



Université de Strasbourg

École Doctorale des Sciences de la Vie et de la Santé

## THÈSE

présentée et soutenue publiquement le 16 septembre 2011

pour l'obtention du

**DOCTORAT DE L'UNIVERSITÉ DE STRASBOURG**

Discipline : Sciences du Vivant

*Spécialité : Aspects Moléculaires et Cellulaires de la Biologie*

# **Évaluation des Protéines Kinases en tant que Cibles Thérapeutiques Application à la Conception Rationnelle d'Inhibiteurs Allostériques Covalents**

Par

Émeline Leproult

**Devant un jury composé de :**

- |                                |   |
|--------------------------------|---|
| Rapporteur externe :           | Dr. Gilles Labesse, CBS, Montpellier              |
| Rapporteur externe :           | Dr. Marc Poirot, CRCT, Toulouse                   |
| Examineur interne :            | Pr. Marcel Hibert, Faculté de Pharmacie, Illkirch |
| Membre invité :                | Pr. Nicolas Winssinger, ISIS, Strasbourg          |
| Membre invité :                | Dr. Olivier Poch, IGBMC, Illkirch                 |
| <b>Directeur de thèse :</b>    | Pr. Jean-Marie Wurtz, IGBMC, Illkirch             |
| <b>Co-directeur de thèse :</b> | Dr. Dino Moras, IGBMC, Illkirch                   |

*Thèse préparée au sein du Département de Biologie Structurale Intégrative  
de l'Institut de Génétique et de Biologie Moléculaire et Cellulaire  
UDS, CNRS UMR7104, INSERM U964, Illkirch*



Je tiens à remercier les membres du jury : le Docteur Gilles Labesse, le Docteur Marc Poirot et le Professeur Marcel Hibert, pour avoir accepté de juger ma thèse et pris le temps de se déplacer (ou de rester) à Strasbourg, malgré leur charge de travail respective.

Je remercie vivement Nicolas Winssinger dont l'expertise dans le domaine des kinases ainsi que des inhibiteurs covalents est à l'origine des travaux développés pendant ma thèse.

Je remercie tout ceux - oui ils sont nombreux - qui ont, à un moment ou un autre, pendant un temps plus ou moins long, d'une façon assidue ou lointaine mais paternelle, encadré ou veillé à mon travail de thèse : Jean-Marie Wurtz, Dino Moras, Olivier Poch, Jean-Paul Renaud, Serge Grisoni, Charles Andrianjara, Philippe Schambel, Jean-Claude Thierry.

Je remercie Dino Moras pour m'avoir accueillie dans son institut, il y a exactement trois ans, et pour avoir accepté la codirection de thèse avant de me connaître.

Je remercie Jean-Marie Wurtz pour avoir accepté de diriger ma thèse, tout en m'accordant la liberté et l'autonomie nécessaires pour développer mes propres idées.

Je remercie Jean Cavarelli pour m'avoir accueillie dans son équipe.

Je remercie Olivier Poch pour son implication vis-à-vis du bon déroulement de ma thèse et pour son enthousiasme communicatif.

Un grand merci à nos collaborateurs chimistes Nicolas Winssinger et Sofia Barluenga, qui m'ont apporté leur confiance et leur intérêt quant à mon travail et mes suggestions.

Je remercie l'équipe de la plateforme Puces à Cellules Transfectées, dirigée par Laurent Brino, pour son implication dans mon travail et nos multiples discussions.

Je serai longtemps redevable envers les personnes qui, chez Pierre Fabre, m'ont donné l'opportunité de faire une thèse, tout en continuant de venir travailler une à deux fois par an à Castres : Charles Andrianjara, Jean-Marc Laurenceau, Philippe Schambel et Serge Grisoni.

Merci à Patrice Gouet pour sa relecture finale, accompagnée de corrections et remarques judicieuses, mais aussi de macarons.

Je lance un clin d'œil à mes compagnons d'infortune et néanmoins collègues, qui m'ont initiée aux vertus apaisantes (pas toujours, finalement) que pouvait avoir une promenade autour du bâtiment, une tasse à la main : Benjamin Schwarz et Yann Lienhart.

Finalement, j'avais beau tenter de m'inciviliser en me cachant derrière mon ordinateur et mes *todo* listes, mais rien n'y a fait. Je vais devoir saluer chaleureusement toute l'équipe d'Olivier Poch. Cela a commencé par une place dans leurs locaux, l'intégration à leurs réunions d'équipe, les discussions de couloirs, le coin café avec ses viennoiseries, les blagues sur les blondes,... et me voilà, trois ans plus tard, à remercier Luc, Laurent-Philippe, Radouhène, Daho, Laëtitia, Véronique, Sophie, Frédéric, Laurent, Yann, Raymond, Julie, Odile, Francesco, Anne, David, Nicolas, Yannick, Benjamin, Hoan, Wolfgang, Nicodème,...

En prenant la fuite au CEBGS, je pensais retrouver une vie monacale mais là encore, on m'a contrainte à faire moult gâteaux et *Bredele* (ou serait-ce l'inverse ?), à manger des sushis, du houmous, de la confiture au lait de chèvre, des insectes (oh non, on a oublié !!!), à m'occuper des plantes des autres (contre bons soins pour mon orchidée), et à jouer au foot ... Et tout ça, je le dois à Marie-Laure (que je ne remercierai jamais assez pour son accueil, son soutien, sa solidarité, ses gâteaux, bref notre amitié), Justine, Stéphanie, Emiko, Denise, Laura, Nada, Claire, Morgane, Pierre, Martino, Maxime/Wassim, Massimo, Alex(s), Quentin, Firas ... C'est donc ça l'auberge espagnole ? Restons dans ce même bâtiment pour que je remercie vivement le reste de mon équipe : Nathalie, Vincent, Martin et Christophe.

Dans une vie, il y a des coups de cœur d'amitié, et Katja, tu en fais partie. C'est fou comme un regard suffit pour que, dès le premier rendez-vous, on s'aperçoive de la foultitude de choses que l'on a en commun... Vous remercier toi et Christian en une série de mots donnerait : balades à vélo et vol de cerises (allemandes !), *Käsewurst*, Berlin, *Deutsche Perfektion*, Wien, *Opa Apflesaft*, barbecues, l'Allemand pour les nuls, *vielen Küchen*, *große Frühstück*, les élastiques à 4 branches... Et puis, je pourrai me vanter de connaître la « World expert in PDZ domains » (et tant d'autres choses).

C'est souvent à trois qu'on fait les meilleurs coups, alors arriva Matthieu, le seul qui se soit retrouvé dans une situation personnelle aussi nulle que la mienne (mais qui s'en sort mieux que moi, la bague au doigt)... On a rapidement su outrepasser cela à coups de *berliner Wiese*, mais surtout lors de nos marathons de la flammekueche. Que j'ai aimé cette franche camaraderie, en particulier les crises de rires aux dépens de notre futur (*perfekt*) filleul...

Cessons de parler du travail, pour que je remercie celui qui m'a fait sillonner le canal de la Marne au Rhin, la forêt d'Illkirch et les abords du Rhin, avec beaucoup de prestance et

souvent plein galop, Kobra (qui vous remercie tous pour les petits pains). Je reste dans le même paragraphe pour saluer mes partenaires de badminton, Olivier (mais si, le munster, c'est bon), Michaël (le seul à comprendre mon enthousiasme pour *La Belle Hélène*), Christophe (arrête de plonger pour le volant, tu sais que je ne vais pas assurer derrière).

Du point de vue des infrastructures, je dresse un bilan positif sur le fonctionnement d'AirFrance et Airlinair : environ 250 avions empruntés, seulement 6 grèves et 4 annulations, 0 crashes et 60 000 *miles* accumulés.

Et puis, il y a Ian, les amis et la famille : « Mais si, ça va le faire ». Merci pour votre soutien, qui ne date pas d'hier. J'aurai bientôt le temps de vous remercier plus longuement et plus personnellement.

À présent, « En route vers de nouvelles aventures ! » ...



*Ne te détourne pas, par lâcheté, du désespoir. Traverse-le. C'est par-delà  
qu'il sied de retrouver motif d'espérance. Va droit. Passe outre.  
De l'autre côté du tunnel tu retrouveras la lumière.*

*André Gide*





# Table des Matières

---

<b>TABLE DES MATIÈRES</b> .....	<b>1</b>
<b>LISTE DES ABRÉVIATIONS</b> .....	<b>5</b>
<b>AVANT PROPOS</b> .....	<b>9</b>
<b>1 INTRODUCTION</b> .....	<b>13</b>
1.1 LE MONDE PHARMACEUTIQUE .....	14
1.1.1 <i>Au tournant de son histoire</i> .....	14
1.1.2 <i>Remise en question de la découverte des médicaments et de la « druggabilité » des protéines cibles</i> .....	15
1.2 LA FAMILLE DES PROTÉINES KINASES HUMAINES .....	19
1.2.1 <i>Rôle physiologique</i> .....	19
1.2.2 <i>Fonction catalytique et conformation Active</i> .....	23
1.2.3 <i>Régulation de l'activité catalytique</i> .....	27
1.2.4 <i>Implication dans les maladies et premiers médicaments</i> .....	33
1.2.5 <i>Limitations rencontrées par les inhibiteurs</i> .....	38
1.2.6 <i>Les stratégies d'inhibition adoptées</i> .....	40
1.2.7 <i>Émergence des inhibiteurs de type covalent</i> .....	47
1.3 RÉACTION COVALENTE EN CHIMIE ET EN BIOLOGIE .....	57
1.3.1 <i>Réaction entre un électrophile et un nucléophile</i> .....	57
1.3.2 <i>Le glutathion pour réguler les métabolites réactifs</i> .....	62
1.3.3 <i>Mode naturel d'inhibition</i> .....	62
1.3.4 <i>Médicaments covalents ou inhibiteurs suicides ?</i> .....	63
1.3.5 <i>Autres usages de la liaison covalente</i> .....	65
1.3.6 <i>Pourquoi les inhibiteurs covalents sont-ils plus spécifiques qu'on ne le pense ?</i> .....	68
<b>2 MATÉRIEL ET MÉTHODES</b> .....	<b>69</b>
2.1 DE L'EXTRACTION À L'EXPLOITATION DES INFORMATIONS SUR LES PROTÉINES KINASES .....	70
2.1.1 <i>Création de la base de données dédiée aux kinases</i> .....	70
2.1.2 <i>Les bases de données interrogées</i> .....	71
2.1.3 <i>Extraction et organisation de l'information</i> .....	72
2.1.4 <i>Enrichissement de la base de données</i> .....	73
2.1.5 <i>Exploitation de la base de données des kinases</i> .....	73
2.2 CONCEPTION RATIONNELLE D'INHIBITEURS.....	75
2.2.1 <i>La cible kinase et les modèles 3D par homologie</i> .....	75
2.2.2 <i>Les petites molécules et futurs inhibiteurs</i> .....	75
2.2.3 <i>Méthodes d'amarrage (docking) et évaluation de l'affinité de l'inhibiteur (scoring)</i> .....	76
2.3 VALIDATIONS EXPÉRIMENTALES .....	79
<b>3 ÉTUDE DE LA DIVERSITÉ CONFORMATIONNELLE AU SEIN DES PROTÉINES KINASES</b> .....	<b>81</b>
3.1 UNE INFORMATION DIFFICILE À OBTENIR.....	82
3.2 RAPATRIEMENT ET ANALYSE DES STRUCTURES.....	83
3.3 STRATÉGIE D'ASSIGNATION DE LA CONFORMATION .....	84
3.4 PREMIÈRE ASSIGNATION DE LA CONFORMATION .....	89
3.5 DE TROIS À CINQ CONFORMATIONS .....	91
3.6 ARTICLE EN COURS DE SOUMISSION .....	92
3.7 QUEL NIVEAU DE DÉTAIL AU SEIN DE CHAQUE CONFORMATION ?.....	93
3.8 CONCLUSION .....	121
<b>4 CARTOGRAPHIE DES CYSTÉINES ET APPLICATION À LA CONCEPTION D'INHIBITEURS COVALENTS</b> .....	<b>125</b>
4.1 CARTOGRAPHIE DES CYSTÉINES SELON LA CONFORMATION DU DOMAINE CATALYTIQUE .....	126
4.2 CONCEPTION DU PREMIER INHIBITEUR COVALENT DE LA KINASE KIT .....	129
4.3 PUBLICATION .....	131
4.4 EFFET ANTIPROLIFÉRATIF SUR DES CELLULES HUMAINES DE LEUCÉMIE À MASTOCYTES.....	146

4.5	CONCEPTION RATIONNELLE D'INHIBITEURS COVALENTS DÉRIVÉS DE L'HYPOTHEMYCINE .....	150
4.5.1	<i>Amélioration du profil de sélectivité de l'hypothemycine parmi les kinases ayant une cystéine avant le motif DFG.....</i>	150
4.5.2	<i>Réadressage de l'hypothemycine en vue de l'inhibition covalente d'EGFR via la cystéine située à la fin de la région charnière.....</i>	156
4.6	CONCLUSION .....	171
<b>5</b>	<b>KINATOWN : UNE INTERFACE WEB DÉDIÉE AUX KINASES ET À LA CONCEPTION D'INHIBITEURS.....</b>	<b>173</b>
5.1	INTRODUCTION .....	174
5.2	ARTICLE EN COURS DE SOUMISSION .....	176
5.3	PERSPECTIVES .....	177
<b>6</b>	<b>CONCLUSIONS ET PERSPECTIVES.....</b>	<b>181</b>
<b>7</b>	<b>ANNEXES.....</b>	<b>189</b>
7.1	SECONDE CAMPAGNE D'INHIBITEURS D'EGFR .....	190
7.2	TRAVAUX SUR D'AUTRES INHIBITEURS COVALENTS DE KINASES D'INTÉRÊT THÉRAPEUTIQUE .....	194
7.2.1	<i>La kinase AXL et un second inhibiteur allostérique covalent en vue .....</i>	194
7.2.2	<i>La kinase NDR1 et une application de la méthodologie de « tethering ».....</i>	196
7.2.3	<i>Une seconde cystéine visée chez la kinase EGFR.....</i>	197
7.2.4	<i>HER3 comme nouvelle cible thérapeutique.....</i>	198
7.3	LA FAMILLE DES RÉCEPTEURS NUCLÉAIRES .....	199
7.3.1	<i>Introduction.....</i>	199
7.3.2	<i>Inhibiteurs covalents connus.....</i>	200
7.3.3	<i>Étude de la diversité conformationnelle.....</i>	201
7.3.4	<i>Cartographie des cystéines.....</i>	205
7.3.5	<i>Perspectives.....</i>	207
	<b>PUBLICATIONS ET COMMUNICATIONS.....</b>	<b>209</b>
	<b>BIBLIOGRAPHIE.....</b>	<b>211</b>



## Liste des abréviations

---

- ABL : *Abelson tyrosine kinase*
- A-loop : *Activation loop*
- AC : *Accession Code* dans *Uniprot*
- ADMET : *Absorption-Distribution-Métabolisation-Excrétion-Toxicité*
- ADP : *adénosine diphosphate*
- AKT2 : *RAC-beta serine/threonine-protein kinase*
- AMM : *Autorisation de Mise sur le Marché*
- ANP $\alpha$  : *Atrial Natriuretic Peptide receptor A*
- aPK : *atypical Protein Kinase*
- ATP : *adénosine triphosphate*
- AVR2B : *Activin receptor type-2B*
- BIRD : *Biological Integration and Retrieval of Data*
- BTK : *Bruton's Tyrosine Kinase*
- cAPK : *cAMP-dépendante protéine kinase*
- CDK2 : *Cell-Division kinase 2* ou *Cyclin-dependant kinase 2*
- Da : *Dalton*
- DFG : *Acide aspartique-Phénylalanine-Glycine*
- CML : *Chronic Myelogenous Leukemia*
- CSD: *Cambridge Structure Database*
- EGFR : *Epidermal Growth Factor Receptor*
- EPHA3 : *Ephrin type-A receptor 3*
- EPHB4 : *Ephrin type-B receptor 4*
- ePK : *eucaryotic Protein Kinase*
- FDA : *Food and Drug Administration*
- GIST : *Gastro-intestinal Stromal Tumors*
- GPCRs : *Récepteurs Couplés aux Protéines G*
- GSH : *Glutathion*
- HER3 : *receptor tyrosine kinase erbb3*

- HMC : *Human Mast Cell leukemia line*
- HOMO : *Highest Occupied Molecular Orbital*
- HTS : *High-Throughput Screening*
- IC<sub>50</sub> : *half maximal inhibitory concentration*
- ILK : *Integrin-Linked Kinase*
- INSR : Récepteur à l'INSuline
- JM : région JuxtaMembranaire
- KID : *Kinase Insert Domain*
- KIT : *Mast/stem cell growth factor receptor*
- K<sub>M, ATP</sub> : affinité de l'ATP
- LCK : *Lymphocyte-specific protein tyrosine kinase*
- LUMO : *Lowest Unoccupied Molecular Orbital*
- MK14 (ou p38α): *Mitogen-activated protein kinase 14*
- MS : *Mass spectrometry*
- NSCLC : *Non-Small Cell Lung Cancer*
- PDB : *Protein Data Bank*
- PDGFR : *Platelet derived Growth Factor Receptor*
- PPilot : *Pipeline Pilot*
- RAL : *Resorcylic Acid Lactone*
- rmsd : *root mean square deviation*
- RTK : Récepteur à activité Tyrosine Kinase
- SAR : *Structure-Activity Relationship*
- SCF : *Stem Cell Factor*





# Avant propos

---

Le durcissement des réglementations auquel est confronté le monde pharmaceutique, en terme d'efficacité et de non toxicité des molécules, rend de plus en plus difficile l'arrivée sur le marché de nouveaux médicaments. Pour remédier à cette difficulté, les industries pharmaceutiques tendent à repenser leur système de validation de cibles en amont des projets thérapeutiques afin de s'assurer de leur « druggabilité ». Il existe deux aspects de la « druggabilité », le premier faisant référence à l'implication de la protéine cible dans une maladie et aux conséquences de son inhibition, y compris dans les cellules saines. Le second aspect consiste à savoir si la cible peut être modulée par une petite molécule hautement affine et spécifique, afin de limiter les problèmes de toxicité, et qui soit insensible aux résistances que la cible peut développer. C'est principalement sur ce second aspect, dans le contexte des protéines kinases humaines et de la conception rationnelle d'inhibiteurs, que mes travaux de thèse ont porté.

La famille des protéines kinases humaines n'a pas toujours été la cible privilégiée des industries pharmaceutiques. Mais depuis l'autorisation de mise sur le marché en 2001 d'un premier inhibiteur de kinase, l'imatinib, l'intégration de cibles kinases au sein des industries pharmaceutiques et le développement de leurs inhibiteurs ont largement augmenté. Pourtant, très peu de ces molécules connaissent le succès de l'imatinib. En effet, un grand nombre d'entre elles voient leur développement arrêté, soit par manque d'efficacité face aux concentrations élevées d'ATP dans les cellules, soit pour des problèmes de sélectivité du fait des similarités de séquence entre les 518 membres que comporte la famille des kinases humaines. Au fur et à mesure de la résolution des structures 3D des complexes kinase/inhibiteur, dont le complexe de la cible ABL avec l'imatinib, il semblerait que la meilleure stratégie pour obtenir un inhibiteur affin et sélectif consiste à prendre avantage de la régulation allostérique à laquelle sont soumises certaines kinases, c'est-à-dire bloquer la cible kinase dans une conformation inactive.

Cependant, certains patients traités avec les inhibiteurs de kinases actuellement commercialisés ont commencé à présenter des phénomènes de résistance, c'est-à-dire des mutations sur le domaine catalytique ciblé, rendant l'inhibiteur inefficace. Une stratégie récente, et à laquelle les industries pharmaceutiques restent réticentes, consiste à inhiber la cible kinase en créant une liaison covalente entre un groupement électrophile positionné sur l'inhibiteur et un résidu nucléophile du site de liaison de l'ATP, le plus souvent une cystéine. Malgré l'aspect hautement réactif de ces inhibiteurs, il a été démontré que le choix du groupe électrophile et son association avec un squelette de molécule déjà sélectif permettaient

d'atteindre l'effet désiré tout en conservant un bon profil de sélectivité. L'entrée récente en phase 3 de développement clinique d'un inhibiteur covalent de la kinase EGFR, l'afatinib, démontre l'efficacité de cette stratégie. La cystéine avec laquelle cette petite molécule réagit lui confère deux avantages majeurs en comparaison de ses analogues non covalents : une efficacité prolongée due à l'élévation du temps de résidence au niveau de la cible, et surtout une bien meilleure efficacité face aux résistances développées par la cible.

Dans ce contexte, mes travaux de thèse tendent à démontrer la « druggabilité » des protéines kinases par la conception et la caractérisation d'inhibiteurs allostériques covalents, au moyen d'outils de bioinformatique et de chémoinformatique combinés à des tests expérimentaux.

Avant de détailler ces travaux, j'introduirai brièvement dans un premier chapitre les nouveaux *challenges* auxquels l'industrie pharmaceutique est confrontée, puis je décrirai plus longuement l'état des connaissances sur cette nouvelle famille de cibles que constituent les protéines kinases. Je terminerai l'introduction en présentant les mécanismes et usages de l'inhibition covalente dans le contexte biologique.

Dans le second chapitre, je présenterai les matériels et méthodes utilisés lors de ce travail de thèse. J'illustrerai notamment la création de la base de données de kinases et son enrichissement avec de nouvelles informations générées par mes travaux de thèse. Je décrirai également les outils utilisés pour exploiter cette base de données, dont les outils de modélisation moléculaire pour la conception rationnelle d'inhibiteurs de kinases.

Les travaux et résultats obtenus durant cette thèse seront détaillés dans trois chapitres. Dans le chapitre 3, je décrirai l'écriture du programme d'analyse des structures 3D des protéines kinases m'ayant permis d'explorer la diversité des conformations adoptées par ces protéines et leurs conséquences sur la forme du site actif. Je développerai dans ce chapitre la caractérisation d'une nouvelle conformation du domaine catalytique et son possible rôle dans le contexte de la régulation allostérique de l'activité des kinases. Un article, qui sera soumis prochainement, sera joint à ce chapitre. Par le biais de cette large étude conformationnelle, j'expliquerai dans le chapitre 4 la méthodologie d'identification des cystéines rendues accessibles dans chaque forme du site actif et les kinases concernées. L'ensemble de ces analyses m'ont permis de procéder à la conception rationnelle d'inhibiteurs covalents, qui ont été ou seront synthétisés dans l'équipe du Professeur Nicolas Winssinger (Institut de Science et d'Ingénierie Supramoléculaires, ISIS, Strasbourg), et que je présenterai dans ce même

chapitre. L'emphase portera sur la conception d'un inhibiteur allostérique covalent de la kinase KIT, inspiré du médicament non covalent imatinib. Les tests biologiques confirmant le mode d'action covalent et la sélectivité parmi les kinases humaines pour cet inhibiteur allostérique feront partie intégrante de ce chapitre. Notre article publié dans *Journal of Medicinal Chemistry* en février 2011 sera joint à ce chapitre. De plus, les derniers résultats des tests cellulaires obtenus, démontrant une efficacité conservée de notre inhibiteur sur l'arrêt de la prolifération de cellules humaines de leucémie à mastocytes, résistantes au traitement par l'imatinib, seront présentés. Deux autres séries d'inhibiteurs covalents seront également détaillées, dont une série visant la kinase EGFR et qui se révèle très prometteuse aux vues des premiers résultats des tests biologiques. Dans le chapitre 5, je présenterai Kinatown, une interface web que j'ai développée à l'intention de la communauté des biologistes et chimistes, et qui est dédiée à l'étude des conformations des kinases et à la conception d'inhibiteurs allostériques et/ou covalents. Un article décrivant cette interface, qui sera prochainement soumis dans le journal *Bioinformatics*, sera joint à ce chapitre.

Enfin, dans le dernier chapitre je discuterai des résultats et des perspectives ouvertes par mes travaux de thèse.

# 1 Introduction

---

1.1 Le monde pharmaceutique.....	14
1.2 La famille des protéines kinases humaines.....	19
1.3 Réaction covalente en chimie et en biologie.....	57

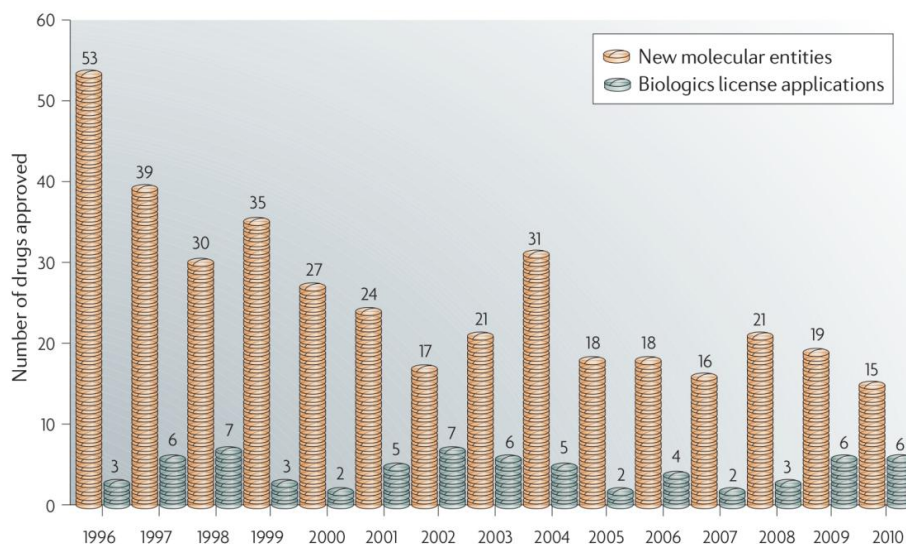
## 1.1 Le monde pharmaceutique

### 1.1.1 Au tournant de son histoire

Au cours de ces trente dernières années, les grands groupes pharmaceutiques ont développé et mis sur le marché des *blockbusters*, c'est-à-dire des médicaments prescrits à des millions de personnes, leur faisant gagner des sommes d'argent considérables. En 2010, 307 milliards de dollars ont été dépensés aux États-Unis pour l'achat de médicaments, dont les classes thérapeutiques les plus profitables sont les anticancéreux, les agents respiratoires, les régulateurs de lipide, les antidiabétiques et les antipsychotiques (Report by the IMS Institute for Healthcare Informatics, 2010). Cependant, l'industrie pharmaceutique arrive à un tournant de son histoire pour plusieurs raisons, énoncées dans les paragraphes suivants (Steiner et Bugen, 2007).

D'une part, les brevets qui protégeaient plusieurs médicaments à succès arrivent à échéance, confrontant les groupes concernés à l'arrivée sur le marché de molécules génériques à prix concurrentiels. L'exemple le plus actuel est celui de Pfizer qui vient d'annoncer la fermeture prochaine du site de Sandwich (UK) pour faire face à l'expiration du brevet de l'un de ses médicaments les plus lucratifs, le Lipitor®, un anticholestérol (Pfizer-Objectif de BPA 2012 confirmé en sabrant les coûts).

D'autre part, les coûts et les risques associés au développement de nouveaux médicaments ne cessent d'augmenter. En effet, de plus en plus de maladies rares ou orphelines sont ciblées alors que les connaissances accumulées dans ces domaines sont encore faibles et demandent plus d'efforts de recherche (Arrowsmith, 2011). Parallèlement, les agences évaluant les futurs médicaments, telles que la *Food and Drug Administration* (FDA) aux États-Unis, ont durci leurs règles pour n'autoriser l'arrivée sur le marché qu'à des molécules qui soient les plus efficaces et les moins toxiques possible (Hughes, 2009). Cela incite les industries pharmaceutiques à caractériser leur candidats-médicaments de façon très rigoureuse, impliquant des tests précliniques et cliniques excessivement onéreux. Une conséquence directe est la diminution du nombre de nouveaux médicaments arrivant sur le marché en 2010 par rapport aux années précédentes (Figure 1) (Mullard, 2011).



**Figure 1 : Évolution du nombre de médicaments approuvés par la FDA entre 1996 et 2010 (Mullard, 2011). Avant l'année 2000, en moyenne 40 petites molécules (*New molecular entities*) devenaient des médicaments chaque année. Ce nombre a diminué au fil des années pour atteindre seulement 15 nouveaux médicaments en 2010.**

Enfin, il faut souligner que l'industrie pharmaceutique est la plus encline à faire face à des poursuites judiciaires, aux vues d'effets secondaires imprévus, et parfois graves, observés chez certains patients. L'exemple actuel est celui du Mediator®, un antidiabétique également prescrit en tant que coupe-faim, commercialisé par les laboratoires Servier, et qui serait responsable du décès de plus de 500 personnes dans le monde (Mediator : Combien de morts ? | France Soir, 2011).

### 1.1.2 Remise en question de la découverte des médicaments et de la « druggabilité » des protéines cibles

La pression est grandissante pour les industries pharmaceutiques qui doivent trouver de nouveaux médicaments. Avec un coût de l'ordre de 800 millions de dollars répartis sur plus d'une dizaine d'années pour qu'une molécule parvienne sur le marché, celles-ci s'interrogent sur le taux élevé des échecs dans les phases de développement clinique, un moment où les sommes d'argent avancées sont les plus grandes (Figure 2) (Paul et al., 2010). Lors de ces phases, le candidat-médicament est d'abord testé sur des volontaires sains pour étudier son comportement dans l'organisme (pharmacocinétique, biodisponibilité) et estimer sa tolérabilité (phase 1), puis il est testé sur des patients pour évaluer son efficacité en parallèle de sa toxicité (phase 2) et doit enfin prouver qu'il est plus efficace et moins toxique que le traitement actuel (phase 3) pour justifier son entrée sur le marché.

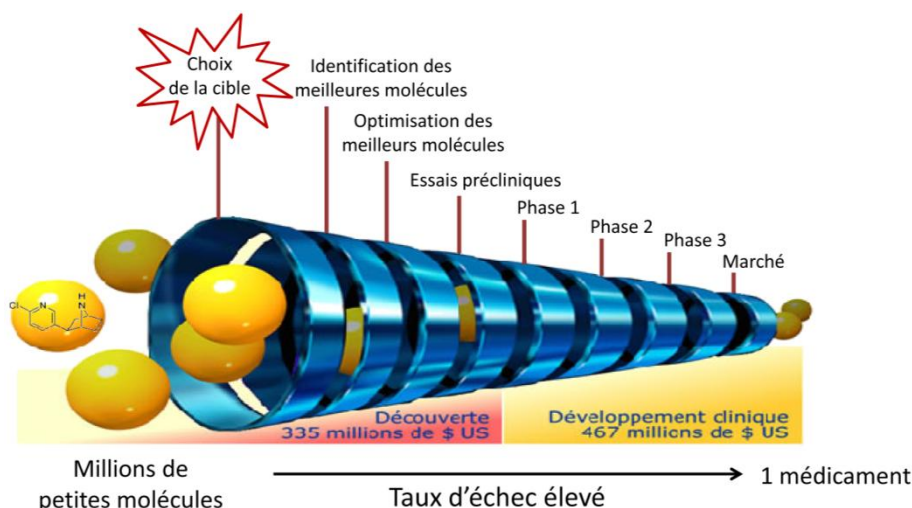


Figure 2: Schéma de développement d'un médicament. Le choix de la cible protéique, au début du processus, est déterminant pour la possibilité de trouver un médicament. L'enjeu économique est maximum lorsque des molécules atteignent les phases 1 à 3 (candidats-médicaments).

Alors que des techniques telles que le criblage à haut débit (HTS) ont été utilisées dès les années 1980 pour augmenter le nombre de molécules entrant dans les phases de développement clinique, un nombre croissant de molécules candidates ont été rejetées car elles présentaient des propriétés pharmacocinétiques inadaptées et une faible biodisponibilité (données de 1991 dans la Figure 3) (Kola et Landis, 2004; Kramer et al., 2007). L'industrie pharmaceutique a dû rapidement remettre en question ses critères d'évaluation des molécules et les raisons des échecs ont évolué. En 2000, le manque d'efficacité et les problèmes toxicologiques, liés aux interactions non-spécifiques entre la molécule et d'autres cibles, ont été le plus mis en cause (données de l'année 2000, Figure 3) (Kola et Landis, 2004). Cette tendance est restée la même en 2010 (Arrowsmith, 2011).

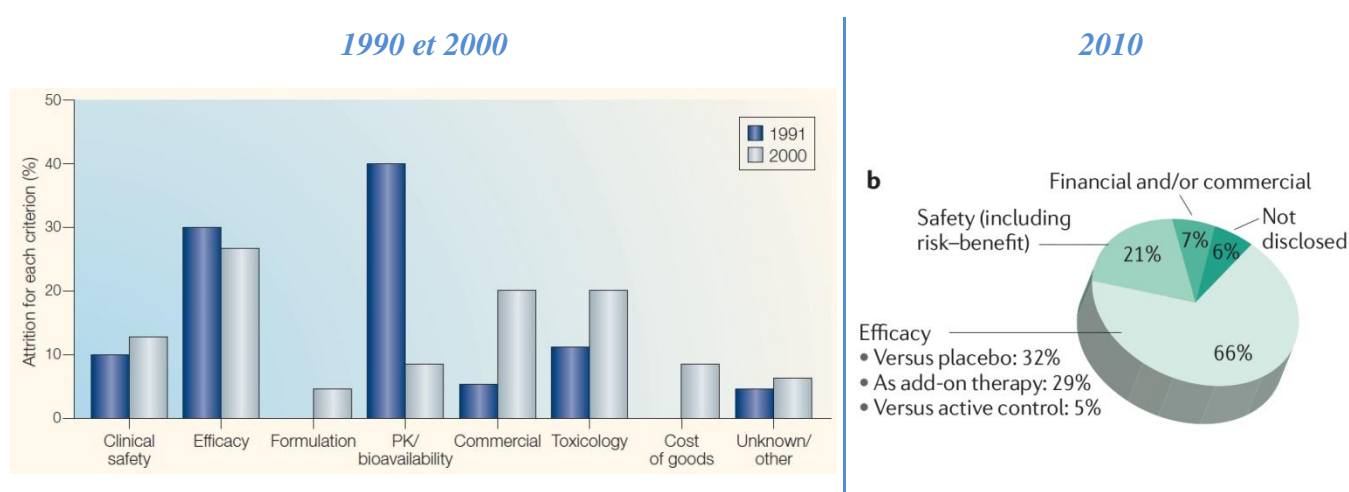
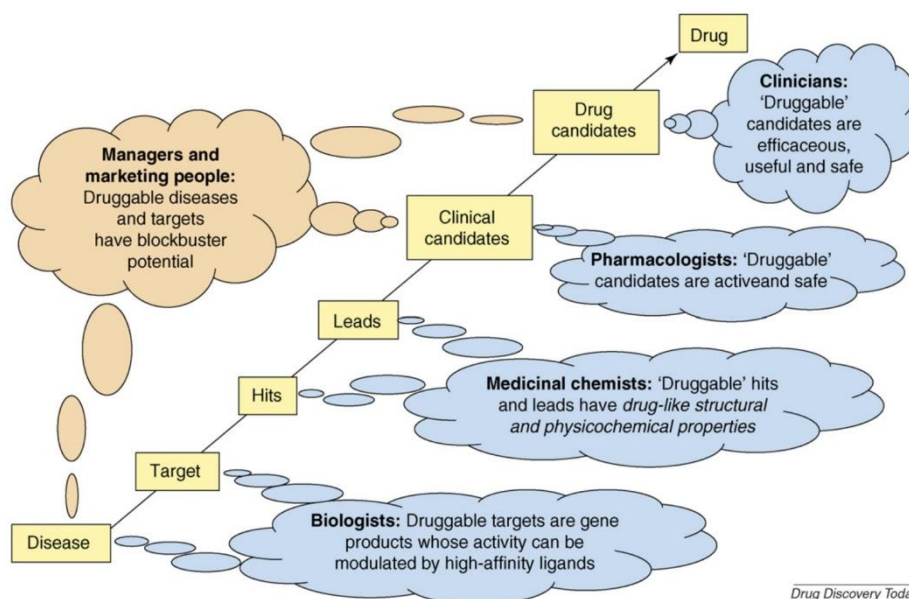


Figure 3 : Évolution des causes d'échecs dans les phases de développement clinique entre 1991 et 2010. En 2000, les échecs liés à la biodisponibilité et aux propriétés pharmacocinétiques sont moindres, tandis que la toxicologie est mise en cause. Les problèmes liés à l'efficacité des molécules restent d'actualité (Kola et Landis, 2004). En 2010, l'efficacité et la toxicité restent les problèmes majeurs (statistiques effectuées sur les molécules échouant en phase 3) (Arrowsmith, 2011).

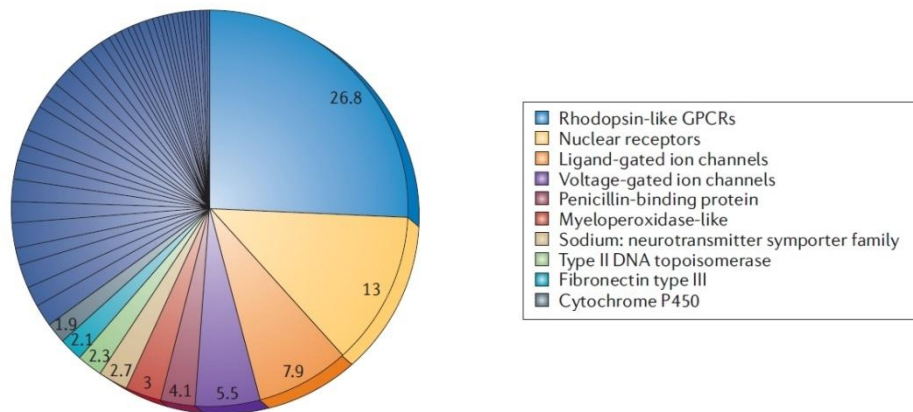


Toutes ces observations remettent en cause le choix de la protéine cible en tant que cible thérapeutique, ou *druggable*, au début du processus de développement d'un médicament (Figure 2) (Paul et al., 2010). Chaque acteur de la recherche est responsable d'un aspect de la « druggabilité » d'une cible : pour les biologistes, il s'agit de s'assurer que la modulation de l'activité de la cible aura l'effet recherché vis-à-vis de la maladie ciblée ; pour les chimistes médicaux et les pharmacologistes, la druggabilité réside dans la capacité à trouver une petite molécule à caractère *drug-like*, c'est-à-dire avec des propriétés pharmacocinétiques optimales et une bonne biodisponibilité orale, qui puisse se lier spécifiquement et de façon très affine à sa cible (Figure 4) (Vistoli et al., 2008).



**Figure 4 :** Les différents acteurs de la recherche traitent divers aspects de la « druggabilité ». Ainsi, les biologistes s'attendent à ce que la modulation de l'activité de la cible par une petite molécule produise le phénotype désiré, tandis que les chimistes recherchent des molécules *drug-like* capables de se lier de façon très affine à la cible. Les cliniciens veulent des molécules qui soient avant tout non-toxiques et qui se révèlent plus efficaces que les traitements actuels (Vistoli et al., 2008).

Parmi les 30 000 protéines que compterait le protéome humain, seules 200 protéines sont les cibles thérapeutiques d'une des quelques 2 000 petites molécules, administrables oralement, actuellement sur le marché (Overington et al., 2006; Pharmaceutical Research News & Drug Development Pipeline Information, 2010). Les familles de protéines les plus ciblées par des médicaments sont les récepteurs couplés aux protéines G (GPCRs) et les récepteurs nucléaires (NRs) (Figure 5) (Overington et al., 2006).



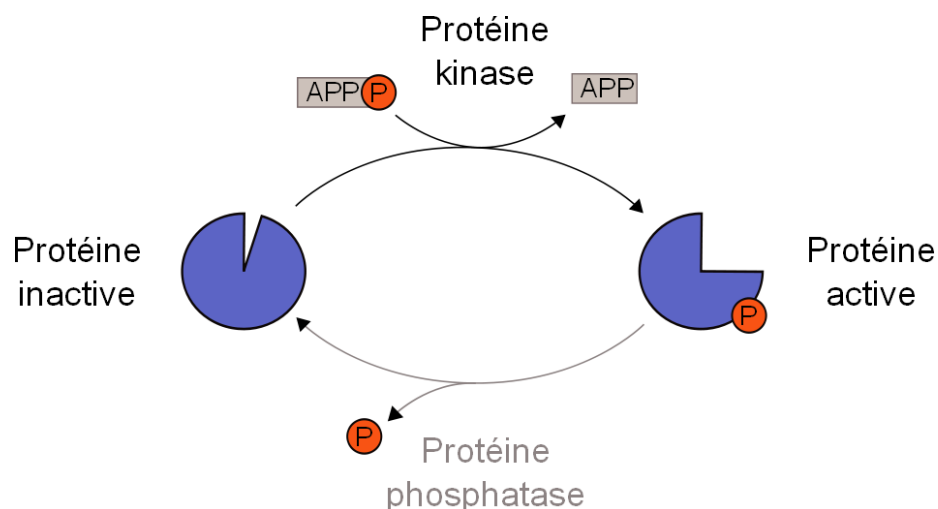
**Figure 5 : Répartition, en pourcentage, des familles de protéines ciblées par des médicaments, en 2006 (Overington et al., 2006). Les familles les plus ciblées sont les GPCRs, suivis des NRs et des canaux ioniques ligand-dépendants.**

Pour des raisons qui seront détaillées dans la partie suivante, la famille des protéines kinases, longtemps considérée comme *non druggable* par l'industrie pharmaceutique, a actuellement le vent en poupe, et tente de rattraper son retard pour rejoindre les GPCRs parmi les protéines les plus ciblées par des médicaments (Cohen, 2002a).

## 1.2 La famille des protéines kinases humaines

### 1.2.1 Rôle physiologique

Les kinases sont des enzymes du groupe des transférase, catalysant les réactions de phosphorylation par l'ajout d'un ion phosphate, provenant du nucléotide adénosine triphosphate (ATP), sur une molécule cible, appelée le substrat, qui peut être une protéine, un acide nucléique, un lipide, un sucre ou même une autre kinase (Figure 6). Dans ce manuscrit, nous nous focaliserons sur les protéines kinases, c'est-à-dire les kinases phosphorylant la chaîne latérale d'une sérine, d'une thréonine ou d'une tyrosine du substrat protéique (Figure 6). Les taux de phosphorylation relatifs de ces trois résidus sont approximativement 1000 : 100 : 1 pour la sérine, la thréonine et la tyrosine, respectivement. La réaction inverse de déphosphorylation est, quant à elle, catalysée par les phosphatases (Pao et al., 2007).



**Figure 6 :** Schéma d'activation d'une protéine via sa phosphorylation par une protéine kinase depuis une molécule d'ATP. La réaction inverse est catalysée par une protéine phosphatase.

La découverte de la phosphorylation en tant que mécanisme de régulation remonte au début des années 1950, avec les travaux de Krebs et Fischer sur la glycogène synthase, qui les ont amenés à isoler la phosphorylase kinase PHK (Krebs, 1998). Pourtant dès le 19<sup>ème</sup> siècle, il est connu que des groupements phosphates peuvent se lier aux protéines. Mais l'existence de ces « phosphoprotéines » était vue comme la conséquence de réactions métaboliques, et les recherches n'étaient pas allées plus loin. Il faudra attendre la fin des années 1970 pour qu'une seconde kinase, responsable de la phosphorylation et de l'activation de la première, soit découverte : la cAMP-dépendante protéine kinase (cAPK) (Walsh et al., 1968). Les travaux

sur les protéines kinases vont ensuite se multiplier, mettant au grand jour l'importance de cette modification post-traductionnelle. En effet, la phosphorylation d'un substrat protéique peut engendrer une amplification ou une inhibition de son activité, provoquer un changement de localisation cellulaire, ou encore un changement de structure qui permette l'association avec d'autres protéines (plus de détails seront donnés dans le paragraphe 1.2.3 « Régulation de l'activité catalytique »). La phosphorylation est synonyme de transduction du signal et fait intervenir des protéines kinases qui vont agir sur différents substrats spécifiques pour induire une réponse appropriée de la cellule. Des analyses des régions environnant le résidu phosphorylé permettent d'expliquer la spécificité pour certains substrats et d'établir des séquences consensus (Kennelly et Krebs, 1991; Protein Kinase Substrate Recognition). Les mécanismes de phosphorylation-déphosphorylation sont extrêmement courants pour réguler l'activité de la cellule (Graves et Krebs, 1999). Ainsi, les protéines kinases peuvent réguler une grande variété de processus cellulaires, parmi lesquels le métabolisme des carbohydrates et des lipides, la biosynthèse de neurotransmetteurs, la transcription et la réplication de l'ADN, le transport des organelles et la différenciation cellulaire (Adams, 2001). On estime à un tiers le nombre de protéines humaines qui seraient régulées de la sorte, faisant de la phosphorylation un des plus importants mécanismes post-traductionnels (Cohen, 2002b). Les recherches sur la phosphorylation ont d'ailleurs été saluées par l'attribution de trois prix Nobel en physiologie ou médecine :

- En 1992, à Edwin G. Krebs et Edmond H. Fischer, pour leurs découvertes sur « *reversible protein phosphorylation as a biological regulatory mechanism* » (The Nobel Prize in Physiology or Medicine 1992)
- En 2000, à Arvid Carlsson, Paul Greengard et Eric R. Kandel pour leurs découvertes sur « *signal transduction in the nervous system* » (The Nobel Prize in Physiology or Medicine 2000)
- En 2001, Leland H. Hartwell, Tim Hunt et Sir Paul M. Nurse pour leurs découvertes sur « *key regulators of the cell cycle* » (The Nobel Prize in Physiology or Medicine 2001)

En 1995, alors que de plus en plus de kinases sont identifiées, les premières analyses génomiques laissent présager que les génomes eucaryotes pourraient contenir plus de 2000 protéines kinases (Hunter, 1994). Finalement, le séquençage du génome humain associé aux travaux de Hanks et Hunter (Hanks et Hunter, 1995) et de Manning et al. (Manning et al.,

2002) recensera 518 protéines kinases humaines. Ce chiffre place les kinases comme l'une des plus larges familles de protéines humaines, constituant environ 1.7% du génome humain.

Parmi les 518 protéines kinases humaines, regroupées sous le mot « kinome », 478 possèdent un (voir deux) domaine catalytique de type eucaryote (ePK), tandis que les 40 autres sont dites « atypiques » (aPK). Pour ces dernières, le domaine catalytique est non similaire à celui des kinases ePK mais l'activité protéine kinase a été démontrée (Manning et al., 2002). Les 491 domaines catalytiques humains ePK se répartissent en neuf groupes phylogénétiques, principalement d'après les identités de séquence entre les domaines catalytiques : AGC, CAMK, CMGC, CK1, STE, TK, TKL, RGC, Autres (Tableau 1, Figure 7) (Manning et al., 2002). Les similarités et les différences du point de vue des séquences et des structures 3D seront discutées dans le paragraphe suivant et dans le chapitre 3 « Étude de la Diversité Conformationnelle au Sein des Protéines Kinases ».

Groupe	Nombre de domaines catalytiques	Informations succinctes	Quelques exemples de kinases, citées dans ce manuscrit
AGC	63	Phosphoryle préférentiellement les résidus sérine et thréonine proches d'acides aminés basiques tels que la lysine ou l'arginine.	cAPK RSK1..4
CAMK	82	Très proche du groupe AGC. Phosphoryle également des substrats contenant des résidus basiques.	PHK, TTN RSK1..4 (second domaine, pseudokinase)
CMGC	61	Large groupe présent chez tous les eucaryotes. Rôles importants dans le cycle cellulaire, la transduction du signal (en bas de la cascade des MAP kinases), etc.	CDK1..11 ERK1..7, MK10, MK14 GSK3 $\alpha$ , $\beta$
CK1	12	Petit groupe, présent chez tous les eucaryotes.	CK1 $\alpha$ .. $\epsilon$
STE	48	Rôles importants dans la transduction du signal (en haut de la cascade des MAP kinases), etc. Certaines sont duales, c'est-à-dire capables de transférer le phosphate aussi bien sur une sérine ou une thréonine que sur une tyrosine.	MAP2K1..7 MAP3K1..8
TK	94	Phosphoryle spécifiquement les tyrosines. Rôles importants dans la transduction du signal intra et intercellulaire. La moitié de ces kinases ont un domaine récepteur transmembranaire avant le domaine catalytique, et sont appelées Récepteurs à activité Tyrosine Kinase (RTKs).	Non RTKs : ABL, SRC, BMX, BTK, TEC, CSK, LCK, FAK, HCK JAK1..3 (un domaine catalytique et un domaine pseudokinase) RTKs : EPHA1..10, EPHB1..6, EGFR/Her1...4, KDR, KIT, PDGFR $\alpha$ , $\beta$ , FMS, INSR, FGFR1..4
TKL	43	Présente une similarité de séquence proche des TKs, mais ne phosphoryle que les résidus sérine ou thréonine.	ILK (pseudokinase), TAK1
RGC	5	Très petit groupe, présentant une similarité de séquence proche des TKs, mais ne phosphoryle que les résidus sérine ou thréonine.	ANP $\alpha$ (pseudokinase)
Autres	83	Faible similarité de séquence avec les 8 autres groupes.	STK6, WNK1..4, CK2 $\alpha$ , PLK1, NEK2, I $\kappa$ B
Atypiques	40	Absence de similarité de séquence bien que capable de phosphoryler des protéines. Pour celles dont la structure 3D est connue, le repliement est globalement conservé avec les ePK.	mTOR

**Tableau 1 : Informations relatives aux neuf groupes de kinases humaines de type ePK ainsi qu'aux Atypiques (aPKs).**

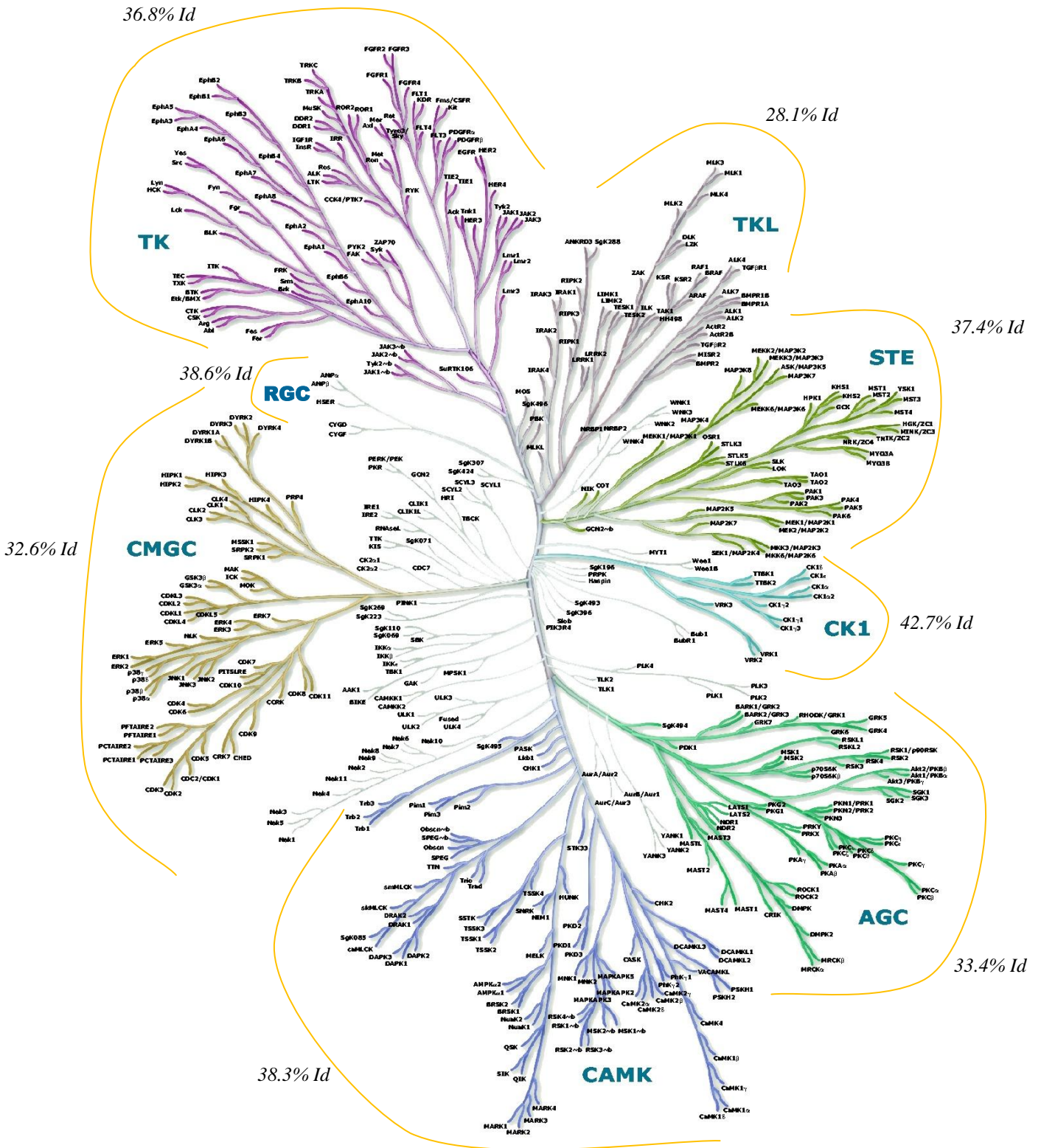


Figure 7 : Arbre phylogénétique des 478 protéines kinases humaines de type ePK soit 491 domaines catalytiques distincts et répartis en 9 groupes, le groupe « Autres » correspondant aux branches les plus claires (Manning et al., 2002). Les pourcentages d'identité de séquence (*Id*) au sein de chaque groupe sont également présentés, le pourcentage d'identité sur la totalité des séquences étant de 22.6%.

Les autres variations existantes entre les protéines kinases se localisent en dehors de ces domaines catalytiques et consistent, la plupart du temps, en une addition de domaines ou de sous-unités ayant un rôle dans la régulation de l'enzyme (Figure 8). Certains de ces domaines additionnels seront discutés plus en détail dans le paragraphe 1.2.3 « Régulation de l'activité catalytique ».

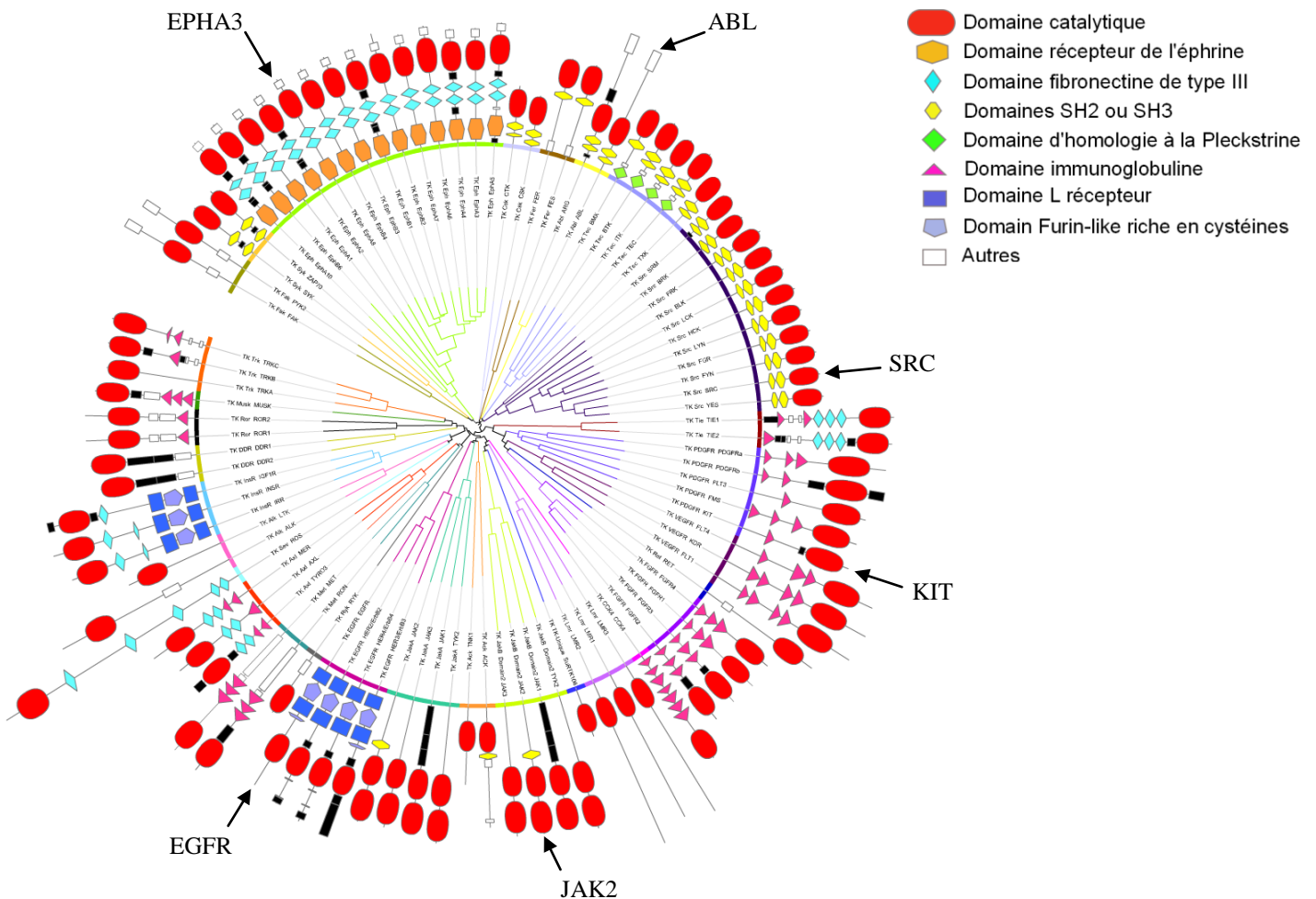
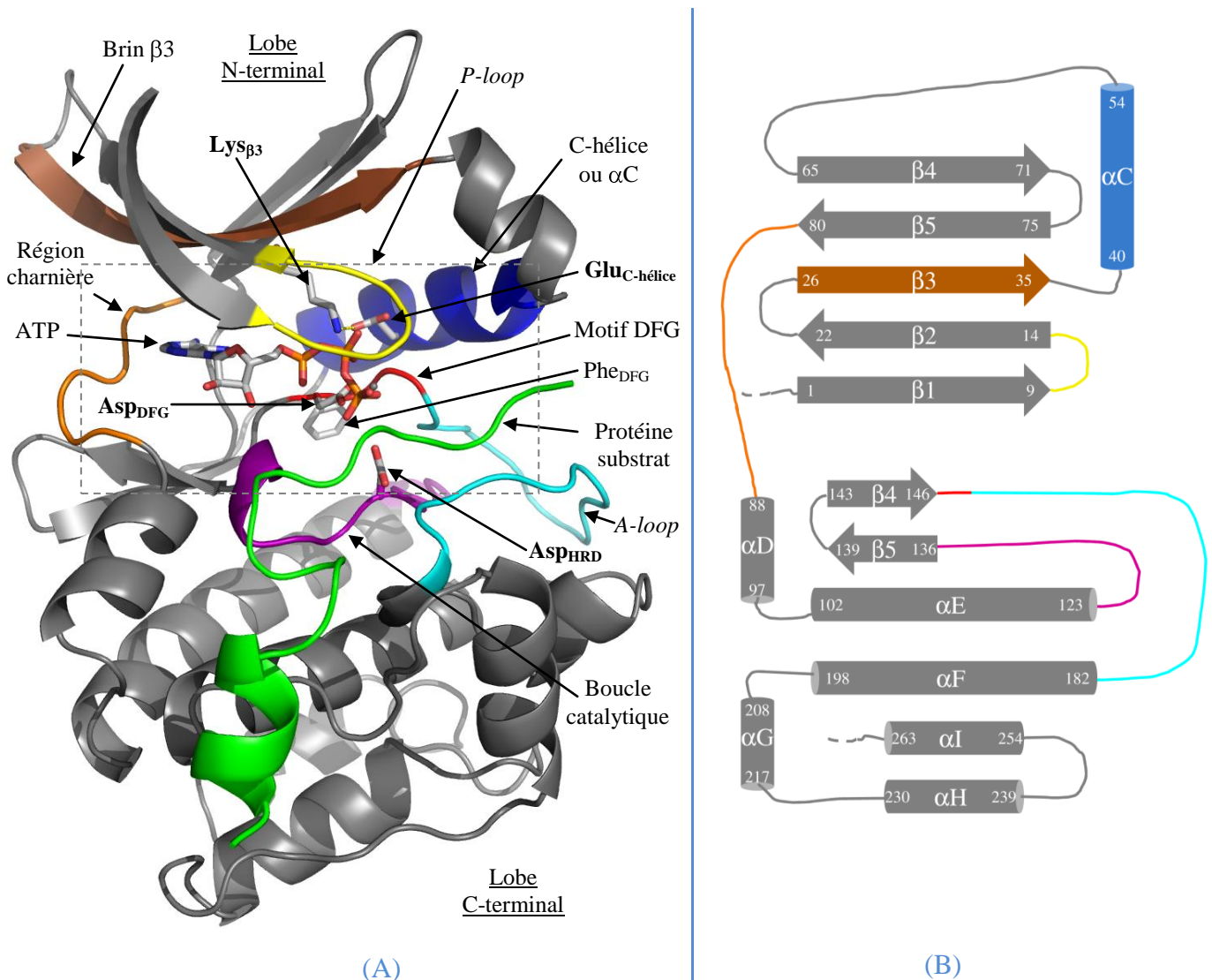


Figure 8 : Arbre phylogénétique du groupe des tyrosines kinases (TK) humaines, avec leur organisation en domaines (construit avec *iTOL* (Letunic et Bork, 2007)). Tandis que certaines kinases telles que JAK2 possèdent deux domaines catalytiques (représentés par des formes ovales rouges), la plupart n'en possèdent qu'un seul. Le nombre et le type de domaines additionnels est également très variable. Quelques kinases citées dans ce manuscrit sont également mises en avant.

### 1.2.2 Fonction catalytique et conformation Active

La première structure 3D d'un domaine catalytique de kinase déposée à la *Protein Data bank* (PDB) est celle de cAPK (groupe AGC), déterminée par cristallographie aux rayons X en 1991 (Knighton et al., 1991) (code PDB : 2CPK). Une seconde structure expérimentale est

rapidement obtenue, où cAPK est cocristallisée avec l'ATP et un peptide substrat (Zheng et al., 1993c) (code PDB : 1ATP ; Figure 9A). La description du repliement du domaine catalytique dans la conformation *Active* ainsi que du mode de fonctionnement du plateau catalytique se révélera commune à l'ensemble des kinases, au fur et à mesure de l'apparition de nouvelles structures 3D. Les similarités observées dans les alignements de séquences (Xu et al., 1995; Robinson et al., 1996; Russo et al., 1996b) sont ainsi confirmées au niveau de la structure tridimensionnelle (codes PDB : 1CSN - CK1 kinase ; 1GOL - Erk2 kinase ; 1JST - CDK2 kinase). Certaines structure 3D laissent également entrevoir les premiers éléments de régulation (Hubbard et al., 1994; Schulze-Gahmen et al., 1996) (codes PDB : 1IRK - INSR kinase ; 1HCL - CDK2 kinase).



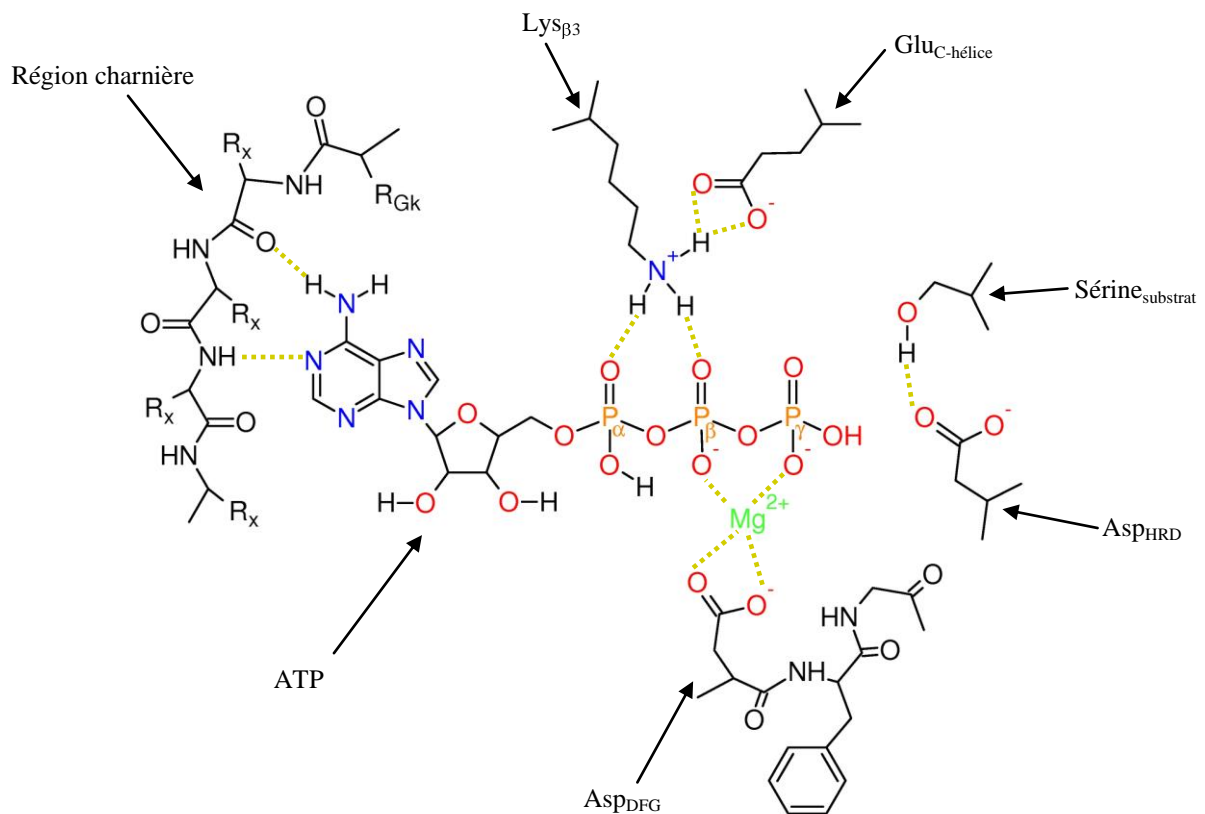
**Figure 9 :** Le domaine catalytique kinase dans la conformation *Active*. (A) Exemple de cAPK (code PDB : 1ATP) représentée en ruban. La région encadrée en pointillés gris correspond aux sites de reconnaissance du substrat (en vert) et de l'ATP (en mode bâton). Les principaux éléments structuraux sont annotés. Les quatre résidus catalytiques sont en gras. Remarque : ce code couleur sera strictement conservé dans les prochaines figures. Une particularité de cAPK est la présence d'une petite hélice avant la C-hélice. (B) Diagramme topologique avec tous les éléments structuraux, leur longueur moyenne sur l'ensemble des kinases humaines et leur connectivité.



Le domaine catalytique est constitué d'environ 270 acides aminés répartis dans deux lobes (Figure 9). Le lobe N-terminal, qui est le plus petit avec environ 80 résidus, est essentiellement constitué d'un feuillet antiparallèle à cinq brins  $\beta$  et d'une hélice nommée « C-hélice » ou  $\alpha_C$ . La boucle reliant les deux premiers brins  $\beta$  est nommée « *P-loop* » ou « *Glycine-rich loop* » ou encore « *Nucleotide binding loop* ». Le lobe C-terminal, long d'environ 190 résidus, est, quant à lui, essentiellement constitué d'hélices  $\alpha$ , mais également de deux brins  $\beta$  et de deux boucles très importantes que sont la boucle catalytique et la boucle d'activation (« *A-loop* »). Les deux lobes sont reliés par une boucle dite « région charnière », qui leur confère une certaine flexibilité de l'un par rapport à l'autre (Johnson et al., 1996).

La phosphorylation du substrat se produit dans le site catalytique, ou site actif, qui se trouve à l'intersection des deux lobes (Figure 9A). Ce site comprend deux poches juxtaposées : l'une pour recevoir l'ATP, l'autre pour accueillir le substrat. A priori, la fixation des deux molécules se fait indépendamment l'une de l'autre (Adams, 2001). Le site de reconnaissance du substrat, formé par la *A-loop* qui est peu conservée et de longueur variable, constitue la principale différence notable dans l'alignement des domaines catalytiques de kinases, et contribue ainsi à la spécificité de chaque enzyme (Adams, 2001). Le positionnement de la *A-loop* est crucial pour permettre l'amarrage du substrat (Johnson et al., 1996) et sera discuté plus en détail dans le paragraphe 1.2.3 « Régulation de l'activité catalytique ». La phosphorylation se fait par transfert du phosphate  $\gamma$  de l'ATP vers l'hydroxyle du résidu sérine (ou thréonine ou tyrosine) du substrat, probablement via un mécanisme dissociatif (Adams, 2001). Pour cela, le positionnement de l'ATP dans son site est crucial (Figure 10). La partie adénine est localisée dans une poche formée de résidus hydrophobes provenant des deux lobes, dans laquelle elle s'oriente grâce à deux liaisons hydrogènes avec la chaîne principale de la région charnière (Figure 10). Le groupement ribose est stabilisé par des liaisons hydrogènes avec des résidus du lobe C-terminal. Les trois phosphates sont alignés pour la catalyse de par leurs interactions avec la *P-loop* ainsi qu'avec quatre acides aminés dit « catalytiques », hautement conservés parmi toutes les kinases humaines, et dont le rôle exact reste débattu (Adams, 2001). Une lysine, située au milieu du brin  $\beta_3$  (Lys $_{\beta_3}$  sur la Figure 9 ; Figure 10), est stabilisée par une liaison ionique avec l'acide glutamique de la C-hélice (Glu $_{C\text{-hélice}}$  sur la Figure 9 ; Figure 10) afin de contribuer au positionnement des phosphates  $\alpha$  et  $\beta$ . L'acide aspartique situé au début de la *A-loop*, plus précisément dans le motif très conservé DFG (Asp $_{DFG}$  sur la Figure 9 ; Figure 10), coordonne l'ion métallique, souvent Mg $^{2+}$ , qui serait nécessaire au positionnement des deux phosphates  $\beta$

et  $\gamma$ . Cet acide aspartique pourrait également servir à masquer la charge du phosphate  $\gamma$  lors du transfert. Quant au rôle de l'acide aspartique situé sur la boucle catalytique, plus précisément dans le motif très conservé HRD ( $Asp_{HRD}$  sur la Figure 9 ; Figure 10), il pourrait s'agir d'orienter l'hydroxyle du substrat avant l'attaque ou bien de faciliter la dissociation après réaction, par repoussement des charges négatives. Une fois la réaction terminée, l'ADP résultant et le substrat phosphorylé sont relâchés.



**Figure 10 :** Représentation 2D du positionnement de l'ATP, et du résidu phosphorylable sur le substrat ( $Sérine_{substrat}$ ), dans le site catalytique et rôles supposés des quatre résidus catalytiques ( $Lys_{\beta 3}$ ,  $Glu_{C-hélice}$ ,  $Asp_{DFG}$ ,  $Asp_{HRD}$ ). Les liaisons hydrogènes sont représentées en pointillés jaunes. Les résidus du lobe C-terminal stabilisant le ribose ne sont pas représentés.

Cependant, il a été démontré que plusieurs domaines ePK n'ont pas d'activité catalytique. Ceci peut être expliqué par l'absence d'un des quatre résidus catalytiques très conservés (Manning et al., 2002). D'après ce critère, environ 10% des kinases n'auraient pas d'activité catalytique, et sont appelées « pseudokinases ». Or, le repliement de ces domaines est bien conservé, ce qui suggère qu'ils ont soit une autre fonction soit un mécanisme catalytique différent (Boudeau et al., 2006). En effet, certaines pseudokinases agissent en tant que substrats d'autres kinases. C'est le cas du récepteur  $ANP\alpha$ , dépourvu de l'acide aspartique situé sur la boucle catalytique, dont la phosphorylation est essentielle pour permettre l'activation de la guanylyl cyclase-A (Potter et Hunter, 1998). D'autres

pseudokinases sont des régulateurs de l'activité catalytique d'autres kinases. Par exemple, l'activité catalytique de la tyrosine kinase JAK2 (Saharinen et al., 2000) qui contient deux domaines kinases, dont un domaine dépourvu des acides aspartique et glutamique de la C-hélice et de la boucle catalytique, est régulée par son domaine pseudokinase. Elles peuvent aussi être un soutien pour l'assemblage de complexes de signalisation, comme pour la kinase ILK (Wickstrom et al., 2010) qui interagit avec le domaine cytoplasmique de l'intégrine  $\beta 1$ . Il est également important de noter que le mécanisme catalytique ne nécessite pas forcément ces quatre résidus, comme il l'a récemment été démontré pour le récepteur HER3 (Shi et al., 2010). Il peut aussi arriver qu'un autre résidu du site catalytique se substitue à l'un des quatre premiers, comme dans le cas de la kinase WNK1 (Xu et al., 2000), où la lysine catalytique est absente du brin  $\beta 3$  et dont le rôle est assuré par une lysine située à la fin de la *P-loop* (code PDB : 3FPQ (Min et al., 2004)) .

### 1.2.3 Régulation de l'activité catalytique

La fonction catalytique des kinases est si déterminante pour la cellule que ces dernières sont soumises à divers modes de régulation, permettant ainsi un contrôle fin des événements de phosphorylation, donc des événements biologiques. Tous les modes de régulation possibles ne sont probablement pas encore connus mais plusieurs ont déjà été identifiés et peuvent se répartir en deux groupes : soit la kinase est catalytiquement active mais d'autres paramètres empêchent l'ATP ou le substrat de se fixer ; soit la kinase est catalytiquement inactive du fait de déplacements d'éléments clés participant au domaine catalytique, on parle alors de régulation allostérique (Johnson et al., 1996; Taylor et Radzio-Andzelm, 1997; Hubbard, 2002). Dans ce dernier cas, les changements conformationnels rendent impossible la fixation de l'ATP ou du substrat protéique. D'une façon générale, il est communément admis que dans l'environnement cellulaire, les kinases transitent entre la conformation catalytiquement active et des conformations inactives (Johnson et al., 1996), sous l'effet de moyens de régulation qui peuvent être complexes. Actuellement, deux conformations inactives communes à plusieurs kinases ont pu être caractérisées : la conformation inactive *C-helix out* et la conformation inactive *DFG out*. Avant de décrire ses deux conformations inactives ainsi que deux autres modes de régulation allostérique, le paragraphe suivant montre des exemples où des phénomènes extérieurs au domaine catalytique jouent le rôle de régulateurs.

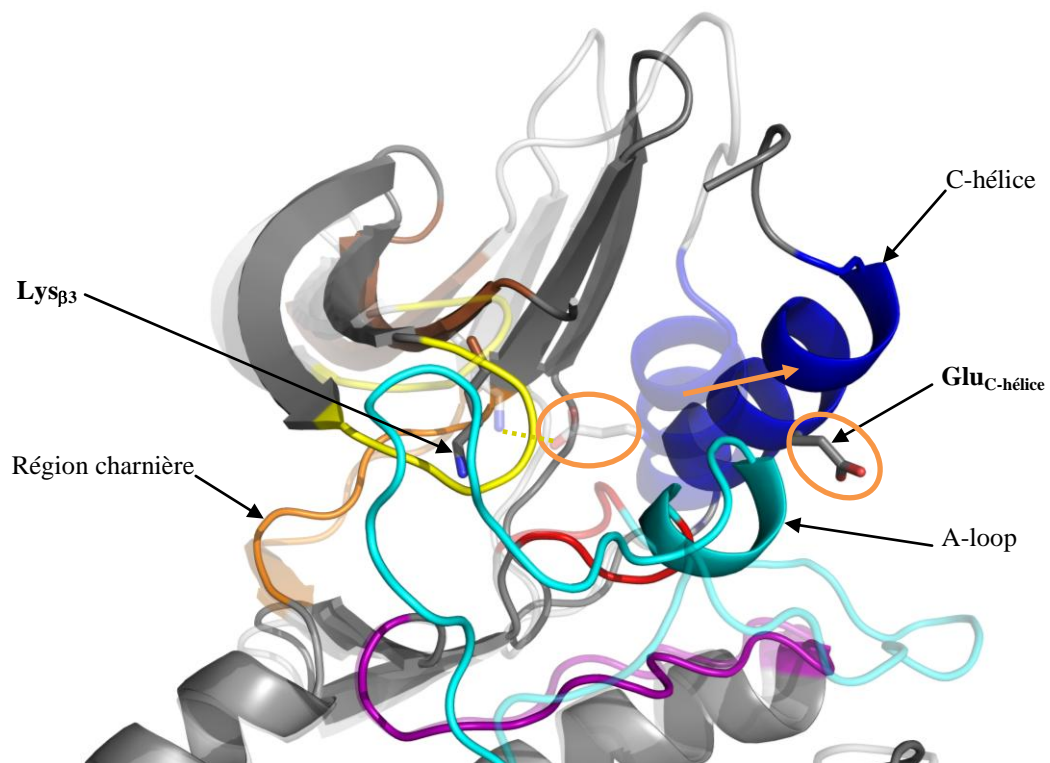
### 1.2.3.a Interaction avec des régulateurs

Dans ce premier mode de régulation, on citera les kinases dont les domaines ou les extensions environnant le domaine catalytique peuvent venir interférer avec les sites de fixation de l'ATP ou du substrat protéique. C'est le cas de la kinase Titin, ou TTN, dont l'extension C-terminale bloque l'accès aux deux sites de fixation (Mayans et al., 1998) (code PDB : 1TKI). Ces phénomènes d'autoinhibition sont relativement courants chez les kinases (Hubbard, 2002). Les mêmes observations ont été faites avec des peptides inhibiteurs endogènes, comme il l'a été reporté pour la kinase cAPK (Zheng et al., 1993a) et le peptide inhibiteur PKI (Dalton et Dewey, 2006) (code PDB : 1ATP), ou encore la kinase CDK2 et le peptide inhibiteur p27<sup>kip1</sup> dont une tyrosine pointe dans le site de fixation de l'ATP (Russo et al., 1996a). Plus subtilement, la kinase peut être séquestrée à un endroit de la cellule où son substrat n'est pas présent (Bar-Sagi et al., 1993). Tel est le cas pour la kinase SRC dont l'association avec la kinase FAK, qui interagit avec les domaines additionnels SH2-SH3 de SRC, est nécessaire à sa localisation subcellulaire et donc à son activité enzymatique (Schaller et al., 1999).

### 1.2.3.b Conformation inactive C-helix out

La première kinase cristallisée dans sa conformation inactive *C-helix out* est la kinase CDK2 (De Bondt et al., 1993) en 1993 (groupe CMGC ; codes PDB : 1HCL et 1HCK ; Figure 11). Cette kinase participe au processus de division cellulaire en orchestrant les transitions d'une phase à une autre, en particulier les phases G1/S et S (Neganova et al., 2011). Bien qu'il s'agisse d'une des plus petites kinases, dans le sens où elle ne possède que le domaine catalytique, la régulation de cette enzyme est primordiale et passe par son association avec une protéine régulatrice, la cycline (Celis et al., 1987). En effet, sans la présence de cette cycline, le domaine catalytique de CDK2 présente des modifications rendant impossible la réaction de phosphorylation, ni même la fixation du substrat. Plus exactement, la liaison ionique entre la lysine catalytique du brin  $\beta_3$  et l'acide glutamique de la C-hélice (Lys $_{\beta_3}$ -Glu $_{C\text{-hélice}}$ ) est rompue, du fait d'un mouvement de recul de l'extrémité N-terminale de la C-hélice par rapport au site catalytique (Figure 11). La *A-loop* subit aussi quelques changements. Elle forme en particulier une courte hélice juste après le motif DFG puis dévie complètement de sa trajectoire adoptée en conformation *Active* (Figure 11).

Il convient de noter que, bien que le déplacement de la *A-loop* empêche le substrat de se fixer, l'ATP a toujours accès à son site de fixation. En 1995, la première structure du complexe CDK2/Cycline, résolue par Jeffrey et al. (Jeffrey et al., 1995), dévoile la zone d'interaction entre les deux protéines. Cette zone est exactement située dans l'environnement de la C-hélice et de la *A-loop*, provoquant ainsi le repositionnement correct des acides aminés catalytiques et de la *A-loop* (code PDB : 1FIN).



**Figure 11 :** Représentation en ruban du domaine catalytique, centré sur le site catalytique, en conformation inactive *C-helix out* de la kinase CDK2 (code PDB : 1HCL). Une autre structure de la même kinase dans la conformation Active est représentée en transparent (code PDB : 1FIN), ce qui permet de souligner les déplacements de la C-hélice (flèche orange, le résidu catalytique  $\text{Glu}_{\text{C-hélice}}$  situé sur la C-hélice est entouré en orange) et de la *A-loop*.

