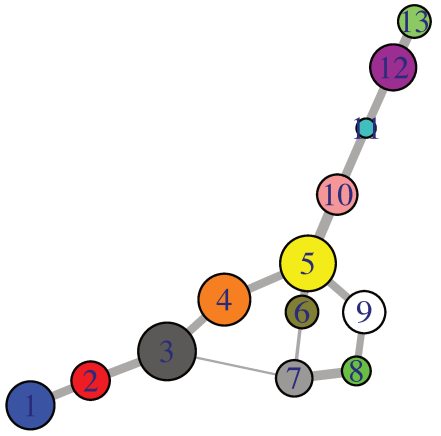


Figure 49. CNA of the F310S system. A. LBD, B. LBD rotated by 180°, C. detailed list of node members (add 229 for isoform 2 numbering), D. network image.

4. Conclusions

In this chapter, we presented a combined MD/Far-IR approach to the study of the PPARγ ligand binding domain. The protein in apo-, holo- form complexed to the agonist GW1929, as well as two mutant proteins in apo form, T475M and F310S, were studied.

Through our collaboration with the team of Prof. Petra Hellwig of the Laboratory of Bioelectrochemistry and Spectroscopy, University of Strasbourg, we demonstrated for the first time the feasibility of using a far-IR/MD approach to measure the effect of ligand binding by a nuclear receptor protein, in this case, the ligand binding domain of PPARγ. The study showed that upon ligand binding, the lowest frequency peaks undergo a shift toward higher frequencies (blue shift) suggesting a strengthening of the vibrational mode.

We carried out extensive molecular dynamics simulations of all the forms studied, from which we generated free energy surfaces based on the distribution of RMSD and radius of gyration. From these surfaces, we identified a number of populated potential wells. From the primary free energy wells on these surfaces, representative structures were extracted and used for the Normal Mode Analysis, followed by calculations of the IR spectra and analysis of the collective motions through correlated motions and community network analysis.

The computational results demonstrated that in the far-IR domain, the calculated IR spectra displayed very similar feature patterns to those observed experimentally, in particular shifts of the low frequency spectral peaks. We confirmed through hydrogen bonding analysis of the simulations that the changes in the experimental spectra coincide with changes in hydrogen bonding. Effects of point mutations on the low frequency spectra were also observed and correlated with changes in hydrogen bonding.

Using the structures extracted from the most populated well for an Ensemble averaged Normal Mode Analysis, we calculated maps of correlated motions of the LBD, which showed regions of coupling within the single domain. We noted that even small perturbations due to ligand binding or point mutations can have an effect on the collective dynamics of the protein, which modulated couplings between physiologically distinctive regions of the LBD.

To further exploit the calculated collective motions, we carried out a community network analysis (CNA) to identify highly correlated regions that are coupled to other highly correlated regions of the LBD. Couplings between highly correlated nodes suggest a means by which information might pass within the protein domain due to changes in the structural dynamics. We identified regions of the LBD that are strongly coupled to the rest of the domain, in particular, the loop between H8 and H9, which is known to be physiologically important as a site that binds cyclin during a phosphorylation cascade identified on some nuclear receptors (Samarut et al. 2011a). We also showed how ligand binding shifts the correlated dynamics toward the region of the LBD that takes part in dimerization with RXR. We showed that small changes related to single point mutations are measurable experimentally, which allowed us to validate our simulations.

We showed that the computational and experimental far-IR approaches are complementary since both approaches displayed similar trends in the measured and calculated spectra. Once the simulations were validated through comparison to this experimental data, we were able to further exploit the correlated motions obtained from our Ensemble-averaged normal mode analysis to get insight into the effects of these perturbations on the intra-domain dynamics. The sensitivity to binding of different ligands, agonists versus antagonists remains to be explored. In the future, we expect to study the effects of point mutations to better understand their effects on the structural dynamics arising from the collective modes.

5. Annex: Experimental methodology

(with permission from Professor Petra Hellwig)

5.1. Mid-Infrared (Mid-IR) Measurements

Five samples were measured immediately after being removed from storage at -70°C. In the mid-IR domain, a mercury cadmium telluride (MCT) detector was used with a scan velocity of 40 kHz. For each sample, five spectra were recorded with a resolution of 4 cm-1 (256 scans) and averaged. The measurements were performed using a Vertex 70 instrument (Bruker).

The infrared transmission spectra were obtained using a transmission cell consisting of two calcium fluoride (CaF2) windows. A volume of 0.5 µL of the protein solution, along with the buffer, was deposited on one window, which was then covered with the other window, forming a defined path length of 17 µm as described by Barth et al. The assembled transmission cell was placed in the optical path of the IR beam. For FTIR-ATR spectra of the peptide, 3 µL of the protein solution was deposited on a diamond crystal (Harrick) and the spectra were recorded until a dry film was obtained.

The transmission spectra were subjected to specific data treatment to isolate the amide I band using OPUS software. This involved baseline correction of the protein and buffer spectra, normalization of both spectra in the amide I band region, and subtraction of the buffer spectrum from the protein spectrum to obtain the difference spectra.

5.2. Far-Infrared (Far-IR) Measurements

Five samples were measured after being removed from -70°C and measured using diamond ATR. For each measurement, 3 µL of the sample was deposited on the crystal and spectra were recorded until dry. In the far-IR domain, a deuterated triglycine sulfate (dTGS) detector was used with a scan velocity of 2.5 kHz. For each sample, five spectra with a resolution of 4 cm-1 (128 scans) were recorded and averaged. These measurements were also conducted using a Vertex 70 instrument (Bruker).

5.3. Determination of Protein Secondary Structure

The amide I and VI bands, which appear in the spectral ranges of 1700 to 1600 cm^{-1} and 590 to 490 cm^{-1} respectively, were analyzed. Key positions of structural elements were identified using second derivative spectra. A straight baseline was subtracted from the spectra in the ranges of 1700 to 1600 cm-1 and 590 to 490 cm-1. Least-squares iterative curve fitting with Gaussian bands was performed using the Peak Fit Analysis Program (Sea-Solve, MA, USA). Each band obtained from the fitting procedure was assigned to a specific secondary structure element. The areas of all bands assigned to a given secondary structure element were summed, and the total area was used to determine the contribution of each structural element to the overall protein structure.

Chapter V - Impact of force field polarization on the collective motions of PPARγ

1. Introduction

While experimental structure determination has shed light on the many conformations a particular protein can exist in, there remains little in the way of experimental exploration of the detailed motions of particular conformations. Of these, long-range correlated motions are considered fundamentally important for key functional properties of proteins such as substrate binding, allostery and catalysis (Kurkcuoglu et al. 2012). Changes in single domain collective motions have been associated to the sensing of ligand binding resulting in the propagation of a signal through the protein to transmit information and alter activity. Studies have suggested that correlated motions of secondary structure elements, such as β-sheets, contribute importantly to protein function (Fenwick et al. 2014). For example, PDZ domains are protein interaction modules that recognize short amino acid motifs at the C-termini of target proteins. Ligand binding affects the transfer of binding information to other domains in the context of PDZ-containing multidomain scaffold proteins. In the PDZ domain, the global network of correlated motions, can lead to the coupling of the N- and C-terminal ends by pathways involving the β-sheets. These motions arise from the low-frequency collective movements of residues and it has been suggested that these protein motions are selected by evolution (Glembo et al. 2012; Orellana 2023).

While the importance of correlated motions has become more apparent and appreciated, they remain difficult to measure experimentally, so one of the principal methods for studying correlated motions is by molecular dynamics simulations. Molecular dynamics simulations of proteins rely on the use of empirical force fields, which are parameterized using, for the most part, experimental data and quantum mechanical calculations. And while this approach has been used with great success over the past decades to study a wide range of topics, there is a constant effort to introduce improvements. One such effort has been to improve the treatment of electrostatic interactions, which in the standard classical force fields, are treated by fixed point charges. Efforts by numerous teams have focused on introducing aspects of electronic polarization. One approach characterizes the charge redistribution within each atom, by either induced dipoles (Shi et al. 2013) or by a Drude oscillator model (Lopes, Huang, Shim, Luo, Li,

Roux, and Mackerell 2013) , and the other approach is based on charge flow between atoms, as implemented in the fluctuating charge (FQ) model (S. Patel, Mackerell Jr., and Brooks III 2004).

Developed largely in the framework of the CHARMM all-atom force field, the Drude oscillator model for protein force fields (Anisimov et al. 2005) is a theoretical framework that introduces an auxiliary particle called the "Drude particle" for each atom representing a loosely bound electron that contributes to the atomic polarizability. A harmonic oscillator function is used to connect the Drude particle to the atom, simulating the restoring force on the electrons. The model introduces a term that represents the polarizability of the atom, allowing for the simulation of electronic response to an external electric field. The development of the Drude model in the context of polarizable force fields in molecular dynamics simulations aims to capture electronic polarization effects that are not explicitly represented in the classical models. This type of model is intended for studying systems where electronic polarization plays a significant role.

The Drude model has been used and benchmarked for a variety of systems (Lopes, Roux, and MacKerell 2009), and several reviews are available (Lemkul, Huang, Roux, and MacKerell 2016; Lemkul 2020). Though the Drude model for polarization has undergone extensive development and application, the analysis associated with the applications has largely focused on aspects of structure, energetics and local dynamics. None of these studies have utilized the Drude model to address questions concerning collective behaviour.

Collective motions occur across the three-dimensional structure of the protein and the principal tool for studying such motions in proteins is molecular dynamics simulations. The question naturally arises, then, to what degree does the inclusion of polarization affect the collective motions of single protein domains. In this work, we address this question through the study of the ligand binding domain of the nuclear receptor PPARγ. We assess the impact of the polarization on various dynamical properties, including their collective motions.

Peroxisome proliferator-activated receptor gamma (PPARγ) is a ligand-dependent transcription factor belonging to the nuclear receptors superfamily (Lazar 2005). PPARγ was presented in the introduction, chapter I, section 5.1. Structures of PPARγ LBD in its apo and

corepressor-bound form, in complex with a peptide from the NCor1 corepressor protein are shown in Fig. 50.

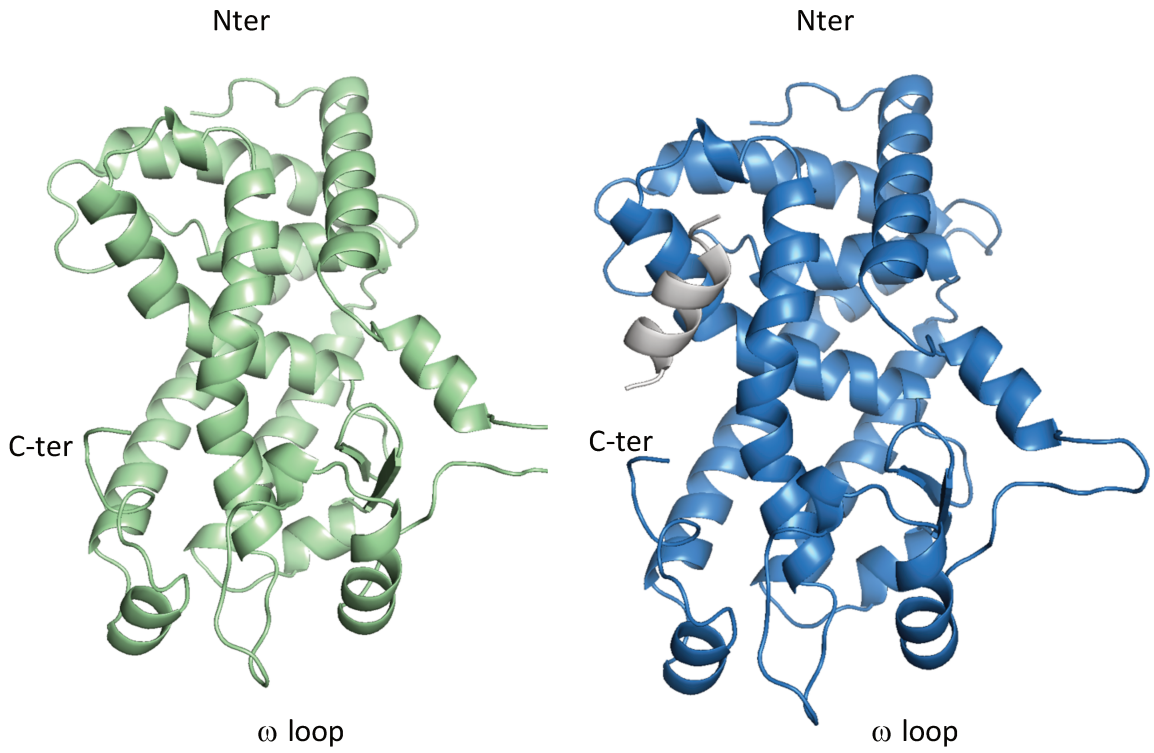


Figure 50. PPARγ ligand binding domain (residues 230 – 505) from the 3.2Å crystal structure PDBID 7WOX chain B. A) Apo form (green) and B) the same LBD (blue) modeled with the corepressor-peptide (gray).

The physiological function of NRs is highly dependent on conformation and structural dynamics modulated by ligand binding. The ligand binding domain acts as a dynamic hub, transmitting binding events to other protein interfaces and domains. The PPARγ activation involves a conformational change of the H12 helix at the C-terminal part of the LBD induced by ligand binding (Weikum, Liu, and Ortlund 2018a). Helix H12 undergoes a transition from a flexible ensemble of conformations to a folded conformation localized on the core of the

LBD. Characterized by numerous crystallographic structures of agonist-bound LBDs, this position of the H12 helix is often referred to as the transcriptionally active conformation (J. P. Renaud and Moras 2000c; Wurtz et al. 1996). In this active conformation, H12, along with helices H3 and H4, constitute a hydrophobic interface called an activation function 2 (AF2). This interface serves as a platform for coactivator protein binding and the recruitment of chromatin modulator complexes as well as other components of basal transcriptional machinery (Raj Kumar, Johnson, and Thompson 2004b). In contrast to the active conformation of the LBD and H12, the inactive conformation of the receptor is not structurally well described. The crystallographic and computational data suggest an ensemble of conformations for H12, meaning that this region, in the absence of an agonist ligand is flexible (Frkic et al. 2023).

The first study of the functional dynamics of the PPARγ ligand binding domain (LBD) was done by Fidelak *et al* (Fidelak et al. 2010b). This study explored the role of allostery in the functioning of the receptor by comparing LBDs in apo and in agonist-bound form. A dynamical pathway linking amino acids which are in topological proximity or at distance was established, explaining correlated motions primarily arising from low-frequency collective motions. The analysis of correlated motions shows coupling between distant regions of the LBD, such as different helices, the N- and C-terminals and the coactivator peptide. In the apo form of PPARγ, the dynamical network plays a role in maintaining the structural integrity and flexibility of the protein. It is also involved in mediating interactions with co-regulator proteins and with PPARγ heterodimeric partner RXRα. The addition of the agonist rosiglitazone contributed to changes in the dynamical network of PPARγ. As a consequence, the changes in this network could impact the ability of LBD to bind ligands and coregulators, and by extension the overall function of PPARγ. Correlated motion calculations and network analysis were later done for the full PPARγ/RXRα heterodimer structure in complex with DNA (Ricci et al. 2016). The results showed the existence of longer range interdomain correlations which were used toward the understanding of allostery in nuclear receptor complexes. Here, we explore the intrinsic dynamics of PPARγ LBD in its apo- and corepressor peptide bound forms using both the classical all-atom additive empirical energy functions and the Drude force field. We provide

quantitative insights into the effects of polarization on the modelling of correlated motions of PPARγ.

2. Methods and Analysis

We study the ligand binding domain (LBD) of human PPARγ because of the importance of collective motions in its function. The system was prepared using the PDB Reader and Manipulator option of the CHARMM-GUI web interface (Jo et al. 2008) to prepare the simulations using the CHARMM all-atom additive force field (AA using the CHARMM36 all-atom parameter set (A. D. MacKerell, Bashford, et al. 1998). The CHARMM-GUI Drude Prepper interface (Kognole et al. 2022) was subsequently used to prepare the systems for simulations using the Drude polarizable force field.

For the PPARγ ligand-binding domain (residues 230 - 505), we used the 3.2 Å resolution crystal structure of chain B from the PDB file 7WOX (Yoshizawa et al. 2022b). Although one chain in this PDB entry is bound to the antagonist MMT-160, the second chain (chain B) did not show any electron density representing a ligand in the binding pocket, so it was taken to be a structure of the apo protein. The protonation states of the histidine residues of this chain were determined using PROPKA method (Olsson et al. 2011; Søndergaard et al. 2011) via the webserver https://server.poissonboltzmann.org/pdb2pqr, followed by manual verification and the structure was further prepared using the CHARMM GUI interface (Jo et al. 2008).

The molecular dynamics simulations were done using the NAMD program under NPT conditions (Phillips et al. 2020c). The protocol consists of four steps, first, the protein was fixed but the water and ions were without constraints. The system was subjected to 1000 steps of energy minimization to allow the water and ions to adjust position in response to the presence of the protein. Next, the system was heated up to 600K, during 23000 steps, again with the protein. This was followed by another energy minimization for 1000 steps. This was followed by a heating to 296.5 K. The constraints on the protein/ligand were removed and the entire system was energy minimized for 2000 steps. The entire system was then heated up to 296.5

K over 15000 steps, followed by an equilibration run of 85 000 steps of dynamics followed by the production phase. A time step of 1 fs was used. The duration of each simulation was 100 stages of $1x10^6$ timesteps, which resulted in 100 ns - long simulations. The last trajectory frame was taken as a starting structure for building the structures for the Drude and AA simulations used in this study.

The two systems we compared were the apo form of PPARγ LBD and the PPARγ LBD complexed to the corepressor peptide NCoR ID1 (12 amino acid sequence GLEDIIRKALMG), identified here as the corepressor-bound form. The coordinates for the corepressor peptide were taken from the crystallographic structure of a PPARγ mutant complexed to the NCoR peptide resolved by our team (unpublished data). The 12 amino acid peptide was added by superposing the apo PPARγ LBD structure on the in-house structure.

For the simulations with the Drude force field, the apo- and corepressor-bound structures were solvated in 100 x 100 x 100 $Å^3$ water box using the SWM4-NDP water model. A minimization for 2000 steps was done followed by an equilibration for 200000 steps using the NAMD program with the time step of 0.5 fs. During the production phase, we used a time step of 1fs. The duration of each simulation was 100 ns. Three replica simulations were carried out for the four PPARγ LBD systems. Similarly, three replica AA simulations were performed. For each simulation, the root-mean-square coordinate difference (RMSD) and residue averaged backbone atomic root mean square fluctuations (RMSF) were calculated. These calculated fluctuations were compared to the atomic fluctuations calculated from experimental B-factors.

In addition to the Community Network Analysis (used and described in chapter IV), we used the Shortest Path Method (SPM) tool. SPM was used to assess the importance of individual residues and their pairwise connections in the structural dynamics of proteins (Casadevall et al. 2024). This is in contrast to the community network analysis, which establishes communities around multiple residues. The SPM method produces a network graph based on mean distances and correlation values, and computes shortest path lengths using Dijkstra algorithm (Dijkstra 1959). The shortest path is the most direct path with the most significant connection between two residues, and shows how the residues are connected in the protein's structural dynamics. The tool is mostly aimed at exploring key residues implicated in

enzymatic activity, but here we use as a way to assess the similarities and differences of simulations using different force fields.

3. Results and Discussion

3.1. Structural dynamics of PPARγ – RMSD and RMSF

The RMSD time series for PPARγ were calculated from the molecular dynamics simulations for the three replicas of the AA and Drude simulations for each system. For each system, the average times series of the three replicas were displayed along with the high/low values at each time point. All four PPARγ systems show stable 100 ns trajectories (Fig. 51). The RMSD mean value of PPARγ apo system simulated with the Drude model was higher than the value of the system simulated with the AA force field with the values being 3 Å (std_dev: 0.09) and 2.5 Å (std_dev: 0.06), respectively (Fig. 51A and B). We notice the same trend when comparing the simulations of PPARγ bound to the corepressor peptide NCoR (Fig. 51C and D). The Drude simulations presented higher values of RMSD, with the mean value of 3.2 Å (std_dev: 0.21) than the non-drude simulations, where the mean value is 2.4 Å (std_dev: 0.04). These results are consistent with the conclusions that the Drude force field allows for a higher conformational flexibility than the standard additive CHARMM force field (Lopes, Huang, Shim, Luo, Li, Roux, and Mackerell 2013). In addition to the overall stability of the PPARγ - corepressor bound system, we see the interaction of two components as stable as the peptide does not

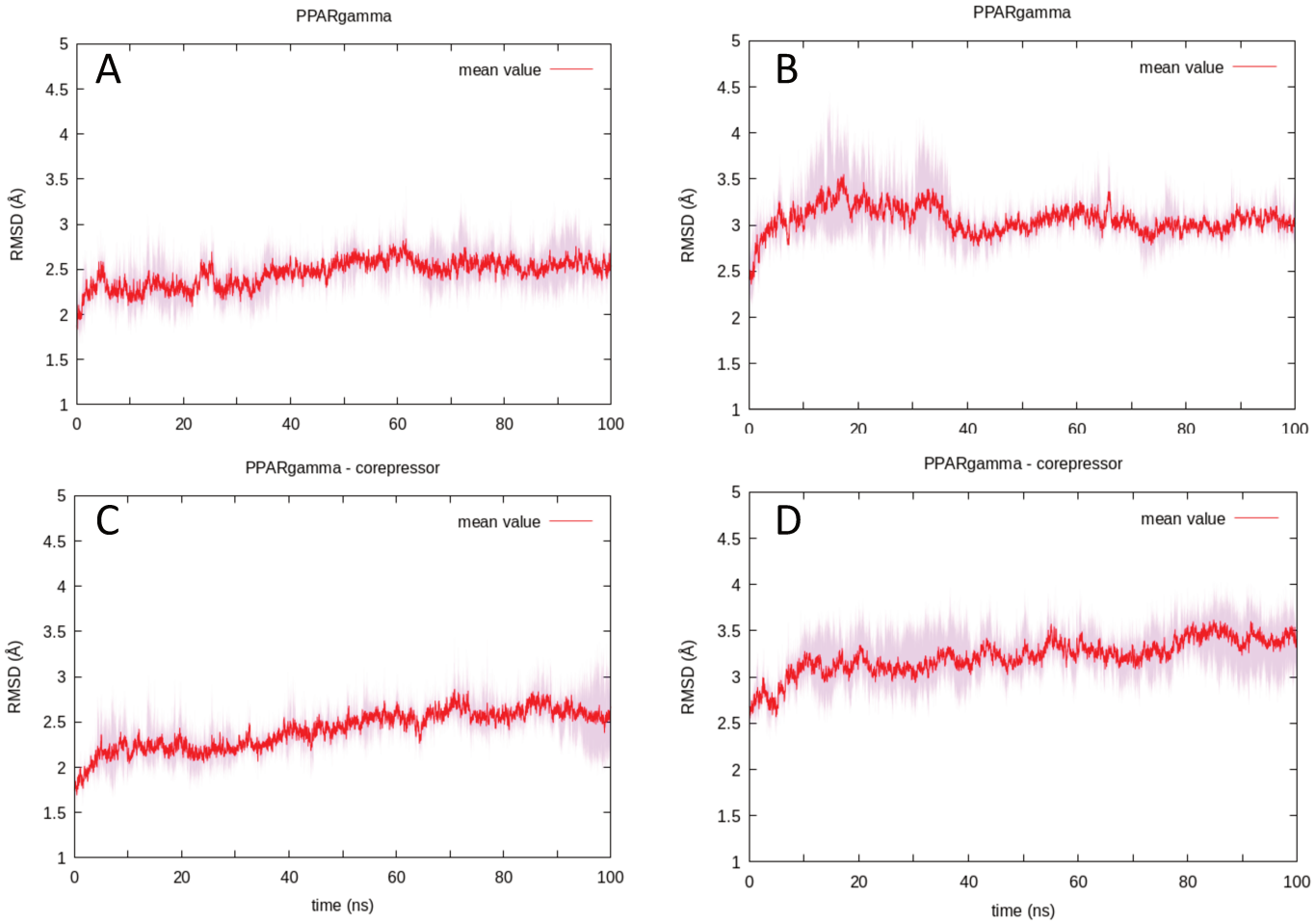


Figure 51. RMSD of PPARγ LBD, left: AA simulations for the apo A) and co-rep complexed C) protein and for the Drude simulations, apo B) and corepressor complexed D). The mean value of 3 replicas is represented as a red line.

dissociate from the PPARγ LBD. In our case, the Drude force field maintains the protein – peptide complex.