La seconde kinase ayant été observée dans la conformation inactive *C-helix out* est la kinase SRC (Xu et al., 1997), cristallisée en 1997, avec ses domaines SH2 et SH3 en interaction avec le domaine catalytique (groupe TK ; code PDB : 1FMK). Dans cette structure, SRC présente une tyrosine phosphorylée sur l'extension C-terminale, contribuant à l'amarrage du domaine SH2 le long du domaine catalytique, et provoquant ainsi l'adoption de la conformation inactive *C-helix out*. C'est plus exactement la boucle reliant le domaine SH2 au domaine catalytique qui vient perturber le lobe N-terminal en logeant la chaîne latérale d'un tryptophane entre le feuillet et la C-hélice, provoquant le déplacement de cette dernière (Xu et al., 1997). Ce mécanisme a été validé par des tests d'activité sur des mutants de HCK, kinase de la même sous-famille que SRC, où le tryptophane - conservé dans la sous-famille - a été remplacé par une alanine (LaFevre-Bernt et al., 1998). HCK a d'ailleurs été la troisième

kinase à être cristallisée en conformation inactive *C-helix out* et montre un assemblage identique à celui de SRC (code PDB : 1AD5) (Sicheri et al., 1997). Inversement, la première structure de SRC en conformation *Active* présente une extension C-terminale non phosphorylée, expliquant la complète dissociation entre le domaine catalytique et les domaines SH2-SH3 (code PDB : 1Y57) (Cowan-Jacob et al., 2005).

### 1.2.3.c Conformation inactive DFG out

Le premier domaine catalytique à avoir été cristallisé dans cette conformation inactive est celui d'une tyrosine kinase : le récepteur à l'insuline INSR en 1994 (groupe TK ; code PDB : 1IRK), soit trois ans après la première structure de la kinase cAPK (Hubbard et al., 1994). Dans cette structure non ligandée et non phosphorylée, la totalité de la *A-loop* subit un déplacement important. Le mouvement le plus remarquable est opéré par le motif conservé DFG, dont la phénylalanine - qui ne participait pas au site catalytique dans la conformation *Active* - subit une rotation de 180° autour de la chaîne principale, soit un déplacement d'environ 10Å de sa chaîne latérale. Ce mouvement amène la phénylalanine à pointer dans le site de fixation de l'ATP, dans la zone située entre la poche de fixation du groupe adénine et celle du ribose. La suite de la *A-loop* étant également déplacée, ni l'ATP ni le substrat ne peuvent se fixer (Figure 12).

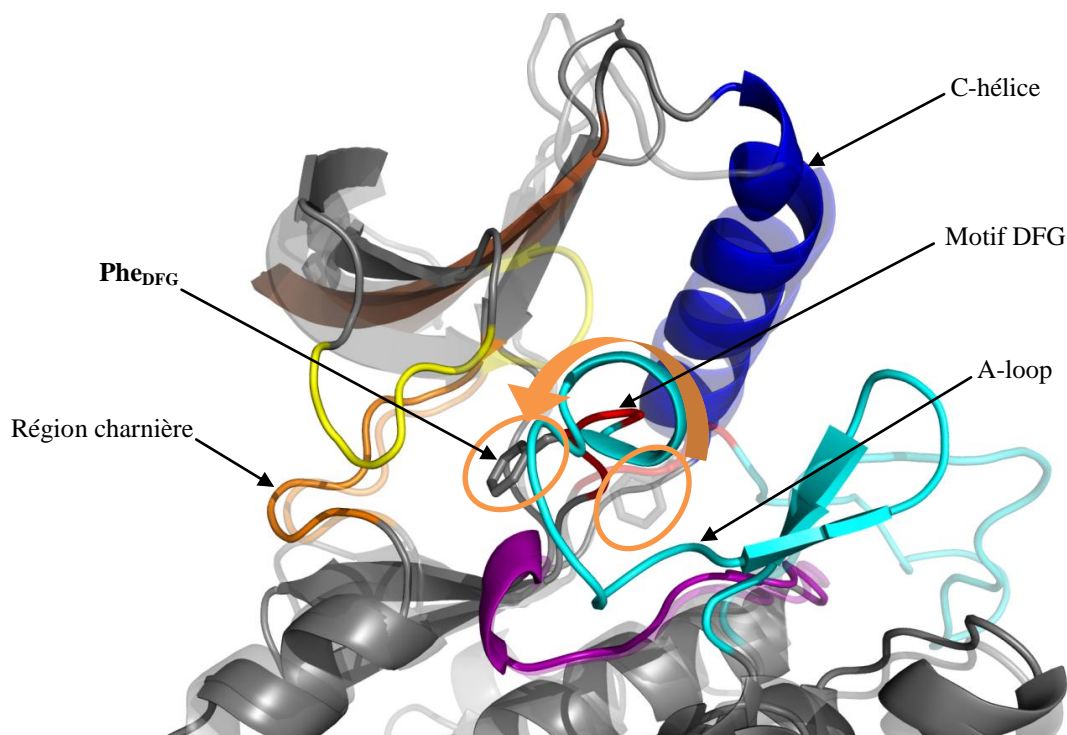


Figure 12 : Représentation en ruban du domaine catalytique, centré sur le site catalytique, en conformation inactive *DFG out* de la kinase ABL (code PDB : 1FPU). Une autre structure de la même kinase dans la conformation *Active* est représentée en transparent (code PDB : 1GAG), ce qui permet de souligner le déplacement du motif DFG (flèche orange, le résidu Phe<sub>DFG</sub> est entouré en orange) et du reste de la *A-loop*.

Le terme *DFG out* sera employé, plusieurs années plus tard, pour décrire les structures d'autres kinases cristallisées dans une conformation similaire. Il faudra en effet attendre la structure de la kinase ABL (groupe TK) cocrystallisée avec l'inhibiteur commercialisé imatinib, réalisée en 2000, pour voir réapparaître cette conformation (code PDB : 1FPU ; Figure 12) (Schindler et al., 2000), contre toute attente (plus de détails seront donnés dans le paragraphe 1.2.5 « Limitations rencontrées par les inhibiteurs »). Puis en 2002, c'est une serine/thréonine kinase, la *mitogen-activated protein kinase 14* dite MK14 ou p38 $\alpha$  (Pargellis et al., 2002), qui est cristallisée dans la conformation inactive *DFG out*, ainsi nommée pour la première fois par les auteurs (codes PDB : 1KV1 et 1KV2).

Toutes les structures cristallographiques précitées ne contiennent que le domaine catalytique, dans un état non phosphorylé. Les auteurs se posèrent donc la question de la viabilité de cette conformation inactive *DFG out* dans la cellule et de la présence possible de sous-unités régulatrices pouvant la provoquer. Un élément de réponse fut dévoilé en 2004, avec la résolution de la structure 3D d'une construction de la tyrosine kinase KIT contenant le domaine catalytique et la région juxtamembranaire le précédant (Mol et al., 2004). Il était déjà admis que la juxtamembrane de KIT jouait un rôle d'autoinhibition, mais le mécanisme précis par lequel KIT était maintenu dans un état autoinhibé restait inconnu. Dans cette structure, on peut voir la juxtamembrane interagir avec le domaine catalytique, en prenant la position de la *A-loop* dans sa configuration active, incitant cette dernière à adopter la conformation inactive *DFG out* (code PDB : 1T45).

### 1.2.3.d Régulation par la « P-loop »

La *P-loop* est une zone très flexible, qui ne se stabilise qu'en présence de l'ATP (Taylor et Radzio-Andzelm, 1997). Quelques structures cristallographiques non ligandées présentent une distorsion de la *P-loop*, de telle façon que l'accès au site de liaison de l'ATP est bloqué. C'est le cas de la kinase CK2 $\alpha$  où la *P-loop* (en particulier une arginine) descend dans le site de liaison de l'ATP (code PDB : 3FWQ) (Raaf et al., 2009). Les auteurs définiront cette conformation comme inactive, cependant de nombreuses structures non ligandées offrent un site de fixation de l'ATP tout à fait accessible, comme le suggère la kinase LCK (code PDB : 3LCK) (Yamaguchi et Hendrickson, 1996). Inversement, dans certaines structures cocrystallisées avec un inhibiteur prenant la place de l'ATP, la *P-loop* vient englober une partie de l'inhibiteur pour augmenter la surface d'interaction de ce dernier avec la protéine. C'est par exemple le cas de la kinase ABL avec l'imatinib (code PDB : 1FPU) (Schindler et

al., 2000). Le rôle régulateur de la *P-loop* est donc toujours débattu. Il en est de même pour l'ensemble du lobe N-terminal. Un grand nombre de publications relatent en effet l'aspect « ouvert » de certains domaines catalytiques cristallisés, où un déplacement du lobe N-terminal par rapport au lobe C-terminal est observé (Zheng et al., 1993b). Cette « ouverture » est souvent perçue comme une conformation inactive or on commence à penser que cet état « ouvert » ou « fermé » ne constitue pas un critère suffisant pour déterminer si la kinase est catalytiquement active ou non (Taylor et Kornev, 2011).

### 1.2.3.e État de phosphorylation

Plusieurs des modes de régulation décrits ci-dessus peuvent s'appliquer à une même kinase, comme démontré avec CDK2 (Andzelm et al., 1995). Mais le mode de régulation le plus répandu est probablement celui de la phosphorylation de la kinase, soit par elle-même, soit par une autre kinase (Figure 13) (Nolen et al., 2004). Les acides aminés phosphorylés peuvent se situer en dehors du domaine catalytique, comme c'est le cas pour le récepteur de l'éphrine, EPHB2, dont la phosphorylation de deux tyrosines sur la région juxtamembranaire précédant le domaine catalytique est nécessaire pour l'activation de la kinase (code PDB : 1JPA) (Wybenga-Groot et al., 2001). Les résidus phosphorylés peuvent également être localisés sur des segments particuliers du domaine catalytique, et principalement sur la *A-loop* (Johnson et al., 1996; Nolen et al., 2004). En effet, l'état de phosphorylation de la kinase sur sa *A-loop* participe à la rendre constitutivement active, en stabilisant la boucle pour la fixation du substrat. Un exemple est celui du récepteur INSR, possédant trois tyrosines phosphorylables sur cette boucle, la dernière étant déterminante pour l'activité de la kinase (code PDB : 1GAG) (Hubbard, 1997).

Cependant, certaines kinases dont beaucoup de récepteurs à activité tyrosine kinase (RTKs) présentent une activité catalytique dite basale, lorsqu'elles ne sont pas phosphorylées. Cette activité augmente au fur et à mesure des phosphorylations (Hubbard, 2004). Au contraire, quelques kinases telles que PKC ne sont pas du tout contrôlées par une phosphorylation sur leur *A-loop* (Ranganathan et al., 2007). Il convient de noter également que les phosphorylations ne sont pas systématiquement activatrices, mais peuvent être inhibitrices, comme c'est le cas pour la tyrosine kinase SRC qui possède une tyrosine phosphorylable située sur son extension C-terminale. Lorsque cette tyrosine est phosphorylée par la kinase CSK, elle contribue à l'amarrage du domaine SH2 de SRC le long du domaine catalytique, provoquant l'adoption de la conformation inactive *C-helix out* (Xu et al., 1997).



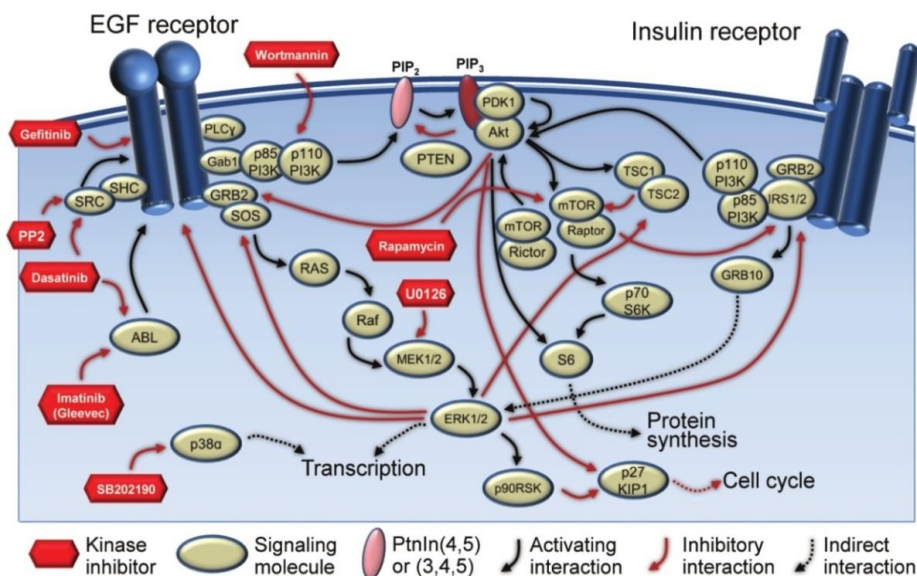


Figure 13 : Diagramme schématique laissant entrevoir la complexité des réseaux de signalisation dans lesquels les kinases réagissent en cascades, en aval des récepteurs EGFR et INSR (Carlson et White, 2011).

#### 1.2.4 Implication dans les maladies et premiers médicaments

Certaines kinases peuvent se retrouver constitutivement activées, dès lors que leurs mécanismes de régulation échouent, qu'elles sont surexprimées ou bien qu'elles subissent des mutations (Engh et Bossemeyer, 2002). Ces événements peuvent avoir des conséquences graves telles que l'apparition de cancers (Dhillon et al., 2007; Santarius et al., 2010; Liegl-Atzwanger et al., 2010; Lugo et al., 1990) mais également des maladies immunologiques, cardiovasculaires, neurologiques, métaboliques ou encore infectieuses (Bennett, 2006; Wang, 2007; Pandya et al., 2005; Huang et al., 2007). Par exemple, lors de l'étude des séquences codantes de kinases dans 210 formes humaines de cancer, environ 120 kinases ont présenté des mutations susceptibles d'être responsables de la maladie (Greenman et al., 2007).

La leucémie myéloïde chronique (CML), un cancer du sang, fut la première maladie grave à être traitée efficacement par un inhibiteur de kinase. Alors que la majorité des médicaments anticancéreux et des chimiothérapies étaient peu efficaces car hautement toxiques vis-à-vis des cellules saines et présentaient des effets secondaires indésirables (nausées, pertes de cheveux, etc), l'imatinib s'est révélée être la première preuve d'efficacité dans le concept de « thérapeutique ciblée » (Blay et al., 2005). C'est en 1990, plus exactement, que l'activité tyrosine kinase constitutive de la protéine de fusion BCR-ABL,

indispensable à la prolifération de la CML, a été mise en évidence (Lugo et al., 1990). Ces résultats n'ont pas tardé à intéresser les chercheurs de Ciba-Geigy (devenant Novartis six ans plus tard) qui se sont mis à développer une multitude de petites molécules chimiques susceptibles d'inhiber l'activité catalytique de la protéine. Deux années plus tard, et après de nombreux échecs de molécules trop toxiques, ils parviennent à optimiser une molécule dérivée du squelette phenylaminopyrimidine : l'imatinib. Cette molécule se révèle capable de stopper efficacement la leucémie, tout en présentant très peu d'effets secondaires. Les premiers tests sur les souris se révélant positifs, l'imatinib a poursuivi son chemin jusqu'aux tests cliniques, où les premiers résultats ont été fulgurants. Dès les premières semaines, les cliniciens ont pu constater un abaissement massif des cellules leucémiques dans le sang de tous les malades. L'autorisation de mise sur le marché (AMM) a été obtenue en 2001 aux États-Unis (Figure 14) et un an plus tard en Europe. L'activité remarquable de l'imatinib a fait envisager son utilisation dans d'autres tumeurs présentant une protéine à activité tyrosine kinase anormalement exprimée. C'est ainsi que l'imatinib s'est révélée efficace contre des tumeurs du tube digestif appelées GIST (acronyme anglais de *Gastro Intestinal Stromal Tumors*) en inhibant la kinase KIT (Liegl-Atzwanger et al., 2010).

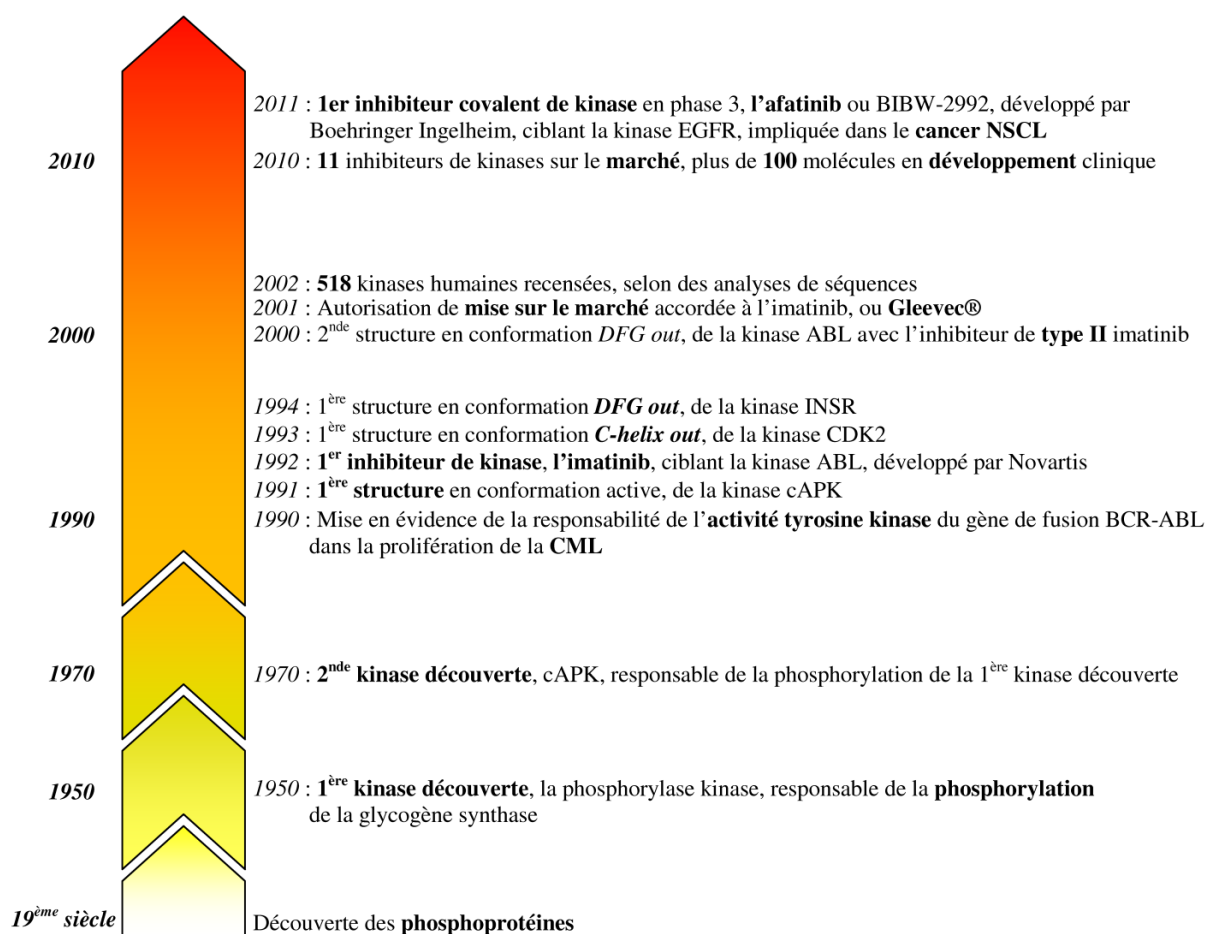


Figure 14 : Chronologie des découvertes importantes sur les protéines kinases et le développement d'inhibiteurs.

Depuis sa mise sur le marché, l'imatinib, vendue sous le nom de Gleevec®, contribue largement au chiffre d'affaire de Novartis, placée au 3<sup>ème</sup> rang mondial des industries pharmaceutiques en 2010 (Classement des dix Leaders de l'industrie pharmaceutique au niveau mondial en 2010 | Pharmactua, 2010).

L'histoire de l'imatinib s'est rapidement propagée dans les autres industries pharmaceutiques, qui se sont alors intéressées de plus près à la thérapeutique ciblée, via l'inhibition de protéines kinases anormalement activées dans des maladies (Figure 14). À ce jour, dans le cas d'une administration par voie orale, une quinzaine de kinases sont les cibles principales de médicaments en vente sur le marché et plus de 140 kinases sont les cibles de molécules situées entre la phase préclinique et à la phase clinique 3 (informations extraites de la base de données Pharmaprojects) (Pharmaceutical Research News & Drug Development Pipeline Information, 2010). Les inhibiteurs de kinases administrables oralement et actuellement commercialisés sont au nombre de 11 : l'imatinib ou Gleevec®, Fasudil®, gefitinib ou Iressa®, erlotinib ou Tarceva®, sorafenib ou Nexavar®, sunitinib ou Sutent®, nilotinib ou Tassigna®, dasatinib ou Sprycel®, lapatinib ou Tykerb®, rapamycine ou Rapamune® et son analogue temsirolimus ou Torisel® (Tableau 2) (Johnson, 2009; Pharmaceutical Research News & Drug Development Pipeline Information, 2010). Bien que très peu de molécules ne soient encore commercialisées, du fait de l'intérêt tardif apporté aux kinases par rapport aux autres cibles, plus d'une centaine d'inhibiteurs sont actuellement dans les phases de développement clinique 1 à 3 (Figure 15) (Pharmaceutical Research News & Drug Development Pipeline Information, 2010).

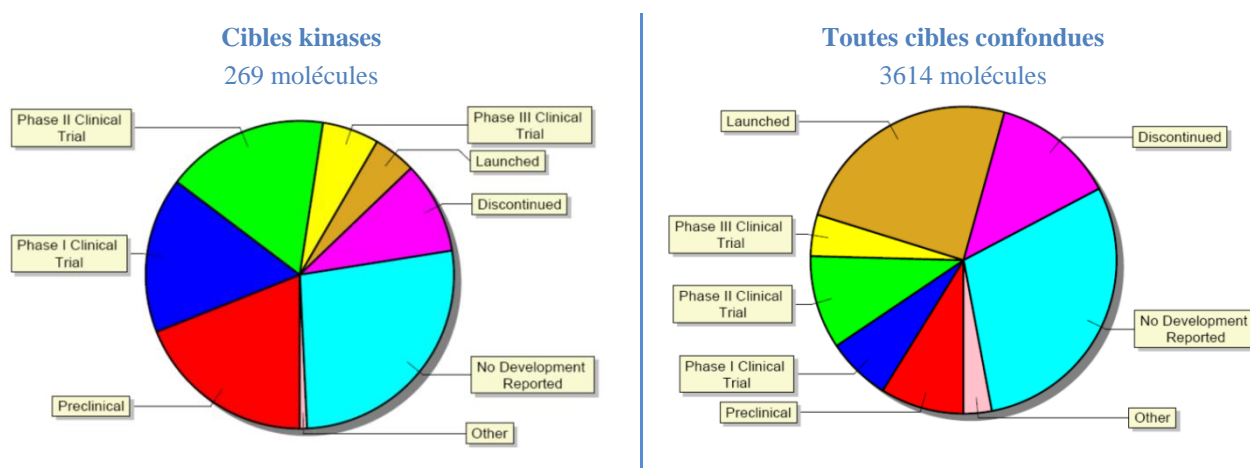
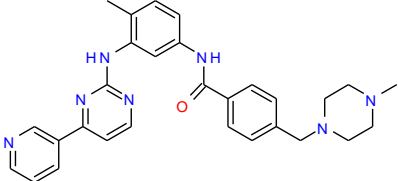
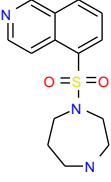
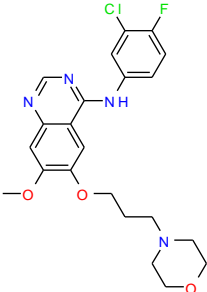
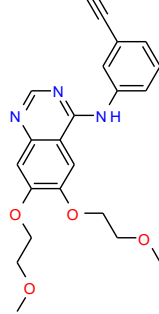
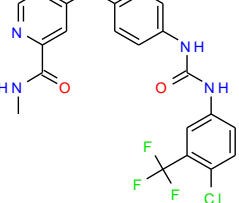
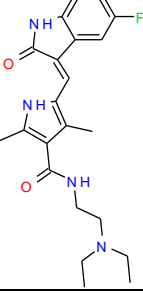


Figure 15 : Distribution du nombre de molécules administrables oralement et ciblant des protéines kinases dans les différentes étapes de développement d'un médicament (extrait de Pharmaprojects). Pour comparaison, la distribution pour tout type de cible confondu est également présentée.

Structure chimique du médicament	Nom générique Nom commercial (Développeur) Année d'approbation aux USA	Indication	Type d'inhibiteur	Cibles kinases principales (Code PDB de la structure avec le médicament)	Conformation visée
	Imatinib Gleevec® (Novartis) 2001	Leucémie (CML), Tumeurs du tube digestif (GIST)	II	ABL (1FPU, 2HYY), KIT (1T46), PDGFR $\alpha$	<i>DFG out</i>
	Fasudil Fasudil® ou Eiril (Asahi Kasei) 2001	Vasospasme cérébral	I	ROCK (2ESM)	<i>Active</i>
	Gefitinib Iressa® (Astra Zeneca) 2004	Cancer du poumon (NSCL)	I ½	EGFR (2ITO)	<i>Active</i>
	Erlotinib Tarceva® (OSI Pharmaceuticals, Genentech et Roche) 2004	Carcinomes du poumon (NSCL) et pancréatiques	I ½	EGFR (1M17)	<i>Active</i>
	Sorafenib Nexavar® (Bayer) 2006	Carcinome des cellules rénales	II	KDR, FLT4, FLT3, RET, BRAF (1UWH)	<i>DFG out</i>
	Sunitinib Sutent® (Pfizer) 2006	Carcinome des cellules rénales, Tumeurs du tube digestif (GIST)	I	KIT (3G0F), PDGFR $\alpha$ , FLT3, KDR, FLT1, ITK (3MIY)	<i>Active</i>

Suite du Tableau 2 sur la page suivante

	<p>Nilotinib Tasigna® (Novartis) 2007</p>	<p>Leucémie (CML)</p>	<p>II</p>	<p>ABL (3CS9), KIT, PDGFR<math>\alpha</math></p>	<p><i>DFG out</i></p>
	<p>Dasatinib Sprycel® (BMS) 2007</p>	<p>Leucémie (CML)</p>	<p>I ½</p>	<p>ABL (2GQG), KIT, PDGFR<math>\alpha</math>, SRC (3G5D), EPHA2</p>	<p><i>Active</i></p>
	<p>Lapatinib Tykerb® (GSK) 2007</p>	<p>Cancer du sein</p>	<p>I ½</p>	<p>EGFR (1XKK), HER2</p>	<p><i>C-helix out</i></p>
	<p>rapamycine Rapamune® (Wyeth) 1999</p> <p>temsirolimus (analogue) Torisel® (Wyeth) 2007</p>	<p>Carcinome des cellules rénales, Transplantation d'organes (immunosuppresseur)</p>	<p>III</p>	<p>mTOR, via la FK-Binding Protéine (2FAP)</p>	<p>N.D.</p>

Tableau 2 : Les 11 médicaments administrables oralement et ciblant des protéines kinases, disponibles à ce jour sur le marché. N.D.=non déterminé.

### 1.2.5 Limitations rencontrées par les inhibiteurs

Avant le succès attribué à l'imatinib, les industries pharmaceutiques étaient très réticentes à l'idée de travailler sur la famille des protéines kinases. En effet, les inhibiteurs doivent entrer en compétition avec l'ATP, présent à des concentrations élevées dans la cellule (de 1 à 5mM), pour accéder au site actif (Knight et Shokat, 2005). Cette compétition était susceptible de limiter considérablement l'efficacité de l'inhibiteur. De plus, les similarités de séquences entre les protéines kinases laissaient présager d'importants problèmes de sélectivité de la part de l'inhibiteur, risquant de provoquer des effets secondaires graves (Karaman et al., 2008).

Alors que Novartis faisait évoluer l'imatinib dans les phases de développement clinique, les autres industries pharmaceutiques ont commencé à intégrer des « projets kinases », avec l'espoir que les problèmes précités seraient plus facilement contournés que prévu. Ainsi, toute une panoplie de molécules destinées à prendre la place de l'ATP dans le site actif de la cible kinase, et ainsi empêcher l'activité catalytique de la kinase de fonctionner, ont été développées (Johnson, 2009). En peu de temps, la famille des protéines kinases est devenue la seconde famille la plus travaillée après les GPCRs dans l'industrie pharmaceutique, comptant pour 20 à 30 % des projets menés (Cohen, 2002a).

Des mesures d'affinité de l'ATP ( $K_{M, ATP}$ ) sur une majorité des 518 kinases ont indiqué quelles étaient les kinases à cibler préférentiellement et quelles étaient celles à éviter (Knight et Shokat, 2005). En effet, l'équation de Cheng-Prusoff traduit le fait que plus le  $K_{M, ATP}$  est petit (haute affinité de l'ATP pour la kinase), plus l' $IC_{50}$  augmente (faible efficacité de l'inhibiteur). Il ressort de ces études qu'une majorité de kinases ont un  $K_{M, ATP}$  compris entre 10 et 100 $\mu$ M donc il faut s'attendre à ce que les inhibiteurs soient actifs dans la cellule à des concentrations devant être 10 à 100 fois supérieures au  $K_d$ .

$$IC_{50} = K_d * (1 + [ATP] / K_{M, ATP}) \quad (\text{Équation de Cheng-Prusoff})$$

$IC_{50}$  = concentration d'inhibiteur nécessaire pour inhiber 50% de l'activité d'une kinase.

$K_d$  = constante de dissociation, permet d'évaluer l'affinité de l'inhibiteur pour sa cible.

$K_{M, ATP}$  = affinité de l'ATP pour sa cible.

Les études de sélectivité sont, quant à elles, plus difficiles à envisager. En 2005, Ambit BioSciences est l'une des premières sociétés de biotechnologie à proposer un test d'affinité en compétition site-ATP-dépendante sur une majorité des 518 kinases, afin d'évaluer le profil de sélectivité de n'importe quel inhibiteur donné (Fabian et al., 2005). L'application de ce criblage pour 38 inhibiteurs, actuellement situés dans les phases de développement clinique ou sur le marché, révèle des profils opposés entre eux (Figure 16) (Karaman et al., 2008).

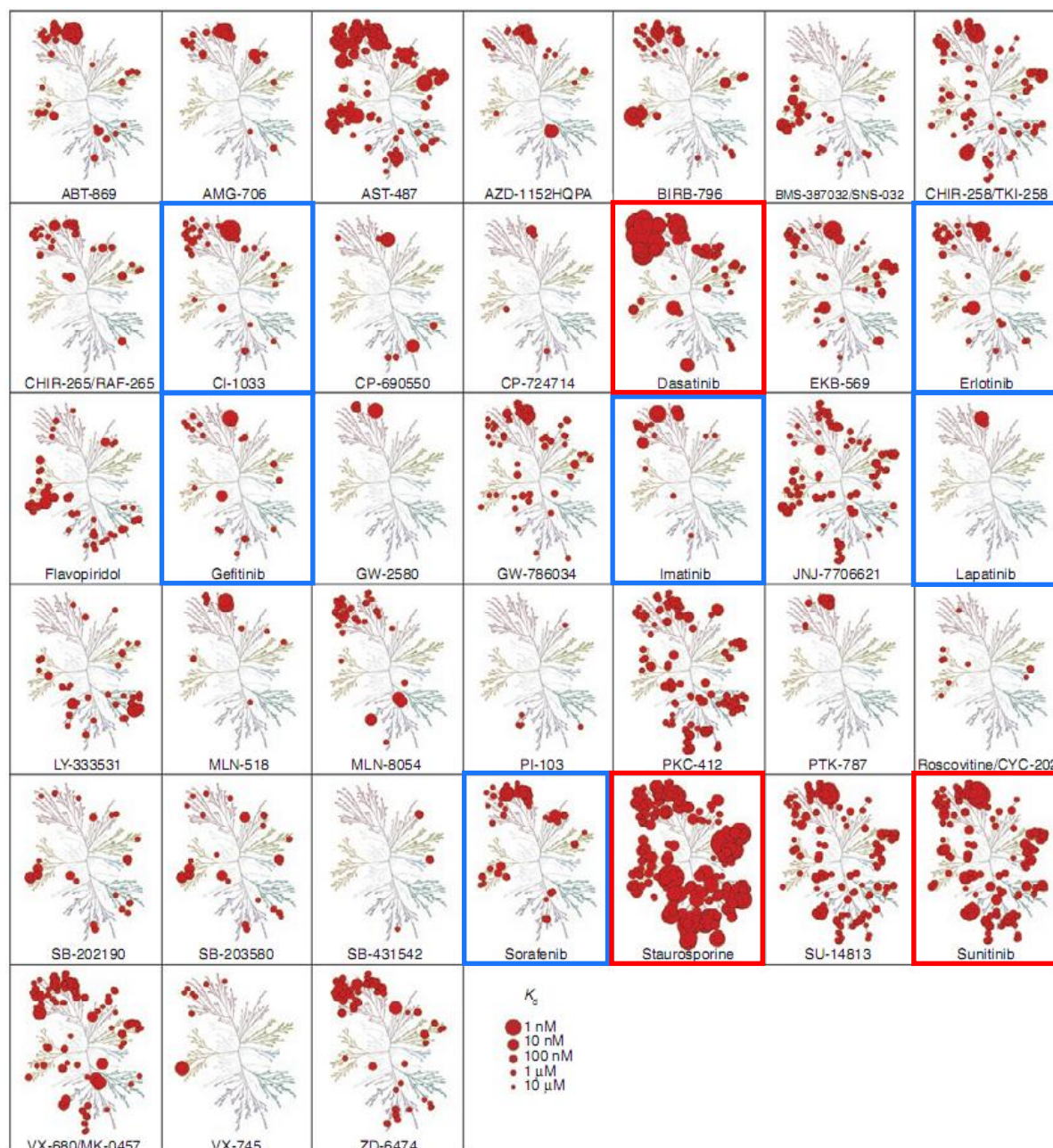


Figure 16 : Profil de sélectivité de 38 inhibiteurs de kinases se trouvant de la phase clinique 1 au marché (Karaman et al., 2008). Les sphères rouges indiquent les kinases humaines inhibées dans l'arbre phylogénétique, avec un  $K_d$  allant de 10  $\mu$ M à 1 nM selon le diamètre croissant de la sphère. Les arbres encadrés en bleu et rouge correspondent respectivement à de bons et mauvais profils de sélectivité de plusieurs inhibiteurs cités dans ce manuscrit. Par exemple, la staurosporine a un très mauvais profil de sélectivité tandis que l'imatinib et le lapatinib sont hautement sélectifs, limitant les problèmes de toxicité.

Alors que l'imatinib inhibe majoritairement cinq kinases ( $K_d < 1\mu\text{M}$ ) avec quelques faibles touches ( $K_d < 10\mu\text{M}$ ) parmi une dizaine d'autres kinases, beaucoup d'inhibiteurs sont au contraire très peu sélectifs (Figure 16). L'exemple le plus évocateur est celui de la staurosporine, qui inhibe quasiment toutes les protéines kinases humaines avec des affinités inférieures à 100nM. Cette extraordinaire non-spécificité (Tanramluk et al., 2009) exclut la staurosporine d'un quelconque usage clinique mais en fait un outil de recherche très intéressant (Zhang et al., 2004). Parmi les autres molécules montrant des profils de sélectivité inquiétants, on citera AST-487, CHIR258, Dasatinib, PKC-412, SU-14813, Sunitinib, VX-680 et ZD-6474 (Figure 16).

Cette étude révèle la difficulté réelle de faire des inhibiteurs sélectifs d'une kinase ou d'un sous-groupe de kinases. Pourtant, la première tentative d'un tel inhibiteur, l'imatinib, a été couronnée de succès ! Les industries pharmaceutiques ne tardent pas à se demander pourquoi certains inhibiteurs sont bien plus sélectifs que d'autres. Le mode d'action de l'imatinib était supposé connu puisqu'il a été démontré qu'il entrait en compétition avec l'ATP, donc probablement en se plaçant dans le site de fixation de l'ATP lorsque la kinase adoptait une conformation *Active*. Mais contre toute attente, la résolution de la structure cristallographique du domaine catalytique d'ABL avec l'imatinib révèle une toute autre vérité : l'imatinib reconnaît une conformation inactive de la kinase ABL, la conformation *DFG out* (code PDB : 1FPU ; voir Figure 12 à la page 30) (Schindler et al., 2000). La kinase ainsi bloquée ne peut plus adopter la conformation *Active* et ne fixe alors plus l'ATP. L'imatinib est donc un inhibiteur allostérique tout en étant ATP-compétitif.

### 1.2.6 Les stratégies d'inhibition adoptées

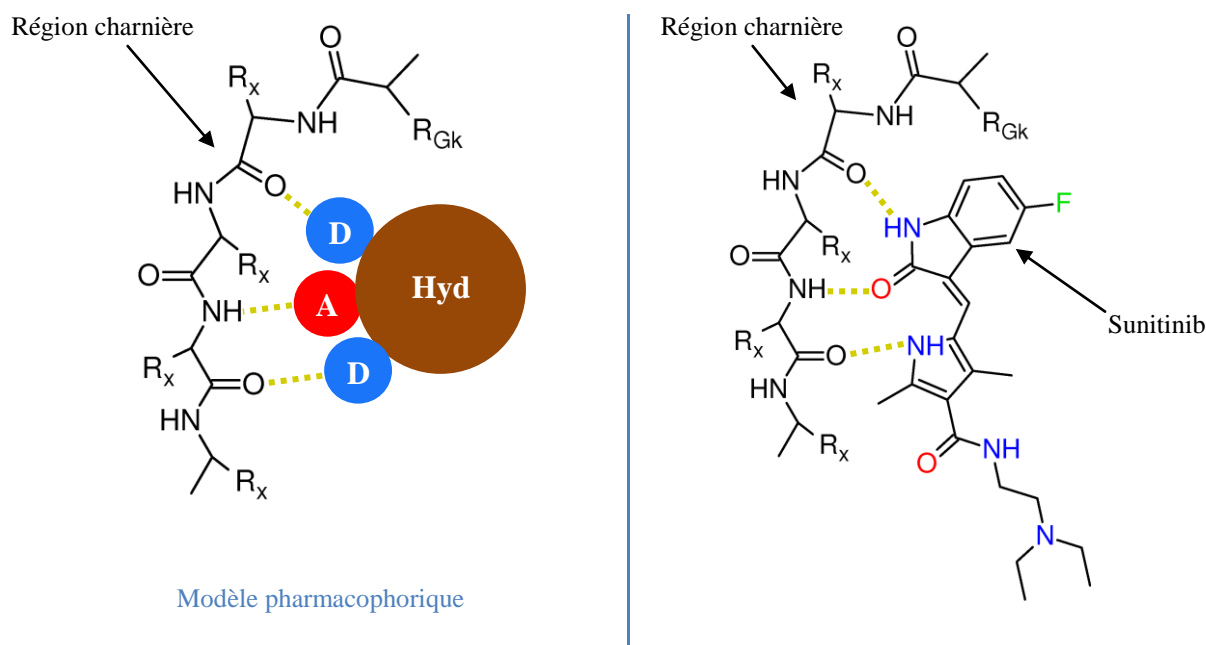
Au fur et à mesure de la résolution de complexes kinase/inhibiteur, on distingue majoritairement trois types d'inhibiteurs ATP-compétitifs, selon les poches additionnelles dans lesquelles ils s'étendent : les inhibiteurs de type I, les inhibiteurs de type I ½ et les inhibiteurs de type II - l'imatinib étant de type II (Zuccotto et al., 2010; Zhang et al., 2009). Les inhibiteurs non ATP-compétitifs sont dits de type III. Ces différents types sont présentés dans les paragraphes suivants. Un autre type d'inhibiteur est celui des inhibiteurs covalents, c'est-à-dire formant une liaison covalente avec un résidu du site catalytique. Ces derniers, qui font l'objet d'une grande partie de ce travail de thèse, seront présentés plus longuement par la suite.



### 1.2.6.a Les inhibiteurs de Type I

Alors que le vrai mode d'action de l'imatinib n'était pas connu, les premiers inhibiteurs de kinases qui ont été développés avaient pour objectif de mimer l'ATP afin d'entrer en compétition avec celui-ci et se mettre dans le site catalytique de la kinase en conformation *Active*. Il s'agit alors des inhibiteurs de type I. Toutes les molécules de type I développées s'inspirent de la partie adénine de l'ATP, c'est-à-dire un centre aromatique, faisant une interaction favorable  $\pi$ - $\pi$  avec les résidus hydrophobes du brin  $\beta$ 2, dans le lobe C-terminal, et du brin  $\beta$ 6, dans le lobe N-terminal, auquel s'ajoute une combinaison parmi deux donneurs et un accepteur de liaisons hydrogènes, pouvant interagir en trois points de la région charnière (Figure 17).

La majorité des inhibiteurs actuellement développés sont de type I, en parti dû au fait qu'initialement, seules les formes actives phosphorylées de kinases étaient utilisées dans les essais biochimiques d'identification de touches et d'optimisation des meilleures molécules (Zuccotto et al., 2010). Un exemple d'inhibiteur de kinases de type I actuellement sur le marché est le sunitinib (Figure 17), ciblant un sous-groupe de tyrosine kinases (voir Tableau 2 à la page 37 ; code PDB : 3MIY (Kutach et al., 2010)).



Il est assez difficile d'obtenir un inhibiteur sélectif de type I, comme le montre le profil de sélectivité du sunitinib (voir Figure 16 à la page 39), du fait de la conservation élevée des acides aminés dans le site catalytique, parmi les 518 kinases. Par ailleurs, cette stratégie est tellement exploitée qu'il devient difficile de trouver des molécules brevetables. Cependant, du point de vue pharmacologique, cette stratégie serait avantageuse dans le cas des maladies où l'équilibre conformationnel est déplacé vers la forme *Active* de la kinase.

### 1.2.6.b Les inhibiteurs de Type I ½

Les inhibiteurs de type I ½ prennent avantage de la petite taille d'un résidu situé juste avant la région charnière pour atteindre une poche additionnelle, hydrophobe, leur permettant d'effectuer des interactions supplémentaires avec la cible kinase (Figure 18) (Zuccotto et al., 2010). Ce résidu, communément appelé *gatekeeper*, est très peu conservé parmi les kinases, ce qui offre une belle opportunité de sélectivité. L'exemple d'un tel inhibiteur est l'erlotinib (Figure 18), qui est assez spécifique et très affiné à sa cible EGFR, qu'il inhibe dans la conformation *Active* (voir Tableau 2 à la page 37 ; voir Figure 16 à la page 39 ; code PDB : 1M17 (Stamos et al., 2002)).

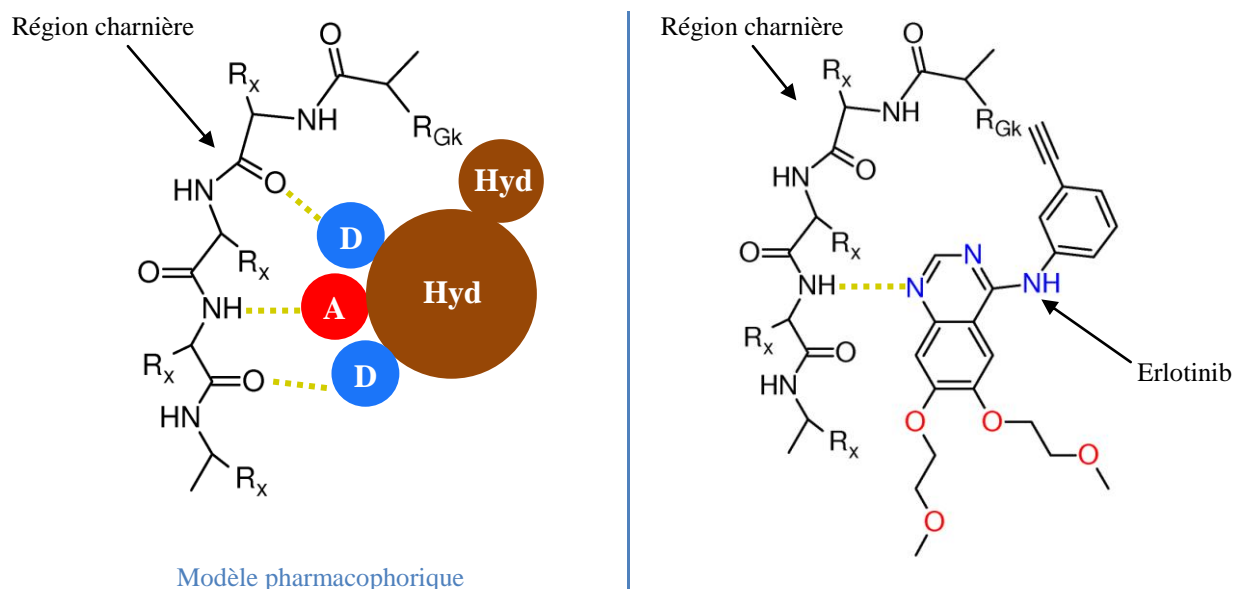


Figure 18 : Les inhibiteurs de type I ½ : modèle pharmacophorique dans le cas d'une kinase en conformation *Active* et exemple de l'erlotinib. Seule la région charnière de la kinase ciblée est représentée. Gk=*gatekeeper*. D/A=donneur/accepteur de liaison hydrogène. Hyd=poche hydrophobe. Les liaisons hydrogènes sont représentées en pointillés. Dans le cas de l'erlotinib, seule une liaison hydrogène - celle faisant intervenir l'accepteur de liaison hydrogène sur le ligand - est formée entre la région charnière et le ligand. Grâce à la petite taille du résidu *gatekeeper*, l'erlotinib peut accéder à une petite poche hydrophobe additionnelle, via son groupe alcyne.

Lorsqu'une kinase est à la fois capable d'adopter la conformation inactive *C-helix out* et présente un résidu *gatekeeper* de petite taille, le profil de sélectivité peut s'avérer optimum. En effet, en conformation inactive *C-helix out*, le recul de l'hélice augmente la taille de la poche additionnelle hydrophobe, offrant ainsi une nouvelle région d'interaction accessible par les inhibiteurs (Figure 19) (Zuccotto et al., 2010). Le plus bel exemple d'un tel inhibiteur allostérique est celui du lapatinib (Figure 19), hautement sélectif pour sa cible EGFR en occupant pleinement le site de liaison de l'ATP et la poche hydrophobe (voir Tableau 2 à la page 37 ; voir Figure 16 à la page 39 ; code PDB : 1XKK (Wood et al., 2004)).

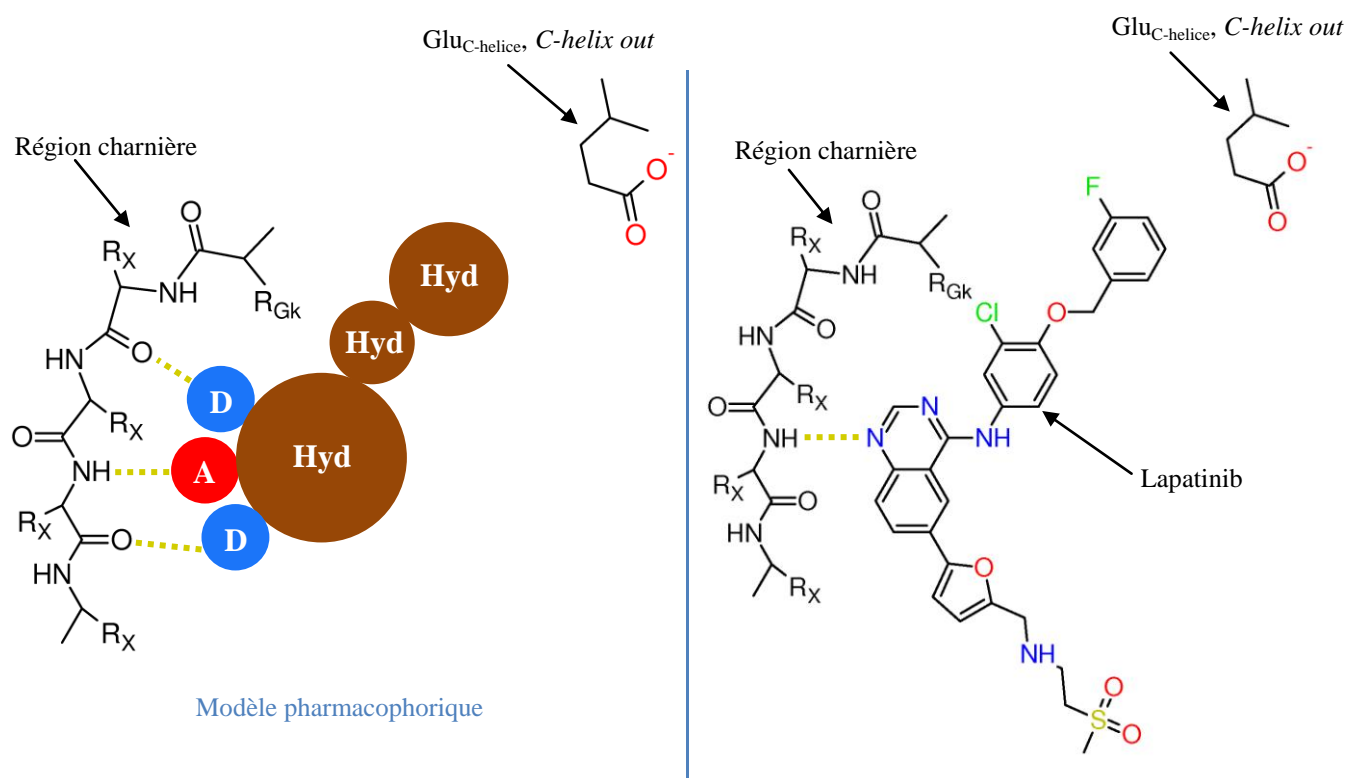
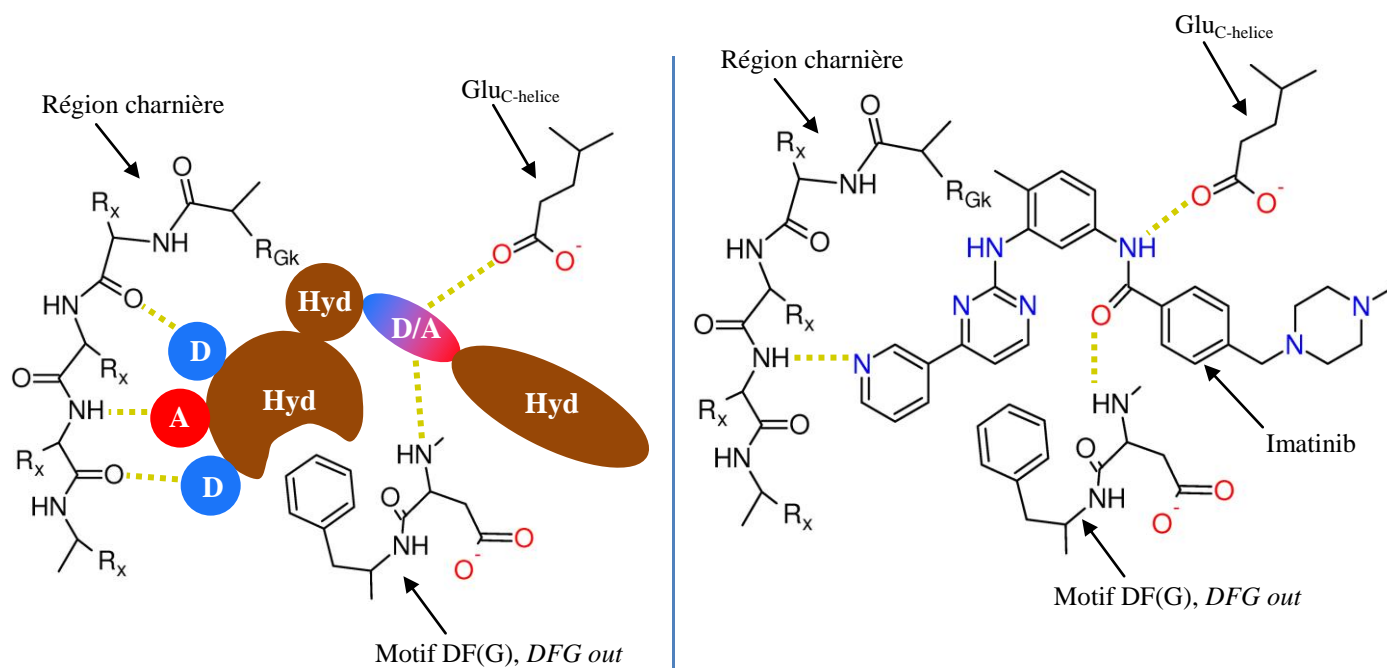


Figure 19 : Les inhibiteurs de type I ½ : modèle pharmacophorique dans le cas d'une kinase en conformation inactive *C-helix out* et exemple du lapatinib. Seule la région charnière et le résidu catalytique  $\text{Glu}_{\text{C-helice}}$  de la kinase ciblée sont représentés. Gk=*gatekeeper*. D/A=donneur/accepteur de liaison hydrogène. Hyd=poche hydrophobe. Les liaisons hydrogènes sont représentées en pointillés. Dans le cas du lapatinib, seule une liaison hydrogène - celle faisant intervenir l'accepteur de liaison hydrogène sur le ligand - est formée entre la région charnière et le ligand. Grâce à la petite taille du résidu *gatekeeper* et au déplacement de la C-hélice, le lapatinib peut accéder à deux poches hydrophobes successives, via les groupes chlorophényle et fluorophényle.

### 1.2.6.c Les inhibiteurs de Type II

Le premier inhibiteur de kinases ayant atteint le marché, l'imatinib, est un inhibiteur dit allostérique de type II, du fait qu'il reconnaît une conformation différente de la forme *Active* : la conformation inactive *DFG out* (voir Tableau 2 à la page 37). Le large déplacement du motif DFG change considérablement la morphologie du site catalytique, en donnant accès à une grande poche hydrophobe, exploitée par les inhibiteurs de type II (Figure 20). Il est

cependant nécessaire que le *gatekeeper* soit un petit résidu pour permettre à l'inhibiteur de s'étendre du site de liaison de l'ATP jusqu'à la poche hydrophobe. Contrairement aux autres types d'inhibiteurs, les inhibiteurs de type II n'occupent pas pleinement la poche de liaison de l'ATP du fait de la présence de la phénylalanine du motif DFG qui réduit le site (Figure 20). La découverte de ce nouveau mode de fixation s'est faite par hasard (Schindler et al., 2000), mais aux vues du succès de l'imatinib (Figure 20 ; voir Tableau 2 à la page 37 ; voir Figure 16 à la page 39), les inhibiteurs de type II devraient prendre le pas sur les inhibiteurs de type I.

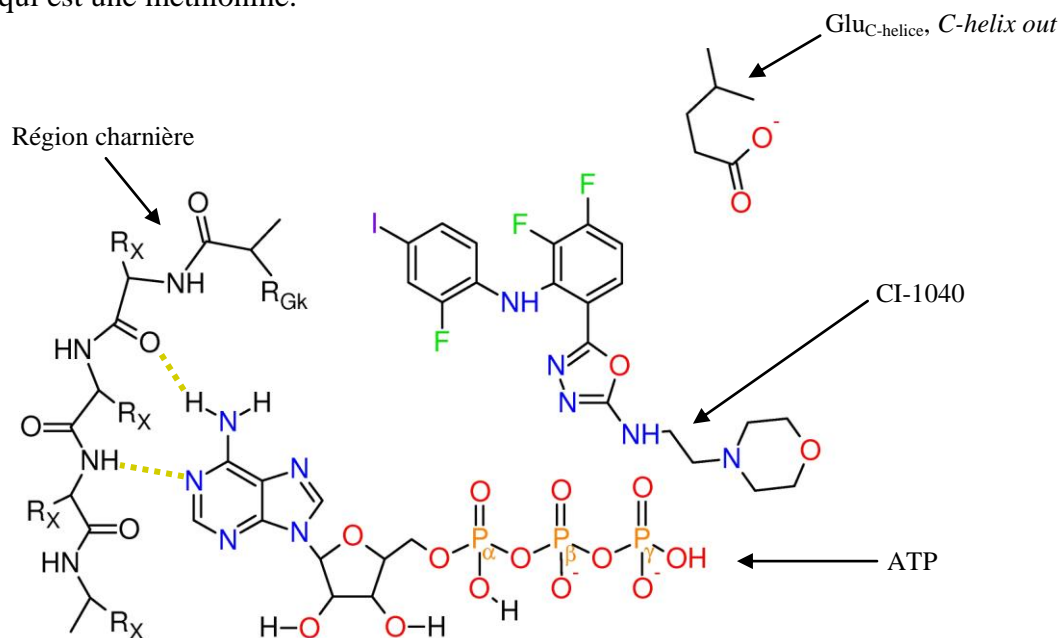


**Figure 20 :** Les inhibiteurs de type II : modèle pharmacophorique dans le cas d'une kinase en conformation inactive *DFG out* et exemple de l'imatinib. Seule la région charnière, le résidu catalytique  $\text{Glu}_{\text{C-helice}}$  et une partie du motif DFG de la kinase ciblée sont représentés. Gk=*gatekeeper*. D/A=donneur/accepteur de liaison hydrogène. Hyd=poche hydrophobe. Le déplacement du motif DFG vient réduire le site de liaison de l'ATP, empêchant l'ATP de se fixer, mais donne accès à une large extension de la poche. Les liaisons hydrogènes sont représentées en pointillés. Dans le cas de l'imatinib, seule une liaison hydrogène - celle faisant intervenir l'accepteur de liaison hydrogène sur le ligand - est formée entre la région charnière et le ligand. Grâce à la petite taille du résidu *gatekeeper*, et au déplacement du motif DFG et du reste de la *A-loop*, l'imatinib peut accéder à toute une extension hydrophobe qui s'ouvre sur le solvant.

### 1.2.6.d Les inhibiteurs de Type III

Récemment, des campagnes de criblage à haut débit ont permis de découvrir des molécules capables d'inhiber l'activité d'une cible kinase sans pour autant empêcher l'ATP de se lier (Bogoyevitch et Fairlie, 2007; Kirkland et McInnes, 2009). Il s'agit des inhibiteurs de type III. Ces molécules se positionneraient soit sur le site de fixation du substrat, soit dans une poche allostérique bloquant la *A-loop* en conformation inactive, et donc gêneraient indirectement la fixation du substrat. Pour ce dernier cas, on citera l'exemple du CI-1040, un inhibiteur des kinases MAP2K1 et MAP2K2 (Figure 21). En effet, grâce aux structures cristallographiques des deux complexes (codes PDB : 1S9I, 1S9J), on peut voir les deux

kinases en conformation inactive *C-helix out* avec l'ATP dans son site et l'inhibiteur fixé dans une poche adjacente délimitée par la C-hélice et la *A-loop* (Ohren et al., 2004). L'inhibiteur peut ainsi s'étendre jusque dans la poche hydrophobe qui a été décrite pour les inhibiteurs de type I ½, tels que le lapatinib (voir Figure 19 à la page 43), sans être gêné par le résidu *gatekeeper*, qui est une méthionine.



**Figure 21 :** Exemple d'inhibiteur de type III ou non ATP-compétitif: le CI-1040. Seule la région charnière et le résidu catalytique acide glutamique situé sur la C-hélice de la kinase ciblée sont représentés. Gk=*gatekeeper*. Les liaisons hydrogènes sont représentées en pointillés. Le CI-1040 est fixé dans une poche adjacente au site de liaison de l'ATP quand la kinase est en conformation inactive *C-helix out*.

Quelques inhibiteurs capables de se lier directement au site du substrat, tels que ON01910 et sa cible PLK1 (Gumireddy et al., 2005), ont également été découverts mais aucune structure cristallographique n'est actuellement disponible.

Les seuls inhibiteurs de type III se trouvant sur le marché sont la rapamycine et son analogue le temsirolimus (voir Tableau 2 à la page 37). La rapamycine, approuvée par la FDA en 1999, a été développée pour son efficacité clinique en tant qu'immunosuppresseur, sans que son mode d'action ne soit clairement identifié. Finalement, la rapamycine se révèle être un inhibiteur de la kinase mTOR, appartenant à un réseau de régulation impliqué dans le contrôle de la progression du cycle cellulaire ainsi que dans la prolifération des lymphocytes T (Dumont et Su, 1995). C'est plus exactement via son association avec la protéine FKBP, que la rapamycine est capable de se lier à une poche hydrophobe du domaine FRB de la kinase mTOR (code PDB : 2FAP) (Choi et al., 1996). Il a parallèlement été démontré que ce domaine FRB est un régulateur positif de l'activité catalytique de la kinase mTOR (Vilella-Bach et al., 1999).

### 1.2.6.e Conclusion sur les divers types d'inhibiteurs

Actuellement, ce sont près de la moitié des inhibiteurs de kinases se trouvant sur le marché qui bloquent leur cible dans une conformation inactive, c'est-à-dire tous les inhibiteurs de type II et quelques inhibiteurs de type I  $\frac{1}{2}$  (voir Tableau 2 à la page 37) (Alton et Lunney, 2008). La résolution de la structure cristallographique du complexe kinase/inhibiteur est l'une des seules méthodes permettant de déterminer si l'inhibiteur agit sur une conformation active ou bien inactive de la kinase. C'est au fur et à mesure de la résolution des complexes kinase/inhibiteur que le rapprochement s'est fait entre inhibiteur allostérique - donc reconnaissant une conformation inactive de la kinase - et bon profil de sélectivité. L'efficacité de cette stratégie peut s'expliquer par le fait qu'il y ait moins de compétition avec l'ATP, qui ne se lie qu'à la forme active de la kinase. De plus, pour expliquer le meilleur profil de sélectivité des inhibiteurs allostériques par rapport aux inhibiteurs de type I, il est envisagé que la conformation inactive ciblée ne puisse pas être adoptée par l'ensemble des kinases humaines. D'autre part, chaque conformation inactive donne accès à un site catalytique élargi et moins conservé en termes de résidus parmi les kinases (Alton et Lunney, 2008).

Cependant, quel que soit le type d'inhibiteur, ils peuvent tous être impactés par des phénomènes de résistance développés par leur cible kinase, induisant une inefficacité du traitement chez les patients (Azam et al., 2008; Gazdar, 2009). Plus exactement, la kinase va développer une mutation qui peut avoir divers effets selon l'endroit où elle se trouve. Elle peut avoir lieu dans le site actif pour augmenter l'affinité de l'ATP, exemplifié avec la kinase EGFR et la mutation T790M (Yun et al., 2008), résistante au gefitinib et à l'erlotinib. La mutation peut également apparaître au niveau du résidu *gatekeeper* pour empêcher l'accès à la poche additionnelle hydrophobe qui est adjacente à ce résidu et qui est exploitée par l'inhibiteur, exemplifié avec la kinase ABL et la mutation T315I résistante à l'imatinib (Mian et al., 2009). La mutation peut encore se trouver en dehors du site de liaison de l'ATP, au niveau de la *A-loop* pour déplacer l'équilibre conformationnel vers la forme *Active*, exemplifié par EGFR et la mutation L858R qui, combinée à la mutation du résidu *gatekeeper*, induit une résistance à de nombreux inhibiteurs d'EGFR (Yeh et al., 2011).

Une nouvelle génération d'inhibiteurs de kinases est actuellement développée, qui est capable de rester efficace face aux résistances développées par leur cible et tend à remplacer les précédentes générations d'inhibiteurs. Il s'agit des inhibiteurs de type covalent, qui sont détaillés plus longuement dans le paragraphe suivant.

## 1.2.7 Émergence des inhibiteurs de type covalent

### 1.2.7.a Les inhibiteurs covalents en phase clinique

L'afatinib ou BIBW-2992, développé par Boehringer Ingelheim, est le premier inhibiteur covalent de kinase à avoir atteint la phase 3 de développement clinique (voir Tableau 3 à la page 50) (Subramaniam et Hwang, 2011). Ce dérivé de quinazoline est un analogue des molécules gefitinib et erlotinib (voir Tableau 2 à la page 37), qui sont actuellement sur le marché dans le but d'inhiber la kinase EGFR, surexprimée dans le cancer du poumon non à petites cellules (cancer NSCLC). La différence majeure entre l'afatinib et ces deux molécules commercialisées est que la première molécule possède un groupement électrophile amide  $\alpha,\beta$ -insaturée lui permettant de réagir de manière covalente et irréversible avec la cystéine 797 située à la fin de la région charnière d'EGFR, en sortie du site de liaison de l'ATP (code PDB : 2J5E ; Figure 22) (Blair et al., 2007).

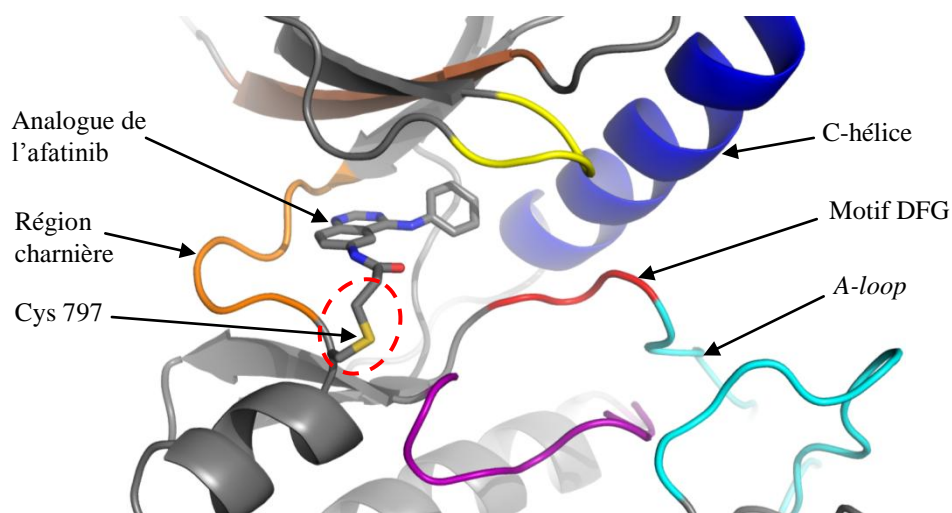


Figure 22 : Représentation du mode de liaison covalent d'un analogue de l'inhibiteur afatinib via la cystéine 797 (entourée en pointillés), située à la fin de la région charnière dans le site catalytique de la kinase EGFR, en conformation Active (code PDB : 2J5E).

L'entrée en phase de développement clinique d'une telle molécule vient de la nécessité de trouver une nouvelle génération d'inhibiteurs de la cible EGFR, suite à l'apparition de résistances développées par EGFR chez certains patients, réduisant l'efficacité du gefitinib et de l'erlotinib (Yun et al., 2008; Yeh et al., 2011). Contrairement à ce que l'on pensait, la mutation T790M du résidu *gatekeeper*, à l'origine de 50% des résistances observées, n'empêche pas les inhibiteurs de se fixer (la méthionine est certes encombrante mais flexible ; code PDB : 2JIV) mais elle augmente l'affinité de l'ATP pour EGFR, rendant la compétition plus forte (Yun et al., 2008). L'afatinib subit la même compétition, à la différence qu'une fois

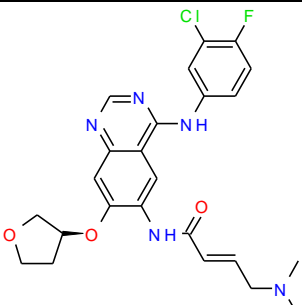
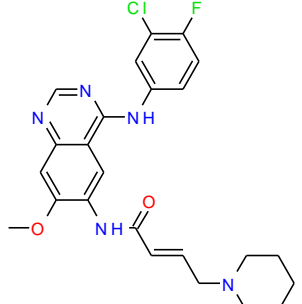
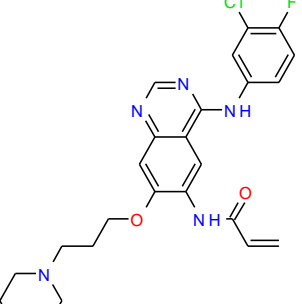
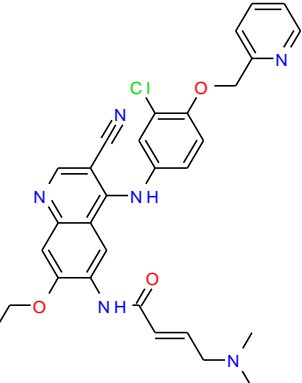
que cette molécule se met dans le site actif, la liaison covalente la bloque définitivement dans le site, expliquant son efficacité par rapport à ces deux prédécesseurs.

Le succès de l'afatinib a rapidement suscité le développement de molécules analogues par d'autres industries pharmaceutiques. Actuellement, cinq analogues covalents de l'afatinib sont dans les phases de développement clinique 1 à 3 : dacomitinib, canertinib, neratinib, pelitinib et AV-412 (Tableau 3) (Subramaniam et Hwang, 2011; Janne et al., 2011; Allen et al., 2003; Sequist et al., 2010; Laheru et al., 2008). Toutes ces molécules sont majoritairement testées sur des patients ayant un cancer du poumon de type NSCLC présentant ou non des formes résistantes d'EGFR. Certaines de ces molécules sont également testées sur des patientes présentant un cancer du sein, pour leur potentiel d'inhibition de la kinase anormalement surexprimée HER2, qui est très proche d'EGFR dans l'arbre phylogénétique (Allen et al., 2003; Burstein et al., 2010).

La seule molécule covalente se distinguant de l'afatinib et de ses analogues est la molécule PCI-32765, présentant un squelette pyrazolopyrimidine mais contenant également un groupement électrophile amide  $\alpha,\beta$ -insaturée (Tableau 3) (Honigberg et al., 2010). Bien que cette molécule réagisse de manière covalente avec une cystéine située à la même position que celle ciblée par les inhibiteurs covalents actuels d'EGFR, cette molécule inhibe plus particulièrement la tyrosine kinase de Bruton (BTK ; code PDB : 3GEN) (Pan et al., 2007), qui joue un rôle important dans la maturation des lymphocytes B. L'efficacité de l'inhibiteur PCI-32765 est actuellement testée en phase 1 de développement clinique sur divers cancers lymphatiques et d'autres maladies auto-immunes (Honigberg et al., 2010).

Cette cystéine que partagent EGFR, HER2 et BTK est commune à 8 autres kinases : BMX, TEC, TXK, ITK, BLK, HER4, JAK3 et MAP2K7, soit 7 tyrosine kinases (groupe TK) et une kinase du groupe phylogénétique STE. Certaines de ces kinases sont également inhibées par les molécules précédemment décrites mais cela ne justifie pas un arrêt de leur développement clinique (Hur et al., 2008). D'autre part, la présence du groupement réactif n'a pas encore suscité de problèmes de toxicité où l'inhibiteur réagirait avec toutes les kinases possédant une cystéine en surface. En fait, le profil de sélectivité de l'afatinib parmi les kinases humaines est comparable à celui de ses analogues non covalents (comparer sur la Figure 16, à la page 39, le profil du canertinib ou CI-1033, qui est un analogue covalent de l'afatinib, avec celui de l'erlotinib).



Structure chimique	Noms génériques (Développeur) Phase de développement clinique	Indication	Type d'inhibiteur	Cibles kinases principales (code PDB de la structure avec le candidat-médicament ou un analogue)	Conformation visée
	Afatinib BIBW-2992 (Boehringer Ingelheim) Phase 3	Cancer NSCLC, cancer du sein, cancer de la prostate, cancer des ovaires	I ½ + covalent	EGFR (2J5E, 2J5F, 2QLQ, 2QQ7, 2HWP), HER2	Active
	Dacomitinib PF-299804 (Pfizer) Phase 3	Cancer NSCLC, autres cancers	I ½ + covalent	EGFR (2J5E, ...), HER2	Active
	Canertinib CI-1033 (ParkeDavis/Pfizer) Phase 2	Cancer NSCLC, cancer du sein	I ½ + covalent	EGFR (2J5E, ...), HER2, HER4	Active
	Neratinib HKI-272 (Wieth/Pfizer) Phase 3	Cancer NSCLC, cancer du sein	I ½ + covalent	EGFR (2JIV), HER2	C-helix out

Suite du Tableau 3 sur la page suivante

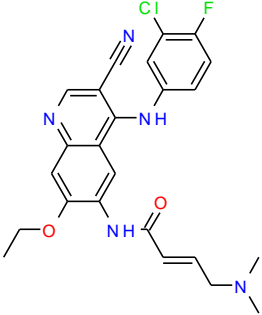
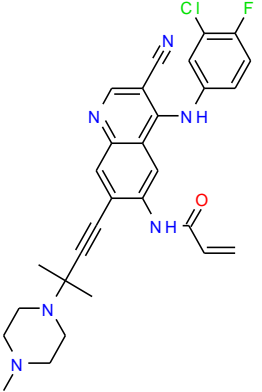
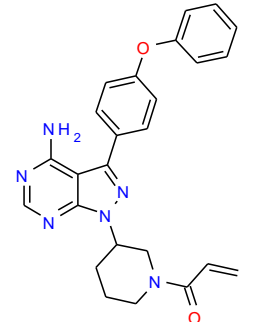
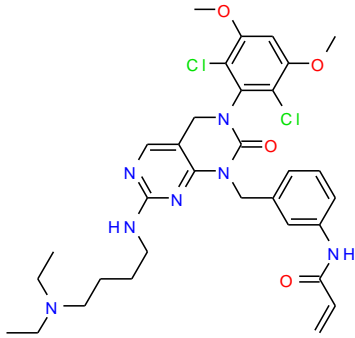
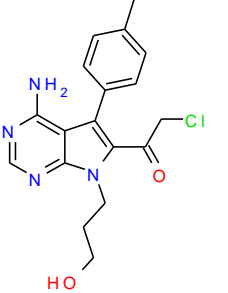
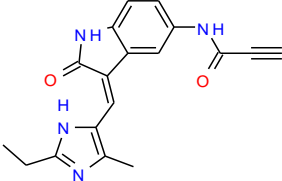
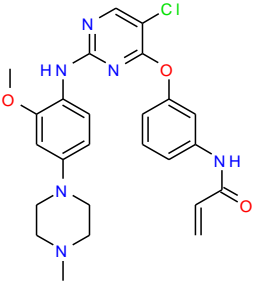
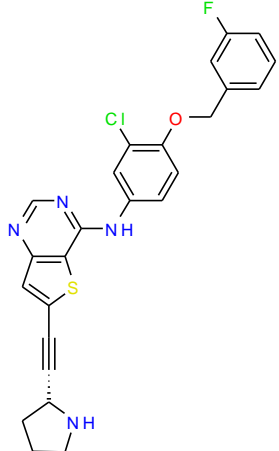
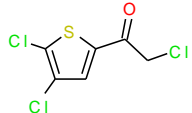
	<p>Pelitinib EKB-569 (Wieth/Pfizer) Phase 2</p>	<p>Cancer NSCLC</p>	<p>I ½ + covalent</p>	<p>EGFR</p>	<p>Active</p>
	<p>AV-412, MP-412 (Aveo) Phase 1</p>	<p>Cancer NSCLC</p>	<p>I ½ + covalent</p>	<p>EGFR, HER2</p>	<p>Active</p>
	<p>PCI-32765 (Pharmacyclics) Phase 1</p>	<p>Cancers lymphatiques, Maladies auto- immunes</p>	<p>I ½ + covalent</p>	<p>BTK (3GEN)</p>	<p>C-helix out</p>

Tableau 3 : Liste des candidats-médicaments se liant de manière covalente à leur cible kinase via une cystéine située à la fin de la région charnière. La majeure partie de ces inhibiteurs sont dérivés de quinazoline ou de quinoline. On remarquera qu'un inhibiteur covalent peut également être de type I ou I ½ selon la conformation adoptée par la kinase et les poches additionnelles dans lesquelles il s'étend.

### 1.2.7.b Les inhibiteurs covalents dans les publications

Les inhibiteurs covalents de kinases actuellement en développement clinique se restreignent quasiment à un seul squelette (quinazoline), une seule cystéine visée (à la fin de la région charnière) et une seule cible inhibée (EGFR). Cependant, la littérature révèle plusieurs autres tentatives prometteuses. Actuellement, trois autres cystéines situées dans la poche de liaison de l'ATP sont les cibles d'inhibiteurs synthétiques et semi-synthétiques covalents (Tableau 4).

Structure chimique	Nom générique	Type d'inhibiteur	Cibles kinases principales (code PDB de la structure avec l'inhibiteur ou un analogue)	Position de la cystéine	Conformation visée
	Fin-1	I ½ + covalent	FGFR1..4	Début de la P-loop	Active
	Cmk	I ½ + covalent	RSK1..4	Milieu du brin β2	Active
	Su16	I + covalent	NEK2	Milieu du brin β2	N.D.
	WZ-4002	I + covalent	EGFR (3IKA)	Fin de la région charnière	Active
	Dérivé de thienopyrimidine	I ½ + covalent	EGFR (2R4B), HER2, HER4	Fin de la région charnière	C-helix out
	Dérivé de thienylhalométhyl-cétone	III + covalent	GSK3b	Avant le motif DFG	N.D.

Suite du Tableau 4 sur la page suivante

	Hypothemycine (naturel)	I + covalent	MAP2K1, ERK2 (3C9W), KDR et plus de 10 autres kinases	Avant le motif DFG	Active
	Radicol A	I + covalent	KDR, VEGFR3, FLT3 et d'autres kinases (3C9W)	Avant le motif DFG	Active
	LL-Z1640-2 (naturel)	I + covalent	TAK1, MAP2K1, PDGFR $\alpha$ et d'autres kinases (3C9W)	Avant le motif DFG	Active
	LL-783277 (naturel)	I + covalent	MAP2K1, KDR et d'autres kinases (3C9W)	Avant le motif DFG	Active
	Prostaglandine de type cyclopentenone (naturel)	III + covalent	I $\kappa$ B	A-loop	N.D.

Tableau 4 : Liste quasi-exhaustive des inhibiteurs covalents de kinases évoqués dans la littérature. Contrairement aux candidats-médicaments covalents de kinases, ces inhibiteurs ne visent pas que les kinases ayant une cystéine située à la fin de la région charnière. N.D.=non déterminé.

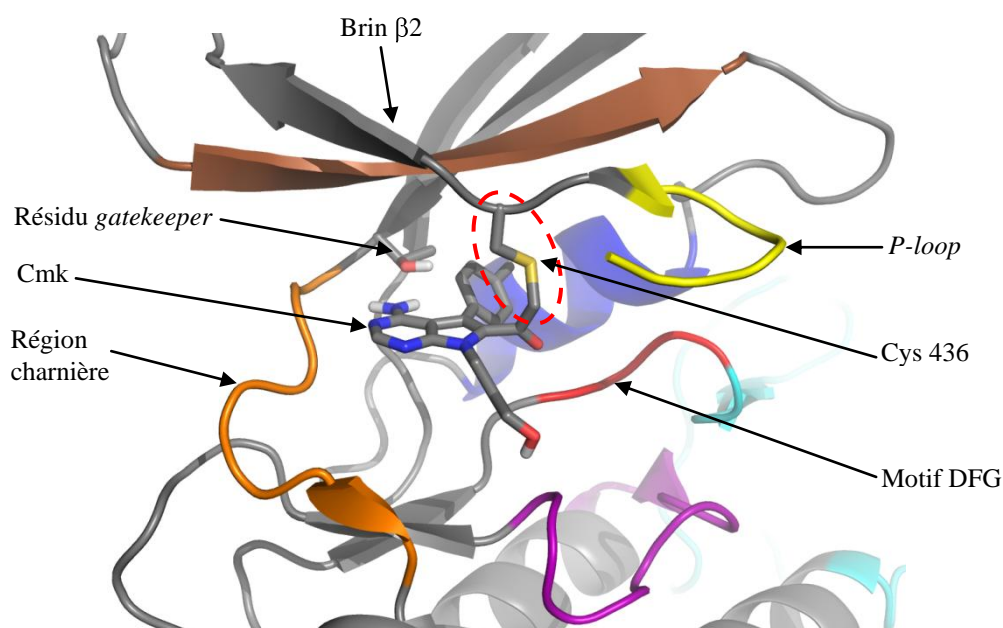
La première cystéine se situe au début de la *P-loop* et est commune à neuf kinases. L'inhibiteur fiin-1 réagissant avec cette cystéine est un dérivé d'une pyrimidine-2-amine et inhibe spécifiquement les quatre tyrosines kinases FGFRs (Figure 23 ; Tableau 4) (Zhou et al., 2010).



Figure 23 : Structure de FGFR1 (code PDB : 2FG1) et proposition d'un mode de liaison pour la molécule fiin-1, liée de manière covalente à la cystéine 488 (entourée en pointillés), située sur la *P-loop*. Par mesure de clarté, le brin  $\beta$ 1 a été supprimé.

Les kinases FGFRs étant impliquées dans divers cancers, plusieurs inhibiteurs réversibles de celles-ci sont actuellement en développement clinique. Les auteurs voient ce premier inhibiteur irréversible, fiin-1, comme une alternative au cas où les patients actuellement traités avec les inhibiteurs réversibles développeraient une forme mutée des FGFRs, au niveau du résidu *gatekeeper* (Zhou et al., 2010), comme observé pour EGFR (Yun et al., 2008).

La seconde cystéine se situe au milieu du brin  $\beta 2$  et est commune à 11 kinases. Afin de démontrer que cette cystéine pouvait jouer le rôle de filtre de sélectivité parmi les 518 kinases humaines, l'inhibiteur cmk, qui est un analogue de l'ATP avec un groupe électrophile chlorométhylcétone, a été développé (Figure 24 ; voir Tableau 4 à la page 52) (Cohen et al., 2005). Cet inhibiteur de type I  $\frac{1}{2}$  s'est révélé encore plus spécifique au sein des 11 kinases, en n'inhibant que les kinases RSKs et MSKs, grâce à un second filtre de sélectivité apporté par le petit *gatekeeper* thréonine, qui est substitué par une leucine ou une isoleucine dans les autres kinases du groupe.



**Figure 24 :** Modèle 3D par homologie de RSK2 humaine depuis RSK2 murine (code PDB : 2QR7) et proposition d'un mode de liaison pour la molécule cmk, liée de manière covalente à la cystéine 436 (entourée en pointillés), située au milieu du brin  $\beta 2$ . Par mesure de clarté, le brin  $\beta 1$  a été supprimé.

La troisième cystéine se trouve juste avant le motif DFG et la *A-loop*, et est commune à 48 kinases. Il s'agit probablement de la cystéine la plus ciblée par des inhibiteurs covalents de kinases. Ces inhibiteurs sont tous dérivés de l'hypothémicine, une substance naturelle de la classe des lactones d'acide résorcyclique (RALs) extraite d'un champignon (Figure 25 ; voir

Tableau 4 à la page 52). Alors que cette molécule était étudiée pour ses propriétés antimalariale et antifongique, elle s'est également révélée capable d'arrêter la croissance de cellules tumorales chez des souris (Tanaka et al., 1999). Des études plus poussées ont permis de mettre en évidence que l'hypothémicine et certains de ses analogues n'avaient pas pour seule cible la protéine chaperonne HSP90, mais également plusieurs protéines kinases telles que MAP2K1, TAK1, KDR et ERK2 (code PDB : 3C9W (Rastelli et al., 2008)) (Schirmer et al., 2006; Zhao et al., 1999; Ninomiya-Tsuji et al., 2003). Une propriété partagée par toutes ces kinases était la présence d'une cystéine conservée avant le motif DFG. Des analogues de l'hypothémicine ne possédant pas de cétone  $\alpha,\beta$ -insaturée, donc pas de groupement électrophile, se sont révélés inefficaces face aux kinases précitées, confirmant la présence et l'importance de la liaison covalente avec leurs cibles (Ninomiya-Tsuji et al., 2003; Jogireddy et al., 2009; Ohori et al., 2007; Hearn et al., 2007). Actuellement, aucun analogue covalent de l'hypothémicine n'est suffisamment sélectif parmi les 48 kinases pour être amené en développement clinique.

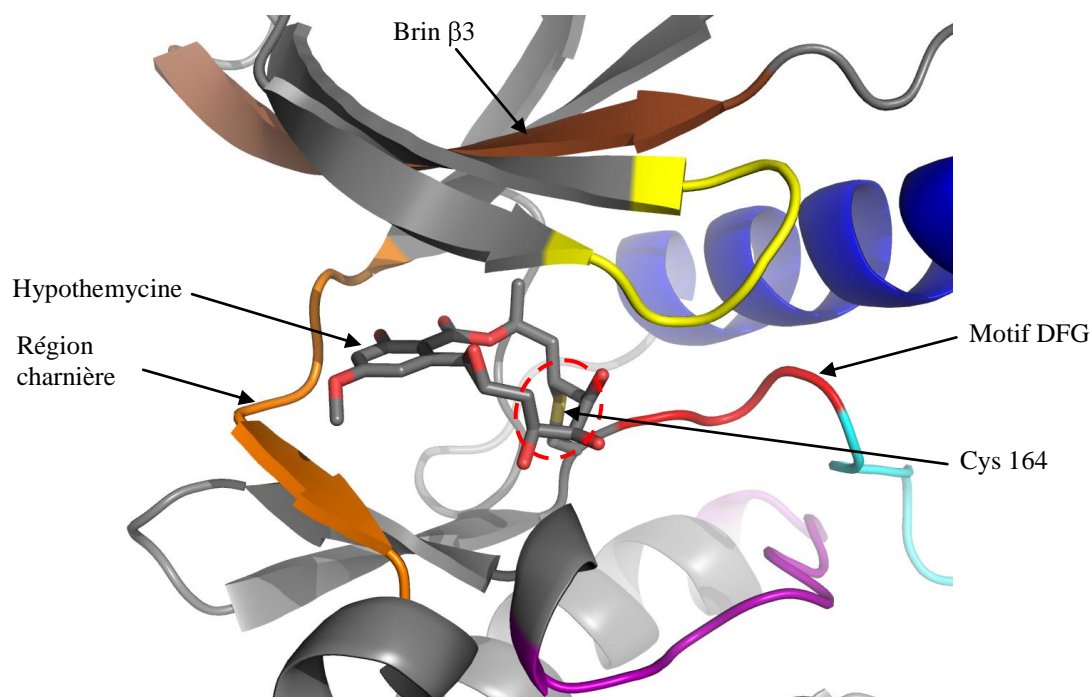


Figure 25 : La kinase ERK2 avec l'inhibiteur hypothémicine lié de manière covalente à la cystéine 164 (entourée en pointillés), située avant le motif DFG (code PDB : 3C9W).

Un autre inhibiteur covalent dérivé de thienylhalométhylcétone, donc très différent de l'hypothémicine, a été récemment rapporté (voir Tableau 4 à la page 52) (Perez et al., 2009). Cet inhibiteur a la particularité d'être de type III parce que non compétitif de l'ATP. Or les auteurs ont démontré qu'il réagissait de manière covalente avec la kinase GSK3 $\beta$ , via la

cystéine située avant le motif DFG donc dans le site de liaison de l'ATP. En effet, cet inhibiteur étant très petit (230 Da), il est probable qu'il parvienne à se loger entre la *gatekeeper* leucine et le motif DFG sans interférer avec l'ATP. Par contre, aucun profil de sélectivité de cet inhibiteur parmi les 47 autres kinases n'a encore été effectué.

La cystéine se trouvant à la fin de la région charnière chez EGFR est déjà la cible de plusieurs inhibiteurs covalents dérivés de quinazoline ou de quinoline et qui se trouvent en développement clinique (voir Tableau 3 à la page 50 ; voir Figure 22 à la page 47). De nouveaux archétypes structuraux ont également fait leur apparition, tels que WZ-4002, dérivé de pyrimidine-2-amine (code PDB : 3IKA) (Zhou et al., 2009) ainsi qu'un dérivé de thienopyrimidine (code PDB : 2R4B ; voir Tableau 4 à la page 52) (Wood et al., 2008). Ces deux molécules ont les mêmes profils d'inhibition et d'efficacité thérapeutique que les inhibiteurs covalents se trouvant dans les phases cliniques mais leur structure chimique pourrait leur apporter des propriétés physico-chimiques différentes. Les bons résultats précliniques du composé WZ-4002 laissent envisager une entrée prochaine en phase 1 (Zhou et al., 2009).

### 1.2.7.c Avantages des inhibiteurs covalents

Alors que la première génération d'inhibiteurs de kinases a été plutôt décevante car beaucoup se sont montrés peu sélectifs (Karaman et al., 2008) ou inefficaces face à l'apparition de résistances dans leur cible kinase (Yun et al., 2008), l'arrivée des inhibiteurs covalents apporte un nouveau souffle dans la recherche de traitements efficaces. En effet, la première qualité attribuée aux inhibiteurs covalents est leur capacité à rester efficaces face aux résistances développées par la cible, contrairement à leurs prédécesseurs (Yun et al., 2008). Une seconde qualité est que les inhibiteurs covalents peuvent rester compétitifs de l'ATP, même quand ce dernier est présent à forte concentration dans la cellule (Yun et al., 2008). Les inhibiteurs covalents sont ainsi très efficaces, et ces diverses qualités se justifient par un temps de résidence plus élevé dans le site de liaison de l'ATP, en comparaison des inhibiteurs non covalents (Copeland et al., 2006). En effet, ces derniers restent souvent peu de temps dans le site, du fait de changements conformationnels de la cible qui déstabilisent la liaison. À l'inverse, les inhibiteurs covalents, lorsqu'ils se lient de manière irréversible à leur cible, ont un temps de résidence qui est infini. La suppression de l'activité de la kinase est alors totale jusqu'à ce que la protéine soit dégradée et resynthétisée (Spicer et Rudman, 2010). Même

dans le cas où la liaison covalente est réversible, un cycle d'étapes de dissociation lente suivi de re-fixation covalente, du fait de la proximité au site de liaison, peut assurer une grande efficacité de la molécule. Cette qualité est très importante quand on passe aux tests *in vivo*, où la concentration et le temps d'exposition nécessaires pour l'efficacité sont cruciaux pour le développement futur de la molécule en tant que médicament. Cependant, la plupart des tests de criblage *in vitro* actuels ne détectent pas cette propriété car le temps d'exposition de la molécule à sa cible est souvent trop court avant que la mesure ne soit prise. De plus, la liaison covalente, générant rapidement une très bonne affinité entre la cible et l'inhibiteur, permet de passer plus de temps sur l'optimisation des propriétés pharmacocinétiques, dites ADMET, du futur médicament (Smith et al., 2009).

En plus de contribuer à son efficacité, le temps de résidence élevé de la molécule au sein de sa cible permet d'augmenter sa sélectivité et donc de diminuer les problèmes de toxicité, en limitant les temps d'interaction possibles avec d'autres cibles, d'autant plus que la dose administrée peut être réduite (Copeland et al., 2006; Smith et al., 2009). Parallèlement, le fait que la cystéine visée soit peu conservée parmi les kinases humaines représente un atout pour son bon profil de sélectivité (Cohen et al., 2005). D'ailleurs, la présence du groupement électrophile dans les inhibiteurs covalents actuellement développés en phases cliniques n'a pas soulevé de problèmes de sélectivité. Leur profil de toxicité est comparable à celui de la première génération d'inhibiteurs sur le marché et ils restent biodisponibles oralement (Spicer et Rudman, 2010).

La stratégie d'inhibition covalente, bien que récente dans son application à la famille des protéines kinases, possède tout un historique inspiré de processus biologiques naturels et qui est détaillé dans le sous-chapitre suivant.



## 1.3 Réaction covalente en chimie et en biologie

### 1.3.1 Réaction entre un électrophile et un nucléophile

La liaison covalente résulte de la mise en commun de deux électrons par deux atomes. Une interaction non covalente entre deux éléments se transforme en liaison covalente dès lors que le premier élément est suffisamment électrophile et le second suffisamment nucléophile. Une liaison covalente est souvent admise comme étant irréversible, or ça n'est pas systématiquement le cas (Potashman et Duggan, 2009). Pour savoir si la réaction est irréversible, une simple mesure de la constante de dissociation ( $K_d$ ) ne suffit pas. Il faut soit intégrer un temps d'exposition long et/ou des lavages, ou bien utiliser des techniques telles que la spectrométrie de masse ou la résolution de structures 3D.

Les protéines et autres molécules biologiques possèdent de nombreux sites nucléophiles, comme les thiols et les amines (Meares, 2008). La cystéine est l'acide aminé le plus nucléophile, mais d'autres résidus sont également capables de réagir avec un électrophile, comme la lysine, l'histidine, la sérine ou la thréonine (Figure 26).

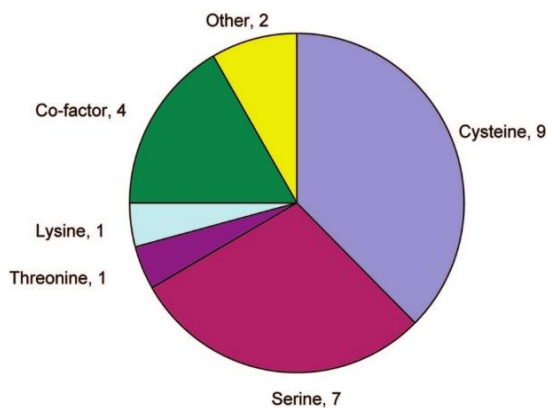


Figure 26 : Distribution des résidus modifiés de manière covalente par des médicaments, la cystéine étant le résidu le plus modifié (Potashman et Duggan, 2009).

Le lien entre la réactivité de la cystéine et ses fonctions biologiques a été établi depuis longtemps. En plus de son activité catalytique observée chez certaines protéases, oxydoréductases et acyltransférases, ce résidu peut également être sujet à des modifications post-traductionnelles oxydatives telles que la sulfénation (SOH), la sulfination (SO<sub>2</sub>H), la glutathionylation et la formation de ponts disulfures (Giles et al., 2003; Weerapana et al., 2010). Toutes ces fonctions font de la cystéine un véritable régulateur des protéines.

L'addition de Michaël et la substitution nucléophile sont deux réactions possibles en milieu biologique pour la formation d'une liaison covalente entre deux entités (Potashman et Duggan, 2009).

Dans l'addition de Michaël, le groupe électrophile est un système conjugué (cétone, aldéhyde, nitriles, dérivés carboxyliques  $\alpha,\beta$ -insaturés ; voir Tableau 5 à la page 61). Dans le cas d'une cétone  $\alpha,\beta$ -insaturée, la délocalisation des électrons  $\pi$  fait apparaître une charge partielle négative sur l'atome d'oxygène du carbonyle et une charge partielle positive sur le carbone en position  $\beta$  (Figure 27). Ce carbone  $\beta$  électrophile peut subir l'attaque d'un nucléophile par le doublet libre de sa HOMO (orbitale moléculaire liante occupée de plus haute énergie, dans l'approximation des orbitales frontières de FUKUI). L'angle d'attaque favorable du nucléophile est de l'ordre de  $107^\circ$ , dans le plan perpendiculaire au plan de la cétone  $\alpha,\beta$ -insaturée, selon les travaux de Bürgi-Dunitz (Burgi et al., 1973). En effet, cette position permet un meilleur recouvrement de la HOMO du nucléophile avec l'orbitale p du carbone  $C_\beta$  de contribution majoritaire dans la LUMO (orbitale moléculaire antiliante  $\pi^*$  vacante de plus basse énergie) du système énone, et limite l'interaction répulsive avec l'orbitale p du carbone  $C_\alpha$ . Ainsi, une liaison  $\sigma$  se forme entre le carbone  $C_\beta$  et le nucléophile (Figure 27). Cependant, les substituants présents sur le carbone  $C_\beta$  peuvent, par effet stérique, provoquer un décalage latéral de la position d'attaque du nucléophile.

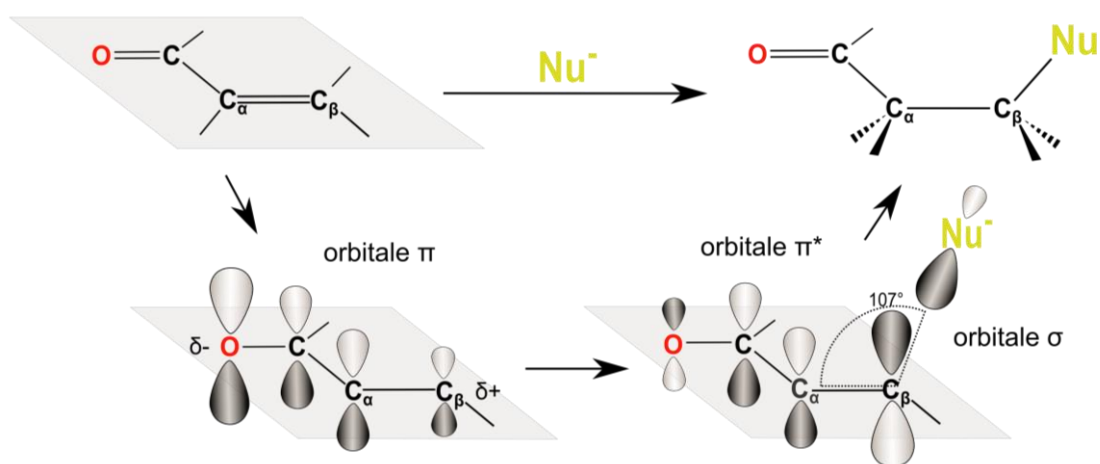


Figure 27 : Représentation de Lewis et représentation orbitaire de l'addition de Michaël sur une cétone  $\alpha,\beta$ -insaturée.

Lors de la substitution nucléophile de type  $S_N2$ , l'attaque par le nucléophile est simultanée au départ du groupe partant. Dans le cas d'une chlorométhylcétone, l'effet électroattracteur du chlore, auquel s'ajoute celui du carbonyle, rend le carbone du groupe

chlorométhyle très électrophile. Le recouvrement de l'orbitale contenant le doublet libre du nucléophile avec l'orbitale antiliante  $\sigma^*$  de la liaison carbone-chlore se fait suivant l'axe de cette liaison, et du côté opposé au groupe partant, le chlore. L'inversion de configuration des substituants, dite effet « parapluie », au cours de la réaction implique un état de transition trigonale du carbone électrophile (Figure 28).

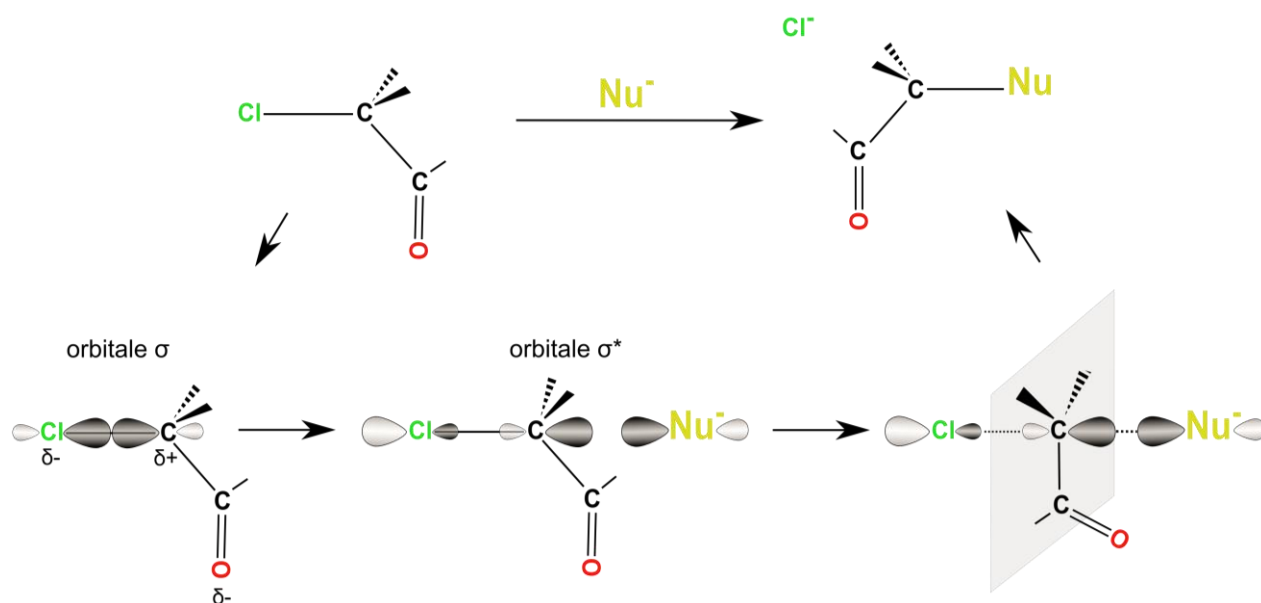
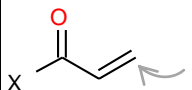
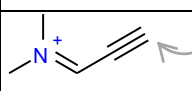
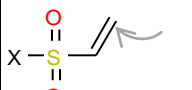
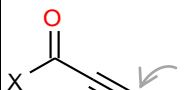
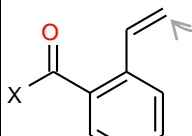
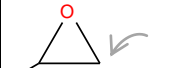
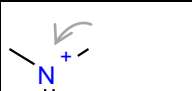
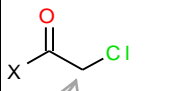
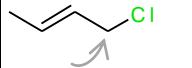
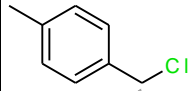


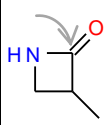
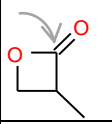
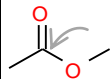
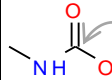
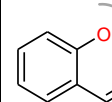
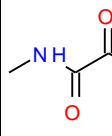
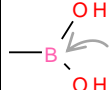
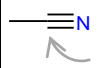
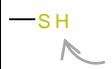
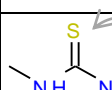
Figure 28 : Représentation de Lewis et représentation orbitaire de la substitution nucléophile de type  $S_N2$  sur une chlorométhylcétone.

La réactivité du groupement électrophile peut varier selon son environnement. Par exemple, dans le cas du développement de sondes covalentes de l'enzyme HCA II, seule la sonde avec l'époxyde, en contraste avec les autres fonctionnalités électrophiles testées, a montré le profil attendu, soit une bonne réactivité et spécificité pour la cible (Chen et al., 2003). L'environnement protéique aussi bien que l'environnement intra-ligand peuvent influencer la réactivité en contribuant à la polarisation de l'atome électrophile (Wood et al., 2008; Vicik et al., 2006).

De même, la réactivité de toute cystéine est influencée par le pH du milieu ainsi que son environnement électrostatique (Britto et al., 2002). À l'état libre, le pKa de la cystéine est de 8. Au sein des protéines, la cystéine subit l'influence de son environnement et peut présenter une forme ionisée. Par exemple, les cystéines les plus réactives ont souvent des acides aminés chargés positivement dans leur environnement. En effet, des charges positives tendent à stabiliser l'anion thiolate et donc diminuer le pKa de la cystéine, la rendant plus réactive. Au contraire, une cystéine ayant des acides aminés chargés négativement à proximité est moins réactive, même si elle se situe en surface.

Mécanisme	Nom du groupement réactif	Structure chimique	Médicaments ou candidats-médicaments covalents (Réversibilité)	Indication	Cible thérapeutique (code PBD de la structure avec le médicament ou un analogue) <i>Résidu modifié</i>
Addition de Michaël	Vinyl carbonyle <i>Cétone <math>\alpha,\beta</math>-insaturée (X=C)</i> ou <i>Amide <math>\alpha,\beta</math>-insaturée (X=N)</i> ou <i>Ester <math>\alpha,\beta</math>-insaturé (X=O)</i>		Gemcitabine (N.D.)	Cancer	Ribonucléoside diphosphate réductase <i>Cystéine</i>
			Floxuridine (Réversible)	Cancer	Thymidylate synthase <i>Cystéine</i>
			Afatinib (Irréversible)	Cancer	EGFR, HER2 kinases (2J5E) <i>Cystéine</i>
			Finasteride (Réversible)	Hyperplasie prostatique bénigne	5- $\alpha$ -réductase <i>Cofacteur</i>
			PX-866 (Wortmannin analogue) (Irréversible)	Cancers	Phosphoinositide-3 kinase (groupe aPK) (1E7U) <i>Lysine</i>
	Iminium acétylénique		Selegiline (Irréversible)	Maladie de Parkinson	MAO-B oxidase <i>Cystéine</i>
Vinyl sulfonyle X= C, N, O					
Ethynyl carbonyle X= C, N, O					
Vinylphényl carbonyle X= C, N, O					
Alkylation (substitution)	Époxyde		Fosfomycine (N.D.)	Infection bactérienne	UDP-N-acetylglucosamine-1-carboxyvinyltransferase (3LTH) <i>Cystéine</i>
	Iminium (état intermédiaire)		D-cycloserine (N.D.)	Infection bactérienne (tuberculose)	Alanine racémase (1NIU) <i>Cofacteur</i>
			Vigabatrine (Irréversible)	Epilepsie	GABA-AT transaminase <i>Cofacteur</i>
	Chlorométhylcarbonyle X= C, N, O				
	Chloropropényl				
Chlorobenzyle					

Suite du Tableau 5 sur la page suivante

Acylation (substitution)	β-lactame		Pénicilline, Amoxiciline, Ampiciline (Irréversible)	Infection bactérienne	Peptidoglycane transpeptidase (1PWC) <i>Sérine</i>
			Clavulanate (Irréversible)	Infection bactérienne	β-lactamase (3IQA) <i>Sérine</i>
	β-lactone		Orlistat (Réversible)	Obésité	Triacylglycerol lipase <i>Sérine</i>
	Ester		Aspirine (Irréversible)	Douleur	COX cyclooxygénase <i>Sérine</i>
	Carbamate		Rivastigmine, ganstigmine (Réversible)	Maladie d' Alzheimer	Acétylcholinestérase (1GQR,2BAG) <i>Sérine</i>
	Coumarine		Warfarine (N.D.)	Anticoagulant	Vitamine K époxyde réductase <i>Cystéine</i>
	Hydrazide	N.D.	Isoniazide (Irréversible)	Infection bactérienne (tuberculose)	Enoyl réductase (1ZID) <i>Cofacteur</i>
Formation d'un hémiacétal	α-cétoamide		VX-950 (Réversible)	Infection virale	Sérine protéase du virus de l'hépatite C <i>Sérine</i>
			SNJ-1945 (Réversible)	Accident vasculaire cérébral	Calpaïne cystéine protéase (2G8J) <i>Cystéine</i>
Fixation d'un métal	Acide boronique		Bortezomib (Réversible)	Cancers	Protéasome (2F16) <i>Thréonine</i>
Réaction de Pinner	Nitrile		Vildagliptine (Réversible)	Diabète	DPP-IV peptidase <i>Sérine</i>
			Odanacatib (Réversible)	Ostéoporose	Cathepsine K <i>Cystéine</i>
Pont disulfure (oxydation)	Thiol		Disulfirame (Irréversible)	Alcoolisme	Aldéhyde déshydrogénase <i>Cystéine</i>
			Oméprazole (Irréversible)	Reflux gastro- œsophagien	H <sup>+</sup> /K <sup>+</sup> ATPase <i>Cystéine</i>
			Clopidogrel (Irréversible)	Antiagrégant plaquettaire	P2Y12 récepteur <i>Cystéine</i>
	Thiourée		Propylthiouracile (N.D.)	Hyperthyroïdisme	Thyroxine 5'-déiodinase <i>Sélenométhionine</i>

**Tableau 5 : Groupes réactifs capables de réagir de manière covalente avec un résidu nucléophile, classés par mécanisme d'action. La plupart des groupes appartiennent à des médicaments et des candidats-médicaments. N.D.=non déterminé. D'autres groupes réactifs, qui ont pu être utilisés au cours de cette thèse, sont également présentés. Les flèches indiquent l'atome qui est attaqué par le nucléophile. Ce tableau est largement inspiré du travail bibliographique de Potashman et Duggan (Potashman et Duggan, 2009).**

### 1.3.2 Le glutathion pour réguler les métabolites réactifs

Le glutathion (GSH) est un tripeptide naturel un peu particulier, de la forme L- $\gamma$ -glutamyl-L-cysteinylglycine (Figure 29).

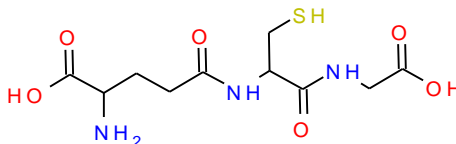


Figure 29 : Le Glutathion

Présent à une concentration élevée dans les cellules, le GSH, grâce à sa fonction thiol, a un rôle antioxydant, capturant toute espèce réactive en solution, tels que les radicaux libres ou les peroxydes, protégeant ainsi les groupements thiols des protéines (Pompella et al., 2003). Pour se faire, le GSH est transformé en sa forme oxydée de glutathion disulfide (GSSG). Normalement, on ne trouve le GSH que sous sa forme réduite car l'enzyme catalysant la réaction inverse, la glutathion réductase, est constitutivement active dans la cellule. Un taux élevé de toxicité dans la cellule peut se traduire par une baisse de la quantité de GSH réduit (Yarbrough et Schultz, 2007). Très souvent, la réactivité des inhibiteurs covalents envers le GSH est testée afin de s'assurer qu'ils ne seront pas capturés avant d'atteindre leur cible (Potashman et Duggan, 2009). C'est, par la même occasion, le moyen de vérifier que l'inhibiteur n'est pas trop réactif.

### 1.3.3 Mode naturel d'inhibition

La petite molécule E-64, extraite d'un champignon en 1978, s'est révélée être un inhibiteur irréversible d'une sous-famille de cystéine-protéases, les papaines (Vicik et al., 2006). Ainsi, en se liant de manière covalente à une cystéine du site actif de l'enzyme via un groupement époxyde, E-64 régule son activité (code PDB : 1PPP) (Kim et al., 1992). La dérégulation des cystéine-protéases ayant été mise en évidence dans diverses maladies immunologiques, plusieurs inhibiteurs irréversibles inspirés de E-64 ont été développés et optimisés pour ne cibler que certaines sous-familles de cystéine-protéases, telles que les cathepsines (Somoza et al., 2002), les calpaines (Cuerrier et al., 2006) ou les caspases (Erlanson et al., 2003) (code PDBs : 1M6D, 2G8J et 1NME, respectivement, pour voir des exemples d'inhibiteurs covalents).

Dans le cas des protéines kinases, Rossi et al. ont mis en évidence, en 2000, l'effet anti-inflammatoire des prostaglandines de type cyclopentenone, via une liaison covalente formée avec la kinase I $\kappa$ B (voir Tableau 4 à la page 52) (Rossi et al., 2000). Cette kinase est en effet responsable de l'activation de NF- $\kappa$ B, un activateur important des gènes impliqués dans l'inflammation et le système immunitaire. Les auteurs ont démontré que le groupement cétone  $\alpha,\beta$ -insaturé de ces prostaglandines pouvait réagir de manière covalente avec une cystéine située sur la *A-loop* de I $\kappa$ B, sans gêner la fixation de l'ATP. Les prostaglandines sont naturellement présentes à faible concentration dans les cellules, sauf en cas d'inflammation aiguë où la concentration locale de prostaglandines atteint le micromolaire.

### 1.3.4 Médicaments covalents ou inhibiteurs suicides ?

Du fait de leur réactivité, les médicaments covalents ont depuis longtemps la mauvaise réputation d'« inhibiteurs suicides », c'est-à-dire d'agents alkylants, rendant très difficile leur acceptation dans les phases de développement clinique (Powers et al., 2002). Cette réputation date des années 1970, où des études ont démontré la responsabilité hépatotoxique de molécules telles que le bromobenzène et l'acétaminophène, via une inhibition covalente (Smith et al., 2009). Cependant, on réalisera plus tard que la toxicité attribuée à la liaison covalente est rarement due à la liaison covalente de la molécule avec sa cible principale, mais aux interactions aspécifiques de la molécule ou d'un métabolite réactif de la molécule avec d'autres cibles. On s'apercevra aussi que des molécules très efficaces, et toujours sur le marché, sont en fait des médicaments covalents irréversibles, soit près d'une quarantaine de molécules (voir Tableau 5 à la page 61). C'est par exemple le cas de l'aspirine, le médicament le plus largement utilisé au monde (Figure 30), de l'oméprazole, l'un des médicaments les plus vendus en 2009, et de la pénicilline et tous les antibiotiques dérivés de  $\beta$ -lactame.

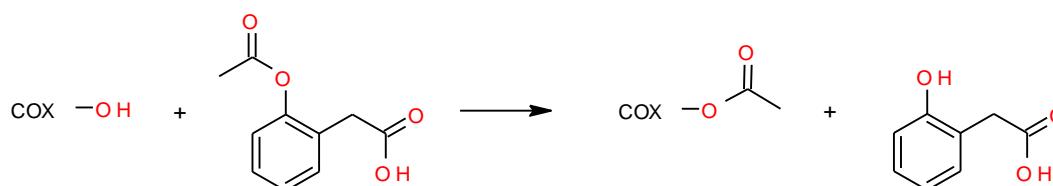


Figure 30 : Schéma réactionnel de l'aspirine se liant irréversiblement à sa cible COX en réagissant avec une sérine.

Finalement, sur les quelques 200 cibles thérapeutiques (Pharmaceutical Research News & Drug Development Pipeline Information, 2010) pour lesquelles des médicaments biodisponibles oralement sont sur le marché, 24 sont les cibles de médicaments covalents ou

de candidats-médicaments covalents, soit 12% (voir Tableau 5 à la page 61) (Potashman et Duggan, 2009; Robertson, 2005). La grande majorité de ces cibles sont des enzymes, dont sept protéases (Figure 31). Les indications thérapeutiques sont variées : douleurs passagères, infection virale, obésité, maladie d'Alzheimer, cancer du sein, etc (Figure 31).

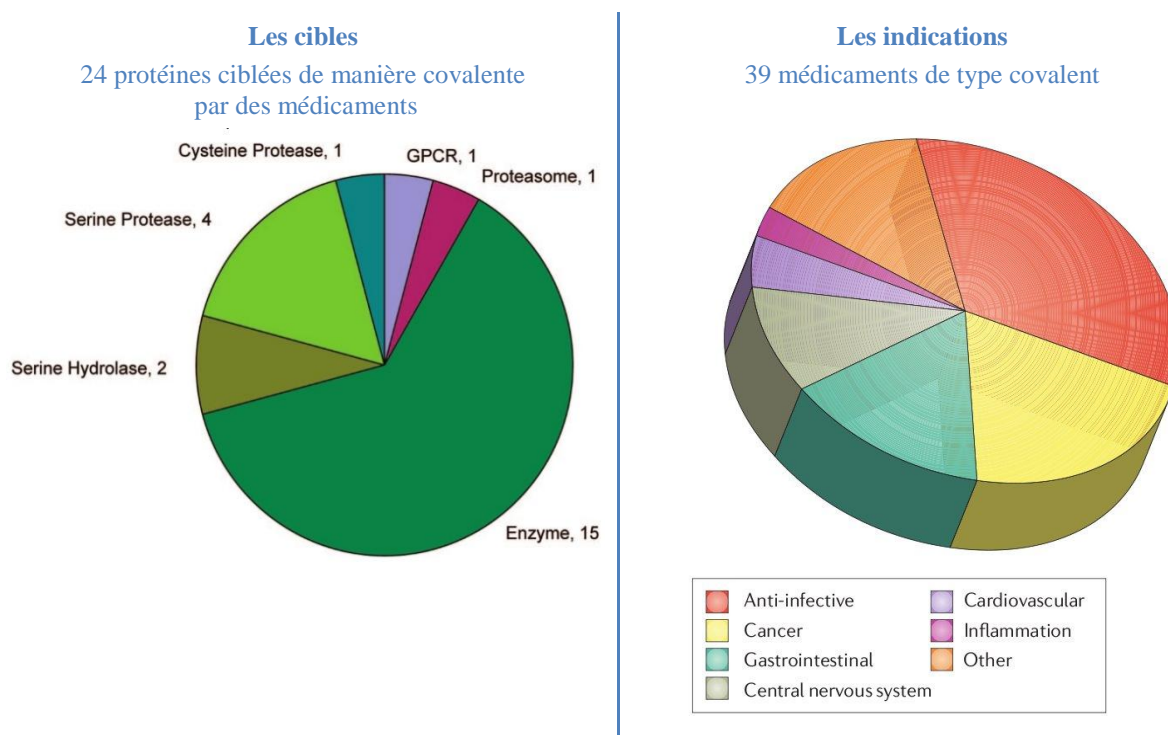


Figure 31 : Distribution des familles de protéines ciblées par des médicaments de type covalent et distribution des indications thérapeutiques (Potashman et Duggan, 2009; Singh et al., 2011).

Les médicaments covalents seraient moins indiqués pour les maladies chroniques car, à usage répété, ils peuvent devenir toxiques et provoquer des réactions allergiques, dites idiosyncratiques (Potashman et Duggan, 2009; Vicik et al., 2006). Par contre, ils sont tout indiqués dans le cas de maladies infectieuses, car la cible est rapidement inhibée et l'agent pathogène n'a pas les moyens de la resynthétiser (Vicik et al., 2006).

L'industrie pharmaceutique recherche des modèles de prédiction de la toxicité pour filtrer les molécules indésirables se trouvant dans leur chimiothèque. Cependant, dans le cas des composés  $\alpha,\beta$ -insaturés, personne n'est encore parvenu à proposer un modèle fiable basé sur des descripteurs physico-chimiques des molécules (Yarbrough et Schultz, 2007). Cela confirme qu'on ne peut pas avoir d'a priori sur des molécules contenant des groupements réactifs, tant que leur réactivité n'a pas été testée expérimentalement.



### 1.3.5 Autres usages de la liaison covalente

La conception de molécules se fixant de manière covalente ne se limite pas à la recherche de nouveaux médicaments mais constitue également des outils chimiques, utilisés en « chémogénomique » ou « protéomique chimique » pour la recherche fondamentale, comme le démontre les exemples suivants.

Alors que le séquençage à haut débit du génome humain a permis d'identifier de nombreux gènes codant pour des protéines, il est à présent nécessaire de trouver la fonction de 30-40% de protéines qui n'ont encore aucun rôle identifié (Heal et al., 2011). Alors que les analyses d'expression de protéines, basées sur des gels d'électrophorèse 2D puis sur de la spectrométrie de masse, ne permettent d'identifier que les protéines présentes en grande quantité dans la cellule, des stratégies chimiques permettant une analyse des protéomes plus détaillée ont été développées. Il s'agit des ABPPs ou *Affinity-based protein probes* (Figure 32) (Chen et al., 2003; Jessani et Cravatt, 2004; Heal et al., 2011).

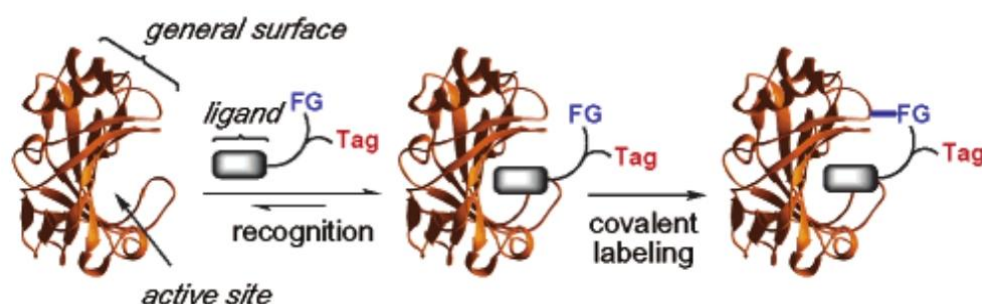
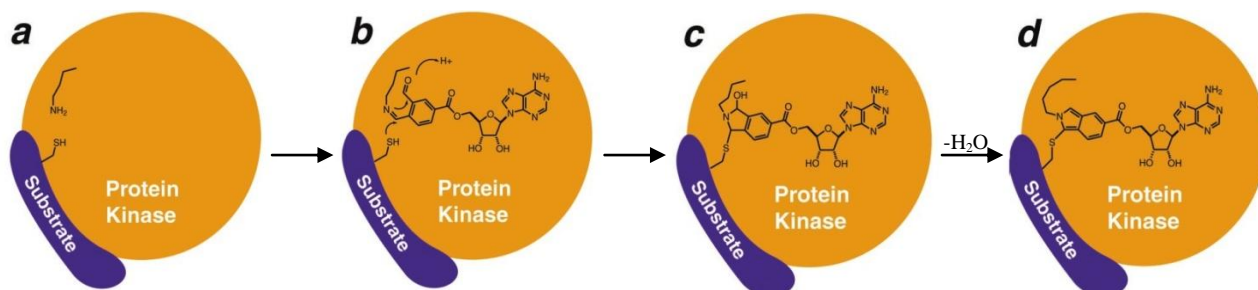


Figure 32 : Concept général des ABPPs. Il s'agit de trouver un ligand capable de se lier irréversiblement et de manière spécifique à une cible, afin de détecter la présence de cette dernière dans des tissus malades, par exemple. FG = groupe fonctionnel, TAG = fluorophore (Chen et al., 2003).

Un ABPP consiste en trois éléments : un groupement réactif qui se lie de manière covalente à une enzyme ou une classe d'enzymes, une région *linker* et un fluorophore (Figure 32). De telles sondes peuvent être utilisées comme biomarqueurs pour l'identification d'enzymes sélectivement exprimées dans des tissus malades. A l'aide de ligands covalents auxquels est attaché un groupe fluorescent, Blair et al. ont montré qu'il était possible d'évaluer la fraction de kinases nécessaires pour la signalisation cellulaire et de quantifier la relation entre l'inhibition de la kinase et l'inhibition de ses substrats (Blair et al., 2007). De la même façon, une cystéine a été introduite dans la proximité du site actif de l'anticorps monoclonal 2D12.5, afin de lier de manière covalente le DOTA (1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid) modifié avec un électrophile, permettant

ainsi un marquage des ions chélatés par le DOTA (Corneillie et al., 2004). La stratégie des ABPPs peut également être utilisée pour permettre la détection de poches et de ligands liés, de manière covalente ou non, à une protéine, sans que ceux-ci ne soient marqués avec un fluorophore (Kim et al., 2004). Cela nécessite la mise en place de deux tests, l'un où la protéine est seule, l'autre où la protéine est mise en présence du ligand testé. Une concentration élevée d'ABPP non spécifique (des agents alkylants du type iodoacétamide ou bromobimane) permet ensuite de marquer toutes les cystéines accessibles. Puis, par digestion tryptique suivie de spectrométrie de masse, les spectres sont comparés et permettent de déterminer la ou les cystéines rendues inaccessibles par la présence du ligand. Une limitation à cette technique est qu'il faut que la protéine contienne au moins une cystéine dans la poche où s'est placé le ligand.

Une méthode de *cross-linking* a été développée par Maly et al, en 2004, pour identifier les kinases responsables de la phosphorylation d'un substrat donné (Figure 33) (Maly et al., 2004). Alors que plusieurs techniques servaient à identifier les protéines phosphorylées, aucune ne permettait de trouver la kinase responsable d'une phosphorylation donnée, du fait que la formation du complexe kinase/substrat soit transitoire et de faible affinité. Le mécanisme de *cross-linking* développé est alors le suivant : le résidu normalement phosphorylé sur le substrat est remplacé par une cystéine. Puis un analogue de l'ATP, capable de former une liaison covalente avec la lysine catalytique de la kinase recherchée (et de toutes les kinases), est synthétisé. Cet analogue est également capable de se lier de façon covalente au substrat, via la cystéine introduite, quand celui-ci est reconnu par la cible recherchée. Enfin, grâce à un fluorophore attaché au substrat, le complexe est localisé sur gel et analysé par spectrométrie de masse (MS) pour déterminer de quelle kinase il s'agit.



**Figure 33 : Représentation schématique de la réaction de *cross-linking* entre la kinase et le substrat, via un analogue de l'ATP réagissant de manière covalente avec une cystéine et la lysine catalytique, respectivement (Maly et al., 2004).**

Dans le contexte d'identification de nouveaux candidats-médicaments, le criblage à haut débit (HTS) de larges bibliothèques de molécules permet d'identifier des touches d'affinité

moyenne ( $K_d$  entre 1 et 10  $\mu\text{M}$ ). Bien que certaines cibles se prêtent facilement à ce criblage, en particulier quand il s'agit d'une librairie dite focalisée, d'autres sont plus problématiques car aucune touche n'est obtenue. Identifier puis optimiser des touches de plus faible affinité pourrait résoudre le problème (Erlanson et al., 2000; Levitsky et al., 2003). Pour ce faire, Erlanson et al. ont développé une librairie de fragments (250 Da) contenant tous un pont disulfure (voir Tableau 5 à la page 61) (Erlanson et al., 2000). Cette librairie a pour but d'être criblée sur une cible où l'on aurait préalablement introduit une cystéine aux abords du site actif. Les fragments liés de façon covalente sont révélés par spectrométrie de masse. L'affinité des quelques fragments identifiés est ensuite remesurée sans la partie disulfide. Si l'affinité se confirme, le fragment est optimisé pour augmenter l'affinité. Ils ont appliqué leur technique dite de *tethering* sur la thymidylate synthase. Cela leur a permis d'obtenir une touche avec une affinité de 1mM, une fois le disulfide enlevé, puis 40nM après la phase d'optimisation. Trois ans plus tard, Erlanson et al. ont étendu leur technique en *double-tethering*, soit un premier criblage permettant de lier de manière covalente des fragments à la cible, via leur fixation préalable dans un premier site, puis un second criblage permettant de lier de manière covalente un second fragment au premier fragment, via sa fixation préalable dans un second site de reconnaissance avec la protéine (Figure 34) (Erlanson et al., 2003).

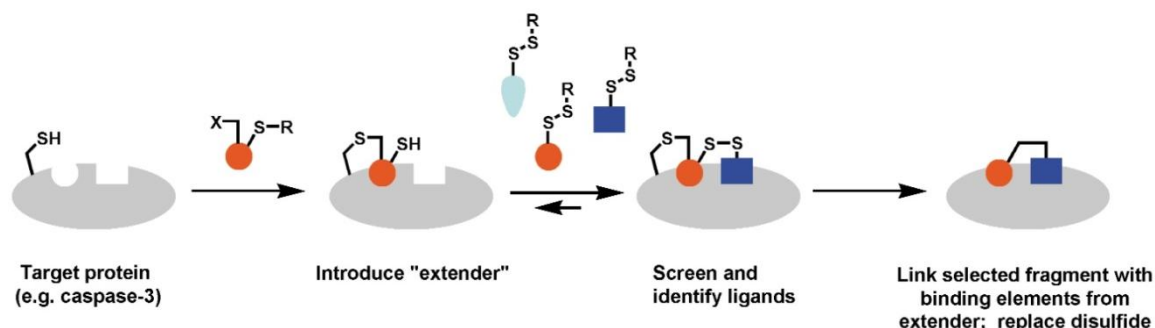


Figure 34 : Représentation schématique de la technique de *double-tethering* (Erlanson et al., 2003).

Enfin, dans le but d'identifier et de caractériser des ligands capables de stabiliser leur cible kinase dans une conformation inactive, Simard et al. (2009) ont récemment développé un test d'affinité basé sur la fluorescence. Ainsi, en attachant de façon covalente un fluorophore sur la *A-loop* de la kinase, les auteurs sont capables de détecter des changements conformationnels et mesurer le  $K_d$  de ligands allostériques (Simard et al., 2009).

### 1.3.6 Pourquoi les inhibiteurs covalents sont-ils plus spécifiques qu'on ne le pense ?

Le groupement électrophile mis sur une molécule, bien qu'appelé parfois *war head*, ne doit pas agir comme une « tête chercheuse » ciblant le premier nucléophile venu (Vicik et al., 2006). Au contraire, il est nécessaire que les deux groupes soient à proximité pendant un certain temps et qu'ils s'orientent favorablement l'un par rapport à l'autre pour que la liaison covalente se forme (Figure 35). Cela revient à dire que le squelette entier du ligand doit contribuer à la première étape de reconnaissance/complémentarité entre le ligand et la protéine. Dans une seconde étape, le groupement électrophile se lie de manière covalente, voir irréversiblement, à la protéine, via un résidu nucléophile peu conservé au sein de la famille. Les inhibiteurs covalents agissent donc comme des *mechanism-based inactivators*, réactifs qu'une fois dans le site visé (Copeland et al., 2006).

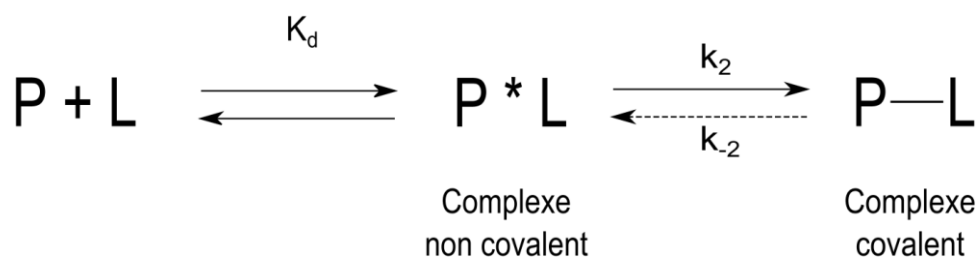


Figure 35 : Schéma réactionnel en deux étapes de la formation d'une liaison covalente entre une protéine et un ligand. Si  $k_2=0$ , la liaison covalente ne se forme pas. Si  $k_{-2}=0$ , la liaison covalente est irréversible.

Pour exemple, des analogues des inhibiteurs covalents d'EGFR ne disposant plus de l'azote sur le squelette quinazoline (dont le rôle est de former une liaison hydrogène avec la région charnière) perdent toute capacité à inhiber EGFR (Carmi et al., 2010). Par ailleurs, en comparant des structures d'EGFR cocrystallisé avec des inhibiteurs covalents et leurs analogues non covalents, Blair et al. ont remarqué deux orientations distinctes des molécules dans la poche : le mode réversible, nécessaire à la reconnaissance de l'inhibiteur ; le mode irréversible, légèrement déplacé par rapport au précédent pour obtenir l'orientation requise pour la réaction covalente (Blair et al., 2007).

## 2 Matériel et Méthodes

---

2.1 De l'extraction à l'exploitation des informations sur les protéines kinases.....	70
2.2 Conception rationnelle d'inhibiteurs.....	75
2.3 Validations expérimentales.....	79

## 2.1 De l'extraction à l'exploitation des informations sur les protéines kinases

### 2.1.1 Création de la base de données dédiée aux kinases

Afin d'évaluer l'étendue des opportunités offertes pour la conception d'inhibiteurs allostériques et/ou covalents parmi les kinases, ainsi que de concevoir de tels inhibiteurs pour des kinases d'intérêt thérapeutique, j'ai dû rassembler et exploiter un grand nombre d'informations, mises sous la forme d'une base de données dédiée aux kinases (Figure 36).

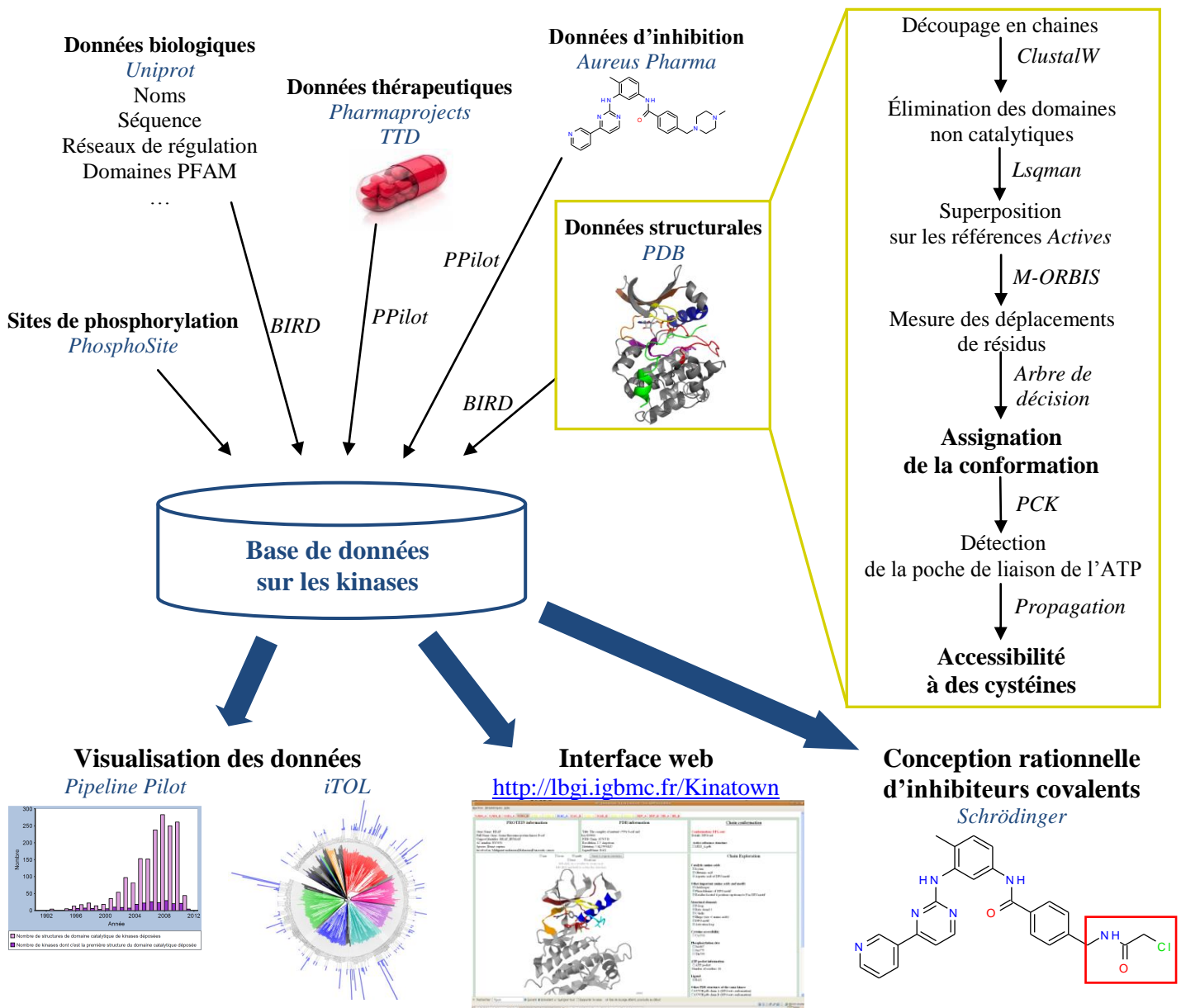


Figure 36 : Schéma global du rapatriement des informations disponibles sur les protéines kinases et de leur organisation dans une base de données qui leur est dédiée. Le traitement des données structurales a permis d'intégrer à la base de données de nouvelles informations sur la conformation que peut adopter chaque kinase et sur l'accessibilité à des cystéines dans le site catalytique. La base de données a facilité l'analyse des informations disponibles sur l'ensemble des kinases et a été exploitée pour concevoir des inhibiteurs allostériques covalents.

### 2.1.2 Les bases de données interrogées

Différents niveaux d'information sur les protéines kinases que j'ai utilisés pour mon travail de thèse ont été extraits de bases de données publiques et commerciales (Figure 36). La base de données *Uniprot* (The UniProt Consortium, 2011) a donné accès à de nombreuses informations biologiques relatives à chaque protéine kinase tels que ses divers identifiants et noms de gène, sa séquence protéique, sa longueur, son organisation en domaines PFAM (Bateman, 2004), son groupe phylogénétique (AGC, TK, etc.), l'espèce d'appartenance, les tissus dans lesquelles elle est exprimée, les réseaux de régulation dans lesquels elle est impliquée, les autres protéines avec lesquelles elle interagit, etc. Elle a également donné accès aux identifiants des structures associées et déterminées par cristallographie aux rayons X (dites « structures 3D » dans la suite de ce manuscrit). Avec cette liste d'identifiants, les structures 3D ont été extraites de la *Protein Data Bank* (PDB) (Berman et al., 2000). Les sites de phosphorylation répertoriés sur les kinases ont été extraits de la base de données publique *PhosphoSite* (Hornbeck et al., 2004). Des données d'inhibition des kinases par de petites molécules ont pu être obtenues grâce à un accès à la base de données commerciale *Aureus Pharma* (Aureus Pharma, 2010). Le but de cette base de données est d'extraire et de référencer toutes les données d'inhibition apparues dans la littérature (nom de la cible kinase, nom et structure chimique de la petite molécule,  $K_d$  ou  $IC_{50}$  mesurés, conditions du test, etc). L'intérêt thérapeutique de chaque kinase a été inféré par l'extraction d'informations contenues dans la base de données commerciale *Pharmaprojects* (Pharmaceutical Research News & Drug Development Pipeline Information, 2010). Cette base de données rassemble des informations sur les petites molécules se trouvant, ou s'étant trouvées, dans les phases de développement clinique. Parmi ces informations, on peut trouver la phase de développement dans laquelle les petites molécules se trouvent, le nom de la cible protéique principale et l'indication thérapeutique. Une autre base de données moins enrichie que *Pharmaprojects* mais qui est publique est la *Therapeutic Target Database* (TTD) (Zhu et al., 2010). On y trouve le nom des protéines pour lesquelles des médicaments ont été développés et le nom de l'indication thérapeutique. Les informations extraites de la TTD, contrairement à celles issues de *Pharmaprojects* ou d'*Aureus Pharma*, peuvent être partagées avec la communauté scientifique.

### 2.1.3 Extraction et organisation de l'information

Les bases de données *Uniprot* et *PDB* sont prises en charge par le système *BIRD*, (Bard et al., 2010; BIRD, 2010) développé dans le Laboratoire de Bioinformatique et de Génomique Intégratives (IGBMC), qui transforme automatiquement leur contenu en bases de données relationnelles. *BIRD* intègre un langage de requête simple et efficace permettant d'accéder rapidement à l'information recherchée. Les données extraites d'*Aureus Pharma* et de *Pharmaprojects* sont le résultat de requêtes lancées depuis des systèmes d'exploitation propres à ces bases de données. Toutes les informations accessibles depuis la *TTD* et *PhosphoSite* sont téléchargeables sous la forme de fichiers textes.

De manière générale, toutes les données que j'ai extraites ont été organisées sous la forme de deux fichiers textes qui constituent, par abus de langage, ma base de données sur les kinases (voir Figure 36 à la page 70). Le premier fichier regroupe les informations propres à chaque kinase, identifiée par son *Accession Code* (identifiant AC dans *Uniprot*). Le second fichier recense les informations propres à chaque structure 3D de kinases.

Il a ainsi fallu organiser les données extraites d'*Aureus Pharma* et de *Pharmaprojects*. J'ai utilisé, dans le cas d'*Aureus Pharma*, la plateforme *Pipeline Pilot* (*PPilot*) (Pipeline Pilot is Accelrys' scientific informatics platform, 2010) pour regrouper par cible kinase toutes les petites molécules et les données d'inhibition associées. J'ai fait de même avec les données de *Pharmaprojects*, où j'ai regroupé par cible kinase les petites molécules en développement (ou commercialisées) et l'indication thérapeutique associée. *PPilot* permet de construire des protocoles, c'est-à-dire une succession d'actions préprogrammées, représentées sous forme de boîtes avec des paramètres à ajuster, qui s'appliquent aux données initiales (Figure 37). Une action peut être un filtre, un tri, un regroupage, etc. En sortie, les données organisées sont enregistrées sous le format souhaité.

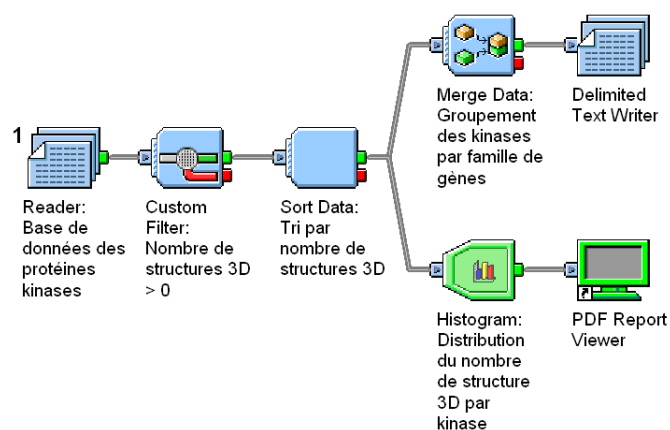


Figure 37: Exemple de protocole construit avec Pipeline Pilot pour organiser mais aussi exploiter les données contenues dans la base de données des kinases.



### 2.1.4 Enrichissement de la base de données

L'utilité de la base de données des kinases, que j'ai construite, n'est pas seulement de regrouper les informations extraites de plusieurs autres bases de données, mais également de contenir de nouvelles informations issues du traitement des premières. Ainsi, le traitement des structures 3D a permis d'ajouter de nouvelles informations à la base de données qui sont, principalement, l'assignation de la conformation (voir chapitre 3 « Étude de la Diversité Conformationnelle au Sein des Protéines Kinases ») et l'accessibilité à des cystéines dans le site catalytique (voir chapitre 4 « Cartographie des Cystéines et Application à la Conception d'Inhibiteurs Covalents » ; voir Figure 36 à la page 70). L'accès à ces nouvelles informations a nécessité l'utilisation de divers programmes (voir Figure 36 à la page 70). Le programme d'alignement de séquences *ClustalW* (Thompson et al., 2002) a permis de filtrer les structures 3D de chaînes protéiques ne correspondant pas l'identifiant AC de kinase voulu. *Lsqman* (Kleywegt, 1996) a été utilisé pour superposer chaque structure 3D sur les structures de référence de kinases en conformation *Active*, comme détaillé dans le chapitre 3. *PCK* (Albou et al., 2009), développé au sein de notre équipe (IGBMC), a permis d'identifier la composition de la poche formant le site catalytique. Des développements du logiciel *M-ORBIS* (Albou et al., 2011), développé dans le Laboratoire de Bioinformatique et Génomique Intégratives (IGBMC), ont été utilisés pour faciliter le parcours des chaînes protéiques et des résidus au sein de chaque structure et ont permis de mesurer rapidement des déplacements d'atomes entre diverses structures ainsi que des distances intramoléculaires. Tous ces programmes et leur paramétrages sont décrits plus en détail dans la partie « *Materials and Methods* » des deux publications insérées dans les chapitres 3 et 4.

Le langage de programmation java a finalement été utilisé pour coordonner l'organisation et l'enrichissement des informations, et aboutir à la création de la base de données des kinases.

### 2.1.5 Exploitation de la base de données des kinases

*PPilot* a permis de visualiser les données contenues dans la base de données des kinases, sous la forme d'histogrammes, de graphiques à deux dimensions, de diagrammes de Venn, etc (voir Figure 36 à la page 70 ; Figure 37). Tous les histogrammes et graphiques présentés dans ce manuscrit ont été réalisés avec *PPilot*.

L'interface web *iTOL* (Letunic et Bork, 2007) a été utilisée afin de comparer les informations quantitatives propres à chaque kinase humaine. *iTOL* permet d'afficher selon une forme circulaire un arbre phylogénétique donné en entrée (dans ce manuscrit, l'arbre des kinases humaines selon Manning (Manning et al., 2002)) et d'associer à chaque branche de l'arbre (et donc à chaque kinase humaine) une information (voir Figure 36 à la page 70).

Les principales informations issues de ce travail de thèse, c'est-à-dire l'assignation de la conformation pour une structure de kinase donnée et l'accessibilité à des cystéines dans le site catalytique, sont visualisables depuis l'interface web « Kinatown » (voir Figure 36 à la page 70). La création de cette interface a nécessité l'apprentissage des langages de développement web : PHP et HTML. L'affichage de la structure 3D passe par l'interface graphique jmol ([www.Jmol.org](http://www.Jmol.org)). J'ai donc également dû utiliser le langage de développement javascript pour faire communiquer les actions effectuées par l'utilisateur du site web et leurs conséquences sur la structure 3D visualisée.

J'ai surtout mis à profit les données accumulées sur la régulation allostérique des kinases et la présence de cystéines dans le site catalytique pour concevoir, de manière rationnelle, c'est-à-dire assistée par ordinateur, des inhibiteurs covalents de kinases (voir Figure 36 à la page 70). Pour cela, j'ai été amenée à utiliser des programmes de modélisation moléculaire et de chémoinformatique que je décris dans les prochains paragraphes.

## 2.2 Conception rationnelle d'inhibiteurs

### 2.2.1 La cible kinase et les modèles 3D par homologie

Le choix de la structure 3D de la cible kinase contre laquelle les ligands devaient être conçus a dépendu de plusieurs critères : une bonne résolution (environ 2Å), pas ou peu de mutations dans le site catalytique et la présence d'un ligand cocrystallisé dans le site catalytique. Avec les programmes fournis par le logiciel de modélisation moléculaire *Schrödinger* (Schrödinger), les résidus manquants ont éventuellement été modélisés et les résidus mutés ont été corrigés (programme *Prime*). La structure 3D a ensuite été minimisée énergétiquement afin de relaxer les interactions entre les résidus de la protéine, et entre la protéine et le ligand (programme *MacroModel*). Le programme *MODELLER* (version 9v7) (Fiser et Sali, 2003) a été utilisé pour générer un modèle 3D par homologie dans les cas où aucune structure 3D n'était disponible pour la cible kinase ou bien que le domaine catalytique cristallisé n'était pas dans la conformation souhaitée. La structure 3D qui était alors prise comme référence correspondait à la kinase partageant la plus grande identité de séquence avec la kinase à modéliser. L'alignement de séquences, nécessaire pour la construction du modèle, entre le domaine catalytique de la kinase à modéliser et celui de la référence choisie, a été généré avec *ClustalW* (Thompson et al., 2002).

### 2.2.2 Les petites molécules et futurs inhibiteurs

La majeure partie des inhibiteurs covalents que j'ai conçus de manière rationnelle sont inspirés de ligands non covalents cocrystallisés avec la protéine cible. Ainsi, la connaissance de leur mode de liaison dans le site catalytique et la proximité avec une cystéine accessible dans le site m'ont permis de concevoir des ligands analogues aux ligands non covalents et présentant un groupe électrophile idéalement placé sur le squelette de la petite molécule.

Certains ligands au squelette un peu particulier (présence d'un macrocycle) sont inspirés de molécules extraites de la *Cambridge Structure Database* (CSD) (Allen, 2002). La CSD donne accès aux structures cristallisées de plus de 500 000 petites molécules, et fournit ainsi l'information de conformation qu'elles peuvent adopter. Chaque conformation que peut adopter une petite molécule est appelée « conformère ».

J'ai ensuite utilisé des méthodes d'amarrage (ou de *docking*) pour valider mes propositions d'inhibiteurs covalents avant de les exposer aux collaborateurs chimistes.

### 2.2.3 Méthodes d'amarrage (*docking*) et évaluation de l'affinité de l'inhibiteur (*scoring*)

La formation du complexe protéine/ligand par liaison covalente se fait en deux étapes, une première étape de reconnaissance, où le ligand entre dans le site actif, puis une seconde étape où le lien covalent se forme. Le *docking* permet d'appréhender la situation lors de la première étape. Le but du *docking* est d'anticiper la possibilité qu'un ligand quelconque puisse ou non se loger dans une poche de la protéine avec une bonne affinité. Ainsi, le *docking* de chaque ligand génère un ensemble de modes de liaison appelés « poses » pour lesquelles un score est calculé. Ce score est une approximation de l'énergie de liaison du complexe (kcal/mol) et donc plus il est négatif, meilleure est l'affinité prédite. Cependant, le score en lui-même est peu informatif et il est plus intéressant de le comparer au score obtenu pour un inhibiteur confirmé de la cible kinase visée, dont l'affinité et le mode de liaison sont connus. Le ligand cocrystallisé représente alors souvent la référence idéale. Ainsi, j'ai systématiquement procédé à l'amarrage du ligand cocrystallisé pour obtenir un score de référence et, par la même occasion, valider la performance prédictive du programme de *docking* utilisé, en comparant la meilleure pose proposée à celle observée dans la structure 3D. Tous les résultats de *docking* évoqués dans ce manuscrit ont été effectués avec le programme *Glide* (Friesner et al., 2004), et ont été analysés via l'interface *Maestro*, ces deux programmes étant fournis par *Schrödinger* (Tableau 6). Le principe de *Glide* et les divers modes d'utilisation sont décrits dans les paragraphes suivants.

Remarque : la réaction covalente ne peut être décrite que par mécanique quantique, lors de la seconde étape où le lien covalent entre l'inhibiteur et une cystéine du site catalytique se forme. Cette technique étant beaucoup trop demandeuse en temps de calcul, surtout dans le cas de grands systèmes tels que les complexes protéine/ligand, il reste envisageable d'étudier la réaction par l'utilisation combinée de la mécanique quantique et de la mécanique moléculaire. Cependant, aucun programme fiable n'est actuellement disponible.

Paragraphe du manuscrit de thèse	Ligands	Cibles kinases (Code PDB, résolution)	Modes de <i>docking</i> utilisés/ Modifications/particularités
« 4.2 Conception du premier inhibiteur covalent de la kinase KIT »	Analogues de l'imatinib avec divers groupements électrophiles	• KIT (1T46, 1.60Å)	❖ <i>Docking</i> rigide ❖ <i>Docking</i> covalent manuel
« 4.5.1 Amélioration du profil de sélectivité de l'hypothemycine parmi les kinases ayant une cystéine avant le motif DFG »	Analogues de l'hypothemycine présentant divers substituants sur le macrocycle	• KDR (3CJG, 2.25Å) • KIT (1PKG, 2.90Å) • GSK3b (1J1B, 1.80Å) • MNK2 (2HW7, 2.71Å) • ERK2 (3C9W, 2.50Å) • TAK1 (2EVA, 2.00Å)	❖ <i>Docking</i> flexible : Optimisation du positionnement de la chaîne latérale du résidu <i>gatekeeper</i> et de la cystéine située avant le motif DFG  ❖ <i>Docking</i> covalent manuel
« 4.5.2 Réadressage de l'hypothemycine en vue de l'inhibition covalente d'EGFR via la cystéine située à la fin de la région charnière »	Dérivés de l'hypothemycine avec un macrocycle ne contenant pas le groupe électrophile	• EGFR (1M17, 2.60Å)	Avant le <i>docking</i> : Éloignement de la chaîne latérale du résidu aspartique sur le motif DFG par rapport au site actif en changeant de rotamère  ❖ <i>Docking</i> rigide : Une molécule d'eau conservée  ❖ <i>Docking</i> covalent manuel
« 4.5.2 Réadressage de l'hypothemycine en vue de l'inhibition covalente d'EGFR via la cystéine située à la fin de la région charnière »	Dérivés de l'hypothemycine sans le macrocycle	• EGFR (1M17, 2.60Å)	❖ <i>Docking</i> rigide : Une molécule d'eau conservée  ❖ <i>Docking</i> covalent manuel

Tableau 6 : Détails des amarrages (ou *dockings*) effectués avec *Glide* et présentés dans ce manuscrit de thèse.

### 2.2.3.a Docking rigide avec *Glide*

*Glide* (Friesner et al., 2004) commence par générer une grille, qui est une représentation simplifiée de la forme et des propriétés physico-chimiques de la poche visée. Différents niveaux d'information contenus dans la grille sont pris en compte au fur et à mesure des étapes de *docking*, afin de fournir un score de *docking* de plus en plus précis. Le score de *docking* final, calculé par *Glide* et permettant d'évaluer les différentes poses, est une approximation de l'énergie de liaison entre le ligand et la protéine, qui prend en compte les contacts hydrophobes, la formation de liaisons hydrogènes, le coût de la désolvatation, les effets entropiques et les possibles clashes stériques.

*Glide* procède par une succession de filtres pour trouver les meilleures poses de chaque ligand. Tout d'abord, plusieurs conformères de chaque ligand à amarrer sont générés, puis positionnés et scorés grossièrement, en n'utilisant que les informations spatiales contenues dans la grille. Cette étape permet d'éliminer rapidement un grand nombre de conformères. La conformation des ligands restants est minimisée dans le contexte de la protéine puis le score est recalculé. Les poses des ligands ayant les meilleurs scores sont encore optimisées via l'exploration stochastique de l'espace conformationnel accessible par le ligand, issue d'un échantillonnage de Monte Carlo.

### 2.2.3.b *Docking flexible avec Glide*

Le *docking* flexible a été utilisé lorsque la position originale de certaines chaînes latérales de résidus a provoqué l'obtention de mauvaises poses lors du *docking* rigide. Ce mode de *docking* s'opère en quatre étapes : la génération des conformères à partir des ligands à amarrer ; le *docking* rigide dans la protéine ; l'optimisation des chaînes latérales proches du ligand ; le *redocking* dans la protéine optimisée. Ce processus est beaucoup plus coûteux en temps mais permet d'obtenir des poses plus précises en adaptant le site protéique au ligand qu'il reçoit. Une conséquence directe est l'amélioration des scores obtenus pour chaque ligand.

### 2.2.3.c *Docking covalent avec Glide*

Il ne s'agit pas réellement d'un *docking*, mais plutôt de la minimisation du ligand, une fois qu'il est attaché de manière covalente à la protéine. Le ligand est en fait considéré comme une chaîne latérale d'un acide aminé dont les diverses positions sont échantillonnées et scorées. Les poses obtenues avec cette méthode ont rarement été convaincantes puisqu'elles faisaient pointer le ligand vers le solvant plutôt que dans le site actif. Ces observations semblent être dues au fait que les cystéines sur lesquelles les ligands étaient attachés se trouvaient proches de la sortie du site actif.

### 2.2.3.d *Docking covalent manuel puis affinement du complexe*

Afin d'évaluer la possibilité de former une liaison covalente entre le ligand et la protéine au niveau de la cystéine visée, le complexe non covalent protéine/ligand obtenu lors

de l'étape de *docking* rigide ou flexible a été utilisé comme point de départ. Le groupe électrophile, si absent du ligand, a été dessiné manuellement dans *Maestro*, dans son état covalent. Le lien covalent entre le soufre de la cystéine et le carbone électrophile du ligand a également été ajouté manuellement. Le complexe a ensuite été soumis à une minimisation d'énergie dans la zone modifiée (programme *MacroModel*). L'impact du lien covalent sur le positionnement du ligand dans le site a été analysé visuellement, par comparaison avec le complexe non covalent initial.

## 2.3 Validations expérimentales

Les ligands covalents conçus pendant mon travail ont été synthétisés (ou sont en projet de synthèse) dans le Laboratoire de Chimie Organique et Bio-Organique dirigé par le Professeur Nicolas Winssinger (Institut de Science et d'Ingénierie Supramoléculaires, ISIS, Strasbourg). Les tests d'inhibition de l'activité enzymatique ont été faits chez *Proqinase* et *Cerep* pour les inhibiteurs covalents des kinases KIT et EGFR, respectivement. Les tests d'affinité et de sélectivité parmi les kinases humaines ont été faits chez *KinomeScan*. Les tests de spectrométrie de masse pour valider la formation de la liaison covalente entre KIT et les inhibiteurs covalents ont été effectués dans le laboratoire du Professeur N. Winssinger. Les tests d'inhibition de la croissance cellulaire sur des lignées de leucémie à mastocytes, présentant une activité constitutive de la kinase KIT, ont été réalisés par la plateforme Pucés à Cellules Transfectées de l'IGBMC, dirigée par le Docteur Laurent Brino. Ces lignées cellulaires ont été généreusement mises à disposition par les Docteurs Andres Garcia Montero, de l'Université de Salamanca, et Joseph Butterfield, de l'Université de Chicago.





# 3 Étude de la Diversité Conformationnelle au Sein des Protéines Kinases

---

3.1 Une information difficile à obtenir.....	82
3.2 Rapatriement et analyse des structures.....	83
3.3 Stratégie d'assignation de la conformation.....	84
3.4 Première assignation de la conformation.....	89
3.5 De trois à cinq conformations.....	91
3.6 Article en cours de soumission.....	92
3.7 Quel niveau de détail au sein de chaque conformation ?.....	93
3.8 Conclusion.....	121

## 3.1 Une information difficile à obtenir

Au fur et à mesure de la résolution de nouvelles structures cristallographiques de kinases, les auteurs se sont généralement référés à la première structure de cAPK (voir Figure 9 à la page 24) pour souligner les similarités de repliement et expliquer le mécanisme catalytique très conservé. Par contre, lorsque les structures révélèrent une perturbation de ce mécanisme, les auteurs ne s'attardaient pas toujours à expliquer les différences, évoquant un artefact de cristallisation (Levinson et al., 2006) ou le fait que les domaines additionnels n'étaient pas présents pour contribuer à la conformation *Active* du domaine catalytique (Taylor et Radzio-Andzelm, 1997). Cette remarque est d'autant plus vraie que l'intérêt d'obtenir la structure 3D d'une kinase a vite évolué. En effet, les structures 3D les plus récentes ont majoritairement pour but d'expliquer l'efficacité d'un inhibiteur cocrystallisé dans le site actif, plutôt que d'étudier la position des résidus catalytiques ou les mécanismes de régulation. Pourtant, l'étude de la diversité conformationnelle des kinases pourrait représenter une source importante d'informations pour la conception d'inhibiteurs sélectifs parmi les kinases.

Actuellement, trois conformations majoritaires ont été décrites dans la littérature : la conformation *Active* (Knighton et al., 1991) et deux conformations inactives, dite *DFG out* (Hubbard et al., 1994) et *C-helix out* (De Bondt et al., 1993) (voir paragraphe 1.2.3 « Régulation de l'activité catalytique » à la page 27). Bien que la majorité des kinases ait été cristallisée en conformation *Active*, il est envisagé que toutes peuvent adopter une voir plusieurs conformations inactives. Pour certains auteurs, chaque kinase aurait une conformation inactive préférentielle, selon son état de phosphorylation et les mécanismes régulant son activité (Cowan-Jacob et al., 2009). Pour d'autres, les kinases adopteraient un continuum d'états inactifs, du fait qu'aucune fonction catalytique ne soit requise contrairement à la conformation *Active*, et chacun de ces états pourrait être la cible d'inhibiteurs de types différents (Alton et Lunney, 2008; Cowan-Jacob, 2006).

Il est donc difficile de dresser la liste des kinases vues en conformations inactives en se basant sur la bibliographie puisque la structure 3D du domaine catalytique ne fait pas toujours l'objet de la publication, que certaines de ces structures n'ont pas de publication associée, et que le nombre de structures 3D de kinases déposées dans la *Protein Data Bank* (PDB) (Berman et al., 2000) a tendance à augmenter chaque année (Figure 38).

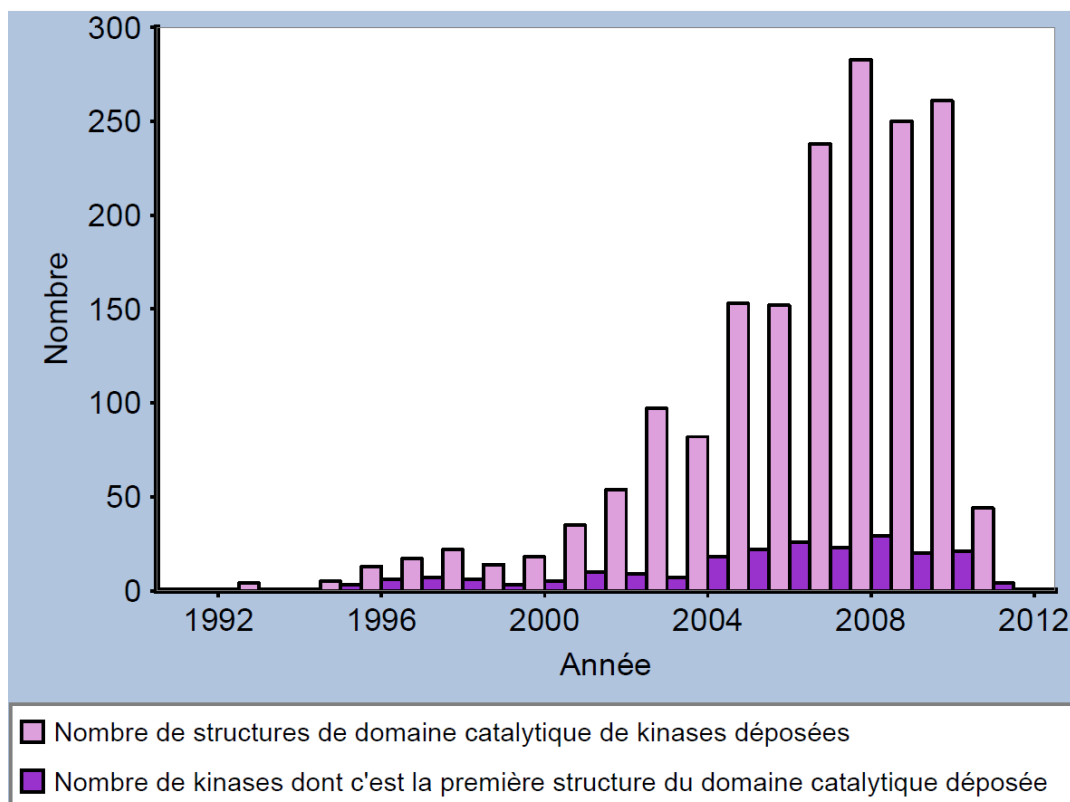


Figure 38 : Histogramme du nombre de structures du domaine catalytique de kinases déposées à la PDB entre 1993, qui est la date de la première structure déposée, et juin 2011.

Pour contourner les limitations précitées, nous avons décidé d'élaborer une stratégie d'assignation automatique de la conformation, applicable à toute structure de domaine catalytique de kinase donnée.

## 3.2 Rapatriement et analyse des structures

En juin 2011, j'ai extrait de la PDB les 1744 structures 3D contenant des domaines catalytiques de kinases. Ces domaines catalytiques recouvrent 221 kinases distinctes (= identifiant AC distinct). En se restreignant à l'espèce humaine, ce sont 172 des 518 kinases distinctes qui ont leur domaine catalytique cristallisé parmi les 1414 structures déposées (Tableau 7). Parmi les neuf groupes phylogénétiques au sein des protéines kinases humaines, les TKs ont été les plus étudiées avec près de la moitié d'entre elles qui ont été cristallisées. Les CMGCs sont également très étudiées, en particulier la kinase CDK2 qui comptabilise 197 structures 3D à elle seule. La majorité de ces structures sont de bonne qualité avec une résolution moyenne de 2.3Å (Figure 39).

Groupe	Nombre de domaines catalytiques humains <u>crystallisés</u>	Nombre de kinases humaines <u>distinctes cristallisées</u>	Fraction cristallisée par groupe (%)
AGC	97	15	26.3
CAMK	221	30	40.5
CMGC	487	25	41.0
CK1	10	6	50.0
STE	55	19	40.4
TK	373	45	50.0
TKL	49	14	32.6
RGC	0	0	0.0
Autres	120	17	20.5
Atypique	2	1	2.5
Total	1414	172	33.2

Tableau 7 : Répartition des structures de domaines catalytiques humains dans les divers groupes phylogénétiques. Plus de 30% des protéines kinases humaines ont leur domaine catalytique cristallisé.