We calculated the root mean square fluctuations (RMSF) of the backbone atoms of PPARγ averaged by residue over all three replicas (Fig. 52). In all of the cases, we observe an RMSF profile that well reflects the structure- in that loops are more flexible than the secondary structure regions. The AA simulations present the highest flexibility in the regions of the loop between H2 – S1, loop H9 - H10 and the H12 (Fig. 52A). The apo Drude simulation shows a higher flexibility of the Ω loop region, and smaller flexibility of the loop H9 - H10 (Fig. 52B), than the AA simulation. The explanation for the lower flexibility of the loop between H9 and H10-11 could be that in both cases of Drude simulations, a salt bridge between residues D411 (H8) and H453 (loop H9-10) is maintained. This salt bridge is present in the crystal structure,

and so it is also in the initial simulation structure His 453 was taken to be a protonated histidine. This salt bridge is known to be conserved in all NRs of class II (Brelivet et al. 2004), so the simulations with the Drude force field lead to the maintaining of this conserved salt bridge, while it was lost in the AA simulations. The second salt bridge characteristic of the class II NRs, between residues E352 (mid H4 – 5) and R425 (loop H8 – 9), is well maintained in both AA and Drude simulations. The RMSF variability, represented by the light pink regions, of three replicas is also higher for Drude simulations.

For the PPARγ - NCoR system, the differences are less prominent, the RMSF curves for both the AA and Drude simulations are similar, albeit with differences in the H9 - H10 loop and the H12 and corepressor peptide region (Fig. 52C and D). Higher variability is found in the AA simulation around H2' and the Ω loop, and also in the loop H9 - H10- the later reflects the loss of contact in salt-bridge D411- H453. Comparing the apo and corepressor bound PPARγ systems, we see the difference in the β-sheet region and H6. For both AA and Drude simulations, adding the corepressor peptide lowered the replica-averaged RMSF. With the Drude simulations, the variability among replicas is also much smaller. In cases of stable structures, like the LBD/corepressor peptide complex, the use of the Drude model seems to provide additional stabilization of the protein complex.

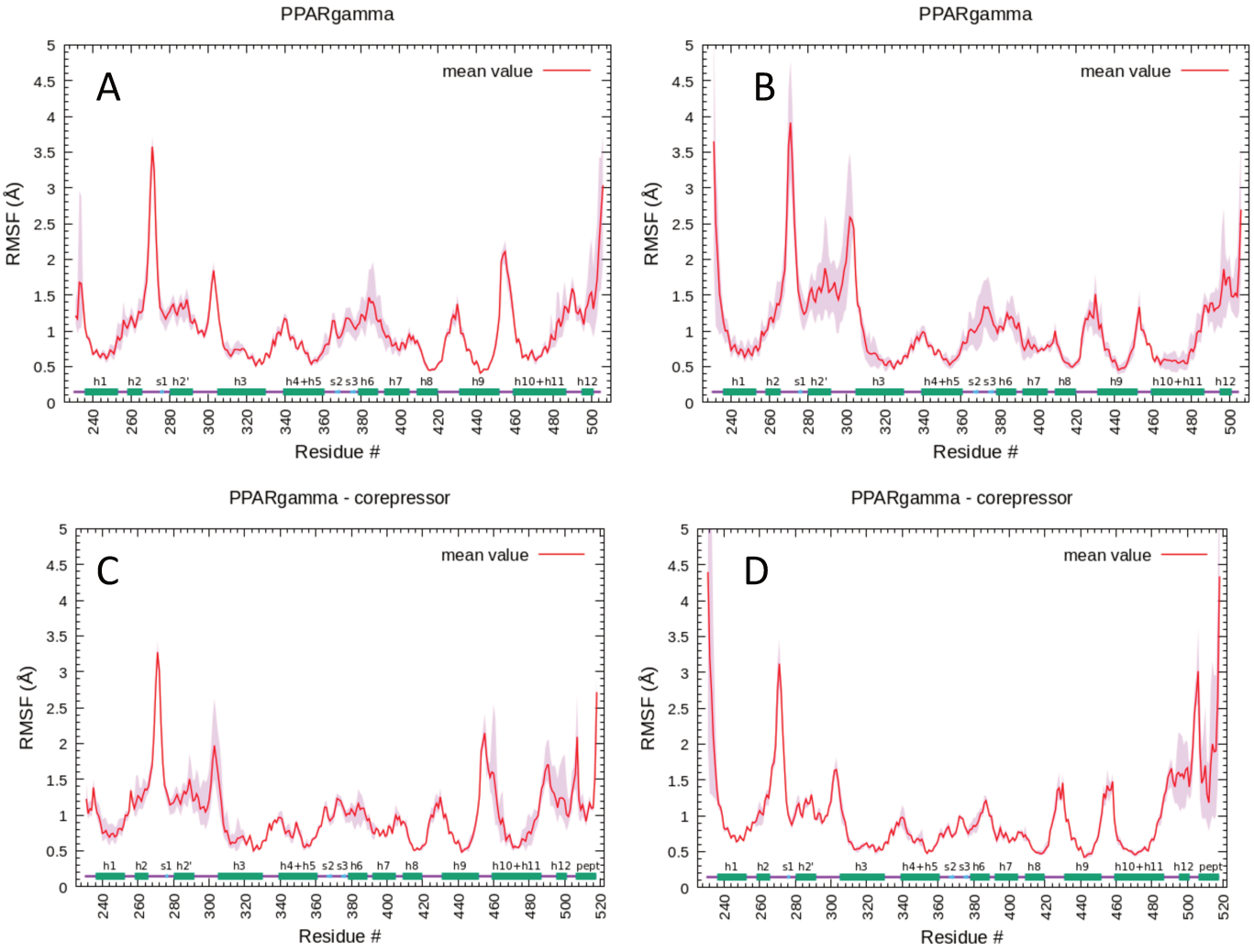


Figure 52. RMSF of PPARγ LBD, left: AA apo A) and AA corep complex C), Drude FF apo B) and copre complex D). The mean value of 3 replicas is represented as a red line. Secondary structure elements are shown on the x axis: alpha helices (h1 – h12) as green, and beta strands (s1 – s3) as blue rectangles. The corepressor peptide is represented as a green rectangle on the far right of C) and D).

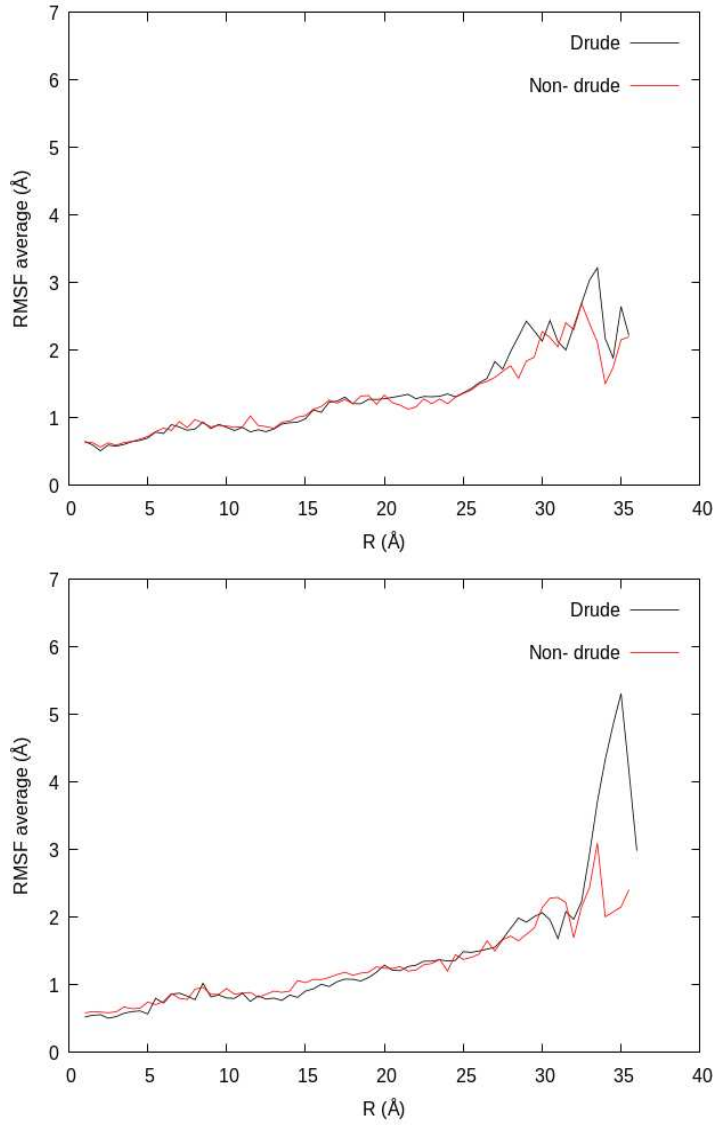


Figure 53. Radially averaged atomic fluctuations of the PPARγ LBD apo form (top), and corepressor bound form (bottom).

The radially averaged fluctuations of all heavy atoms were calculated (Fig. 53). We notice the increase in the fluctuations as we move towards the residues located further from the centre of geometry. The fluctuations are higher in the Drude simulations than in the AA (non-drude) simulations, reaching up to 3.3 Å in the apo system and 5.3 Å in the corepressor-bound system near the surface of the protein. The radius of gyration is similar for all four systems, with values of 20.1 Å (std_dev: 0.04) for AA apo system, 20.1 Å (std_dev: 0.01) for AA corepressor-bound system, 19.9 Å (std_dev: 0.15) for the drude apo system, and 20 Å (std_dev: 0.11) for the drude corepressor-bound system.

3.2. Electric dipole moments

Electric dipole moments of proteins contribute to the structural dynamics and function by influencing how proteins interact with their environment. Accurate modelling of the dipole moment can thus improve the overall representation of intermolecular interactions. The dipole moments of PPARγ were calculated along the trajectories (Fig. 54). The data are presented as time series of the dipole moment of the full protein. We further calculated the average dipole moment of the protein backbone by-residue (Fig. 55).

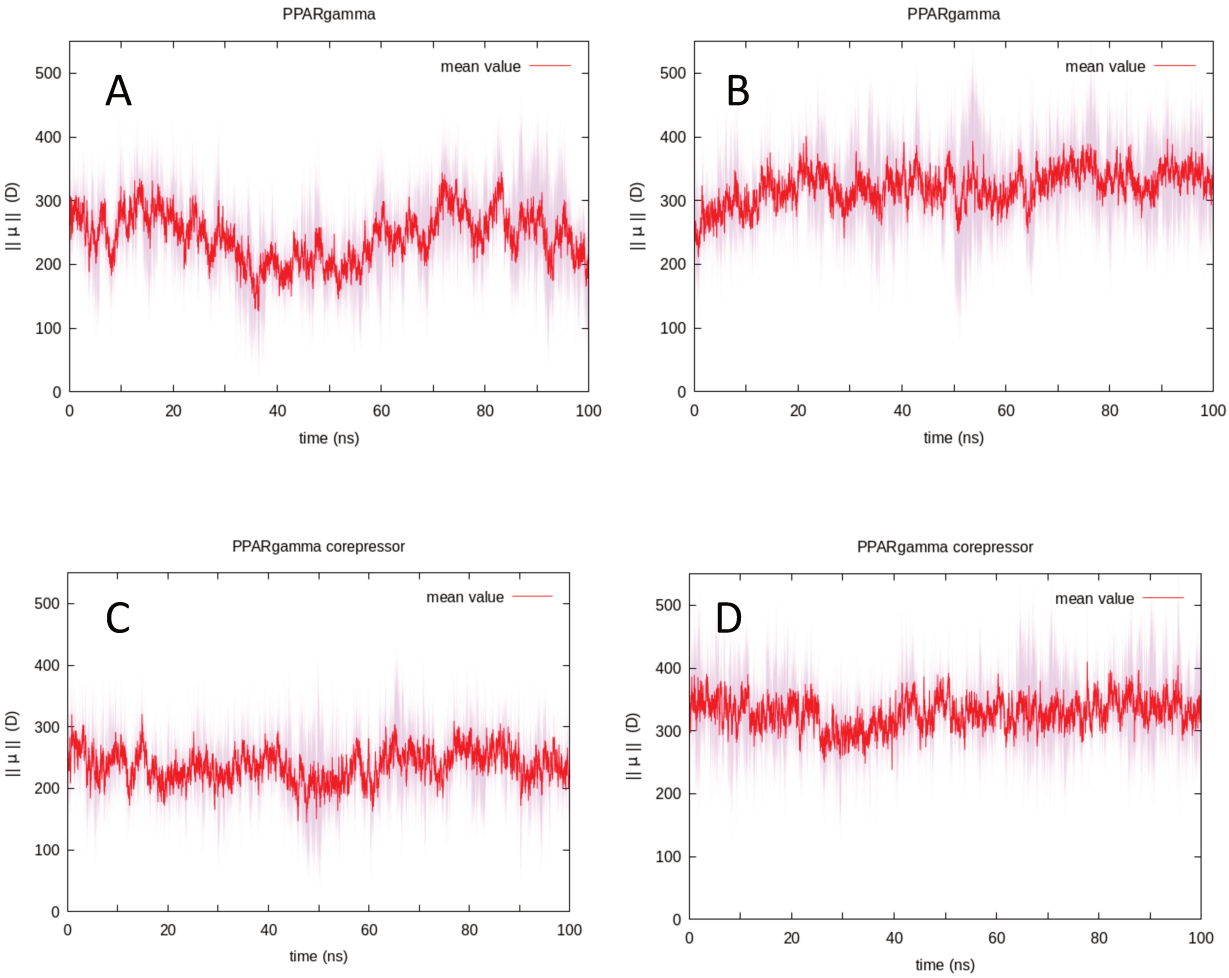


Figure 54. Protein dipole moment timeseries of the PPARγ LBD. A. in the apo form with AA force field, B. in the apo form with the Drude FF, C. in complex with co-repressor and AA force field, D. in complex with co-repressor and the Drude FF.

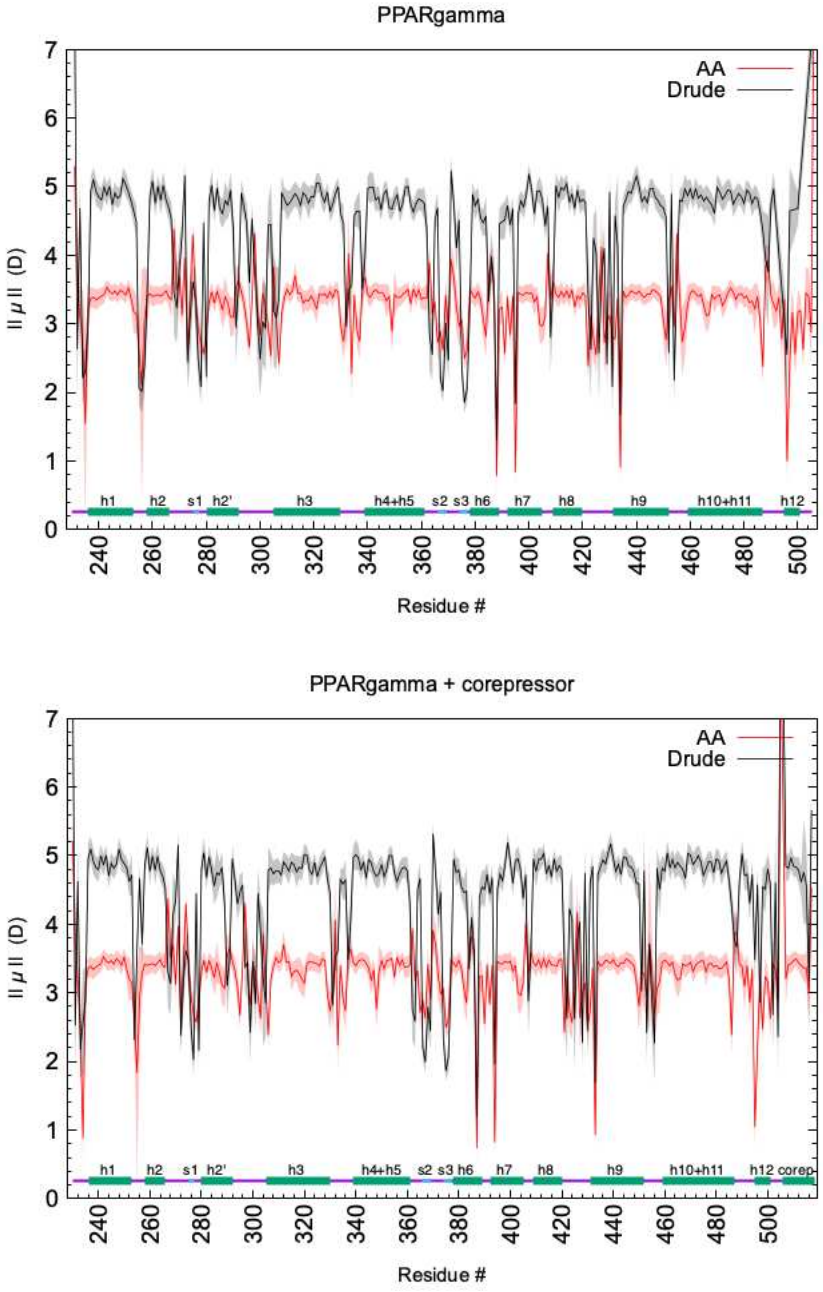


Figure 55. By-residue dipole moments of PPARg. Top figure is for the apo protein and the bottom figure is for the LBD in complex with the co-repressor peptide.

We calculated dipole moment timeseries from the AA and Drude simulations of the apo PPARγ protein and the PPARγ protein complexed to the co-repressor peptide (Fig 54). For the PPARγ apo system, we see lower dipole values in simulations with AA force field, with the average of 247 D, compared to the Drude force field, where the average value is 329 D. The

calculations of PPARγ corepressor bound system follow the same trend, where the average values are 240D and 332D for the AA and the Drude FF, respectively.

Concerning the by-residue average dipole moment (Fig. 55), systematically, the dipole moments in alpha helices are larger in the Drude simulations than in the AA force field simulations for PPARγ.

3.3. Correlated motions

Correlated motions are important for understanding how the motions in different regions of the protein are coupled to other regions and how those coupled motions change in response to different perturbations, such as ligand binding. Changes in the correlated motions can effectively occur over long distances. In its simplest interpretation, this could correspond to allostery in the absence of conformational changes (Cooper and Dryden 1984a). It is therefore important to identify the residues involved in this transmission of structural dynamic information. This information can be obtained by calculating the cross-correlations, which complement the fluctuation analysis presented above by providing information on correlated motions as calculated by Eq. 38. From the C_{ij} correlation coefficients, which are organized as a matrix, a cross-correlation map was calculated using a color-coded 2D representation. These calculations find use in many different applications (Cote et al. 2017; Gaillard et al. 2007; Fidelak et al. 2010).

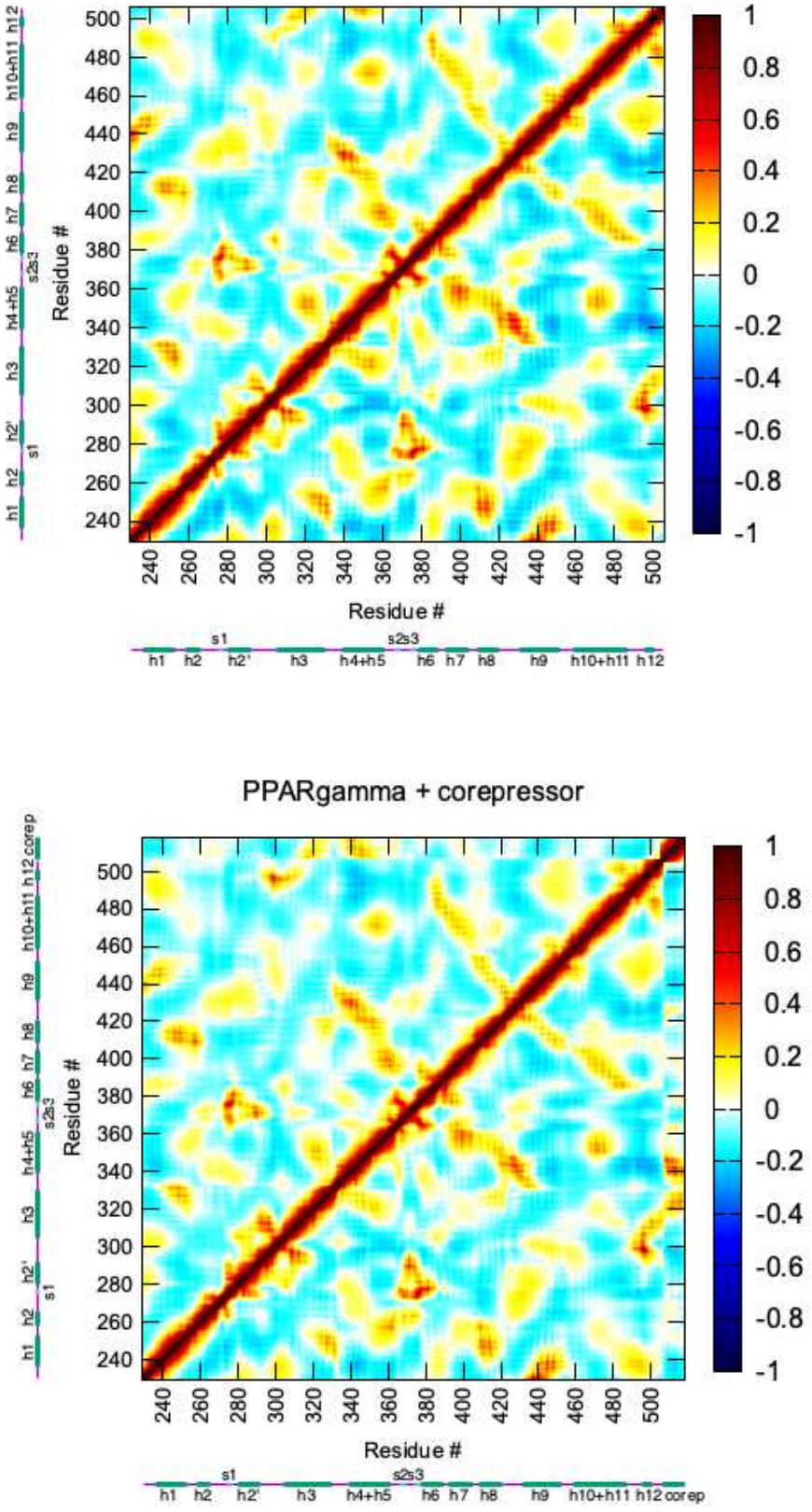


Figure 56. Correlated motions calculated from the simulation of the PPARγ LBD apo form (top), and corepressor bound form (bottom). Correlated motion maps are represented with a color code related to the sign and intensity of correlations (ranging from dark blue for perfect anticorrelations to dark red for perfect correlations). The secondary structural elements are indicated(a) from the Drude simulation (upper triangle), compared to the AA simulations, lower triangle.

We assessed the effects of polarization on the collective motions of the PPARγ LBD by calculating the cross-correlations from the MD simulations (Fig. 56). The lower right triangle of the maps corresponds to the correlation map for PPARγ calculated from the simulation using the AA force field, while the upper left triangle corresponds to the correlated motions of PPARγ calculated from the Drude simulations. The calculations were done for both the apo (top) and corepressor-bound forms (bottom). Regarding the general aspect of the correlation maps, for both forms, we notice a great similarity between the two. The differences are noticeable regarding the intensities of the correlations, as they appear damped in the Drude simulations, to the point that in particular regions, correlation islands disappear. In the case of the PPARγ apo system, the most significant isles represent the correlation between H12 - Ω loop residues, and H1-H9. While strongly present in the AA maps, only traces of the isles are present in the Drude simulations. Concerning the PPARγ bound to the corepressor, the most important distinction in the correlated motions is observed between helices 1 and 10, and in the motions between the H12 – H3 and H12 - H4. The peptide itself seems to be positively correlated to the C-ter of H3 and the N-ter of H4, which is a functionally important interaction, considering that the H3, H4 and H12 constitute the platform for corepressor binding.