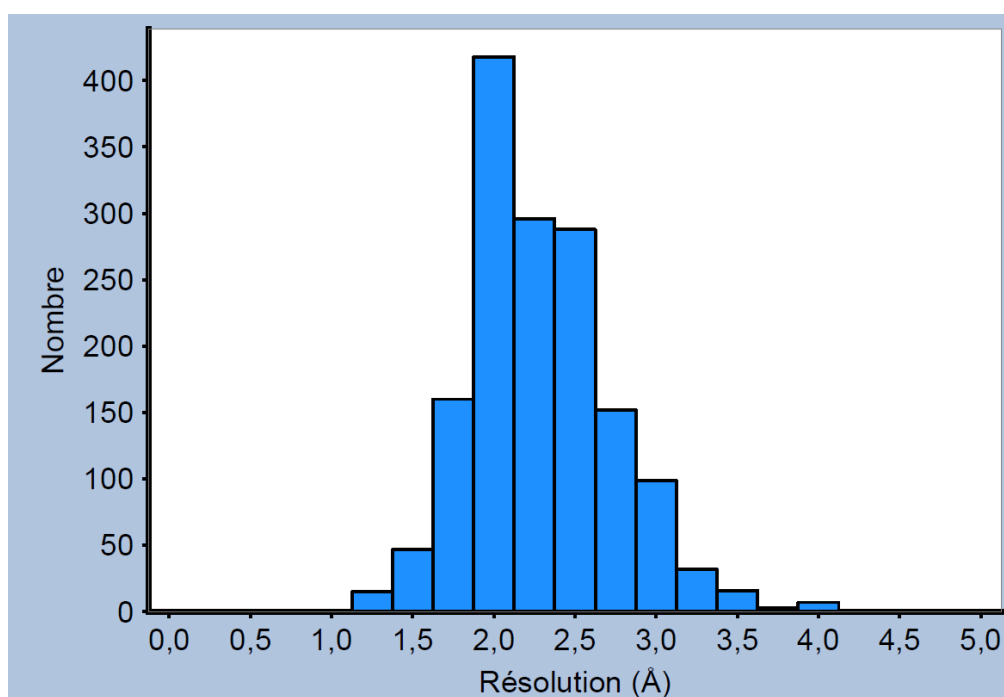


Figure 39 : Répartition de la résolution des domaines catalytiques de kinases.

### 3.3 Stratégie d'assignation de la conformation

Afin de procéder à la classification des structures cristallographiques des domaines catalytiques de kinases, j'ai commencé par effectuer une analyse des fluctuations existantes entre des kinases provenant de différents groupes phylogénétiques (AGC, TK, STE, etc.), au sein de la conformation *Active*. Pour cela, j'ai sélectionné une référence pour chaque groupe

de kinases, dont la structure en conformation *Active* était disponible et de très bonne qualité, soit une résolution de l'ordre de 2Å (Tableau 8). Il convient de noter que le groupe RGC est le seul groupe à ne disposer d'aucune structure 3D.

Code PDB (chaîne)	Kinase	Groupe	Résolution (Å)
2JDO (A)	AKT2	AGC	1.80
1JKS (A)	DAPK1	CAMK	1.50
2CMW (A)	CSNK1G1	CK1	1.75
3BHV (A)	CDK2	CMGC	2.10
2HY8 (I)	PAK1	STE	2.00
3LCK (A)	LCK	TK	1.70
2QLU (A)	ACVR2B	TKL	2.00

Tableau 8 : Les sept structures 3D en conformation *Active* choisies comme références, une par groupe phylogénétique.

Les sept structures 3D ont été superposées à 3LCK (référence choisie pour le groupe TK ; Tableau 8), servant ainsi de référence principale (Figure 40). Les déplacements des carbones  $\alpha$  de chaque résidu par rapport aux résidus équivalents de la référence ont ensuite été mesurés (voir Figure 41 à la page 87).

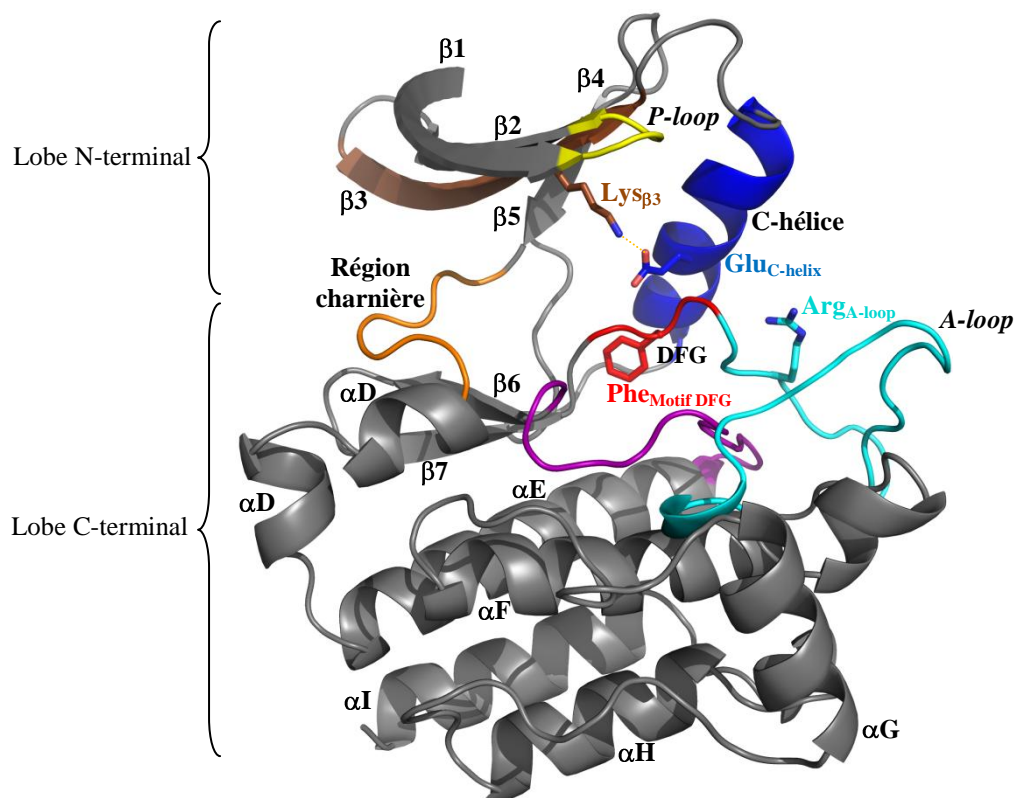


Figure 40 : Structure du domaine catalytique de la tyrosine kinase LCK dans la conformation *Active* et dénomination (en noir) de tous les éléments de structures secondaires (code PDB : 3LCK). Dénomination des résidus impliqués dans les critères géométriques pour l'assignation de la conformation (Phe<sub>DFG</sub> en rouge, Glu<sub>C-hélix</sub> en bleu, Lys<sub>β3</sub> en marron et Arg<sub>A-loop</sub> en cyan).

Cette analyse (Figure 41) démontre une forte conservation du repliement (déplacement des carbones  $\alpha$  inférieur à 1.5Å) au niveau des brins  $\beta 2$  à  $\beta 5$ , de l'extrémité C-terminale de la C-hélice, de la boucle entre la C-hélice et le brin  $\beta 4$ , de la région charnière, de la boucle catalytique, des brins  $\beta 6$  et  $\beta 7$ , du début de la *A-loop* (motif DFG + les quatre résidus suivants) et des hélices  $\alpha D$ ,  $\alpha E$  et  $\alpha F$ . Par contre, d'importantes fluctuations (déplacement des carbones  $\alpha$  supérieur à 4Å, Figure 41) sont observées au niveau du brin  $\beta 1$ , de la *P-loop*, de la boucle reliant  $\beta 2$  à  $\beta 3$ , de la région reliant  $\beta 3$  à la C-hélice, du côté N-terminal de la C-hélice, de la boucle reliant  $\beta 4$  à  $\beta 5$ , de la boucle reliant  $\alpha D$  à  $\alpha E$ , de la *A-loop* (après les sept premiers résidus), de la boucle reliant  $\alpha F$  à  $\alpha G$ , et des hélices  $\alpha G$ ,  $\alpha H$  et  $\alpha I$ . Toutes ces fluctuations s'expliquent par des longueurs de boucles différentes entre les groupes (e.g. AVR2B a souvent des boucles plus longues - de 5 jusqu'à 13 résidus - que LCK ; CDK2 a une insertion de boucle de 30 résidus entre les hélices  $\alpha G$  et  $\alpha H$ ), ainsi que la présence d'hélices supplémentaires (e.g. AKT2 a une petite hélice reliant  $\beta 3$  à la C-hélice, ce qui déplace un peu l'extrémité C-terminale de la C-hélice ; LCK a deux petites hélices au lieu d'une seule hélice  $\alpha D$ , Figure 40). Par conséquent, il est préférable de travailler avec plusieurs structures de référence afin de prendre en compte les différences entre les divers groupes phylogénétiques.

L'analyse précédente ayant permis d'identifier les zones de fluctuation entre diverses structures en conformation *Active*, l'étape suivante a consisté à déterminer quels sont les meilleurs descripteurs permettant de distinguer les conformations inactives de la conformation *Active*. Pour cela, j'ai pris une structure en conformation inactive *DFG out* et une structure en conformation inactive *C-helix out* que j'ai ensuite superposées à la structure *Active* de référence 3LCK (Figure 41). Une région où seule la structure inactive *C-helix out* diverge des autres structures, y compris de la structure inactive *DFG out*, correspond à la C-hélice, et plus particulièrement à son centre où se situe l'acide glutamique catalytique, qui est alors trop éloigné de la position active pour interagir avec la lysine du brin  $\beta 3$  (voir paragraphe 1.2.3.b « Conformation inactive C-helix out » à la page 28). L'unique région où la structure inactive *DFG out* se distingue des autres conformations correspond à l'exacte position du motif DFG, avec le plus grand déplacement observé pour le carbone  $\alpha$  de la phénylalanine, puisque ce résidu pointe dans le site de fixation de l'ATP (voir paragraphe 1.2.3.c « Conformation inactive DFG out »). Puis, dans la zone située juste après le motif DFG, la structure inactive *C-helix out* subit un déplacement du même ordre de grandeur (déplacement supérieur à 4Å) que la structure inactive *DFG out*, exprimant ainsi



Figure 41 : Graphique montrant les fluctuations des carbones  $\alpha$  (Å) le long de la séquence protéique entre six des sept références *Actives* (courbes jaunes) et la référence *Active* 3LCK d'une part, et les fluctuations des carbones  $\alpha$  entre les deux conformations *DFG out* (courbe rouge) et *C-helix out* (courbes bleue et verte) et la référence 3LCK d'autre part. Les zones les plus conservées correspondent à un déplacement des carbones  $\alpha$  inférieur à 1.5Å, les fluctuations les plus importantes correspondent à un déplacement des carbones  $\alpha$  supérieur à 4Å (traits horizontaux en pointillés gris). Les éléments de structure secondaire sont indiqués sous les courbes et sont encadrés selon le code couleur de la Figure 40. Les quatre résidus faisant partie des critères géométriques pour l'assignation de la conformation sont encadrés en pointillés marron, bleu, rouges et cyan.

une divergence nette des deux structures inactives par rapport aux sept structures *Actives*. C'est en effet cette première partie de la *A-loop* (motif DFG + les quatre résidus suivants) qui est censée former le site de fixation du substrat. Un déplacement élevé des sept premiers résidus de la *A-loop* est donc synonyme d'une conformation inactive.

Ces diverses analyses m'ont permis de mettre au point une stratégie d'assignation de la conformation adoptée par une structure 3D de kinase donnée, en commençant par sa superposition sur l'une des sept références *Actives* avec laquelle la structure a le maximum de résidus alignés. Puis, j'ai automatisé le calcul du déplacement de certains résidus équivalents entre les deux structures, plus particulièrement l'acide glutamique porté par la C-hélice ( $\text{Glu}_{\text{C-hélice}}$ ), la phénylalanine du motif DFG ( $\text{Phe}_{\text{Motif DFG}}$ ) et l'arginine ( $\text{Arg}_{\text{A-loop}}$ ) localisée quatre résidus en aval de la phénylalanine sur la *A-loop* (voir Figure 40 à la page 85 ; Figure 42). J'ai également mesuré la distance entre la lysine catalytique du brin  $\beta_3$  ( $\text{Lys}_{\beta_3}$ ) et l'acide glutamique de la C-hélice (voir Figure 40 à la page 85 ; Figure 42), une autre caractéristique permettant de discriminer la conformation inactive *C-helix out* des autres conformations. Cette dernière mesure est indépendante de la superposition.

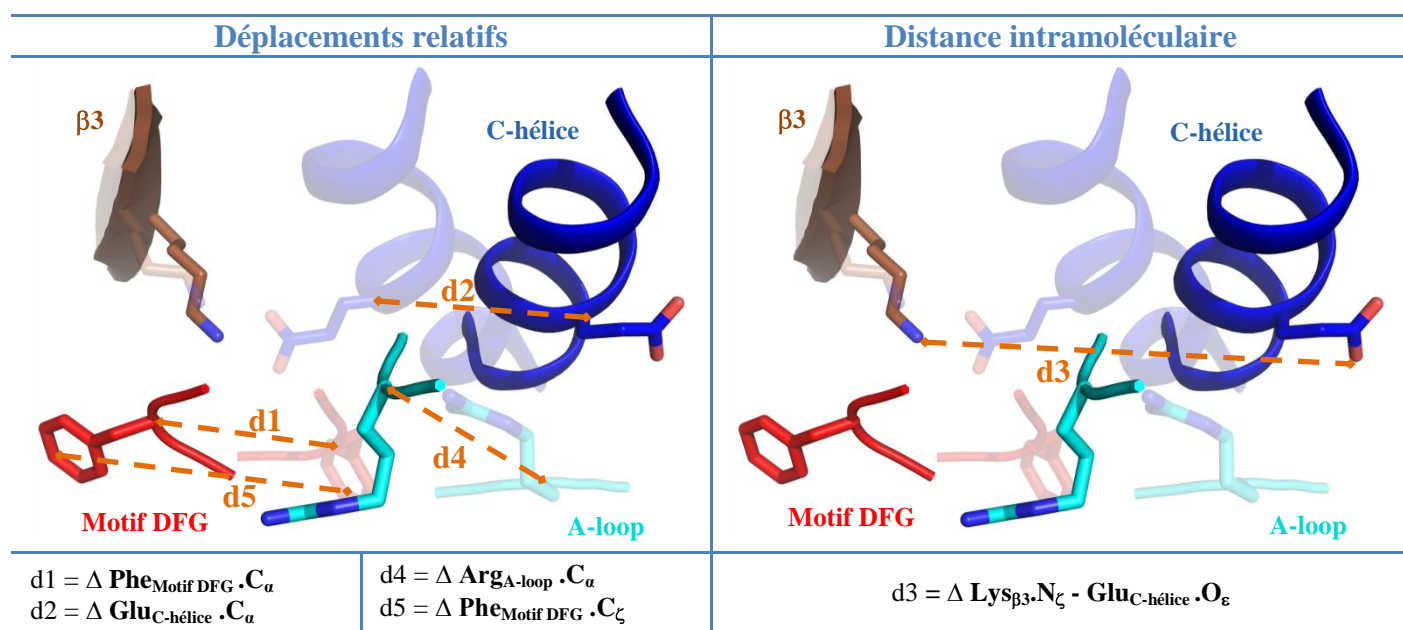


Figure 42 : Description des déplacements mesurés, après superposition des domaines catalytiques, entre les résidus équivalents de la référence (structure transparente) et de la chaîne protéique dont on veut assigner la conformation (structure opaque).



### 3.4 Première assignation de la conformation

Parmi les 14739 identifiants AC de kinases, toute espèce confondue, présents dans la base de données *Uniprot* (The UniProt Consortium, 2011), 554 identifiants ont des structures cristallographiques associées, se répartissant parmi 17 espèces dont une majorité de structures humaines. Cela représente un total de 3015 structures qui ont été rapatriées. Ces structures ont été découpées en chaînes protéiques (5928 chaînes) avant d'être superposées sur les sept structures de références, ces dernières étant en conformation *Active*. Les chaînes présentant moins de 50 résidus alignés ont été éliminées car elles ne correspondaient pas au domaine catalytique, mais à un domaine additionnel cristallisé. Au final, ce sont 2587 chaînes de domaine catalytique de kinases pour lesquelles on a voulu assigner une conformation.

Si l'on compare ces résultats de superposition à ceux obtenus en ne prenant qu'une seule structure de référence, on observe dans le second cas que 28% des chaînes protéiques ont moins de 100 résidus alignés avec la référence, contre 2% seulement avec sept références (Figure 43). L'usage de sept références permet d'améliorer à la fois le nombre de résidus alignés et la qualité de la superposition, évaluée par le critère *rmsd* (*root mean square deviation*) (Tableau 9 ; Figure 43). De plus, on observe une meilleure superposition au sein du lobe C-terminal, comparé au lobe N-terminal (Tableau 9).

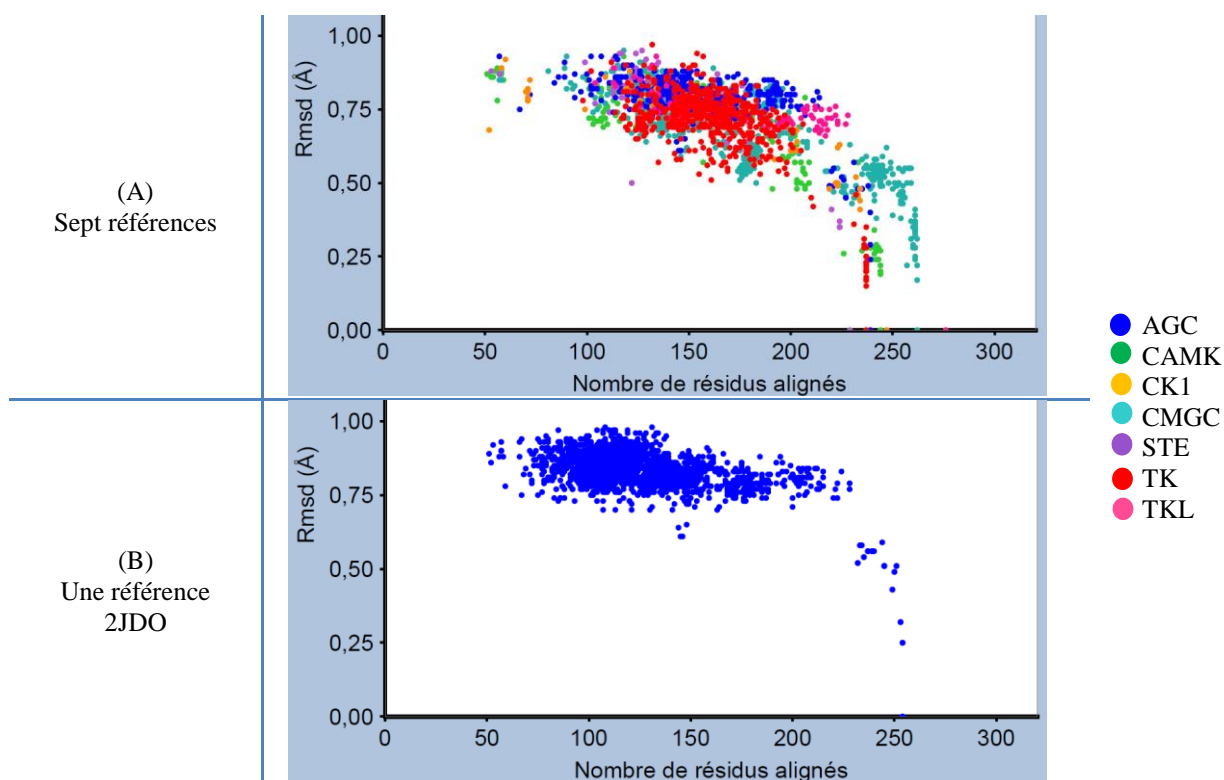


Figure 43 : Impact du nombre de références sur la distribution du *rmsd* en fonction du nombre de résidus alignés (atomes C $\alpha$  distants de moins de 1,5Å). Chaque point correspond à une chaîne protéique et la couleur correspond au groupe phylogénétique de la meilleure référence.

Nombre de références	Nombre moyen de résidus alignés	Rmsd moyen (Å)	Pourcentage moyen de résidus alignés dans le lobe N-terminal	Pourcentage moyen de résidus alignés dans le lobe C-terminal
7	162 (164)	0.72 (0.70)	44.34 (43.62)	68.95 (70.27)
1	117 (114)	0.83 (0.83)	40.12 (40.92)	47.94 (47.03)

Tableau 9 : Statistiques sur la qualité de la superposition et l'influence du nombre de références *Actives*. L'usage de sept références plutôt qu'une seule permet d'améliorer la qualité de la superposition. Les résidus sont considérés comme étant alignés si l'écart entre les C $\alpha$  est inférieur à 1.5Å. Les chiffres entre parenthèses correspondent aux calculs sur les chaînes protéiques humaines uniquement.

Un premier graphe (Figure 44), représentant le déplacement du résidu Glu<sub>C-hélice</sub> en fonction du résidu Phe<sub>Motif DFG</sub> (voir distances d1 et d2 dans la Figure 42 à la page 88), révèle trois populations de chaînes protéiques correspondant grossièrement aux trois conformations majeures : la conformation *Active*, localisée près de l'origine du graphique ; la conformation inactive *C-helix out*, située le long de l'axe X et éloignée de l'origine ; la conformation inactive *DFG out*, située le long de l'axe Y et éloignée de l'origine. Cette distribution a permis l'élaboration d'un premier arbre de décision pour l'assignation des conformations (Figure 45).

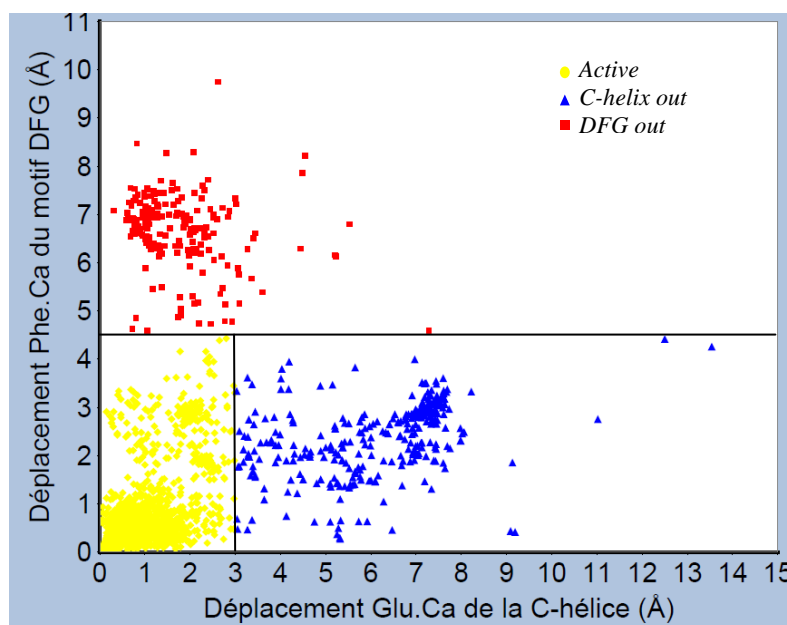


Figure 44 : Graphe représentant le déplacement  $\Delta$  Phe<sub>Motif DFG</sub>.C $\alpha$  en fonction du déplacement  $\Delta$  Glu<sub>C-hélice</sub>.C $\alpha$ . Chaque point correspond à une chaîne protéique et la couleur correspond à la conformation assignée avec le premier arbre de décision (Figure 45).

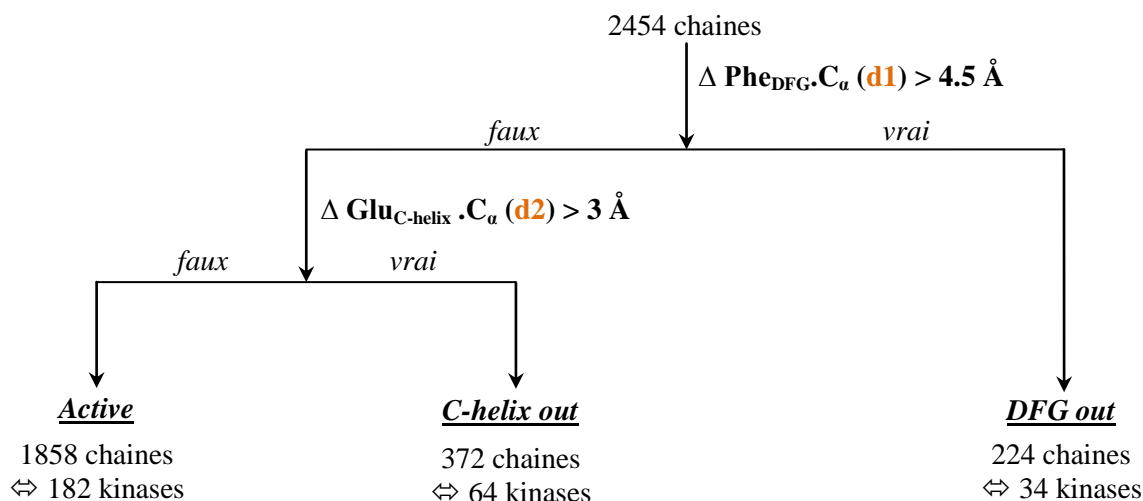


Figure 45 : Premier arbre de décision permettant une classification grossière de chaque chaîne protéique dans l'un des trois grands groupes de conformations.

Après analyse visuelle des chaînes *C-helix out* se trouvant proches de la population des chaînes *Actives*, il ressort que certaines structures 3D ont un lobe N-terminal décalé par rapport à la référence, induisant un déplacement du résidu  $\text{Glu}_{\text{C-hélice}}$ . Le résidu  $\text{Lys}_{\beta 3}$  est également déplacé dans ces structures et son interaction conservée avec le résidu  $\text{Glu}_{\text{C-hélice}}$  invalide l'assignation de la conformation inactive *C-helix out*. Cette observation a justifié la prise en compte du critère de la distance intramoléculaire  $\text{Lys}_{\beta 3}$ - $\text{Glu}_{\text{C-hélice}}$  (voir distance d3 dans la Figure 42 à la page 88) en supplément des critères de déplacements relatifs entre la chaîne et sa référence pour l'assignation de la conformation.

### 3.5 De trois à cinq conformations

L'analyse visuelle des chaînes classées dans la conformation *Active* révèle certaines particularités. Alors que dans la majorité des chaînes protéique *Actives*, la *A-loop* - lorsqu'elle est visible - présente un *rmsd* moyen de 4Å, d'autres chaînes a priori *Actives* se distinguent en présentant un déplacement de la *A-loop* beaucoup plus large (18Å en moyenne), et donc incompatible avec la fixation du substrat. Ces chaînes adoptent donc une conformation inactive. Ce déplacement de la *A-loop* n'a cependant pas conduit à l'assignation d'une des deux conformations inactives, dans la première version de l'arbre de décision, car ni la  $\text{Glu}_{\text{C-hélice}}$  ni la  $\text{Phe}_{\text{Motif DFG}}$  ne se sont suffisamment déplacées. Par contre, en prenant en compte le déplacement du résidu  $\text{Arg}_{\text{A-loop}}$  situé quatre résidus en aval de la  $\text{Phe}_{\text{Motif DFG}}$  (voir distance d4 dans la Figure 42 à la page 88), on peut distinguer les chaînes adoptant réellement la conformation *Active* de celles dont la *A-loop* ne permet pas un amarrage du substrat. Cela

m'a amenée à parfaire l'arbre de décision et à nommer ce nouvel état conformationnel « *A-loop out* ».

Par ailleurs, au sein de la conformation inactive *DFG out*, quelques chaînes présentent un déplacement important du résidu Glu<sub>C-hélice</sub>, accompagné d'une perte d'interaction entre les résidus Lys<sub>β3</sub> et Glu<sub>C-hélice</sub>. Autrement dit, ces chaînes présentent à la fois les caractéristiques d'une conformation inactive *DFG out* et celles d'une conformation inactive *C-helix out*. Cela m'a amenée à préciser encore l'arbre de décision et à nommer ce nouvel état conformationnel *C-helix and DFG out*.

Plus de détails quant à cette stratégie de classification ainsi qu'une analyse des diverses conformations obtenues font l'objet d'un article, inséré dans le paragraphe 3.6, qui sera soumis prochainement. Dans cet article, une étude plus approfondie de la conformation *A-loop out* dans le contexte des récepteurs de l'éphrine suggère qu'il pourrait s'agir d'un état intermédiaire entre la conformation inactive *DFG out* et la conformation *Active*. D'autre part, nous avons relevé un phénomène de glissement du feuillet situé dans le lobe N-terminal et propre à certaines kinases adoptant la conformation inactive *C-helix out*, dont la kinase CDK2.

### 3.6 Article en cours de soumission

## TITLE

Structural Analysis of the Conformational Diversity inside the Kinase Family: A Drug Design Perspective

## AUTHORS

Leproult E.<sup>a</sup>, Moras D.<sup>a</sup>, Poch O.<sup>a</sup>, Wurtz JM.<sup>a</sup>

## AFFILIATION

<sup>a</sup> Institut de Génétique et de Biologie Moléculaire et Cellulaire – CNRS (UMR 7104), INSERM (U964), 1 rue Laurent Fries, Université de Strasbourg, 67400 Illkirch, France.

## CONTACT INFORMATION

Jean-Marie Wurtz, Institut de Génétique et de Biologie Moléculaire et Cellulaire, 67400 Illkirch, France, +33 3 88 65 57 96, [wurtz@igbmc.fr](mailto:wurtz@igbmc.fr).

## RUNNING TITLE

From Kinase Conformational Analysis To Drug Design

## SUMMARY

Protein kinases have been revealed to be a challenging class of therapeutic targets for the design of efficient and selective inhibitors. The best strategy consists in the inhibition of the kinase target adopting an inactive conformation. But up to now, it is not known which kinases are amenable to such strategy. The methodology presented here aims to classify 1744 crystal structures of 221 distinct kinases into one active and four inactive conformations. Moreover, peculiar features have also been annotated for each kinase structure, giving further insight for drug design. All together, the present study reveals that 40% of crystallized kinases adopt an inactive conformation, suggesting that a large proportion of kinases may be targeted in an inactive conformation. This data has already been successfully exploited to design an allosteric covalent inhibitor of KIT kinase.

## HIGHLIGHTS

- 3D structure classification of kinase catalytic domains into five conformations
- 3D structure annotation to highlight remarkable motions
- 40% of crystallized kinases adopt inactive conformations
- A continuum of conformations may be targeted by inhibitors

## INTRODUCTION

The protein kinase family is one of the largest families in the human proteome, containing 518 members (Manning et al., 2002). Protein kinases have a fundamental role in signal transduction pathways, by catalysing the phosphorylation of a protein substrate. It has been shown that in order to transfer the phosphate to the substrate protein, the kinase adopts the so-called active conformation (Figure 1A) (Zheng et al., 1993). Aberrant kinase activity has been observed in many diseases (Lahiry et al., 2010; Santarius et al., 2010) but pharmaceutical companies were reluctant to consider the members of this family as therapeutic targets due to the selectivity problem among kinases, as well as competition with ATP (Knight et Shokat, 2005; Karaman et al., 2008). The trend has changed in recent years, and kinase inhibition has become a major area for therapeutic intervention since the success of imatinib in the treatment of leukemia (Schindler et al., 2000). This compound was the first kinase inhibitor that reached the market; it exhibits high selectivity and affinity for its target kinase ABL. The crystal structure of the imatinib/ABL complex revealed a kinase conformation that differed significantly from the previously observed ATP/kinase complexes (Figure 1A) (Schindler et al., 2000). In contrast to the ATP/kinase complexes (Zheng et al., 1993), ABL adopts a conformation that hampers the substrate phosphorylation, and as such, has been described as being an inactive conformation (Figure 1C) (Schindler et al., 2000). Imatinib in this complex binds into a new pocket that overlaps with the ATP binding site. To date, several strategies have been devised to develop kinase inhibitors giving rise to different compound classes encompassing type I or type II inhibitors, targeting the active or inactive conformations of kinases, respectively (Zhang et al., 2009). More recently, a new generation of compounds has been developed to overcome the resistance problem encountered in patients treated with the first generations of inhibitors (Yun et al., 2008). These new molecules bind irreversibly to their target by reacting with a cysteine residue. BIBW2992 is the most advanced ligand of this family of inhibitors; it reached recently phase III clinical trials for the treatment of lung cancer, by covalently targeting the active conformation of the EGFR kinase (Subramaniam et Hwang, 2011).

In the light of these data, our aim was to design a new class of compounds that combines the advantages of an irreversible inhibitor with those of an inhibitor targeting an inactive conformation. Such a strategy required the exploitation of the wealth of structural data available in the protein databank (PDB) to classify the structures into active or inactive complexes and to localize the putative reactive cysteines in the respective pockets. Such an analysis will provide a lower bound estimate of the proportion of kinases susceptible to

adopting an inactive conformation, this number will indicate whether this strategy is applicable to a larger set of kinases.

To date, two inactive conformations of the catalytic domain have been described, that prevent substrate binding by displacing the activation loop: the “C-helix out” and the “DFG out” conformations (Figure 1B,C) (Schindler et al., 2000; De Bondt et al., 1993). Moreover, in the “C-helix out” conformation, the C-helix of the N-terminal lobe moves away from the ATP binding site, disrupting an ionic interaction between catalytic residues (K72<sub>β3</sub> - E91<sub>C-helix</sub>) involved in phosphate transfer (Figure 1B; residue numbering will refer to cAPK kinase nomenclature in the rest of the article) (De Bondt et al., 1993). In the “DFG out” conformation, a conserved motif, the so-called DFG motif (D184-F185-G186), located at the beginning of the A-loop, rotates by ~180° around the main chain (Figure 1C) (Schindler et al., 2000). In consequence, the phenylalanine (F185<sub>DFG motif</sub>) is dislodged from its position and points towards the ATP binding site, hindering ATP and substrate binding.

A few studies aimed at classifying kinase crystal structures in the three previously described conformations have already been published (Jacobs et al., 2008; Brooijmans et al., 2010). Similar to our study, conserved secondary structure elements are identified among the kinase family, such as the DFG motif, the A-loop, the hinge region, the C-helix (Figure 1A), and the displacement of such motifs are used to discriminate between the active and inactive conformations. These displacements can be either measured with respect to a reference structure (relative displacement) or can be measured internally to the kinase (intramolecular distance). Despite this similarity, the two previous studies differ significantly from our approach both at the methodological level and the focus of the study. In the first study, in 2007, they used one reference kinase structure to classify the family whereas we introduce seven references for each of the phylogenetic groups described by Manning (Jacobs et al., 2008). These seven reference structures have been introduced in order to take into consideration the plasticity of this large kinase family. Moreover, the initial classification of the kinase structures that they proposed needs to be updated, since then the number of deposited crystal structures has doubled (Figure SI 1). More recently, Brooijmans et al. (2010) published a study where they focused on the interaction between the ligand and the kinase, as they wanted this data to be integrated into their kinase knowledge database. The kinase classification was a required property for their system. It is worth to note that their study and the present work have been conducted independently over the same period, and give rise to a different classification of the kinase family. This difference is due to an alternate

strategy in defining the features that discriminate the active conformation from the inactive ones.

Our analysis of the kinase conformations has already been exploited for the design of a covalent inhibitor targeting an inactive conformation that has been published recently (Leproult et al., 2011). In the present paper, as well as describing the classification strategy, our study reveals a more diverse structural landscape than expected. Such observations could be of importance, on one hand, for the understanding of the kinase structure/function relationship and, on the other hand, for drug design. Among the relevant features described in the present study, it is worth underlining the displacement of the  $\beta$ -sheet from the N-terminal domain that modifies the size and shape of the ATP binding site, as well as the assignment of a novel inactive conformation termed “A-loop out”.

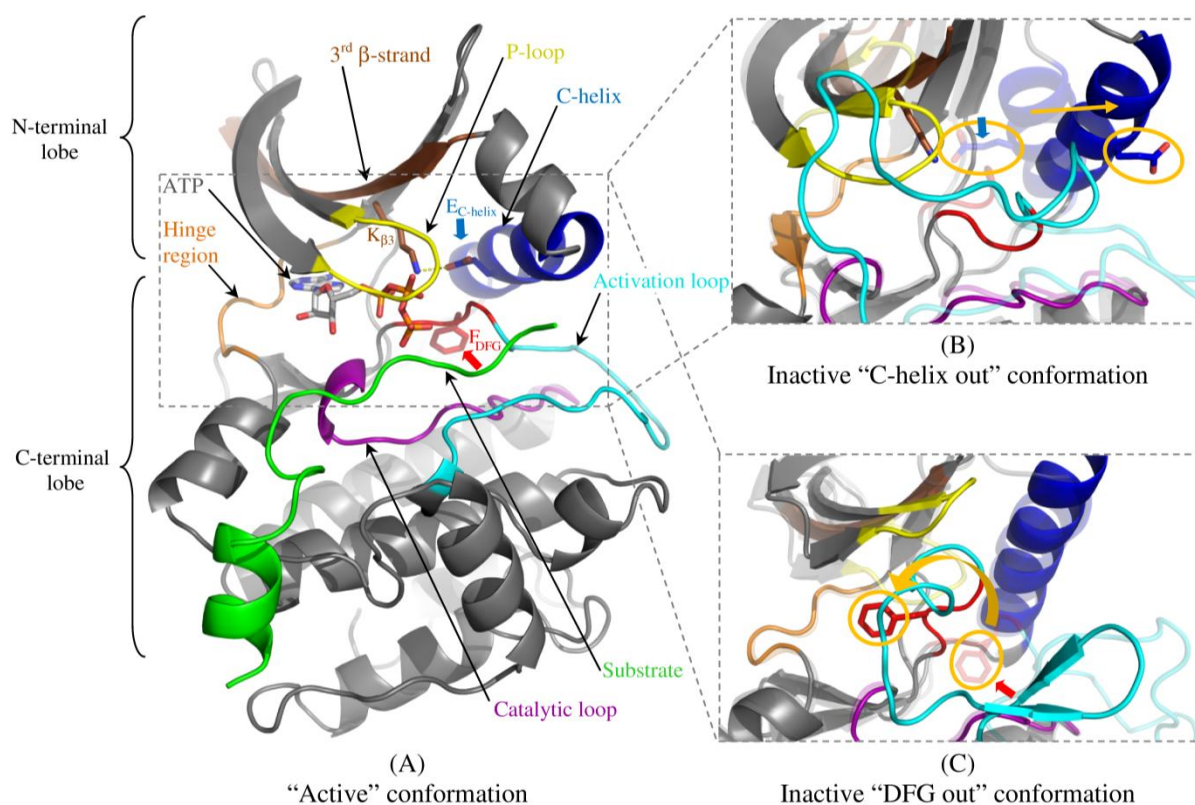


Figure 1: The kinase catalytic domain and its secondary structure elements. (A) cAPK kinase (PDB code: 1ATP) in the “Active” conformation. Color scheme in this figure is kept throughout this article: the P-loop is in yellow; the third  $\beta$ -strand of the  $\beta$ -sheet, bearing the catalytic lysine  $K72_{\beta 3}$ , is in brown; the C-helix, bearing the catalytic glutamic acid residue  $E91_{C\text{-helix}}$  (marked with a blue arrow), is in blue; the catalytic loop is in purple; the activation loop is in cyan; the DFG motif ( $F185_{\text{DFG motif}}$ , marked with a red arrow) starting the activation loop is in red; the protein substrate is in green; ATP, located in its binding site, is represented in sticks. Oxygen and nitrogen atoms are color-coded in red and blue, respectively. Carbons are color-coded according to the secondary structure element color. (B) Close up view of CDK2 kinase catalytic site: superimposition of the inactive “C-helix out” conformation (PDB code: 1PXN) on the “Active” one (PDB code: 3BHV, translucent). In the “C-helix out” conformation, the catalytic glutamic acid residue ( $E91_{C\text{-helix}}$ , circled in orange in the two structures) is moved away from the ATP binding site, due to the disruption of an ionic interaction between  $E91_{C\text{-helix}}$  and  $K72_{\beta 3}$ , involving a general displacement of the C-helix (orange arrow). Except for the DFG motif, the activation loop is highly displaced. (C) Close up view of KIT kinase catalytic site: superimposition of the inactive “DFG out” conformation (PDB code: 1T46) with the “Active” one (PDB code: 1PKG, translucent). The phenylalanine of the DFG motif is circled in orange in the two structures, in order to highlight its large displacement in the “DFG out” conformation, due to a general movement of the activation loop (orange arrow).



## RESULTS

### **Kinase classification strategy**

In total, 1744 protein kinase crystal structures containing the catalytic domain of 221 different protein kinases, alone or in complex with partner proteins, have been extracted from the Protein Data Bank (Berman et al., 2000). This dataset gives rise to 2587 kinase chains shared among 17 species, of which 2073 chains extracted from 1414 crystal structures correspond to 172 human protein kinases.

*Kinase superimposition.* Our strategy relies on the superimposition of each chain onto seven references adopting the active conformation and which are representative for each phylogenetic group described by Manning et al. (2002) (Table 1A). For all the chains, only the best superimposition of the kinase domain with one of the references was kept (See Method Section). All the superimpositions have an rmsd (root mean square deviation) below 1Å (Figure SI 2) and 98% of the chains had more than 100 residues aligned over the whole domain (~270 residues). In the superimposition process, on average, 69% of the residues belonging to the C-terminal lobe (125 residues) matched, whereas, on average, only 45% of the residues aligned in the N-terminal lobe (37 residues) which is known to be more flexible (Figure SI 2) (Taylor et Radzio-Andzelm, 1997). It is interesting to note that most of the chains preferentially aligned on the reference corresponding to their own phylogenetic group (Table 1B), except for the CAMK group where 50% of the chains aligned best with the AGC reference. The proximity in the phylogenetic tree of the CAMK and AGC members most likely explains the structural similarity. The “Others” group, spread evenly onto the phylogenetic kinase tree, displayed the largest selection of references (Table 1B).

PDB code (Chain)	Gene name	Kinase phylogenetic group	Resolution (Å)	Catalytic Domain (residue numbering)	Catalytic Domain without the P-loop	Lys $\beta_3$	Gluc-helix	Phe <sub>DFG</sub> motif	Arg <sub>A-loop</sub>
2JDO (A)	AKT2	AGC	1.80	152-405	167-405	K 181	E 200	F 294	K 298
1JKS (A)	DAPK1	CAMK	1.50	13-271	28-271	K 42	E 64	F 162	H 166
2CMW (A)	CSNK1G1	CK1	1.75	46-313	65-313	K 73	E 87	F 188	K 192
3BHV (A)	CDK2	CMGC	2.10	4-282	19-282	K 33	E 51	F 146	R 150
2HY8 (I)	PAK1	STE	2.00	270-517	289-517	R 299	E 315	F 408	A 412
3LCK (A)	LCK	TK	1.70	245-496	260-496	K 273	E 288	F 383	R 387
2QLU (A)	ACVR2B	TKL	2.00	190-480	205-480	K 217	E 230	F 340	V 344

(A)

SUPERIMPOSED CHAINS				
Group	Number of chains	Reference group	Number of chains	Percentage of group
AGC	144	AGC	107	74
		CAMK	35	24
		STE	2	2
CAMK	366	AGC	182	50*
		CAMK	181	49*
		STE	3	1*
CK1	18	CK1	18	100
CMGC	632	AGC	4	1
		CMGC	624	99
		STE	4	0
STE	80	AGC	12	14
		CAMK	3	4
		CMGC	8	10
		STE	54	68
		TK	3	4
TK	543	CAMK	2	0
		CMGC	16	3
		TK	518	95
		TKL	7	2
TKL	131	CAMK	14	11
		STE	2	2
		TK	1	0
		TKL	114	87
Others	159	AGC	52	32
		CAMK	44	27
		CK1	1	1
		CMGC	49	31
		STE	11	7
		TK	2	1
Total	2073		2073	

(B)

Table 1: The seven reference structures in the “Active” conformation. (A) One reference per kinase phylogenetic group has been chosen, based on the adopted conformation and structure quality. The designation of relevant residues used for the conformation classification is given. (B) Summary of references that have been selected by the human chains according to the superimposition quality. Most of the time, chains preferentially aligned on the reference corresponding to their own phylogenetic group. An exception is observed for chains belonging to the CAMK group since only 49% selected the CAMK reference. 50% of CAMK chains preferentially superimposed onto the AGC reference. This structural similarity is mostly due to proximity between the two groups in the phylogenetic tree. The 1% remaining CAMK chains superimposed onto the STE reference (see numbers highlighted with \*).

*Kinase conformation assignment with two features.* The superimposed kinase domains form a continuum of structures ranging from the “Active” conformation (Zheng et al., 1993) (Figure 1A) to the inactive “C-helix out” and “DFG out” conformations (Schindler et al., 2000; De Bondt et al., 1993) (Figure 1B,C) that have already been described in the literature.

To detect these different conformations, structural features were defined that measure the displacements of the C-helix and the DFG motif with respect to the “Active” reference. These two features, the displacements of the phenylalanine residue in the DFG motif ( $\Delta\text{Phe}_{\text{DFG motif}}$ ; F185<sub>DFG motif</sub>) and the glutamic acid residue in the C-helix ( $\Delta\text{Glu}_{\text{C-helix}}$ ; E91<sub>C-helix</sub>) (Figure 2A), are sufficient to reveal the three major conformational groups: “Active”, “C-helix out” and “DFG out” as illustrated in the 2D graph (Figure 3A). In order to make the kinase conformation assignation automatic, cut-offs were assigned according to the distribution of these displacements (Figure 2B). Additional features were required to fine-tune the classification of the kinase chains as illustrated below.

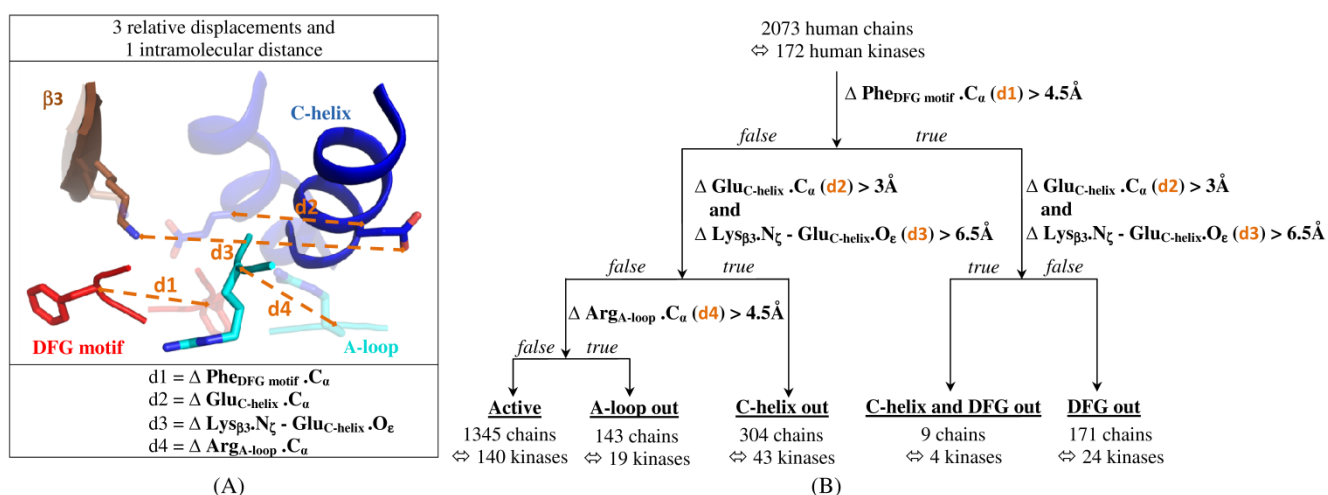


Figure 2: Structural features used for the kinase conformation classification. (A) Three displacements relative to the active reference (d1, d2, d4) and one intramolecular distance (d3) are automatically measured to assign the conformation. The secondary structure elements of the active reference are represented in translucent colors. The secondary structure elements of a kinase structure, that has been superimposed onto its reference structure, are represented in opaque. (B) Decision tree used to assign the kinase conformation adopted by any 3D chain handled in the present study.

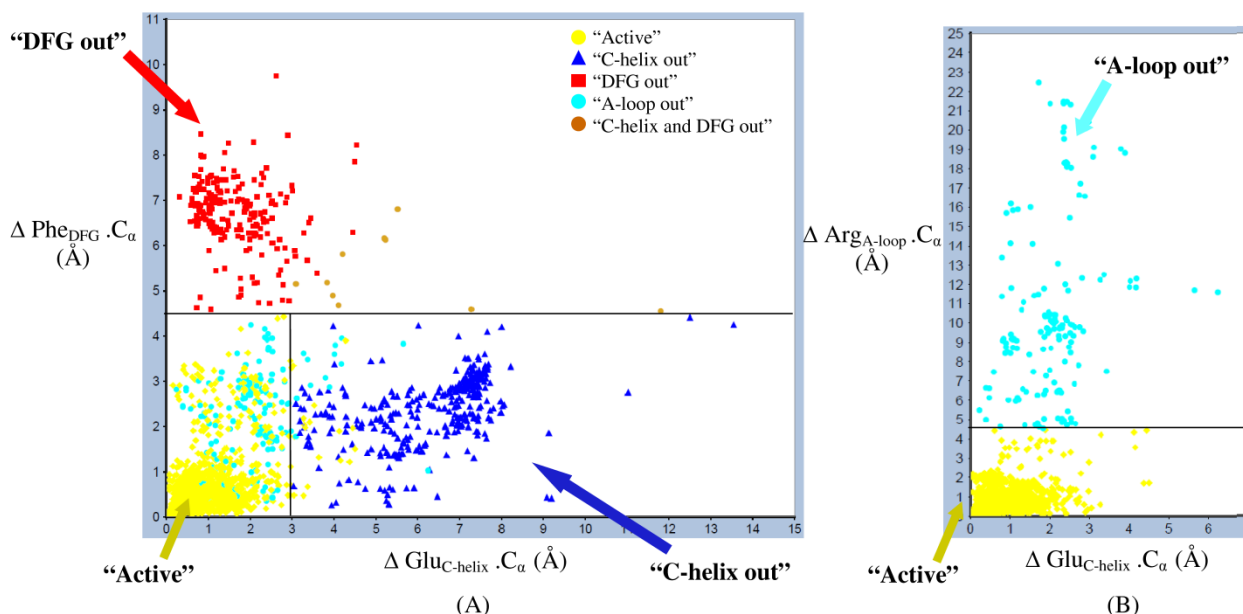


Figure 3: Conformational landscape of kinases revealed by 2D graphs. (A) This 2D graph represents each kinase chain by a point where the X and Y axes correspond to the displacements of  $\text{Glu}_{\text{C-helix}} \cdot \text{C}_\alpha$  and  $\text{Phe}_{\text{DFG motif}} \cdot \text{C}_\alpha$ , respectively. Depending on their localization on the graph, chains have been assigned to the “Active”, “C-helix out” or “DFG out” conformations and color-coded in yellow, blue or red respectively. This conformation assignment follows the rules given in Figure 2. (B) To distinguish the “A-loop out” conformation from the “Active” conformation, an additional feature has been introduced ( $\text{Arg}_{\text{A-loop}} \cdot \text{C}_\alpha$ ; see Figure 2).

*Additional features to improve the classification.* Most of the chains belonging to the “Active” cluster present an ordered activation loop (A-loop) that adopts a similar position (mean A-loop rmsd of about 4Å) except for some members that exhibit an activation loop that underwent a large displacement (mean activation loop rmsd of about 18Å). Such a displacement is incompatible with the binding of the substrate, suggesting that these members have been assigned to the wrong “Active” cluster with the two criteria only. Consequently, the displacement of this loop has been introduced as an additional criteria (residue located four residues upstream to the residue  $\text{Phe}_{\text{DFG motif}}$  that is mainly an arginine or a lysine ( $\Delta \text{Arg}_{\text{A-loop}}$ ;  $\text{K189}_{\text{A-loop}}$ ; Figure 2) and all these misclassified “Active” chains are relabeled as a new inactive conformation: the “A-loop out” conformation (Figure 3). An in depth analysis will be described in a later paragraph.

Some chains were also misclassified as “C-helix out” conformation. Among these chains, the glutamic acid residue ( $\text{E91}_{\text{C-helix}}$ ) exhibits a large displacement, but the residue was still interacting with the conserved lysine residue from the third  $\beta$ -strand ( $\text{K72}_{\beta 3}$ ), suggesting a correlated motion between the C-helix and the  $\beta$ -sheet. The intramolecular distance between the lysine and the glutamic acid residues has been taken as an additional criteria to assign the kinase conformation ( $\Delta \text{Lys}_{\beta 3} \cdot \text{Glu}_{\text{C-helix}}$ ; Figure 2). Note that  $\beta$ -sheet displacements will be analysed in more details in the next section.

Finally, an additional inactive conformation has been introduced that exhibits both “DFG out” and “C-helix out” features and has been named “C-helix and DFG out” conformation (Figure 2; Figure 3A).

### **Kinase classification and remarkable conformations**

*Five conformations for the kinase family.* In the rest of this article the numbers are given for human kinases only, the numbers for all species are given in Table SI 1. Out of the 2587 chains, 2073 correspond to human kinases. The classification assigned 1345 chains as “Active” that correspond to 140 different kinases (Figure 2). Similarly, 304 and 171 chains were classified as “C-helix out” and “DFG out” respectively, which correspond to 43 and 24 distinct human kinases, respectively (Figure 2; Figure SI 3). 19 different kinases adopt the “A-loop out” conformation, observed in 143 chains (Figure 2). The rare “C-helix and DFG out” conformation has been assigned to 9 chains representing 4 different kinases (Figure 2). Note that a kinase can be present in more than one conformational cluster (Figure SI 3; Figure 4). A total of 101 chains could not be assigned because of missing feature residues in the structures.

The five conformations “Active”, “C-helix out”, “DFG out”, “A-loop out” and “C-helix and DFG out” were mapped onto the human kinase phylogenetic tree (Figure SI 4; Figure SI 5). The mapping reveals that the tyrosine kinases (TK) are the most investigated phylogenetic group in the structure field and many of them adopt inactive conformations. Moreover, each inactive conformation is not specific to a kinase phylogenetic group.

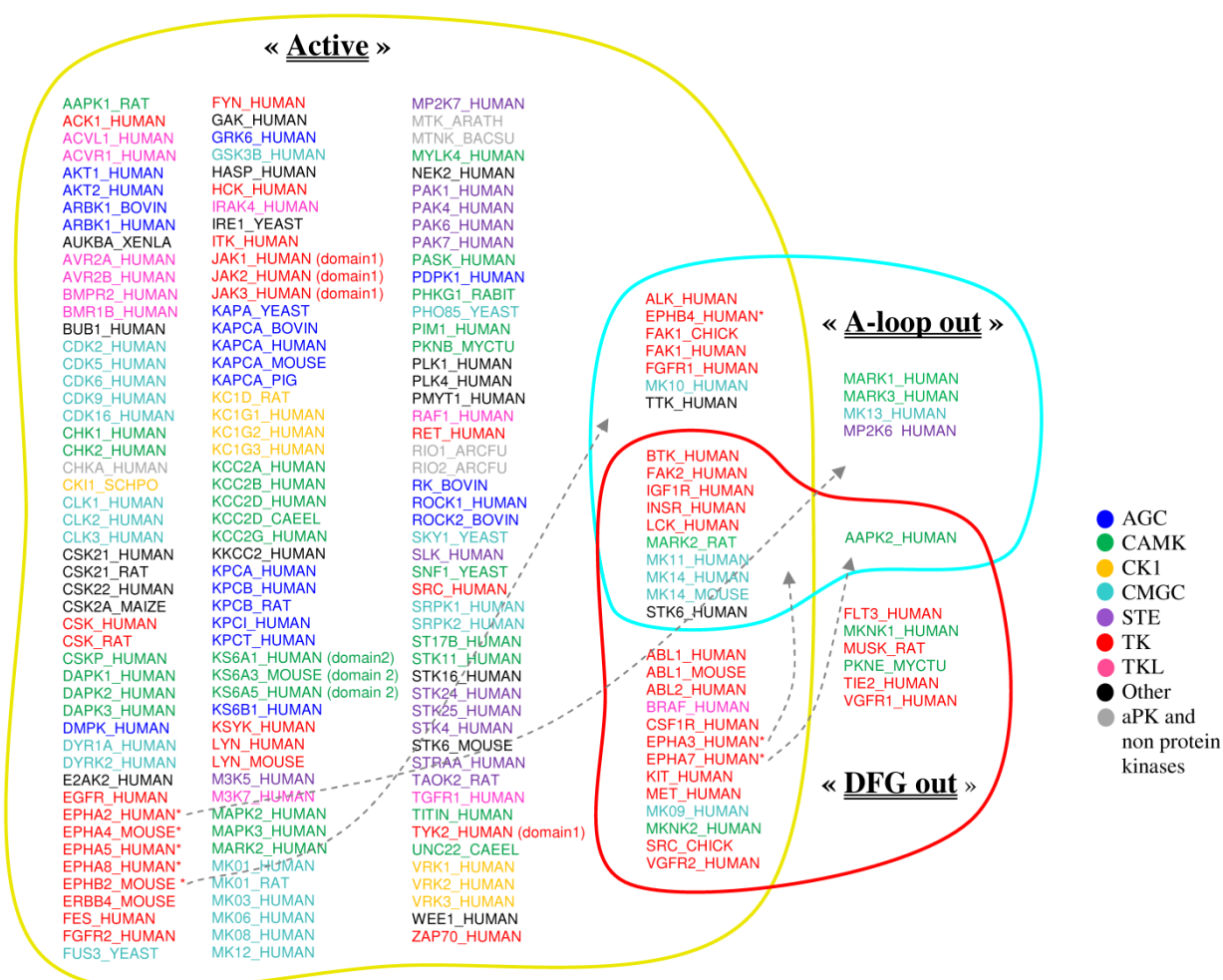


Figure 4: Distribution of kinases among three conformations, “Active”, “A-loop out” and “DFG out”, illustrated by a Venn diagram. Kinases are present in this diagram if a structure has been assigned to one of the three conformations. Note that only three conformations have been selected for clarity. An additional diagram is available in supplemental information for the “Active”, “C-helix out” and “DFG out” conformations (Figure SI 3). The dashed arrows represent members of the Ephrin receptor family (marked with \*) that have been manually reassigned to the “A-loop out” conformation, depending on the activation sensor position.

“C-helix out” conformation:  $\beta$ -sheet displacement in the N-terminal lobe. Compared to the “Active” conformation, 18 distinct human kinases adopting the “C-helix out” conformation, including BTK kinase (Figure SI 6), present a C-helix sliding away from the ATP binding site and perpendicular to the conserved  $\beta$ -sheet lining the ATP binding site ( $\Delta\beta 3 < 1.5\text{\AA}$ ; Figure 5A). On the contrary, 24 other kinases adopting the “C-helix out” conformation exhibit a correlated displacement between the C-helix and the  $\beta$ -sheet, the later is largely shifted towards the ATP binding site, especially the third  $\beta$ -strand which is displaced by more than  $3\text{\AA}$  (Figure 5A). Interestingly, such a displacement is concomitant not only with the repositioning of the C-helix but also with the hinge region, resulting in a reduced ATP binding site, as observed among EGFR (Figure SI 6) and CDK2 kinase structures (PDB code: 1PXN (Wang et al., 2004); Figure 5). For CDK2, this effect is emphasized due to its shorter

hinge region (5 residues compared to 6 for most kinases). Consequently, the CDK2 complexes in the “C-helix out” conformation exhibit quite small ligands as compared to other members. Interestingly, the “C-helix out” CDK2/staurosporine complex (PDB code: 1AQ1 (Lawrie et al., 1997)) exhibits an intermediary conformation, with the  $\beta$ -sheet and the hinge region showing a smaller displacement ( $\Delta\beta_3 \approx 2.3\text{\AA}$ ), probably due to the bulkiness of the ligand (Figure 5B). Recently, two CDK2 structures adopting the “C-helix out” conformation have been crystallized with allosteric inhibitors bound in a pocket located between the C-helix and the third  $\beta$ -strand, and adjacent to the ATP binding site (PDB codes: 3PXF, 3PXZ (Betzi et al., 2011)). In consequence, the  $\beta$ -sheet is constrained and, this time, no displacement towards the ATP binding site is observed. Together the data suggests that a continuum of conformations could be achieved by exploiting the size, shape and type of ligands.

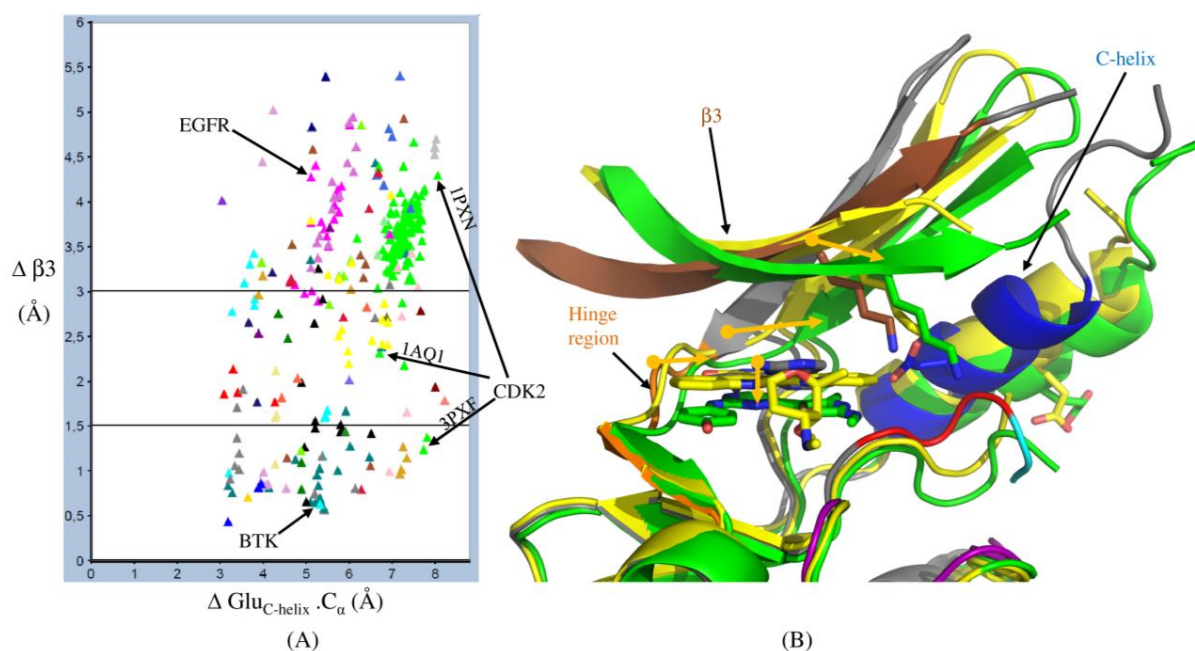


Figure 5: N-terminal lobe displacements observed for kinases adopting the “C-helix out” conformation. (A) A 2D graph showing the displacements of the  $\text{Glu}_{\text{C-helix}}\text{-C}_\alpha$  and the third  $\beta$ -strand along the X and Y axes respectively. Triangles of the same color correspond to chains of the same kinase. Some kinases such as BTK (see Figure SI 6) have only the C-helix moved away when adopting the “C-helix out” conformation ( $\Delta\beta_3 < 1.5\text{\AA}$ ), in contrast to kinases such as EGFR (see Figure SI 6) and CDK2 that exhibit also a displacement of the  $\beta$ -sheet towards the ATP binding site ( $\Delta\beta_3 > 3\text{\AA}$ ). (B) Most “C-helix out” CDK2 chains have a  $\beta$ -sheet displaced towards the ATP binding site and present an hinge region that is significantly displaced (see orange arrows; PDB code: 1PXN, in green), when compared to the “Active” reference (PDB code: 3BHV, color scheme as in Figure 1B). Such displacements induce the small size of the ATP binding site and shift the ligand in its binding site (orange arrow on the ligand). One CDK2 “C-helix out” structure adopts an intermediate conformation (PDB code: 1AQ1, in yellow) where the  $\beta$ -sheet and the hinge region nearly didn’t move ( $\Delta\beta_3 \approx 2.3\text{\AA}$ ), mostly due to the bulky staurosporine bound into the ATP binding site. Note that the two first  $\beta$ -strands and part of the activation loop are not shown for clarity.

*The inactive “A-loop out” conformation.* The chains belonging to the “A-loop out” cluster had been initially assigned to the “Active” conformation, but due to the large displacement of

the activation loop (after the DFG motif) which hampers the substrate binding, these chains have been relabeled. This loop, in many cases, forms extensive interactions with the C-helix (MARK1 kinase, PDB code: 2HAK (Marx et al., 2006); FAK1 kinase, PDB code: 3BZ3 (Roberts et al., 2008)). For the remaining cases, this loop adopts a conformation with the DFG motif similar to the “Active” conformation (MK14 kinase, PDB code: 1BMK (Wang et al., 1998)), but with the rest of the loop blocking substrate binding in a similar way to the “DFG out” conformation. It is interesting to note that the phenylalanine of the DFG motif (F185<sub>DFG motif</sub>) adopts different orientations among the chains that are classified as “A-loop out”.

In summary, a total of 143 chains adopt the “A-loop out” conformation and concern 19 human kinases. It is worth to note that more than half of these kinases also adopt the inactive “DFG out” conformation (Figure 4).

#### **“A-loop out” conformation: an intermediate state**

*The Ephrin receptor family as a case study.* Four of the 19 human kinases adopting the “A-loop out” conformation (EPHB4, FGFR1, IGF1R and INSR) are members of the receptor tyrosine kinase family (RTK) that are also observed in the “Active” and “DFG out” conformations (Figure 4; Figure SI 5). Focusing on EPHB4 (Ephrin type-B receptor 4 kinase) and other members of the Ephrin receptor family reveals that they exhibit an extra domain, the juxtamembrane (JM), that regulates their catalytic activity. The presence of the juxtamembrane indirectly prevents the activation loop from adopting an “Active” conformation (Davis et al., 2008) by pushing the activation sensor (Y<sub>Activation sensor</sub>; Y742 in EPHA3; Figure 6A). Upon phosphorylation of the juxtamembrane (Y<sub>JM</sub>; Y596 in EPHA3; Figure 6A), a major rearrangement of the sensor residue occurs that allows the active conformation (Figure 6A,B). For many Ephrin receptor structures, part of the activation loop containing the Arg<sub>A-loop</sub> residue was missing but the position of the activation sensor suggested that the activation loop cannot adopt an “Active” position (Figure 6B), so these structures were relabeled as “A-loop out” conformation (Figure 4; among them, EPHA3 with PDB code: 2QOQ (Davis et al., 2008); Figure 6A). A similar mechanism, where the activation sensor is pushed away, is observed for the “DFG out” conformation for the same kinase (PDB code: 3DZQ) but with a much larger displacement of the Phe<sub>DFG motif</sub> and the activation loop (Figure 6C).

The structural proximity of the “Active” and “A-loop out” conformations on one side and the “A-loop out” and “DFG out” conformations on the other side suggest that, for the RTK family



at least, the “A-loop out” conformation represents an intermediate conformation between the “Active” and the inactive “DFG out” conformations.

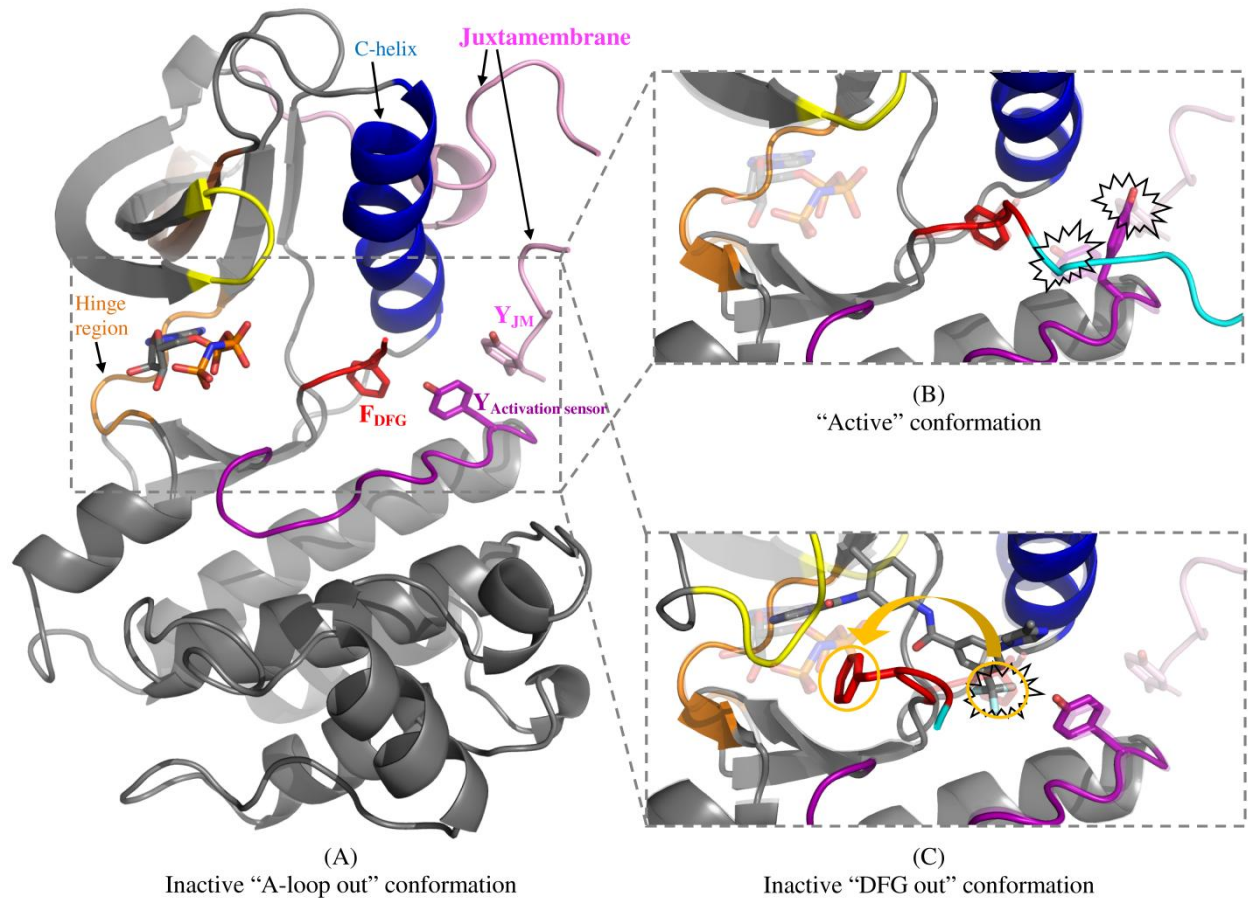


Figure 6 : Illustration of the “A-loop out” intermediate state: a case study with Ephrin type-A receptor 3 kinase (EPHA3). (A) Catalytic domain of EPHA3 in the “A-loop out” conformation (color scheme as in Figure 1) and the juxtamembrane region (JM; in pink) (PDB code: 2QOQ). The position of the “activation sensor” tyrosine ( $Y_{\text{Activation sensor}}$  or Y742, in purple sticks) located on the catalytic loop is constrained by the presence of an unphosphorylated tyrosine located on the JM ( $Y_{\text{JM}}$  or Y596, in pink sticks). The ATP ligand is bound in the “A-loop out” structure. (B) Close up view of the catalytic site of EPHA3 in two different conformations, “Active” (color scheme as in Figure 1; PDB code: 3FY2) and “A-loop out” (color scheme as in Figure 1 and translucent; PDB code: 2QOQ). The  $F_{\text{DFG}}$  residue adopts the same position in the two conformations. Note the steric clash between  $Y_{\text{JM}}$ , belonging to the “A-loop out” conformation, and  $Y_{\text{Activation Sensor}}$ , belonging to the “Active” conformation. Another steric clash exists between  $Y_{\text{Activation Sensor}}$ , belonging to the “A-loop out” conformation, and the position of the activation loop in the “Active” conformation. (C) Close up view of the EPHA3 catalytic site in two different inactive conformations, “A-loop out” (color scheme as in Figure 1 and translucent; PDB code: 2QOQ) and “DFG out” (color scheme as in Figure 1; PDB code: 3DZQ). The  $Y_{\text{Activation Sensor}}$  residue adopts the same position in the two inactive conformations. The ligand in the “DFG out” structure is bound into an extended ATP binding site. Such a binding mode cannot exist in the “A-loop out” conformation due to a steric clash between the ligand and the beginning of the activation loop. The largest displacement of the activation loop occurs in the “DFG out” conformation. All together, the “A-loop out” conformation could be an intermediate state between the “Active” and the “DFG out” conformations.

*A-loop out intermediate state outside the RTKs.* Furthermore, kinases outside the RTK group (FAK1 and FAK2), but nevertheless close to this group in the phylogenetic tree (TK group), also exhibit the same three conformations. These kinases (FAK1 and FAK2) do not exhibit any region equivalent to the juxtamembrane. Thus, another mechanism, different from the one observed for the RTK family, is required to displace the activation loop. A similar observation

is made for members more distant in the phylogenetic tree and belonging to the CMGC kinase group (MK11 and MK14). These kinases exhibit a C-terminal extension, reminiscent of the N-terminal juxtamembrane for RTKs, involved in an auto-inhibition mechanism requiring the so-called L16-loop and its unphosphorylated tyrosine (Diskin et al., 2007). Together these observations strongly suggest that the inactive “A-loop out” conformation represents an intermediate conformation between the “Active” and the inactive “DFG out” conformations that holds for kinase members beyond the RTK family.

## DISCUSSION

The classification strategy presented here aims to classify all the crystal structures of kinase catalytic domains into active and inactive conformations. The kinases are superimposed to seven “Active” references representative for the phylogenetic groups defined by Manning et al. (2002) and the best of them is retained for the classification (Table 1). With four criteria (Figure 2), 166 distinct human kinase catalytic domains (1972 chains belonging to 1358 structures) were classified into five major conformations: “Active”, “C-helix out”, “DFG out”, “A-loop out” and “C-helix and DFG out” (Figure 2B). A majority of human kinases are observed in the “Active” conformation (140 distinct kinases; 84% of assigned kinases; Figure SI 3), consistent with the idea that any kinase can adopt the “Active” conformation. Interestingly, 67 kinases are classified into one of the four inactive conformations, they represent 40% of assigned kinases. Among them, two new conformations (“C-helix and DFG out”; “A-loop out”) have been characterized in this study. The “C-helix and DFG out” conformation is a combination between the inactive “C-helix out” and “DFG out” conformations, and is a small class with only 2% of the assigned kinases adopting it (NEK2, CDK6, MET, PDPK1). On the other hand, the “A-loop out” conformation is more common (11% of assigned kinases; 143 chains; Figure 4). In this conformation, the activation loop undergoes a large displacement allowing ATP to bind, but hampering substrate binding (hence the name “A-loop out” conformation).

The comparison between the active and inactive conformations reveals that the “A-loop out” conformation represents an intermediate state that is close, on one side, to the “Active” conformation, and on the other side, to the “DFG out” conformation (Figure 6). This observation is relevant for the Ephrin family and other members of the TK phylogenetic group. Furthermore, it can be extended to more distant members in the phylogenetic tree such

as MK14 (CMGC group), suggesting that some members of the kinase family may adopt a continuum of conformations ranging from the active to inactive conformations.

In addition to the class assignation, inside each conformational class, kinase structures have been annotated according to structural peculiarities that may be relevant for drug design approaches. Two structural peculiarities are particularly relevant, the  $\beta$ -sheet displacement inside the “C-helix out” class, and the multiple orientations of the phenylalanine in the DFG motif. In the latter case, depending on the orientation of the residue, it may participate to an additional pocket in which ligands can extend, as exemplified by the lapatinib/EGFR complex (PDB code: 1XKK (Wood et al., 2004)). In the former case, half of the structures, classified as “C-helix out”, exhibit such a displacement that narrows the ATP binding site (Figure 5). Consequently, smaller ligands bind to these kinase members, as exemplified by the CDK2 complexes (Figure 5B).

All together, kinase conformations and annotations are gathered in one resource that is available through Kinatown, a web interface ([lbg.igbmc.fr/Kinatown](http://lbg.igbmc.fr/Kinatown)) that will be published elsewhere.

Our initial aim to develop covalent inhibitors targeting kinases in their inactive conformation raised an important issue concerning the proportion of kinases amenable to such a strategy. The present analysis reveals that, all together, 40% of distinct kinases adopt at least one of the four inactive conformations: 43 in “C-helix out”, 24 in “DFG out”, 19 in “A-loop out” and 4 in “C-helix and DFG out” (Figure 2B; Figure SI 3). These inactive conformations are spread evenly over the human phylogenetic tree, indicating that they are not specific to a kinase group (Figure SI 4). The present data suggests that a large proportion of kinases are targetable through an inactive conformation and that different alternatives may exist to achieve this goal, and as such may be a valuable information in drug discovery programs.

The present classification has already been exploited to specifically target the “DFG out” inactive conformation of KIT by a covalent analogue of imatinib (Leprout et al., 2011). This latter study explores in depth the presence of reactive cysteine residues in the extended ATP pocket observed in the inactive conformations.

## SIGNIFICANCE

Human protein kinases are involved in many cellular processes, being responsible for the phosphorylation of about one third of the proteome. Deregulation of some kinases has been observed in many diseases but, for a long time, they were not considered as relevant

therapeutic targets. This changed recently with the drug imatinib that proved to be selective among all kinases and able to compete with high ATP concentration in cells. Indeed, imatinib is an allosteric inhibitor that owes its good profile to the fact that it recognizes an inactive conformation adopted by its primary target ABL. We used the wealth of available 3D structures of kinases to investigate which kinase adopts an inactive conformation. In addition to the two well-described inactive conformations “C-helix out” and “DFG out”, we characterized two new inactive conformations. One of these two conformations, the “A-loop out” conformation, is adopted by 11% of crystallized kinases, and could represent a suitable target for the design of potent inhibitors. About 40% of crystallized kinases adopt at least one inactive conformation and 11% of crystallized kinases adopt more than one inactive conformation. These data reveals that the activity of many kinases could be modulated by allosteric inhibitors, and that a continuum of inactive conformations can be targeted for the same kinase. The present study provides, on one hand, for the biologist a detailed kinase conformational landscape to understand sequence/structure/function relationships, and, on the other hand, for the chemist the kinases that can be targeted in drug design projects in one of the four inactive conformations in order to be selective.

## EXPERIMENTAL PROCEDURES

### **Protein kinase dataset and cleaning**

The UniProt database (The UniProt Consortium, 2011) allowed the collection of 14739 reviewed protein kinases using the advanced search option: “protein family = kinase”. Among them, 554 proteins had associated crystal structures, leading to a total amount of 3015 structures. All the structures were downloaded using the *BIRD* system (Bard et al., 2010). 5928 chains were extracted from these 3D structures using some java developments of the *M-Orbis* software (Albou et al., 2011). In order to only keep chains coding for the related kinases (since protein partners can be present in the structure), the residue sequence alignment between the chain and the kinase was performed using *ClustalW* (Thompson et al., 2002). Chains were not retained when they poorly aligned with the kinase sequence (score less than 80% of identity, since mutations or deletions are accepted) or if less than 150 residues were aligned (low coverage of the kinase catalytic domain). An additional cleaning step checked whether the catalytic domain was present, since other domains of the same protein kinase could have been crystallised. This step was performed once the superimposition with an “Active” reference fold was done (see next paragraph), by retaining chains having more than

50 residues aligned on the “Active” reference. All together, this process resulted in 2587 chains containing the expected kinase catalytic domain, corresponding to 1744 crystal structures and 221 unique kinases shared among 17 species.

### **Superimposition on seven reference kinase structures adopting the “Active” conformation**

The structures of seven kinase catalytic domain, one per phylogenetic kinase group (Manning et al., 2002), were selected as “Active” reference structures (Table 1). Note that no reference was taken for the 8<sup>th</sup> and smallest group, the “RGC” group, since no structure is available. Moreover, an additional group exists among the kinase family, the “Others” group, corresponding to members that are dissimilar to the first eight groups. These reference structures were chosen based on the following requirements: a full kinase catalytic domain, adopting an “Active” conformation (Figure 1A) and having a resolution better than 2.2Å. The entire catalytic domain of the seven references, except for the P-loop, was used for the superimposition. Indeed, the position of this loop is known to be variable (Taylor et Radzio-Andzelm, 1997) regardless of the adopted conformation. The *Lsqman* program (Kleywegt, 1996) was used to superimpose each chain on the “Active” references. It aligns fragments with a minimum length of 50 residues and a maximum distance of 1.5Å between each pair of C<sub>α</sub> atoms. The output provides the number of aligned residues and the corresponding root mean square deviation. In order to find the best reference for a given protein chain, the superimposition leading to the highest number of aligned residues was kept as the best one.

### **Geometric criteria and conformation assignment**

The kinase classification relies on the displacements of C<sub>α</sub> atoms from relevant residues (Glu<sub>C-helix</sub> ; Phe<sub>DFG motif</sub> ; Arg<sub>A-loop</sub> ; Figure 2) between the reference and the superimposed chain, thus three relative distances are obtained (Figure 2). An additional intramolecular distance (Lys<sub>β3</sub>-Glu<sub>C-helix</sub>, Figure 2) is measured as well for this chain. For the seven reference structures, these residues are known (Table 1A). For the chains to be assigned, the equivalent residues are identified by using the sequence alignment inferred from the structural superimposition (obtained using the “global-superposition-distance-based Needleman-Wunsch” alignment method (Kleywegt, 1996)). Once these residues are known, geometrical features can be calculated and used for the classification according to the decision tree (Figure 2B).