3.3.1. Community Network Analysis

To further interpret the consequences of the long-range correlated motions, we performed a community network analysis (CNA). Maps from a CNA are derived from a functional clustering of correlated motions obtained from MD simulations. It has been shown that this type of analysis, based on the Gervan-Newman algorithm, can be used to interpret long range communication and dynamic allostery of proteins (Madan et al. 2023; Lesgidou and Vlassi 2024). Community maps can help interpret how different parts of proteins move together and how changes in one part of a protein can affect the dynamics of distant sites. Communities highlight regions of the protein that exhibit collective movements and may represent functionally important domains or allosteric communication pathways. Using correlation matrices calculated from our MD simulations and the R package *bio3D* (Grant, Skjærven, and

Yao 2021), we obtained coarse grained networks of dynamically coupled communities. Community network maps are depicted using coloured spheres mapped onto the average 3D structure, in tube representation. The size of a node is related to the size of its community and larger spheres depict larger number of residues. The edges connect coupled communities, where thicker edges correspond to higher degree of correlation.

The CNA of the PPARγ apo system using the correlated motions calculated and averaged from the 3 replicas of molecular dynamics simulations detected 13 nodes for the AA simulations and 11 nodes for Drude simulations (Fig. 57). The specific compositions of the nodes are given in Table 8. The nodes correspond largely to entire secondary structure elements, mostly helices, however, the AA simulations have four nodes encompassing spatially adjacent residues belonging to different helices: the first community regroups the N-terminal residues with residues from H9 (node 1, 23 residues (230:234, 432:449)), the second one regroups the large node containing H2' and Ω loop, the beta sheet and helix 6 (node 4, 45 residues (274:296, 363:384)), the third one regroups the Ω loop C-terminal residues with H12 residues (node 6, 27 residues (298 to 306, 488 to 505)), and the fourth one regroups the loop between H3 and H4 along with H4 with the H8 – H9 loop (node 8, 25 residues (333:348, 423:431)). Node 4 is the one most coupled to other nodes in the apo structure. Interestingly, node 4 shows a relatively weak direct coupling to node 6, which contains the transcriptionally important H12, but it has a strong coupling via node 5, which encompasses the N-terminal end of H3. This lack of strong direct coupling may be due to the fact that the spatially near loop in node 6 is quite flexible. There is also a relatively strong coupling between the loop H8-H9 (node 8) with the rest of the protein. Interestingly, this loop is known to interact with cyclin H in other nuclear receptors, in particular RARα (Samarut et al. 2011b). The PPARγ LBD is known to interact with cyclin D (Sarruf et al. 2005) in the context of regulating adipogenesis. The nodes encompassing the loop regions at either end of Helix 9 are well connected to the rest of the protein and are known to be important in the allostery related to phosphorylation in other receptors (Chebaro et al. 2017; 2013; Samarut et al. 2011b).

Table 8. Composition of the nodes from the community network analysis. Isoform PPARγ2 numbering of LBD residues: 230 - 505. Corepressor peptide is numbered from 506 - 517. Size represents the number of amino acids in a node. Node members are the numbers of amino acids belonging to a node.

PPAR_AA_apo				**PPAR_AA_corep**		
node id	size	members		node id	size	members
1	23	c(230:234, 432:449)		1	28	c(230:232, 431:455)
2	22	235:256		2	22	233:254
3	17	257:273		3	19	255:273
4	45	c(274:296, 363:384)		4	32	c(274:289, 372:378)
5	14	c(297, 307:319)		5	38	c(290:308, 487:505)
6	27	c(298:306, 488:505)		6	24	309:332
7	13	320:332		7	17	333:349
8	25	c(333:348, 423:431)		8	13	350:362
9	14	349:362		9	13	379:391
10	21	385:405		10	16	392:407
11	17	406:422		11	23	408:430
12	15	450:464		12	31	456:486
13	23	465:487		13	12	506:517
PPAR_Drude_apo				**PPAR_Drude_corep**		
node id	size	members		node id	size	members
1	25	230:254		1	25	230:254
2	20	255:274		2	20	255:274
3	33	c(275:284, 362:384)		3	25	c(275:277, 362:383)
4	21	285:305		4	13	278:290
5	26	306:331		5	14	291:304
6	30	332:361		6	27	305:331
7	22	385:406		7	30	332:361
8	15	407:421		8	23	384:406
9	28	422:449		9	16	407:422
10	17	450:502		10	31	423:453
11	39	467:505		11	34	454:487
				12	18	488:505
				13	12	506:517

The CNA of the Drude simulation of the apo PPARγ LBD generally shows smaller nodes than those observed in the AA simulation. The largest node, node 3 (residues 274:283, 361:383) encompasses the residues of H6, the β sheet and some of the Ω loop; the equivalent node in the AA simulation is node 4, however, the node from the Drude simulation is smaller. Many of the other nodes are along secondary structure elements. As in the AA simulation, there is no direct coupling between node 3 and the helix 12 region of PPAR. In the Drude results, the coupling passes through 3 to 4 nodes depending on the path, while in the AA simulation, the coupling was either direct (weak) or through just one additional node. In addition to the couplings being different between the AA and the Drude simulation, these results suggest that the coupling between different regions of the PPAR ligand binding domain is less strong in the Drude simulations than in the AA simulations.

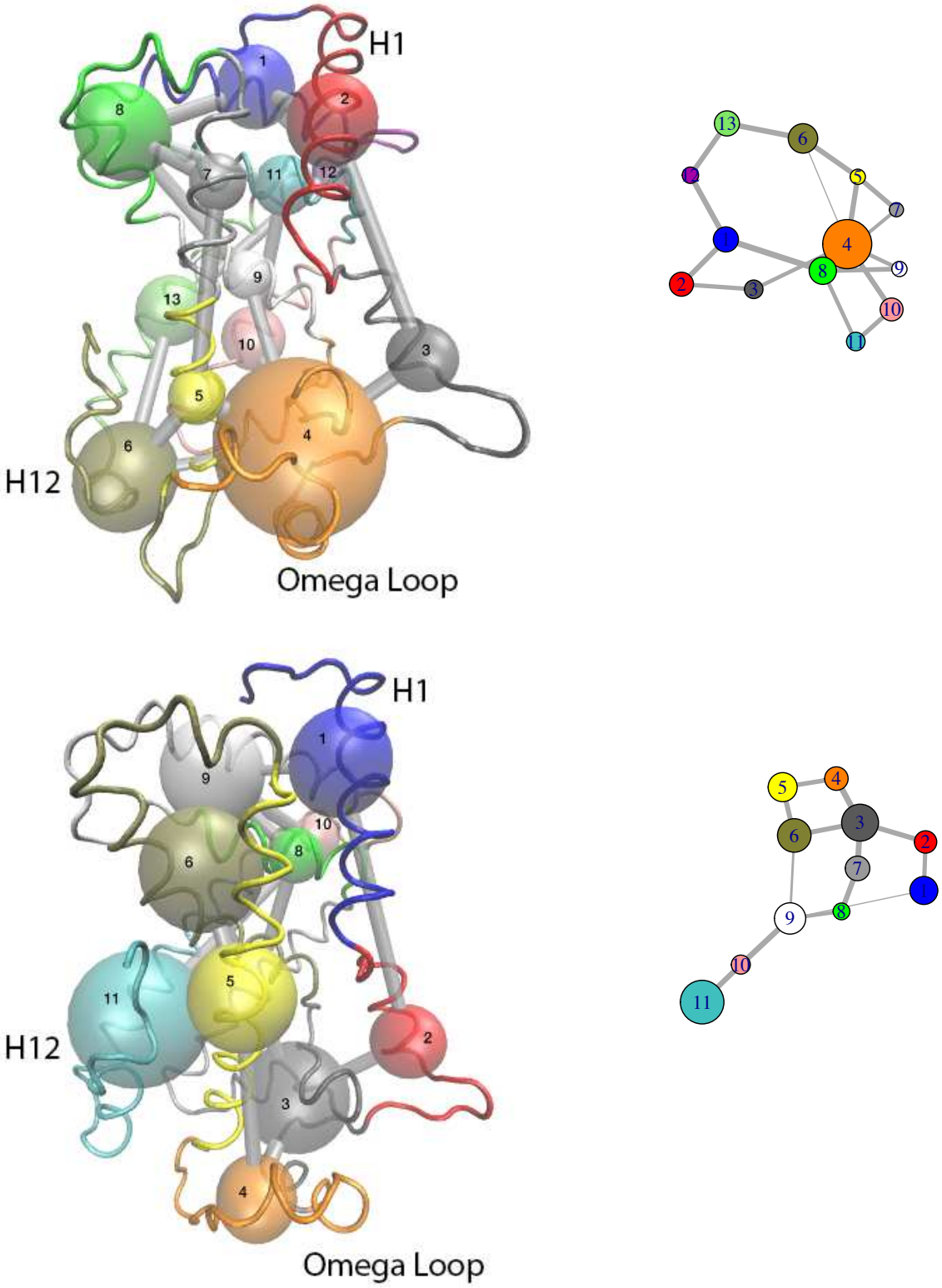


Figure 57. Community network analysis of the PPARγ LBD. On the left, the colored nodes are superposed on the protein backbone structure, represented as a tube and colored according to the nodes. The edges are denoted as grey connections between the nodes, where the thickness indicates the strength of the correlation between two nodes. On the right is the network representation. In (top) are the results from the AA simulation, and in (bottom) are the results from the Drude simulations.

For the PPARγ/corepressor complex, the CNA identified 13 community nodes for correlated motions from both the AA and the Drude simulations (Table 8, Fig. 58). The specific node compositions are provided in supplementary material. In both cases, the protein has 12 nodes, and the corepressor peptide forms its own node. More of the nodes identified in the AA simulations include sequentially distant, but spatially near resides (nodes 1, 4 and 5), while in the Drude simulation, there is only one node that includes sequentially distant, but spatially near residues (node 3). In the AA system, there are three nodes which connect to neighbouring nodes: the first groups N-terminal residues with H9 (node 1, 229:231, 430:454), the second groups the N-terminal of the Ω loop with the β sheet (node 4, 32 residues, (45:60, 134:149)), and the third one associates the Ω loop with H12 (node 5, 38 residues (61:79, 258:276)). In the Drude simulation, node 3 encompasses the β sheet and the N-terminal end of the Ω loop.

Concerning the interconnectivity of the nodes, the CNA shows that many of the node interconnections are along the secondary structure elements, but the AA simulations display several edges between nodes beyond secondary structure. The node organization for AA is more global as opposed to the Drude simulation results, where there is less interconnectivity, and the network is more extended. Central to the interconnectivity in PPARγ is the node that encompasses the β sheet and part of the Ω loop region. This node forms a hub through which many edges connect.

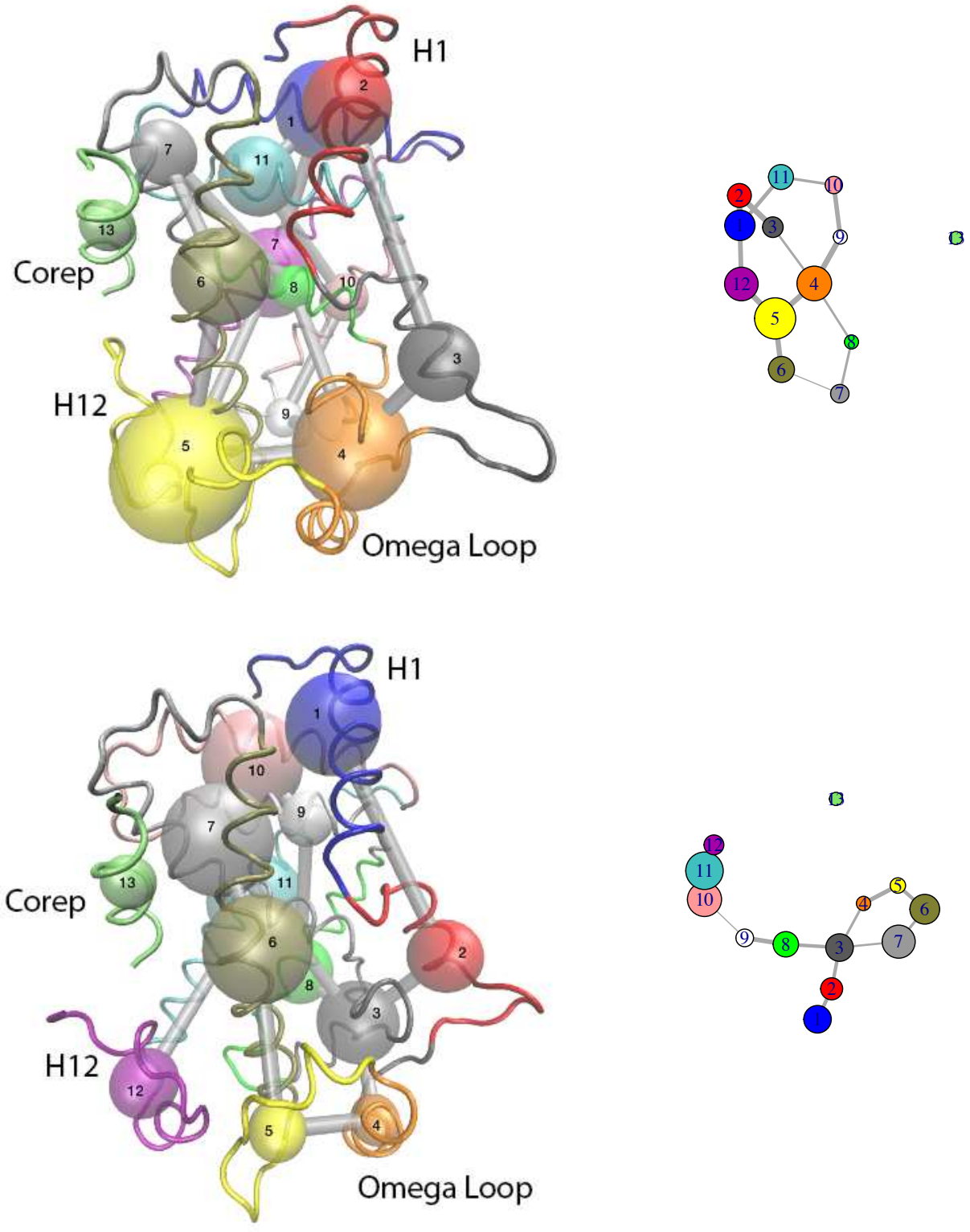


Figure58. Community network analysis of PPARgamma with corepressor peptide. On the left, the colored nodes are superposed on the protein backbone structure, represented as a tube and colored according to the nodes. The edges are denoted as grey connections between the nodes, where the thickness indicates the strength of the correlation between two nodes. On the right is the network representation. Representation on the PPARγ LBD corepressor bound form, from AA simulation (top), and Drude simulation (bottom).

CNA analysis for both the AA and the Drude simulations show the corepressor peptide as a single node. And we further notice that this node does not form any edges to any nodes in the ligand binding domain of PPARγ. If we look back at the correlation plots of AA simulations, we see positive correlation between the corepressor peptide residues and the N-terminal of H4, while the Drude simulations did not capture these correlations. The correlations were weak and as they did do not go over the 0.5 threshold, they are not represented by an edge. The addition of the corepressor peptide in the AA and Drude simulations does not seem to break the community of the H3 – 4 loop and H4 residues, which forms the corepressor binding platform. In the AA results, the presence of the peptide seems to increase the H12 and Ω loop community, passing from 27 (node 6) to 38 (node 5) residues, and reinforcing their correlations. In the Drude simulations, the addition of the peptide seems to decouple two different communities. First, the H11 – H12 community is split into two separate ones, connected by an edge (from node 11 to nodes 11 and 12). The second community, built around the Ω loop (node 4), is divided into 2 separate nodes (nodes 4 and 5), connected by edges. Other nodes do not seem to be affected by the corepressor addition.

One significant distinction between simulations with two force field concerns the community that represents helix H12. In the AA simulations of both apo and corepressor-bound systems, H12 and one part of the Ω loop are coupled and are therefore represented by one community. This node is of medium size, with 27 residues for the apo form and with 38 residues for the corepressor-bound form. In the Drude simulations, H12, together with H11 make up an individual community, represented by a node containing 39 residues. This community is decoupled from the node encompassing the Ω loop in both systems simulated by the Drude FF meaning there are no edge connections between them. This suggests that the correlations in the Drude simulations are not sufficiently strong to result in the CNA analysis detecting direct communication between them. In the corepressor-bound form, H12 is further decoupled from the H11, having its own community of helix residues connected by an edge to the H11. Furthermore, in both systems simulated with AA FF, the N-terminal residues of PPARγ are grouped in the same community with H9 residues, while in the Drude simulations, these N-terminal residues are in the same community as H1 residues. This coincides with high

RMSF values for the N-terminal residues of the LBD in both Drude simulated systems, see Fig. 52.

Helix H12 represents the activation function 2 in LBDs and therefore is physiologically important for the regulation of PPARγ's transcriptional activity. In the transcriptionally inactive form studied here (apo or corepressor bound forms), the H12 exhibits higher flexibility and is capable of exploring multiple conformations (Chrisman et al. 2018b). In the community analysis of the Drude simulations, we notice the decoupling of the H12 from the other regions, notably the Ω loop and the H11. This suggests that these regions explore different movements which are not directly correlated and display different conformational dynamics. The lack of high correlating communities and the presence of communities largely representative of individual alpha-helices is apparent in the correlation maps, where the Drude simulations display attenuated colours, and thus smaller correlations. It is generally appreciated that around the ligand binding pocket of the PPARγ LBD, the region is flexible in the absence of a ligand, so we would expect a low degree of correlated motion is this area, notably of the functionally relevant helix H12 and the conformationally flexible Ω loop.

3.3.2. Shortest Path Method (SPM)

A second approach for interpreting correlated motions is the Shortest Path Method (SPM), which was used through the online webserver (Casadevall et al. 2024). This tool was used to assess the importance of individual residues, and their pairwise connections, in the structural dynamics of the two proteins. This is in contrast to the community network analysis, which establishes communities around multiple residues. The SPM method produces a network graph based on mean distances and correlation values. The shortest path lengths were calculated using the Dijkstra algorithm (Casadevall et al. 2024). The shortest path is the most direct path following the most significant connection between two residues and shows how the residues are connected in the protein's structural dynamics. The tool is mostly aimed at exploring key residues implicated in enzymatic activity, but here we use it to assess the similarities and differences of simulations using different force fields.

The PPARγ apo system (Fig. 59) shows a graph network connecting different nodes corresponding to same secondary structure elements. For example, if we look at the side view of the structure from AA simulation, we can see a network spanning across the entire helix H10-11, and then continuing connecting the loop and H12, and even further the Ω loop. This suggests a correlation and coupling of these secondary structure elements. On the other hand, the Drude simulation shows no such connection and the functionally important H12 is not coupled to Ω loop movements. Similar observations were made from the community network analysis.

Compared to the apo PPARγ, the SPM paths of PPARγ bound to the corepressor peptide are relatively different for both the AA and Drude calculations (Fig. 60). In this case, we again discern in the case of the AA simulation, the SPM spanning throughout the 'upper' region of the LBD and the one in the 'bottom' region with respect to the illustration, where H12 and the Ω loop are connected. Interestingly, we see short paths between alpha helices.

Drawing similar conclusions, the use of the Drude force field leads to a decoupling of H12 and the Ω loop region in the apo protein; the same observation was made from the community network analysis. The corepressor peptide, even though it was included in the SPM network calculation, does not appear to participate in the shortest path representation. Despite the somewhat correlated motions between the corepressor peptide and helices H3 and H4, the co-repressor peptide is not connected to the rest of the protein in this analysis. A similar conclusion was made the community network analysis, the corepressor peptide does not enter into any communication network. We also notice the absence of the SPM path in the regions of the loop H3 – H4, probably caused by the addition of the corepressor peptide. This suggests that the presence of the peptide, while not directly implicated in a network, will perturb the underlying communication network of PPARγ.

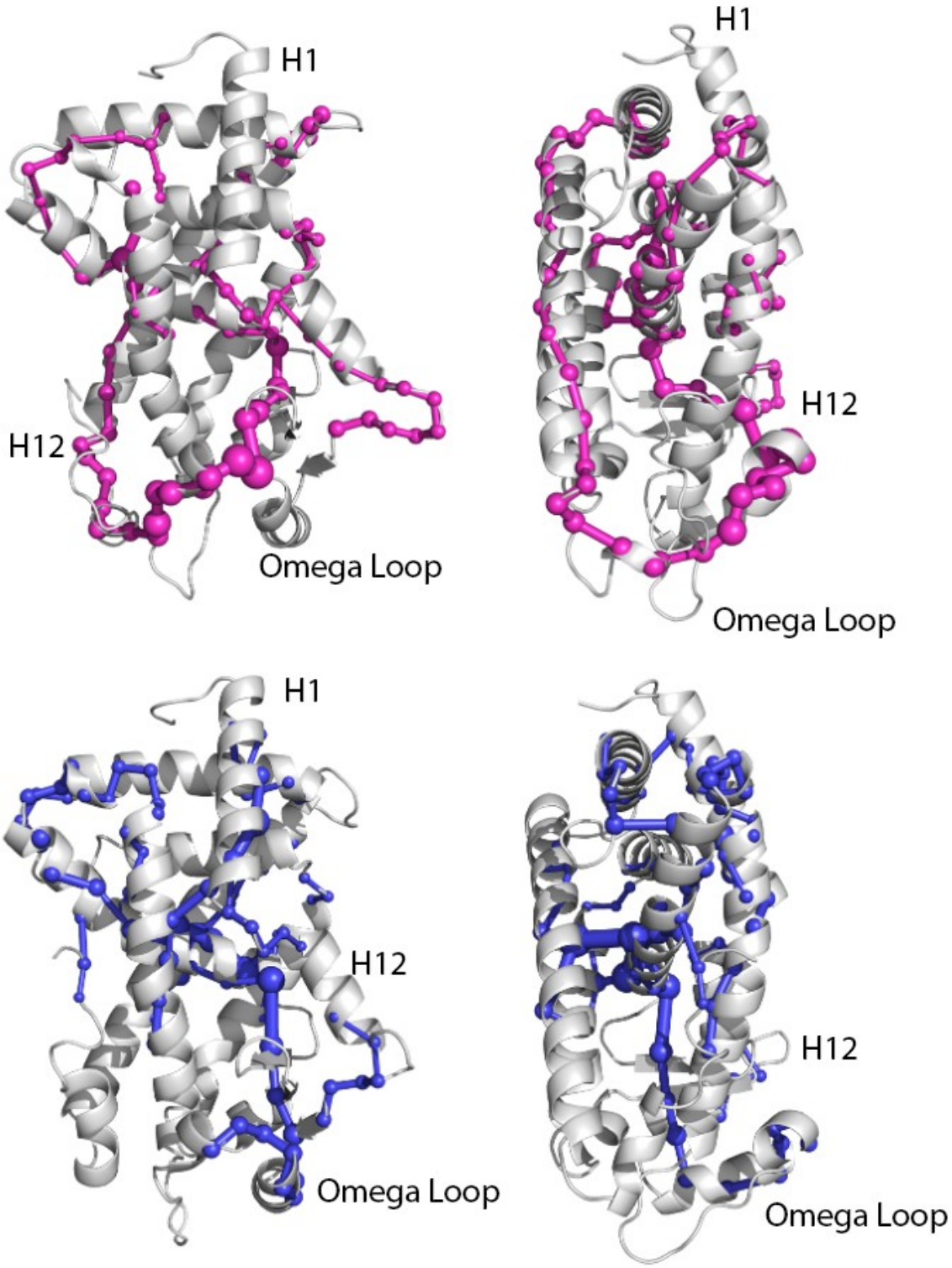


Figure 58. Shortest Path Method ball and stick representation mapped on the front and side views of PPARγ LBD apo form. The AA simulation is at the top (magenta) and the Drude simulation on the bottom (blue).