## Annotations

Further analyses among chains required the extraction of features measured relative to the “Active” reference, such as the displacement of the third  $\beta$ -strand from the  $\beta$ -sheet in the N-terminal lobe ( $\Delta\beta_3$ ). The  $C\alpha$  atoms that have been taken into account correspond to three downstream residues and two upstream residues with respect to  $Lys_{\beta_3}$  ( $K72_{\beta_3}$ ). A-loop displacement was also studied by measuring the relative displacement of  $C\alpha$  atoms for the residues located between  $Phe_{DFG\ motif}$  and 20 residues upstream to  $Phe_{DFG\ motif}$ .

## Data processing

All the steps discussed previously were handled through one java program that was in charge of successively launching the external programs for structure download (*BIRD* (Bard et al., 2010)), chain extraction (*M-ORBIS* (Albou et al., 2011)), chain cleaning (*ClustalW* (Thompson et al., 2002)) and superimposition (*Lsqman* (Kleywegt, 1996)). The same java program processed the output of each external program, this data has been exploited at the different steps discussed previously. Finally, the decision tree rules, implemented in this program (Figure 2B), were applied for the conformation assignment.

## ACKNOWLEDGEMENTS

The Institut de Recherche Pierre Fabre (IRPF) is gratefully acknowledged for its support.

## REFERENCES

- Albou, L., Poch, O., et Moras, D. (2011). M-ORBIS: Mapping of mOlecular Binding sites and Surfaces. *Nucleic Acids Research* 39, 30 -43.
- Bard, N., Bolze, R., Caron, E., Desprez, F., Heymann, M., Friedrich, A., Moulinier, L., Nguyen, N. H., Poch, O., et Tournel, T. (2010). Décryphon grid - grid resources dedicated to neuromuscular disorders. *Stud Health Technol Inform* 159, 124-133.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., et Bourne, P. E. (2000). The Protein Data Bank. *Nucl. Acids Res.* 28, 235-242.
- Betzi, S., Alam, R., Martin, M., Lubbers, D. J., Han, H., Jakkaraj, S. R., Georg, G. I., et Schönbrunn, E. (2011). Discovery of a Potential Allosteric Ligand Binding Site in CDK2. *ACS Chemical Biology* 6, 492-501.
- Brooijmans, N., Chang, Y., Mobilio, D., Denny, R. A., et Humblet, C. (2010). An enriched structural kinase database to enable kinome-wide structure-based analyses and drug discovery. *Protein Sci* 19, 763-774.
- Davis, T. L., Walker, J. R., Loppnau, P., Butler-Cole, C., Allali-Hassani, A., et Dhe-Paganon, S. (2008). Autoregulation by the Juxtamembrane Region of the Human Ephrin Receptor Tyrosine Kinase A3 (EphA3). *Structure* 16, 873-884.
- De Bondt, H. L., Rosenblatt, J., Jancarik, J., Jones, H. D., Morgant, D. O., et Kim, S. (1993). Crystal structure of cyclin-dependent kinase 2. *Nature* 363, 595-602.
- Diskin, R., Lebendiker, M., Engelberg, D., et Livnah, O. (2007). Structures of p38[alpha] Active Mutants Reveal Conformational Changes in L16 Loop that Induce Autophosphorylation and Activation. *Journal of Molecular Biology* 365, 66-76.
- Jacobs, M. D., Caron, P. R., et Hare, B. J. (2008). Classifying protein kinase structures guides use of ligand-selectivity profiles to predict inactive conformations: Structure of Ick/imatinib complex. *Proteins: Structure, Function, and Bioinformatics* 70, 1451-1460.
- Karaman, M. W., Herrgard, S., Treiber, D. K., Gallant, P., Atteridge, C. E., Campbell, B. T., Chan, K. W., Ciceri, P., Davis, M. I., Edeen, P. T., et al. (2008). A quantitative analysis of kinase inhibitor selectivity. *Nat Biotech* 26, 127-132.

- Kleywegt, G. J. (1996). Use of non-crystallographic symmetry in protein structure refinement. *Acta Crystallogr. D Biol. Crystallogr* 52, 842-857.
- Knight, Z. A., et Shokat, K. M. (2005). Features of Selective Kinase Inhibitors. *Chemistry & Biology* 12, 621-637.
- Lahiry, P., Torkamani, A., Schork, N. J., et Hegele, R. A. (2010). Kinase mutations in human disease: interpreting genotype-phenotype relationships. *Nat Rev Genet* 11, 60-74.
- Lawrie, A. M., Noble, M. E., Tunnah, P., Brown, N. R., Johnson, L. N., et Endicott, J. A. (1997). Protein kinase inhibition by staurosporine revealed in details of the molecular interaction with CDK2. *Nat Struct Mol Biol* 4, 796-801.
- Leproult, E., Barluenga, S., Moras, D., Wurtz, J., et Winssinger, N. (2011). Cysteine Mapping in Conformationally Distinct Kinase Nucleotide Binding Sites: Application to the Design of Selective Covalent Inhibitors. *Journal of Medicinal Chemistry* 54, 1347-1355.
- Manning, G., Whyte, D. B., Martinez, R., Hunter, T., et Sudarsanam, S. (2002). The Protein Kinase Complement of the Human Genome. *Science* 298, 1912-1934.
- Marx, A., Nugoor, C., Müller, J., Panneerselvam, S., Timm, T., Bilanz, M., Mylonas, E., Svergun, D. I., Mandelkow, E., et Mandelkow, E. (2006). Structural Variations in the Catalytic and Ubiquitin-associated Domains of Microtubule-associated Protein/Microtubule Affinity Regulating Kinase (MARK) 1 and MARK2. *Journal of Biological Chemistry* 281, 27586 -27599.
- Roberts, W. G., Ung, E., Whalen, P., Cooper, B., Hulford, C., Autry, C., Richter, D., Emerson, E., Lin, J., Kath, J., et al. (2008). Antitumor Activity and Pharmacology of a Selective Focal Adhesion Kinase Inhibitor, PF-562,271. *Cancer Research* 68, 1935 -1944.
- Santarius, T., Shipley, J., Brewer, D., Stratton, M. R., et Cooper, C. S. (2010). A census of amplified and overexpressed human cancer genes. *Nat Rev Cancer* 10, 59-64.
- Schindler, T., Bornmann, W., Pellicena, P., Miller, W. T., Clarkson, B., et Kuriyan, J. (2000). Structural Mechanism for STI-571 Inhibition of Abelson Tyrosine Kinase. *Science* 289, 1938 -1942.
- Subramaniam, D. S., et Hwang, J. (2011). BIBW 2992 in non-small cell lung cancer. *Expert Opin. Investig. Drugs* 20, 415-422.
- Taylor, S. S., et Radzio-Andzelm, E. (1997). Protein kinase inhibition: natural and synthetic variations on a theme. *Current Opinion in Chemical Biology* 1, 219-226.
- The UniProt Consortium (2011). Ongoing and future developments at the Universal Protein Resource. *Nucleic Acids Res* 39, D214-D219.
- Thompson, J. D., Gibson, T. J., et Higgins, D. G. (2002). Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics Chapter 2*, Unit 2.3.
- Wang, Z., Canagarajah, B. J., Boehm, J. C., Kassisà, S., Cobb, M. H., Young, P. R., Abdel-Meguid, S., Adams, J. L., et Goldsmith, E. J. (1998). Structural basis of inhibitor selectivity in MAP kinases. *Structure* 6, 1117-1128.
- Wang, S., Meades, C., Wood, G., Osnowski, A., Anderson, S., Yuill, R., Thomas, M., Mezna, M., Jackson, W., Midgley, C., et al. (2004). 2-Anilino-4-(thiazol-5-yl)pyrimidine CDK Inhibitors: Synthesis, SAR Analysis, X-ray Crystallography, and Biological Activity. *Journal of Medicinal Chemistry* 47, 1662-1675.
- Wood, E. R., Truesdale, A. T., McDonald, O. B., Yuan, D., Hassell, A., Dickerson, S. H., Ellis, B., Pennisi, C., Horne, E., Lackey, K., et al. (2004). A Unique Structure for Epidermal Growth Factor Receptor Bound to GW572016 (Lapatinib). *Cancer Research* 64, 6652 -6659.
- Yun, C., Mengwasser, K. E., Toms, A. V., Woo, M. S., Greulich, H., Wong, K., Meyerson, M., et Eck, M. J. (2008). The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proceedings of the National Academy of Sciences* 105, 2070-2075.
- Zhang, J., Yang, P. L., et Gray, N. S. (2009). Targeting cancer with small molecule kinase inhibitors. *Nat Rev Cancer* 9, 28-39.
- Zheng, J., Knighton, D. R., Ten Eyck, L. F., Karlsson, R., Xuong, N., Taylor, S. S., et Sowadski, J. M. (1993). Crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with magnesium-ATP and peptide inhibitor. *Biochemistry* 32, 2154-2161.

SUPPLEMENTAL INFORMATION

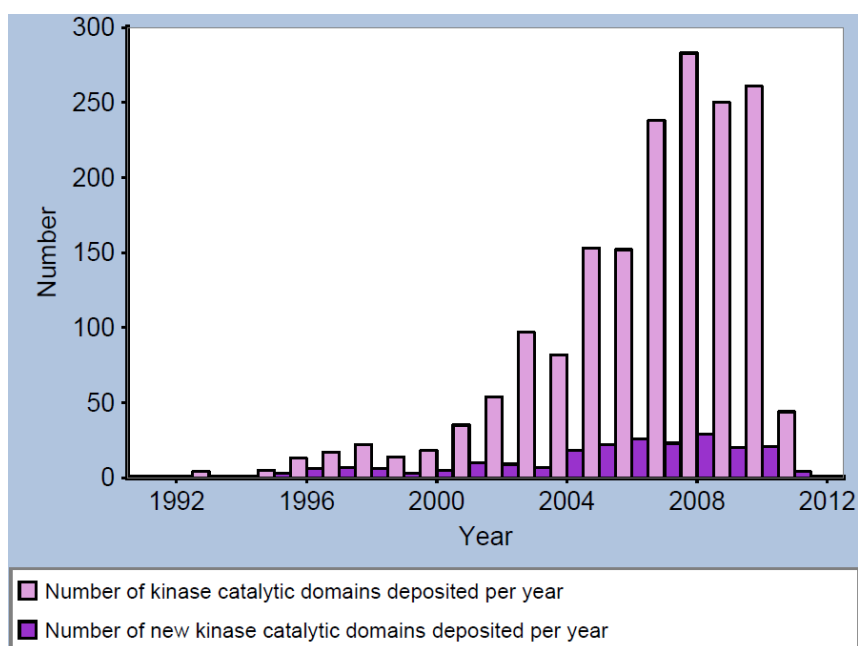
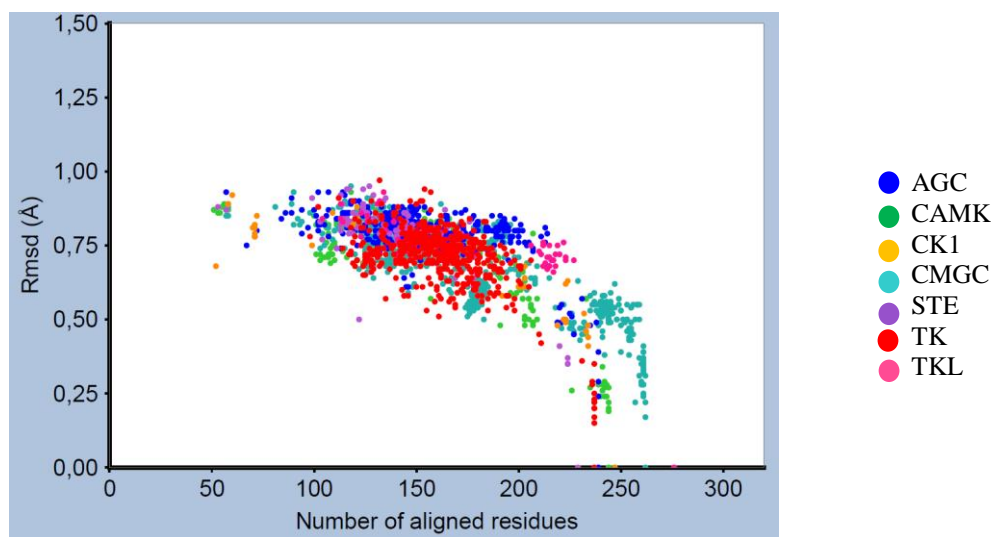


Figure SI 1: Evolution of the number of deposited structures containing a protein kinase catalytic domain, in the Protein Data Bank, between 1993 - date of the first deposited structure – and June 2011. In total, 1744 structures have been deposited. During the two last years, about 500 structures have been deposited.



(A)

Number of references	Mean number of aligned residues	Mean rmsd (Å)	Mean percent of aligned residues from N-terminal lobe	Mean percent of aligned residues from C-terminal lobe
7	162 (164)	0.72 (0.71)	44.34 (43.62)	68.95 (70.27)

(B)

Figure SI 2: Superimposition quality by using seven “Active” reference. (A) A 2D graph showing the number of aligned residues and the rmsd (Å) along the X and Y axis respectively, for the 2587 kinase catalytic domain chains. Each dot corresponds to a chain, color coded by the reference group. The use of seven “Active” references, one per kinase group, allows taking into account a certain plasticity inside the “Active” kinase fold. Only the best superimposition among the seven references is drawn. (B) Statistics on the 2587 superimpositions. The use of seven references reveals a better superimposition inside the C-terminal lobe, compared to the N-terminal lobe, known to be more flexible. Data for human chains are given in parenthesis.



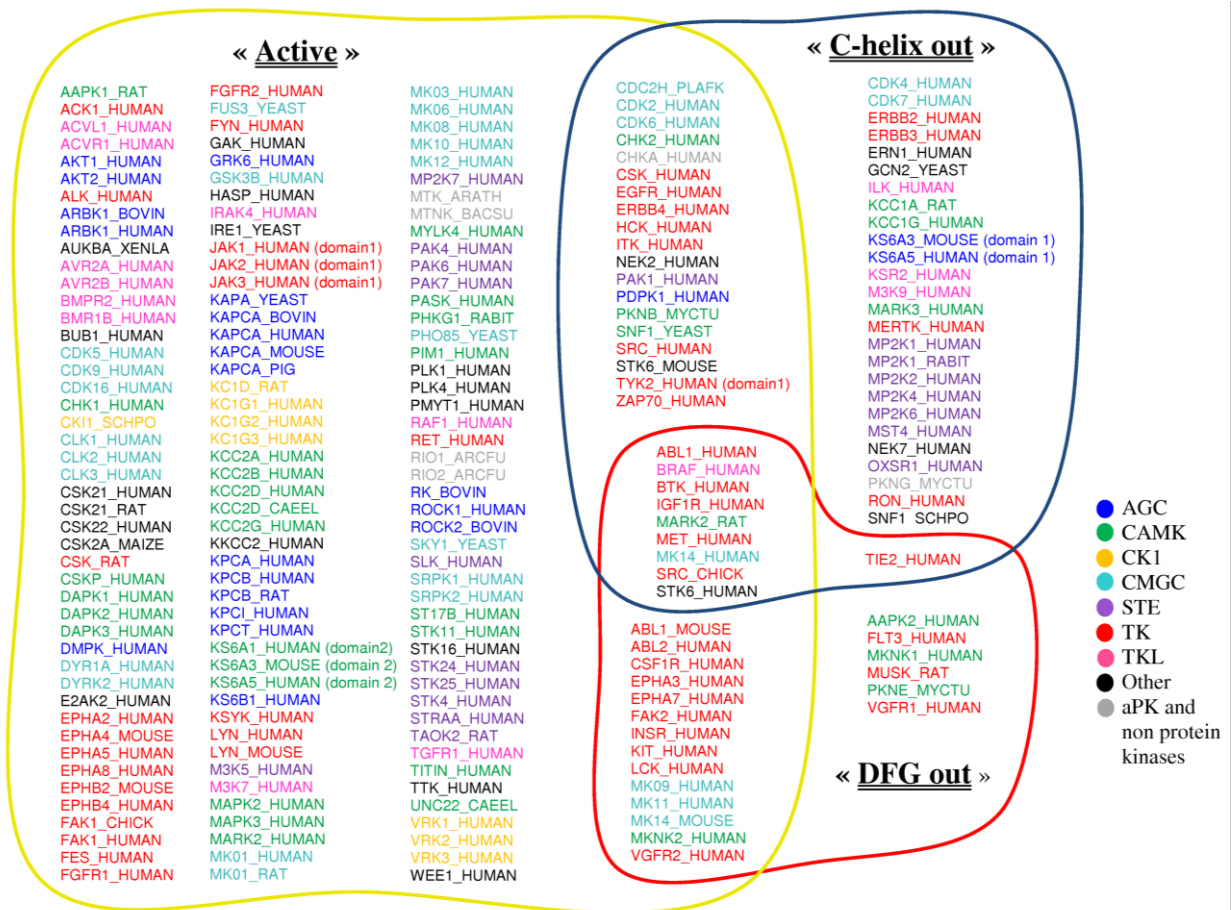


Figure SI 3: Venn diagram for the kinases crystallized in at least one of the three major conformations: “Active” (181 kinases), inactive “C-helix out” (55 kinases) and inactive “DFG out” (30 kinases).

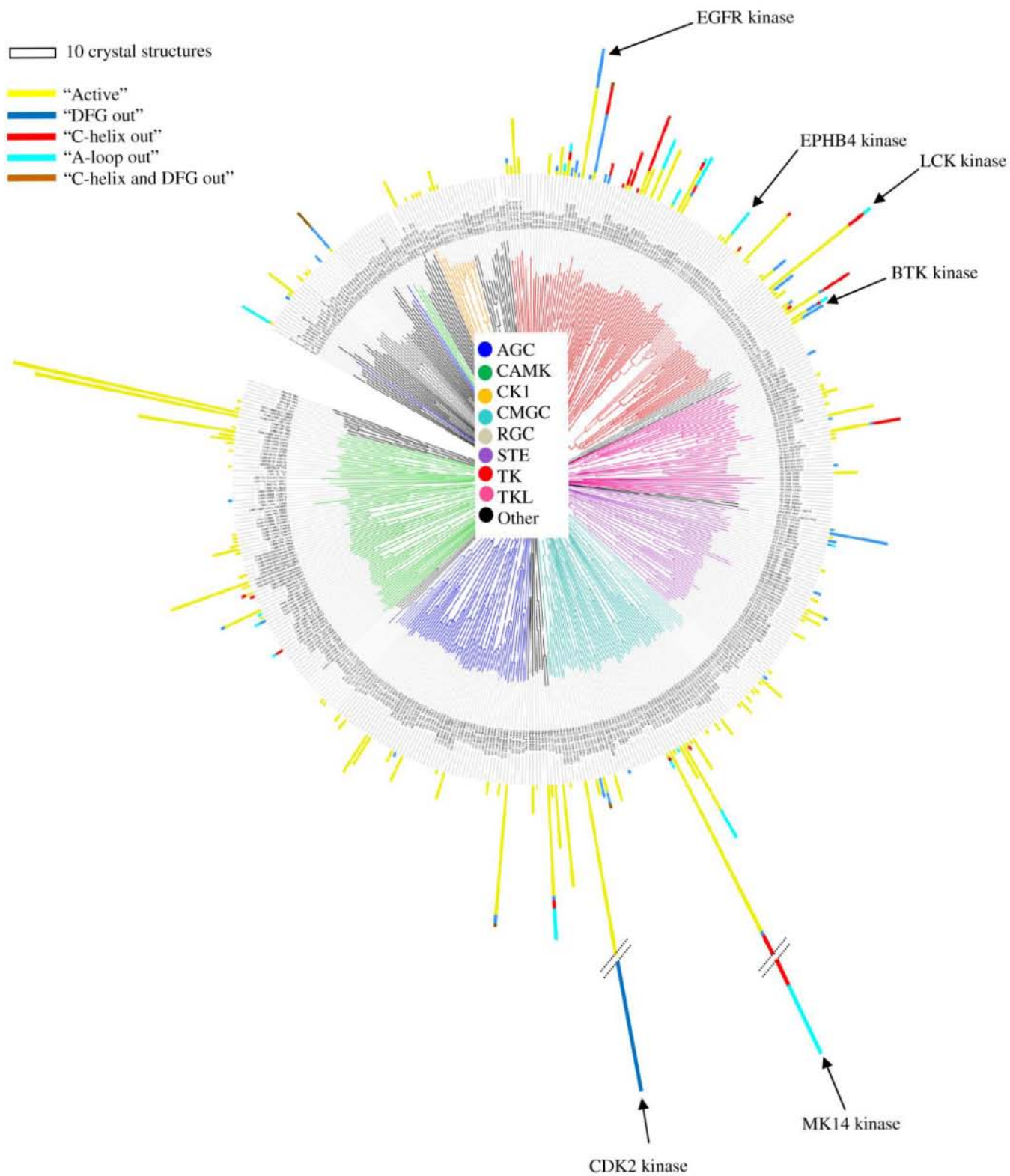


Figure SI 4 : The human phylogenetic tree with the distribution of crystal structure number per conformation. Tree representation made with Itol (Letunic and Bork, 2006).

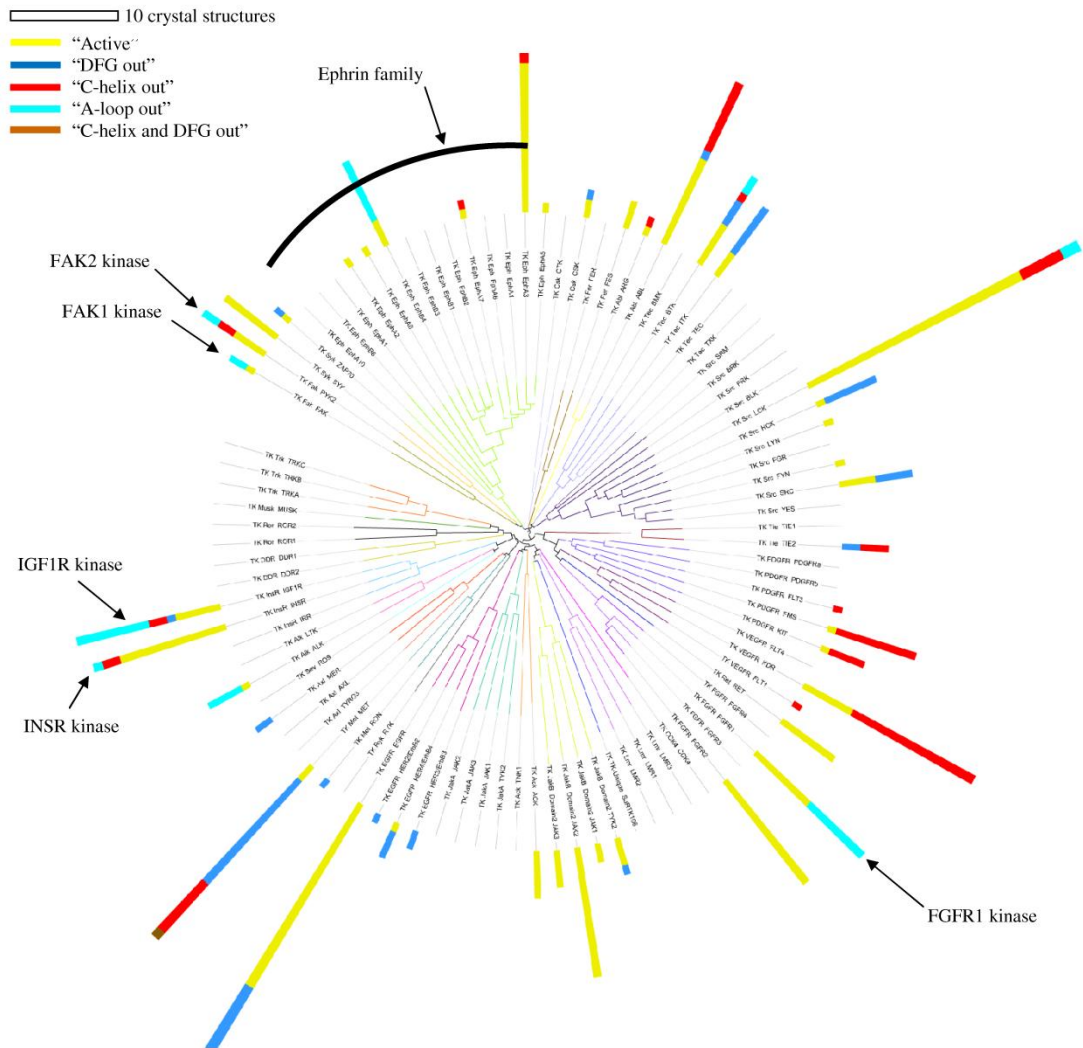


Figure SI 5 : The human tyrosine kinase phylogenetic tree with the distribution of crystal structure number per conformation. Tree representation made with Itol (Letunic and Bork, 2006).

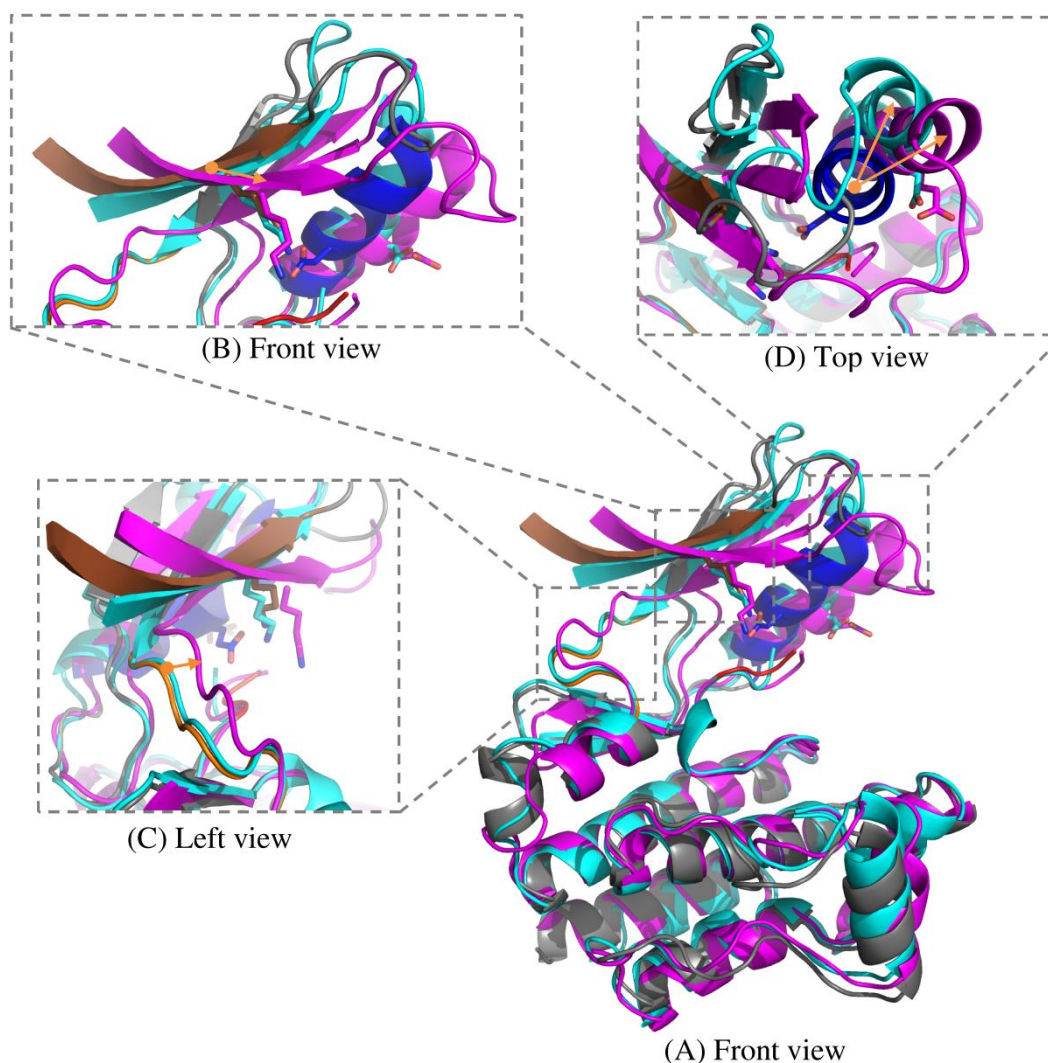


Figure SI 6 : N-terminal lobe movements inside kinases adopting the “C-helix out” conformation. (A) Superimposition of BTK and EGFR kinases on the “Active” reference LCK. BTK kinase (PDB code: 1K2P) is in cyan, EGFR kinase (PDB code: 2GS7) is in magenta, and the “Active” reference (PDB code: 3LCK) is color-coded as in Figure 1 (in grey apart from the  $\beta_3$  in brown, the C-helix in blue and the hinge region in orange). Lys $_{\beta_3}$  and Glu $_{C-helix}$  are shown in sticks in all structures. Both BTK and EGFR kinases adopt the “C-helix out” conformation but EGFR kinase has (B) the  $\beta$ -sheet displaced towards the ATP binding site compared to BTK and the “Active” reference, impacting also the positioning of the (C) hinge region and the (D) C-helix. The two first  $\beta$ -strands, part of the activation loop and ligands are not shown for clarity.

Conformations	Species	
	All species	human
“Active”	1732 (181)	1345 (140)
“C-helix out”	342 (55)	304 (43)
“DFG out”	213 (30)	171 (24)
“A-loop out”	156 (22)	143 (19)
“C-helix and DFG out”	11 (6)	9 (4)
Total of assigned chains	2454 (214)	1972 (166)
Chains missing relevant residues	133	101
Total	2587 (221)	2073 (172)

Table SI 1 : Distribution of chains in the five conformations, for all species and human only. The number of chains that could not be classified, since they missed relevant residues used in the conformation assignment procedure, is also given. The number of distinct kinases encoded by the chains is given in parenthesis.

### 3.7 Quel niveau de détail au sein de chaque conformation ?

Les analyses visuelles des structures au sein de chaque conformation ont fait ressortir des particularités qui nous ont paru importantes de souligner, du fait de leur possible impact sur la conception d'inhibiteurs. Cela m'a amenée à ajouter des critères de déplacement permettant l'assignation de sous-conformations au sein des cinq grands groupes de conformations.

Le déplacement du brin  $\beta 3$  a été décrit dans la publication ci-dessus et a permis d'étudier la diversité du déplacement du feuillet au sein de la conformation inactive *C-helix out*. Un autre critère, qui n'a pas été cité dans la publication, est le déplacement de l'atome  $C\zeta$  de la phénylalanine situé dans le motif DFG (voir distance  $d5$  dans la Figure 42 à la page 88). En effet, son déplacement combiné à celui de l'atome  $C\alpha$  sur le même résidu ont permis de caractériser divers rotamères adoptés par la phénylalanine, au sein des conformations *Active*, *A-loop out* et *C-helix out* (Tableau 10 ; Figure 46). Ces observations sont importantes du point de vue de la conception d'inhibiteurs car certains rotamères amènent la phénylalanine à participer au site actif, donc pouvant potentiellement former des interactions additionnelles avec le ligand. Il convient de souligner que la dénomination de toutes ces sous-conformations, hormis *DFG out-like* (Liao, 2007), relève de nos observations et non de la littérature (Tableau 10).

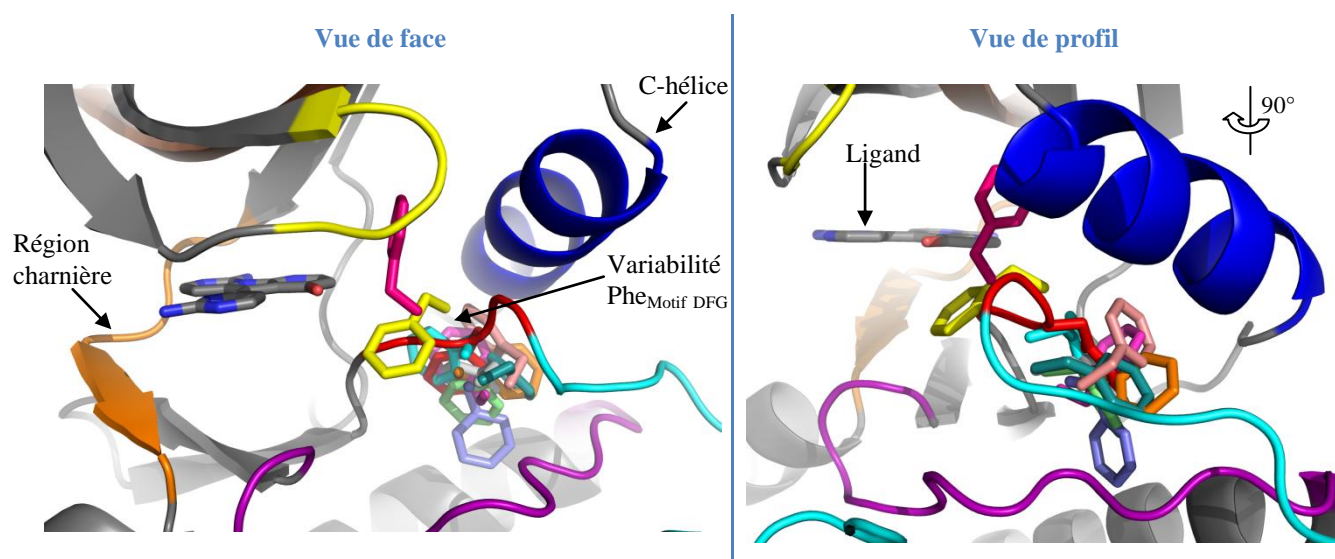


Figure 46 : Superposition de plusieurs structures de kinases présentant diverses sous-conformations du résidu Phe<sub>Motif DFG</sub>. L'ensemble du domaine catalytique d'une seule kinase (CDK2, code PDB : 3BHV, Phe<sub>Motif DFG</sub> en rouge) est représenté, avec son ligand. La sous-conformation des résidus Phe<sub>Motif DFG</sub> est donnée dans le Tableau 10, via le code couleur correspondant.

Sous-conformation Phe <sub>Motif DFG</sub>	Description par rapport à la référence	Nb chaînes (%) en conformation <i>Active</i> (1732 chaînes)	Nb chaînes (%) en conformation <i>A-loop out</i> (156 chaînes)	Nb chaînes (%) en conformation <i>C-helix out</i> (342 chaînes)	Exemples de PDB
<i>Active</i>	Inchangée, donc pointant vers le lobe C-terminal	87.1	14.6	2.4	3BHV
<i>A bit displaced</i>	Déplacée d'environ 2Å	4	1.5	0	1K3A
<i>Pan-like</i>	Ressemble à une poêle	3.4	28.5	36	3FMK
<i>DFG out-like</i>	Rotation de 180° autour de la chaîne principale	2	10.2	0	1OPL
<i>Rotated and displaced at the right</i>	Rotation de 90° puis translaturée vers la droite	1.6	10.2	0	3ELJ
<i>Displaced forward</i>	Translatée d'environ 2Å vers l'arrière	0.8	14.6	0	1R3C
<i>Highly displaced at the bottom</i>	Translatée d'environ 3.5Å vers le lobe C-terminal	0.5	15.3	34	1B38
<i>Pan-like rotated at the right</i>	Ressemble à une poêle pointant vers la droite	0	0	12.1	3DAK
<i>Displaced at the bottom</i>	Translatée d'environ 2Å vers le bas	0.4	0.7	5.7	2FVD
<i>Displaced at the right</i>	Translatée d'environ 1.5Å vers la droite	0.4	2.2	5.4	3CC6
<i>Highly displaced at the top</i>	Déplacée d'environ 3.5Å et pointant vers le lobe N-terminal	0	0	2.7	2W5B
<i>Unclassified</i>	Inclassable	0	2.2	1.7	

Tableau 10 : Variabilité de la sous-conformation du résidu Phe<sub>Motif DFG</sub> au sein des conformations *Active*, *A-loop out* et *C-helix out*. Les orientations *top*, *right*, etc. se réfèrent à notre choix d'orientation où le lobe N-terminal est en haut et la région charnière à gauche (voir Figure 40 à la page 85). Un exemple de chaque sous-conformation est donné dans la Figure 46, avec le code couleur correspondant.

Selon la conformation du domaine catalytique, certaines positions du résidu Phe<sub>Motif DFG</sub> sont plus fréquentes que d'autres (Tableau 10). Au sein de la conformation *Active*, le résidu Phe<sub>Motif DFG</sub> pointe très majoritairement vers le lobe C-terminal, comme c'était le cas pour les sept références. Cependant, il lui arrive d'adopter des positions atypiques telles que la sous-conformation *DFG out-like*, où la phénylalanine subit une rotation de 180° autour de la *A-loop*, tandis que le reste de la *A-loop* est inchangé (Liao, 2007). Au sein de la conformation *A-loop out*, la variabilité devient un peu plus importante avec la Phe<sub>Motif DFG</sub> qui peut adopter les sous-conformations *Pan-like*, *DFG out-like*, *Rotated and displaced at the right*, *Displaced forward* et *highly displaced at the bottom*, entre autres. Cet éloignement de la position *Active* de la Phe<sub>Motif DFG</sub>, supérieur à 2Å, est en accord avec notre hypothèse que la conformation *A-loop out* serait un état intermédiaire entre la conformation *Active* et la conformation inactive *DFG out* (voir paragraphe 3.6 « Article en cours de soumission » à la page 92). Dans la conformation inactive *C-helix out*, très peu de chaînes ont une Phe<sub>Motif DFG</sub> en position *Active*. Au contraire, pour la majeure partie d'entre elles, la Phe<sub>Motif DFG</sub> adopte l'une des deux sous-conformations *Pan-like* et *Highly displaced at the bottom*. La sous-conformation *Pan-like* est intéressante car elle permet au cycle aromatique de la Phe<sub>Motif DFG</sub> de participer à la poche additionnelle hydrophobe située près du résidu *gatekeeper*. Ainsi, dans la structure

1XKK, l'inhibiteur lapatinib (voir Tableau 2 à la page 21) forme une interaction favorable  $\pi$ - $\pi$  avec la Phe<sub>Motif DFG</sub> (Figure 47) (Wood et al., 2004).

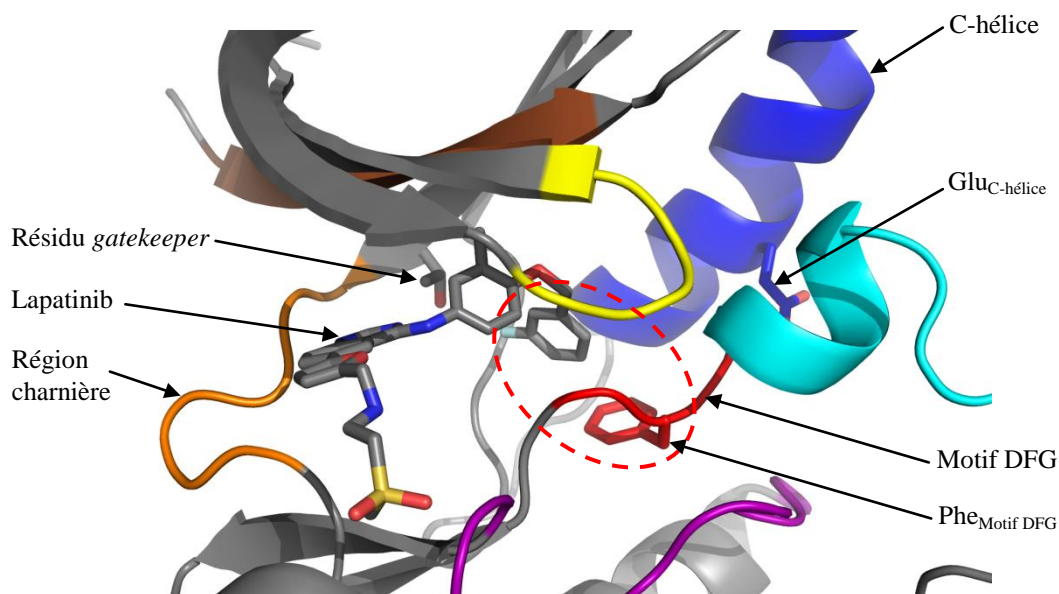


Figure 47 : Structure 3D de la kinase EGFR en conformation inactive *C-helix out* (code PDB : 1XKK). L'inhibiteur lapatinib forme une interaction favorable  $\pi$ - $\pi$  (entourée en pointillés) avec le résidu Phe<sub>Motif DFG</sub>, adoptant la sous-conformation *pan-like*, au niveau de la poche additionnelle hydrophobe située, ici, à l'arrière du *gatekeeper* thréonine.

Les conformations inactives *A-loop out* et *C-helix out* présentent les mêmes sous-conformations majoritaires de la Phe<sub>Motif DFG</sub> : *Pan-like* et *Highly displaced at the bottom*. Cette similarité va dans le sens d'une étude récente où il a été démontré pour la kinase ABL, par dynamique moléculaire, que le passage de la conformation *Active* à la conformation inactive *DFG out* requiert un passage par un état intermédiaire où l'hélice recule, correspondant à la conformation inactive *C-helix out* (Shan et al., 2009). Plus précisément, cette conformation permet au résidu Phe<sub>Motif DFG</sub> d'utiliser la poche additionnelle hydrophobe située près du résidu *gatekeeper* pour tourner autour de la *A-loop*. La conformation inactive *C-helix out* serait donc un état intermédiaire, comme la conformation *A-loop out*, pour aller vers la conformation inactive *DFG out*. Cette hypothèse va de paire avec un travail expérimental effectué sur les récepteurs de l'éphrine, dont l'étude par résonance magnétique nucléaire (RMN) a permis d'identifier un déplacement de la C-hélice lorsque la kinase passe de la conformation inactive *DFG out* à la conformation *Active*, suite à la phosphorylation de la région juxtamembranaire (Wiesner et al., 2006). D'ailleurs, nous avons nous-mêmes identifié sept kinases cristallisées dans les deux conformations inactives *A-loop out* et *C-helix out* : IGF1R, STK6, MAP2K6, MARK2 (rat), MARK3, BTK et MK14.

Au sein de la conformation inactive *C-helix out*, plus d'une dizaine de chaînes correspondant à une seule et même kinase se distinguent des autres. Il s'agit de la tyrosine kinase MET, qui a la particularité de faire participer la *A-loop* au site de liaison du ligand, amenant la tyrosine 1230 - située six résidus en aval du motif DFG - à pointer vers le site actif, la protégeant ainsi de la phosphorylation (Figure 48). L'assignation de cette sous-conformation, que j'ai nommée *C-helix out MET-specific*, résulte de la détection d'une position particulière du résidu Phe<sub>Motif DFG</sub>, proche de la sous-conformation *Pan-like*, ainsi que d'un déplacement élevé du résidu Arg<sub>A-loop</sub> ( $\Delta\text{Arg}_{A-loop}\cdot\text{C}_\alpha > 10\text{\AA}$ ). La singularité de cette sous-conformation fait déjà l'objet de plusieurs projets avancés de recherche d'inhibiteurs au sein des industries pharmaceutiques, qui prennent avantage de la position du résidu Tyr1230 pour effectuer des interactions favorables  $\pi$ - $\pi$  entre le noyau aromatique de la tyrosine et celui présent sur le ligand (Figure 48) (Asses et al., 2009; Porter et al., 2009). Ces inhibiteurs sont majoritairement de type I car le *gatekeeper* leucine de la kinase MET empêche l'accès à la poche additionnelle hydrophobe située près de lui, depuis la poche de l'adénine. Cependant, parce que les inhibiteurs interagissent avec la *A-loop* et donc s'étendent dans la zone de liaison des phosphates, ils pourraient atteindre la fameuse poche hydrophobe et se transformer en ligand de type I  $\frac{1}{2}$ , ce qui augmenterait la surface d'interaction et la spécificité (Figure 48).

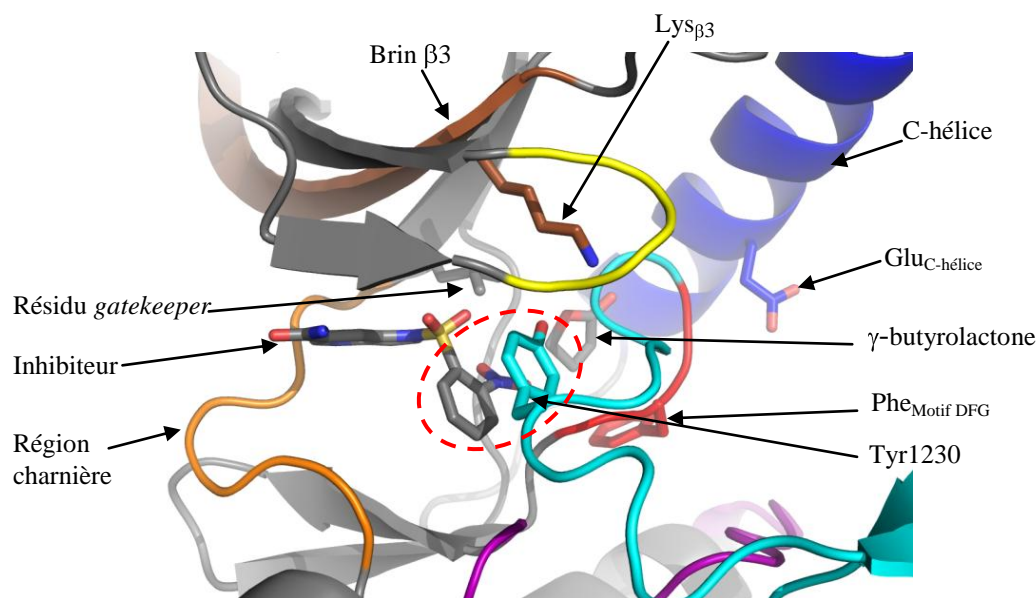


Figure 48 : Structure d'un inhibiteur dérivé d'azaindole et formant une interaction favorable  $\pi$ - $\pi$  (entourée en pointillés) avec le résidu Tyr1230 situé sur la *A-loop* de la kinase MET, adoptant la conformation inactive *C-helix out* qui lui est propre (sous-conformation *C-helix out MET-specific*). La poche additionnelle hydrophobe située près du *gatekeeper* leucine, occupée par un groupe  $\gamma$ -butyrolactone issu de la solution tampon, est atteignable par l'inhibiteur depuis le groupe NO<sub>2</sub> (code PDB : 2WD1).

Une interaction Lys<sub>β3</sub>-Glu<sub>C-hélice</sub> interrompue (Lys<sub>β3</sub>-Glu<sub>C-hélice</sub> > 4.5 Å) ou un déplacement un peu élevé du résidu Glu<sub>C-hélice</sub> ( $\Delta\text{Glu}_{C-hélice\alpha} > 2.5\text{\AA}$ ), observés en dehors de la



conformation inactive *C-helix out*, ont également été rapportés. De même, un déplacement global du lobe N-terminal, caractérisé par un faible pourcentage de résidus alignés dans cette zone, peut être identifié.

## 3.8 Conclusion

Étonnamment, seuls deux travaux de classification automatique des structures de kinases, extraites de la PDB, sont rapportés à ce jour dans la littérature. Le premier, réalisé par Jacobs et al. (Jacobs et al., 2008), date de 2008 et traite 71 kinases humaines distinctes. Parmi elles, les auteurs classifient 16 kinases dans la conformation inactive *DFG out* et 11 kinases dans la conformation inactive *C-helix out*. Plusieurs points sont à relever concernant ce travail : les structures 3D sans *A-loop* complète ont été éliminées (or une partie de cette zone est souvent désordonnée dans les structures) ; seules les structures de kinases humaines ont été étudiées ; la superposition se fait sur seulement 30 résidus d'une seule référence *Active* ; plusieurs structures sont admises comme étant mal classées par les auteurs ; et enfin, la liste des structures ayant permis ces classifications n'est pas accessible. La seconde classification date de 2010 et a été réalisée par Brooijmans et al. (Brooijmans et al., 2010), au sein de Wyeth Research. Leur travail s'inscrit dans le contexte plus général de la création d'une base de données d'inhibiteurs de kinases. Ainsi, leurs recherches portent plus sur une étude des interactions protéine/ligand dans les structures 3D que sur la description détaillée de la conformation adoptée par la kinase. Leur stratégie d'assignation des conformations inactives *DFG out* et *C-helix out* repose sur l'utilisation de deux critères intramoléculaires : la distance  $\text{Lys}_{\beta 3}\text{-Glu}_{\text{C-hélice}}$  et la distance  $\text{Phe}_{\text{Motif DFG}}\text{-région charnière}$ . Après comparaison entre nos classes et les leurs, il nous apparaît très clairement que le choix de leurs critères a généré d'importantes erreurs, allant jusqu'à classer plusieurs structures de kinases à la fois en conformation inactive *DFG out* et en conformation inactive *C-helix out*.

Finalement, d'autres études de classification des structures ont dû être menées au sein des industries pharmaceutiques, comme le laisse entendre Alton et Lunney, travaillant chez Pfizer (Alton et Lunney, 2008) mais leurs résultats sont restés confidentiels. Ces derniers évoquent l'idée que chaque kinase n'adopte pas une seule conformation inactive mais bien un continuum d'états conformationnels auxquels des inhibiteurs de divers types peuvent accéder.

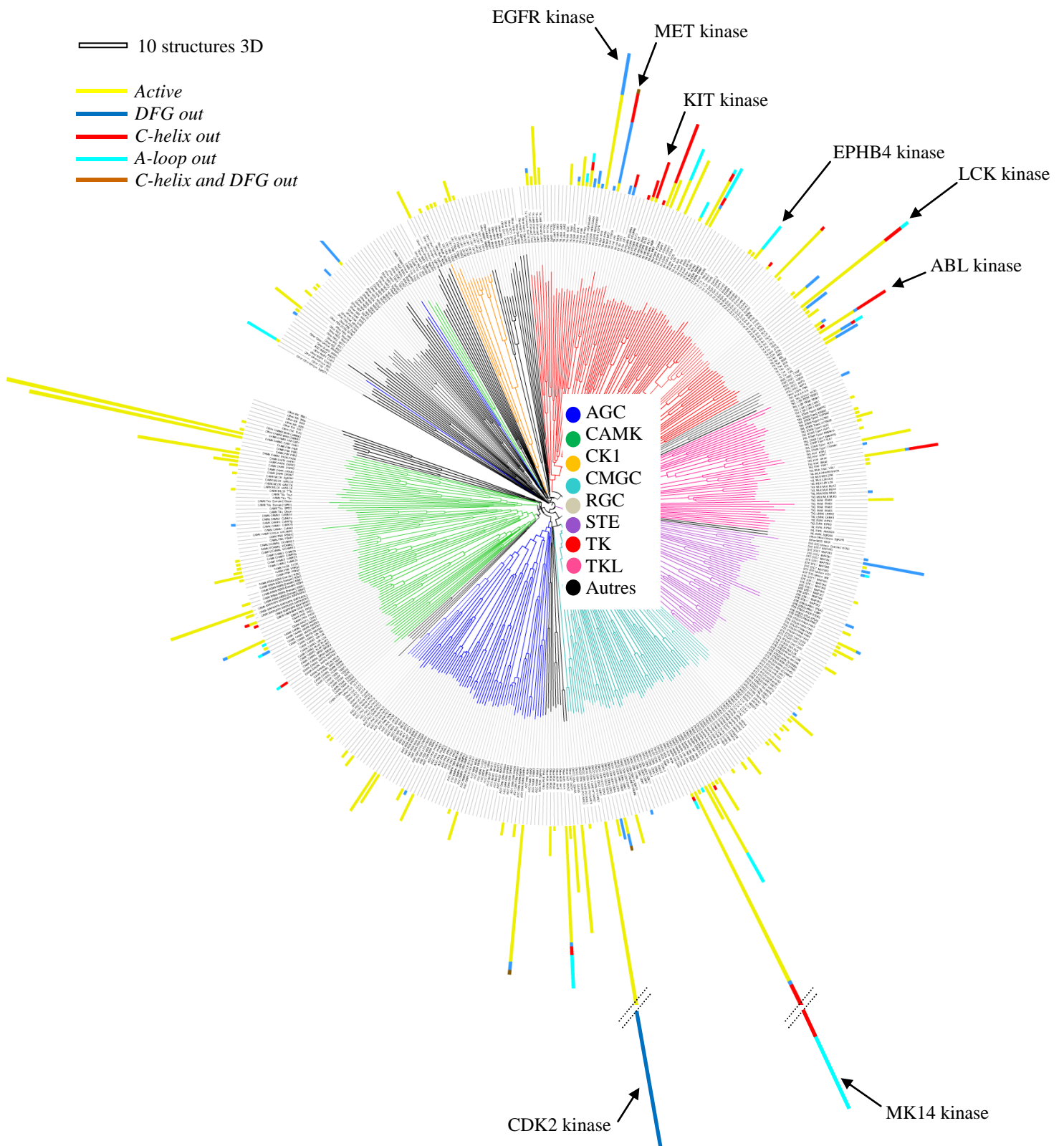


Figure 49 : Arbre phylogénétique des kinases humaines ePK avec la distribution du nombre de structures par conformation et par kinase (iTOL (Letunic et Bork, 2007)).

En effet, suite à notre classification des chaînes protéiques dans les cinq classes de conformation *Active*, *C-helix out*, *DFG out*, *A-loop out* et *C-helix and DFG out*, on peut remarquer qu'une même kinase peut avoir été cristallisée dans plusieurs conformations inactives, via des chaînes protéiques extraites de différentes structures 3D (Figure 49). Le groupe TK, qui est le plus travaillé de par le nombre de structures déposées, présente les quatre conformations inactives que nous avons décrites. Aucune de ces conformations inactives n'est cependant spécifique des TKs puisqu'on les retrouve quasiment toutes dans les autres groupes, malgré le plus petit nombre de structures cristallographiques disponibles (Figure 49). Finalement, ce sont 39% des 221 kinases distinctes cristallisées, qui sont classifiés dans une des quatre conformations inactives. Ces résultats confirment que beaucoup de kinases adoptent une, voir plusieurs conformations inactives, et nous avons montré qu'il existait même de la variabilité au sein de ces conformations (déplacement du feuillet ; conformations particulières du résidu Phe<sub>Motif DFG</sub> ; *A-loop* atypique). Ces observations ouvrent de multiples possibilités pour la conception d'inhibiteurs allostériques, qui seront alors plus sélectifs que les inhibiteurs de type I, visant la conformation *Active*.

Une particularité de mon travail de classification fut la caractérisation de la conformation *A-loop out*, adoptée par 22 des 221 kinases cristallisées, soit presque autant que de kinases classées en conformation inactive *DFG out*. Les similarités entre cette conformation et la conformation *Active* d'une part, et la conformation inactive *DFG out* d'autre part, nous laissent penser qu'il s'agit d'un état intermédiaire. La première structure de la kinase ALK, déposée à la PDB en 2010, a reçu l'assignation *A-loop out* (code PDB : 3LCS) (Lee et al., 2010). Les auteurs décrivent la position de la *A-loop* et la qualifient de similaire à la conformation *Active*, sans pour autant permettre la liaison du substrat. Leur description est donc identique à la nôtre. Ils ajoutent que la singularité de cette conformation en fait une cible attractive pour le développement d'inhibiteurs spécifiques de la kinase ALK. Nous rejoignons leur point de vue, à la différence que nous savons que cette conformation n'est pas aussi unique qu'ils le pensent.

De nombreuses protéines kinases seraient donc susceptibles d'adopter des conformations inactives. Comme décrit par Liu et Gray (Liu et Gray, 2006), certaines kinases ont pu être piégées dans une de leurs conformations inactives uniquement parce qu'un inhibiteur a été trouvé pour bloquer celle-ci. Afin d'optimiser la recherche d'inhibiteurs allostériques, les tests d'affinité et d'activité ont commencé à évoluer. Par exemple, Simard et al. ont développé un test de détection d'inhibiteurs allostériques de kinases, via l'accrochage d'un fluorophore à un résidu situé au début de la *A-loop* (Simard et al., 2009).

Plus exactement, lorsque qu'un inhibiteur de type II se fixe sur la cible kinase MK14 préalablement modifiée, la *A-loop* est déplacée et le changement d'environnement en acides aminés induit une différence d'intensité émise par le fluorophore. La mise en pratique de ces nouveaux tests devrait permettre l'identification d'un plus grand nombre d'inhibiteurs allostériques, et donc l'inhibition de kinases nouvellement bloquées dans des conformations inactives.

# 4 Cartographie des Cystéines et Application à la Conception d'Inhibiteurs Covalents

---

4.1 Cartographie des cystéines selon la conformation du domaine catalytique.....	126
4.2 Conception du premier inhibiteur covalent de la kinase KIT.....	129
4.3 Publication.....	131
4.4 Effet antiprolifératif sur des cellules humaines de leucémie à mastocytes.....	146
4.5 Conception rationnelle d'inhibiteurs covalents dérivés de l'hypothémicine.....	150
4.6 Conclusion.....	171

## 4.1 Cartographie des cystéines selon la conformation du domaine catalytique

Beaucoup de kinases sont susceptibles d'adopter une voir plusieurs conformations inactives, en plus de la conformation *Active*, comme nous avons pu l'observer en analysant toutes les structures 3D extraites de la PDB (voir chapitre 3 « Étude de la Diversité Conformationnelle au Sein des Protéines Kinases »). Chacune de ses conformations donne accès à un site catalytique plus ou moins étendu. Cette particularité est exploitée par les industries pharmaceutiques afin de trouver des inhibiteurs allostériques, c'est-à-dire susceptibles de bloquer leur cible kinase dans l'une des conformations inactives *C-helix out* ou *DFG out*. Ces inhibiteurs, dits de type I ½ ou de type II, auraient l'avantage d'être plus spécifiques que les inhibiteurs de type I, qui se lient à la conformation *Active* (voir paragraphe 1.2.6 « Les stratégies d'inhibition adoptées »). Une stratégie plus récente, et donc moins exploitée, est celle des inhibiteurs covalents. Plusieurs de ces inhibiteurs, se liant irréversiblement à une cystéine située à la fin de la région charnière chez la kinase EGFR, sont actuellement dans les dernières phases de développement clinique, laissant présager une autorisation prochaine de mise sur le marché (voir paragraphe 1.2.7 « Émergence des inhibiteurs de type covalent » à la page 47 ; voir Tableau 3 à la page 50) (Subramaniam et Hwang, 2011; Janne et al., 2011; Allen et al., 2003; Sequist et al., 2010; Laheru et al., 2008). Le plus grand bénéfice apporté par ces inhibiteurs réside dans leur aptitude à rester efficace face à une résistance développée par EGFR suite aux premiers traitements avec les inhibiteurs erlotinib et gefitinib, de type I ½ (voir Tableau 2 à la page 37) (Yun et al., 2008). La formation de cette liaison covalente leur procure également une grande affinité pour leur cible et un bon profil de sélectivité, assuré en partie par la faible conservation de la cystéine visée parmi les kinases humaines.

Le succès des inhibiteurs covalents relance la recherche d'inhibiteurs de kinases dans une autre voie que les voies habituelles et surexploitées que sont les inhibiteurs de type I à II. Nous avons alors voulu connaître le nombre de kinases humaines, autres que EGFR, qui présentaient une cystéine dans leur site catalytique et pouvaient être les cibles d'inhibiteurs covalents. Nous avons identifié, dans la littérature, seulement trois autres positions distinctes de cystéines ciblées par des inhibiteurs covalents de kinases (voir paragraphe 1.2.7.b « Les inhibiteurs covalents dans les publications » à la page 50 ; voir Tableau 4 à la page 52) (Subramaniam et Hwang, 2011; Zhou et al., 2010; Cohen et al., 2005; Jogireddy et al., 2009). Ces positions sont toutes les quatre accessibles dans le site catalytique de leur kinase

respective, lorsque celle-ci adopte la conformation *Active*. Une analyse bioinformatique succincte, menée en 2009, avait révélé qu'environ 200 protéines kinases humaines avaient une cystéine dans leur site catalytique et que ces cystéines se répartissaient entre 14 positions distinctes (Zhang et al., 2009). Cependant, les auteurs n'avaient pas effectué de recherches sur les kinases pouvant présenter des cystéines accessibles dans les poches additionnelles qui viennent élargir le site catalytique lorsque les kinases adoptent l'une ou l'autre des conformations inactives *C-helix out* ou *DFG out*. D'autre part, nous avons relevé des erreurs sur les résultats de leur analyse.

Afin de compléter et corriger ces résultats, j'ai voulu prédire, pour chacune des 518 kinases humaines, les résidus participant au site catalytique selon la conformation adoptée par le domaine catalytique et en déduire les kinases ayant une cystéine accessible. Pour cela, j'ai poursuivi le développement du programme présenté dans le chapitre précédent en automatisant la détection du site catalytique dans chaque chaîne protéique, à l'aide du programme *PCK* précédemment développé au sein de notre équipe. Ainsi, grâce à la connaissance des structures 3D de kinases, j'ai pu établir, pour chacun des trois grands groupes de conformations (*Active*, *C-helix out*, *DFG out*), une liste robuste des acides aminés dont la chaîne latérale participe au site catalytique. Chacune de ces trois listes a été propagée sur l'ensemble des séquences alignées des domaines catalytiques de kinases humaines (alignement adapté des travaux de Manning et al. (Manning et al., 2002)). Les kinases ayant au moins une cystéine dans le site catalytique ont ainsi pu être identifiées, dévoilant l'étendue des opportunités offertes pour la conception d'inhibiteurs covalents ciblant des kinases d'intérêt thérapeutique. En effet, il apparaît que 310 des 518 kinases humaines disposent d'au moins une cystéine dans leur site catalytique, toute conformation du domaine catalytique confondue. Dans la conformation *Active*, 27 positions de cystéines ont été trouvées et correspondent à 211 kinases humaines (Figure 50A). Dans les conformations inactives *C-helix out* et *DFG out*, les mêmes positions de cystéines sont retrouvées. 6 positions supplémentaires de cystéines propres aux extensions et modifications du site catalytique sont observées pour la conformation *C-helix out* (Figure 50B), tandis que 10 positions supplémentaires sont identifiées pour la conformation *DFG out* (Figure 50C). Ces deux derniers groupes de cystéines concernent 66 et 127 kinases, respectivement. Il est intéressant de souligner que le nombre de kinases ayant une cystéine à une position donnée du site catalytique n'est pas homogène parmi toutes les positions. Ainsi, la cystéine située au milieu de la région charnière est commune à 98 kinases, soit 20% du kinome humain, tandis que la cystéine située à la fin de la région charnière n'est commune qu'à 11 kinases (Figure 50A).

Les détails sur la stratégie de cartographie des cystéines sont présentés dans l'article « *Cysteine Mapping in Conformationally Distinct Kinase Nucleotide Binding Sites: Application to the Design of Selective Covalent Inhibitors* » publié en février 2011 dans *Journal of Medicinal Chemistry*, et qui est inséré dans le paragraphe 4.3 « Publication » à la page 131 (Leprout et al., 2011).

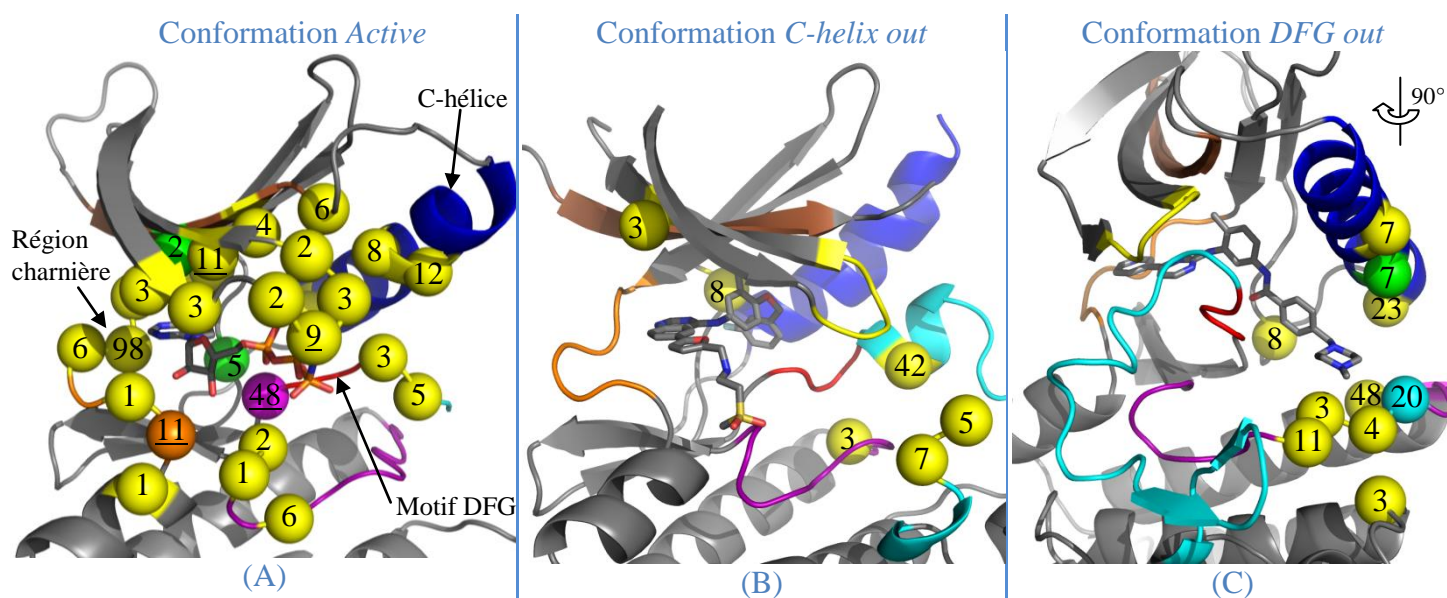


Figure 50 : Cartographie des positions de cystéines accessibles dans le site catalytique des kinases humaines, en fonction de la conformation du domaine catalytique. (A) Conformation Active et les 27 positions de cystéines accessibles dans le site catalytique. (B) Conformation inactive *C-helix out* et les 6 positions supplémentaires de cystéines accessibles dans le site catalytique. (C) Conformation inactive *DFG out* et les 10 positions supplémentaires de cystéines accessibles dans le site catalytique. Chaque sphère et le numéro inscrit dans la sphère correspondent à une position de cystéine et au nombre de kinases humaines qui la partagent. Les quatre nombres soulignés, au sein de la conformation Active, correspondent à des positions déjà ciblées par des inhibiteurs covalents, d'après la littérature. Les trois sphères orange, magenta et cyan correspondent aux positions de cystéines ciblées de manière covalente par des inhibiteurs que j'ai développés pendant mon travail de thèse. Les trois sphères vertes correspondent à d'autres positions exploitées pendant ma thèse, qui seront brièvement développées en annexe.

Au début de cette étude, seules quatre positions de cystéines, toutes spécifiques de la conformation *Active* des kinases visées, étaient répertoriées comme cibles d'inhibiteurs covalents. J'ai donc voulu procéder à la conception rationnelle d'inhibiteurs covalents, ciblant des kinases d'intérêt thérapeutique, en réagissant si possible avec d'autres cystéines que celles rapportées dans la littérature. Mon travail le plus avancé repose sur la conception d'un analogue covalent de l'imatinib, capable d'inhiber spécifiquement la kinase KIT en réagissant de manière covalente avec une cystéine située dans l'extension du site catalytique, lorsque la kinase est en conformation inactive *DFG out* (sphère cyan dans la Figure 50C). Vient ensuite tout un travail autour de l'hypothémicine, une molécule naturelle se liant de manière covalente à plusieurs des 48 kinases présentant une cystéine juste avant le motif DFG (sphère magenta dans la Figure 50A). Un premier travail a consisté à améliorer le profil de sélectivité



de l'hypothémicine parmi les 48 kinases en proposant des modulations de sa structure chimique. Un second travail a consisté à modifier de manière radicale le profil d'inhibition de l'hypothémicine. Pour cela, j'ai imaginé supprimer son groupe électrophile pour ensuite le repositionner sur l'inhibiteur de telle façon que la cystéine visée n'est plus située avant le motif DFG mais elle est située à la fin de la région charnière (sphère orange dans la Figure 50A), c'est-à-dire celle pour laquelle des inhibiteurs covalents d'EGFR sont actuellement en développement clinique. Quelques autres travaux ont été entrepris pour cibler d'autres kinases d'intérêt thérapeutique (sphères vertes dans la Figure 50A et C) mais ne seront pas détaillés explicitement dans ce manuscrit, si ce n'est une brève description en annexe (voir paragraphe 7.2 « Travaux sur d'autres inhibiteurs covalents de kinases d'intérêt thérapeutique » à la page 194).

## 4.2 Conception du premier inhibiteur covalent de la kinase KIT

Pour la première application du travail de cartographie, j'ai voulu obtenir un inhibiteur allostérique susceptible de se lier de manière covalente à une cystéine située dans une extension du site de liaison, lorsque la kinase est en conformation inactive. Une combinaison originale de plusieurs stratégies d'inhibition pourrait ainsi permettre d'atteindre un excellent profil de sélectivité et une grande efficacité de l'inhibiteur.

Mon travail a donc commencé par la conception rationnelle d'un inhibiteur covalent du récepteur à activité tyrosine kinase KIT. L'inhibition de KIT présente un intérêt thérapeutique majeur. En effet, son activité catalytique constitutive a été mise en cause dans diverses maladies graves, dont des tumeurs du tube digestif dites GIST et une forme de leucémie de type AML (*Acute Myelogenous Leukemia*) (Liegl-Atzwanger et al., 2010; Orfao et al., 2007). Des mutations de KIT sont également responsables de tumeurs des cellules germinales et de la mastocytose, une maladie rare où les mastocytes s'accumulent dans certains organes (voir paragraphe 4.4 « Effet antiprolifératif sur des cellules humaines de leucémie à mastocytes » à la page 146) (Orfao et al., 2007; Kemmer et al., 2004). Parmi les quatre cystéines identifiées dans le site catalytique de KIT, la cystéine 788 située sur la boucle catalytique n'est accessible que lorsque la kinase adopte la conformation inactive *DFG out* (sphère cyan dans la Figure 50C). Un inhibiteur de type II bien connu de KIT est l'imatinib (Figure 51 ; voir Tableau 2 à la page 37), également capable d'inhiber plusieurs autres kinases dont ABL (voir paragraphe 1.2.4 « Implication dans les maladies et premiers médicaments » à la page 33). En

analysant le positionnement de l'imatinib dans le site catalytique de KIT (code PDB : 1T46) et en observant sa proximité avec la cystéine 788, j'ai pu proposer aux collaborateurs chimistes des modifications où la substitution du groupement méthylpipérazine par divers groupes électrophiles permettait d'approcher le nucléophile avec une orientation et une distance susceptibles de favoriser la formation d'une liaison covalente. Le groupe méthylpipérazine, bien que contribuant à l'affinité de l'imatinib pour KIT ( $K_d$  de 100nM, extrait d'*Aureus Pharma*), n'est pas primordial pour observer une affinité. En effet, je me suis préalablement assurée que sa substitution par d'autres groupements non électrophiles permettait toujours d'observer une affinité pour KIT, bien que diminuée ( $K_d$  d'environ 1 $\mu$ M, extrait d'*Aureus Pharma*, d'après les inhibitions rapportées pour plusieurs analogues de KIT). Suite à mes propositions d'analogues covalents de l'imatinib, cinq molécules ont été synthétisées (Figure 51).

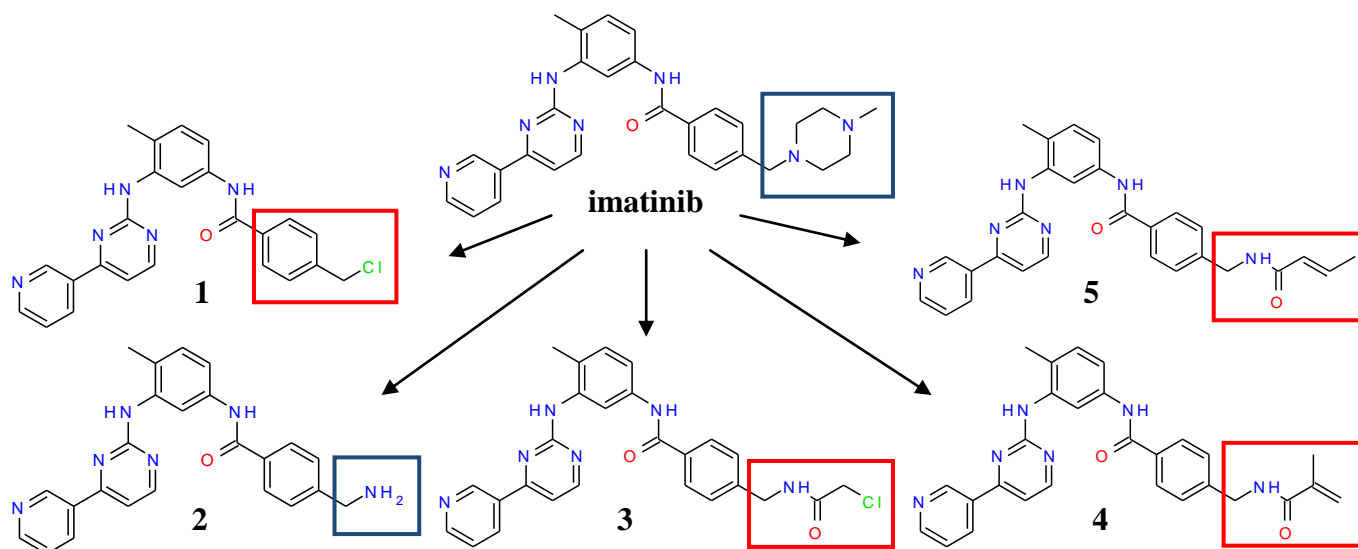
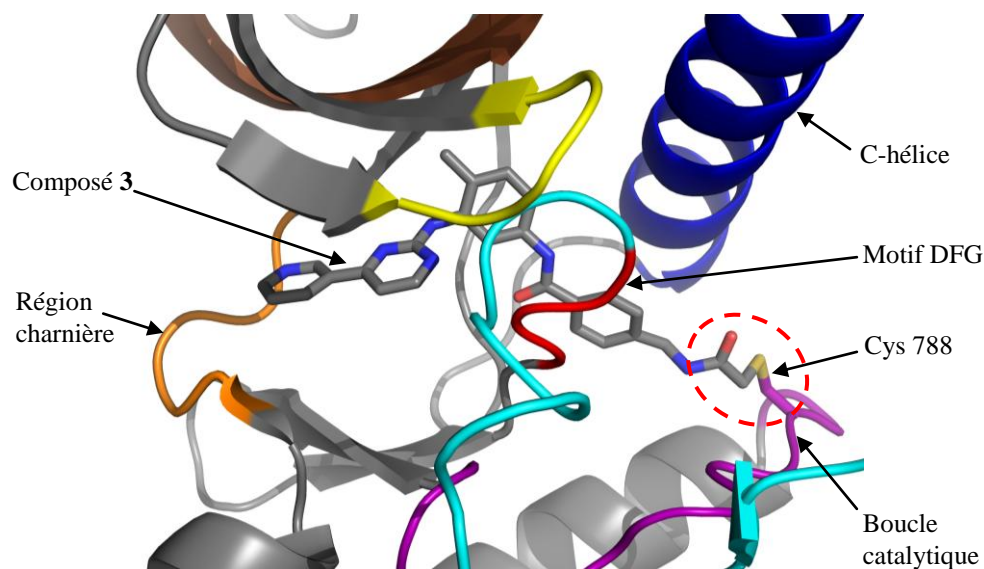


Figure 51 : Structure chimique de l'imatinib et des cinq molécules analogues synthétisées. Le groupe méthylpipérazine, encadré en bleu sur l'imatinib a été remplacé par divers groupes électrophiles encadrés en rouge sur les molécules 1, 3, 4 et 5. La molécule 2, avec le groupe amine qui n'est pas électrophile, a également été synthétisée.

La formation d'une liaison covalente entre l'une des molécules synthétisées, le composé **3** (Figure 51 ; Figure 52), et la cystéine 788 a été validée au moyen d'une digestion tryptique, suivie de la fragmentation MS/MS du peptide contenant la cystéine. Par la suite, nous avons obtenu le profil de sélectivité de ce premier inhibiteur covalent de KIT sur une majorité des 518 kinases humaines. D'une part, ce composé semble plus sélectif que l'imatinib. Par exemple, il n'inhibe pas ABL, qui ne possède pas de cystéine sur la boucle catalytique. D'autre part, la présence du groupe électrophile n'a pas provoqué l'inhibition d'autres kinases présentant des cystéines situées à d'autres endroits du site catalytique. Enfin, le squelette de l'imatinib était suffisamment sélectif, dès le départ, pour que la majeure partie

des 19 kinases autres que KIT et possédant une cystéine sur la boucle catalytique (sphère cyan dans la Figure 50C à la page 128) ne soient pas inhibées.



Les premiers résultats des tests biologiques associés à cet inhibiteur covalent de la kinase KIT sont détaillés dans le même article que celui expliquant la cartographie des cystéines, inséré dans le paragraphe suivant.

## 4.3 Publication

« *Cysteine Mapping in Conformationally Distinct Kinase Nucleotide Binding Sites: Application to the Design of Selective Covalent Inhibitors* ». Emeline Leproult, Sofia Barluenga, Dino Moras, Jean-Marie Wurtz, Nicolas Winssinger. *Journal of Medicinal Chemistry*, **54**, 1347-1355.

Remarque 1 : Suite à la publication de cet article (Leproult et al., 2011) en février 2011, celui-ci a fait l'objet d'un *Research Highlights* dans la rubrique *Drug Discovery* du journal *Nature Chemical Biology* paru en avril 2011 (Our choices from the recent literature, 2011).

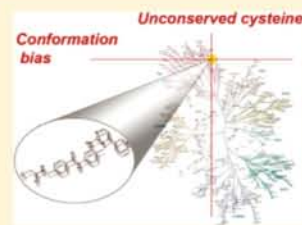
Remarque 2 : Alors que nous rédigeons l'article sur ce premier inhibiteur allostérique de type II et covalent, un brevet contenant un inhibiteur covalent de KIT analogue au nôtre était déposé par Avila Therapeutics, une jeune société de biotechnologie américaine qui se revendique comme étant spécialisée dans la conception d'inhibiteurs covalents (Singh et al., 2010).

## Cysteine Mapping in Conformationally Distinct Kinase Nucleotide Binding Sites: Application to the Design of Selective Covalent Inhibitors

Emeline Leproult,<sup>†</sup> Sofia Barluenga,<sup>‡</sup> Dino Moras,<sup>†</sup> Jean-Marie Wurtz,<sup>\*,†</sup> and Nicolas Winssinger<sup>\*,†</sup><sup>†</sup>Institut de Génétique et de Biologie Moléculaire et Cellulaire—CNRS (UMR 7104), INSERM (U964), 1 rue Laurent Fries, Université de Strasbourg, 67400 Illkirch, France<sup>‡</sup>Institut de Science et Ingénierie Supramoléculaire—CNRS (UMR 7006), Université de Strasbourg, 8 allée Gaspard Monge, 67000 Strasbourg, France

## Supporting Information

**ABSTRACT:** Kinases have emerged as one of the most prolific therapeutic targets. An important criterion in the therapeutic success of inhibitors targeting the nucleotide binding pocket of kinases is the inhibitor residence time. Recently, covalent kinase inhibitors have attracted attention since they confer terminal inhibition and should thus be more effective than reversible inhibitors with transient inhibition. The most robust approach to design irreversible inhibitors is to capitalize on the nucleophilicity of a cysteine thiol group present in the target protein. Herein, we report a systematic analysis of cysteine residues present in the nucleotide binding site of kinases, which could be harnessed for irreversible inhibition, taking into consideration the different kinase conformations. We demonstrate the predictive power of this analysis with the design and validation of an irreversible inhibitor of KIT/PDGFR kinases. This is the first example of a covalent kinase inhibitor that combines a pharmacophore addressing the DFG-out conformation with a covalent trap.