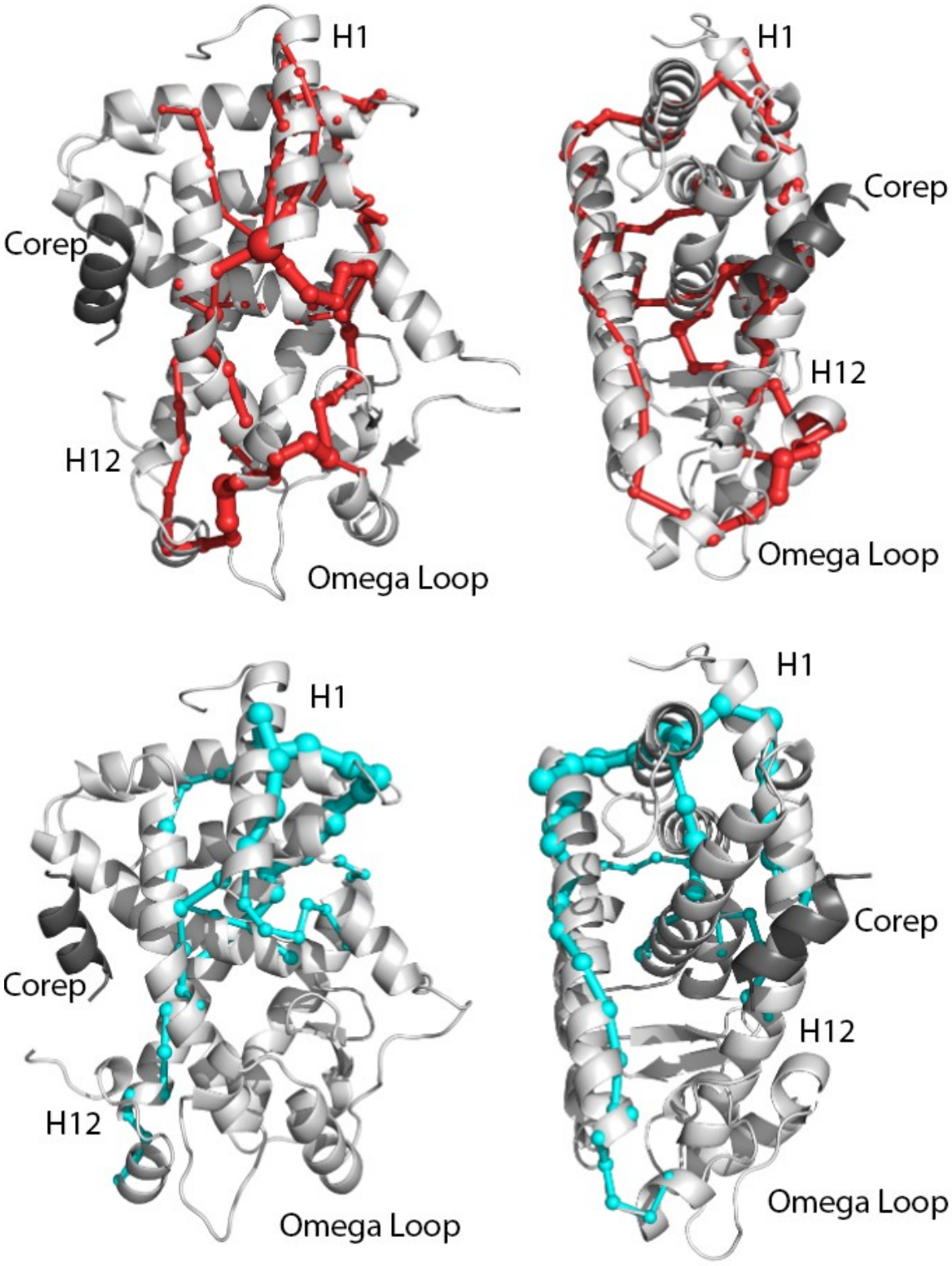


Figure 59. Shortest Path Method ball and stick representation mapped on the front and side views of PPARγ LBD with corepressor peptide bound. The AA simulation is at the top (red) and the Drude simulation on the bottom (cyan).

4. Conclusions

In this work, we used the Drude polarizable force field in molecular dynamics simulations of the ligand binding domain of human PPARγ and compared the results to simulations using the CHARMM all atom atomic force field. We examined the effect of the Drude force field on standard measures of structural dynamics, such as RMSD and RMSF via comparison to simulations using a classical, AA force field. We generally found conformational change leading to a higher RMSD and, in flexible regions of the proteins, greater flexibility. But overall, the trends remained the same. Looking at dipole moments, we confirmed the effect of polarization on amino acid dipole moments, which is influenced by secondary structure.

We also characterized for the first time the effects of using the Drude force field on correlated motions, which have been implicated in the biological function of proteins. The correlated motions were characterized by correlation maps generated by molecular dynamic simulations, by community network analysis and shortest path method (SMP) analysis. The latter two interpret information from the correlated motions obtained from the simulations. The CNA distinguished regions of the proteins where residues interact strongly with each other, and are placed in the same community, from those that interact more weakly. The latter are placed outside of the community and if they are part of another community with sufficiently strong correlations to another community, the information is indicated by connection between the nodes. The analysis further reveals paths through which signals, structural or through interactions with other proteins, can propagate from one region to another. The SPM approach provide more residue-to-residue mapping of the correlated motions, but both approaches provided insights into how protein dynamics map onto the modular organization of the protein and reveal residues or communities of residues that display coordinated motions. Such coordinated motions may underpin allosteric communication.

Analysis of the correlated motions through examination of the correlation maps, the CNA and the SMP analysis based on these correlations show that the use of polarization via the Drude force field affects the low-frequency collective modes by decoupling the motions and generally softening the correlated motions. This suggests that, perhaps the simulated

protein will be less responsive to perturbations introduced, for example, by ligand binding or by the introduction of point mutations. So, perhaps there is an advantage to using the AA force field if one is interested in the studying allosteric behaviour of proteins based on their collective motions. In conclusion, we notice that the major difference arises in regions of the protein that are known from simulations to exhibit more significant flexibility that in other regions of the protein, more precisely at the lower half of the protein, concerning ligand binding pocket, and aforementioned helix H12.

Chapter VI - Conclusions and Perspectives

Increasing attention is being focused on the role of protein structural dynamics in crucial cellular signalling pathways and modulating structural dynamics is becoming an important avenue of exploitation for the discovery of new therapeutic compounds. However, there remains a serious paucity of techniques that permit one to obtain relevant data related to structural dynamics on appropriate timescales. New approaches are needed to both elucidate and measure physical properties directly related to structural dynamics.

In this thesis, the primary objective was to develop an integrated far infrared absorption spectroscopy/molecular dynamics simulations for quantifying and studying collective motions in proteins. Integrating far-infrared (far-IR) measurements with molecular dynamics (MD) simulations offers a powerful, multi-dimensional approach for investigating protein dynamics. Far-IR spectroscopy is sensitive to low-frequency vibrational modes of molecules, including protein motions, conformational changes, and collective dynamics that are difficult to capture using conventional structure-based techniques, but the results are difficult to interpret. Far-IR data also provides crucial experimental validation for molecular dynamics simulations, which offer insights into protein conformational changes, motions and low frequency collective dynamics of proteins. The combination of the two techniques can provide more insight than either technique alone. This was previously demonstrated in a first integrated study of the response of a PDZ domain to the binding of a small peptide, where it was shown that exploitable information concerning changes in low frequency collective motions could be obtained even for proteins where there is no substantial conformational change upon ligand binding (Cote et al. 2017b). The integrated approach allowed for quantification of a mechanism of allostery in a PDZ domain.

We aimed to enlarge the field of application of this integrated approach and to characterize the dynamics of nuclear receptor (NR) proteins. NRs constitute a superfamily of proteins that function as DNA-binding, ligand dependent transcription factors. Being a large, complex protein implicated in transcriptional cascades that underlie many physiological phenomena make them one of the major signal transduction paradigms in metazoans. Indeed, evidence suggests that there exist multiple mechanisms exploiting structural dynamics and allostery that implicate ligand, DNA, co-activator and co-repressor binding, as well as post-

translational modifications. Central to these mechanisms is the ligand binding domain (LBD), which acts as an allosteric hub, transmitting binding events to other protein interfaces and, eventually, other intra- and intermolecular domains.

As the nuclear receptor LBD is a large protein and presents a complex allostery, new developments to the earlier computational approach were introduced to address the complex structural dynamics more accurately. While the approach developed here maintains its reliance on normal mode analysis, which a powerful computational tool for studying protein dynamics, particularly for understanding low-frequency, collective motions, to better adapt to long molecular dynamics simulations, we developed a conformational ensemble approach that includes a large number of structures extracted from long molecular dynamics simulations. This ensemble approach differs from the original approach in that the latter relied on the normal mode analysis of a single structure extracted from a molecular dynamics simulation.

Our ensemble approach, presented in chapter III, leads to more robust results in terms of quantities calculated from normal mode analysis. In chapter IV, we applied the method to the nuclear receptor protein Peroxisome Proliferator-Activated Receptor gamma (PPARγ), a nuclear receptor that is a particularly important target for development of therapeutic compounds for multiple diseases, including diabetes and cancer. We focused on the ligand binding domain in both apo- and holo-wild-type forms, as well as two mutant forms, where we characterized the effects of agonist ligand binding, as well as the effects of gain-of-function and loss-of-function mutations on the structural dynamics. Both far infrared absorption spectroscopy (done by our collaborators) and molecular dynamics simulations were applied to the same systems, giving a unique opportunity to synergistically study the systems. We studied the PPARγ LBD in various physiological states, including apo and holo forms, as well as apo forms of gain- and loss-of function mutations implicated in bladder cancer. The results of the calculations were compared to results from mid- and far-IR measurements. This is the first time that such an approach has been applied to a protein of this complexity. Beyond comparing the calculated IR spectra with the experimental spectra, which provided a robust validation of the computational results, we were able to calculate, from our ensemble approach, the long-range correlated motions that we used in in further analysis by community network analysis, a

tool that is powerful in pulling out information concerning correlated modes and identifying local regions that are highly correlated and their interactions with other highly correlated regions. The important conclusions were that physiologically important regions were identified as being highly correlated to the rest of the domain. In particular, the loop between H8 and H9 formed a highly correlated node that was well connected to the rest of the protein. This region of the LBD is important for signalling cascades in nuclear receptor proteins and its degree of coupled correlation to the rest of the LBD has not been elucidated to such a degree in earlier studies. We also found the regions such as the beta sheet regions were also well connected and that the introduction of agonist ligands shifted the corelation profile toward the dimer interface, which would be important for information transfer during the process of gene transcription.

The results presented in this thesis rely on an accurate representation of the collective motions of the protein. We confirmed (chapter III) that these collective motions are well captured by low-frequency vibrational modes. This raises an important question, particularly in light of current interest in polarizable force fields: What effect does polarization have on the collective motions of a protein, as calculated in molecular dynamics simulations? In chapter V, we presented a study where we carried out a simulation study of the nuclear receptor protein PPARγ and compared the calculated correlated motions from simulations that did not use a polarizable force field to those that did. While in many studies, polarizable force fields show good performance for structural features, such as RMSD or atomic fluctuations, we found in our study that polarization softened the calculated correlations. The correlated motions were fed to a community network analysis and we found there were fewer correlations between different regions of the LBD compared to the classic all atom force field. With softened correlations, interpretations pertaining to allostery are more difficult. Additionally, community network analysis yielded less definitive results when polarization was included and some features that made physiological sense in the context of nuclear receptor proteins were lost. This raises the point that one should be cautious when using a polarizable force field for studies where low frequency collective motions are important.

Through the use of molecular dynamics simulations, the research presented here has provided a deeper understanding of the intrinsic intradomain collective motions of nuclear

receptor proteins. Our results underscore the complex interplay within a single receptor domain as a function of ligand or mutations. These insights are critical for advancing the development of novel therapeutic strategies targeting nuclear receptors, particularly in the context of diseases such as cancer, metabolic disorders, and endocrine diseases. This work demonstrates the power of computational approaches, especially when complementing experimental techniques such as IR spectroscopy. This work may pave the way for more rational drug design based on molecular-level understanding of receptor dynamics.

While the molecular dynamics simulations conducted in this thesis have provided valuable insights into the structural dynamics of nuclear receptor proteins, several aspects remain to be explored. Future studies could expand upon this work by incorporating more diverse ligand libraries to probe a wider range of binding modes and receptor subtypes, particularly in the context of therapeutic design. Additionally, inclusion of coregulator proteins could also enhance our understanding of the couplings between different domains. More extensive studies will be needed before one can predict whether a particular compound will function as a good agonist or antagonist. But more immediately, within a particular family of compounds, one will be able to predict trends in pharmaceutical responses. A particular pipeline could involve the testing of a series of compounds against a nuclear receptor protein and the more promising candidates could undergo further testing by far-IR spectroscopy before being used in more expensive biological tests.

Other improvements could involve the integration of enhanced sampling simulations to access an even larger distribution of structures as well as including structures from multiple minimum energy wells. Continued experimental validation of the computational findings will also be essential for further refining our understanding of nuclear receptor function. Finally, investigating the interactions of nuclear receptors with coactivators and corepressors will be crucial for understanding the full spectrum of receptor-mediated signalling.

This study has implications for drug discovery, particularly in the context of nuclear receptors. By providing a detailed analysis of receptor dynamics, this thesis lays the groundwork for an approach that can contribute to the rational design of small molecules that can modulate receptor activity through structural dynamics. The findings emphasize the

importance of targeting structural dynamics of the receptor rather than relying solely on static models, which may overlook critical functional states. The insights gained here could aid in the design of more potent nuclear receptor modulators for the treatment of diseases such as cancer, diabetes, and autoimmune disorders. Moreover, the computational approaches applied in this work can be extended to other protein systems, further advancing the field of structure-based drug design. Ultimately, the work presented here contributes to the growing understanding of nuclear receptor structural biology and provides a foundation for future investigations into receptor-ligand interactions, eventually with therapeutic implications.

Résumé

Étude de la dynamique structurale du récepteur nucléaire PPARgamma

Chapitre I

Le bon fonctionnement des cellules vivantes repose sur la régulation complexe de l'expression des gènes (Cramer, 2019). Cette régulation est assurée par des protéines appelées facteurs de transcription. Les récepteurs nucléaires (RN) forment la plus grande famille de facteurs de transcription chez les métazoaires. Ils contrôlent de nombreux processus liés au cycle cellulaire, à la différenciation, à l'apoptose, au développement, à la reproduction et à l'homéostasie (Renaud and Moras 2000). Ils régulent l'expression des gènes suite à la fixation de divers ligands lipophiles. Cette activité dépendante du ligand fait des récepteurs nucléaires des cibles centrales pour le développement de médicaments dans de nombreuses maladies, telles que le diabète, l'artériosclérose, les maladies inflammatoires chroniques, le cancer et d'autres. Les récepteurs nucléaires sont des protéines à domaines multiples qui présentent une structure modulaire commune où différentes régions correspondant à des domaines autonomes exécutent des fonctions différentes (Fig. 1A et 1B). Cette organisation structurale et fonctionnelle canonique commence par le domaine N terminal A/B, également appelé domaine de la fonction d'activation 1 (AF- 1), suivi du domaine C, ou domaine de liaison à l'ADN (DBD), de la région charnière D et se terminant par le domaine E ou domaine de liaison au

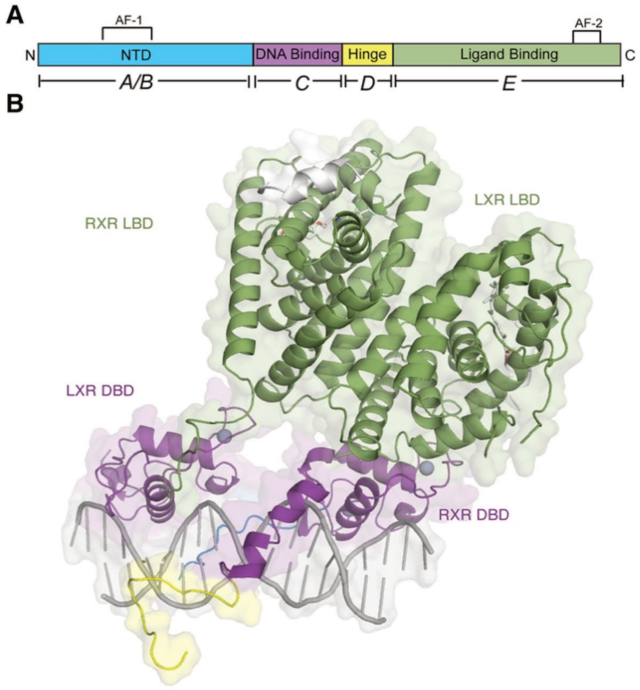


Figure 1.

ligand (LBD), également appelé domaine de la fonction d'activation 2 (AF-2). Certains récepteurs nucléaires possèdent également une extension C-terminale appelée domaine F, mais ce n'est pas le cas pour le PPARγ.

Le domaine A/B N-terminal des récepteurs nucléaires est un domaine intrinsèquement désordonné et contient la fonction d'activation autonome AF-1 qui assure la médiation du recrutement de multiples protéines corégulatrices de la transcription des gènes (Shamilov and Aneskievich 2019). Le domaine de liaison à l'ADN (DBD), ou domaine C, permet la reconnaissance spécifique et la liaison à des séquences d'ADN cibles (Claessens and Gewirth 2004). Il s'agit, d'un point de vue structurel et fonctionnel, d'un domaine hautement conservé, composé de deux hélices α avec des brins β courts et antiparallèles. Il contient deux motifs en doigt de zinc qui sont responsables de la reconnaissance des séquences d'ADN. Afin de réguler la transcription, les récepteurs nucléaires se lient à l'ADN en reconnaissant des séquences spécifiques appelées éléments de réponse aux hormones (HRE).

Le domaine de liaison au ligand (LBD) est un domaine multifonctionnel dont la structure est conservée au sein d'une même espèce et d'une espèce à l'autre. Il se compose de 11 à 13 hélices alpha et d'un petit feuillet bêta (Moras and Gronemeyer 1998). La poche de liaison au ligand du LBD peut accueillir une variété de ligands, y compris des agonistes, des antagonistes et des agonistes inverses. Ce domaine a donc fait l'objet de nombreuses études biochimiques, biophysiques et structurales, qui ont permis d'établir des aspects essentiels de la relation structure-fonction des récepteurs nucléaires. L'activation d'un récepteur nucléaire et le

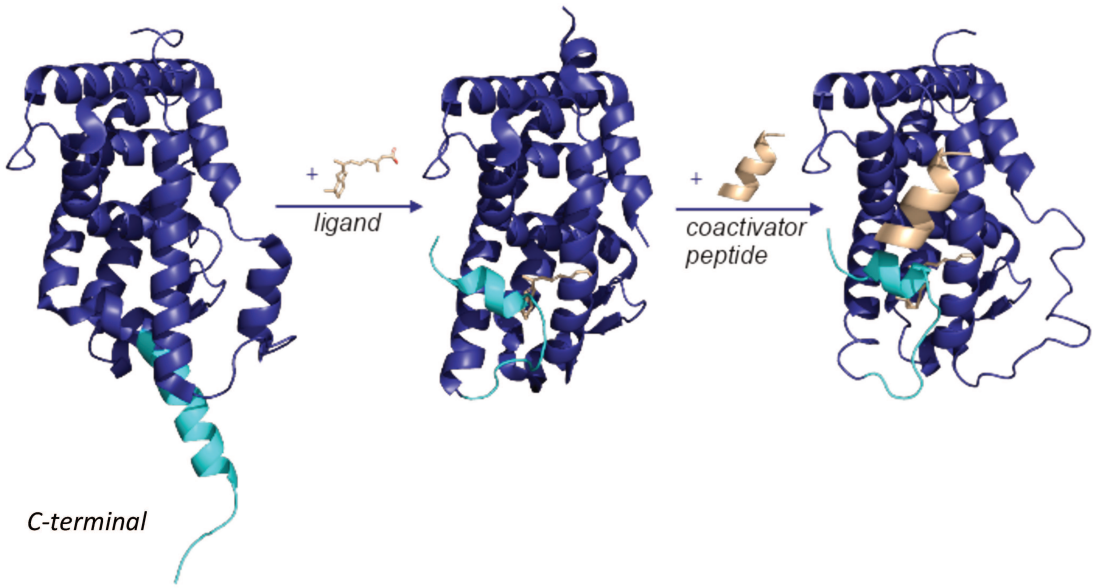


Figure 2.

déclenchement de la transcription d'une protéine sont liés à un changement de conformation de l'hélice H12 de la partie C-terminale du LBD (Fig. 2, l'hélice 12 est colorée en cyan).

L'inclinaison de l'hélice H12 vers le corps du LBD permet à ce domaine d'atteindre une conformation active, liée à la formation d'une plate-forme permettant la liaison de protéines régulatrices appelées coactivateurs, qui modulent positivement la transcription. En revanche, la conformation répressive des récepteurs nucléaires se caractérise par des traits structuraux qui la distinguent de l'état actif. Pour assurer une configuration répressive stable, l'hélice 12 peut soit adopter une conformation où elle est stabilisée contre le LBD, tout en créant une surface compatible avec des protéines à activité corépresseurs (mais incompatible avec les co-activateurs), soit montrer une variation conformationnelle importante, liée à des mouvements où elle peut prendre de multiples conformations, y compris être étendue en solution.

Les travaux de cette thèse portent sur le récepteur nucléaire appelé Peroxisome proliferator - activated receptor gamma (PPARγ), et plus spécifiquement l'isoforme PPARgamma2, qui compte 505 résidus, et qui est considérée comme la forme complète de la protéine. Dans l'organisme, le récepteur PPARγ est principalement exprimé dans les tissus adipeux bruns et blancs, où il joue un rôle crucial dans la régulation de l'adipogenèse et du métabolisme du glucose en favorisant la différenciation des préadipocytes en adipocytes matures. Cette protéine stimule l'absorption du glucose en régulant la sécrétion d'adipocytokines - les médiateurs de l'action de l'insuline (Janani and Ranjitha Kumari 2015). En tant que tel, il présente une cible particulièrement importante pour le développement de traitements contre le diabète, le cancer, l'inflammation et l'hypertension (Berger and Moller 2002). Avec une poche de liaison du ligand d'environ 1300 $Å^3$, PPARγ peut accueillir une

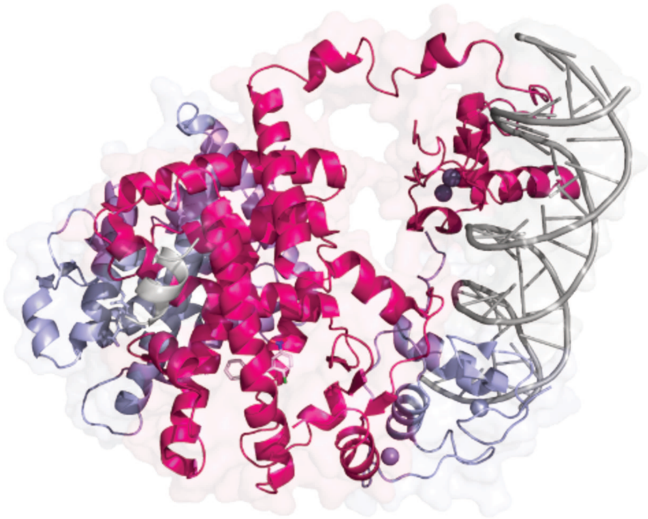


Figure 3.

variété de ligands lipophiles - des acides gras polyinsaturés (AGPI) et leurs dérivés oxydés appelés eicosanoïdes. En l'absence de ligand, PPARγ interagit avec des complexes corépresseurs, tels que NCoR ou SMRT, qui recrutent des enzymes modifiant la chromatine (HAT, ayant une activité histone acétyltransférase) et répriment activement la transcription. Après avoir fixé un ligand agoniste, PPARγ forme un hétérodimère avec le récepteur rétinoïde X (RXR) (Fig. 3, la protéine PPARγ « full-length » en rose). L'hétérodimère peut recruter des coactivateurs tels que le coactivateur PPARγ 1-α (PGC-1α) ou la protéine de liaison E1A p300 (EP300), qui sont capables d'ouvrir la structure de la chromatine et de permettre l'assemblage de complexes protéiques modulateurs de transcription.