## INTRODUCTION

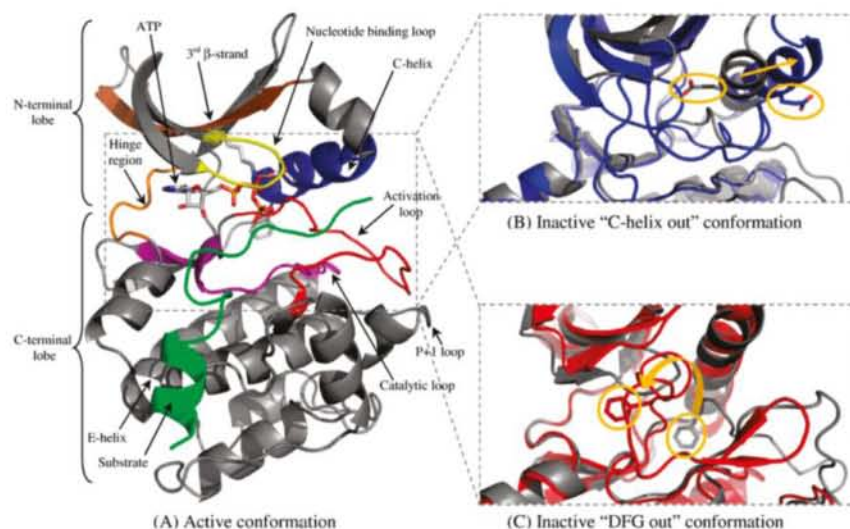
The sequencing of the human genome resulted in the identification of 518 protein kinases, which constitutes about 1.7% of all human genes.<sup>1</sup> Most protein kinases belong to a single superfamily containing a eukaryotic protein kinase<sup>1</sup> catalytic domain. This domain is composed of approximately 270 amino acids and promotes the transfer of the ATP  $\gamma$  phosphate to serine, threonine, or tyrosine residues of protein substrates (Figure 1A). Kinases are responsible for the phosphorylation of about one-third of the proteome, endowing them with a central role in regulating biochemical pathways. Dysfunction in kinase regulation is responsible for numerous pathologies ranging from inflammation to cancers<sup>2–6</sup> and has been the subject of numerous drug discovery efforts. Thus far, the most successful approaches to designing selective kinase inhibitors targeting the nucleotide binding pocket have been based on the presence of an inactive conformation adopted by some kinases, which provides an enhanced level of discrimination within the kinome. Indeed, “type 1½”<sup>7</sup> and “type 2”<sup>8</sup> inhibitors, such as lapatinib<sup>9</sup> and imatinib,<sup>10</sup> which target the inactive “C-helix-out”<sup>11</sup> (Figure 1B) and “DFG-out”<sup>12</sup> (Figure 1C) conformations, respectively, have been reported to be more specific than inhibitors targeting the active conformation,<sup>13</sup> which is highly conserved throughout the kinome. Alternatively, efforts have focused on the unique position of cysteine residues within the nucleotide binding site to develop irreversible inhibitors (Figure 2). Indeed, the most selective kinase inhibitor reported to date takes advantage of two selectivity filters, a bulky gatekeeper residue in

conjunction with a cysteine residue present in only 11 kinases, to achieve inhibition of ribosomal protein S6 kinases (RSKs) only.<sup>14</sup> Irreversible inhibitors have a long history in therapeutic intervention,<sup>15</sup> and several efforts in the kinase area have progressed to clinical trials: HKI-272,<sup>16</sup> CI-1033,<sup>17</sup> EKB-569,<sup>18</sup> and BIBW 2992,<sup>19</sup> which are directed against lung cancer by inhibiting the epidermal growth factor receptor (EGFR) kinase.<sup>20</sup> Encouragingly, these irreversible inhibitors overcame some of the limitations of reversible EGFR inhibitors as they retain activity against the otherwise debilitating T790-M mutation,<sup>16</sup> while maintaining a very good selectivity profile.<sup>21,22</sup>

To date, only four distinct cysteine positions in the ATP binding site have been harnessed for irreversible inhibition (the positions are marked in orange in Figure 3A), and efforts have been restricted to cysteines accessible in the “active conformation”. A cysteine, located just after the hinge region and present in 11 kinases, is targeted with covalent EGFR<sup>16,23–26</sup> or Bruton's tyrosine kinase (BTK)<sup>27</sup> inhibitors. A second cysteine, located on the second  $\beta$ -strand, is targeted by the covalent inhibitor CMK,<sup>14</sup> directed against RSK kinases. Covalent inhibitors exploiting a third cysteine shared by 48 kinases, which is localized before the DFG motif, have been discovered from a natural product family (resorcylic acid lactones such as hypothemycin).<sup>28–31</sup> The most recent covalent inhibition that has been reported targets a

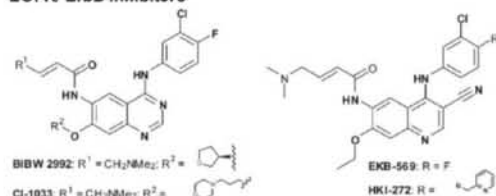
Received: October 28, 2010

Published: February 15, 2011

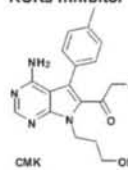


**Figure 1.** Catalytic kinase domain and its structural elements. (A) cAMP-dependent kinase (PDB code 1ATP) in the active conformation. The nucleotide binding loop is in yellow; the third  $\beta$ -strand of the  $\beta$ -sheet, bearing the catalytic lysine, is in brown; the C-helix, bearing the catalytic glutamic acid residue, is in blue; the catalytic loop is in purple; the activation loop, starting with the DFG motif, is in red; the peptide substrate is in green; ATP is represented in sticks and located in its binding site. Nucleotide and substrate binding sites are outlined in dashed line. (B) Cell division protein kinase 2 (CDK2): superimposition of the inactive C-helix-out conformation (in blue; PDB code 1PXN) with the active one (in gray; PDB code 3BHV). In the C-helix-out conformation, the glutamic acid residue (shown in stick in the two structures), located on the C-helix, is moved away from the ATP binding site due to the ionic interaction disruption with the catalytic lysine residue, involving a general displacement of the C-helix. Apart from the DFG motif, the activation loop is also highly displaced. (C) KIT kinase: superimposition of the inactive DFG-out conformation (in red; PDB code 1T46) with the active one (in gray; PDB code 1PKG). The phenylalanine of DFG motif is shown in stick in the two structures, to see its displacement in the DFG-out conformation, due to a general movement of the activation loop.

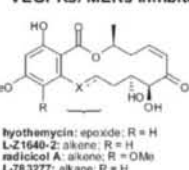
#### EGFR/ ErbB inhibitors



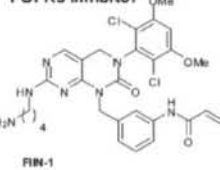
#### RSKs inhibitor



#### VEGFRs/ MEKs inhibitors



#### FGFRs inhibitor



**Figure 2.** Representative irreversible kinase inhibitors reported to date.

cysteine located on the nucleotide binding loop of basic fibroblast growth factor receptor (FGFR) kinases.<sup>32</sup>

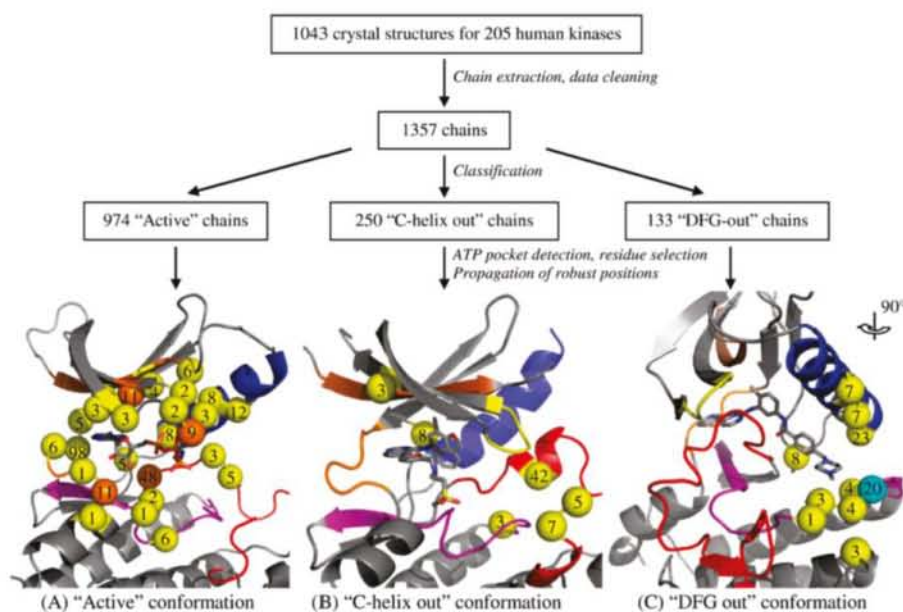
In their review, Zhang et al.<sup>8</sup> identified about 200 kinases bearing a cysteine in the nucleotide pocket, thus highlighting the broad applicability of this approach. While several reports have described the use of bioinformatics approaches<sup>8,14,23,28</sup> to identify a targetable cysteine, a systematic analysis taking into account the kinase conformation and accessibility to the thiol has not been reported.

## RESULTS

**Workflow Overview.** The workflow presented here allows access to all residues whose side chain robustly participates in the

kinase nucleotide binding site for a large number of protein kinase crystal structures (Figure 3). The beginning of the workflow is tuned to process all available crystal structures of the kinase catalytic domain and classify them into conformations. As a result, the structures were mainly classified into three conformations: active,<sup>13</sup> inactive C-helix-out,<sup>11</sup> and inactive DFG-out.<sup>12</sup> This classification is important for the subsequent analysis since the shape of the ATP binding pocket varies in function of the conformation.<sup>13</sup> The workflow allowed side chains that were not pointing toward the ATP binding site to be discarded. Furthermore, positions with low occurrence in the sequence alignment (non-robust positions) were also not included. At the end of the workflow, a list of residue positions robustly participating with the ATP binding site was obtained for each conformation. Such information was propagated to kinases without structural data associated. A first application consisted in finding out which kinases have accessible cysteines in their ATP binding site, with the objective to design covalent inhibitors.

**Cysteine Residues Inside the ATP Binding Site.** As a result of the workflow, 33 robust residue positions were obtained to describe the ATP binding site of kinases adopting the active conformation. In such a conformation, 211 proteins (43% of eukaryotic protein kinases, Table S11 in the Supporting Information) have at least one cysteine in the ATP binding site, distributed over 27 positions (Table 1 and Figure 3A). We then turned our attention to the other conformations. To date, 42% of crystallized human kinases adopt inactive conformations (to be submitted, Leproult et al.). The inactive C-helix-out conformation confers a globally larger ATP pocket than in the active conformation, and six additional cysteine positions were identified (Table 1 and Figure 3B) resulting in 66 kinases with a targetable cysteine (Table S12 in the Supporting Information). For the inactive DFG-out conformation,



**Figure 3.** Workflow overview and mapping of the cysteine positions with respect to the kinase conformation. The number inside each sphere denotes the number of kinases sharing a given cysteine position. Orange spheres denote previously targeted cysteine positions. (A) Active conformation, 27 cysteine positions; (B) C-helix-out conformation, six additional positions; and (C) DFG-out conformation, 10 additional positions. The position marked with a cyan sphere corresponds to the targeted position in the present work.

**Table 1.** Number of Kinases Having Cysteines in the Nucleotide Binding Pocket and Number of Positions Where Cysteines Are Located, Related to the Kinase Conformation<sup>a</sup>

Conformation	Nc	Nk	Ns									Nx
			AGC	CAMK	CK1	CMGC	RGC	STE	TK	TKL	OTHER	
Active	27	211 (47)	15	36	2	16	5	28	42	13	54	44
DFG-out	10*	127 (7)	5	35	1	26	1	19	20	5	15	11
C-helix-out	6*	66 (2)	32	6	1	3	0	10	7	1	6	1

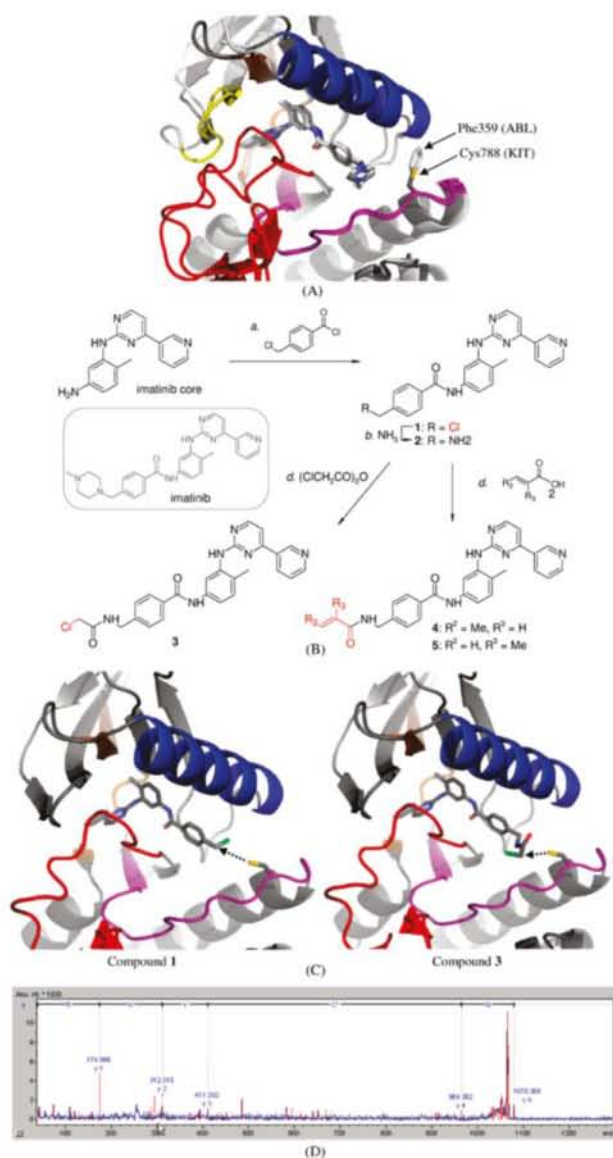
<sup>a</sup> Nc, number of cysteines positions; \*, specific positions of the inactive kinase conformations of which all positions found for the active conformation have to be added; Nk, number of kinases having at least one cysteine (number of kinases having more than one cysteine); Ns, number of kinases per kinase subfamily; and Nx, number of kinases having crystal structures in the desired conformation.

10 additional cysteine positions (Table 1) with respect to the active conformation were found (Figure 3C). These are all located in the extended pocket characteristic of the DFG-out conformation. Globally, 127 kinases have at least one cysteine located in this extension (Table S13 in the Supporting Information), which is not accessible in the active conformation and provides opportunities that have thus far not been explored.

**Design of a Covalent Kinase Inhibitor Targeting the Inactive DFG-out Conformation.** The analysis of cysteine positioning revealed a unique opportunity to discriminate among kinases targeted by pharmacophores designed for the DFG-out conformation, for example, imatinib<sup>10</sup> (Figure 4B). This conformation has been observed for 16% of the crystallized kinases, and indeed, while imatinib was designed to selectively target Abelson tyrosine kinase (ABL), it is now known also to inhibit two other therapeutic targets:<sup>33,34</sup> the mast/stem cell growth factor receptor (KIT) and the platelet-derived growth factor receptor (PDGFR). Among these three kinases, only KIT and PDGFR kinases bear a cysteine at the beginning of the catalytic loop (position equivalent to Cys788/Cys814 in KIT/PDGFR $\alpha$  kinases, Table S13 in the Supporting Information), which could

be exploited as a second selectivity filter to further refine the selectivity of an inhibitor and provide terminal inhibition. This position is located in the extension of the ATP binding site when the kinase adopts the inactive DFG-out conformation (the position is marked in cyan in Figure 3C).

Two crystal structures of the KIT and ABL kinases exhibiting the DFG-out conformation in complex with imatinib are available in the Protein Data Bank (PDB)<sup>35</sup> (Figure 4A; codes: 1T46 and 2HYY, respectively). The cysteine suggested by our study is solvent exposed and should thus be accessible for covalent bond formation. The striking proximity between this cysteine and the methylpiperazine group of imatinib in the KIT crystal structure suggested that modification of imatinib to include a suitably positioned electrophilic moiety should result in a reaction with cysteine. As shown in Figure 4B, starting from the imatinib heteroaromatic core structure, the aniline was acylated with 4-chloromethyl benzoyl chloride to afford compound 1. The benzyl chloride was then converted into the benzyl amine (compound 2) under the action of ammonia and acylated with chloroacetic anhydride, crotyl anhydride, or methylacrylic anhydride to provide compounds 3, 4, and 5, respectively. This



**Figure 4.** Design, synthesis, and characterization of irreversible inhibitors of KIT and PDGFR kinases in the DFG-out conformation. (A) Superimposition of two kinase crystal structures adopting the inactive DFG-out conformation in complex with imatinib: ABL (in light gray; PDB code 2HY7) and KIT (in dark gray; PDB code 1T46). (B) Synthesis of compounds 1–5 (electrophilic traps are shown in red). (C) Homology modeling of PDGFR kinase in complex with molecular models of imatinib analogues 1 and 3 bearing different electrophilic traps. (D) Sequencing by MS/MS fragmentation of PDGFR tryptic digest showing the expected shift in molecular weight for the cysteine adduct having reacted with compound 3.

synthesis thus afforded four compounds with slightly different distances between the heteroaromatic core and the electrophilic trap as well as two different reactivity modes (Michael addition or nucleophilic substitution). While molecular models suggested that compound 3 (Figure 4C) should have the optimal alignment, we reasoned that a certain level of mobility in the side chain may in fact favor other geometries. The inhibition  $IC_{50}$  of these five compounds, together with imatinib, was measured for ABL1, KIT, and PDGFR $\alpha$  (Table 2). While this assay measures kinase

activity and thus requires the phosphorylated state that favors the active conformation, we reasoned that the comparison between the five compounds and the imatinib would remain relevant and would provide an indication whether the modifications had had a deleterious impact. As the kinase substrate is added simultaneously with the kinase and the assay is allowed to proceed for only 60 min, we did not anticipate a pronounced impact stemming from covalent inhibition. Aside from compound 1, all other derivatives lost their activity against ABL, presumably due to the loss of ionic interaction between the positively charged methylated nitrogen of the piperazine ring of imatinib<sup>36</sup> and the carbonyl groups of Ile 789 and His 790 (KIT kinase numbering). On the other hand, the compounds containing the different electrophilic traps retained their activity against KIT and PDGFR. To confirm the formation of a covalent adduct between the imatinib analogues and the  $\gamma$ -sulfur of the targeted cysteine residue, tryptic digests of the kinases incubated with the different inhibitors were analyzed by mass spectrometry (MS). Compound 3, bearing the chloroacetamide group, indeed led to the disappearance of the fragment containing the targeted cysteine residue and to the appearance of a new peak corresponding to the peptide–inhibitor adduct for both KIT and PDGFR $\alpha$ . Further sequencing by MS/MS fragmentation indicated the expected shift in molecular weight (450 amu) for the cysteine adduct having reacted with compound 3 (Figure 4D). Other compounds failed to show similar covalent adducts. This can be rationalized by a less optimal positioning of the electrophile (for example, the angle between compound 1 warhead and the sulfur atom is unfavorable for  $S_N2$  substitution; Figure 4C) or a lack of sensitivity in the mass spectrometry assay.

To further assess the specificity of compound 3, which did form a covalent adduct with KIT and PDGFR $\alpha$ , it was profiled in a competition assay with 440 kinases (376 distinct human kinases and mutants or kinases of pathogenic organisms<sup>33,37</sup>) measuring residual binding of kinases to competitive binders. From this profile, nine kinases [KIT, PDGFR $\alpha/\beta$ , c-Jun NH2-terminal kinases (JNK) 1–3, epithelial discoidin domain-containing receptor 1 (DDR1), BRAF (V600E), and macrophage colony-stimulating factor 1 receptor (CSF1R)] were identified as having less than 10% residual binding (i.e., 90% binding inhibition; see Table S14 in the Supporting Information for a complete list). Among the 20 kinases that have a cysteine at the suitable position, 17 were tested, and only four show significant binding to compound 3 (KIT, PDGFR $\alpha/\beta$ , and CSF1R; Table 3). When comparing the binding of kinases with less than 10% residual binding (i.e., 90% binding inhibition) for either compound 3 or imatinib<sup>33</sup> (Table 4), the modification of the methylpiperazine group for the chloroacetamide affords an improved affinity for the JNK1–3 as well as the mutant of BRAF(V600E) despite the lack of a predictable covalent interaction.

## DISCUSSION

The workflow presented here allowed the analysis of a large number of crystal structures of human protein kinases, which takes into account the conformational diversity within the nucleotide binding pocket. Indeed, depending on the conformation adopted by the kinase catalytic domain, the ATP binding site is more or less extended in different directions, hence providing opportunities to refine inhibitor selectivity.<sup>13</sup> Nevertheless, despite the success of kinase inhibition, there has been a high attrition of kinase inhibitors in clinical development due to off-target toxicity<sup>33</sup> and lack of efficacy.<sup>16,38,39</sup> The preponderance of

Table 2. IC<sub>50</sub> Inhibition of Kinase Activity by Imatinib and Five Analogues, of Which Four Have an Electrophilic Moiety

Compound	IC <sub>50</sub> (M) (residue at the position corresponding to the cysteine targeted by electrophilic trap)			Electrophilic moiety
	ABL1 wt (phenylalanine)	KIT wt (cysteine)	PDGFR $\alpha$ wt (cysteine)	
imatinib	$2.7 \times 10^{-6}$	$3.70 \times 10^{-7}$	$1.60 \times 10^{-7}$	no
1	$6.95 \times 10^{-6}$	$2.45 \times 10^{-7}$	$1.39 \times 10^{-7}$	yes
2	$>2 \times 10^{-5}$	$7.88 \times 10^{-7}$	$1.54 \times 10^{-6}$	no
3	$>2 \times 10^{-5}$	$7.88 \times 10^{-7}$	$1.08 \times 10^{-6}$	yes
4	$>2 \times 10^{-5}$	$4.71 \times 10^{-6}$	$4.89 \times 10^{-6}$	yes
5	$>2 \times 10^{-5}$	$2.74 \times 10^{-6}$	$4.11 \times 10^{-6}$	yes

Table 3. Affinity Profile of Compound 3 for the 20 Kinases Containing the Suitably Positioned Cysteine<sup>a</sup>

kinase	compound 3	imatinib
AnpA	ND	ND
CDKL1	100	ND
CDKL4	ND	ND
CSF1R	5	0
Fer	95	91
Fes	100	78
Fgfr1	100	80
Fgfr2	92	100
Fgfr3	100	97
Fgfr4	100	99
Fltl	86	78
Flt3	81	76
Flt4	100	85
Jak3 (domain 1)	100	100
Kit	0	0
PDGFR $\alpha$	0.7	0.15
PDGFR $\beta$	0	0
Pyk2	68	93
STLK6	ND	ND
Vegfr2 (KDR)	64	57

<sup>a</sup>The assay was performed as described<sup>33</sup> using 1  $\mu$ M compound 3 and compared to the results reported for imatinib (the reported data were obtained using 10  $\mu$ M imatinib). The data represent residual binding of a competitive binder to the kinases. Thus, the lowest values reflect the strongest inhibition.

cysteine residues within the nucleotide binding pocket offers the opportunity to rationally design inhibitors capable of terminally inhibiting such kinases. While several inhibitors in clinical development have already highlighted the potential of covalent kinase inhibition, with the most advanced candidates in phase III, reported efforts have been restricted to very few kinases. In addition, a systematic analysis of the cysteine landscape taking into account kinase conformation and solvent accessibility of the cysteine's thiol has not been reported. While the endogenous function of most of these cysteines remains uncharted, at least in the case of the I $\kappa$ b signaling pathway, one of the natural modulation of the I $\kappa$ b kinase activity is leveraged on covalent inhibition by a prostaglandin bearing a Michael acceptor.<sup>40</sup>

Clearly, the covalent link between an inhibitor and its target generates long dissociation half-lives,<sup>41</sup> which prolongs efficacy beyond the clearance of the inhibitor, thus potentially allowing a reduction in drug exposure and a decreased risk of off-

Table 4. Table of Kinases for Which Compound 3 or Imatinib Resulted in Less than 10% Residual Binding of Competitor (i.e., More Than 90% Binding Inhibition)<sup>a</sup>

kinase	residue at the position corresponding to the cysteine targeted by electrophilic trap	compound	
		3	imatinib
ABL1	Phe	67	1.4
ABL2	Phe	49	0.4
BLK	Ser	69	0.8
BRAF (V600E)	Ile	4.6	20
CAMKK1	Ile	100	1.4
CAMKK2	Ile	100	2.1
CSF1R	Cys	5	0
DDR1	Phe	0.5	0.2
DDR2	Phe	31	7.4
DRAK1	Val	100	5.2
EPHA8	Tyr	100	7.2
GAK	Ile	18	5.1
Jnk1	Ile	0.45	14
Jnk2	Ile	4.6	75
Jnk3	Ile	1.9	5.6
Kit	Cys	0	0
Kit (V559D)	Cys	0	0
Kit (V559D, V654A)	Cys	7	0.3
LCK	Tyr	64	1
LYN	Tyr	76	3.2
Mark2	Ile	100	8.8
Mst2	Lys	100	4.8
PDGFR $\alpha$	Cys	0.7	0.15
PDGFR $\beta$	Cys	0	0

<sup>a</sup>The assay was performed as described<sup>33</sup> using 1  $\mu$ M compound 3 and compared to the results reported for imatinib (the reported data were obtained using 10  $\mu$ M imatinib). The data represent residual binding of a competitive binder to the kinases. Thus, the lowest values reflect the strongest inhibition.

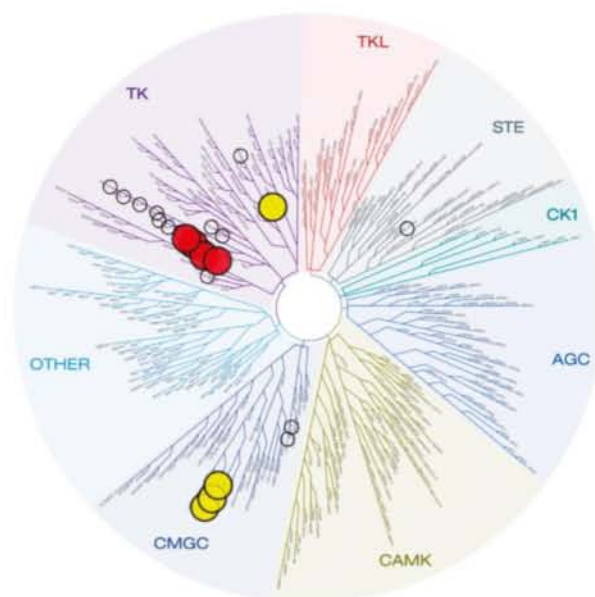
target effects.<sup>42</sup> Moreover, covalent inhibitors can be efficacious against kinases endowed with high  $k_m$  value for ATP,<sup>16,43</sup> by ultimately shifting the equilibrium between the free and the inhibitor-bound forms.<sup>38</sup> Another important advantage of covalent inhibitors is the opportunity to refine their selectivity profile by exploiting unconserved cysteine among kinase family,<sup>15,42</sup> a strategy that is complementary and synergic with other design approaches.



The design of such covalent inhibitors requires the identification of all human kinases having cysteine residues in their ATP binding site. The active kinase conformation gives access to 27 cysteine positions in the ATP binding site (Figure 3A) shared between 211 kinases (Table 1 and Table S11 in the Supporting Information). The C-helix-out and DFG-out kinase conformations give access to 6 (Figure 3B) and 10 (Figure 3C) additional cysteine positions shared between 66 and 127 kinases, respectively (Table 1 and Tables S12 and S13 in the Supporting Information). Kinases bearing equivalent cysteines are not necessarily from the same kinase subfamily. For example, the majority of the 11 kinases sharing the cysteine position equivalent to Cys797 in EGFR kinase (Table S11 in the Supporting Information) belong to the tyrosine kinase (TK) subfamily, unlike the five kinases sharing the cysteine position equivalent to Leu792 in EGFR kinase (Table S11 in the Supporting Information), which belong to TK, calcium/calmodulin-dependent kinases (CAMK), AGC, and other subfamilies. Interestingly, the number of kinases having one cysteine in a particular position of the nucleotide binding site is not homogeneous for all positions (Figure 3A–C). While most cysteine positions are populated with more than one kinase, other factors may contribute to the selectivity of inhibitors targeting a particular cysteine residue. For example, broad profiles of the resorcylic acid lactones targeting a cysteine residue present in 48 kinases show a marked selectivity for vascular endothelial growth factor receptors (VEGFRs) and mitogen-activated protein kinase kinases (MEKs). Surprisingly, despite the numerous opportunities offered by cysteines in the nucleotide binding pocket, only a few cysteine positions have thus far been targeted by covalent inhibitors<sup>14,18,29,32</sup> (Figure 2 and Figure 3A).

The results of this structural analysis were exploited to design compounds 1 and 3–5 to retune selectivity through covalent trapping. The targeted cysteine appears in the nucleotide binding pocket only when these kinases adopt the inactive DFG-out conformation. Among the kinases that can adopt the required conformation to accommodate the imatinib pharmacophore, only KIT, PDGFRs, and CSF1R have a suitably positioned cysteine residue at the beginning of the catalytic loop. Indeed, compound 3 that forms a covalent adduct with KIT and PDGFR did inhibit their activity but not that of ABL kinase. A broad profile of compound 3's affinity against a large panel of kinases reveals that the electrophilic trap did not confer promiscuity for other cysteine-containing kinases (see Figure 5 for graphical representation of selectivity). Indeed, among the kinases bearing the suitably positioned cysteine residue (20 kinases), only a small subset (KIT, PDGFRs, and CSF1R) are known to accommodate the imatinib pharmacophore and were targeted by compound 3. The substitution of the methylpiperazine group in imatinib for the chloroacetamide in compound 3 led in one case to an improved affinity (JNKs) among the kinases that are known to bind imatinib (over 20), albeit the cellular relevance of such reversible and transient interactions remains to be established. This is the first example of a covalent kinase inhibitor that exploits a cysteine not available in the active conformation in combination with a pharmacophore broadly selective for the DFG-out conformation.

In conclusion, the structural analysis described herein provides a thorough landscape of cysteine positions that can be exploited for irreversible inhibition. Taking into account the different conformations of kinases has revealed previously unrecognized opportunities. The identification of an inhibitor that



**Figure 5.** Graphical representation of compound 3's selectivity. Red- and yellow-filled circles represent covalently and reversibly targeted kinases (less than 10% residual binding of competitor), respectively. Empty circles represent kinases bearing the suitably positioned cysteine but devoid of binding due to pharmacophore specificity.

discriminates between ABL and KIT/PDGFRs illustrates the opportunities to enhance selectivity by harnessing unconserved cysteine residues. Taken together, the results provide a framework to extrapolate this approach to other members of the kinome.

## EXPERIMENTAL SECTION

**Human Kinase Data Set.** Sequences corresponding to the 491 kinase catalytic domains of the 478 human eukaryotic protein kinases and their sequence alignment file were retrieved from the published data generated by Manning et al.<sup>1</sup> This alignment has been refined for the purpose of this study (available upon request). For each kinase, the corresponding crystal structures were downloaded (July 2010) from the Protein Data Bank<sup>35</sup> via the BIRD system.<sup>44,45</sup> The part of the workflow (Figure 3) explaining the data cleaning of all chains extracted from crystal structures, as well as the classification into conformations, is detailed in Leproult et al. (to be submitted). This led to a total of 1357 chains, of which 974, 250, and 133 chains were classified into active, C-helix-out, and DFG-out conformations, respectively.

**ATP Binding Site Pocket Detection.** Each chain was prepared with the PDB2PQR program<sup>46</sup>, which was efficient for automatically removing ligands, solvent, and ions. This was also used to add hydrogens to all atoms in a manner consistent with favorable hydrogen bonding. To automatically detect empty pockets inside the chains, the in-house Pck program<sup>47</sup> was used. This program is dedicated to the geometric-based detection and characterization of pockets. This uses the alpha shape theory<sup>48</sup> to represent the surface of proteins and then implements algorithms specific of some types of pockets. Among these algorithms, CAST<sup>49</sup> was preferentially chosen due to its ability to detect partially buried pockets, as observed for the kinase ATP binding site. Because the classification into conformations required the superimposition of all chains (to be submitted, Leproult et al.), the pocket corresponding to the ATP binding site was located by testing whether it contains a virtually created 3D point placed in the ATP binding site nearby the

hinge region. In the end, only 23 chains were reported as having no detected ATP binding site pocket. Further inspection of these chains revealed that most of them have either a small pocket or a side chain that blocks a part of the ATP binding site, resulting in an autoinhibition phenomenon.

**Amino Acid Selection Inside the ATP Binding Site.** Among the amino acids having atoms participating in the detected ATP pocket, those having at least one atom of the side chain that participates in the pocket are selected for the next part of the workflow. For those that only have main chain atoms that participate in the binding site, the secondary structure information using DSSP program<sup>50</sup> is taken into account. Indeed, if the amino acid is located on a flexible structural element, such as a loop or a turn, then the amino acid is selected. The only peculiar case concerns glycine amino acid when located on a nonflexible structural element. Because a glycine amino acid does not have a side chain, the two hydrogens attached to the C $\alpha$  atom are investigated. If the one generating an R chirality when mutated to the side chain of a cysteine participates in the detected ATP binding site pocket, then the glycine is selected.

**Robust Amino Acid Positions and Propagation to All Human Kinases.** Each previously selected amino acid is highlighted in the sequence of the corresponding kinase, inside the sequence alignment file.<sup>1</sup> This gives access to an amino acid position. Once the amino acid positions have been obtained for all chains, the ones appearing in more than 30% of all chains, with respect to the conformational class, are considered as robust positions. This step avoids keeping rare positions located in abnormally extended ATP binding sites. Indeed, the detection of some large ATP binding sites by Pck was due to the proximity of the ATP binding site with other domains or the lack of structural elements forming the ATP binding site in the chain. Next, robust amino acid positions are propagated to all human kinases using the sequence alignment. This allows the inference of amino acids having a side chain subject to participation in the ATP binding site for every human kinase in every conformation. The robust cysteine positions are the focus of this study and can be visualized in 3D on the following website: <http://lbg.igbmc.fr/Kinatwon>.

**Inhibition IC<sub>50</sub> for ABL1 wt, KIT wt, and PDGFR $\alpha$  wt Kinases.** A radiometric protein kinase assay (<sup>33</sup>PanQinase Activity Assay) was used for measuring the kinase activity of ABL1 wt, KIT wt, and PDGFR $\alpha$  wt kinases (ProQinase, Freiburg, Germany). All kinase assays were performed in 96-well FlashPlates from Perkin-Elmer (Boston, MA) using 50  $\mu$ L of assay buffer (60 mM HEPES-NaOH, pH 7.5, 3 mM MgCl<sub>2</sub>, 3 mM MnCl<sub>2</sub>, 3  $\mu$ M Na<sub>3</sub>VO<sub>4</sub>, 1.2 mM DTT, 50  $\mu$ g mL<sup>-1</sup> PEG2000, and 1  $\mu$ M [ $\gamma$ -<sup>33</sup>P]ATP), 20 ng of kinase, and a generic substrate (polyGluTyr for KIT and polyAlaGluLsTyr for ABL and PDGFR $\alpha$ ) with 1% DMSO. The test compound concentration ranged from 20  $\mu$ M to 0.1 nM (semilog dilution). The assays were performed by premixing the ATP solution with the test compound and addition of this solution to the kinase/substrate solution. After 60 min at 30 °C, the reaction was stopped with 50  $\mu$ L of 2% H<sub>3</sub>PO<sub>4</sub>, plates were aspirated and treated with 200  $\mu$ L of 0.9% NaCl (2 $\times$ ), and the level of <sup>33</sup>P incorporation was determined with a microplate scintillation counter (MicroBeta, Wallac).

**Mass Spectrometry of KIT Kinase and PDGFR $\alpha$  Kinase Complexes.** A 50  $\mu$ M solution of each compound was incubated with 2  $\mu$ g of each kinase in 60 mM Hepes-NaOH, pH 7.5 (3 mM MgCl<sub>2</sub> and 3 mM MnCl<sub>2</sub>), for 12 h. The proteins were then isolated by one-dimensional sodium dodecyl sulfate–polyacrylamide gel electrophoresis gel, subjected to in-gel trypsin digest, and analyzed by matrix-assisted laser desorption/ionization. As compound 2 cannot form a covalent adduct, this sample was used as a negative control. Only compound 3 showed a new peak corresponding to the expected molecular weight from the peptide adduct (peptide containing Cys 466, MW = 1092.487 for recombinant KIT Kinase, and Cys 491 for recombinant

PDGFR $\alpha$  kinase, MW = 1078.472). To further confirm the identity of the peptide sequence of the new peak, MS/MS experiments were performed, which indeed provide the expected degradation product corresponding to the following sequences: KIT = NC(compound 3)IHR; PDGFR $\alpha$  = NC(compound 3)VHR.

**Kinase Selectivity Profile.** The selectivity profile was measured by using the KinomeScan technology (<http://www.kinomescan.com/>) based on active site-dependent competition binding assays.<sup>33,37</sup> The results are expressed as a percentage of signal between a negative (DMSO) and a positive (known binder) control. Residual binding (%) = [(tested compound – positive control)/(tested compound – negative control)  $\times$  100].

## ■ ASSOCIATED CONTENT

**S Supporting Information.** Tables, figure, synthetic protocols and full characterization of compounds 1–5, peptide sequences, and mass spectrometric characterization of kinase–inhibitor adduct. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\* (J.-M.W.) Tel: +33 388 655 796. E-mail: [wurtz@igbmc.fr](mailto:wurtz@igbmc.fr). (N.W.) Tel: +33 368 855 112. E-mail: [wingssinger@unistra.fr](mailto:wingssinger@unistra.fr).

## ■ ACKNOWLEDGMENT

The Institut Universitaire de France (IUF) and the Institut de Recherche Pierre Fabre (IRPF) are gratefully acknowledged for their support.

## ■ ABBREVIATIONS USED

ABL, Abelson tyrosine kinase; BTK, Bruton's tyrosine kinase; CAMK, calcium/calmodulin-dependent kinases; CDK2, cell division protein kinase 2; CSF1R, macrophage colony-stimulating factor 1 receptor; DFG, aspartate–phenylalanine–glycine; DDR1, epithelial discoidin domain-containing receptor 1; EGFR, epidermal growth factor receptor; FGFR, basic fibroblast growth factor receptor; JNK, c-Jun NH2-terminal kinase; KIT, mast/stem cell growth factor receptor; MEK, mitogen-activated protein kinase kinase; MS, mass spectrometry; PDB, Protein Data Bank; PDGFR, platelet-derived growth factor receptor; RSK, ribosomal protein S6 kinase; TK, tyrosine kinases; VEGFR, vascular endothelial growth factor receptor

## ■ REFERENCES

- (1) Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. The Protein Kinase Complement of the Human Genome. *Science* **2002**, *298*, 1912–1934.
- (2) Bennett, B. L. c-Jun N-terminal kinase-dependent mechanisms in respiratory disease. *Eur. Respir. J.* **2006**, *28*, 651–661.
- (3) Dhillon, A. S.; Hagan, S.; Rath, O.; Kolch, W. MAP kinase signalling pathways in cancer. *Oncogene* **2007**, *26*, 3279–3290.
- (4) Wang, Y. Mitogen-Activated Protein Kinases in Heart Development and Diseases. *Circulation* **2007**, *116*, 1413–1423.
- (5) Cuenda, A.; Rousseau, S. p38 MAP-kinases pathway regulation, function and role in human diseases. *Biochim. Biophys. Acta* **2007**, *1773*, 1358–1375.
- (6) Pandya, N.; Santani, D.; Jain, S. Role of mitogen-activated protein (MAP) kinases in cardiovascular diseases. *Cardiovasc. Drug Rev.* **2005**, *23*, 247–254.

- (7) Zuccotto, F.; Ardini, E.; Casale, E.; Angiolini, M. Through the "Gatekeeper Door": Exploiting the Active Kinase Conformation. *J. Med. Chem.* **2010**, *53*, 2681–2694.
- (8) Zhang, J.; Yang, P. L.; Gray, N. S. Targeting cancer with small molecule kinase inhibitors. *Nat. Rev. Cancer* **2009**, *9*, 28–39.
- (9) Scaltriti, M.; Verma, C.; Guzman, M.; Jimenez, J.; Parra, J. L.; Pedersen, K.; Smith, D. J.; Landolfi, S.; Ramon y Cajal, S.; Arribas, J.; Baselga, J. Lapatinib, a HER2 tyrosine kinase inhibitor, induces stabilization and accumulation of HER2 and potentiates trastuzumab-dependent cell cytotoxicity. *Oncogene* **2009**, *28*, 803–814.
- (10) Buchdunger, E.; Zimmermann, J.; Mett, H.; Meyer, T.; Müller, M.; Druker, B. J.; Lydon, N. B. Inhibition of the Abl protein-tyrosine kinase in vitro and in vivo by a 2-phenylaminopyrimidine derivative. *Cancer Res.* **1996**, *56*, 100–104.
- (11) Andzelm, E. R.; Lew, J.; Taylor, S. Bound to activate: Conformational consequences of cyclin binding to CDK2. *Structure* **1995**, *3*, 1135–1141.
- (12) Nagar, B.; Hantschel, O.; Young, M. A.; Scheffzek, K.; Veach, D.; Bornmann, W.; Clarkson, B.; Superti-Furga, G.; Kuriyan, J. Structural Basis for the Autoinhibition of c-Abl Tyrosine Kinase. *Cell* **2003**, *112*, 859–871.
- (13) Liao, J. J. Molecular Recognition of Protein Kinase Binding Pockets for Design of Potent and Selective Kinase Inhibitors. *J. Med. Chem.* **2007**, *50*, 409–424.
- (14) Cohen, M. S.; Zhang, C.; Shokat, K. M.; Taunton, J. Structural Bioinformatics-Based Design of Selective, Irreversible Kinase Inhibitors. *Science* **2005**, *308*, 1318–1321.
- (15) Potashman, M. H.; Duggan, M. E. Covalent Modifiers: An Orthogonal Approach to Drug Design. *J. Med. Chem.* **2009**, *52*, 1231–1246.
- (16) Yun, C.; Mengwasser, K. E.; Toms, A. V.; Woo, M. S.; Greulich, H.; Wong, K.; Meyerson, M.; Eck, M. J. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 2070–2075.
- (17) Allen, L. F.; Eiseman, I. A.; Fry, D. W.; Lenehan, P. F. CI-1033, an irreversible pan-erbB receptor inhibitor and its potential application for the treatment of breast cancer. *Semin. Oncol.* **2003**, *30*, 65–78.
- (18) Laheru, D.; Croghan, G.; Bukowski, R.; Rudek, M.; Messersmith, W.; Erlichman, C.; Pelley, R.; Jimeno, A.; Donehower, R.; Boni, J.; Abbas, R.; Martins, P.; Zacharchuk, C.; Hidalgo, M. A Phase I Study of EKB-569 in Combination with Capecitabine in Patients with Advanced Colorectal Cancer. *Clin. Cancer Res.* **2008**, *14*, S602–S609.
- (19) Spicer, J.; Rudman, S. EGFR inhibitors in non-small cell lung cancer (NSCLC): The emerging role of the dual irreversible EGFR/HER2 inhibitor BIBW 2992. *Targeted Oncol.* **2010**, *5*, 245–255.
- (20) For another example (structure undisclosed), see Engelman, J. A.; Zejnullahu, K.; Gale, C.; Lifshits, E.; Gonzales, A. J.; Shimamura, T.; Zhao, F.; Vincent, P. W.; Naumov, G. N.; Bradner, J. E.; Althaus, I. W.; Gandhi, L.; Shapiro, G. L.; Nelson, J. M.; Heymach, J. V.; Meyerson, M.; Wong, K.; Janne, P. A. PF00299804, an Irreversible Pan-ERBB Inhibitor, Is Effective in Lung Cancer Models with EGFR and ERBB2 Mutations that Are Resistant to Gefitinib. *Cancer Res.* **2007**, *67*, 11924–11932.
- (21) Cha, M. Y.; Lee, K.; Kim, J. W.; Lee, C. G.; Song, J. Y.; Kim, Y. H.; Lee, G. S.; Park, S. B.; Kim, M. S. Discovery of A Novel Her-1/Her-2 Dual Tyrosine Kinase Inhibitor for the Treatment of Her-1 Selective Inhibitor-Resistant Non-small Cell Lung Cancer. *J. Med. Chem.* **2009**, *52*, 6880–6888.
- (22) Rabindran, S. K.; Discifani, C. M.; Rosfjord, E. C.; Baxter, M.; Floyd, M. B.; Golas, J.; Hallett, W. A.; Johnson, B. D.; Nilakantan, R.; Overbeek, E.; Reich, M. F.; Shen, R.; Shi, X.; Tsou, H.; Wang, Y.; Wissner, A. Antitumor Activity of HKI-272, an Orally Active, Irreversible Inhibitor of the HER-2 Tyrosine Kinase. *Cancer Res.* **2004**, *64*, 3958–3965.
- (23) Blair, J. A.; Rauh, D.; Kung, C.; Yun, C.; Fan, Q.; Rode, H.; Zhang, C.; Eck, M. J.; Weiss, W. A.; Shokat, K. M. Structure-guided development of affinity probes for tyrosine kinases using chemical genetics. *Nat. Chem. Biol.* **2007**, *3*, 229–238.
- (24) Zhou, W.; Ercan, D.; Chen, L.; Yun, C.; Li, D.; Capelletti, M.; Cortot, A. B.; Chiriac, L.; Iacob, R. E.; Padera, R.; Engen, J. R.; Wong, K.; Eck, M. J.; Gray, N. S.; Janne, P. A. Novel mutant-selective EGFR kinase inhibitors against EGFR T790M. *Nature* **2009**, *462*, 1070–1074.
- (25) Fry, D. W.; Bridges, A. J.; Denny, W. A.; Doherty, A.; Greis, K. D.; Hicks, J. L.; Hook, K. E.; Keller, P. R.; Leopold, W. R.; Loo, J. A.; McNamara, D. J.; Nelson, J. M.; Sherwood, V.; Smaill, J. B.; Trumpp-Kallmeyer, S.; Dobrusin, E. M. Specific, irreversible inactivation of the epidermal growth factor receptor and erbB2, by a new class of tyrosine kinase inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12022–12027.
- (26) Wood, E. R.; Shewchuk, L. M.; Ellis, B.; Brignola, P.; Brashear, R. L.; Caferro, T. R.; Dickerson, S. H.; Dickson, H. D.; Donaldson, K. H.; Gaul, M.; Griffin, R. J.; Hassell, A. M.; Keith, B.; Mullin, R.; Petrov, K. G.; Reno, M. J.; Rusnak, D. W.; Tadepalli, S. M.; Ulrich, J. C.; Wagner, C. D.; Vandervall, D. E.; Waterson, A. G.; Williams, J. D.; White, W. L.; Uehling, D. E. 6-Ethynylthieno[3,2-d]- and 6-ethynylthieno[2,3-d]pyrimidin-4-anilines as tunable covalent modifiers of ErbB kinases. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 2773–2778.
- (27) Pan, Z.; Scheerens, H.; Li, S.; Schultz, B. E.; Sprengeler, P. A.; Burrill, L. C.; Mendonca, R. V.; Sweeney, M. D.; Scott, K. C. K.; Grothaus, P. G.; Jeffery, D. A.; Spoerke, J. M.; Honigberg, L. A.; Young, P. R.; Dalrymple, S. A.; Palmer, J. T. Discovery of selective irreversible inhibitors for Bruton's tyrosine kinase. *ChemMedChem* **2007**, *2*, 58–61.
- (28) Ohori, M.; Kinoshita, T.; Yoshimura, S.; Warizaya, M.; Nakajima, H.; Miyake, H. Role of a cysteine residue in the active site of ERK and the MAPKK family. *Biochim. Biophys. Res. Commun.* **2007**, *353*, 633–637.
- (29) Jogireddy, R.; Dakas, P.; Valot, G.; Barluenga, S.; Winssinger, N. Synthesis of a Resorcylic Acid Lactone (RAL) Library Using Fluorous-Mixture Synthesis and Profile of its Selectivity Against a Panel of Kinases. *Chem.—Eur. J.* **2009**, *15*, 11498–11506. Dakas, P.-Y.; Barluenga, S.; Totzke, F.; Zirrgiebel, U.; Winssinger, N. Modular Synthesis of Radicicol A and Related Resorcylic Acid Lactones, Potent Kinase Inhibitors. *Angew. Chem., Int. Ed.* **2007**, 6899–6902.
- (30) Rastelli, G.; Rosenfeld, R.; Reid, R.; Santi, D. V. Molecular modeling and crystal structure of ERK2-hypothemycin complexes. *J. Struct. Biol.* **2008**, *164*, 18–23.
- (31) Schirmer, A.; Kennedy, J.; Murli, S.; Reid, R.; Santi, D. V. Targeted covalent inactivation of protein kinases by resorcylic acid lactone polyketides. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 4234–4239.
- (32) Zhou, W.; Hur, W.; McDermott, U.; Dutt, A.; Xian, W.; Ficarro, S. B.; Zhang, J.; Sharma, S. V.; Brugge, J.; Meyerson, M.; Settleman, J.; Gray, N. S. A structure-guided approach to creating covalent FGFR inhibitors. *Chem. Biol.* **2010**, *17*, 285–295.
- (33) Karaman, M. W.; Herrgard, S.; Treiber, D. K.; Gallant, P.; Atteridge, C. E.; Campbell, B. T.; Chan, K. W.; Ciceri, P.; Davis, M. L.; Edeen, P. T.; Faraoni, R.; Floyd, M.; Hunt, J. P.; Lockhart, D. J.; Milanov, Z. V.; Morrison, M. J.; Pallares, G.; Patel, H. K.; Pritchard, S.; Wodicka, L. M.; Zarrinkar, P. P. A quantitative analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* **2008**, *26*, 127–132.
- (34) Santarius, T.; Shipley, J.; Brewer, D.; Stratton, M. R.; Cooper, C. S. A census of amplified and overexpressed human cancer genes. *Nat. Rev. Cancer* **2010**, *10*, 59–64.
- (35) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (36) Aleksandrov, A.; Simonson, T. A molecular mechanics model for imatinib and imatinib:kinase binding. *J. Comput. Chem.* **2010**, *31*, 1550–1560.
- (37) Fabian, M. A.; Biggs, W. H.; Treiber, D. K.; Atteridge, C. E.; Azimioara, M. D.; Benedetti, M. G.; Carter, T. A.; Ciceri, P.; Edeen, P. T.; Floyd, M.; Ford, J. M.; Galvin, M.; Gerlach, J. L.; Grotzfeld, R. M.; Herrgard, S.; Insko, D. E.; Insko, M. A.; Lai, A. G.; Lelias, J.; Mehta, S. A.; Milanov, Z. V.; Velasco, A. M.; Wodicka, L. M.; Patel, H. K.; Zarrinkar, P. P.; Lockhart, D. J. A small molecule-kinase interaction map for clinical kinase inhibitors. *Nat. Biotechnol.* **2005**, *23*, 329–336.
- (38) Michalczyk, A.; Klüter, S.; Rode, H. B.; Simard, J. R.; Grütter, C.; Rabiller, M.; Rauh, D. Structural insights into how irreversible

inhibitors can overcome drug resistance in EGFR. *Bioorg. Med. Chem.* **2008**, *16*, 3482–3488.

(39) Blencke, S.; Zech, B.; Engkvist, O.; Greff, Z.; Orfi, L.; Horváth, Z.; Kéri, G.; Ullrich, A.; Daub, H. Characterization of a Conserved Structural Determinant Controlling Protein Kinase Sensitivity to Selective Inhibitors. *Chem. Biol.* **2004**, *11*, 691–701.

(40) Rossi, A.; Kapahi, P.; Natoli, G.; Takahashi, T.; Chen, Y.; Karin, M.; Santoro, M. G. Anti-inflammatory cyclopentenone prostaglandins are direct inhibitors of I[ $\kappa$ ]B kinase. *Nature* **2000**, *403*, 103–118.

(41) Copeland, R. A.; Pompliano, D. L.; Meek, T. D. Drug-target residence time and its implications for lead optimization. *Nat. Rev. Drug Discovery* **2006**, *5*, 730–739.

(42) Smith, A. J. T.; Zhang, X.; Leach, A. G.; Houk, K. N. Beyond Picomolar Affinities: Quantitative Aspects of Noncovalent and Covalent Binding of Drugs to Proteins. *J. Med. Chem.* **2009**, *52*, 225–233.

(43) Knight, Z. A.; Shokat, K. M. Features of Selective Kinase Inhibitors. *Chem. Biol.* **2005**, *12*, 621–637.

(44) Toursel, T.; Bard, N.; Bolze, R.; Caron, E.; Desprez, F.; Heymann, M.; Friedrich, A.; Moulinier, L.; Nguyen, N. H.; Poch, O. Décryphon grid - grid resources dedicated to neuromuscular disorders. *Stud. Health Technol. Inform.* **2010**, *159*, 124–133.

(45) BIRD 2010; <http://alnitak.u-strasbg.fr/wikili/index.php/BIRD>.

(46) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. PDB2PQR: An automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Res.* **2004**, *32*, W665–W667.

(47) Albou, L.; Schwarz, B.; Poch, O.; Wurtz, J. M.; Moras, D. Defining and characterizing protein surface using alpha shapes. *Proteins: Struct., Funct., Bioinf.* **2009**, *76*, 1–12.

(48) Edelsbrunner, H.; Facello, M.; Liang, J. On the definition and the construction of pockets in macromolecules. *Pac. Symp. Biocomput.* **1996**, 272–287.

(49) Liang, J.; Edelsbrunner, H.; Woodward, C. Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design. *Protein Sci.* **1998**, *7*, 1884–1897.

(50) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577–2637.

# Cysteine Mapping in Conformationally Distinct Kinase Nucleotide Binding Sites: Application to the Design of Selective Covalent Inhibitors.

*Emeline Leproult<sup>a</sup>, Sofia Barluenga<sup>b</sup>, Dino Moras<sup>a</sup>, Jean-Marie Wurtz<sup>a\*</sup>,  
Nicolas Winssinger<sup>b\*</sup>*

<sup>a</sup> Institut de Génétique et de Biologie Moléculaire et Cellulaire – CNRS (UMR 7104), INSERM (U964), 1 rue Laurent Fries, Université de Strasbourg, 67400 Illkirch, France; + 33 388 655 796; [wurtz@igbmc.fr](mailto:wurtz@igbmc.fr)

<sup>b</sup> Institut de Science et Ingénierie Supramoléculaire - CNRS (UMR 7006), Université de Strasbourg, 8 allée Gaspard Monge, 67000 Strasbourg, France; +33 368 855 112 ; [winssinger@unistra.fr](mailto:winssinger@unistra.fr)

## SUPPORTING INFORMATION (*partiel*)

Table SI 1: The exact list of human kinases having a cysteine in the ATP binding site, considering that all human kinases can adopt the active conformation. AGC, CAMK, CK1, CMGC, STE TK, TKL, RGC and Other human kinase subfamilies are colour-coded in dark blue, green, orange, light blue, violet, red, pink, grey and black, respectively.....	142
Table SI 2: The exact list of human kinases having a cysteine (or more) in the enlarged ATP binding site, considering that all human kinases can adopt the inactive “C-helix-out” conformation. AGC, CAMK, CK1, CMGC, STE TK, TKL, RGC and Other human kinase subfamilies are colour-coded in dark blue, green, orange, light blue, violet, red, pink, grey and black, respectively. ....	144
Table SI 3: The exact list of human kinases having a cysteine (or more) in the extension of the ATP binding site, considering that all human kinases can adopt the inactive “DFG-out” conformation. AGC, CAMK, CK1, CMGC, STE TK, TKL, RGC and Other human kinase subfamilies are colour-coded in dark blue, green, orange, light blue, violet, red, pink, grey and black, respectively. ....	145

Table SI 1: The exact list of human kinases having a cysteine in the ATP binding site, considering that all human kinases can adopt the active conformation. AGC, CAMK, CK1, CMGC, STE TK, TKL, RGC and Other human kinase subfamilies are colour-coded in dark blue, green, orange, light blue, violet, red, pink, grey and black, respectively.

Structural elements in the catalytic domain	Residues corresponding to cysteine positions on EGFR kinase (PDB code: 2ITV)	Human kinase list	Human kinase number
β1	LEU718	Other_NKF3__SgK269 TKL_MLK_MLK_ZAK Other_TBCK__TBCK	3
	GLY719	Other_NKF3__SgK223 Other_NKF3__SgK269 TK_TK-Unique__SuRTK106	3
P-loop	GLY721	TK_Ror__ROR1 Other_NKF2__PINK1	2
	ALA722	TK_Src__SRC TK_Src__YES TK_Src__FGR TK_FGFR__FGFR1 TK_FGFR__FGFR2 TK_FGFR__FGFR3 TK_FGFR__FGFR4 TK_Ack__TNK1 TKL_LISK_LIMK_LIMK1	9
	PHE723	CAMK_RAD53__CHK2 STE_STE7__MAP2K7 TK_TK-Unique__SuRTK106	3
	GLY724	CAMK_CAMK-Unique__VACAMKL Other_BUB__BUBR1	2
β2	VAL726	CAMK_RSKb_MSKb_Domain2_MSK1 CAMK_RSKb_MSKb_Domain2_MSK2 CAMK_RSKb_RSKb_Domain2_RSK1 CAMK_RSKb_RSKb_Domain2_RSK2 CAMK_RSKb_RSKb_Domain2_RSK3 CAMK_RSKb_RSKb_Domain2_RSK4 Other_PLK__PLK1 Other_PLK__PLK2 Other_PLK__PLK3 STE_STE11__MAP3K1 Other_NEK__NEK2	11
β3	ALA743	TK_EGFR__HER3/ErbB3 Other_PEK__HRI	2
	LYS745	Other_Wnk__Wnk1 Other_Wnk__Wnk2 Other_Wnk__Wnk3 Other_Wnk__Wnk4	4
	LEU747	Other_IKK__IKKa Other_IKK__IKKb TK_Fer__FER TK_Fer__FES TK_Fak__FAK TK_Fak__PYK2	6
C-helix	GLU758	AGC_PKC_Alpha_PKCa AGC_PKC_Alpha_PKCb AGC_PKC_Alpha_PKCg AGC_PKC_Eta_PKCh AGC_PKC_Eta_PKCe AGC_PKC_Delta_PKCd AGC_PKC_Delta_PKCe AGC_DMPK_GEK_MRCKa AGC_DMPK_GEK_MRCKb AGC_DMPK_GEK_DMPK1 AGC_DMPK_GEK_DMPK2 CAMK_Trbl__Trb2	12
	ILE759	AGC_RSK__SgK494 CAMK_PSK__PSKH1 CAMK_PSK__PSKH2 CAMK_PIM__PIM2 Other_NEK__NEK6 Other_NEK__NEK7 Other_NAK__AAK1 Other_NAK__BIKE	8
	MET766	CAMK_CAMK2__CaMK2a CAMK_CAMK2__CaMK2b CAMK_CAMK2__CaMK2d CAMK_CAMK2__CaMK2g CAMK_CASK__CASK CAMK_Trbl__Trb2 TKL_IRAK__IRAK2 STE_STE-Unique__NIK	8
β4	CYS775	CAMK_CAMKL_MELK_MELK TK_EGFR__EGFR CK1_TTBK__TTBK1 CK1_TTBK__TTBK2 Other_TBCK__TBCK	5
β5/gatekeeper	THR790	AGC_RSK__SgK494 CMGC_RCK__MOK	2
Hinge region	LEU792	AGC_RSK_p70_p70S6b CAMK_MAPKAPK_MAPKAPK_MAPKAPK2 CAMK_MAPKAPK_MAPKAPK_MAPKAPK3 TK_FGFR__FGFR4 Other_TTK__TTK	5
	MET793	Other_IRE__IRE1 Other_IRE__IRE2 AGC_PKG__PKG1 AGC_PKG__PKG2 AGC_RSK__SgK494 CAMK_CAMK-Unique__STK33 CAMK_CAMKL_MELK_MELK CAMK_CAMKL_HUNK_HUNK CAMK_CAMKL_LKB_LKB1 CAMK_CAMKL_CHK1_CHK1 Other_PLK__PLK3 Other_PLK__PLK2 Other_PLK__PLK1 Other_PLK__PLK4 CAMK_Trio__SPEG CAMK_Trio__Obscn CAMK_Trio__Domain2_SPEG CAMK_Trio__Domain2_Obscn CMGC_CDK_CDK5_CDK5 CMGC_CDK_CDK10_CDK10 CMGC_CDK_CDK9_CDK9 CMGC_CDKL__CDKL1 CMGC_CDKL__CDKL4 Other_IKK__IKKa Other_IKK__IKKb Other_IKK__IKKe Other_IKK__TBK1 STE_STE11__MAP3K4 STE_STE20_KHS_KHS1 STE_STE20_KHS_KHS2 STE_STE20_KHS_GCK STE_STE20_KHS_HPK1 STE_STE20_MSN_ZC1/HGK STE_STE20_MSN_ZC3/MINK STE_STE20_MSN_ZC2/TNIK STE_STE20_MSN_ZC4/NRK STE_STE20_NinaC_MYO3B STE_STE20_NinaC_MYO3A STE_STE20_MST_MST1 STE_STE20_MST_MST2 STE_STE20_SLK_SLK STE_STE20_SLK_LOK STE_STE20_TAO_TAO2 STE_STE20_TAO_TAO1 STE_STE20_TAO_TAO3 Other_NKF4__CLIK1 Other_NKF4__CLIK1L Other_ULK__ULK1 Other_ULK__ULK2 Other_ULK__ULK3 Other_ULK__ULK4 Other_NEK__NEK1 Other_NEK__NEK5 Other_NEK__NEK3 Other_NEK__NEK4 Other_NEK__NEK2 Other_NEK__NEK9 Other_NEK__NEK11 Other_TLK__TLK1 Other_TLK__TLK2 Other_NAK__AAK1 Other_NAK__BIKE Other_NAK__GAK Other_PEK__PEK Other_PEK__PKR Other_PEK__GCN2 Other_WEE__Wee1 Other_WEE__Wee1B Other_WEE__MYT1 RGC_RGC__ANPa RGC_RGC__ANPb RGC_RGC__HSER RGC_RGC__CYGD RGC_RGC__CYGF TK_VEGFR__KDR TK_VEGFR__FLT1 TK_VEGFR__FLT4 TK_PDGF	98

Cartographie des Cystéines et Application à la Conception d'Inhibiteurs Covalents

		R_FMS TK_PDGFR_KIT TK_PDGFR_PDGFRa TK_PDGFR_PDGFRb TK_PDGFR_FLT3 TK_Ror_ROR2 TK_Fak_FAK TK_Lmr_LMR1 TK_Lmr_LMR3 TK_Lmr_LMR2 TKL_MLk_LZK_DLK TKL_MLk_LZK_LZK TKL_RAF_BRAF TKL_RAF_RAF1 TKL_RAF_ARAF TKL_RAF_KSR1 TKL_RAF_KSR2 Other_PEK_HRI Other_TBCK_TBCK TKL_TKL-Unique_MLKL Other_Other-Unique_RNaseL	
	PRO794	Other_NKF3_SgK269 TK_Src_FGR TK_PDGFR_FMS TK_PDGFR_KIT TK_PDGFR_FLT3 TK_Met RON	6
	GLY796	TK_Eph_EphB3	1
	CYS797	TK_Tec_BTk TK_Tec_BMX TK_Tec_TEC TK_Tec_TXK TK_Tec_ITK TK_Src_BLK TK_EGFR_EGFR TK_EGFR_HER2/ErbB2 TK_EGFR_HER4/ErbB4 TK_JakA_JAK3 STE_STE7_MAP2K7	11
D-helix	ASP800	CAMK_Trio_Domain2_SPEG	1
Catalytic loop	ALA839	Other_NKF3_SgK223 TK_JakB_Domain2_JAK1 TK_JakB_Domain2_TYK2 TK_JakB_Domain2_JAK2 Other_SCY1_SCYL1 Other_SCY1_SCYL3	6
	ARG841	CAMK_Trbl_Trbl3	1
	ASN842	CAMK_CASK_CASK Other_BUB_BUBR1	2
Before DFG motif	THR854	CAMK_MAPKAPK_MAPKAPK_MAPKAPK5 CAMK_MAPKAPK_MNK_MNK1 CAMK_MAPKAPK_MNK_MNK2 CAMK_RSKb_RSKb_Domain2_RSK2 CAMK_RSKb_RSKb_Domain2_RSK3 CAMK_RSKb_RSKb_Domain2_RSK1 CAMK_RSKb_RSKb_Domain2_RSK4 CAMK_PKD_PKD1 CAMK_PKD_PKD3 CAMK_PKD_PKD2 CAMK_Trio_SPEG CAMK_Trio_Obscn CMGC_MAPK_ERK_Erk1 CMGC_MAPK_ERK_Erk2 CMGC_MAPK_nmo_NLK CMGC_MAPK_Erk7_Erk7 CMGC_CDKL_CDKL2 CMGC_CDKL_CDKL3 CMGC_CDKL_CDKL1 CMGC_CDKL_CDKL4 CMGC_CDKL_CDKL5 CMGC_GSK_GSK3A CMGC_GSK_GSK3B CMGC_DYRK_PRP4_PRP4 STE_STE7_MAP2K1 STE_STE7_MAP2K2 STE_STE7_MAP2K5 STE_STE7_MAP2K6 STE_STE7_MAP2K3 STE_STE7_MAP2K4 STE_STE7_MAP2K7 Other_ULK_Fused Other_NAK_AAK1 Other_NAK_BIKE Other_NAK_GAK TKL_STKR_Type2_TGFbR2 TK_VEGFR_KDR TK_VEGFR_FLT1 TK_VEGFR_FLT4 TK_PDGFR_KIT TK_PDGFR_PDGFRa TK_PDGFR_PDGFRb TK_PDGFR_FLT3 TKL_MLk_TAK1_TAK1 TKL_MLk_MLk_ZAK STE_STE-Unique_NIK TK_TK-Unique_SuRTK106 Other_TOPK_PBK	48
A-loop	GLY857	Other_ULK_ULK4 CAMK_CAMK-Unique_SgK495 Other_SCY1_SCYL2	3
	ARG858	STE_STE11_MAP3K4 STE_STE11_MAP3K8 RGC_RGC_HSER Other_MOS_MOS Other_NKF2_PINK1	5

Table SI 2: The exact list of human kinases having a cysteine (or more) in the enlarged ATP binding site, considering that all human kinases can adopt the inactive “C-helix-out” conformation. AGC, CAMK, CK1, CMGC, STE TK, TKL, RGC and Other human kinase subfamilies are colour-coded in dark blue, green, orange, light blue, violet, red, pink, grey and black, respectively.

Structural elements in the catalytic domain	Residues corresponding to cysteine positions on EGFR kinase (PDB code: 1XKK)	Human kinase list	Human kinase number
β2	LYS728	CAMK_CAMKL_SNRK_SNRK CAMK_Trbl_Trbl TK_Eph_EphA10	3
β4	LEU777	CAMK_MLCK_smMLCK STE_STE7_MAP2K7 TK_Alk_ALK TK_Alk_LTK TK_Lmr_LMR1 TK_Lmr_LMR2 TK_Lmr_LMR3 CK1_VRK_VRK3	8
Catalytic loop	HIS835	AGC_RSKL_RSKL1 AGC_RSKL_RSKL2 TKL_IRAK_IRAK3	3
Activation loop	ALA859	Other_IRE_IRE1 Other_IRE_IRE2 AGC_AKT_AKT1 AGC_AKT_AKT2 AGC_AKT_AKT3 AGC_SGK_SGK AGC_SGK_SGK3 AGC_SGK_SGK2 AGC_PKC_Alpha_PKCα AGC_PKC_Alpha_PKCβ AGC_PKC_Alpha_PKCγ AGC_PKC_Eta_PKCη AGC_PKC_Eta_PKCε AGC_PKC_Delta_PKCδ AGC_PKC_Delta_PKCθ AGC_PKC_Iota_PKCι AGC_PKC_Iota_PKCζ AGC_PKN_PKN1 AGC_PKN_PKN2 AGC_PKN_PKN3 AGC_RSK_p70_p70S6K AGC_RSK_p70_p70S6Kb AGC_DMPK_GEK_MRCKa AGC_DMPK_GEK_MRCKb AGC_DMPK_GEK_DMPK1 AGC_DMPK_GEK_DMPK2 AGC_DMCK_PCK1 AGC_DMCK_PCK2 AGC_NDR_NDR1 AGC_NDR_NDR2 AGC_NDR_LATS1 AGC_NDR_LATS2 CAMK_CAMKL_MELK_MELK CMGC_RCK_RCK STE_STE20_PAKA_PAK1 STE_STE20_PAKA_PAK2 STE_STE20_PAKA_PAK3 STE_STE20_PAKB_PAK4 STE_STE20_PAKB_PAK5 STE_STE20_PAKB_PAK6 Other_Other-Unique_SgK496 Other_NKF2_PINK1	42
	VAL876	CMGC_GSK_GSK3A CMGC_GSK_GSK3B TK_Eph_EphB6 Other_NAK_MPSK1 Other_Other-Unique_SgK493	5
	PRO877	STE_STE7_MAP2K3 STE_STE7_MAP2K4 STE_STE7_MAP2K6 STE_STE7_MAP2K7 CAMK_Trbl_Trbl1 CAMK_Trbl_Trbl2 CAMK_Trbl_Trbl3	7

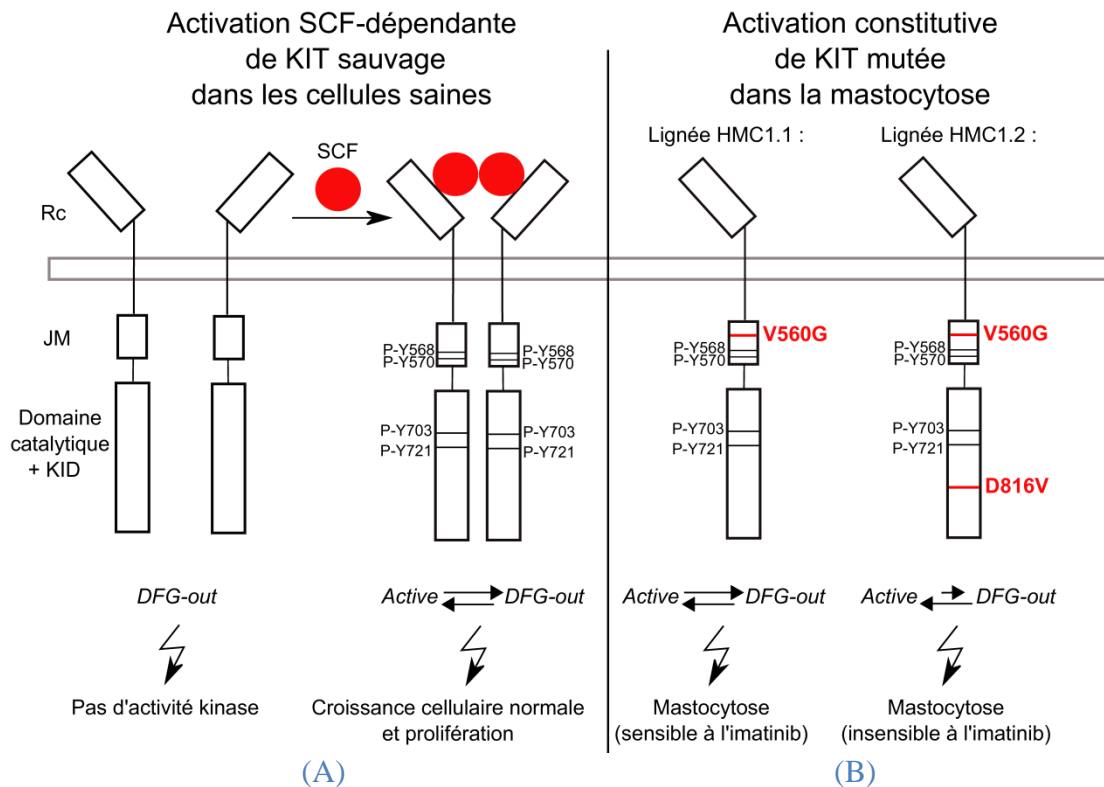


Table SI 3: The exact list of human kinases having a cysteine (or more) in the extension of the ATP binding site, considering that all human kinases can adopt the inactive “DFG-out” conformation. AGC, CAMK, CK1, CMGC, STE TK, TKL, RGC and Other human kinase subfamilies are colour-coded in dark blue, green, orange, light blue, violet, red, pink, grey and black, respectively.

Structural elements in the catalytic domain	Residues corresponding to cysteine positions on KIT kinase (PDB code: 1T46)	Human kinase list	Human kinase number
C-helix	SER639	AGC_PKN_PKN1 AGC_PKN_PKN2 AGC_PKN_PKN3 STE_STE11_MAP3K2 STE_STE11_MAP3K3 TK_Eph_EphA7 CK1_VRK_VRK2	7
	VAL643	CAMK_CAMKL_SNRK_SNRK TK_Ax1_AXL TK_Ax1_MER TK_Ax1_TYRO3 Other_CDC7_CDC7 Other_Other-Unique_RNaseL Other_Slob_Slob	7
	LEU647	CAMK_MAPKAPK_MAPKAPK_MAPKAPK5 CAMK_MAPKAPK_MNK_MNK1 CAMK_MAPKAPK_MNK_MNK2 CAMK_RSKb_MSKb_Domain2_MSK1 CAMK_RSKb_MSKb_Domain2_MSK2 STE_STE20_FRAY_OSR1 STE_STE20_FRAY_STLK3 STE_STE20_KHS_KHS1 STE_STE20_KHS_KHS2 STE_STE20_KHS_GCK STE_STE20_KHS_HPK1 STE_STE20_MST_MST1 STE_STE20_MST_MST2 STE_STE20_YSK_MST3 STE_STE20_YSK_YSK1 STE_STE20_YSK_MST4 STE_STE20_SLK_SLK STE_STE20_SLK_LOK STE_STE7_MAP2K1 STE_STE7_MAP2K2 STE_STE7_MAP2K5 TKL_IRAK_IRAK2 TKL_IRAK_IRAK4	23
β4	ILE653	CMGC_CLK_CLK1 CMGC_CLK_CLK2 CMGC_CLK_CLK3 CMGC_CLK_CLK4 CMGC_DYRK_PRP4 Other_WEE_MYT1 Other_VPS15_PIK3R4 Other_Haspin_Haspin	8
E-helix	LEU783	CAMK_CAMK2_CaMK2a CAMK_CAMK2_CaMK2b CAMK_CAMK2_CaMK2d CAMK_CASK_CASK CAMK_MAPKAPK_MAPKAPK_MAPKAPK5 CAMK_CAMKL_AMPK_AMPKa1 CAMK_CAMKL_AMPK_AMPKa2 CAMK_CAMKL_BRSK_BRSK2 CAMK_CAMKL_BRSK_BRSK1 CAMK_CAMKL_MARK_MARK1 CAMK_CAMKL_MARK_MARK2 CAMK_CAMKL_MARK_MARK3 CAMK_CAMKL_MARK_MARK4 CAMK_CAMKL_QIK_QIK CAMK_CAMKL_QIK_SIK CAMK_CAMKL_QIK_QSK CAMK_CAMKL_NuaK_NuaK1 CAMK_CAMKL_NuaK_NuaK2 CAMK_CAMKL_SNRK_SNRK CAMK_TSSK_TSSK1 CAMK_TSSK_TSSK2 CAMK_TSSK_TSSK3 CAMK_PIM_PIM1 CAMK_PIM_PIM2 CAMK_PIM_PIM3 Other_AUR_Aur1 Other_AUR_AurB Other_AUR_AurC CMGC_CDK_CDC2_CDK2 CMGC_CDK_CDC2_CDK3 CMGC_CDK_CDC2_CDC2 CMGC_CDK_CDK5_CDK5 CMGC_CDK_TAIRE_PCTAIRE1 CMGC_CDK_TAIRE_PCTAIRE2 CMGC_CDK_TAIRE_PCTAIRE3 CMGC_CDK_CCRK CMGC_CDK_CRK7_CHEJ CMGC_CDK_CRK7_CRK7 CMGC_CDKL_CDKL1 CMGC_CDKL_CDKL2 CMGC_CDKL_CDKL4 CMGC_CDKL_CDKL5 Other_CK2_CK2a1 Other_CK2_CK2a2 Other_NEK_NEK2 CAMK_Trbl_Trbl1 CAMK_Trbl_Trbl2 CAMK_Trbl_Trbl3	48
Catalytic loop	CYS788	CMGC_CDKL_CDKL1 CMGC_CDKL_CDKL4 STE_STE20_STLK_STLK6 RGC_RGC_ANPa TK_Fer_FER TK_Fer_FES TK_FGFR_FGFR1 TK_FGFR_FGFR2 TK_FGFR_FGFR3 TK_FGFR_FGFR4 TK_VEGFR_KDR TK_VEGFR_FLT1 TK_VEGFR_FLT4 TK_PDGFR_FMS TK_PDGFR_KIT TK_PDGFR_PDGFRa TK_PDGFR_PDGFRb TK_PDGFR_FLT3 TK_Fak_PYK2 TK_JakA_JAK3	20
	ILE789	CAMK_CAMKL_BRSK_BRSK1 CAMK_CAMKL_BRSK_BRSK2 CMGC_GSK_GSK3A CMGC_GSK_GSK3B	4
	HIS790	AGC_RSKL_RSKL1 AGC_RSKL_RSKL2 TKL_IRAK_IRAK3	3
	ARG791	CAMK_PKD_PKD1 CAMK_PKD_PKD2 CAMK_PKD_PKD3 CMGC_DYRK_Dyrk1_DYRK1A CMGC_DYRK_Dyrk1_DYRK1B CMGC_DYRK_Dyrk2_DYRK2 CMGC_DYRK_Dyrk2_DYRK3 CMGC_DYRK_Dyrk2_DYRK4 Other_ULK_ULK4 Other_NKF3_SgK269 TKL_LRRK_LRRK1	11
P+1 loop	TYR846	TKL_STKR_Type2_BMPR2 Other_CDC7_CDC7 Other_Other-Unique_SgK493	3

## 4.4 Effet antiprolifératif sur des cellules humaines de leucémie à mastocytes

L'activité catalytique de la kinase KIT est essentielle pour le développement normal des cellules humaines de mastocytes. L'activation normale de KIT est déclenchée par la liaison d'une protéine appelée *Stem Cell Factor* (SCF) sur le domaine récepteur de la kinase, conduisant à la dimérisation et à l'autophosphorylation en *trans* de KIT sur sa région juxtamembranaire (tyrosines 568 et 570) (Figure 53A) (Roskoski Jr., 2005). En conséquence, le domaine catalytique, qui était bloqué dans la conformation inactive *DFG out*, retrouve un équilibre conformationnel entre les formes *Active* et *DFG out*. Le domaine d'insertion ou KID (une insertion d'une soixantaine de résidus dans le lobe C-terminal de KIT) est également phosphorylé (tyrosines 703 et 721 ; Figure 53A). Ces divers sites de phosphorylation permettent le recrutement de protéines qui sont ensuite phosphorylées par KIT et induisent la croissance cellulaire (Roskoski Jr., 2005).



**Figure 53 :** Comparaison entre l'activation normale de KIT et deux mutants de KIT ayant une activité catalytique constitutive. (A) L'activation normale de KIT est initiée par la liaison du *Stem Cell Factor* (SCF) sur le domaine récepteur de la kinase (Rc), conduisant à la dimérisation et à l'autophosphorylation en *trans* de KIT sur sa région juxtamembranaire (JM) et sur son domaine d'insertion (KID). Le domaine catalytique de KIT, initialement bloqué dans la conformation inactive *DFG out*, retrouve un équilibre conformationnel entre les formes *Active* et inactive *DFG out*. L'activité catalytique de KIT contribue alors à la croissance normale de la cellule. (B) Dans plusieurs maladies, les mutations KIT-V560G (lignée HMC1.1) et KIT-D816V (lignée HMC1.2) conduisent à l'activité constitutive de KIT, indépendamment de la liaison du SCF. De plus, la localisation de la mutation D816V sur la A-loop augmente l'activité de KIT en déplaçant l'équilibre conformationnel du domaine catalytique vers la conformation *Active*.

L'efficacité du composé **3** (voir Figure 51 à la page 130), capable de se lier covalamment à KIT (voir paragraphe 4.3 « Publication » à la page 131), a été testée sur deux lignées de cellules humaines de leucémie à mastocytes (HMC), une variante de la mastocytose (Orfao et al., 2007; Furitsu et al., 1993). Ces lignées présentent des mutations de la kinase KIT favorisant son activation sans nécessiter la liaison du SCF (Figure 53B). La première lignée, HMC1.1, présente une forme mutée de KIT sur sa région juxtamembranaire (V560G). Malgré cette mutation, HMC1.1 reste sensible à l'imatinib, qui est capable de réduire la prolifération cellulaire en quelques heures (Akin et al., 2003). La seconde lignée cellulaire, HMC1.2, exprime un double mutant de KIT : V560G, D816V. Cette deuxième mutation, observée chez 90% des patients, se situe sur la *A-loop* et a pour effet de déplacer l'équilibre conformationnel vers la forme *Active* du domaine catalytique (Figure 53B) (Orfao et al., 2007). En conséquence, l'imatinib est inefficace sur cette lignée (Akin et al., 2003).

L'IC<sub>50</sub> d'environ 100nM mesurée sur HMC1.1 confirme que le composé **3** est capable de réduire la prolifération des cellules de leucémie à mastocytes, soit une efficacité légèrement inférieure à l'imatinib, dont l'IC<sub>50</sub> mesurée au cours de cette expérience est de l'ordre de 40nM (Figure 54).

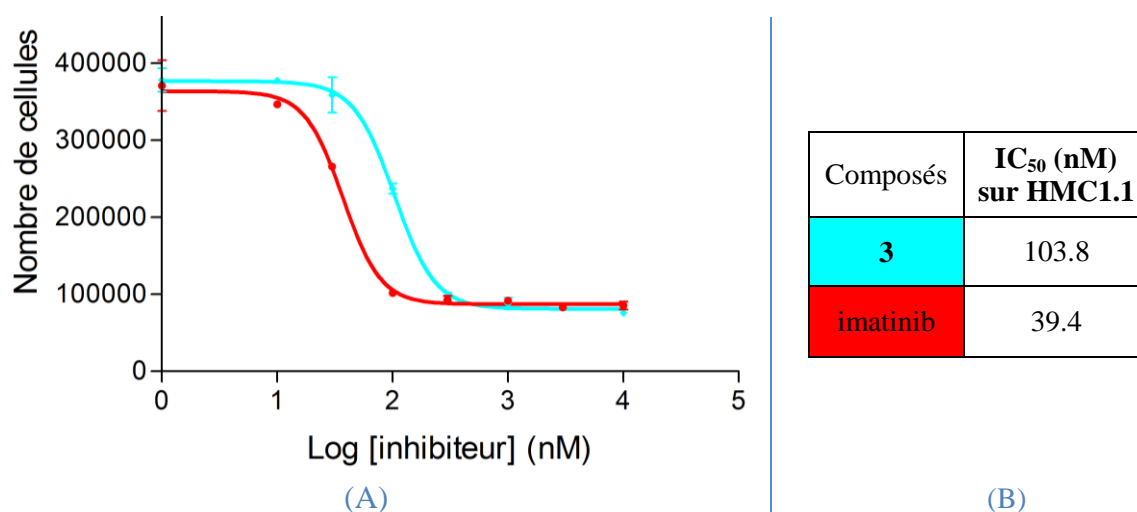


Figure 54 : (A) Profil d'inhibition de la croissance cellulaire dans la lignée HMC1.1 mesuré pour le composé **3** (courbe cyan) et l'imatinib (courbe rouge), en fonction de leur concentration. (B) Déduction de l'IC<sub>50</sub>.

Cependant, ces tests cellulaires ne permettent pas de savoir si l'inhibition observée est due à la formation d'une liaison covalente entre notre inhibiteur et la cible KIT. Dans ce but, nous avons mis au point un protocole, issu d'une réflexion commune avec l'équipe du Docteur L. Brino (IGBMC) et s'inspirant de publications exposant des tests cellulaires avec des inhibiteurs covalents (Honigberg et al., 2010; Zhou et al., 2010; Spicer et Rudman, 2010). Ainsi, les cellules de la lignée HMC1.1 ont été traitées avec le composé **3** et l'imatinib pendant une durée de quatre heures et à une concentration cent fois supérieure à leur IC<sub>50</sub>

respective et précédemment mesurée (Figure 54). Au bout de ces quatre heures, les cellules ont été centrifugées afin de pouvoir retirer le milieu et, par conséquent, éliminer les inhibiteurs qui ne se seraient pas liés de manière covalente à KIT. Du milieu frais a été ensuite remis. Cette étape de « lavage » a été effectuée successivement trois fois, puis une quatrième fois où les cellules ont été laissées pendant une heure dans le milieu avant d'être recentrifugées. Cette dernière étape de lavage permet d'éliminer les inhibiteurs qui se seraient liés (de façon non covalente) à KIT au cours des premiers lavages et qui se seraient ensuite retrouvés libérés dans le milieu du fait des échanges entre les conformations *Active* et inactive *DFG out* de la kinase.

Quarante-huit heures après ces lavages, nous avons pu observer une efficacité prolongée de la molécule **3** sur HMC1.1 (Figure 55). En effet, la molécule **3** est parvenue à ralentir la prolifération cellulaire, contrairement à l'imatinib dont l'efficacité est comparable à celle du DMSO (Figure 55). Autrement dit, l'imatinib a été complètement retirée du milieu au cours des lavages tandis que le composé **3**, qui s'est probablement lié de façon covalente à KIT, n'a pas été éliminé et a pu poursuivre l'inhibition de la kinase. L'évolution de la croissance cellulaire a également été mesurée 72H après le traitement les inhibiteurs. On peut alors observer une reprise de la croissance cellulaire pour les cellules traitées avec le composé **3**, qui peut s'expliquer par la resynthèse de la protéine KIT (Figure 55).

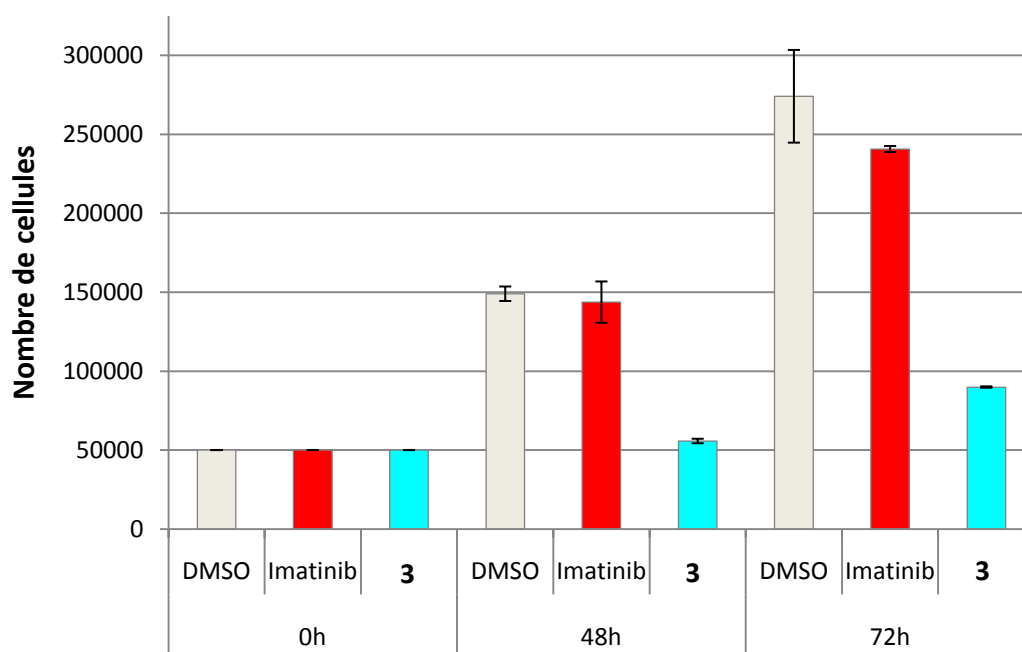


Figure 55 : Évolution de la croissance cellulaire de la lignée HMC1.1 après un temps de traitement de 4H avec les composés suivi de 4 lavages. 48H après traitement, les cellules traitées avec l'imatinib ont poussé normalement tandis que les cellules traitées avec le composé **3** n'ont pas poussé. 72H après traitement avec le composé **3**, la croissance cellulaire semble repartir.

Un avantage sérieux que pourrait avoir le composé **3** vis-à-vis de l'imatinib serait que la formation de cette liaison covalente lui permette de rester efficace face à des formes résistantes de KIT, à l'image des inhibiteurs covalents d'EGFR (Yun et al., 2008). Pour vérifier cela, son  $IC_{50}$  sur la lignée HMC1.2 a été mesurée (Figure 56). Les résultats obtenus confirment une meilleure efficacité de la part de l'inhibiteur covalent **3** sur la lignée HMC1.2, comparé à l'imatinib. Le composé **3** présente une  $IC_{50}$  vingt fois supérieure à celle mesurée sur HMC1.1, soit un ratio proche de celui observé pour des inhibiteurs covalents d'EGFR dans des conditions similaires (Michalczyk et al., 2008). Au contraire, l'imatinib présente une  $IC_{50}$  qui est cinq cent fois supérieure à celle mesurée sur HMC1.1, confirmant la résistance à ce traitement observée chez les patients présentant la mutation D816V de KIT (Akin et al., 2003).

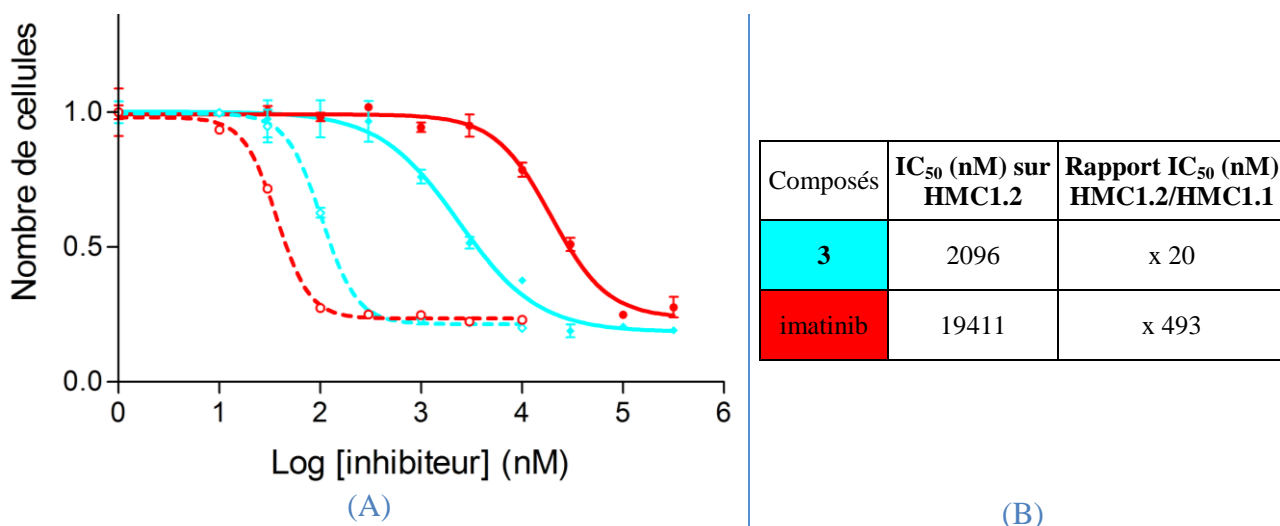


Figure 56 : (A) Profil d'inhibition de la croissance cellulaire sur la lignée cellulaire HMC1.2 mesuré pour le composé **3** et l'imatinib (traits pleins cyan et rouge, respectivement), comparé aux profils obtenus sur HMC1.1 (traits en pointillés cyan et rouge pour le composé **3** et l'imatinib, respectivement). (B) Déduction de l' $IC_{50}$  et comparaison avec l' $IC_{50}$  obtenue sur HMC1.1 (Figure 54). Alors que l'imatinib a perdu son efficacité ( $IC_{50}$  multipliée par 493), le composé **3** est moins sensible à la mutation D816V et conserve une efficacité ( $IC_{50}$  multipliée par 20).

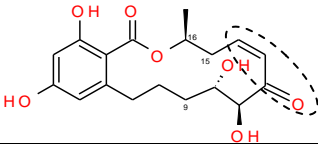
Cette observation fournit une seconde validation pour la formation de la liaison covalente entre le composé **3** et KIT, qui semble être l'explication la plus probable de l'inhibition observée, contrairement à l'imatinib. Un tel résultat supporte l'idée que l'utilisation d'inhibiteurs covalents est très prometteuse dans des cas où le développement de mutations ou de résistances rend inefficace l'utilisation d'inhibiteurs non covalents.

Les détails de ces tests cellulaires, accompagnés d'autres tests validant le mode d'action du composé **3** en comparaison de deux médicaments que ce sont l'imatinib et le dasatinib, devraient faire l'objet d'un prochain article.

## 4.5 Conception rationnelle d'inhibiteurs covalents dérivés de l'hypothemycine

### 4.5.1 Amélioration du profil de sélectivité de l'hypothemycine parmi les kinases ayant une cystéine avant le motif DFG

L'hypothemycine (voir paragraphe 1.2.7.b « Les inhibiteurs covalents dans les publications » à la page 50 ; voir Tableau 4 à la page 52) est une substance naturelle formée d'un acide benzoïque et d'un macrocycle contenant une cétone  $\alpha,\beta$ -insaturée capable de se lier de façon covalente avec une majorité des 48 kinases (Tableau 11 ; voir Figure 25 à la page 54) présentant une cystéine située avant le motif DFG (sphère magenta dans la Figure 50A à la page 128), soit près de 10% du kinome humain (Schirmer et al., 2006). Plusieurs de ces kinases ayant un intérêt thérapeutique, l'équipe du Professeur N. Winssinger (ISIS, Strasbourg) s'est spécialisée dans la synthèse d'inhibiteurs dérivés de l'hypothemycine, dans le but d'atteindre un meilleur profil de sélectivité parmi les kinases humaines inhibées (Jogireddy et al., 2009; Dakas et al., 2009; Barluenga et al., 2010). En particulier, une librairie d'analogues avait été développée en 2009, en faisant essentiellement varier la présence de substituants du type méthyle ou hydroxyle aux positions 9, 15 et 16 du macrocycle, ainsi que la taille du macrocycle (voir la numérotation des atomes sur l'analogue **1afh** dans le Tableau 11) (Jogireddy et al., 2009). Les tests d'IC<sub>50</sub> avaient permis d'identifier plusieurs molécules aux affinités comparables à celles obtenues pour l'hypothemycine, soit des valeurs inférieures à 1 $\mu$ M pour 7 des 10 kinases testées et ayant une cystéine avant le motif DFG (Jogireddy et al., 2009). Ces résultats ont également révélé la difficulté de trouver des molécules spécifiques parmi les cibles thérapeutiques, telles que KDR, KIT, MNK1 et MAP2K1 (Jogireddy et al., 2009). Pour exemple, l'activité résiduelle d'une partie des 48 kinases avait été mesurée après incubation avec la molécule **1afh**, soit l'une des meilleures molécules de la librairie identifiées après la mesure des IC<sub>50</sub> (Tableau 11). De ce criblage, il ressort que **1afh** inhibe 23 kinases (activité résiduelle inférieure à 30%) parmi lesquelles se trouvent les quatre cibles thérapeutiques précitées. Afin de comprendre et améliorer ce profil de sélectivité, j'ai procédé à une analyse des relations structure-activité (SAR) à partir de ces résultats de criblage. Pour cela, j'ai généré les modèles par homologie des kinases pour lesquelles aucune structure en conformation *Active* n'était disponible, soit 40 des 48 kinases. Toutes les structures ont été superposées sur le complexe ERK2/hypothemycine (3C9W).

Kinases ayant une cystéine avant le motif DFG			Particularités de la poche ATP		Activité résiduelle après incubation à 1µM de la molécule <b>1afh</b> (%)
Groupe	Sous-famille	Nom	Résidu <i>Gatekeeper</i>	Longueur de la région charnière	
CAMK	MAPKAPK	MAPKAPK5	M	6	63
	MNK	MNK1*	F	6	4
		<b>MNK2</b>	F	6	0
	RSKb_Domaine2	RSK1*	M	6	59
		RSK2	T	6	N.D.
		RSK3	T	6	67
		RSK4	T	6	8.1
	PKD	PKD1*	M	5	38
		PKD2	M	5	8
		PKD3	M	5	22
	Trio	OBSCN	L	6	N.D.
		SPEG	T	6	N.D.
CMGC	MAPK	ERK1*	Q	5	94
		<b>ERK2*</b>	Q	5	85 (mais inhibé par hypothémicine !)
		NLK	T	5	N.D.
		ERK7	F	5	65
	CDKL	CDKL1	F	5	N.D.
		CDKL2	F	5	83
		CDKL3	F	5	69
		CDKL4	F	5	N.D.
		CDKL5	F	5	100
	GSK	GSK3a	L	6	70
		<b>GSK3b*</b>	L	6	100 (mais inhibé par hypothémicine !)
	DYRK	PRP4	F	6	100
STE	STE7	MAP2K1*	M	6	0
		MAP2K2*	M	6	1
		MAP2K3*	M	5	49
		MAP2K4*	M	5	0
		MAP2K5*	T	6	N.D.
		MAP2K6*	M	5	28
		MAP2K7*	M	5	N.D.
	STE	NIK	M	6	N.D.
TK	VEGFR	FLT1*	V	6	5.8
		FLT4*	V	6	12
		<b>KDR*</b>	V	6	20
	PDGFR	FLT3*	F	6	7
		<b>KIT*</b>	T	6	8.8
		KIT mutée (T670I)	I	6	68
		PDGFRα*	T	6	5
		PDGFRβ*	T	6	1
TK-Unique	SURTK106	L	6	N.D.	
TKL	MLK	ZAK	T	6	15
		<b>TAK1</b>	M	6	3.4
	STKR	TGFBR2*	T	6	10
Autres	ULK	FUSED	T	6	3
	NAK	AAK1	M	6	17
		BIKE	M	6	26
		GAK	T	6	2.8
	TOPK	PBK	M	6	N.D.

**Tableau 11** : Liste des 48 kinases humaines ayant une cystéine située avant le motif DFG, dont 21 cibles thérapeutiques (marquées par \*). L'activité résiduelle (en %, N.D.=non déterminée) après incubation avec la molécule 1afh, un analogue de l'hypothémicine ayant été développé avant le travail de thèse, est présentée en parallèle de quelques caractéristiques du site catalytique de chaque cible. Le groupe électrophile sur la molécule 1afh est entouré en pointillés. Six des huit kinases possédant une structure cristallographique en conformation *Active* (surlignées en gras) ont été utilisées pour la conception rationnelle de ligands plus sélectifs que l'hypothémicine et 1afh.

La partie du site actif montrant le plus de divergence est la zone du résidu *gatekeeper*. À première vue, la majeure partie des kinases présentant un *gatekeeper* petit (valine, thréonine) sont très bien inhibées par **1afh** (activité résiduelle inférieure à 30%): KDR, KIT, GAK, ZAK, etc. tandis que celles ayant un *gatekeeper* plus encombrant (isoleucine, méthionine, glutamine, leucine, phénylalanine, etc.) le sont beaucoup moins (activité résiduelle supérieure à 60%) : KIT et la mutation du *gatekeeper* T670I, MAPKAPK5, ERK1, GSK3b, CDKL2, etc. Cependant, pour quelques kinases telles que MAP2K4, MNK1, l'encombrement a priori généré par le résidu *gatekeeper* n'empêche pas l'inhibition de la kinase. En analysant les structures de ses diverses kinases, une explication possible serait que le résidu *gatekeeper*, selon son environnement (constitué de deux résidus, l'un situé sur la C-hélice et l'autre situé au début du brin  $\beta 4$ ) et sa flexibilité, pourrait se repositionner dans le site actif afin d'accommoder le ligand. Une phénylalanine comme *gatekeeper*, bien qu'encombrante, ne semble pas être discriminante pour l'inhibition (MNKs, FLT3), sauf lorsque la région charnière est courte (CDKLs, ERK7), soit cinq résidus quand la majorité des kinases en présentent six (Tableau 11). Cette observation rejoint celle faite pour la kinase CDK2 (voir paragraphe 3.6 « Article en cours de soumission » à la page 92), dont la combinaison entre la région charnière courte et le *gatekeeper* phénylalanine semble expliquer la faible envergure de ses inhibiteurs (5.5Å), une caractéristique que ne présente pas l'hypothémicine (8.7Å).

Afin de trouver des molécules plus spécifiques que l'hypothémicine et la molécule **1afh**, j'ai imaginé transformer l'hypothémicine, qui est un ligand de type I, en un ligand de type I  $\frac{1}{2}$ , à l'image de plusieurs inhibiteurs sélectifs se trouvant sur le marché (voir Tableau 2 à la page 37). Autrement dit, j'ai voulu prendre avantage de la variabilité du résidu *gatekeeper*, en termes d'encombrement et de flexibilité, parmi les 48 cibles kinases pour concevoir des inhibiteurs pouvant atteindre la poche additionnelle hydrophobe à laquelle seuls des *gatekeepers* peu encombrants donnent accès. Pour exemple, ce résidu est respectivement une valine, une thréonine, une leucine, une phénylalanine et une méthionine chez les cinq cibles thérapeutiques KDR, KIT, GSK3b, MNK1 et MAP2K1, soit des résidus présentant des propriétés physico-chimiques bien distinctes (Tableau 11). Deux positions sur le macrocycle de l'hypothémicine, 15 et 16, sont proches du *gatekeeper* et de la poche hydrophobe à laquelle ce résidu peut donner accès. L'ajout d'un substituant sur la position 15 n'a pas été exploré ici du fait qu'il serait orienté vers les résidus catalytiques Lys $\beta 3$  et Glu<sub>C-hélice</sub>, hautement conservés parmi l'ensemble des kinases humaines, plutôt qu'en direction du résidu *gatekeeper*. De plus, un analogue de l'hypothémicine possédant un groupement hydroxyle en



position 15 n'a montré aucune inhibition vis-à-vis des principales cibles thérapeutiques (Jogireddy et al., 2009). Afin d'établir une liste de substituants possibles en position 16 et qui soient capables de pointer vers le *gatekeeper*, j'ai modélisé le positionnement de l'hypothemycine dans les structures cristallographiques en conformation *Active* de 6 des 48 kinases ayant une cystéine avant le motif DFG, soit un représentant par résidu *gatekeeper* possible (voir Tableau 11 à la page 151). Cela m'a permis de procéder à la conception rationnelle de quatre analogues de l'hypothemycine présentant un substituant en position 16 plus ou moins hydrophobe et plus ou moins encombrant pour ne cibler que les kinases ayant le résidu *gatekeeper* compatible et la poche additionnelle accessible (Tableau 12).

Afin d'appuyer mes hypothèses, j'ai procédé au *docking* avec le programme *Glide* (voir paragraphe 2.2.3 « Méthodes d'amarrage (*docking*) et évaluation de l'affinité de l'inhibiteur (*scoring*) » à la page 76), d'abord en mode rigide puis flexible, des quatre analogues sur chacune des six kinases (Tableau 12). Le *docking* flexible s'est révélé nécessaire pour prendre en considération les divers rotamères possibles du résidu *gatekeeper*, afin d'augmenter la taille du site actif mais également favoriser la création de nouvelles interactions avec le ligand. Par ailleurs, le *docking* flexible, qui ne permet pas d'intégrer la liaison covalente entre le ligand et la protéine, à l'avantage de proposer divers rotamères de la cystéine située avant le motif DFG qui ne gênent pas le positionnement du macrocycle dans la poche. Les poses et les scores obtenus, en comparaison de ceux obtenus pour l'hypothemycine, confortent l'hypothèse que plus le substituant en position 16 sur le ligand est encombrant, plus les kinases présentant un résidu *gatekeeper* lui-même encombrant ne sont pas inhibées (Tableau 12). Il convient de remarquer que plusieurs des analogues proposés, y compris l'hypothemycine, sont susceptibles d'inhiber la kinase GSK3b (score de *docking* autour de -8.9) alors qu'il a été démontré expérimentalement que l'analogue **1afh** n'inhibe pas GSK3b (100% d'activité résiduelle après incubation avec 1 $\mu$ M de **1afh** ; voir Tableau 11 à la page 151). Cependant, ces résultats de *docking* vont dans le sens d'autres analyses expérimentales extraites de la littérature, qui indiquent une activité résiduelle de 40% après incubation de GSK3b avec 2 $\mu$ M d'hypothemycine (Schirmer et al., 2006). Une hypothèse qui n'a pas été validée par le *docking* flexible est la possibilité que l'analogue **2** proposé effectue une interaction favorable  $\pi$ - $\pi$  entre son groupe alcyne et le cycle aromatique du *gatekeeper* phénylalanine de MNK2 (Tableau 12).

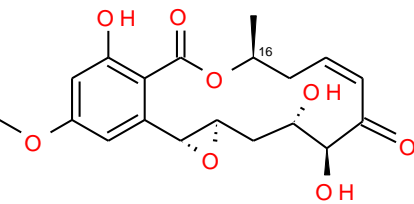
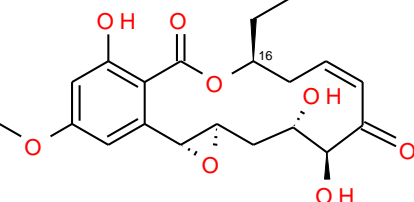
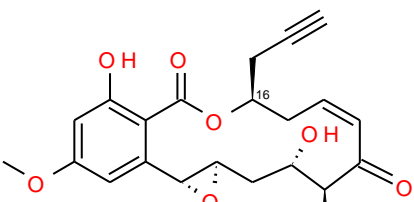
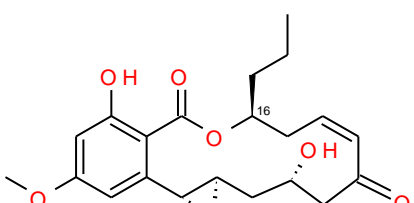
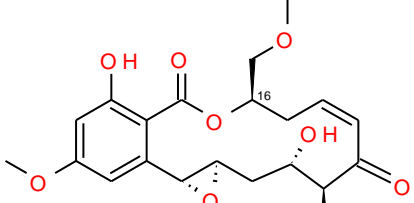
Hypothemycine et Propositions d'analogues sélectifs	Propositions	Sous-groupe de kinases ciblées	Score de <i>docking</i> (kcal/mol)					
			KDR	KIT	GSK3b	TAK1	ERK2	MNK2
			V	T	L	M	Q	F
	hypothemycine	Selon plusieurs publications, une majorité des 48 kinases possédant la cystéine avant le motif DFG sont inhibées	-8.3	-9.6	-8.9	-7.6	-7.0	-6.7
	1	Doit favoriser l'inhibition des kinases ayant un résidu <i>gatekeeper</i> hydrophobe petit ou flexible : VEGFRs, TAK1, STE7s, etc.  Peut défavoriser l'inhibition des kinases avec un résidu <i>gatekeeper</i> encombrant et rigide : MNKs, CDKLs, ERKs, MAPKAPK5, AAK1, BIKE	-8.9	-9.8	-9.1	-	-	-6.0
	2	Doit favoriser l'inhibition des kinases ayant un petit résidu <i>gatekeeper</i> : VEGFRs. Pourrait favoriser l'inhibition de FLT3, MNKs et CDKLs (interaction $\pi$ - $\pi$ avec le <i>gatekeeper</i> phenylalanine)  Doit défavoriser l'inhibition des kinases ayant un résidu <i>gatekeeper</i> encombrant même si flexible: MNKs, STE7s, CDKLs, ERKs, TAK1, MEKs, etc.	-7.9	-9.2	-8.4	-	-	-
	3	Doit favoriser l'inhibition des kinases ayant un petit <i>gatekeeper</i> hydrophobe : VEGFRs.  Doit défavoriser l'inhibition des kinases ayant un résidu <i>gatekeeper</i> encombrant même si flexible: MNKs, STE7s, CDKLs, ERKs, TAK1, MEKs, etc.	-8.1	-10.6	-7.7	-	-	-
	4	Doit favoriser l'inhibition des kinases ayant un <i>gatekeeper</i> thréonine : KIT, PDGFR $\alpha$ et $\beta$ , GAK, etc.  Doit défavoriser l'inhibition des kinases présentant un résidu <i>gatekeeper</i> encombrant et/ou hydrophobe, même si flexible : MNKs, STE7s, GSKs, PKDs, CDKLs, KDR, etc.	-7.2	-10.7	-	-	-	-

Tableau 12 : Propositions d'analogues de l'hypothemycine spécifiques de divers sous-groupes de kinases humaines possédant une cystéine avant le motif DFG et une variabilité du résidu *gatekeeper*. Les scores de *docking* des quatre propositions mesurés sur six kinases, en parallèle des résultats obtenus pour l'hypothemycine, sont également présentés. Plus le score est négatif, plus la molécule est prédite comme étant affine pour sa cible. « - » signifie qu'aucune pose satisfaisante n'a été obtenue.

Parmi les quatre propositions d'analogues, la proposition **4** offre le meilleur profil de sélectivité parmi les six cibles utilisées dans le *docking* flexible (Tableau 12) et semble être spécifique de la cible thérapeutique KIT. En effet, elle obtient un très bon score sur KIT (-10.7 contre -9.6 pour l'hypothémicine), un score un peu moins bon sur KDR (-7.2 contre -8.3 pour l'hypothémicine), et aucune pose correcte n'est obtenue sur les autres cibles. Ainsi, grâce à l'encombrement stérique apporté par le groupe méthoxyméthyle substitué en position 16, la proposition **4** devrait permettre de réduire l'inhibition des kinases ayant un résidu *gatekeeper* encombrant et rigide, comme les kinases d'intérêt thérapeutique GSK3b, MAP2K1 et MNK1. D'autre part, l'oxygène du groupe méthoxyméthyle substitué en position 16 est susceptible de former une liaison hydrogène avec le groupe hydroxyle du *gatekeeper* thréonine, propre à KIT (T670 dans KIT humaine), ce qui pourrait contribuer à la spécificité de ce composé par rapport aux autres kinases disposant d'un petit *gatekeeper* hydrophobe, tel que KDR, qui dispose d'une valine (Figure 57). Cette observation est appuyée par les poses et les scores de *docking* (Figure 57 ; Tableau 12).

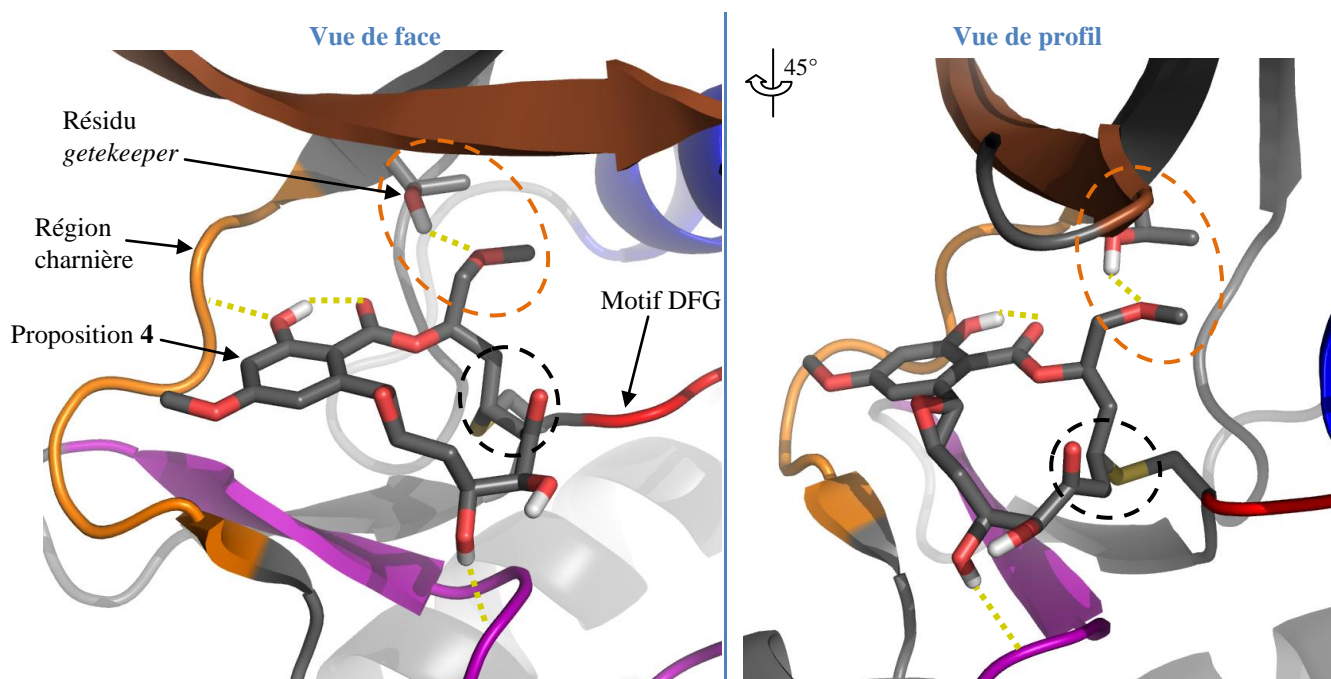


Figure 57 : Modèle 3D du mode liaison de la proposition **4** dans le site actif de KIT humaine (code PDB : 1PKG), après formation de la liaison covalente avec la cystéine située avant le motif DFG (zone entourée en pointillés noirs). Les liaisons hydrogènes sont représentées en pointillés jaunes. La zone où le substituant en position 16 sur le ligand interagit avec le *gatekeeper* T670 et entre dans la poche additionnelle hydrophobe à laquelle ce résidu donne accès est entourée en pointillés orange.

Ces molécules sont en projet de synthèse dans le laboratoire du Professeur N. Winssinger.

L'hypothemycine est capable d'inhiber une majorité des 48 kinases humaines ayant une cystéine avant le motif DFG (Schirmer et al., 2006). Le fait que toutes ces kinases appartiennent à des groupes phylogénétiques distincts (voir Tableau 11 à la page 151) rend compte du caractère aspécifique du squelette de l'hypothemycine. Dans le paragraphe suivant, je décris comment nous avons pris avantage de cette aspécificité pour « réadresser » le squelette de l'hypothemycine (c'est-à-dire pour changer ses cibles) de telle façon qu'il réagisse de manière covalente avec une nouvelle cible kinase humaine, EGFR, via la cystéine 797 située à la fin de la région charnière (sphère orange dans la Figure 50A à la page 128) et non plus avec les kinases ayant une cystéine avant le motif DFG (sphère magenta dans la Figure 50A à la page 128).

#### 4.5.2 Réadressage de l'hypothemycine en vue de l'inhibition covalente d'EGFR via la cystéine située à la fin de la région charnière

La tyrosine kinase humaine EGFR est l'une des seules kinases pour lesquelles des inhibiteurs covalents, tel que l'afatinib, sont dans des phases de développement clinique (voir paragraphe 1.2.7.a « Les inhibiteurs covalents en phase clinique » à la page 47 ; voir Tableau 3 à la page 50). Cette nouvelle génération d'inhibiteurs a la capacité de rester efficace face aux résistances développées par EGFR, à la suite des traitements effectués avec la première génération d'inhibiteurs non covalents (Yun et al., 2008). Tous contiennent le squelette quinazoline ou quinoline auquel est ajouté une amide  $\alpha,\beta$ -insaturée qui réagit de manière covalente avec la cystéine 797, située à la fin de la région charnière chez EGFR (sphère orange dans la Figure 50A à la page 128) et commune à dix autres kinases.

Compte-tenu de l'aspécificité de l'hypothemycine, j'ai voulu réadresser son squelette pour qu'il réagisse avec la cystéine située à la fin de la région charnière de la cible thérapeutique EGFR (sphère orange dans la Figure 50A à la page 128). En effet, les superpositions des kinases ERK2 et EGFR, dans les complexes ERK2/hypothemycine (code PDB : 3C9W) (Rastelli et al., 2008) et EGFR/analogue de l'afatinib (2J5E) (Blair et al., 2007), montrent un très bon recouvrement des deux inhibiteurs, qui présentent un mode de positionnement similaire dans le site catalytique (Figure 58A). Les oxygènes Ox, Oy et Oz de l'hypothemycine (Figure 58B) sont équivalents en position aux azotes Nx, Ny et Nz de l'analogue de l'afatinib (Figure 58C) (distances inférieures à 0.5Å pour les trois paires d'atomes équivalents). Une liaison hydrogène entre les oxygènes Ox et Oy permet d'orienter

les doublets libres de Ox vers la région charnière pour former une liaison hydrogène avec le groupement NH de la chaîne principale, comme le fait l'azote Nx de l'analogue de l'afatinib (Figure 58A). Le groupement électrophile de l'analogue de l'afatinib est une amide  $\alpha,\beta$ -insaturée, attachée au carbone C5 du groupe quinazoline (Figure 58C). Cette position correspond exactement au carbone C5 du cycle aromatique de l'hypothémicine (Figure 58B). On peut donc envisager que l'attachement d'une amide  $\alpha,\beta$ -insaturée sur le carbone C5 de l'hypothémicine, à l'image de l'afatinib, permette de lier de manière covalente la cystéine 797 située à la fin de la région charnière d'EGFR.

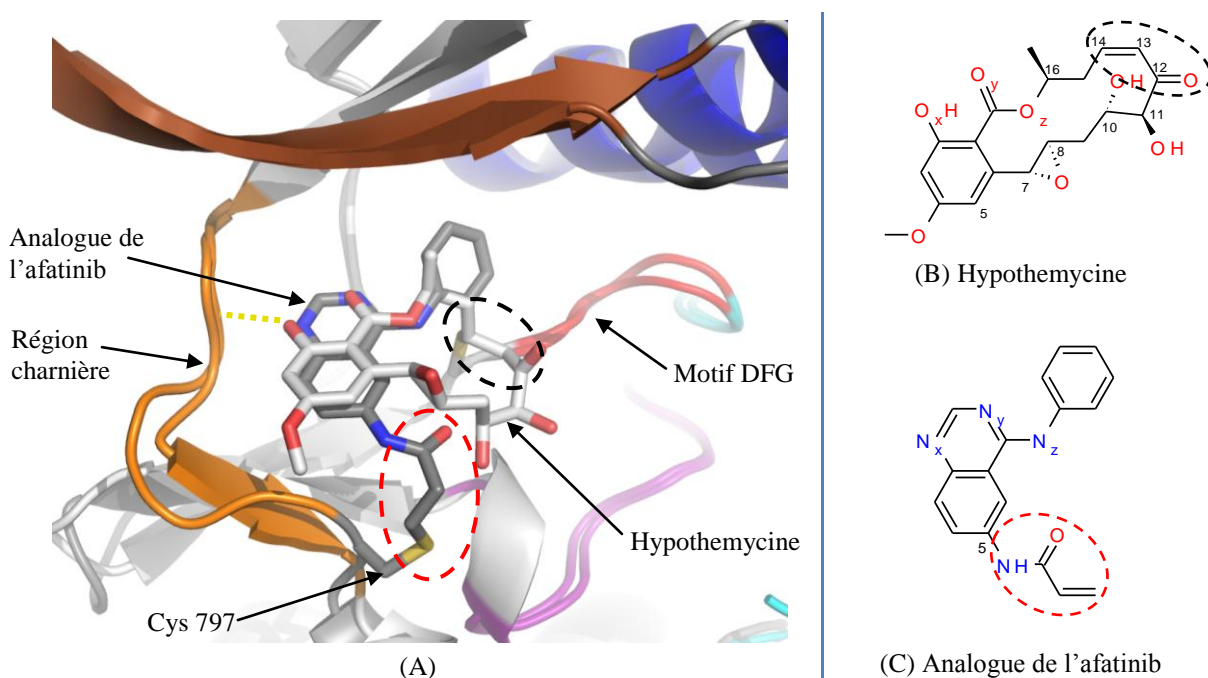


Figure 58 : (A) Superposition des domaines catalytiques des complexes ERK2/hypothémicine (ligand gris clair) et EGFR/analogue de l'afatinib (ligand gris foncé) (codes PDB : 3C9W et 2J5E). La liaison hydrogène entre chacun des deux ligands et la région charnière de leur cible respective est représentée en pointillés jaunes. La zone du lien covalent entre l'hypothémicine et ERK2 est entourée en pointillés noirs. La zone du lien covalent entre l'analogue de l'afatinib et EGFR est entourée en pointillés rouges. (B) Structure chimique de l'hypothémicine, avant qu'elle ne soit liée de manière covalente à ERK2. Le groupe électrophile est entouré en pointillés noirs. (C) Structure chimique de l'analogue de l'afatinib, avant qu'il ne soit lié de manière covalente à EGFR. Le groupe électrophile est entouré en pointillés rouges.

Toutefois, l'ajout d'un groupement en position C5 de l'hypothémicine, et donc à proximité du macrocycle, pourrait modifier la conformation globale de la molécule (Figure 58A) et l'empêcher d'entrer dans le site catalytique. Cependant, des études expérimentales rapportent le criblage du radical A (voir Tableau 4 à la page 52), un analogue de l'hypothémicine possédant un groupement méthoxy attaché au carbone C5 (numérotation selon les atomes de l'hypothémicine, Figure 58B) et une double liaison entre les carbones C7 et C8 (contre un groupement époxyde pour l'hypothémicine). Lors de ce criblage, le radical A a inhibé fortement plusieurs kinases possédant une cystéine avant le

motif DFG, indiquant que la présence d'un groupement électrophile en position 5 ne devrait pas perturber la conformation du macrocycle (Dakas et al., 2007).

Pour effectuer ce réadressage, il a été nécessaire de supprimer le groupe électrophile que possède l'hypothemycine, une cétone  $\alpha,\beta$ -insaturée (Figure 58B), afin de ne plus inhiber de manière covalente le groupe des 48 kinases possédant la cystéine située avant le motif DFG. En effet, le groupe électrophile se situe sur le macrocycle, or son remplacement par un autre groupe non réactif risque de modifier la conformation bioactive du macrocycle et la nouvelle molécule pourrait ne plus pouvoir entrer dans le site catalytique. Afin de trouver un groupe non réactif qui puisse remplacer le groupe électrophile de l'hypothemycine sans changer la conformation du macrocycle, j'ai recherché dans la littérature des analogues de l'hypothemycine ne disposant pas du groupe électrophile mais capable d'inhiber, même faiblement, des protéines kinases. Peu de ces molécules sont répertoriées dans la littérature et quasiment toutes n'ont aucune activité inhibitrice détectée (Ohori et al., 2007; Hofmann, 2008; Ninomiya-Tsuji et al., 2003). Cependant, trois molécules ont montré une faible inhibition : Aigialomycin D (Dakas et al., 2007), Zeaenol (Ninomiya-Tsuji et al., 2003) et HMY-18 (Hearn et al., 2007). Les deux premières ont la particularité de présenter une configuration *trans* de la double liaison située entre les atomes C13 et C14 (Tableau 13).

Structure chimique	Nom générique	Inhibition de kinases dans la littérature
	Aigialomycine D	IC <sub>50</sub> = 5.7 μM sur CDK1 (Dakas et al., 2007)
	Zeaenol	IC <sub>50</sub> = 69 μM sur TAK1 (Ninomiya-Tsuji et al., 2003)
	HMY-18	IC <sub>50</sub> (sur lignée cellulaire) = 32 μM (Hearn et al., 2007)

Tableau 13 : Analogues de l'hypothemycine sans groupe électrophile mais dont l'activité inhibitrice est rapportée dans la littérature. Le groupement remplaçant le groupe électrophile est entouré en pointillés noirs.

La *Cambridge Structure Database* (CSD) contient les structures cristallographiques de plus de 500 000 petites molécules. Une recherche dans la CSD m'a permis d'extraire des molécules analogues de l'hypothemycine, mais ne disposant pas du groupe électrophile, afin de trouver des molécules capables d'adopter une conformation similaire à la conformation bioactive de l'hypothemycine observée dans le site actif de la kinase ERK2 (code PDB : 3C9W). Les structures 3D de vingt petites molécules analogues ont ainsi été rapatriées et superposées à l'hypothemycine, dans le contexte du complexe cristallisé ERK2/hypothemycine, via le groupe acide benzoïque commun (Tableau 14). La structure de l'hypothemycine, disponible dans la CSD sous le nom de PEFRUN01 (molécule 9 dans le Tableau 14), a également été rapatriée (Hussain et al., 2007). Avec un *rmsd* de 0.95Å entre les macrocycles de PEFRUN01 et l'hypothemycine liée (Figure 59), il ressort que l'hypothemycine se lie à sa cible dans une conformation de basse énergie qui subit peu de réarrangements une fois la liaison covalente effectuée, si ce n'est un léger décalage (environ 1.2Å) du macrocycle en direction de la cystéine située avant le motif DFG, avec laquelle l'atome C14 est lié de manière covalente (Figure 59).

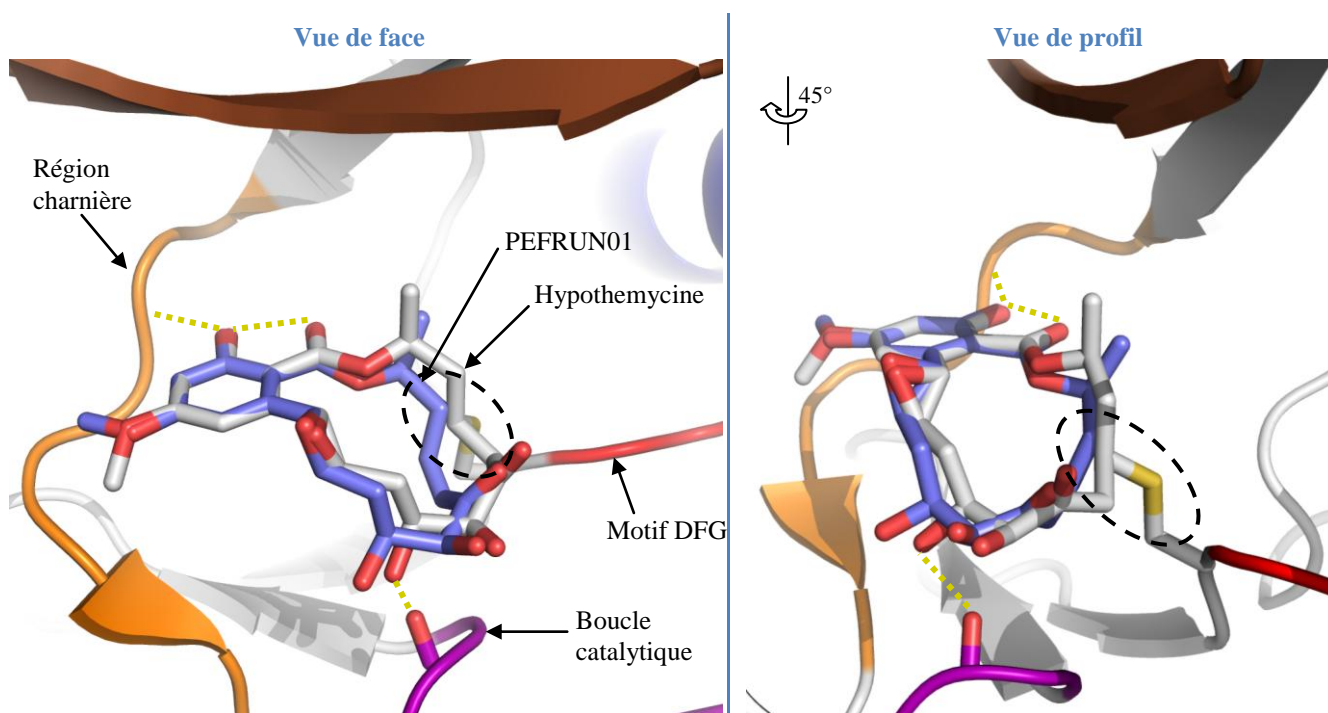
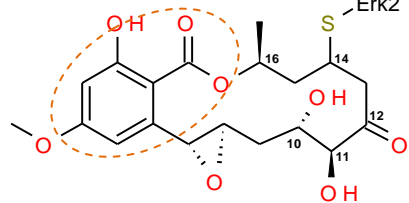
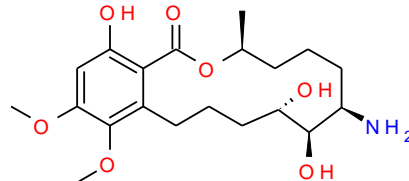
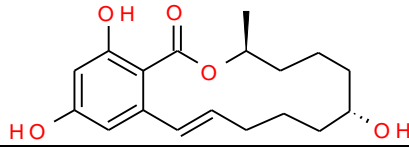
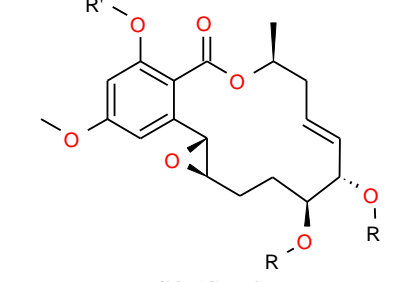
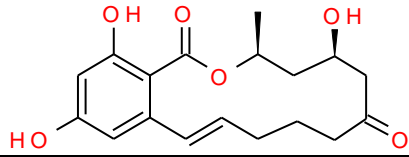
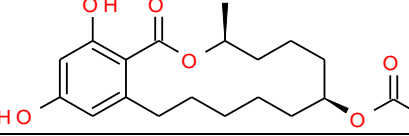
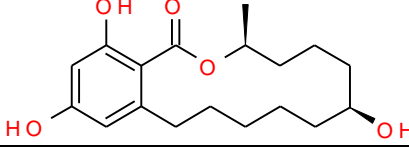
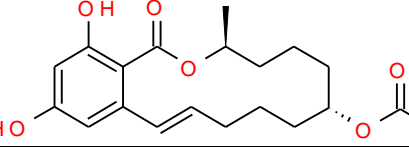
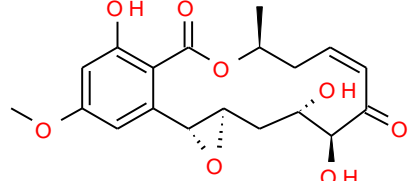
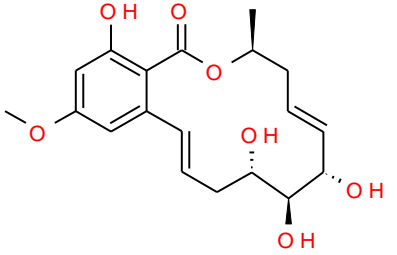
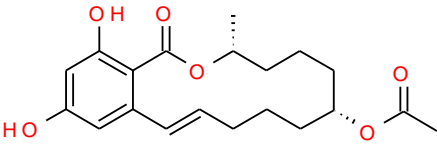
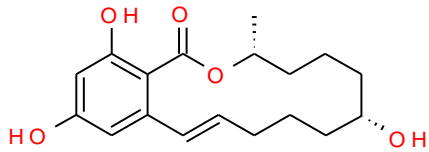
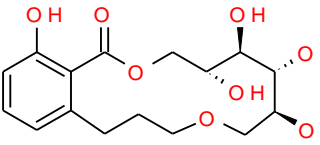
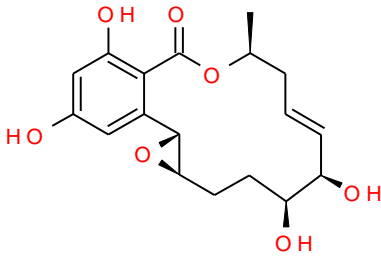
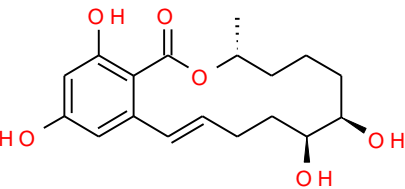
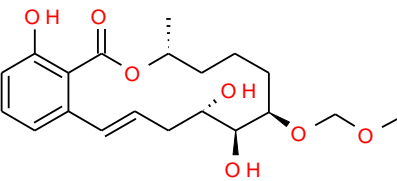
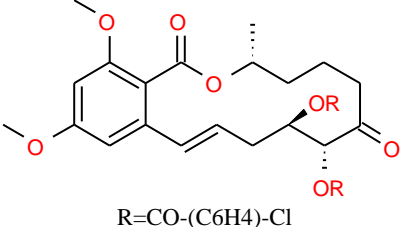
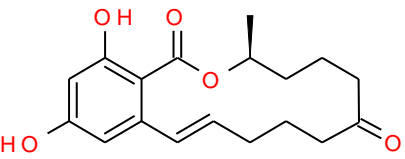


Figure 59 : Structure d'ERK2 avec l'hypothemycine (gris clair) liée de manière covalente à la cystéine située avant le motif DFG (code PDB : 3C9W). La molécule PEFRUN01 (violet), extraite de la CSD et correspondant à l'hypothemycine non liée, a été superposée à l'hypothemycine via le groupement acide benzoïque commun. Un *rmsd* de 1Å environ sépare les deux macrocycles, du fait de la liaison covalente qui rapproche le macrocycle de l'hypothemycine vers la cystéine située avant le motif DFG. Les liaisons hydrogènes sont représentées en pointillés jaunes. Notez la liaison hydrogène avec un groupe carbonyle situé sur la boucle catalytique. La zone de la liaison covalente entre l'hypothemycine et ERK2 est entourée en pointillés noirs.

Structure chimique	Code CSD Numéro	Rmsd du macrocycle par rapport à l'hypothemycine liée à ERK2 (Å)
	Hypothemycine, à l'état covalent (code PDB : 3C9W)	0 (Référence)
	CATWIE 1	0.74
	BEGDAS 2	0.74
 R=CO-(C6H4)-Br R'=CO-(C6H4)-(NO2) <sub>2</sub>	IFAPIO 3	0.79
	HXZEAR10 4	0.81
	YERVAS 5	0.86
	YERTUK YERTO 6,7	0.87
	DUHZIP 8	0.90
	PEFRUN01 (hypothemycine) 9	0.95

Suite du Tableau 14 à la page suivante



	<p>JUWCIN <b>10</b></p>	<p>1.00</p>
	<p>YERTIY <b>11</b></p>	<p>1.13</p>
	<p>TEGXEI YERTEU <b>12,13</b></p>	<p>1.25</p>
	<p>CIBWEQ <b>14</b></p>	<p>1.35</p>
	<p>SERBUN <b>15</b></p>	<p>1.36</p>
	<p>WOBFAV <b>16</b></p>	<p>1.36</p>
	<p>WEWFUA <b>17</b></p>	<p>1.41</p>
 <p>R=CO-(C6H4)-Cl</p>	<p>CBZEAL <b>18</b></p>	<p>1.48</p>
	<p>ZEARLN01 <b>19</b></p>	<p>1.51</p>

Suite du Tableau 14 à la page suivante

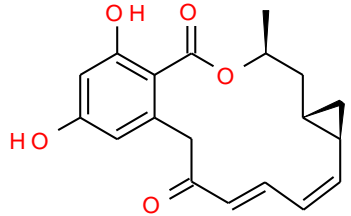
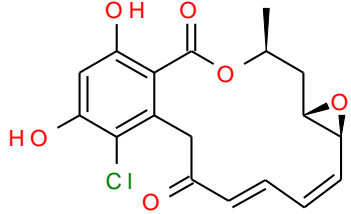
	VAQPAF 20	2.22
	ZZZGUC01 21	2.23

Tableau 14 : Liste exhaustive des molécules analogues de l'hypothemycine extraites de la *Cambridge Structure Database* et comparaison de leur conformation avec celle adoptée par l'hypothemycine dans le complexe cristallographique ERK2/hypothemycine (code PDB : 3C9W). Les molécules ont été alignées sur le groupement acide benzoïque commun (entouré en pointillés orange sur l'hypothemycine) et le calcul du *rmsd* s'est fait sur les 14 atomes communs formant le macrocycle. Les molécules sont triées par *rmsd* croissant. Plus les molécules sont substituées différemment de l'hypothemycine, plus la conformation du macrocycle s'éloigne de celle de l'hypothemycine.

La configuration absolue de l'hypothemycine est 7R, 8R, 10S, 11S, 16S (Figure 58B) (Hussain et al., 2007). La position et la configuration R ou S des carbones asymétriques sur les macrocycles des analogues de l'hypothemycine extraits de la CSD influence la conformation de ceux-ci. En général, les molécules moins substituées que l'hypothemycine (pas de OH en positions 10 ou 11, telles que BEGDAS (2), HXZEAR10 (4), YERVAS (5), YERTUK (6) et DUHZIP (8)) ou substituées aux mêmes positions avec les mêmes configurations R ou S (CATWIE (1) et PEFRUN01 (9)) présentent une conformation du macrocycle proche de celle de l'hypothemycine, soit un *rmsd* inférieur à 1Å (Tableau 14 ; Figure 60). Au contraire, des molécules avec des substituants supplémentaires par rapport à l'hypothemycine (CIBWEQ (14)) ou des configurations inversées aux positions 10, 11 ou 16 (CBZEAL (18), TEGXEY (12), WEWFUA (17)) s'éloignent de l'hypothemycine, soit un *rmsd* supérieur à 1.2Å (Tableau 14 ; Figure 60). Ces observations vont dans le sens des analyses conformationnelles publiées sur le radicicol (un autre analogue de l'hypothemycine) et des analogues ayant des configurations inversées par rapport au radicicol (Moulin et al., 2005).

Le groupement électrophile cétone  $\alpha,\beta$ -insaturée est absent dans toutes les molécules, sauf dans PEFRUN01 (9), et est majoritairement remplacé par une cétone (HXZER10 (4)), un alcool (BEGDAS (2)) ou un *trans*-énol (JUWCIN (10)). Dans le cas de l'alcool et de l'énol, la configuration R ou S du carbone chiral C12 ne semble pas discriminante pour la conformation globale du macrocycle (CATWIE (1), BEGDAS (2), JUWCIN(10)) (Tableau 14). Par contre, la chiralité de cette position influe sur l'orientation des groupes OH en positions 10 et 11 (JUWCIN (10) comparé à CATWIE (1), Figure 60). Or, d'après le complexe

ERK2/hypothémicine, le groupe OH en position 10, forme une liaison hydrogène avec un groupe carbonyle de la chaîne principale, situé à la fin de la boucle catalytique (voir Figure 59 à la page 159). Il paraît donc important de conserver l'orientation de ce groupe OH, ce qui est le cas lorsque le carbone C12 est en configuration S (JUWCIN (**10**); Figure 60). Il a d'ailleurs été montré que lorsque cet OH est remplacé par un groupement méthoxy, l' $IC_{50}$  est multipliée par 100 et passe de 80nM à 8 $\mu$ M (Hearn et al., 2007). Par contre, une méthylation du OH en position 11 est moins discriminante ( $IC_{50}$  multipliée par 5) (Hearn et al., 2007). Cependant, dans le cas de la molécule CATWIE (**1**) où le carbone C12 est en configuration R, le groupe OH en position 11 pourrait remplacer le groupe OH en position 10 pour former la liaison hydrogène avec le groupe carbonyle de la chaîne principale (Figure 60). Autrement dit, les deux configurations R ou S du carbone C12 devraient permettre de maintenir une liaison hydrogène entre le macrocycle et la fin de la boucle catalytique.

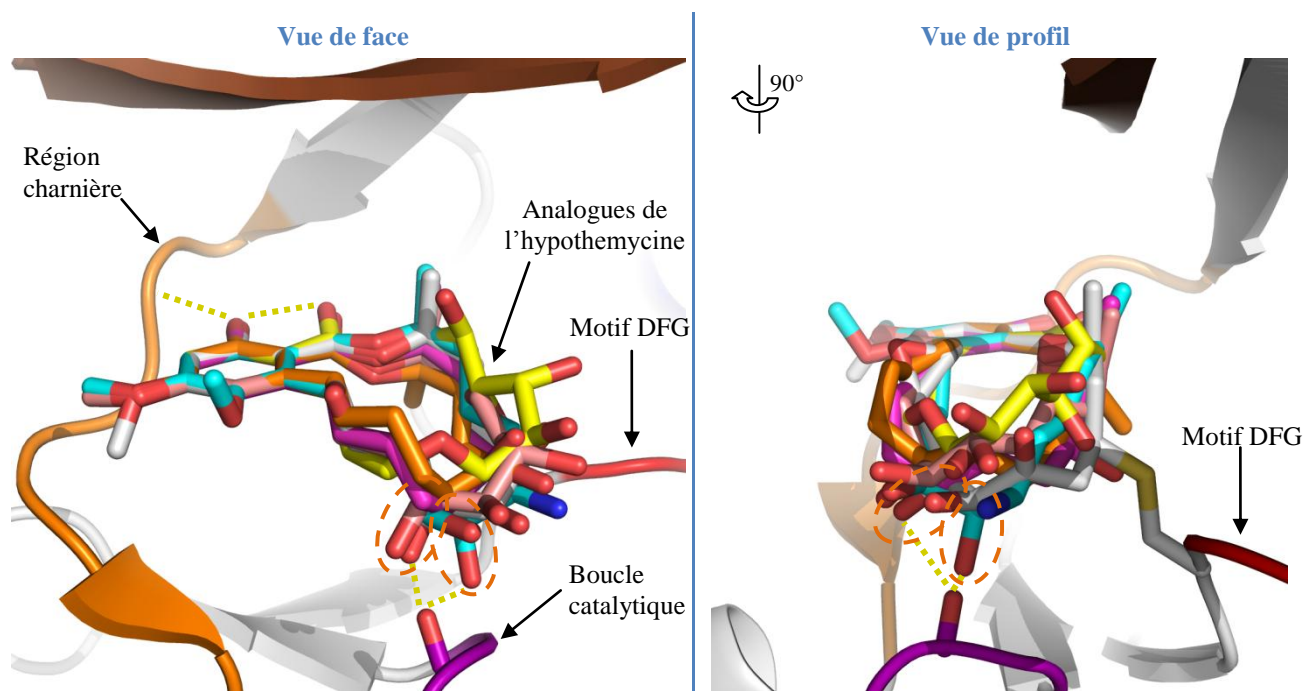


Figure 60 : Structure d'ERK2 avec l'hypothémicine (gris clair) liée de manière covalente à la cystéine située avant le motif DFG (code PDB : 3C9W). Les molécules HXZEAR10 (**4**) (rose foncé), CIBWEQ (**14**) (jaune), TEGXEI (**12**) (orange), JUWCIN (**10**) (rose clair) et CATWIE (**1**) (cyan) ont été superposées à l'hypothémicine via le groupement acide benzoïque commun. Les liaisons hydrogènes sont représentées en pointillés jaunes. Le groupe OH en position 10 de JUWCIN (**10**) et le groupe OH en position 11 de CATWIE (**1**) (entourés en pointillés orange) sont tous les deux susceptibles d'effectuer une liaison hydrogène avec un groupe carbonyle de la chaîne principale, sur la boucle catalytique.

Compte-tenu des observations issues de la littérature et de la CSD, j'ai pu proposer quatre macrocycles différents où le groupe électrophile initial a été remplacé (Tableau 15). J'ai ajouté le nouveau groupe électrophile, une amide  $\alpha,\beta$ -insaturée, sur la position C5 de chacune des nouvelles molécules proposées, afin de compléter le réadressage de ces molécules pour qu'elles inhibent irréversiblement EGFR. Il convient de noter que toutes les

molécules proposées présentent une double liaison entre les carbones 7 et 8. En effet, bien que l'hypothemycine et ses analogues LL-Z1640-2 et LL-783277 (voir Tableau 4 à la page 52), ayant respectivement un groupe époxyde, une double liaison et une simple liaison entre les carbones 7 et 8, sont tous les trois des inhibiteurs covalents, les deux premiers inhibiteurs présentent des affinités beaucoup plus élevés pour leurs cibles comparé à la molécule LL-783277 (Jogireddy et al., 2009; Ohori et al., 2007). De plus, la double liaison a été favorisée dans les propositions car la présence de l'oxygène du groupe époxyde risque de déstabiliser la position du carbonyle du groupe électrophile.

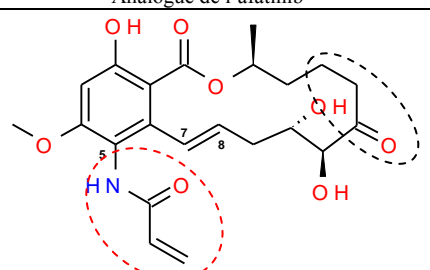
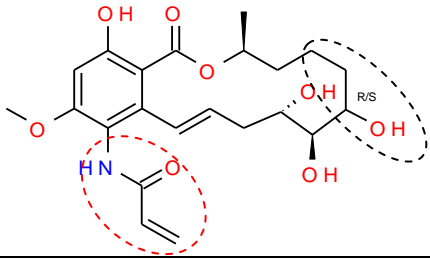
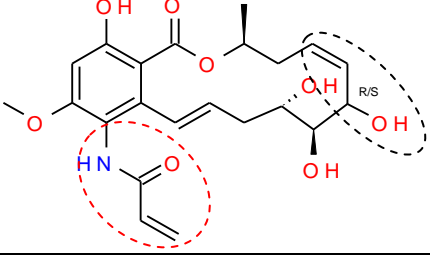
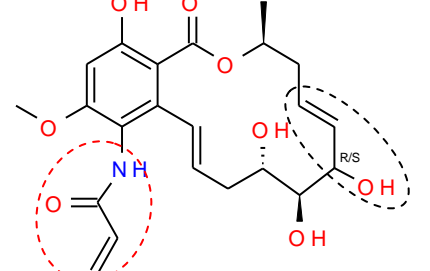
Structure chimique	Proposition	Inspirée par	Rmsd du macrocycle /hypothemycine (Å)	Score de docking (kcal/mol)
Analogue de l'afatinib		Ligand cristallographique		-8.8
	1	Afatinib Hypothemycine Radicicol A HXZEAR10	0.98	-7.8
	2	Afatinib Hypothemycine Radicicol A CATWIE BEGDAS YERVAS YERTUK DUHZIP	R : 0.97 S : 0.98	R : -7.7 S : -7.7
	3	Afatinib Hypothemycine Radicicol A HMY-18	R : 0.91 S : 0.87	R : -7.9 S : -7.6
	4	Afatinib Hypothemycine Radicicol A Aigialomycine D Zeaenol JUWCIN IFAPIO	R : 1.19 S : 1.17	R : -7.2 S : -7.0

Tableau 15 : Les quatre propositions de macrocycles dérivés de l'hypothemycine et susceptibles d'entrer dans le site actif de la kinase EGFR pour ensuite réagir de manière covalente avec la cystéine 797, située à la fin de la région charnière. Les groupements non réactifs qui remplacent le groupe électrophile situé sur le macrocycle de l'hypothemycine sont entourés en pointillés noirs. Le groupe électrophile ajouté en position 5 pour réagir avec la cystéine située à la fin de la région charnière est entouré en pointillés rouges. Les scores de docking obtenus avec la cible EGFR et comparés au ligand cristallographique, un analogue de l'afatinib, sont présentés.

Ces molécules ont été minimisées puis amarrées dans le site actif d'EGFR, en présence d'une molécule d'eau très souvent observée dans l'environnement du résidu *gatekeeper* d'EGFR. Les poses obtenues ainsi que les scores de *docking* proches de celui calculé pour le ligand cristallographique - un analogue de l'afatinib - confortent nos propositions (Tableau 15). Afin de prédire le positionnement des molécules proposées dans EGFR, après la formation de la liaison covalente avec la cystéine 797, des essais de *docking* covalent ont été menés mais se sont révélés infructueux, avec des poses proposées systématiquement éloignées du site actif. Ceci peut s'expliquer par le fait que la cystéine, qui se trouve à la fin de la région charnière, n'est pas enfouie et se trouve à la sortie du site. J'ai donc procédé à l'attachement manuel, après le *docking*, du ligand avec la cystéine 797 d'EGFR, et procédé à une minimisation d'énergie. La pose obtenue s'éloigne peu de la pose de *docking* initiale, bien que légèrement déplacée pour approcher la cystéine 797 avec laquelle elle est liée, ce qui contribue par ailleurs à éloigner le macrocycle du résidu situé avant le motif DFG, une thréonine chez EGFR (Figure 61).

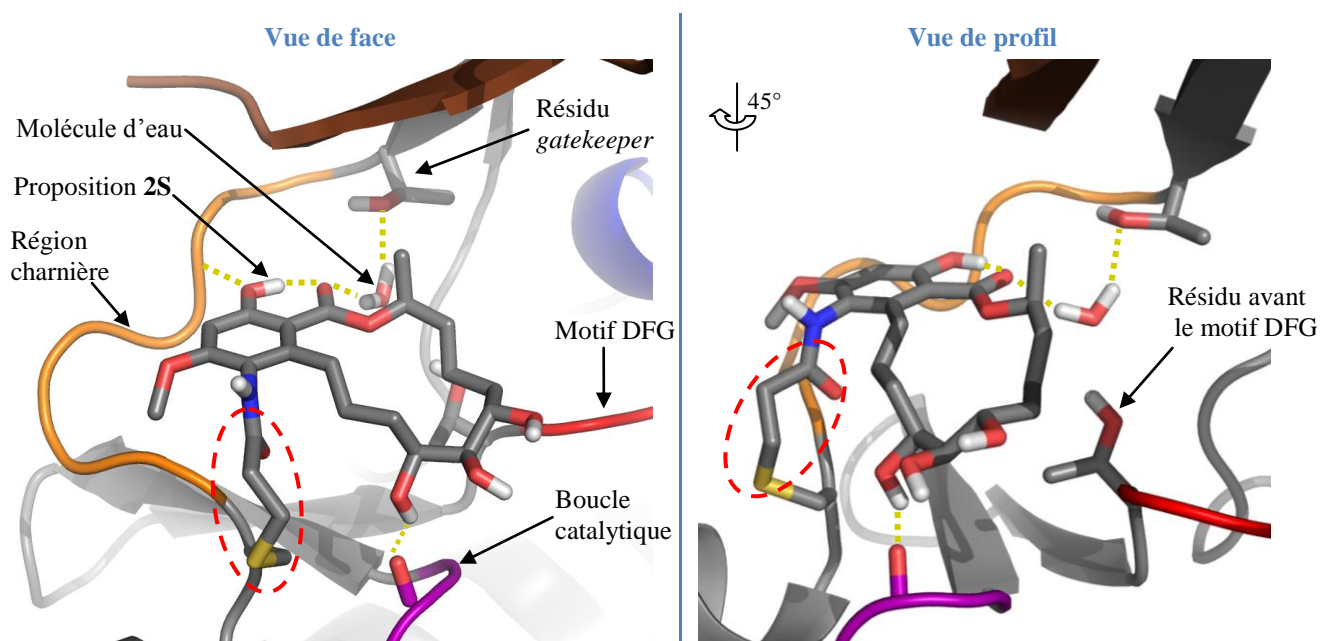


Figure 61 : Modèle 3D du mode liaison de la proposition 2S dans le site actif d'EGFR humaine (code PDB : 1M17), après formation de la liaison covalente avec la cystéine C797. La région de la liaison covalente est entourée en pointillés rouges. Le *gatekeeper* thréonine T790, situé avant la région charnière et la molécule d'eau environnante sont représentés en bâton. Les liaisons hydrogènes sont représentées en pointillés jaunes. La chaîne latérale de la thréonine T854 située avant le motif DFG (équivalente à la cystéine chez ERK2 et les cibles de l'hypothémicine) est représentée en bâton.

A l'heure actuelle, la synthèse de ces molécules n'est pas terminée car elles demandent beaucoup d'effort aux chimistes pour refermer les nouveaux macrocycles, une fois le groupe électrophile placé sur le cycle aromatique.

Cette difficulté, ainsi que les autres contraintes évoquées plus haut (élimination de l’électrophile situé sur le macrocycle, conservation de la conformation bioactive du macrocycle), ont été contournées en supprimant le macrocycle pour ne garder que le nouveau groupe électrophile et la partie de la molécule interagissant avec la région charnière, soit l’acide 2-hydroxybenzoïque (Figure 62A) ou l’analogue 2-hydroxybenzamide (Figure 62B). Il est important de noter que les atomes constituant chacun des deux squelettes se trouvent dans un même plan.

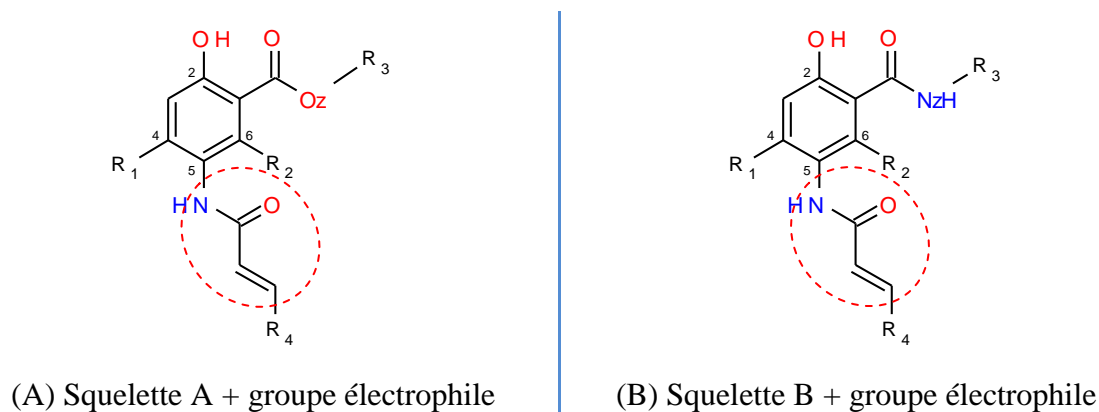


Figure 62 : Propositions de nouveaux squelettes de petites molécules dérivées de l’hypothemycine sans le macrocycle et présentant un groupe électrophile en position 5 (entouré en pointillé rouges), par analogie avec les inhibiteurs covalents d’EGFR, tels que l’afatinib. (A) Structure chimique de l’acide 2-hydroxybenzoïque (squelette A) avec le groupe électrophile qui est une amide  $\alpha,\beta$ -insaturée. (B) Structure chimique du 2-hydroxybenzamide (squelette B) avec le groupe électrophile qui est une amide  $\alpha,\beta$ -insaturée.

Les poses obtenues lors des essais de *docking* rigide avec la kinase EGFR prédisent un mode de positionnement de ces nouveaux squelettes qui est comparable à celui de l’hypothemycine dans le site catalytique de ERK2 (code PDB : 3C9W). Les scores obtenus sont cependant moins bons comparés au ligand cristallographique (de l’ordre de -6 contre -8.8 pour le ligand cristallographique), mais cette observation vient du fait que les molécules sont petites (poids moléculaire de 240 Da contre 400 Da pour le ligand cristallographique). Une première série de molécules a donc été synthétisée avec divers groupes alkyles attachés à l’oxygène Oz (ou Nz pour le groupe 2-hydroxybenzamide). Ces molécules ont ensuite été criblées sur EGFR (Tableau 16).

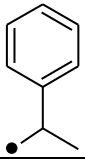
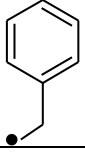
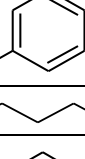
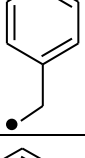
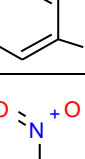
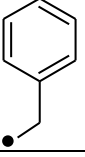
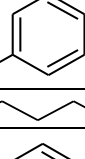
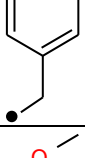
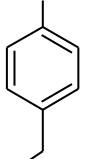
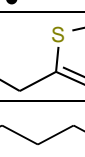
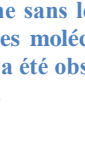

Nom	Squelette	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	IC <sub>50</sub> (μM)
1	A	OH	CH <sub>3</sub>		H	3.2
2	A	H	H		H	7.1
3	A	OH	CH <sub>3</sub>		H	22
4	A	H	H		H	33
5	A	OH	CH <sub>3</sub>		H	40
6	B	OH	CH <sub>3</sub>		H	-
7	A	OH	CH <sub>3</sub>		H	-
8	B	OH	CH <sub>3</sub>		H	-
9	A	OH	CH <sub>3</sub>		H	-
10	B	OH	CH <sub>3</sub>		H	>30
11	A	OH	CH <sub>3</sub>		H	-
12	B	OH	CH <sub>3</sub>		H	>30
13	A	H	H		CH <sub>3</sub>	-

Tableau 16 : Les treize premières molécules dérivées de l'hypothémicine sans le macrocycle et avec un groupement électrophile en position 5, synthétisées et testées *in vitro* sur EGFR. Les molécules 1 à 5 se sont révélées capables d'inhiber la kinase EGFR (IC<sub>50</sub> inférieure à 40μM). Aucune inhibition n'a été observée pour les molécules 6 à 13 (-) ou bien l'inhibition observée était insuffisante pour calculer l'IC<sub>50</sub> (>30μM).

Les premiers tests d'activité effectués sur EGFR ont été prometteurs, avec deux molécules présentant une  $IC_{50}$  inférieure à  $10\mu M$  (molécules **1** et **2**, Tableau 16) pour un poids moléculaire de l'ordre de 300 Da seulement. Cependant, aucune inhibition n'a été détectée pour des molécules a priori similaires (molécules **6** à **13**, Tableau 16). Les analyses SARs révèlent que le retrait des groupements OH et  $CH_3$ , sur les positions 4 et 6 du cycle aromatique, contribue à améliorer l' $IC_{50}$  (molécule **5** comparée à **2**, molécule **9** comparée à **4**, Tableau 16). Ces deux groupements, situés de part et d'autre du groupement électrophile attaché en position 5 du cycle aromatique, pourraient donc perturber le positionnement du carbonyle appartenant au groupe électrophile, car le OH est susceptible d'effectuer une liaison hydrogène avec le carbonyle, tandis que le  $CH_3$  génère un encombrement stérique. En conséquence, l'orientation de l'électrophile vis-à-vis de la cystéine n'est pas optimale et l'affinité est affectée, même si la liaison covalente est formée. On peut donc espérer améliorer l'affinité de la meilleure molécule de ce criblage, la molécule **1** (Tableau 16), si ces deux groupements lui sont retirés. La présence de ce même groupement  $CH_3$  expliquerait également qu'aucune des quatre molécules dérivées du squelette 2-hydroxybenzamide (squelette B) ne présente une  $IC_{50}$  mesurable (molécules **6**, **8**, **10** et **12**, Tableau 16). En effet, la proximité du  $CH_3$  avec le groupe  $NzH$  n'est pas favorable du point de vue des interactions électrostatiques et il est probable que la planarité du squelette benzamide soit perturbée, empêchant l'entrée de la molécule dans le site actif. Cette observation expliquerait le différentiel observé entre les dérivés d'acide hydroxybenzoïque et les dérivés d'hydroxybenzamide équivalents (molécule **3** comparée à **8**, molécule **5** comparée à **10**, Tableau 16). La seule molécule dont le groupe électrophile amide  $\alpha,\beta$ -insaturée possède un  $CH_3$  sur la position  $\beta$  (molécule **13** comparée à **4**) n'a aucun effet d'inhibition détecté, révélant une gêne stérique qui empêche la formation de la liaison covalente. Enfin, les résultats de criblage montrent que l'encombrement hydrophobe attaché à l'oxygène  $Oz$  des molécules dérivées de l'acide 2-hydroxybenzoïque est bénéfique pour l'inhibition (molécule **1** comparée aux molécules **3**, **4** et **5** ; Tableau 16).

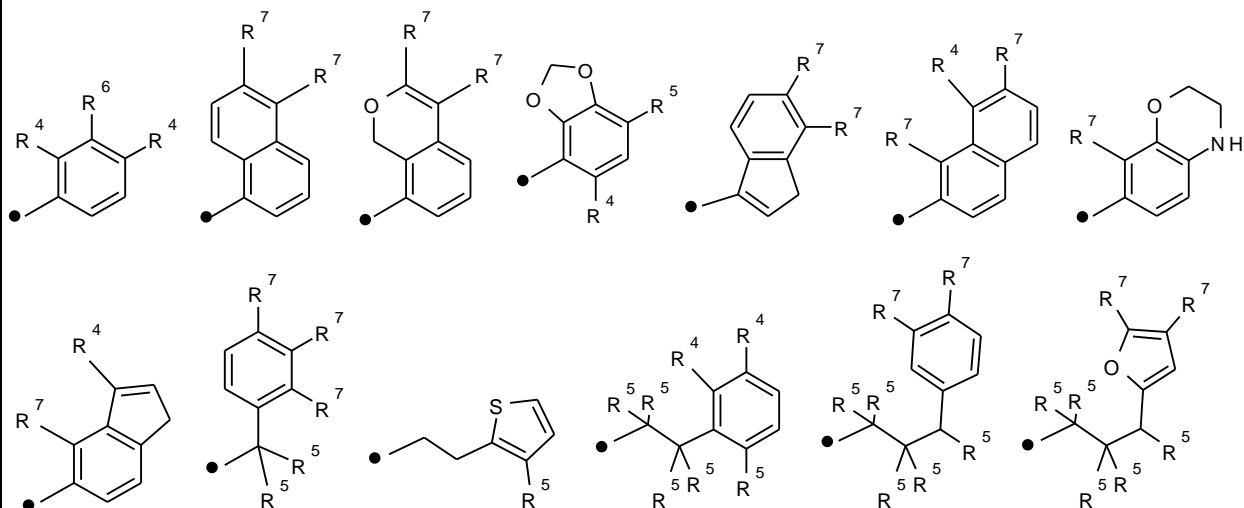
Ces diverses observations m'ont permis de proposer aux chimistes des molécules plus élaborées au niveau des substituants sur l'oxygène  $Oz$ , susceptibles d'occuper plus largement le site catalytique en allant jusqu'à s'étendre dans la petite poche hydrophobe à laquelle le *gatekeeper* thréonine d'EGFR donne accès (Figure 63). Pour cela, je me suis inspirée des inhibiteurs de type I  $\frac{1}{2}$  d'EGFR, covalents ou non, rapportés dans la littérature et/ou cocrystallisés avec EGFR (voir Tableau 2 et Tableau 3 aux pages 37 et 50 ; *Aureus Pharma*).



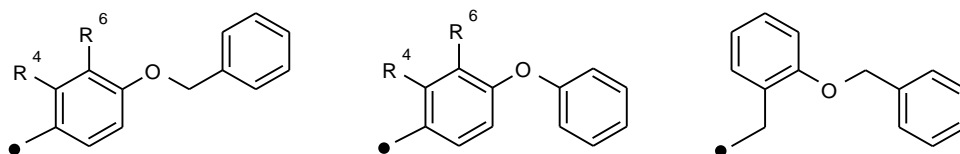
$R_1 = H$

$R_2 = H$

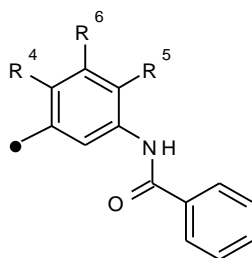
$R_3 =$  groupements destinés à inhiber EGFR dans sa conformation *Active* =



$R_3 =$  groupements destinés à inhiber EGFR dans sa conformation inactive *C-helix out* =



$R_3 =$  groupement destiné à inhiber EGFR dans sa conformation inactive *DFG out*, en supposant qu'elle soit capable d'adopter cette conformation =



$R^4 = H$  ou F ou Cl ou Br ou  $CH_3$

$R^5 = H$  ou  $CH_3$

$R^6 = R^4$  ou  $CH_2-CH_3$  ou  $CH_2=CH_2$

$R^7 = H$  ou F

Figure 63 : Amplification de la série précédente (squelettes A et B dans la Figure 62) et propositions d'inhibiteurs covalents d'EGFR, dérivés de l'hypothémicine sans le macrocycle et inspirés des inhibiteurs d'EGFR connus.

Remarque : certaines de ces molécules sont inspirées de la molécule commercialisée lapatinib (voir Tableau 2 à la page 37) et pourraient ainsi permettre d'inhiber EGFR dans sa conformation inactive *C-helix out*, plutôt qu'*Active* comme le font tous les inhibiteurs

précédemment proposés. Cette propriété devrait leur conférer un profil de sélectivité encore meilleur comparé aux inhibiteurs de type I, visant la conformation *Active* (voir Figure 16 à la page 39).