Il existe trois structures cristallines du PPARγ complet et plus de 300 structures de son domaine LBD. Les structures du domaine LBD sont principalement résolues dans des positions actives, tandis que la structure du récepteur sous forme inactive reste sujette à débat. Les changements de conformation sont liés à la dynamique structurale des RNs, qui s'est donc rapidement révélée comme un élément indispensable pour comprendre leur fonction. En effet, la dynamique structurale des protéines participe de manière essentielle à leur activité biochimique. La dynamique collective, où différentes régions d'une protéine peuvent se déplacer de manière corrélée, a été identifiée comme un facteur important de la fonction des protéines, par exemple, contribuant à la régulation allostérique (Smith et al. 2016). Ces mouvements collectifs sont des mouvements globaux d'un grand nombre d'atomes composant la protéine, caractérisés par des vibrations de basses fréquences, propres à tout système protéique. Bien que la dynamique structurale des protéines ait été étudiée à l'aide de techniques expérimentales, celles-ci présentent des limitations, et la quantification des changements structuraux et de leurs effets, reste un défi majeur. L'objectif de cette thèse était de développer de nouvelles approches pour quantifier les propriétés physiques directement liées aux changements dans la dynamique structurale collective à basse fréquence en réponse à différents effets, tels que les changements de conformation à grande échelle, la liaison de ligands et les mutations ponctuelles physiologiquement pertinentes.

Chapitre III

La dynamique structurale du domaine de liaisons au ligand du PPARγ a été explorée à l'aide d'approches de modélisation moléculaire et de simulations de dynamique moléculaire. Cette approche, que nous appelons « Ensemble moyenné des modes normaux » permet d'intégrer la spectroscopie d'absorption dans l'infrarouge lointain avec les simulations de dynamique

moléculaire. Les spectroscopies d'absorption dans l'infrarouge lointain sondent la région des basses fréquences du spectre de vibrations et révèlent les modes vibrationnels collectifs dans la région spectrale allant de 0,06 THz (2 cm^{-1}) à 10 THz (333 cm^{-1}), une région de fréquence accessible à l'analyse vibrationnelle basée sur des calculs de mécanique et de dynamique moléculaire. Notre hypothèse est que les techniques spectroscopiques telles que la spectroscopie d'absorption dans l'IR lointain, lorsqu'elles sont combinées à la modélisation moléculaire et à l'analyse structurale et biophysique, peuvent être utilisées pour faire progresser notre compréhension de la réponse des protéines à la liaison d'un ligand, même lorsque les changements structuraux sont minimes ou inexistants (Cooper and Dryden 1984) mais que la dynamique est impactée. Non limitée par la taille des protéines, la mise en place et l'intégration de la spectroscopie dans l'IR lointain avec d'autres approches biophysiques fournira un moyen innovant de quantifier une empreinte vibrationnelle à basse fréquence de la protéine ainsi que les changements de cette empreinte en fonction de la liaison du ligand.

Ce projet est fondé sur des travaux étudiant la liaison des ligands et l'allostérie dans le domaine PDZ de MAGI1 (Cote et al. 2017). Dans ce travail, la spectroscopie d'absorption dans l'IR lointain et les simulations de dynamique moléculaire ont été combinées pour étudier la réponse du domaine PDZ à la liaison d'un petit ligand peptidique et son impact sur la dynamique structurale.

Les modes normaux explorent les mouvements collectifs au sein d'une molécule en examinant ses modes vibrationnels. Ces modes représentent les mouvements naturels à basse fréquence que la molécule peut subir sans modification significative de sa structure. Les modes normaux peuvent jouer un rôle important lorsqu'ils sont utilisés en conjonction avec des simulations de dynamique moléculaire, par exemple en faisant partie intégrante d'algorithmes d'échantillonnage qui améliorent l'exploration de l'espace conformationnel. Nous présentons ici une méthode qui s'appuie sur l'utilisation de structures multiples pour les modes normaux et qui est mieux adaptée à l'analyse de simulations plus longues. Notre approche par l'ensemble moyenné consiste à combiner les données de plusieurs structures issues d'une simulation de dynamique moléculaire extraites du puits le plus peuplé d'une surface d'énergie libre. La surface d'énergie libre (Fig. 4A) est générée à partir de la RMSD et du rayon de giration, comme dans (Cote et al. 2017), les structures sont réorientées dans le même référentiel et une NMA est effectuée sur chaque structure. Nous calculons les modes normaux pour les structures extraites et faisons ensuite la moyenne des résultats. À partir des modes normaux, nous calculons la moyenne de la RMSF (Fig. 4B), les mouvements corrélés (Fig. 4C) et les spectres de l'IR lointain (Fig. 4D) que nous comparerons à l'expérience dans les chapitres suivants. La

combinaison des spectres de plusieurs structures (dans nos études, 100 structures) nous permet d'obtenir un spectre IR moyen d'ensemble. Globalement, cette approche est plus robuste et moins dépendante d'une structure individuelle.

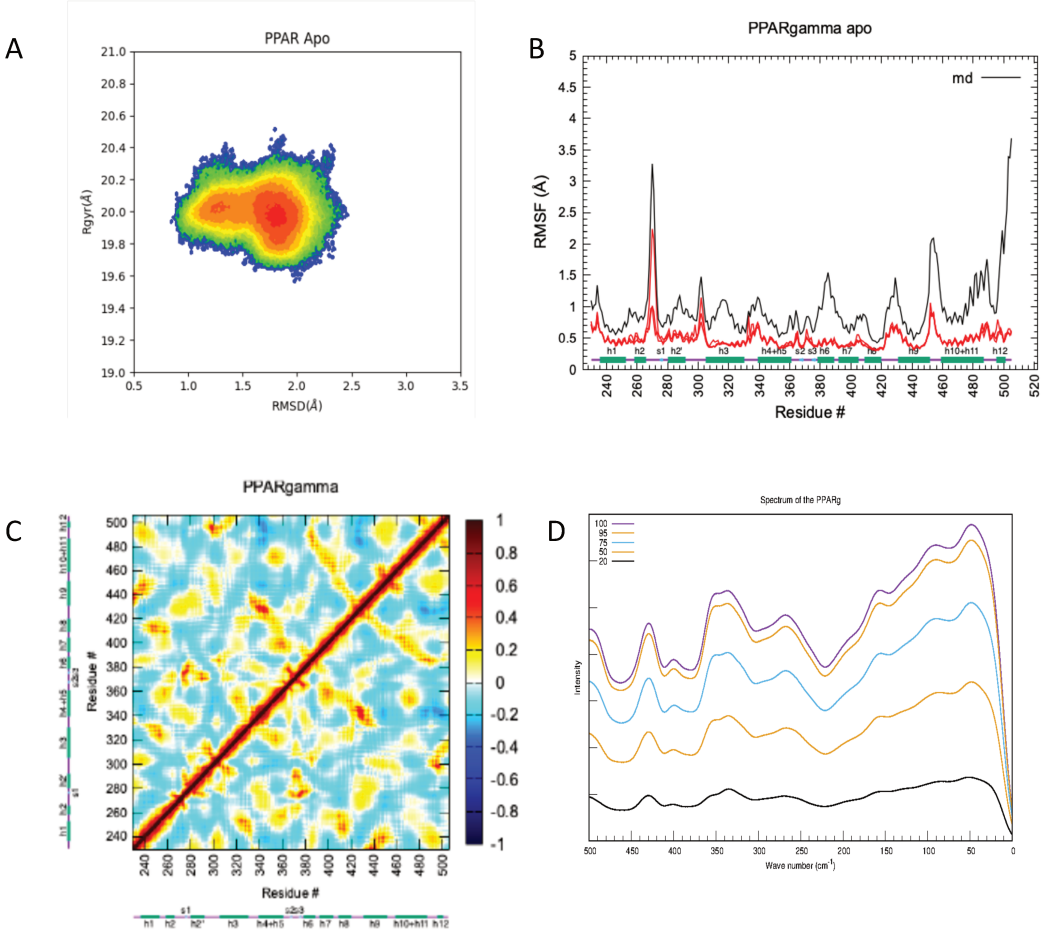


Figure 4.

Chapitre IV

La première étude réalisée au cours de ma thèse a été l'application de cette approche des modes normaux moyennées, au système PPARγ, en combinaison avec les études de spectroscopie IR. Dans ce chapitre, nous présentons les simulations de dynamique moléculaire de PPARγ sous les formes WT apo (Fig. 5) et holo, où PPARγ a été complexé au ligand GW1929. En plus des formes apo et holo WT, nous avons également étudié deux formes mutantes - T475M, qui est la forme mutante connue pour avoir une activité transcriptionnelle accrue même sous forme apo (non ligandée) (Rochel et al. 2019), et F310S, qui est connue pour être une mutation de perte de fonction (Coutos-Thévenot et al. 2019). Les mutations T475M et F310S sont

impliquées dans le cancer de la vessie luminal et basal, respectivement. Grâce à notre collaboration avec l'équipe expérimentale, nous avons démontré pour la première fois la faisabilité de l'utilisation d'une approche IR lointain/MD pour mesurer l'effet de la liaison du ligand par une protéine de la famille des récepteurs nucléaires. L'étude a montré que lors de la liaison d'un ligand, les pics de fréquence les plus bas subissent un déplacement vers des fréquences plus élevées, ce qui suggère que le ligand renforce les interactions et impacte les mouvements vibrationnels.

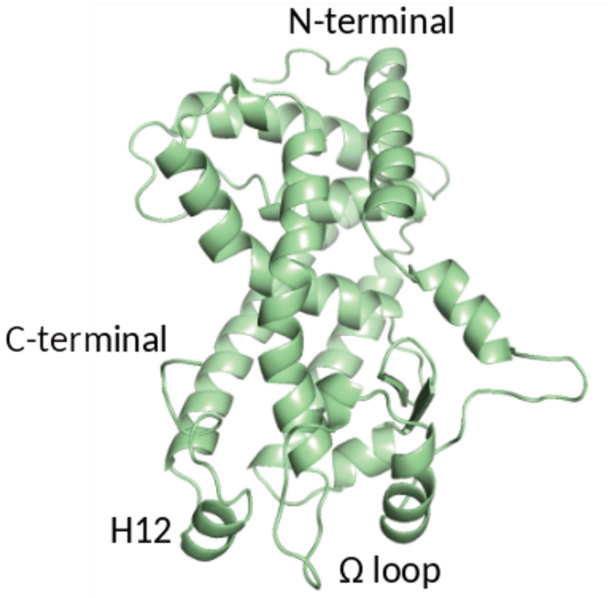


Figure 5.

Pour commencer, nous avons effectué des simulations approfondies de dynamique moléculaire pour toutes les formes étudiées, à partir desquelles nous avons généré des surfaces d'énergie libre basées sur la distribution de la RMSD et du rayon de giration. À partir de ces surfaces, nous avons identifié un certain nombre de puits de potentiel. À partir de ces puits, des structures représentatives ont été extraites et utilisées pour l'analyse des modes normaux, pour les calculs des spectres IR et l'analyse de l'effet collectif par le biais d'une analyse des mouvements corrélés et du réseau d'atomes dont la dynamique est corrélée. Les résultats des simulations ont démontré que dans le domaine de l'IR lointain, les spectres IR calculés présentaient des caractéristiques très similaires à celles observées expérimentalement, en particulier des déplacements des pics spectraux en présence de ligand. Nous avons complété ce résultat par l'analyse des liaisons hydrogène et confirmé que les changements dans les spectres expérimentaux coïncident avec des changements dans les réseaux de liaisons

hydrogènes. Les effets des mutations ponctuelles sur les spectres à basse fréquence ont également été observés et corrélés avec les changements dans la liaison hydrogène. En utilisant notre approche sur les différents systèmes du PPARγ, nous avons calculé des cartes de mouvements corrélés du LBD, qui ont montré des régions de couplage au sein du domaine unique. Nous avons noté que même de petites perturbations dues à la fixation d'un ligand ou à des mutations ponctuelles peuvent avoir un effet sur la dynamique collective de la protéine, ce qui a modulé les couplages entre les régions physiologiquement distinctives du LBD. Pour mieux exploiter les mouvements collectifs calculés, nous avons effectué une analyse de réseau communautaire (CNA, analyse des régions dont les mouvements sont corrélés et couplés) afin d'identifier les régions fortement corrélées qui sont couplées à d'autres régions fortement corrélées du LBD (Fig. 6, à gauche: PPARγ apo, à droite: PPARγ avec GW1929 ligand). Les couplages entre les nœuds fortement corrélés suggèrent un moyen par lequel l'information pourrait passer au sein du domaine protéique en fonctions de changements dans la dynamique structurale.

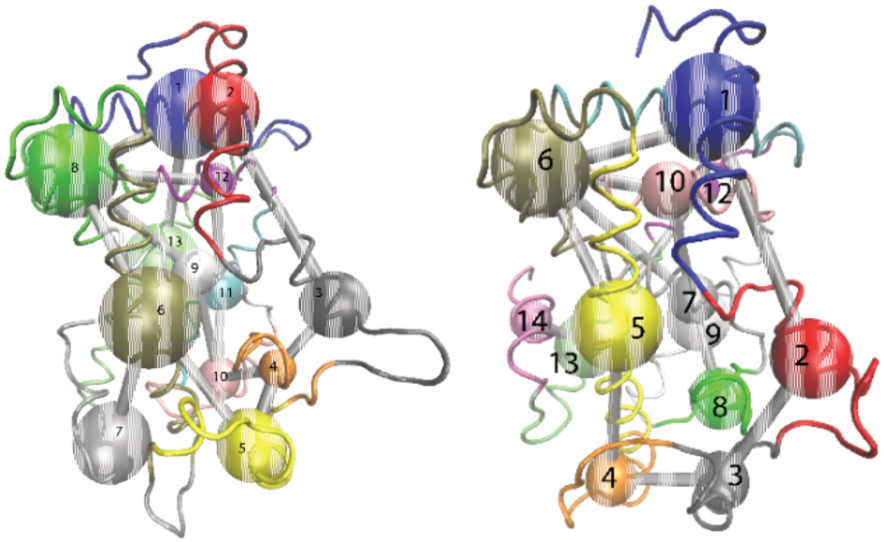


Figure 6.

Nous avons identifié des régions du LBD qui sont fortement couplées au reste du domaine, en particulier la boucle entre H8 et H9, qui est connue pour être physiologiquement importante en tant que site qui lie la cycline au cours d'une cascade de phosphorylation identifiée sur certains récepteurs nucléaires. De plus, nous avons également montré comment la liaison d'un ligand déplace la dynamique corrélée vers la région du LBD qui participe à la dimérisation avec le RXR. En comparant nos résultats avec les résultats de spectroscopie, nous avons vu que des

petits changements liés à des mutations ponctuelles sont mesurables expérimentalement, ce qui nous a permis de valider nos simulations. Nous avons montré que les approches informatiques et expérimentales de l'IR lointain sont complémentaires puisque les deux approches ont montré des tendances similaires dans les spectres mesurés et calculés. Une fois les simulations validées par comparaison avec ces données expérimentales, nous avons pu exploiter davantage les mouvements corrélés obtenus à partir de notre analyse des modes normaux moyennés afin de mieux comprendre la dynamique intra-domaine.

Chapitre V

La seconde étude portait sur l'analyse théorique des effets de la polarisation sur les mouvements collectifs des protéines, introduits via le champ de force polarisable de Drude (Lopes et al., 2013). Dans ce travail, nous avons utilisé le champ de force polarisable de Drude dans des simulations de dynamique moléculaire du domaine de liaison au ligand du PPARγ humain afin de les comparer aux résultats des simulations utilisant le champ de force CHARMM AA "tout atome" (MacKerell et al. 1998). Les deux systèmes de PPARγ LBD utilisés étaient la forme 'apo' (Fig. 5) et la forme liée au peptide corepresseur (Fig 7, le peptide corepresseur en gris). Nous avons examiné l'effet de champ de force Drude sur les mesures



Figure 7.

standard de la dynamique structurale, telles que RMSD et RMSF, en les comparant à des simulations utilisant un champ de force AA classique. Nous avons généralement constaté que le champ de force polarisable montrait une RMSD plus élevée et, dans les régions flexibles des protéines, une plus grande flexibilité. En ce qui concerne les moments dipolaires, nous avons

confirmé l'effet de la polarisation sur les moments dipolaires des acides aminés, qui est influencé par la structure secondaire. Nous avons également caractérisé pour la première fois les effets de l'utilisation du champ de force de Drude sur les mouvements corrélés, qui sont impliqués dans la fonction biologique des protéines. Les mouvements corrélés ont été caractérisés par des cartes de corrélation générées par des simulations de dynamique moléculaire, par l'analyse de réseaux communautaires (CNA) et par l'analyse de la carte du plus court chemin (SMP). Ces deux dernières interprètent les informations des mouvements corrélés obtenues à partir des simulations. L'analyse CNA distingue les régions des protéines où les résidus interagissent fortement entre eux, et sont placés dans la même communauté, de celles qui interagissent plus faiblement. Ces dernières sont placées au sein d'une autre communauté et l'information est indiquée par une connexion entre les nœuds. L'analyse révèle en outre les chemins par lesquels les signaux peuvent se propager d'une région à l'autre.

L'approche SPM fournit davantage de cartographie résidu à résidu des mouvements corrélés, mais les deux approches ont permis de comprendre comment la dynamique des protéines s'inscrit dans l'organisation modulaire de la protéine et révèle des résidus ou des communautés de résidus qui présentent des mouvements coordonnés (Fig. 8, exemple de système PPARγ 'apo', en rose le SPM des simulations avec le champ de force additif, en bleu le SPM des simulations utilisant le champ de force polarisé Drude).

Ces mouvements coordonnés peuvent être à la base de la communication allostérique. L'analyse des mouvements corrélés par l'examen des cartes de corrélation, l'analyse CNA et l'analyse SMP basée sur ces corrélations montrent que l'utilisation de la polarisation via le champ de force de Drude affecte les modes collectifs à basse fréquence en découplant les mouvements et en adoucissant généralement les mouvements corrélés. Cela suggère que la protéine simulée sera peut-être moins sensible aux perturbations introduites, par exemple, par

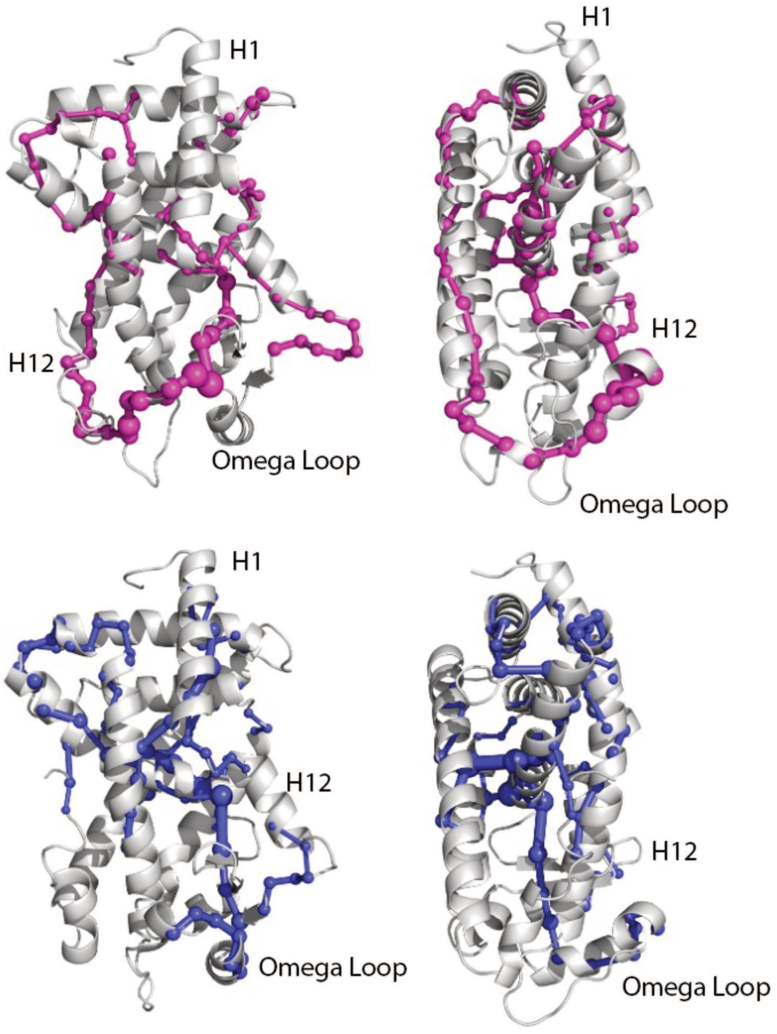


Figure 8.

la liaison d'un ligand ou par l'introduction de mutations ponctuelles. En conclusion, nous remarquons que la différence majeure survient dans des régions de la protéine qui sont connues, d'après les simulations, pour présenter une flexibilité plus importante que dans d'autres régions de la protéine, plus précisément dans la moitié du LBD de la protéine, qui contient la poche de liaison au ligand, et l'hélice H12.

Chapitre VI

En conclusion, l'objectif principal de cette thèse était de développer des protocoles intégrant la spectroscopie d'absorption dans l'infrarouge lointain et des simulations de dynamique moléculaire pour quantifier et étudier les mouvements collectifs dans les protéines. L'intégration de mesures dans l'infrarouge lointain (IR lointain) avec des simulations de dynamique moléculaire (MD) offre une approche puissante et multidimensionnelle pour étudier la dynamique des protéines. Les données de l'IR lointain fournissent également une validation expérimentale cruciale pour les simulations de dynamique moléculaire, qui donnent un aperçu des changements de conformation des protéines, des mouvements et de la dynamique collective à basse fréquence des protéines. La combinaison des deux techniques peut fournir plus d'informations qu'une technique seule. Cela a été démontré précédemment dans une première étude intégrée de la réponse d'un domaine PDZ à la liaison d'un petit peptide, où il a été démontré que des informations exploitables concernant les changements dans les mouvements collectifs à basse fréquence pouvaient être obtenues même pour les protéines où il n'y a pas de changement de conformation substantiel lors de la liaison du ligand. L'approche intégrée a permis de quantifier un mécanisme d'allostérie dans un domaine PDZ. Nous avons voulu élargir le champ d'application de cette approche intégrée et caractériser la dynamique des protéines des récepteurs nucléaires (NR). Les RN constituent une superfamille de protéines qui fonctionnent comme des facteurs de transcription dépendants des ligands et qui se lient à l'ADN. Le fait qu'il s'agisse d'une protéine complexe de grande taille impliquée dans des cascades transcriptionnelles qui sous-tendent de nombreux phénomènes physiologiques en fait l'un des principaux paradigmes de transduction du signal chez les métazoaires. Le domaine de liaison au ligand d'un récepteur nucléaire étant une protéine de grande taille par rapport au domaine PDZ et présentant une allostérie complexe, de nouveaux développements de l'approche computationnelle antérieure ont été introduits afin d'aborder la dynamique structurale complexe avec plus de précision.

Bien que l'approche développée ici continue de s'appuyer sur l'analyse en mode normaux, qui est un outil informatique puissant pour étudier la dynamique des protéines, en particulier pour comprendre les mouvements collectifs à basse fréquence, afin de mieux s'adapter aux longues simulations de dynamique moléculaire, nous avons développé une approche d'ensemble conformationnel qui inclut un grand nombre de structures extraites de longues simulations de dynamique moléculaire. Cette approche d'ensemble diffère de l'approche originale dans la mesure où cette dernière reposait sur l'analyse des modes normaux d'une seule structure extraite d'une simulation de dynamique moléculaire. Notre approche d'ensemble conduit à des résultats plus robustes en termes de quantités calculées à partir de l'analyse des modes normaux. Nous avons ensuite appliqué cette approche à la protéine PPARγ, un récepteur nucléaire. Nous nous sommes concentrés sur le domaine de liaison du ligand dans les formes apo et holo-sauvage, ainsi que dans deux formes mutantes, où nous avons caractérisé les effets de la liaison du ligand agoniste, ainsi que les effets des mutations de gain de fonction et de perte de fonction sur la dynamique structurale. La spectroscopie d'absorption dans l'infrarouge lointain (réalisée par nos collaborateurs) et les simulations de dynamique moléculaire ont été appliquées aux mêmes systèmes, ce qui a donné une occasion unique d'étudier les systèmes de manière synergique. Les résultats des calculs ont été comparés aux résultats des mesures dans l'infrarouge moyen et lointain. C'est la première fois qu'une telle approche est appliquée à une protéine de cette complexité.

Outre la comparaison des spectres IR calculés avec les spectres expérimentaux, qui a fourni une validation solide des résultats informatiques, nous avons pu calculer, à partir de notre approche d'ensemble, les mouvements corrélés à longue portée que nous avons utilisés dans une analyse plus poussée par l'analyse du réseau communautaire, un outil pour extraire des informations concernant les modes corrélés et identifier les régions locales qui sont fortement corrélées et leurs interactions avec d'autres régions. Les principales conclusions sont que des régions importantes sur le plan physiologique ont été identifiées comme étant fortement corrélées au reste du domaine. En particulier, la boucle entre H8 et H9 forme un nœud fortement corrélé qui est bien connecté au reste de la protéine. Cette région du LBD est importante pour les cascades de signalisation dans les protéines des récepteurs nucléaires et son degré de corrélation couplée au reste du LBD n'a pas été élucidé à un tel degré dans des études antérieures. Nous avons également constaté que les régions telles que les régions du feuillet bêta étaient également bien connectées et que l'introduction de ligands agonistes

déplaçait le profil de corrélation vers l'interface du dimère, ce qui serait important pour le transfert d'informations au cours du processus de transcription des gènes.

Ensuite, nous avons présenté une étude de simulation de la protéine du récepteur nucléaire PPARγ et comparé les mouvements corrélés calculés à partir de simulations qui n'utilisaient pas de champ de force polarisable à celles qui en utilisaient un. Alors que dans de nombreuses études, les champs de force polarisables montrent de bonnes performances pour les caractéristiques structurales, telles que la RMSD ou les fluctuations atomiques, nous avons constaté dans notre étude que la polarisation adoucissait les corrélations calculées. Les mouvements corrélés ont été soumis à l'analyse de réseau communautaire (CNA), et nous avons constaté qu'il y avait moins de corrélations entre les différentes régions du LBD par rapport au champ de force classique pour tous les atomes. Avec des corrélations adoucies, les interprétations relatives à l'allostérie sont plus difficiles. En outre, l'analyse du réseau communautaire a donné des résultats moins définitifs lorsque la polarisation a été incluse et certaines caractéristiques qui avaient un sens physiologique dans le contexte des protéines des récepteurs nucléaires ont été perdues. Notre étude a donc démontré l'intérêt de comparer les champs de force polarisables et classiques dans les études dans lesquelles les mouvements collectifs à basse fréquence sont importants, afin d'obtenir une vision plus robuste des réseaux corrélés.

References

Abdelmalak, Silvia, Ningning Yang, and Sidhartha D. Ray. 2024. “Peroxisome Proliferators.” In *Encyclopedia of Toxicology (Fourth Edition)*, edited by Philip Wexler, 399–407. Oxford: Academic Press. https://doi.org/10.1016/B978-0-12-824315-2.00218-9.

Agarwal, Pratul K., Salomon R. Billeter, P. T. Ravi Rajagopalan, Stephen J. Benkovic, and Sharon Hammes-Schiffer. 2002. “Network of Coupled Promoting Motions in Enzyme Catalysis.” *Proceedings of the National Academy of Sciences* 99 (5): 2794–99. https://doi.org/10.1073/pnas.052005999.

Ahsan, Mohd, Chinmai Pindi, and Sanjib Senapati. 2020. “Electrostatics Plays a Crucial Role in HIV-1 Protease Substrate Binding, Drugs Fail to Take Advantage.” *Biochemistry* 59 (36): 3316–31. https://doi.org/10.1021/acs.biochem.0c00341.

Allen, M. P., and D. J. Tildesley. 1989. *Computer Simulation of Liquids*. Clarendon Press.

Allinger, N. L. 1982. “Molecular Mechanics.” American Chemical Society.

Anisimov, Victor M., Guillaume Lamoureux, Igor V. Vorobyov, Niu Huang, Benoît Roux, and Alexander D. MacKerell. 2005. “Determination of Electrostatic Parameters for a Polarizable Force Field Based on the Classical Drude Oscillator.” *Journal of Chemical Theory and Computation* 1 (1): 153–68. https://doi.org/10.1021/ct049930p.

Bahar, Ivet, Mary Hongying Cheng, Ji Young Lee, Cihan Kaya, and She Zhang. 2015. “Structure-Encoded Global Motions and Their Role in Mediating Protein-Substrate Interactions.” *Biophysical Journal* 109 (6): 1101–9. https://doi.org/10.1016/j.bpj.2015.06.004.

Batista, Paulo Ricardo, Charles Herbert Robert, Jean-Didier Maréchal, Meriam Ben Hamida-Rebaï, Pedro Geraldo Pascutti, Paulo Mascarello Bisch, and David Perahia. 2010. “Consensus Modes, a Robust Description of Protein Collective Motions from Multiple-Minima Normal Mode Analysis—Application to the HIV-1 Protease.” *Physical Chemistry Chemical Physics* 12 (12): 2850–59. https://doi.org/10.1039/B919148H.

Berger, Joel, and David E. Moller. 2002. “The Mechanisms of Action of PPARs.” *Annual Review of Medicine* 53 (1): 409–35. https://doi.org/10.1146/annurev.med.53.082901.104018.

Berman, Helen M., John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. 2000. “The Protein Data Bank.” *Nucleic Acids Research* 28 (1): 235–42. https://doi.org/10.1093/nar/28.1.235.

Bianchetti, Laurent, Bianca Wassmer, Audrey Defosset, Anna Smertina, Marion L. Tiberti, Roland H. Stote, and Annick Dejaegere. 2018. “Alternative Dimerization Interfaces in the Glucocorticoid Receptor-α Ligand Binding Domain.” *Biochimica et Biophysica Acta (BBA) - General Subjects* 1862 (8): 1810–25. https://doi.org/10.1016/J.BBAGEN.2018.04.022.

Bour, Gaétan, Emilie Gaillard, Nathalie Bruck, Sébastien Lalevée, Jean-Luc Plassat, Didier Busso, Jean-Pierre Samama, and Cécile Rochette-Egly. 2005a. “Cyclin H Binding to the RARα Activation Function (AF)-2 Domain Directs Phosphorylation of the AF-1

Domain by Cyclin-Dependent Kinase 7." *Proceedings of the National Academy of Sciences* 102 (46): 16608–13. https://doi.org/10.1073/pnas.0505556102.

———. 2005b. "Cyclin H Binding to the RARα Activation Function (AF)-2 Domain Directs Phosphorylation of the AF-1 Domain by Cyclin-Dependent Kinase 7." *Proceedings of the National Academy of Sciences* 102 (46): 16608–13. https://doi.org/10.1073/pnas.0505556102.

Brelivet, Yann, Sabrina Kammerer, Natacha Rochel, Olivier Poch, and Dino Moras. 2004. "Signature of the Oligomeric Behaviour of Nuclear Receptors at the Sequence and Structural Level." *EMBO Reports* 5 (4): 423–29. https://doi.org/10.1038/sj.embor.7400119.

Brooks, B, and M Karplus. 1983. "Harmonic Dynamics of Proteins: Normal Modes and Fluctuations in Bovine Pancreatic Trypsin Inhibitor." *Proceedings of the National Academy of Sciences* 80 (21): 6571–75. https://doi.org/10.1073/pnas.80.21.6571.

Brooks, B. R., C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, et al. 2009. "CHARMM: The Biomolecular Simulation Program." *Journal of Computational Chemistry* 30 (10): 1545–1614. https://doi.org/10.1002/jcc.21287.

Brooks, Bernard R., Robert E. Bruccoleri, Barry D. Olafson, David J. States, S. Swaminathan, and Martin Karplus. 1983. "CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations." *Journal of Computational Chemistry* 4 (2): 187–217. https://doi.org/10.1002/jcc.540040211.

Brooks, C. L., M. Karplus, and B. Montgomery Pettitt. 1988. "Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics" 71 (1): 259. https://doi.org/10.1063/1.2810459.

Brünger, Axel, Charles L. Brooks, and Martin Karplus. 1984. "Stochastic Boundary Conditions for Molecular Dynamics Simulations of ST2 Water." *Chemical Physics Letters* 105 (5): 495–500. https://doi.org/10.1016/0009-2614(84)80098-6.

Bruning, J B, M J Chalmers, S Prasad, S A Busby, T M Kamenecka, Y He, K W Nettles, and P R Griffin. 2007. "Partial Agonists Activate PPARgamma Using a Helix 12 Independent Mechanism." *Structure* 15 (10): 1258–71.

Bugge, Anne, Lars Grøntved, Mads M. Aagaard, Rehannah Borup, and Susanne Mandrup. 2009. "The PPARγ2 A/B-Domain Plays a Gene-Specific Role in Transactivation and Cofactor Recruitment." *Molecular Endocrinology* 23 (6): 794–808. https://doi.org/10.1210/me.2008-0236.

Burley, Stephen K, Charmi Bhikadiya, Chunxiao Bi, Sebastian Bittrich, Henry Chao, Li Chen, Paul A Craig, et al. 2023. "RCSB Protein Data Bank (RCSB.Org): Delivery of Experimentally-Determined PDB Structures alongside One Million Computed Structure Models of Proteins from Artificial Intelligence/Machine Learning." *Nucleic Acids Research* 51 (D1): D488–508. https://doi.org/10.1093/nar/gkac1077.

Casadevall, Guillem, Jordi Casadevall, Cristina Duran, and Sílvia Osuna. 2024. "The Shortest Path Method (SPM) Webserver for Computational Enzyme Design." Edited by

Petersen Eva. *Protein Engineering, Design and Selection* 37 (January):gzae005. https://doi.org/10.1093/protein/gzae005.

Chalmers, Michael J, Scott A Busby, Bruce D Pascal, Graham M West, and Patrick R Griffin. 2011. "Differential Hydrogen/Deuterium Exchange Mass Spectrometry Analysis of Protein–Ligand Interactions." *Expert Review of Proteomics* 8 (1): 43–59. https://doi.org/10.1586/epr.10.109.

Chandra, V, P Huang, Y Hamuro, S Raghuram, Y Wang, T P Burris, and F Rastinejad. 2008. "Structure of the Intact PPAR-Gamma-RXR-Alpha Nuclear Receptor Complex on DNA." *Nature*, 350–56.

Chebaro, Yassmine, Ismail Amal, Natacha Rochel, Cécile Rochette-Egly, Roland H. Stote, and Annick Dejaegere. 2013. "Phosphorylation of the Retinoic Acid Receptor Alpha Induces a Mechanical Allosteric Regulation and Changes in Internal Dynamics." *PLOS Computational Biology* 9 (4): e1003012. https://doi.org/10.1371/journal.pcbi.1003012.

Chebaro, Yassmine, Serena Sirigu, Ismail Amal, Régis Lutzing, Roland H. Stote, Cécile Rochette-Egly, Natacha Rochel, and Annick Dejaegere. 2017. "Allosteric Regulation in the Ligand Binding Domain of Retinoic Acid Receptorγ." *PLOS ONE* 12 (1): e0171043. https://doi.org/10.1371/journal.pone.0171043.

Choi, J H, A S Banks, J L Estall, S Kajimura, P Bostrom, D Laznik, J L Ruas, et al. 2010. "Anti-Diabetic Drugs Inhibit Obesity-Linked Phosphorylation of PPARgamma by Cdk5." *Nature* 466 (7305): 451–56.

Chrisman, Ian M., Michelle D. Nemetchek, Ian Mitchelle S. de Vera, Jinsai Shang, Zahra Heidari, Yanan Long, Hermes Reyes-Caballero, et al. 2018a. "Defining a Conformational Ensemble That Directs Activation of PPARγ." *Nature Communications* 9 (1): 1794. https://doi.org/10.1038/s41467-018-04176-x.

———. 2018b. "Defining a Conformational Ensemble That Directs Activation of PPARγ." *Nature Communications* 9 (1): 1794. https://doi.org/10.1038/s41467-018-04176-x.

Claessens, Frank, and Daniel Gewirth. 2004. "DNA Recognition by Nuclear Receptors." *Essays in Biochemistry* 40 (February):59–72. https://doi.org/10.1042/bse0400059.

Cooper, A., and D. T. Dryden. 1984a. "Allostery without Conformational Change. A Plausible Model." *European Biophysics Journal: EBJ* 11 (2): 103–9. https://doi.org/10.1007/BF00276625.

Cooper, A., and D. T. F. Dryden. 1984b. "Allostery without Conformational Change." *European Biophysics Journal* 11 (2): 103–9. https://doi.org/10.1007/BF00276625.

Cossins, Benjamin P., and Alastair D.G. Lawson. 2015. "Small Molecule Targeting of Protein–Protein Interactions through Allosteric Modulation of Dynamics." *Molecules 2015, Vol. 20, Pages 16435-16445* 20 (9): 16435–45. https://doi.org/10.3390/MOLECULES200916435.

Costa, Mauricio G. S., Paulo R. Batista, Antoniel Gomes, Leonardo S. Bastos, Maxime Louet, Nicolas Floquet, Paulo M. Bisch, and David Perahia. 2023. "MDexciteR:

Enhanced Sampling Molecular Dynamics by Excited Normal Modes or Principal Components Obtained from Experiments." *Journal of Chemical Theory and Computation* 19 (2): 412–25. https://doi.org/10.1021/acs.jctc.2c00599.

Cote, Yoann, Yves Nominé, Juan Ramirez, Petra Hellwig, and Roland H. Stote. 2017a. "Peptide-Protein Binding Investigated by Far-IR Spectroscopy and Molecular Dynamics Simulations." *Biophysical Journal* 112 (12): 2575–88. https://doi.org/10.1016/j.bpj.2017.05.018.

———. 2017b. "Peptide-Protein Binding Investigated by Far-IR Spectroscopy and Molecular Dynamics Simulations." *Biophysical Journal* 112 (12): 2575–88. https://doi.org/10.1016/j.bpj.2017.05.018.

Coutos-Thévenot, Laure, Syrine Beji, Hélène Neyret-Kahn, Quentin Pippo, Jacqueline Fontugne, Judith Osz, Clémentine Krucker, Clarice Dos Santos Groeneveld, et al. 2019. "PPARγ Is a Tumor Suppressor in Basal Bladder Tumors Offering New Potential Therapeutic Opportunities." bioRxiv. https://doi.org/10.1101/868190.

Coutos-Thévenot, Laure, Syrine Beji, Hélène Neyret-Kahn, Quentin Pippo, Jacqueline Fontugne, Judith Osz, Clémentine Krucker, Clarice Dos, et al. 2019. "PPARγ Is a Tumor Suppressor in Basal Bladder Tumors Offering New Potential Therapeutic Opportunities." *bioRxiv*, December, 868190. https://doi.org/10.1101/868190.

Cramer, Patrick. 2019. "Organization and Regulation of Gene Transcription." *Nature* 573 (7772): 45–54. https://doi.org/10.1038/s41586-019-1517-4.

Dijkstra, E. W. 1959. "A Note on Two Problems in Connexion with Graphs." *Numerische Mathematik* 1 (1): 269–71. https://doi.org/10.1007/BF01386390.

"dkNET | dkNET Data Archive: NURSA (Nuclear Receptor Signaling Atlas)." 2024. November 5, 2024. https://dknet.org/about/NURSA_Archive.

Dubois, Vanessa, Jérôme Eeckhoute, Philippe Lefebvre, and Bart Staels. 2020. "Distinct but Complementary Contributions of PPAR Isotypes to Energy Homeostasis." *The Journal of Clinical Investigation* 127 (4): 1202–14. https://doi.org/10.1172/JCI88894.

Ekins, Sean, Sandhya Kortagere, Manisha Iyer, Erica J. Reschly, Markus A. Lill, Matthew R. Redinbo, and Matthew D. Krasowski. 2009. "Challenges Predicting Ligand-Receptor Interactions of Promiscuous Proteins: The Nuclear Receptor PXR." *PLOS Computational Biology* 5 (12): e1000594. https://doi.org/10.1371/journal.pcbi.1000594.

Evans, Ronald M., and David J. Mangelsdorf. 2014. "Nuclear Receptors, RXR, and the Big Bang." *Cell* 157 (1): 255–66. https://doi.org/10.1016/j.cell.2014.03.012.

Fenwick, R. Bryn, Laura Orellana, Santi Esteban-Martín, Modesto Orozco, and Xavier Salvatella. 2014. "Correlated Motions Are a Fundamental Property of β-Sheets." *Nature Communications* 5 (June):4070. https://doi.org/10.1038/ncomms5070.

Fernandez, Elias J. 2018. "Allosteric Pathways in Nuclear Receptors - Potential Targets for Drug Design." *Pharmacology & Therapeutics* 183 (March):152–59. https://doi.org/10.1016/j.pharmthera.2017.10.014.

Fidelak, Jeremy, Silvia Ferrer, Michael Oberlin, Dino Moras, Annick Dejaegere, and Roland H. Stote. 2010a. “Dynamic Correlation Networks in Human Peroxisome Proliferator-Activated Receptor-γ Nuclear Receptor Protein.” *European Biophysics Journal* 39 (11): 1503–12. https://doi.org/10.1007/s00249-010-0608-9.

———. 2010b. “Dynamic Correlation Networks in Human Peroxisome Proliferator-Activated Receptor-γ Nuclear Receptor Protein.” *European Biophysics Journal* 39 (11): 1503–12. https://doi.org/10.1007/s00249-010-0608-9.

Foggia, Michele Di, Paola Taddei, Armida Torreggiani, Monica Dettin, and Anna Tinti. n.d. “Self-Assembling Peptides for Biomedical Applications: IR and Raman Spectroscopies for the Study of Secondary Structure.”

Fratev, Filip. 2016. “PPARγ Helix 12 Exhibits an Antagonist Conformation.” *Physical Chemistry Chemical Physics* 18 (13): 9272–80. https://doi.org/10.1039/C5CP06729D.

Fratev, Filip, Ivanka Tsakovska, Merilin Al Sharif, Elina Mihaylova, and Ilza Pajeva. 2015. “Structural and Dynamical Insight into PPARγ Antagonism: In Silico Study of the Ligand-Receptor Interactions of Non-Covalent Antagonists.” *International Journal of Molecular Sciences* 16 (7): 15405–24. https://doi.org/10.3390/ijms160715405.

Freedman, Leonard P., Ben F. Luisi, Z. Richard Korszun, Ravi Basavappa, Paul B. Sigler, and Keith R. Yamamoto. 1988. “The Function and Structure of the Metal Coordination Sites within the Glucocorticoid Receptor DNA Binding Domain.” *Nature* 334 (6182): 543–46. https://doi.org/10.1038/334543a0.

Frkic, Rebecca L., Jordan L. Pederick, Aimee J. Horsfall, Blagojce Jovcevski, Elise E. Crame, Wioleta Kowalczyk, Tara L. Pukala, et al. 2023. “PPARγ Corepression Involves Alternate Ligand Conformation and Inflation of H12 Ensembles.” *ACS Chemical Biology* 18 (5): 1115–23. https://doi.org/10.1021/acschembio.2c00917.

Gabba, Matteo, Simón Poblete, Tobias Rosenkranz, Alexandros Katranidis, Daryan Kempe, Tina Züchner, Roland G. Winkler, Gerhard Gompper, and Jörg Fitter. 2014. “Conformational State Distributions and Catalytically Relevant Dynamics of a Hinge-Bending Enzyme Studied by Single-Molecule FRET and a Coarse-Grained Simulation.” *Biophysical Journal* 107 (8): 1913–23. https://doi.org/10.1016/j.bpj.2014.08.016.

Gaillard, Emilie, Nathalie Bruck, Yann Brelivet, Gaétan Bour, Sébastien Lalevée, Annie Bauer, Olivier Poch, Dino Moras, and Cécile Rochette-Egly. 2006. “Phosphorylation by PKA Potentiates Retinoic Acid Receptor α Activity by Means of Increasing Interaction with and Phosphorylation by Cyclin H/Cdk7.” *Proceedings of the National Academy of Sciences* 103 (25): 9548–53. https://doi.org/10.1073/pnas.0509717103.

Germain, Pierre, and William Bourguet. 2013. “Chapter 2 - Dimerization of Nuclear Receptors.” In *Methods in Cell Biology*, edited by P. Michael Conn, 117:21–41. Receptor-Receptor Interactions. Academic Press. https://doi.org/10.1016/B978-0-12-408143-7.00002-5.

Germain, Pierre, Bart Staels, Catherine Dacquet, Michael Spedding, and Vincent Laudet. 2006. “Overview of Nomenclature of Nuclear Receptors.” *Pharmacological Reviews* 58 (4): 685–704. https://doi.org/10.1124/pr.58.4.2.

Gervois, Philippe, Inés Pineda Torra, Jean-Charles Fruchart, and Bart Staels. 2000. “Regulation of Lipid and Lipoprotein Metabolism by PPAR Activators” 38 (1): 3–11. https://doi.org/10.1515/CCLM.2000.002.

Gewirth, D. T., and P. B. Sigler. 1995. “The Basis for Half-Site Specificity Explored through a Non-Cognate Steroid Receptor-DNA Complex.” *Nature Structural Biology* 2 (5): 386–94. https://doi.org/10.1038/nsb0595-386.

Glembo, Tyler J., Daniel W. Farrell, Z. Nevin Gerek, M. F. Thorpe, and S. Banu Ozkan. 2012. “Collective Dynamics Differentiates Functional Divergence in Protein Evolution.” *PLOS Computational Biology* 8 (3): e1002428. https://doi.org/10.1371/journal.pcbi.1002428.

Goswami, Devrishi, Celetta Callaway, Bruce D. Pascal, Raj Kumar, Dean P. Edwards, and Patrick R. Griffin. 2014. “Influence of Domain Interactions on Conformational Mobility of the Progesterone Receptor Detected by Hydrogen/Deuterium Exchange Mass Spectrometry.” *Structure* 22 (7): 961–73. https://doi.org/10.1016/j.str.2014.04.013.

Gowers, Richard J., Max Linke, Jonathan Barnoud, Tyler J. E. Reddy, Manuel N. Melo, Sean L. Seyler, Jan Domański, et al. 2016. “MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations.” In *Proceedings of the 15th Python in Science Conference*, edited by Sebastian Benthall and Scott Rostrup, 98–105. https://doi.org/10.25080/Majora-629e541a-00e.

Gowers, Richard, Max Linke, Jonathan Barnoud, Tyler Reddy, Manuel Melo, Sean Seyler, Jan Domański, et al. 2016. “MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations.” In , 98–105. Austin, Texas. https://doi.org/10.25080/Majora-629e541a-00e.

Grant, Barry J., Lars Skjærven, and Xin-Qiu Yao. 2021. “The Bio3D Packages for Structural Bioinformatics.” *Protein Science : A Publication of the Protein Society* 30 (1): 20. https://doi.org/10.1002/pro.3923.

Gronemeyer, Hinrich, Jan-Åke Gustafsson, and Vincent Laudet. 2004. “Principles for Modulation of the Nuclear Receptor Superfamily.” *Nature Reviews Drug Discovery* 3 (11): 950–64. https://doi.org/10.1038/nrd1551.

Grudinin, Sergei, Elodie Laine, and Alexandre Hoffmann. 2020. “Predicting Protein Functional Motions: An Old Recipe with a New Twist.” *Biophysical Journal* 118 (10): 2513–25. https://doi.org/10.1016/j.bpj.2020.03.020.

Haile, J.M. 1997. *Molecular Dynamics Simulation: Elementary Methods*. Wiley.

Harastani, Mohamad, Rémi Vuillemot, Ilyes Hamitouche, Nima Barati Moghadam, and Slavica Jonic. 2022. “ContinuousFlex: Software Package for Analyzing Continuous Conformational Variability of Macromolecules in Cryo Electron Microscopy and

Tomography Data." *Journal of Structural Biology* 214 (4): 107906. https://doi.org/10.1016/j.jsb.2022.107906.

Harder, Edward, Victor M. Anisimov, Igor V. Vorobyov, Pedro E. M. Lopes, Sergei Y. Noskov, Alexander D. MacKerell, and Benoît Roux. 2006. "Atomic Level Anisotropy in the Electrostatic Modeling of Lone Pairs for a Polarizable Force Field Based on the Classical Drude Oscillator." *Journal of Chemical Theory and Computation* 2 (6): 1587–97. https://doi.org/10.1021/ct600180x.

Heinig, Matthias, and Dmitrij Frishman. 2004a. "STRIDE: A Web Server for Secondary Structure Assignment from Known Atomic Coordinates of Proteins." *Nucleic Acids Research* 32 (Web Server issue): W500. https://doi.org/10.1093/nar/gkh429.

———. 2004b. "STRIDE: A Web Server for Secondary Structure Assignment from Known Atomic Coordinates of Proteins." *Nucleic Acids Research* 32 (suppl_2): W500–502. https://doi.org/10.1093/nar/gkh429.

Heppner, Thomas J., Adrian D. Bonev, Delrae M. Eckman, Maria F. Gomez, Georgi V. Petkov, and Mark T. Nelson. 2004. "Novel PPARγ Agonists GI 262570, GW 7845, GW 1929, and Pioglitazone Decrease Calcium Channel Function and Myogenic Tone in Rat Mesenteric Arteries." *Pharmacology* 73 (1): 15–22. https://doi.org/10.1159/000081070.

Hernandez-Quiles, Miguel, Marjoleine F. Broekema, and Eric Kalkhoven. 2021a. "PPARgamma in Metabolism, Immunity, and Cancer: Unified and Diverse Mechanisms of Action." *Frontiers in Endocrinology* 12 (February). https://doi.org/10.3389/fendo.2021.624112.

———. 2021b. "PPARgamma in Metabolism, Immunity, and Cancer: Unified and Diverse Mechanisms of Action." *Frontiers in Endocrinology* 12 (February). https://doi.org/10.3389/fendo.2021.624112.

Holzer, Guillaume, Gabriel V. Markov, and Vincent Laudet. 2017. "Chapter One - Evolution of Nuclear Receptors and Ligand Signaling: Toward a Soft Key–Lock Model?" In *Current Topics in Developmental Biology*, edited by Douglas Forrest and Sophia Tsai, 125:1–38. Nuclear Receptors in Development and Disease. Academic Press. https://doi.org/10.1016/bs.ctdb.2017.02.003.

Hong, Liang, Nitin Jain, Xiaolin Cheng, Ana Bernal, Madhusudan Tyagi, and Jeremy C. Smith. 2016. "Determination of Functional Collective Motions in a Protein at Atomic Resolution Using Coherent Neutron Scattering." *Science Advances* 2 (10): e1600886. https://doi.org/10.1126/sciadv.1600886.

Hu, Erding, Jae Bum Kim, Pasha Sarraf, and Bruce M. Spiegelman. 1996. "Inhibition of Adipogenesis Through MAP Kinase-Mediated Phosphorylation of PPARγ." *Science* 274 (5295): 2100–2103. https://doi.org/10.1126/science.274.5295.2100.

Hu, Xiao, and Mitchell A. Lazar. 1999. "The CoRNR Motif Controls the Recruitment of Corepressors by Nuclear Hormone Receptors." *Nature* 402 (6757): 93–96. https://doi.org/10.1038/47069.

Hughes, Travis S., Michael J. Chalmers, Scott Novick, Dana S. Kuruvilla, Mi Ra Chang, Theodore M. Kamenecka, Mark Rance, et al. 2012a. “Ligand and Receptor Dynamics Contribute to the Mechanism of Graded PPARγ Agonism.” *Structure* 20 (1): 139–50. https://doi.org/10.1016/j.str.2011.10.018.

———. 2012b. “Ligand and Receptor Dynamics Contribute to the Mechanism of Graded PPARγ Agonism.” *Structure* 20 (1): 139–50. https://doi.org/10.1016/j.str.2011.10.018.

Huppunen, Johanna, Gerd Wohlfahrt, and Piia Aarnisalo. 2004. “Requirements for Transcriptional Regulation by the Orphan Nuclear Receptor ERRgamma.” *Molecular and Cellular Endocrinology* 219 (1): 151–60. https://doi.org/10.1016/j.mce.2004.01.002.

Issemann, Isabelle, and Stephen Green. 1990. “Activation of a Member of the Steroid Hormone Receptor Superfamily by Peroxisome Proliferators.” *Nature* 347 (6294): 645–50. https://doi.org/10.1038/347645a0.

Janani, C., and B. D. Ranjitha Kumari. 2015. “PPAR Gamma Gene – A Review.” *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* 9 (1): 46–50. https://doi.org/10.1016/j.dsx.2014.09.015.

Jiang, M., and N. Yang. 2014. “Peroxisome Proliferators.” In *Encyclopedia of Toxicology (Third Edition)*, edited by Philip Wexler, 815–19. Oxford: Academic Press. https://doi.org/10.1016/B978-0-12-386454-3.00346-8.

Jiang, Wei, David J. Hardy, James C. Phillips, Alexander D. MacKerell, Klaus Schulten, and Benoît Roux. 2011. “High-Performance Scalable Molecular Dynamics Simulations of a Polarizable Force Field Based on Classical Drude Oscillators in NAMD.” *The Journal of Physical Chemistry Letters* 2 (2): 87–92. https://doi.org/10.1021/jz101461d.

Jo, Sunhwan, Taehoon Kim, Vidyashankara G. Iyer, and Wonpil Im. 2008. “CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM.” *Journal of Computational Chemistry* 29 (11): 1859–65. https://doi.org/10.1002/jcc.20945.

Johnson, B A, E M Wilson, Y Li, D E Moller, R G Smith, and G Zhou. 2000. “Ligand-Induced Stabilization of PPARgamma Monitored by NMR Spectroscopy: Implications for Nuclear Receptor Activation.” *Journal of Molecular Biology* 298 (2): 187–94. https://doi.org/10.1006/jmbi.2000.3636.

Johnson, Bruce A., Eileen M. Wilson, Ying Li, David E. Moller, Roy G. Smith, and Gaochao Zhou. 2000. “Ligand-Induced Stabilization of PPARgamma Monitored by NMR Spectroscopy: Implications for Nuclear Receptor Activation.” *Journal of Molecular Biology* 298 (2): 187–94. https://doi.org/10.1006/JMBI.2000.3636.

Jorgensen, William L. 1981. “Quantum and Statistical Mechanical Studies of Liquids. 10. Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers. Application to Liquid Water.” *Journal of the American Chemical Society* 103 (2): 335–40. https://doi.org/10.1021/ja00392a016.

Kallenberger, Bettina C., James D. Love, V. Krishna K Chatterjee, and John W R Schwabe. 2003. “A Dynamic Mechanism of Nuclear Receptor Activation and Its Perturbation in a Human Disease.” *Nature Structural Biology* 10 (2): 136–40. https://doi.org/10.1038/nsb892.

Khan, Sabab Hasan, Sean M Braet, Stephen John Koehler, Elizabeth Elacqua, Ganesh Srinivasan Anand, and C Denise Okafor. 2022. “Ligand-Induced Shifts in Conformational Ensembles That Describe Transcriptional Activation.” Edited by Donald Hamelberg, José D Faraldo-Gómez, and Argyris Politis. *eLife* 11 (October):e80140. https://doi.org/10.7554/eLife.80140.

Khoury, Youssef El, and Petra Hellwig. 2017. “Far Infrared Spectroscopy of Hydrogen Bonding Collective Motions in Complex Molecular Systems.” *Chemical Communications* 53 (60): 8389–99. https://doi.org/10.1039/C7CC03496B.

Kishimoto, Masahiko, Ryoji Fujiki, Shinichiro Takezawa, Yasumasa Sasaki, Takashi Nakamura, Kazuyoshi Yamaoka, Hirochika Kitagawa, and Shigeaki Kato. 2006. “Nuclear Receptor Mediated Gene Regulation through Chromatin Remodeling and Histone Modifications.” *Endocrine Journal* 53 (2): 157–72. https://doi.org/10.1507/endocrj.53.157.

Kitao, Akio, Steven Hayward, and Nobuhiro Go. 1998. “Energy Landscape of a Native Protein: Jumping-among-Minima Model.” *Proteins: Structure, Function, and Bioinformatics* 33 (4): 496–517. https://doi.org/10.1002/(SICI)1097-0134(19981201)33:4<496::AID-PROT4>3.0.CO;2-1.

Kognole, Abhishek A., Jumin Lee, Sang-Jun Park, Sunhwan Jo, Payal Chatterjee, Justin A. Lemkul, Jing Huang, Alexander D. MacKerell Jr., and Wonpil Im. 2022. “CHARMM-GUI Drude Prepper for Molecular Dynamics Simulation Using the Classical Drude Polarizable Force Field.” *Journal of Computational Chemistry* 43 (5): 359–75. https://doi.org/10.1002/jcc.26795.

Kroker, Alice J., and John B. Bruning. 2015. “Review of the Structural and Dynamic Mechanisms of PPARγ Partial Agonism.” *PPAR Research* 2015 (1): 816856. https://doi.org/10.1155/2015/816856.

Kumar, R., and E. B. Thompson. 2012. “Folding of the Glucocorticoid Receptor N-Terminal Transactivation Function: Dynamics and Regulation.” *Molecular and Cellular Endocrinology*, Nuclear Receptor Structure: Dynamics and Function, 348 (2): 450–56. https://doi.org/10.1016/j.mce.2011.03.024.

Kumar, Raj, Betty H Johnson, and E Brad Thompson. 2004a. “Overview of the Structural Basis for Transcription Regulation by Nuclear Hormone Receptors.” Edited by Iain J. McEwan. *Essays in Biochemistry* 40 (June):27–39. https://doi.org/10.1042/bse0400027.

———. 2004b. “Overview of the Structural Basis for Transcription Regulation by Nuclear Hormone Receptors.” Edited by Iain J. McEwan. *Essays in Biochemistry* 40 (June):27–39. https://doi.org/10.1042/bse0400027.

Kurkcuoglu, Zeynep, Ahmet Bakan, Duygu Kocaman, Ivet Bahar, and Pemra Doruker. 2012. "Coupling between Catalytic Loop Motions and Enzyme Global Dynamics." *PLoS Computational Biology* 8 (9): e1002705. https://doi.org/10.1371/journal.pcbi.1002705.

Lamas Bervejillo, María, and Ana María Ferreira. 2019. "Understanding Peroxisome Proliferator-Activated Receptors: From the Structure to the Regulatory Actions on Metabolism." In *Bioactive Lipids in Health and Disease*, edited by Andres Trostchansky and Homero Rubbo, 1127:39–57. Advances in Experimental Medicine and Biology. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-11488-6_3.

Lamoureux, Guillaume, Edward Harder, Igor V. Vorobyov, Benoît Roux, and Alexander D. MacKerell. 2006. "A Polarizable Model of Water for Molecular Dynamics Simulations of Biomolecules." *Chemical Physics Letters* 418 (1): 245–49. https://doi.org/10.1016/j.cplett.2005.10.135.

Lamoureux, Guillaume, and Benoît Roux. 2003. "Modeling Induced Polarization with Classical Drude Oscillators: Theory and Molecular Dynamics Simulation Algorithm." *The Journal of Chemical Physics* 119 (6): 3025–39. https://doi.org/10.1063/1.1589749.

Lange, Oliver F., Nils-Alexander Lakomek, Christophe Farès, Gunnar F. Schröder, Korvin F. A. Walter, Stefan Becker, Jens Meiler, Helmut Grubmüller, Christian Griesinger, and Bert L. de Groot. 2008. "Recognition Dynamics up to Microseconds Revealed from an RDC-Derived Ubiquitin Ensemble in Solution." *Science (New York, N.Y.)* 320 (5882): 1471–75. https://doi.org/10.1126/science.1157092.

Laudet, Vincent, and Hinrich Gronemeyer. 2001. *The Nuclear Receptor FactsBook*. Academic Press.

Lazar, Mitchell A. 2005. "PPARγ, 10 Years Later." *Biochimie* 87 (1): 9–13. https://doi.org/10.1016/j.biochi.2004.10.021.

Leach, Andrew R. 2001a. "Molecular Modelling : Principles and Applications." *Computers* 21 (3): 784. https://doi.org/10.1016/S0097-8485(96)00029-0.

Leach, Andrew R. 2001b. *Molecular Modelling: Principles and Applications*. Pearson Education.

Lemkul, Justin A. 2020. "Pairwise-Additive and Polarizable Atomistic Force Fields for Molecular Dynamics Simulations of Proteins." *Progress in Molecular Biology and Translational Science* 170:1–71. https://doi.org/10.1016/bs.pmbts.2019.12.009.

Lemkul, Justin A., Jing Huang, Benoît Roux, and Jr Alexander D. MacKerell. 2016. "An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications." Review-article. ACS Publications. American Chemical Society. World. January 27, 2016. https://doi.org/10.1021/acs.chemrev.5b00505.

Lemkul, Justin A., Jing Huang, Benoît Roux, and Alexander D. MacKerell. 2016. "An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model:

Development History and Recent Applications." *Chemical Reviews* 116 (9): 4983–5013. https://doi.org/10.1021/acs.chemrev.5b00505.

Lesgidou, Nastazia, and Metaxia Vlassi. 2024. "Community Analysis of Large-Scale Molecular Dynamics Simulations Elucidated Dynamics-Driven Allostery in Tyrosine Kinase 2." *Proteins: Structure, Function, and Bioinformatics* 92 (4): 474–98. https://doi.org/10.1002/prot.26631.

Levy, Ronald M., and Martin Karplus. 1979. "Vibrational Approach to the Dynamics of an α-Helix." *Biopolymers* 18 (10): 2465–95. https://doi.org/10.1002/bip.1979.360181008.

Lewandowski, Józef R., Hans Jürgen Sass, Stephan Grzesiek, Martin Blackledge, and Lyndon Emsley. 2011. "Site-Specific Measurement of Slow Motions in Proteins." *Journal of the American Chemical Society* 133 (42): 16762–65. https://doi.org/10.1021/ja206815h.

Lonard, David M., and Bert W. O'Malley. 2007. "Nuclear Receptor Coregulators: Judges, Juries, and Executioners of Cellular Regulation." *Molecular Cell* 27 (5): 691–700. https://doi.org/10.1016/j.molcel.2007.08.012.

Lopes, Pedro E. M., Jing Huang, Jihyun Shim, Yun Luo, Hui Li, Benoît Roux, and Alexander D. Mackerell. 2013. "Force Field for Peptides and Proteins Based on the Classical Drude Oscillator." *Journal of Chemical Theory and Computation* 9 (12): 5430–49. https://doi.org/10.1021/ct400781b.

Lopes, Pedro E. M., Jing Huang, Jihyun Shim, Yun Luo, Hui Li, Benoît Roux, and Alexander D. Jr. MacKerell. 2013. "Polarizable Force Field for Peptides and Proteins Based on the Classical Drude Oscillator." *Journal of Chemical Theory and Computation* 9 (12): 5430–49. https://doi.org/10.1021/ct400781b.

Lopes, Pedro E. M., Benoit Roux, and Alexander D. MacKerell. 2009. "Molecular Modeling and Dynamics Studies with Explicit Inclusion of Electronic Polarizability: Theory and Applications." *Theoretical Chemistry Accounts* 124 (1): 11–28. https://doi.org/10.1007/s00214-009-0617-x.

MacKerell, A. D. Jr., D. Bashford, M. Bellott, R. L. Jr. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, et al. 1998. "All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins." *The Journal of Physical Chemistry B* 102 (18): 3586–3616. https://doi.org/10.1021/jp973084f.

MacKerell, A. D., D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, et al. 1998. "All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins." *The Journal of Physical Chemistry. B* 102 (18): 3586–3616. https://doi.org/10.1021/jp973084f.

MacKerell, A D, Jr, D Bashford, M Bellott, R L Dunbrack, J D Evanseck, M J Field, et al. 1998. "All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins." *The Journal of Physical Chemistry B*, 3586–3616.

Madan, Lalima K., Colin L. Welsh, Alexandr P. Kornev, and Susan S. Taylor. 2023. "The 'Violin Model': Looking at Community Networks for Dynamic Allostery." *The Journal of Chemical Physics* 158 (8): 081001. https://doi.org/10.1063/5.0138175.

Mahajan, Swapnil, and Yves-Henri Sanejouand. 2015. “On the Relationship between Low-Frequency Normal Modes and the Large-Scale Conformational Changes of Proteins.” *Archives of Biochemistry and Biophysics* 567 (February):59–65. https://doi.org/10.1016/j.abb.2014.12.020.

Maire, Albane le, Catherine Teyssier, Cathie Erb, Marina Grimaldi, Susana Alvarez, Angel R. de Lera, Patrick Balaguer, et al. 2010. “A Unique Secondary-Structure Switch Controls Constitutive Gene Repression by Retinoic Acid Receptor.” *Nature Structural & Molecular Biology* 17 (7): 801–7. https://doi.org/10.1038/nsmb.1855.

Marion-Letellier, Rachel, Guillaume Savoye, and Subrata Ghosh. 2016a. “Fatty Acids, Eicosanoids and PPAR Gamma.” *European Journal of Pharmacology*, Immunopharmacology of fatty acids, 785 (August):44–49. https://doi.org/10.1016/j.ejphar.2015.11.004.

———. 2016b. “Fatty Acids, Eicosanoids and PPAR Gamma.” *European Journal of Pharmacology*, Immunopharmacology of fatty acids, 785 (August):44–49. https://doi.org/10.1016/j.ejphar.2015.11.004.

Maston, Glenn, Sara Evans, and Michael Green. 2006. “Transcriptional Regulatory Elements in the Human Genome.” *Annual Review of Genomics and Human Genetics* 7 (February):29–59. https://doi.org/10.1146/annurev.genom.7.080505.115623.

McCammon, Andrew, J., and Harvey Stephen C. 1987. *Dynamics of Proteins and Nucleic Acids*. Cambridge Uuniversity Press.

McKenna, Neil J., Rainer B. Lanz, and Bert W. O’Malley. 1999. “Nuclear Receptor Coregulators: Cellular and Molecular Biology*.” *Endocrine Reviews* 20 (3): 321–44. https://doi.org/10.1210/edrv.20.3.0366.

Meijer, Femke A., Iris A. Leijten-van de Gevel, Rens M. J. M. de Vries, and Luc Brunsveld. 2019. “Allosteric Small Molecule Modulators of Nuclear Receptors.” *Molecular and Cellular Endocrinology* 485 (April):20–34. https://doi.org/10.1016/j.mce.2019.01.022.

Meireles, Lidio, Mert Gur, Ahmet Bakan, and Ivet Bahar. 2011. “Pre-Existing Soft Modes of Motion Uniquely Defined by Native Contact Topology Facilitate Ligand Binding to Proteins.” *Protein Science* 20 (10): 1645–58. https://doi.org/10.1002/pro.711.

Michael P. Allen. 2004. *Computational Soft Matter: From Synthetic Polymers to Proteins; Winter School, 29 February - 6 March 2004, Gustav-Stresemann-Institut, Bonn, Germany. 2: Lecture Notes*. Edited by Gustav-Stresemann-Institut e.V. für Übernationale Bildung und Europäische Zusammenarbeit, John von Neumann-Institut für Computing, Johannes Gutenberg-Universität Mainz, and Max-Planck-Institut für Biophysikalische Chemie. NIC Series 23. Jülich: NIC.

Miglioli, Angelica, Laura Canesi, Isa D. L. Gomes, Michael Schubert, and Rémi Dumollard. 2021. “Nuclear Receptors and Development of Marine Invertebrates.” *Genes* 12 (1): 83. https://doi.org/10.3390/genes12010083.

Millard, Christopher J., Peter J. Watson, Louise Fairall, and John W. R. Schwabe. 2013. “An Evolving Understanding of Nuclear Receptor Coregulator Proteins,” December. https://doi.org/10.1530/JME-13-0227.

Mitsis, Thanasis, Louis Papageorgiou, Aspasia Efthimiadou, Flora Bacopoulou, Dimitrios Vlachakis, George P. Chrousos, and Elias Eliopoulos. 2019. “A Comprehensive Structural and Functional Analysis of the Ligand Binding Domain of the Nuclear Receptor Superfamily Reveals Highly Conserved Signaling Motifs and Two Distinct Canonical Forms through Evolution.” *World Academy of Sciences Journal* 1 (6): 264–74. https://doi.org/10.3892/wasj.2020.30.

Mlynek, Georg, Kristina Djinović-Carugo, and Oliviero Carugo. 2024. “B-Factor Rescaling for Protein Crystal Structure Analyses.” *Crystals* 14 (5): 443. https://doi.org/10.3390/cryst14050443.

Moras, Dino, and Hinrich Gronemeyer. 1998. “The Nuclear Receptor Ligand-Binding Domain: Structure and Function.” *Current Opinion in Cell Biology* 10 (3): 384–91. https://doi.org/10.1016/S0955-0674(98)80015-X.

Mosure, Sarah A., Paola Munoz-Tello, Kuang-Ting Kuo, Brian MacTavish, Xiaoyu Yu, Daniel Scholl, Christopher C. Williams, et al. 2024. “Structural Basis of Interdomain Communication in PPARγ.” bioRxiv. https://doi.org/10.1101/2022.07.13.499031.

Na, Hyuntae, Konrad Hinsen, and Guang Song. 2021. “The Amounts of Thermal Vibrations and Static Disorder in Protein X-Ray Crystallographic B-Factors.” *Proteins* 89 (11): 1442–57. https://doi.org/10.1002/prot.26165.

Nader, Nancy, Sinnie Sin Man Ng, George I. Lambrou, Panagiota Pervanidou, Yonghong Wang, George P. Chrousos, and Tomoshige Kino. 2010. “AMPK Regulates Metabolic Actions of Glucocorticoids by Phosphorylating the Glucocorticoid Receptor through P38 MAPK.” *Molecular Endocrinology* 24 (9): 1748–64. https://doi.org/10.1210/me.2010-0192.

Nagy, L. 2004. “Mechanism of the Nuclear Receptor Molecular Switch.” *Trends in Biochemical Sciences* 29 (6): 317–24. https://doi.org/10.1016/j.tibs.2004.04.006.

Nielsen, Ronni, Thomas Åskov Pedersen, Dik Hagenbeek, Panagiotis Moulos, Rasmus Siersbæk, Eva Megens, Sergei Denissov, et al. 2008. “Genome-Wide Profiling of PPARγ:RXR and RNA Polymerase II Occupancy Reveals Temporal Activation of Distinct Metabolic Pathways and Changes in RXR Dimer Composition during Adipogenesis.” *Genes & Development* 22 (21): 2953–67. https://doi.org/10.1101/gad.501108.

Novac, Natalia, and Thorsten Heinzel. 2004. “Nuclear Receptors: Overview and Classification.” *Current Drug Target -Inflammation & Allergy* 3 (4): 335–46. https://doi.org/10.2174/1568010042634541.

Olsson, Mats H. M., Chresten R. Søndergaard, Michal Rostkowski, and Jan H. Jensen. 2011. “PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions.” *Journal of Chemical Theory and Computation* 7 (2): 525–37. https://doi.org/10.1021/ct100578z.

Orellana, Laura. 2023. “Are Protein Shape-Encoded Lowest-Frequency Motions a Key Phenotype Selected by Evolution?” *Applied Sciences* 13 (11): 6756. https://doi.org/10.3390/app13116756.

Orlov, Igor, Natacha Rochel, Dino Moras, and Bruno P Klaholz. 2012. “Structure of the Full Human RXR/VDR Nuclear Receptor Heterodimer Complex with Its DR3 Target DNA.” *The EMBO Journal* 31 (2): 291–300. https://doi.org/10.1038/emboj.2011.445.

Owen, G. I., and A. Zelent. 2000. “Origins and Evolutionary Diversification of the Nuclear Receptor Superfamily.” *Cellular and Molecular Life Sciences CMLS* 57 (5): 809–27. https://doi.org/10.1007/s000180050043.

Patel, Abdul Kareem Mohideen, Pierre Vilela, Tajith Baba Shaik, Alastair G McEwen, Isabelle Hazemann, Karl Brillet, Eric Ennifar, et al. 2023. “Asymmetric Dimerization in a Transcription Factor Superfamily Is Promoted by Allosteric Interactions with DNA.” *Nucleic Acids Research* 51 (16): 8864–79. https://doi.org/10.1093/nar/gkad632.

Patel, Sandeep, Alexander D. Mackerell Jr., and Charles L. Brooks III. 2004. “CHARMM fluctuating charge force field for proteins: II Protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model.” *Journal of Computational Chemistry* 25 (12): 1504–14. https://doi.org/10.1002/jcc.20077.

Pawlak, Michal, Philippe Lefebvre, and Bart Staels. 2012. “General Molecular Biology and Architecture of Nuclear Receptors.” *Current Topics in Medicinal Chemistry* 12 (6): 486. https://doi.org/10.2174/156802612799436641.

Phillips, James C., David J. Hardy, Julio D. C. Maia, John E. Stone, João V. Ribeiro, Rafael C. Bernardi, Ronak Buch, et al. 2020a. “Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD.” *The Journal of Chemical Physics* 153 (4): 044130. https://doi.org/10.1063/5.0014475.

———. 2020b. “Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD.” *The Journal of Chemical Physics* 153 (4): 044130. https://doi.org/10.1063/5.0014475.

———. 2020c. “Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD.” *The Journal of Chemical Physics* 153 (4): 044130. https://doi.org/10.1063/5.0014475.

Pohl, Ehmke, and Charles W. E. Tomlinson. 2020. “Chapter Seven - Classical Pathways of Gene Regulation by Retinoids.” In *Methods in Enzymology*, edited by Ehmke Pohl, 637:151–73. Retinoid Signaling Pathways. Academic Press. https://doi.org/10.1016/bs.mie.2020.03.008.

Ponzoni, Luca, and Ivet Bahar. 2018. “Structural Dynamics Is a Determinant of the Functional Significance of Missense Variants.” *Proceedings of the National Academy of Sciences* 115 (16): 4164–69. https://doi.org/10.1073/pnas.1715896115.

Rapaport, D. C. 2004. *The Art of Molecular Dynamics Simulation*. Cambridge University Press.

Rastinejad, Fraydoon, Pengxiang Huang, Vikas Chandra, and Sepideh Khorasanizadeh. 2013. "Understanding Nuclear Receptor Form and Function Using Structural Biology," December. https://doi.org/10.1530/JME-13-0173.

Ray, David W. 2022. "Circadian Rhythm and Nuclear Receptors." *Advances in Experimental Medicine and Biology* 1390:143–53. https://doi.org/10.1007/978-3-031-11836-4_8.

Renaud, J. P., and D. Moras. 2000a. "Structural Studies on Nuclear Receptors." *Cellular and Molecular Life Sciences CMLS* 57 (12): 1748–69. https://doi.org/10.1007/PL00000656.

Renaud, J P, and D Moras. 2000b. "Structural Studies on Nuclear Receptors." *Cellular and Molecular Life Sciences* 57 (12): 1748–69. https://doi.org/10.1007/PL00000656.

———. 2000c. "Structural Studies on Nuclear Receptors." *Cellular and Molecular Life Sciences* 57 (12): 1748–69. https://doi.org/10.1007/PL00000656.

Renaud, Jean-Paul, Natacha Rochel, Marc Ruff, Valéria Vivat, Pierre Chambon, Hinrich Gronemeyer, and Dino Moras. 1995. "Crystal Structure of the RAR-γ Ligand-Binding Domain Bound to All-Trans Retinoic Acid." *Nature* 378 (6558): 681–89. https://doi.org/10.1038/378681a0.

Ricci, Clarisse G., Rodrigo L. Silveira, Ivan Rivalta, Victor S. Batista, and Munir S. Skaf. 2016. "Allosteric Pathways in the PPARγ-RXRα Nuclear Receptor Complex." *Scientific Reports* 6 (1): 19940. https://doi.org/10.1038/srep19940.

Rischel, Christian, Diane Spiedel, Justin P. Ridge, Michael R. Jones, Jacques Breton, Jean-Christophe Lambry, Jean-Louis Martin, and Marten H. Vos. 1998. "Low Frequency Vibrational Modes in Proteins: Changes Induced by Point-Mutations in the Protein-Cofactor Matrix of Bacterial Reaction Centers." *Proceedings of the National Academy of Sciences* 95 (21): 12306–11. https://doi.org/10.1073/pnas.95.21.12306.

Rochel, N, C Krucker, L Coutos-Thevenot, J Osz, R Zhang, E Guyon, W Zita, et al. 2019. "Recurrent Activating Mutations of PPARgamma Associated with Luminal Bladder Tumors." *Nat Commun* 10 (1): 253. https://doi.org/10.1038/s41467-018-08157-y.

Rochel, Natacha, Clémentine Krucker, Laure Coutos-Thévenot, Judit Osz, Ruiyun Zhang, Elodie Guyon, Wayne Zita, et al. 2019. "Recurrent Activating Mutations of PPARγ Associated with Luminal Bladder Tumors." *Nature Communications* 10 (1): 253. https://doi.org/10.1038/s41467-018-08157-y.

Rosenfeld, Michael G., Victoria V. Lunyak, and Christopher K. Glass. 2006. "Sensors and Signals: A Coactivator/Corepressor/Epigenetic Code for Integrating Signal-Dependent Programs of Transcriptional Response." *Genes & Development* 20 (11): 1405–28. https://doi.org/10.1101/gad.1424806.

Safer, Joshua D., Ronald N. Cohen, Anthony N. Hollenberg, and Fredric E. Wondisford. 1998. "Defective Release of Corepressor by Hinge Mutants of the Thyroid Hormone Receptor Found in Patients with Resistance to Thyroid Hormone *." *Journal of Biological Chemistry* 273 (46): 30175–82. https://doi.org/10.1074/jbc.273.46.30175.

Salvi, Nicola, Anton Abyzov, and Martin Blackledge. 2017. “Atomic Resolution Conformational Dynamics of Intrinsically Disordered Proteins from NMR Spin Relaxation.” *Progress in Nuclear Magnetic Resonance Spectroscopy* 102–103 (November):43–60. https://doi.org/10.1016/j.pnmrs.2017.06.001.

Samarut, Eric, Ismail Amal, Gabriel V. Markov, Roland Stote, Annick Dejaegere, Vincent Laudet, and Cécile Rochette-Egly. 2011a. “Evolution of Nuclear Retinoic Acid Receptor Alpha (RARα) Phosphorylation Sites. Serine Gain Provides Fine-Tuned Regulation.” *Molecular Biology and Evolution* 28 (7): 2125–37. https://doi.org/10.1093/molbev/msr035.

———. 2011b. “Evolution of Nuclear Retinoic Acid Receptor Alpha (RARα) Phosphorylation Sites. Serine Gain Provides Fine-Tuned Regulation.” *Molecular Biology and Evolution* 28 (7): 2125–37. https://doi.org/10.1093/molbev/msr035.

Santos, Guilherme M., Louise Fairall, and John W.R. Schwabe. 2011. “Negative Regulation by Nuclear Receptors: A Plethora of Mechanisms.” *Trends in Endocrinology & Metabolism* 22 (3): 87–93. https://doi.org/10.1016/j.tem.2010.11.004.

Sarruf, David A., Irena Iankova, Anna Abella, Said Assou, Stéphanie Miard, and Lluis Fajas. 2005. “Cyclin D3 Promotes Adipogenesis through Activation of Peroxisome Proliferator-Activated Receptor γ.” *Molecular and Cellular Biology* 25 (22): 9985–95. https://doi.org/10.1128/MCB.25.22.9985-9995.2005.

Schoonjans, K, B Staels, and J Auwerx. 1996. “Role of the Peroxisome Proliferator-Activated Receptor (PPAR) in Mediating the Effects of Fibrates and Fatty Acids on Gene Expression.” *Journal of Lipid Research* 37 (5): 907–25. https://doi.org/10.1016/S0022-2275(20)42003-6.

Seo, Moon-Hyeong, Jeongbin Park, Eunkyung Kim, Sungchul Hohng, and Hak-Sung Kim. 2014. “Protein Conformational Dynamics Dictate the Binding Affinity for a Ligand.” *Nature Communications* 5 (1): 3724. https://doi.org/10.1038/ncomms4724.

Shamilov, Rambon, and Brian J. Aneskievich. 2019. “Intrinsic Disorder in Nuclear Receptor Amino Termini: From Investigational Challenge to Therapeutic Opportunity.” *Nuclear Receptor Research* 6. https://doi.org/10.32527/2019/101417.

Shi, Yue, Zhen Xia, Jiajing Zhang, Robert Best, Chuanjie Wu, Jay W. Ponder, and Pengyu Ren. 2013. “The Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins.” *Journal of Chemical Theory and Computation* 9 (9): 4046–63. https://doi.org/10.1021/ct4003702.

Shukla, Rohit, Harish Shukla, and Timir Tripathi. 2018. “Activity Loss by H46A Mutation in *Mycobacterium Tuberculosis* Isocitrate Lyase Is Due to Decrease in Structural Plasticity and Collective Motions of the Active Site.” *Tuberculosis* 108 (January):143–50. https://doi.org/10.1016/j.tube.2017.11.013.

Siersbæk, Majken S., Anne Loft, Mads M. Aagaard, Ronni Nielsen, Søren F. Schmidt, Natasa Petrovic, Jan Nedergaard, and Susanne Mandrup. 2012. “Genome-Wide Profiling of Peroxisome Proliferator-Activated Receptor γ in Primary Epididymal, Inguinal, and Brown Adipocytes Reveals Depot-Selective Binding Correlated with Gene

Expression." *Molecular and Cellular Biology* 32 (17): 3452–63. https://doi.org/10.1128/MCB.00526-12.

Skjaerven, Lars, Siv M. Hollup, and Nathalie Reuter. 2009. "Normal Mode Analysis for Proteins." *Journal of Molecular Structure: THEOCHEM*, Theoretical treatment of large molecular systems, 898 (1): 42–48. https://doi.org/10.1016/j.theochem.2008.09.024.

Smith, Colin A., David Ban, Supriya Pratihar, Karin Giller, Maria Paulat, Stefan Becker, Christian Griesinger, Donghan Lee, and Bert L. de Groot. 2016. "Allosteric Switch Regulates Protein–Protein Binding through Collective Motion." *Proceedings of the National Academy of Sciences* 113 (12): 3269–74. https://doi.org/10.1073/pnas.1519609113.

Søndergaard, Chresten R., Mats H. M. Olsson, Michał Rostkowski, and Jan H. Jensen. 2011. "Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values." *Journal of Chemical Theory and Computation* 7 (7): 2284–95. https://doi.org/10.1021/ct200133y.

Subramaniam, Nanthakumar, William Cairns, and Sam Okret. 1998. "Glucocorticoids Repress Transcription from a Negative Glucocorticoid Response Element Recognized by Two Homeodomain-Containing Proteins, Pbx and Oct-1*." *Journal of Biological Chemistry* 273 (36): 23567–74. https://doi.org/10.1074/jbc.273.36.23567.

Swope, William C., Hans C. Andersen, Peter H. Berens, and Kent R. Wilson. 1982. "A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters." *The Journal of Chemical Physics* 76 (1): 637–49. https://doi.org/10.1063/1.442716.

Tagami, Tetsuya, Youngkyu Park, and J. Larry Jameson. 1999. "Mechanisms That Mediate Negative Regulation of the Thyroid-Stimulating Hormone α Gene by the Thyroid Hormone Receptor*." *Journal of Biological Chemistry* 274 (32): 22345–53. https://doi.org/10.1074/jbc.274.32.22345.

Tao, Lian Jing, Dong Eun Seo, Benjamin Jackson, Natalia B. Ivanova, and Fabio Rinaldo Santori. 2020. "Nuclear Hormone Receptors and Their Ligands: Metabolites in Control of Transcription." *Cells* 9 (12): 2606. https://doi.org/10.3390/cells9122606.

Tasumi, M., H. Takeuchi, S. Ataka, A. M. Dwivedi, and S. Krimm. 1982. "Normal Vibrations of Proteins: Glucagon." *Biopolymers* 21 (3): 711–14. https://doi.org/10.1002/bip.360210318.

Tee, Wei-Ven, and Igor N. Berezovsky. 2024. "Allosteric Drugs: New Principles and Design Approaches." *Current Opinion in Structural Biology* 84 (February):102758. https://doi.org/10.1016/j.sbi.2023.102758.

Thole, B. T. 1981. "Molecular Polarizabilities Calculated with a Modified Dipole Interaction." *Chemical Physics* 59 (3): 341–50. https://doi.org/10.1016/0301-0104(81)85176-2.

Tompa, Peter. 2016. "The Principle of Conformational Signaling." *Chemical Society Reviews* 45 (15): 4252–84. https://doi.org/10.1039/C6CS00011H.

Torchia, Joseph, Christopher Glass, and Michael G Rosenfeld. 1998a. “Co-Activators and Co-Repressors in the Integration of Transcriptional Responses.” *Current Opinion in Cell Biology* 10 (3): 373–83. https://doi.org/10.1016/S0955-0674(98)80014-8.

———. 1998b. “Co-Activators and Co-Repressors in the Integration of Transcriptional Responses.” *Current Opinion in Cell Biology* 10 (3): 373–83. https://doi.org/10.1016/S0955-0674(98)80014-8.

Umemoto, Tomoe, and Yukio Fujiki. 2012. “Ligand-Dependent Nucleo-Cytoplasmic Shuttling of Peroxisome Proliferator-Activated Receptors, PPARα and PPARγ.” *Genes to Cells* 17 (7): 576–96. https://doi.org/10.1111/j.1365-2443.2012.01607.x.

UniProt. 2024. “PPARG - Peroxisome Proliferator-Activated Receptor Gamma - Homo Sapiens.” November 2024. https://www.uniprot.org/uniprotkb/P37231/entry.

Vanommeslaeghe, K., and A. D. MacKerell. 2015a. “CHARMM Additive and Polarizable Force Fields for Biophysics and Computer-Aided Drug Design.” *Biochimica et Biophysica Acta* 1850 (5): 861–71. https://doi.org/10.1016/j.bbagen.2014.08.004.

———. 2015b. “CHARMM Additive and Polarizable Force Fields for Biophysics and Computer-Aided Drug Design.” *Biochimica et Biophysica Acta* 1850 (5): 861–71. https://doi.org/10.1016/j.bbagen.2014.08.004.

Vanommeslaeghe, K., and A. D. Jr. MacKerell. 2012. “Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing.” *Journal of Chemical Information and Modeling* 52 (12): 3144–54. https://doi.org/10.1021/ci300363c.

Vera, Ian Mitchelle S. de, Jie Zheng, Scott Novick, Jinsai Shang, Travis S. Hughes, Richard Brust, Paola Munoz-Tello, et al. 2017. “Synergistic Regulation of Coregulator/Nuclear Receptor Interaction by Ligand and DNA.” *Structure* 25 (10): 1506-1518.e4. https://doi.org/10.1016/j.str.2017.07.019.

Vlijmen, Herman W. T. van, and Martin Karplus. 1999. “Analysis of Calculated Normal Modes of a Set of Native and Partially Unfolded Proteins.” *The Journal of Physical Chemistry B* 103 (15): 3009–21. https://doi.org/10.1021/jp983452v.

Walinda, Erik, Daichi Morimoto, and Kenji Sugase. 2018. “Overview of Relaxation Dispersion NMR Spectroscopy to Study Protein Dynamics and Protein-Ligand Interactions.” *Current Protocols in Protein Science* 92 (1): e57. https://doi.org/10.1002/cpps.57.

Wanders, Ronald J. A., Myriam Baes, Daniela Ribeiro, Sacha Ferdinandusse, and Hans R. Waterham. 2023. “The Physiological Functions of Human Peroxisomes.” *Physiological Reviews* 103 (1): 957–1024. https://doi.org/10.1152/physrev.00051.2021.

Wärnmark, Anette, Eckardt Treuter, Anthony P. H. Wright, and Jan-Åke Gustafsson. 2003. “Activation Functions 1 and 2 of Nuclear Receptors: Molecular Strategies for Transcriptional Activation.” *Molecular Endocrinology* 17 (10): 1901–9. https://doi.org/10.1210/me.2002-0384.

Weikum, Emily R., Xu Liu, and Eric A. Ortlund. 2018a. “The Nuclear Receptor Superfamily: A Structural Perspective.” *Protein Science* 27 (11): 1876–92. https://doi.org/10.1002/pro.3496.

———. 2018b. “The Nuclear Receptor Superfamily: A Structural Perspective.” *Protein Science : A Publication of the Protein Society* 27 (11): 1876. https://doi.org/10.1002/pro.3496.

Wiberg, Kenneth. 1965. “A Scheme for Strain Energy Minimization. Application to the Cycloalkanes.” *Journal of the American Chemical Society* 87 (5).

Williams, J.E., P., and P.R. Schleyer. 1968. “Physical Organic Chemistry: Quantitative Conformational Analysis; Calculation Methods” 19:531–58.

Wingender, Edgar, Torsten Schoeps, and Jürgen Dönitz. 2013. “TFClass: An Expandable Hierarchical Classification of Human Transcription Factors.” *Nucleic Acids Research* 41 (Database issue): D165-170. https://doi.org/10.1093/nar/gks1123.

Wurtz, Jean-Marie, William Bourguet, Jean-Paul Renaud, Valérie Vivat, Pierre Chambon, Dino Moras, and Hinrich Gronemeyer. 1996. “A Canonical Structure for the Ligand-Binding Domain of Nuclear Receptors.” *Nature Structural Biology* 3 (1): 87–94. https://doi.org/10.1038/nsb0196-87.

Xu, H. Eric. 2015. “Family Reunion of Nuclear Hormone Receptors: Structures, Diseases, and Drug Discovery.” *Acta Pharmacologica Sinica* 36 (1): 1–2. https://doi.org/10.1038/aps.2014.140.

York, Darrin M., Tom A. Darden, and Lee G. Pedersen. 1993. “The Effect of Long-range Electrostatic Interactions in Simulations of Macromolecular Crystals: A Comparison of the Ewald and Truncated List Methods.” *The Journal of Chemical Physics* 99 (10): 8345–48. https://doi.org/10.1063/1.465608.

Yoshizawa, Mami, Tomomi Aoyama, Toshimasa Itoh, and Hiroyuki Miyachi. 2022a. “Arylalkynyl Amide-Type Peroxisome Proliferator-Activated Receptor γ (PPARγ)-Selective Antagonists Covalently Bind to the PPARγ Ligand Binding Domain with a Unique Binding Mode.” *Bioorganic & Medicinal Chemistry Letters* 64 (May):128676. https://doi.org/10.1016/j.bmcl.2022.128676.

———. 2022b. “Arylalkynyl Amide-Type Peroxisome Proliferator-Activated Receptor γ (PPARγ)-Selective Antagonists Covalently Bind to the PPARγ Ligand Binding Domain with a Unique Binding Mode.” *Bioorganic & Medicinal Chemistry Letters* 64 (May):128676. https://doi.org/10.1016/j.bmcl.2022.128676.

Yu, Xinzhe, Ping Yi, Ross A. Hamilton, Hong Shen, Muyuan Chen, Charles E. Foulds, Michael A. Mancini, Steven J. Ludtke, Zhao Wang, and Bert W. O’Malley. 2020. “Structural Insights of Transcriptionally Active, Full-Length Androgen Receptor Coactivator Complexes.” *Molecular Cell* 79 (5): 812-823.e4. https://doi.org/10.1016/j.molcel.2020.06.031.

Zhang, Dawn X., and Christopher K. Glass. 2013. “Towards an Understanding of Cell-Specific Functions of Signal-Dependent Transcription Factors,” December. https://doi.org/10.1530/JME-13-0216.

Zhang, Haitao, Lili Chen, Jing Chen, Hualiang Jiang, and Xu Shen. 2011. “Structural Basis for Retinoic X Receptor Repression on the Tetramer.” *The Journal of Biological Chemistry* 286 (28): 24593. https://doi.org/10.1074/jbc.M111.245498.

Zuegg, J., and J. E. Gready. 1999. “Molecular Dynamics Simulations of Human Prion Protein: Importance of Correct Treatment of Electrostatic Interactions.” *Biochemistry* 38 (42): 13862–76. https://doi.org/10.1021/bi991469d.

MILINSKI Ana

Dynamique structurale du récepteur nucléaire PPARgamma

Résumé

PPARγ est un régulateur du métabolisme des lipides et de l'énergie, et les changements dans sa dynamique structurale sont impliqués dans de nombreux processus physiologiques et pathologiques. L'étude au niveau atomique des mécanismes moléculaires qui sous-tendent ces effets dynamiques implique de caractériser les mouvements, et en particulier les mouvements collectifs de PPARγ.

L'objectif de cette thèse était de développer une nouvelle approche pour mesurer les propriétés physiques directement liées aux changements dans la dynamique structurale collective à basse fréquence des protéines. A cette fin, nous avons développé une approche intégrée de spectroscopie dans l'infrarouge lointain (Far-IR) et de simulation de dynamique moléculaire (MD), appelée « Ensemble Averaged Normal Modes ». Nous nous sommes appuyés sur l'analyse des modes normaux (NMA) en mettant en œuvre un échantillonnage MD, ce qui nous a permis d'étudier les fluctuations, les spectres IR et les mouvements corrélés. Ces analyses ont été complétées par les analyses de la connectivité et des évaluations de la structure secondaire et des liaisons hydrogène.

Le chapitre III décrit le protocole développé, qui a été appliqué dans le chapitre IV à PPARγ, en particulier son domaine de liaison au ligand (LBD) sauvage dans les formes apo et holo (lié à l'agoniste GW1929) et portant deux mutations associées au cancer (T4757M et F310S). Le chapitre V décrit une deuxième étude qui portait sur des effets de la polarisation, à travers l'implémentation d'un champ de force polarisable, sur les mouvements collectifs de PPARγ, dans sa forme apo et dans sa forme liée à un peptide corépresseur.

En conclusion, nous présentons une méthodologie originale qui combine des simulations de dynamique moléculaire et de la spectroscopie IR lointain, ainsi qu'une application du protocole développé à l'étude de PPARγ, ainsi qu'une étude de l'impact de la polarisation électronique sur les mouvements collectifs de PPARγ.

Mots clés : PPARgamma, récepteur nucléaire, dynamique moléculaire, mutations, ligand, mouvements corrélés, polarisation

Abstract

PPARγ is a regulator of lipid and energy metabolism and is implicated in many different physiological and pathological processes. It acts as an allosteric hub receiving chemical signals that are then translated into biological responses. This action depends on changes in its conformation and its structural dynamics, the latter being difficult to quantify. An atomic level understanding of the molecular mechanisms of both physiological and pathological activity of PPARγ can be can be understood by studying the underlying structural dynamics, particularly those of collective motions.

We developed an integrated approach to study physical properties directly related to low frequency collective structural dynamics of proteins. It is a combined far infrared (Far-IR) spectroscopy – molecular dynamics (MD) simulation approach that relies on a method we call Ensemble Averaged Normal Modes. This allowed us to characterise protein fluctuations, computed IR spectra, and correlated motions. These analyses were complemented by community network analysis, secondary structure- and hydrogen-bond evaluations.

Developed protocol was applied to PPARγ systems, including its wild-type ligand-binding domain (LBD) in apo and holo forms (bound to agonist ligand GW1929) and two cancer-associated mutants (gain-of-function T4757M and loss-of-function F310S). Chapter V describes a separate study of the effects of polarization, through the implementation of the polarizable Drude force field, on the collective motions of PPARγ in its apo form and when bound to a corepressor peptide.

Keywords: PPARgamma, nuclear receptor, molecular dynamics, sampling, mutations, ligand, collective motions, polarization