Soixante composés ont ainsi été synthétisés dans le laboratoire du Professeur N. Winssinger et les  $IC_{50}$  déduites des pourcentages d'inhibition sur EGFR viennent d'être calculées. Une vingtaine de ces composés présente une  $IC_{50}$  inférieure à  $1\mu M$  (voir en annexe le Tableau 17 à la page 193) dont sept ont une  $IC_{50}$  inférieure à 100nM, parmi lesquels le composé **651** (Figure 64). L'ajout de groupements hydrophobes et encombrants sur l'oxygène  $O_z$  a été bénéfique, ainsi que le retrait des groupements OH et CH<sub>3</sub> aux positions 4 et 6 puisque parmi les meilleurs composés apparaissent des dérivés du squelette benzamide (voir en annexe le Tableau 17 à la page 193).

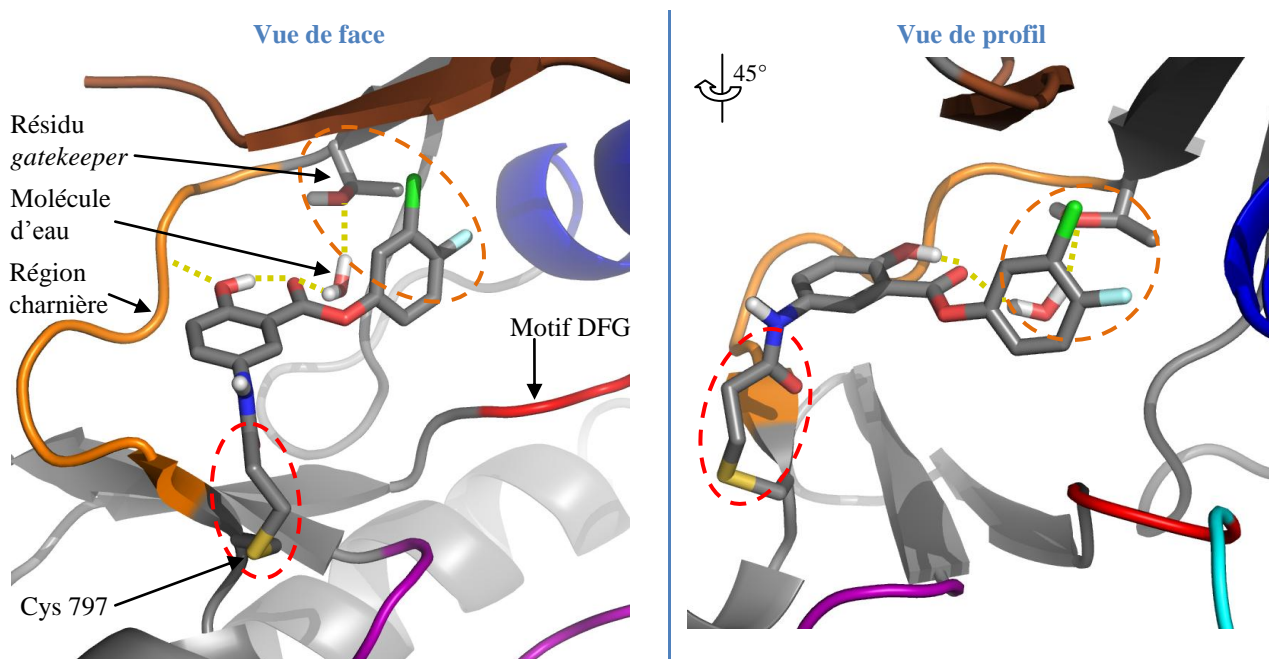


Figure 64 : Modèle 3D du mode liaison dans le site actif d'EGFR humaine (code PDB : 1M17) d'une des meilleures molécules issues de la seconde campagne de criblage sur EGFR, après formation de la liaison covalente avec la cystéine C797. La région de la liaison covalente est entourée en pointillés rouges. Le *gatekeeper* thréonine T790, situé avant la région charnière et la molécule d'eau environnante sont représentés en bâton. Les liaisons hydrogènes sont représentées en pointillés jaunes. La région où la petite molécule interagit avec la poche hydrophobe proche du résidu *gatekeeper* est entourée en pointillés orange.

Les prochaines étapes débiteront par une mesure plus précise de l' $IC_{50}$  pour les meilleures molécules et la validation de la formation de la liaison covalente par spectrométrie de masse. Ensuite, l'obtention d'un profil de sélectivité des meilleurs composés sur l'ensemble des kinases humaines rendra compte de la nécessité d'optimiser leur profil et de

proposer de nouvelles molécules. Enfin, la recherche de lignées cellulaires de cancer du poumon du type NSCLC présentant une activité anormale de la kinase EGFR, si possible avec et sans la mutation du résidu *gatekeeper* T790M (Yun et al., 2008), permettra d'étudier l'efficacité des meilleurs composés. Cette mutation est à l'origine de l'inefficacité des traitements actuels avec les inhibiteurs non covalents erlotinib et gefitinib (voir Tableau 2 à la page 37) et les inhibiteurs covalents pourraient être le seul traitement efficace pour les patients (voir paragraphe 1.2.7.c « Avantages des inhibiteurs covalents » à la page 55).

## 4.6 Conclusion

Ces résultats démontrent que l'utilisation des inhibiteurs covalents de kinases dans des maladies graves comme les cancers est une stratégie qui peut être très efficace. D'une part, la liaison covalente peut leur permettre de faire face à des concentrations élevées d'ATP, puisqu'il leur suffit d'entrer une seule fois dans le site catalytique pour qu'ils y restent attachés de manière covalente. D'autre part, la liaison covalente permet d'avoir un temps de résidence élevé au niveau de la cible, ce qui est directement corrélé à l'efficacité de ce type d'inhibiteur. En effet, même après clairance apparente de l'inhibiteur dans le sang, la cible reste inhibée en continu tant qu'elle n'a pas été dégradée puis resynthétisée par la cellule. Le profil de sélectivité des inhibiteurs covalents doit être excellent pour éviter leur liaison à des cibles secondaires et l'apparition de toxicités. Il est donc préférable de s'inspirer d'un inhibiteur non covalent dont le squelette chimique lui confère déjà une bonne sélectivité, pour ensuite transformer cet inhibiteur en inhibiteur covalent. C'est pourquoi le premier inhibiteur qui a été conçu lors de ce travail de thèse est dérivé d'un inhibiteur allostérique. En effet, les inhibiteurs allostériques sont ceux qui offrent les meilleurs profils de sélectivité du fait qu'ils ciblent une conformation inactive qui n'est peut-être pas adoptée par l'ensemble des kinases humaines, et que cette conformation donne accès à un site catalytique étendu dont les acides aminés sont moins conservés parmi l'ensemble des kinases. Nous avons également montré qu'un inhibiteur allostérique covalent peut se révéler encore plus sélectif que l'inhibiteur allostérique dont il est dérivé dès lors que la cystéine visée est peu conservée parmi les cibles de l'inhibiteur allostérique.



## 5 Kinatown : une Interface Web Dédiée aux Kinases et à la Conception d'Inhibiteurs

---

5.1 Introduction.....	174
5.2 Article en cours de soumission.....	176
5.3 Perspectives.....	177

## 5.1 Introduction

J’ai développé une interface web nommée « Kinatown », afin de rendre disponibles les résultats des travaux décrits dans les chapitres précédents, soit l’assignation de la conformation d’une structure 3D de kinase et la présence de cystéines dans le site actif. Cette interface, accessible à l’adresse <http://lbg.igbmc.fr/Kinatown>, permet aussi bien de visualiser une seule structure 3D donnée que l’ensemble des structures associées à un identifiant AC de kinase donné (Figure 65).

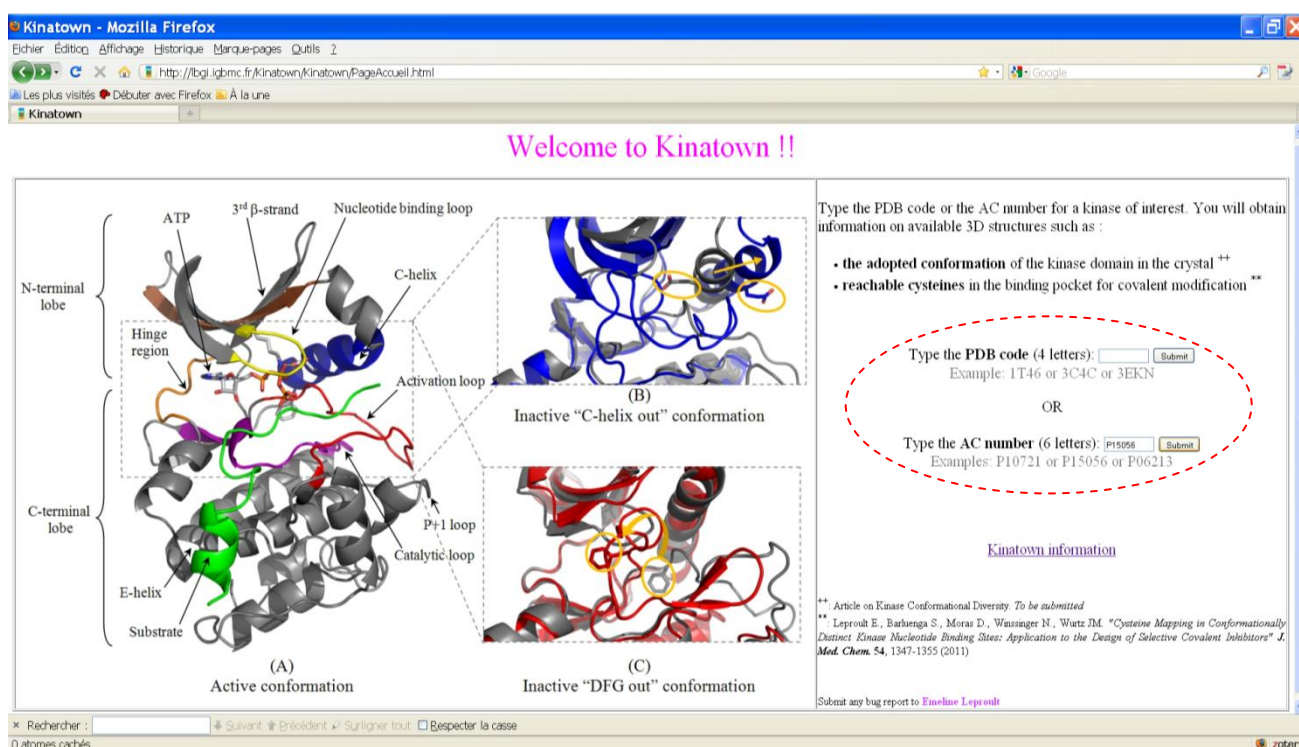


Figure 65 : Page d’accueil de l’interface web Kinatown. L’utilisateur a la possibilité d’entrer le code PDB d’une structure de kinase ou bien de mettre l’identifiant AC d’une kinase (zone entourée en pointillés rouges). Dans ce dernier cas, toutes les structures associées à cet identifiant seront affichées dans des onglets distincts.

Pour chacune de des structures 3D visualisées, l’utilisateur accède à l’assignation de la conformation observée (*Active*, *C-helix out*, *DFG out*, *A-loop out* ou *C-helix and DFG out*) ainsi qu’à un descriptif plus détaillé de cette conformation, c’est-à-dire la sous-conformation (orientation particulière du résidu conservé phénylalanine appartenant au motif DFG ; distance anormalement élevée entre les deux résidus catalytiques Lys<sub>β3</sub> et Glu<sub>C-hélice</sub> ; abaissement du feuillet constituant le lobe N-terminal vers le site catalytique ; etc.) (Figure 66). L’utilisateur accède également à la liste des cystéines accessibles (s’il y en a) et à leur localisation dans le site catalytique de la structure observée (Figure 66). Afin de connaître la conservation de ces cystéines parmi les kinases humaines, la liste des autres kinases humaines partageant ces cystéines est également donnée.

The screenshot displays the Kinatown web interface for the protein BRAF (AC P15056). The interface is organized into several sections:

- PROTEIN information:**
  - Gene Name: BRAF
  - Full Name Gene: Serine/threonine-protein kinase B-raf
  - Uniprot Identifier: BRAF\_HUMAN
  - AC number: P15056
  - Species: Homo sapiens
  - Involved in: Malignant melanoma|Melanoma|Pancreatic cancer
- PDB information:**
  - Title: B-raf kinase in complex with plx4720
  - PDB Chain: 3C4C A
  - Resolution: 2.57 Angstrom
  - Mutation: ILE543ALA|ILE544SER|ILE551LYS|GLN562ARG|GLEU588ASN|LYS630SER|PHE667GLU|TYR673SER|ALA688ARG|GLEU706SER|GLN709ARG|SER713GLU|LEU716GLU|SER729GLU
  - LigandName: 324
- Chain conformation:**
  - Conformation: **Chelix-out**
  - Details: Chelix-out + PHE (DFG motif) pan-like + Beta sheet intact
  - Active reference structure:  IJKS\_A.pdb
- Chain Exploration:**
  - Catalytic amino acids:**
    - Lysine
    - Glutamic acid
    - Aspartic acid of DFG motif
  - Other important amino acids and motifs:**
    - Gatekeeper
    - Phenylalanine of DFG motif
  - Structural elements:**
    - P-loop
    - Beta strand 3
    - C-helix
    - Hinge (size: 6 amino acids)
    - DFG motif
    - Activation loop
  - Cysteine accessibility:**
    - Cys532
  - Phosphorylation sites:**
    - Ser465
    - Ser579
    - Ser614
  - ATP pocket information:**
    - ATP pocket
    - Number of residues: 32
  - Ligand:**
    - 324
  - Other PDB structures of the same kinase:**
    - 1UWH.pdb chain A (DFG-out conformation)
    - 1UWH.pdb chain B (DFG-out conformation)
    - 1UWJ.pdb chain A (DFG-out conformation)
    - 1UWJ.pdb chain B (DFG-out conformation)
    - 2FBS.pdb chain A (Active conformation)
    - 2FBS.pdb chain B (Active conformation)
    - 3C4C.pdb chain B (DFG-out conformation)
    - 3C4D.pdb chain A (Active conformation)
    - 3C4D.pdb chain B (DFG-out conformation)
    - 3D4Q.pdb chain A (Active conformation)
    - 3D4Q.pdb chain B (Active conformation)
    - 3IDP.pdb chain A (DFG-out conformation)
    - 3IDP.pdb chain B (DFG-out conformation)
    - 3IIS.pdb chain A (DFG-out conformation)
    - 3IIS.pdb chain B (DFG-out conformation)

The central part of the interface shows a 3D ribbon representation of the protein structure, with the active site residues highlighted in yellow and blue. The interface also includes navigation controls like 'spin', 'zoom', 'quality', 'trace', and 'cartoons', and a 'Send a new request' button.

Figure 66 : Page de résultats de l’interface web Kinatown obtenu pour l’identifiant AC « P15056 », correspondant à la kinase humaine BRAF. Chaque onglet, en haut de la page, correspond à une structure. La couleur de l’onglet dépend de la conformation assignée. Ici, la structure 3C4C (chaîne A) est affichée. L’onglet est bleu, ce qui correspond à la conformation *C-helix out*. Les informations liées à la conformation et la sous-conformation adoptées par cette structure sont données dans la zone entourée en pointillés bleus. L’information de présence de cystéines dans le site catalytique est disponible dans la même section que les résidus et autres éléments importants de la structure, qui peuvent être affichés au gré de l’utilisateur (zone entourée en pointillés orange). Il est possible de superposer à la structure observée les autres structures de la même kinase, qui ont pu être cristallisées dans d’autres conformations (zone entourée en pointillés verts). Des informations plus générales sur la kinase et la structure observée sont données en haut de la page web (zone entourée en pointillés violets).

Un autre intérêt de cette interface est son aspect didactique, puisque les éléments de structure secondaire (la région charnière, la C-hélice, etc.) et les résidus importants (les résidus catalytiques, le résidu *gatekeeper*, les sites de phosphorylation, etc.) peuvent être localisés sur la structure affichée, au gré de l'utilisateur (Figure 66). La forme du site catalytique (détectée par le programme *PCK*, développé dans l'équipe) peut être affichée, ainsi que le ligand (s'il y en a un). Il est également possible d'afficher la structure 3D *Active* de référence ayant permis l'assignation de la conformation après superposition, ce qui permet d'en comprendre les différences vis-à-vis de la structure 3D observée. De même, les autres structures associées à la kinase d'intérêt (si elles existent) peuvent être superposées à la structure 3D observée (Figure 66). Des informations plus générales sur la kinase, telles que son implication dans des maladies, et des informations sur la structure observée, telles que sa résolution et les mutations présentes, sont affichées au-dessus la structure 3D (Figure 66).

Toutes les informations auxquelles l'utilisateur a accès ont été préalablement calculées pour l'ensemble des structures 3D de kinases répertoriées en juin 2011. Ces informations sont contenues dans notre base de données des kinases dont le contenu est donné dans les chapitres 2, 3 et 4). Toutefois, si l'identifiant de structure donné par l'utilisateur est inconnu de la base de donnée, celui-ci est traité à la volée, ce qui demande un temps de calcul de l'ordre de trois minutes avant l'affichage des résultats.

Le contenu détaillé de l'interface web Kinatown est décrit dans l'article: « *Kinatown: a web interface for kinase structure exploration and subsequent design of allosteric and covalent inhibitors* », qui sera soumis dans le journal *Bioinformatics*.

## 5.2 Article en cours de soumission



# Kinatown: a web interface for kinase structure exploration and subsequent design of allosteric and covalent inhibitors

Emeline Leproult<sup>1,\*</sup> and Jean-Marie Wurtz<sup>1</sup>

<sup>1</sup>Département de Biologie et Génomique Structurale, Institut de Génétique et de Biologie Moléculaire et Cellulaire, 1 Rue Laurent Fries, BP 10142, 67404 ILLKIRCH CEDEX, France.

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXXX

## ABSTRACT

**Summary:** The quest to develop covalent inhibitors towards inactive kinase conformations led us to setup a pipeline that automatically classifies the 3D structures into five conformations. Additional annotations are added that highlight structural peculiarities. Residues in the ATP pocket and its extensions induced by the inactive conformations are derived and reactive cysteine residues are mapped. The Kinatown web interface provides a unique set of capabilities for exploration of protein kinase PDB structures in their active or inactive forms, giving clues to the design of a novel class of kinase inhibitors.

**Availability and Implementation:** Kinatown web interface is freely accessible at <http://lbgi.igbmc.fr/Kinatown>

**Contact:** [leprout@igbmc.fr](mailto:leprout@igbmc.fr)

## 1 INTRODUCTION

Being responsible for the phosphorylation of about one third of the proteome, the 518 human protein kinases play pivotal roles in many aspects of cell behavior such as signal transduction, metabolism, transcription and cell cycle progression (Manning et al., 2002). Among the regulation processes that apply on these proteins, one of them consists in the deformation of the kinase fold, in order to impede the substrate binding to occur. Up to now, two distinct inactive conformations have been reported in the literature: The “C-helix out” conformation (Andzelm et al., 1995) and the “DFG out” conformation (Nagar et al., 2003) (Fig. 1). Such transient inactive conformations are subject of intensive research in the pharmaceutical industries since they give access to extended and less conserved ATP binding pockets, useful for the design of selective inhibitors (J. Zhang et al., 2009). The most famous representative of this inhibitor design strategy is the drug imatinib, targeting ABL and KIT kinases in their “DFG out” conformation (Buchdunger et al., 1996).

More recently, the acceptance into phase III of BIBW 2992 (Subramaniam et Hwang, 2011), an inhibitor that owes its efficiency to the covalent reaction it does with a cysteine residue located in the ATP pocket of EGFR kinase, is attracting much attention. Indeed, the targeted cysteine residue contributes to the

good selectivity profile of this inhibitor (Karaman et al., 2008) as well as its maintained efficiency toward resistant mutations of EGFR (Yun et al., 2008). To date, only a few studies have been reported on covalent kinase inhibitors (Jogireddy et al., 2009; Zhou et al., 2010; Henise et Taunton, 2011).

Targeting a kinase in its inactive conformation and by reacting with an unconserved cysteine of its ATP binding pocket are among the best reported strategies for the design of efficient and selective kinase inhibitors. This led us to present Kinatown, a new web interface dedicated to the broad community of chemists and biologists, that allows to access all available PDB structures for a given protein kinase and gives information relative to the adopted kinase catalytic domain conformation as well as cysteine residue accessibility and conservation in the binding pocket.

## 2 METHODS

### 2.1 Kinase conformation assignment

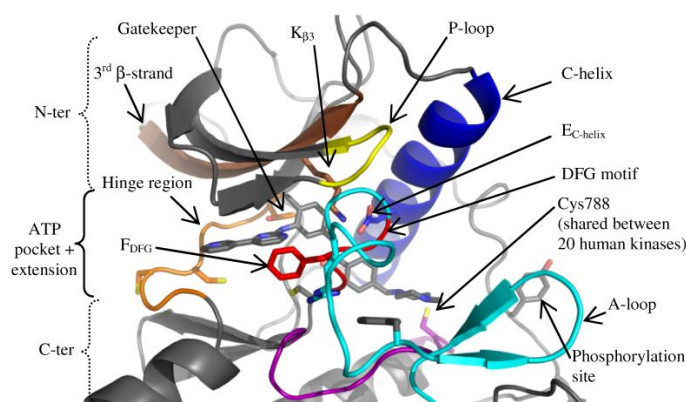
The conformation assignment of a given kinase structure was made possible after its superimposition on one of the seven reference structures adopting the “Active” conformation (one per human kinase phylogenetic group, more details are given in *Leproult et al., submitted paper*). The number of references was justified by the small differences that exist in the secondary structure between each phylogenetic group, despite the global conservation of the “Active” fold. The conformation assignment relies on a decision tree based on the measurement of three displacements of  $C_{\alpha}$  atoms between the reference and the target structures, and one intramolecular distance in the target structure. More precisely, the displacements of the catalytic glutamic acid residue belonging to the C-helix ( $E_{C\text{-helix}}$ ), the phenylalanine residue belonging to the DFG motif ( $F_{DFG}$ ) and a residue located in the activation loop ( $R_{A\text{-loop}}$ ) were measured, as well as the distance between the  $N_{\zeta}$  atom of the catalytic lysine residue and the nearest  $O_{\epsilon}$  atom of the catalytic glutamic acid residue ( $K_{\beta 3}\text{-}E_{C\text{-helix}}$ ). These measurements are very well suited to pinpoint the major fluctuations that are relevant of each conformation. Interestingly, two others conformations can be assigned: the “A-loop out” conformation and the “C-helix and DFG out” conformation. More details are given in *Leproult et al., submitted paper*. Moreover, singularities inside each domain conformation, such as a  $\beta$ -sheet

\*To whom correspondence should be addressed.

displacement, an unusual  $F_{DFG}$  conformation or a non-interacting  $K_{\beta 3}$ - $E_{C-helix}$ , are specified in the conformation details.

## 2.2 Cysteine mapping in the ATP binding pocket

The in-house program Pck (<http://alnitak.u-strasbg.fr/Pck>) has been used to detect the ATP pocket inside each kinase structure so that positions of residues whose side chain participates to the pocket were kept. Once all structures had been processed, the most recurrent positions - by regard of the adopted conformation - were considered as robustly constitutive of the pocket. Then, using a large sequence alignment of all human kinase domains, these positions have been propagated so that each human kinase having a well oriented cysteine residue in its binding pocket could be identified. This analysis revealed numerous opportunities for the design of covalent inhibitor since more than 300 human kinases have at least one cysteine in their ATP pocket and extensions, by regard of the adopted conformation. A more detailed description of this study has recently been published (Leprout et al., 2011).



**Fig. 1.** An adapted view of the catalytic domain orientation and representation mode used by Kinatown to pinpoint important residues and structural elements. The PDB code 1T46 of KIT kinase, corresponding to a “DFG out” conformation with the drug imatinib bound, is represented here.

## 3 IMPLEMENTATION

Kinatown has been implemented as a web application combining a Java program and a Jmol viewer (URL: <http://www.Jmol.org>). It works on all major browsers, but requires the installation of the freely available Jmol applet. Detailed help pages and examples are available online ([Kinatown/KinatownInformation.html](http://Kinatown/KinatownInformation.html)).

### 3.1 Input

The server requires a PDB identifier or a UNIPROT identifier. The server is optimized for processing all chains contained in each PDB file. The first step in the Java program consists in checking if the files have been processed and if the results are already in the database. If not, it will take a few minutes with the detail of all accomplished steps to see the results.

### 3.2 Results and visualization

Each tab corresponds to a processed PDB chain and is color-coded depending on the assigned conformation. Residues having allowed the conformation assignment are highlighted in the structure and can even be compared to their equivalent in the active reference structure (Fig. 1). In addition, phosphorylation sites interpreted from PhosphositePlus database (Hornbeck et al., 2004) are shown (Fig. 1). In a more didactic view about kinase domain folding, the

most important residues (catalytic residues, gatekeeper residue ...) and structural elements (activation loop, hinge region ...) are pinpointed (Fig. 1). Cysteine residues whose side chain participates to the binding pocket can be displayed, with an enclosed note about the number of other protein kinases exhibiting a cysteine at the same position, with their names (Fig. 1). The latter information is available for human kinases only. Disease-related information extracted from the Therapeutic Target Database (Zhu et al., 2010) is also reported, when available.

## CONCLUSION

Kinatown gives a comprehensive view of the structural elements involved in the catalytic function and kinase regulation. Furthermore, Kinatown provides fruitful information for those who might be interested in designing allosteric kinases inhibitors able to covalently react with an unconserved cysteine. We used it ourselves for the design of the first reported covalent inhibitor targeting the “DFG out” conformation of KIT kinase and reacting with Cys788, located in the allosteric pocket (Leprout et al., 2011) (Fig. 1).

## ACKNOWLEDGEMENTS

We would like to thank Pr Patrice Gouet for his advices and comments regarding the website content and rendering.

**Funding:** This work was supported by the Institut de Recherche Pierre Fabre.

**Conflict of Interest:** none declared.

## REFERENCES

- Andzelm,E.R. et al. (1995) Bound to activate: conformational consequences of cyclin binding to CDK2. *Structure*, **3**, 1135-1141.
- Buchdunger,E. et al. (1996) Inhibition of the Abl protein-tyrosine kinase in vitro and in vivo by a 2-phenylaminopyrimidine derivative. *Cancer Res*, **56**, 100-104.
- Henise,J.C. et Taunton,J. (2011) Irreversible Nek2 Kinase Inhibitors with Cellular Activity. *Journal of Medicinal Chemistry*, **54**, 4133-4146.
- Hornbeck,P.V. et al. (2004) PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics*, **4**, 1551-1561.
- Jogireddy,R. et al. (2009) Synthesis of a Resorcylic Acid Lactone (RAL) Library Using Fluorous-Mixture Synthesis and Profile of its Selectivity Against a Panel of Kinases. *Chemistry - A European Journal*, **15**, 11498-11506.
- Karaman,M.W. et al. (2008) A quantitative analysis of kinase inhibitor selectivity. *Nat Biotech*, **26**, 127-132.
- Leprout,E. et al. (2011) Cysteine Mapping in Conformationally Distinct Kinase Nucleotide Binding Sites: Application to the Design of Selective Covalent Inhibitors. *Journal of Medicinal Chemistry*, **54**, 1347-1355.
- Manning,G. et al. (2002) The Protein Kinase Complement of the Human Genome. *Science*, **298**, 1912-1934.
- Nagar,B. et al. (2003) Structural Basis for the Autoinhibition of c-Abl Tyrosine Kinase. *Cell*, **112**, 859-871.
- Subramaniam,D.S. et Hwang,J. (2011) BIBW 2992 in non-small cell lung cancer. *Expert Opin. Investig. Drugs*, **20**, 415-422.
- Yun,C. et al. (2008) The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proceedings of the National Academy of Sciences*, **105**, 2070-2075.
- Zhang,J. et al. (2009) Targeting cancer with small molecule kinase inhibitors. *Nat Rev Cancer*, **9**, 28-39.
- Zhou,W. et al. (2010) A structure-guided approach to creating covalent FGFR inhibitors. *Chem. Biol*, **17**, 285-295.
- Zhu,F. et al. (2010) Update of TTD: Therapeutic Target Database. *Nucleic Acids Res*, **38**, D787-D791.

## 5.3 Perspectives

Kinatown a déjà été testé par des biologistes et des chimistes, et leurs remarques ont été prises en compte dans la version actuelle de l’interface. La principale amélioration à venir concerne la visualisation des partenaires cellulaires présents dans les structures 3D de complexes, puisqu’ils peuvent permettre de mieux comprendre les conformations des kinases observées (exemple de la cycline avec la kinase CDK2). Il est envisagé également de donner accès à la séquence protéique, afin de repérer plus facilement des résidus particuliers sur la structure. D’un point de vue pratique, nous voudrions également donner la possibilité à l’utilisateur de sauver les structures et les informations associées, ou bien lui fournir un accès personnel permettant de disposer des résultats quand il le souhaite. Afin d’assurer la pérennité de la base de données, nous devrions également mettre en place un système de mise à jour assurant le traitement et l’intégration automatiques dans la base de données des kinases de toute nouvelle structure déposée dans la PDB.



## 6 Conclusions et Perspectives

---

La famille des protéines kinases humaines est encore loin de connaître le succès des récepteurs couplés aux protéines G (GPCRs) ou des récepteurs nucléaires (NRs), en tant que famille la plus ciblée par des médicaments. Jusque dans les années 1990, l'industrie pharmaceutique considérait cette famille comme *non druggable*. Il était en effet inconcevable de pouvoir trouver des inhibiteurs capables d'entrer en compétition avec les concentrations cellulaires élevées de l'ATP. Par ailleurs, même si le nombre exact de protéines dans le kinome était encore inconnu, on savait qu'elles étaient nombreuses et que leur site catalytique leur permettant de lier l'ATP devait être extrêmement conservé.

Au milieu des années 1990, le regard des industries pharmaceutiques vis-à-vis des kinases a changé lorsque Novartis a rapporté l'efficacité de leur inhibiteur en développement clinique, l'imatinib. Cette molécule s'est révélée capable de réduire la prolifération de cellules leucémiques dans le sang des patients, en inhibant l'activité de la kinase ABL, et ne présentait que très peu d'effets secondaires, exprimant une très bonne sélectivité. Les autres industries pharmaceutiques ont alors rapidement intégré des kinases dans leurs choix de cibles thérapeutiques et les premiers inhibiteurs ont été développés.

En 2011, soit quinze années plus tard, le bilan est mince avec seulement onze inhibiteurs de kinases commercialisés et surtout de nombreuses molécules dont le développement a dû rapidement être arrêté. Ce qui était à craindre se confirme, beaucoup d'inhibiteurs ne sont pas assez efficaces ou sont trop peu sélectifs. La résolution de la structure 3D de kinases cocrystallisées avec divers inhibiteurs a permis de comprendre les divergences de profil entre des inhibiteurs ATP-compétitifs d'une même kinase et ainsi de classer les inhibiteurs de kinases par types. La majorité des inhibiteurs sont de type I, c'est-à-dire qu'ils se lient à leur cible kinase dans sa conformation catalytiquement *Active*. Au contraire, les inhibiteurs de type II reconnaissent une conformation inactive de la cible kinase. Dans une telle conformation, le site catalytique reste accessible mais sa forme est modifiée et donne accès à une large poche additionnelle dans laquelle les inhibiteurs de type II s'étendent. Quatre des onze inhibiteurs de kinases commercialisés se révèlent être de type II, dont l'imatinib. On peut supposer que leur succès est attribuable au fait qu'ils reconnaissent une conformation inactive car la compétition avec l'ATP est moins élevée, que cette conformation inactive n'est peut être pas adoptée par toutes les kinases et qu'elle donne accès à une poche additionnelle dont les résidus sont moins conservés au sein des kinases.

Ces dernières années, des mutations et des résistances développées par des cibles kinases ont été observées chez certains patients. Sur ces patients, les traitements avec les

inhibiteurs de type II sont devenues inefficaces, en particulier lorsque la mutation favorise l'adoption de la conformation *Active* par le domaine catalytique plutôt que la conformation inactive. Quelques inhibiteurs de type I sont également impactés lorsqu'il s'agit d'une mutation augmentant l'affinité de l'ATP pour le site catalytique. Une nouvelle génération d'inhibiteurs de kinases a été développée du fait qu'elle conserve une efficacité face à une mutation exprimée par la cible kinase EGFR suite aux premiers traitements avec des inhibiteurs de type I (Subramaniam et Hwang, 2011; Janne et al., 2011; Allen et al., 2003; Sequist et al., 2010; Laheru et al., 2008). Il s'agit d'inhibiteurs irréversibles, capables de se lier de façon covalente avec une cystéine du site actif d'EGFR.

Le contenu de mon travail de thèse s'inspire des observations précédemment exposées et tend à démontrer que les kinases sont des cibles *druggables* lorsqu'elles sont inhibées par des inhibiteurs covalents et, si possible, allostériques. Les avantages associés à un inhibiteur covalent peuvent être multiples. D'une part, il est peu affecté par la compétition avec l'ATP car il lui suffit de se lier une seule fois dans le site, de façon covalente, pour exercer une inhibition prolongée. Pour cette même raison, un inhibiteur covalent devrait être moins sensible que ses analogues non covalents face aux mutations qui améliorent l'affinité de l'ATP ou déplacent l'équilibre conformationnel de la kinase. Du point de vue du patient, l'efficacité et la durée d'action de tels inhibiteurs permettent de réduire la dose administrée et la fréquence des prises. En effet, même après clairance apparente du médicament dans le sang, ce dernier reste efficace tant que la cible n'a pas été resynthétisée par la cellule. Mes travaux montrent que l'on peut combiner la stratégie d'inhibition covalente à la stratégie d'inhibition allostérique (inhibiteurs de type II) pour obtenir un inhibiteur présentant ces multiples propriétés.

Une première partie de mon travail a consisté à analyser de manière systématique les structures 3D de kinases pour évaluer la diversité des conformations que peut adopter le domaine catalytique, et voir ainsi les possibilités offertes pour la conception d'inhibiteurs allostériques. Bien que deux conformations inactives, appelées *DFG out* et *C-helix out*, aient été rapportées à ce jour pour plusieurs kinases, on ne sait pas si l'ensemble des 518 protéines kinases humaines sont capables de les adopter. De même, on ne connaît pas l'étendue possible des autres conformations inactives. L'analyse conformationnelle décrite dans ce manuscrit m'a permis d'assigner une conformation à chacun des 1673 domaines catalytiques cristallisés et correspondant à 214 kinases distinctes. D'après cette analyse, environ 35% des 214 kinases

ont été cristallisés au moins une fois dans l'une des deux conformations inactives *C-helix out* et *DFG out*. Ce pourcentage élevé laisse à penser que beaucoup de kinases adoptent l'une de ces deux conformations inactives. Dix d'entre elles ont même été cristallisées dans les deux conformations inactives. Lors de l'analyse conformationnelle, j'ai également été amenée à caractériser deux autres conformations inactives : *A-loop out* et *C-helix and DFG out*. La seconde conformation est très rarement observée (6 kinases distinctes) tandis que la première est adoptée par 10% des 214 kinases cristallisées. Des similitudes observées entre les conformations *A-loop out* et *Active* d'une part, et entre *A-loop out* et les conformations inactives *DFG out* et *C-helix out* laissent envisager un rôle transitoire de la conformation *A-loop out*. En conclusion, la PDB contient beaucoup d'informations sur la flexibilité du domaine catalytique des kinases. Aux vues des premières statistiques, on peut déduire que la majorité des kinases humaines sont sujettes à la régulation allostérique, et donc que la modulation de leur activité par un inhibiteur allostérique est tout à fait envisageable.

En faisant l'hypothèse que toutes les kinases humaines peuvent adopter les conformations *Active*, *DFG out* et *C-helix out*, j'ai procédé à la détection automatique des kinases pouvant présenter des cystéines accessibles dans leur site catalytique, sous l'influence de la conformation adoptée. J'ai ainsi pu constater que plus de 300 kinases possèdent au moins une cystéine accessible, parmi les 518 kinases humaines. Les opportunités pour la conception d'inhibiteurs covalents paraissent importantes. Cependant, plusieurs questions se posent avant de créer un tel inhibiteur : Combien de kinases partagent la cystéine identifiée ? Ces kinases sont-elles très similaires au niveau du site de liaison de l'ATP, en particulier au niveau du résidu *gatekeeper* ? L'une de ces kinases est-elle une cible thérapeutique connue ? Quel est le rôle de cette kinase dans la maladie où elle est impliquée ? L'inhiber en bloquant son site de fixation de l'ATP aura-t-il un effet bénéfique ? Adopte-t-elle au moins une conformation inactive ? Est-elle majoritairement en conformation *Active* dans le tissu malade ? Est-elle sujette à des mutations ou des résistances ?

Une telle démarche m'a conduite à concevoir de manière rationnelle le premier inhibiteur covalent et spécifique de la conformation inactive *DFG out* rapporté à ce jour. Cet inhibiteur est un analogue de l'imatinib qui a la particularité de pouvoir se lier de manière covalente à la cible KIT en réagissant avec une cystéine située dans l'extension du site catalytique lorsque la kinase adopte la conformation inactive *DFG out*. Le mode d'action de cet inhibiteur a été validé par spectrométrie de masse. Sa capacité à inhiber KIT en présence



d'ATP a été confirmée lors de tests d'activité enzymatique. L'excellent profil de sélectivité de notre inhibiteur parmi l'ensemble des kinases humaines résulte de la combinaison de deux propriétés : l'utilisation du squelette de l'imatinib qui était déjà sélectif, auquel s'ajoute la faible conservation de la cystéine parmi les cibles de l'imatinib. Pour exemple, ABL, l'autre cible principale de KIT, n'est pas inhibée par notre analogue covalent. L'efficacité de notre inhibiteur covalent est actuellement comparée à celle de l'imatinib sur des lignées de leucémie à mastocytes présentant une activité constitutive de KIT, due à des mutations de la kinase. Dans la première lignée, l'IC<sub>50</sub> de notre inhibiteur est comparable à celle mesurée pour l'imatinib, et est de l'ordre de 100nM. Des tests de lavage des cellules mimant le processus de clairance, c'est-à-dire d'élimination de l'inhibiteur par l'organisme, montrent que seules les cellules traitées avec l'inhibiteur covalent cessent de proliférer tandis que celles traitées avec l'imatinib poursuivent leur prolifération. Les premiers résultats sont encourageants dans la seconde lignée puisque notre inhibiteur parvient à inhiber la croissance cellulaire (IC<sub>50</sub> de l'ordre de 2µM), alors qu'il s'agit d'une lignée résistante à l'imatinib du fait d'une mutation de KIT (D816V) qui favorise l'adoption de la conformation *Active* par rapport à la conformation inactive *DFG out*. Cela signifierait que malgré la mutation de KIT dans cette lignée, notre inhibiteur est parvenu à se placer au moins une fois dans le site catalytique et à inhiber définitivement KIT. Des tests d'immunoprécipitation seront prochainement effectués pour vérifier que l'efficacité de notre inhibiteur est liée à l'inhibition de la kinase KIT dans les deux lignées cellulaires. Des IC<sub>50</sub> seront également mesurées pour notre inhibiteur sur d'autres lignées cellulaires pour rendre compte de son caractère toxique et déterminer une fenêtre thérapeutique dans laquelle l'utilisation de notre inhibiteur sur des patients peut être envisagée. On peut également envisager d'apporter des modifications à la structure chimique de notre inhibiteur, pour améliorer les IC<sub>50</sub> sur les deux lignées. On pourrait également tenter de moduler sa réactivité en changeant de groupement électrophile pour limiter les risques de toxicité.

En parallèle du travail effectué sur l'inhibiteur allostérique covalent de KIT, j'ai développé plusieurs autres séries d'inhibiteurs covalents de kinases d'intérêt thérapeutique, certains étant également allostériques (voir paragraphe 7.2 « Travaux sur d'autres inhibiteurs covalents de kinases d'intérêt thérapeutique » à la page 194). Le travail pour lequel les synthèses chimiques et les validations expérimentales sont les plus avancées concerne la conception d'inhibiteurs covalents de la kinase EGFR, qui sont issus de modifications apportées au squelette de la molécule naturelle hypothémicine. J'ai, en effet, pris avantage de la faible sélectivité apparente de cette molécule parmi plusieurs kinases pour la

fonctionnaliser de telle sorte qu'elle se lie de manière covalente avec la cystéine située en fin de région charnière chez EGFR. Cette cystéine étant commune à dix autres kinases, j'ai également apporté d'autres modifications permettant à la molécule de s'étendre dans une petite poche hydrophobe, rendue accessible grâce à la petite taille du résidu *gatekeeper* thréonine dans le site catalytique d'EGFR. Cette particularité d'EGFR devrait permettre d'obtenir un très bon profil de sélectivité pour les inhibiteurs développés. Les derniers résultats d'inhibition de l'activité enzymatique d'EGFR ont révélé sept molécules avec une  $IC_{50}$  inférieure à 100nM. La validation de ces composés devrait passer par des étapes similaires à celles effectuées sur l'inhibiteur allostérique covalent de KIT avec, pour finalité, des tests cellulaires sur des lignées exprimant une forme résistante d'EGFR à l'encontre des inhibiteurs actuellement sur le marché.

La méthodologie d'analyse des structures au sein de la famille des kinases, suivie de la cartographie des cystéines et de la conception rationnelle d'inhibiteurs covalents a déjà été en partie appliquée à d'autres familles de protéines. Dans le cadre des récepteurs nucléaires, j'ai pu évaluer la présence de cystéines et la diversité des positions occupées dans le site de fixation du ligand, en regard de la conformation agoniste ou antagoniste adoptée par le domaine. La famille des récepteurs nucléaires étant beaucoup étudiée à l'IGBMC, une application concrète passant par la conception de ligands covalents devrait bientôt se profiler. Ce travail est décrit plus en détail dans la partie Annexe « 7.3 La famille des récepteurs nucléaires » à la page 199.

Dans la course au développement d'inhibiteurs de kinases, la bataille est déjà bien engagée entre les inhibiteurs de type I et les inhibiteurs de type II. Alors que les défenseurs des inhibiteurs de type I prônent leur avantage dans les maladies où la cible kinase développe des mutations activatrices ou des résistances qui stabilisent la conformation *Active* du domaine catalytique, les inhibiteurs de type II restent favorisés lorsque la kinase n'est pas sujette à mutation car les résidus formant l'extension du site sont moins conservés et la compétition avec l'ATP est réduite, du fait que ce soit une conformation inactive qui est visée (Wodicka et al., 2010). Dans tous les cas, l'efficacité d'un médicament relève plus du temps de résidence au niveau de sa cible que de son affinité apparente (Copeland et al., 2006). De ce point de vue, les inhibiteurs covalents irréversibles semblent être la stratégie idéale pour soigner une maladie, à condition que la cible présente un résidu nucléophile accessible dans la

poche visée. Ainsi, les inhibiteurs covalents de kinases font actuellement l'objet de toutes les intentions. Cette nouvelle génération d'inhibiteurs pourraient statuer définitivement sur le caractère *druggable* des kinases, en combinant les avantages des inhibiteurs de type I à ceux de type II. Cependant, le recul sur les conséquences de tels traitements est inexistant et il faudra attendre quelques années pour que cette stratégie d'inhibition des kinases soit définitivement intégrée dans les industries pharmaceutiques. Certains groupes sont déjà confiants, puisque des méthodes et des brevets de développement d'inhibiteurs covalents commencent à apparaître, provenant aussi bien de laboratoires publics (Katritch et al., 2007; Del Rio et al., 2010) que d'industries privées (Singh et al., 2010). Récemment, la jeune société Avila Therapeutics, spécialisée dans la conception d'inhibiteurs covalents, a même signé un partenariat avec Sanofi-Aventis, 5<sup>ème</sup> industrie pharmaceutique mondiale, pour le développement d'inhibiteurs covalents sur six cibles impliquées dans des cancers (Sanofi-aventis signe avec Avila Therapeutics une alliance mondiale de recherche sur des médicaments anticancéreux covalents ciblés - Paperblog).



## 7 Annexes

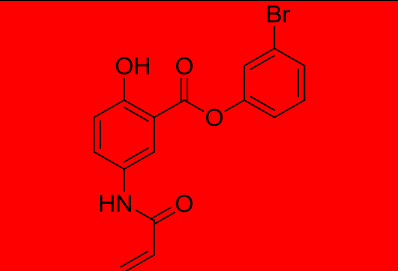
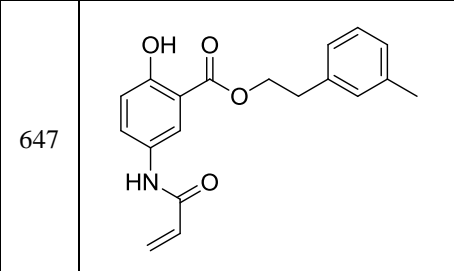
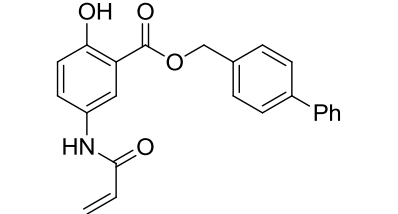
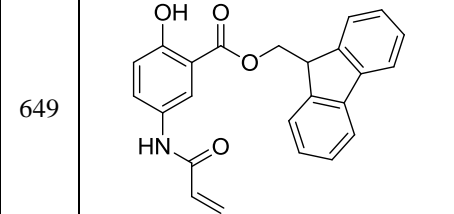
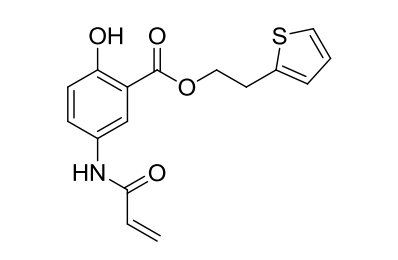
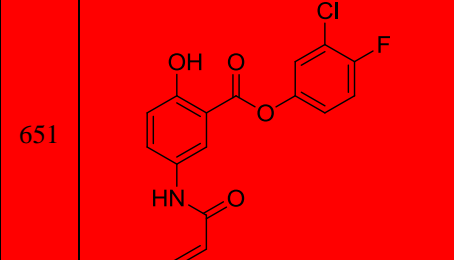
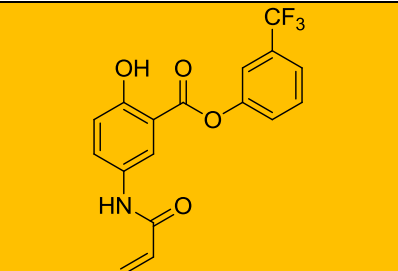
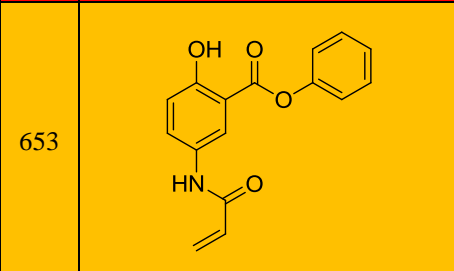
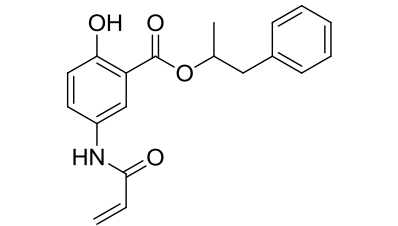
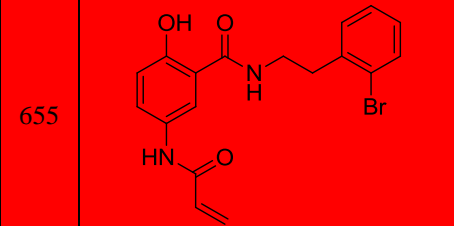
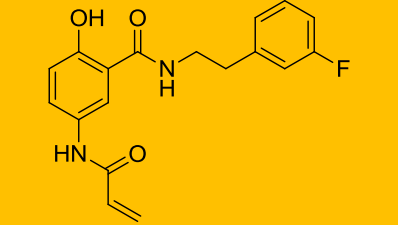
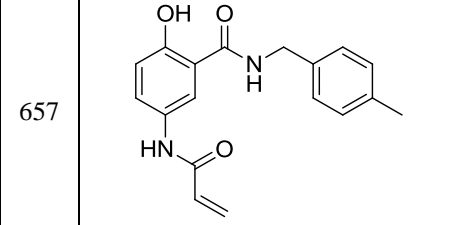
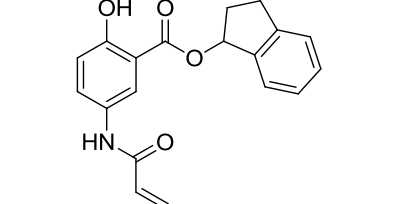
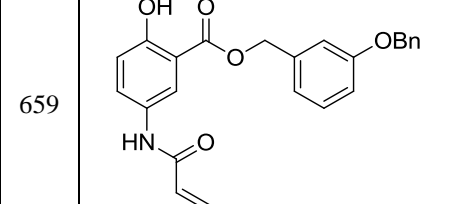
---

7.1 Seconde campagne d'inhibiteurs d'EGFR.....	190
7.2 Travaux sur d'autres inhibiteurs covalents de kinases d'intérêt thérapeutique.....	194
7.3 La famille des récepteurs nucléaires.....	199

## 7.1 Seconde campagne d'inhibiteurs d'EGFR

Id	Structure chimique	% inhibition (concentration d'inhibiteur)
632		11% (100nM) 23% (1µM) 60% (10µM)
634		15% (100nM) 24% (1µM) 57% (10µM)
636		17% (100nM) 11% (1µM) 18% (10µM)
638		1% (100nM) 5% (1µM) 37% (10µM)
640		10% (100nM) 11% (1µM) 37% (10µM)
642		7% (100nM) 12% (1µM) 22% (10µM)
644		6% (100nM) 1% (1µM) 18% (10µM)

Id	Structure chimique	% inhibition (concentration d'inhibiteur)
633		20% (100nM) 54% (1µM) 87% (10µM)
635		23% (100nM) 62% (1µM) 88% (10µM)
637		20% (100nM) 27% (1µM) 73% (10µM)
639		11% (100nM) 25% (1µM) 59% (10µM)
641		2% (100nM) 8% (1µM) 12% (10µM)
643		39% (100nM) 79% (1µM) 96% (10µM)
645		71% (100nM) 93% (1µM) 97% (10µM)

646		83% (100nM) 96% (1µM) 99% (10µM)
647		8% (100nM) 2% (1µM) 6% (10µM)
648		-7% (100nM) -1% (1µM) -3% (10µM)
649		-11% (100nM) 7% (1µM) 3% (10µM)
650		12% (100nM) 48% (1µM) 83% (10µM)
651		70% (100nM) 94% (1µM) 86% (10µM)
652		28% (100nM) 51% (1µM) 94% (10µM)
653		36% (100nM) 76% (1µM) 95% (10µM)
654		-1% (100nM) 4% (1µM) 65% (10µM)
655		51% (100nM) 90% (1µM) 96% (10µM)
656		34% (100nM) 78% (1µM) 94% (10µM)
657		15% (100nM) 42% (1µM) 70% (10µM)
658		14% (100nM) 16% (1µM) 36% (10µM)
659		16% (100nM) 23% (1µM) 36% (10µM)

660		19% (100nM) 15% (1µM) 26% (10µM)
662		11% (100nM) 7% (1µM) 11% (10µM)
664		32% (100nM) 64% (1µM) 92% (10µM)
666		37% (100nM) 80% (1µM) 94% (10µM)
668		8% (100nM) 15% (1µM) 18% (10µM)
670		11% (100nM) 19% (1µM) 25% (10µM)
672		8% (100nM) 10% (1µM) 24% (10µM)
674		8% (100nM) 9% (1µM) 15% (10µM)

661		17% (100nM) 29% (1µM) 44% (10µM)
663		32% (100nM) 61% (1µM) 87% (10µM)
665		40% (100nM) 62% (1µM) 90% (10µM)
667		7% (100nM) 11% (1µM) 10% (10µM)
669		20% (100nM) 14% (1µM) 19% (10µM)
671		15% (100nM) 13% (1µM) 15% (10µM)
673		7% (100nM) 12% (1µM) 21% (10µM)
675		11% (100nM) 17% (1µM) 27% (10µM)



676		15% (100nM) 8% (1µM) 20% (10µM)	677		10% (100nM) -7% (1µM) 14% (10µM)
678		8% (100nM) 36% (1µM) 73% (10µM)	679		18% (100nM) 42% (1µM) 90% (10µM)
680		82% (100nM) 97% (1µM) 100% (10µM)	681		15% (100nM) 48% (1µM) 83% (10µM)
682		29% (100nM) 70% (1µM) 95% (10µM)	683		41% (100nM) 75% (1µM) 97% (10µM)
684		46% (100nM) 81% (1µM) 96% (10µM)	685		17% (100nM) 19% (1µM) 36% (10µM)
686		23% (100nM) 54% (1µM) 90% (10µM)	687		10% (100nM) 9% (1µM) 27% (10µM)
688		19% (100nM) 31% (1µM) 61% (10µM)	689		26% (100nM) 37% (1µM) 81% (10µM)
690		70% (100nM) 90% (1µM) 98% (10µM)	691		73% (100nM) 94% (1µM) 99% (10µM)

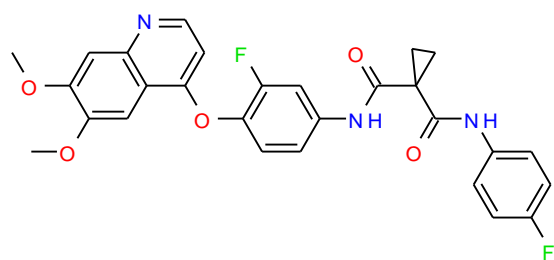
Tableau 17 : Pourcentages d'inhibition de l'activité d'EGFR mesurés pour diverses concentrations d'inhibiteurs. Parmi la soixantaine d'inhibiteurs testés, 21 ont une IC<sub>50</sub> inférieure à 1µM (surlignés en orange) dont 7 ont une IC<sub>50</sub> inférieure à 100nM (surlignés en rouge).

## 7.2 Travaux sur d'autres inhibiteurs covalents de kinases d'intérêt thérapeutique

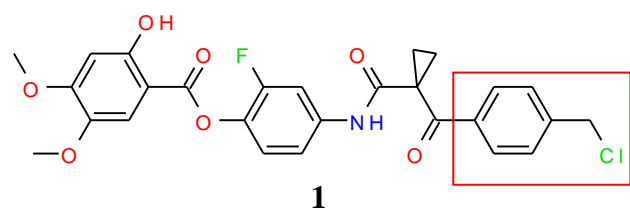
### 7.2.1 La kinase AXL et un second inhibiteur allostérique covalent en vue

La surexpression du récepteur à activité tyrosine kinase AXL, ainsi que des deux autres membres de la sous-famille TAM (TYRO3 et MER), a été mise en cause dans plusieurs cancers (Linger et al., 2008). Aucun inhibiteur de la kinase humaine AXL n'a encore atteint les phases de développement clinique, si ce n'est la molécule foretinib (Figure 67A), dont la cible principale est la tyrosine kinase MET (Eder et al., 2010). La kinase AXL, partageant 36% d'identité de séquence avec le domaine catalytique de MET, s'est en effet révélée être une des cibles secondaires du foretinib. La structure du complexe MET/foretinib a été résolue en 2010 (code PDB : 3LQ8). Le domaine catalytique de MET adopte la conformation *DFG out* et le foretinib s'étend de la poche de liaison de l'ATP jusque dans la poche allostérique rendue accessible par le déplacement de la *A-loop*. Le foretinib est donc un inhibiteur de type II de la kinase MET et l'on peut supposer qu'il adopte le même mode d'inhibition avec la kinase AXL. En effet, aucune structure 3D d'AXL n'est disponible dans la PDB.

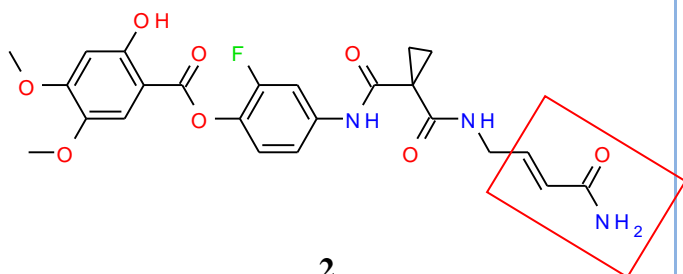
À l'image du travail effectué sur la kinase KIT et la découverte d'un premier inhibiteur allostérique covalent dérivé de l'imatinib, nous avons voulu développer un inhibiteur spécifique et de haute affinité pour la kinase AXL. Le travail de cartographie des cystéines a révélé qu'AXL dispose d'une cystéine sur la C-hélice, la cystéine 588, qui n'est accessible dans le site catalytique que si la kinase adopte la conformation *DFG out* (sphère verte dans la Figure 50C à la page 128). J'ai ainsi inféré le mode de positionnement du foretinib dans un modèle 3D de la kinase AXL adoptant la conformation *DFG out*. J'ai ensuite pu proposer plusieurs modifications du foretinib, avec divers groupements électrophiles qui approchent plus ou moins la cystéine 588 (Figure 67B et C).



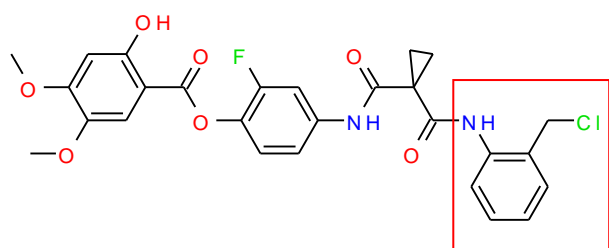
(A) Foretinib



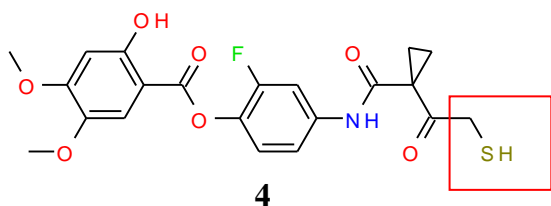
1



2

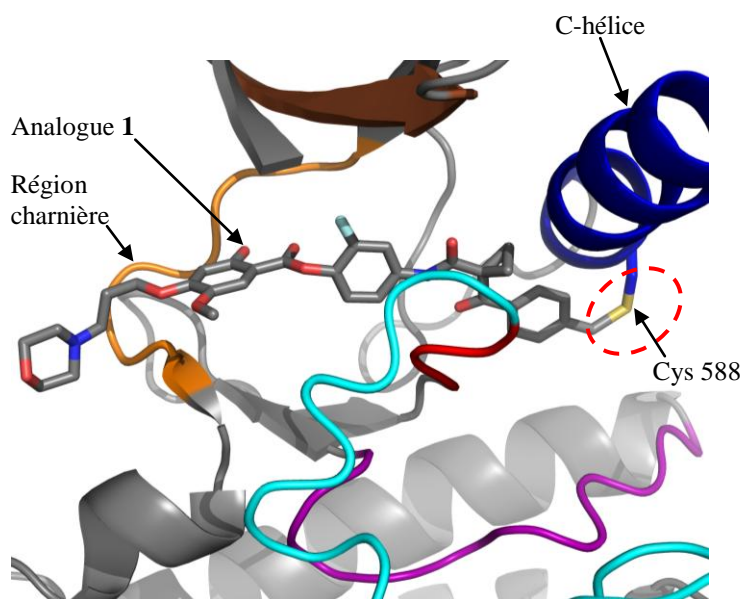


3



4

(B) Propositions d'analogues covalents



(C) Modèle 3D du complexe AXL/Analogue covalent 1

Figure 67 : Propositions d'inhibiteurs allostériques covalents de la kinase d'intérêt thérapeutique AXL. (A) Structure chimique du foretinib. (B) Propositions d'analogues covalents. Les groupes électrophiles sont encadrés en rouge. (C) Modèle 3D du complexe AXL/analogue covalent 1. AXL adopte la conformation *DFG out*. La zone de la liaison covalente avec la cystéine 588 est entourée en pointillés rouges. Par mesure de clarté, la *P-loop* a été supprimée.

Ces molécules sont actuellement synthétisées dans le laboratoire du Professeur N. Winssinger. Elles seront ensuite criblées sur AXL lors de tests *in vitro*. Si les molécules se révèlent efficaces pour inhiber AXL, la validation de leur mode d'inhibition covalente se fera

par spectrométrie de masse. Le profil de sélectivité du meilleur composé covalent révélera sa spécificité parmi les cibles du foretinib ainsi que parmi les six autres cibles disposant d'une cystéine sur la C-hélice.

### 7.2.2 La kinase NDR1 et une application de la méthodologie de « *tethering* »

L'équipe de biologistes du Docteur Ali Hamiche, à l'IGBMC, a mis en évidence l'implication de la kinase NDR1 dans les défauts mitotiques. L'inhibition de cette kinase pourrait permettre de rétablir la division cellulaire. A l'heure actuelle, aucune petite molécule inhibitrice de la kinase humaine NDR1 n'est rapportée dans la littérature, ni même pour les autres trois autres membres de la sous-famille des NDRs kinases (données *Aureus*). Afin de développer de tels inhibiteurs, nous avons pensé adopter la méthode de *tethering* (voir paragraphe 1.3.5 « Autres usages de la liaison covalente » à la page 65). Plus exactement, nous avons proposé aux biologistes d'introduire une cystéine dans le site catalytique de NDR1, à la fin de la région charnière, qui remplace l'acide aspartique D173. Cette position a été choisie car elle correspond exactement à celle ciblée par nos inhibiteurs covalents d'EGFR (voir paragraphe 4.5.2 « Réadressage de l'hypothémicine en vue de l'inhibition covalente d'EGFR via la cystéine située à la fin de la région charnière » à la page 156). Certains de ces inhibiteurs étant de petite taille (240 Da), nous espérons parvenir à les fixer de manière covalente dans le site catalytique de NDR1-D173C. Viendra ensuite toute une étape d'optimisation de la molécule liée de manière covalente en augmentant les interactions possibles avec le site actif, afin d'améliorer significativement l'IC<sub>50</sub> obtenue. Finalement, on retirera le groupe électrophile de la molécule optimisée et l'IC<sub>50</sub> sera remesurée sur NDR1 non mutée pour confirmer l'obtention d'un inhibiteur non covalent de haute affinité. L'inhibiteur sera ensuite testé sur des cellules présentant des défauts mitotiques dans lesquels NDR1 est impliquée.

### 7.2.3 Une seconde cystéine visée chez la kinase EGFR

Des inhibiteurs covalents d'EGFR sont déjà dans des phases de développement clinique. Tous visent la cystéine 797 située à la fin de la région charnière et commune à dix autres kinases. Pour obtenir des inhibiteurs ayant une meilleure sélectivité pour la kinase EGFR, j'ai conçu des molécules covalentes de type I (Figure 68A) susceptibles de réagir avec une autre cystéine, la cystéine 775, située avant le brin  $\beta 4$  dans EGFR (voir l'une des sphères vertes dans la Figure 50A à la page 128). En effet, cette cystéine est commune à seulement quatre autres kinases humaines, EGFR étant la seule tyrosine kinase et la seule cible thérapeutique dans ce groupe. Les molécules proposées sont toutes dérivées de l'acide benzoïque, à l'image des inhibiteurs que nous avons déjà développés pour EGFR (voir paragraphe 4.5.2 « Réadressage de l'hypothémicine en vue de l'inhibition covalente d'EGFR via la cystéine située à la fin de la région charnière » à la page 156).

L'accès à cette cystéine, située à proximité du résidu *gatekeeper*, est facilité si la kinase adopte la conformation *C-helix out*. J'ai alors envisagé de modifier l'inhibiteur lapatinib (Tableau 2 à la page 37) pour qu'il se lie de manière covalente à la cystéine 775 (Figure 68B et C).

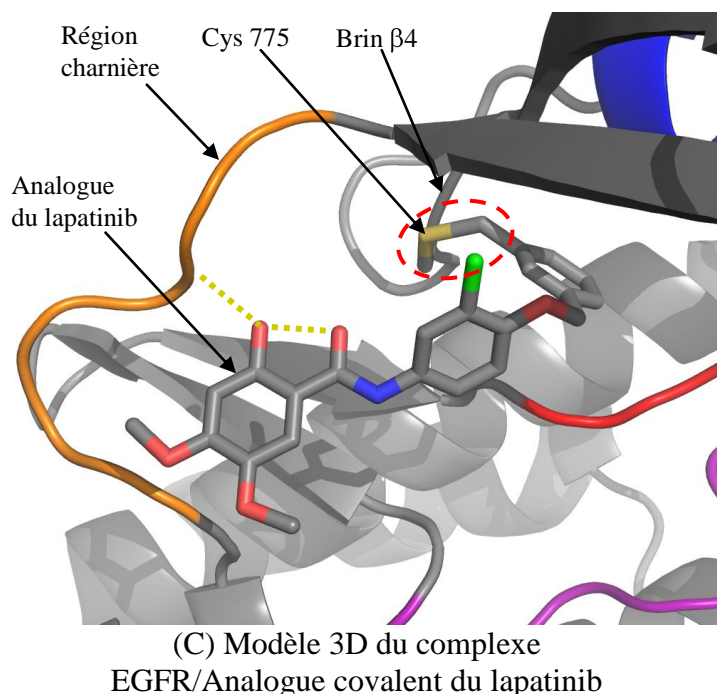
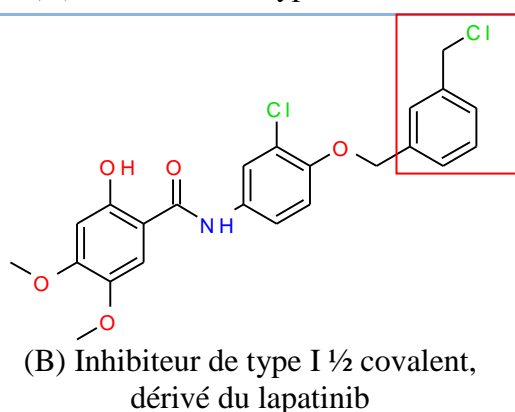
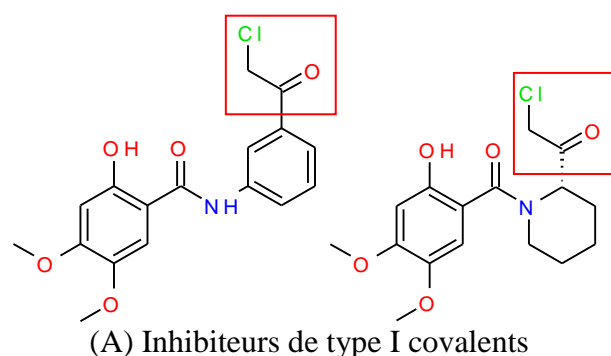
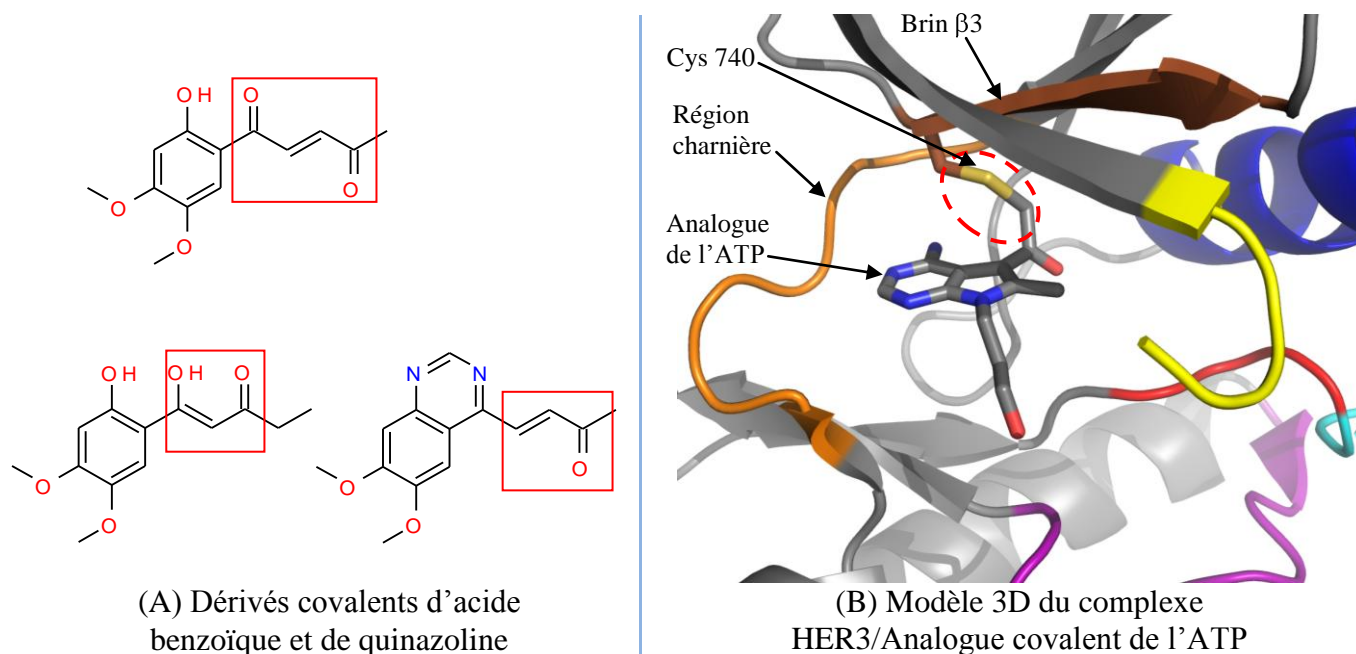


Figure 68 : Propositions d'inhibiteurs covalents de la kinase d'intérêt thérapeutique EGFR, dérivés de l'acide benzoïque et réagissant avec la cystéine située avant le brin  $\beta 4$ . Les groupes électrophiles sont encadrés en rouge. (A) Proposition d'inhibiteurs de type I covalents. (B) Proposition d'inhibiteur allostérique covalent inspiré du lapatinib. (C) Modèle 3D du complexe EGFR/analogue covalent du lapatinib. EGFR adopte la conformation *C-helix out*. La zone de la liaison covalente avec la cystéine 775 est entourée en pointillés rouges. Par mesure de clarté, les brins  $\beta 1$  et  $\beta 2$  ont été supprimés.

## 7.2.4 HER3 comme nouvelle cible thérapeutique

La kinase humaine HER3 appartient à la famille des récepteurs à activité tyrosine kinase et sa surexpression a été observée dans divers cancers (Reschke et al., 2008). Cette kinase est principalement connue pour activer d'autres membres de sa sous-famille, EGFR et HER2, par hétérodimérisation. La particularité de HER3 est qu'il s'agit d'une pseudokinase. En effet, elle présente une histidine à la place du résidu catalytique acide glutamique situé normalement sur la C-hélice. Contre toute attente, il a récemment été démontré que HER3 pouvait lier l'ATP et s'autophosphoryler (Shi et al., 2010). Ces découvertes pourraient lancer la recherche d'inhibiteurs ATP-compétitifs de HER3.

HER3, à l'instar d'EGFR et HER2, ne possède pas de cystéine à la fin de la région charnière. Par contre, elle dispose d'une cystéine, la cystéine 740, située au début du brin  $\beta 3$  et qu'elle ne partage qu'avec une seule autre kinase humaine, HRI (groupe Autres). J'ai procédé à la conception rationnelle de possibles inhibiteurs covalents de HER3, dérivés de l'ATP, de l'acide benzoïque ou de la quinazoline (Figure 69) et visant cette cystéine. Les deux structures du domaine catalytique de HER3 présentes dans la PDB adoptent la conformation *C-helix out*. J'ai donc modélisé le domaine catalytique de HER3 dans la conformation *Active*, à partir d'une structure *Active* d'EGFR.



**Figure 69 :** Propositions d'inhibiteurs covalents de la kinase d'intérêt thérapeutique HER3. (A) Propositions de dérivés covalents d'acide benzoïque et de quinazoline. Les groupes électrophiles sont encadrés en rouge. (B) Modèle 3D du complexe HER3/analogue covalent de l'ATP. La zone de la liaison covalente avec la cystéine 740 est entourée en pointillés rouges. Par mesure de clarté, le brin  $\beta 1$  a été supprimé.

## 7.3 La famille des récepteurs nucléaires

### 7.3.1 Introduction

Les 48 récepteurs nucléaires (RNs) humains forment la plus large famille de régulateurs de la transcription (Renaud et al., 2006). La plupart des RNs fonctionnent sous forme d'homodimères ou d'hétérodimères avec le RN RXR. Les RNs sont activés par des ligands lipophiles se logeant dans le domaine de liaison du ligand (LBD) et sont capables de se lier à l'ADN via un domaine très conservé qui est le domaine de liaison à l'ADN (DBD). Le repliement du LBD, bien que modérément conservé en terme de séquence, est également très bien conservé parmi les RNs. Le LBD est non seulement responsable de la reconnaissance spécifique du ligand mais également du recrutement de cofacteurs (coactivateurs, corépresseurs, etc.). Cette modulation ligand-dépendante de leur activité rend les RNs très attractifs pour l'industrie pharmaceutique.

Le LBD est majoritairement composé de douze hélices et deux brins  $\beta$ . Sa conformation varie principalement au niveau du positionnement de l'hélice terminale H12, selon le type de ligand lié (Figure 70). Ainsi, la liaison d'un ligand agoniste force le déplacement de l'hélice H12 et permet au récepteur de devenir actif, en déstabilisant la surface de liaison du corépresseur ou bien en reformant la surface de liaison du coactivateur, formée des hélices H3, H4 et de l'hélice H12 repositionnée. Il s'agit de la conformation dite « agoniste », en opposition à la conformation inactive dite « antagoniste » où l'hélice H12 occupe le site de liaison du coactivateur.

À l'image du travail d'exploration effectué sur les kinases, nous avons voulu investiguer les récepteurs nucléaires et les possibilités de développer des inhibiteurs covalents, allostériques ou non.

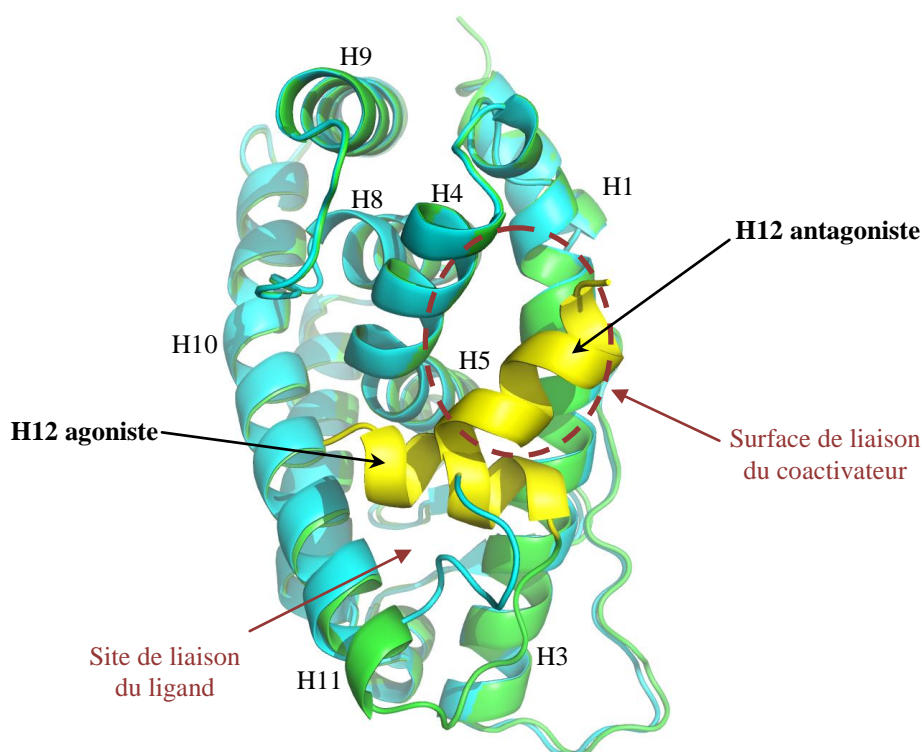


Figure 70: Superposition de deux structures de LBD du récepteur nucléaire ERR3 en conformations agoniste (vert, code PDB : 2ZKC) et antagoniste (bleu, code PDB : 2GPU). L'hélice H12, principale différence entre les deux conformations, est représentée en jaune. H=hélice.

### 7.3.2 Inhibiteurs covalents connus

La littérature regorge d'articles décrivant de nouveaux agonistes ou antagonistes synthétiques de divers RNs. Étonnamment, peu d'articles relatent de molécules se liant de manière covalente au LBD. Ainsi, dès 1987, un marqueur d'affinité covalent du récepteur aux glucocorticoïdes (GCR) a été développé (Simons, 1987). La même technique a été employée pour le récepteur aux œstrogènes ESR1, dont le marqueur réagit avec une cystéine de l'hélice H12 (Harlow et al., 1989). Plus récemment, en 2009, un antagoniste covalent des récepteurs thyroïdiens THA et THB, a été rapporté (Hwang et al., 2009). Il se lierait à une cystéine située sur l'hélice H4. La même année, le mode d'action du trybutyltin (TBT), un contaminant environnemental, a été découvert (le Maire et al., 2009). Il activerait le complexe RXRA-PPARA en se liant de manière covalente à une cystéine située sur l'hélice H10 de RXRA (code PDB : 3E94). Le marquage covalent de VDR, via une cystéine située sur le brin  $\beta$ 1, a également été réalisé (Kaya et al., 2009). Enfin, un autre article relate de PPARG et de la possibilité que son ligand naturel, jusqu'alors inconnu, soit un acide gras comportant une cétone  $\alpha,\beta$ -insaturée (code PDB : 2VV3) réagissant avec une cystéine de l'hélice H3 (Itoh et al., 2008).



Il s'agirait donc d'une stratégie en voie de développement pour cette famille de cibles thérapeutiques, pourtant étudiée depuis bien plus longtemps que les kinases dans l'industrie pharmaceutique (voir Figure 5 à la page 18).

### 7.3.3 Étude de la diversité conformationnelle

Les deux grandes conformations observées chez les RNs sont la conformation agoniste et la conformation antagoniste. Afin d'explorer la diversité conformationnelle au sein du LBD des différents RNs cristallisés, j'ai repris et adapté le programme développé pour les kinases (voir chapitre 3 « Étude de la Diversité Conformationnelle au Sein des Protéines Kinases »). Ainsi, 489 structures de LBD ont été extraites de la PDB, dont 417 correspondent à 36 des 48 NRs humains (Figure 71).

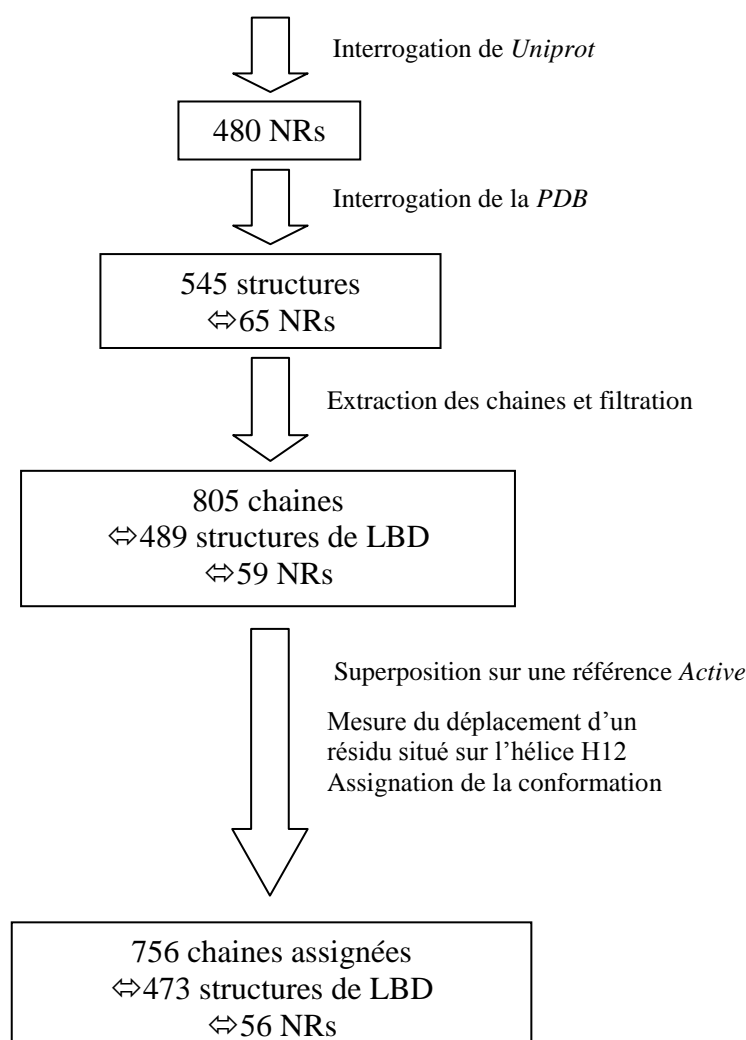


Figure 71 : Succession d'étapes automatisées permettant l'extraction et l'assignation de la conformation de toutes les chaînes de LBD cristallisées.

Un LBD en conformation agoniste a été choisi pour référence (chaîne A dans 1ZGY.pdb, PPARG). L'hélice H12 étant rapportée comme le seul élément structural qui subisse des déplacements entre les différentes conformations, j'ai intégré le calcul du déplacement d'un résidu situé au milieu de cette hélice (Ile472, PPARG) :  $\Delta \text{Ile}_{\text{H12}}\text{-C}_\alpha$  (Figure 72).

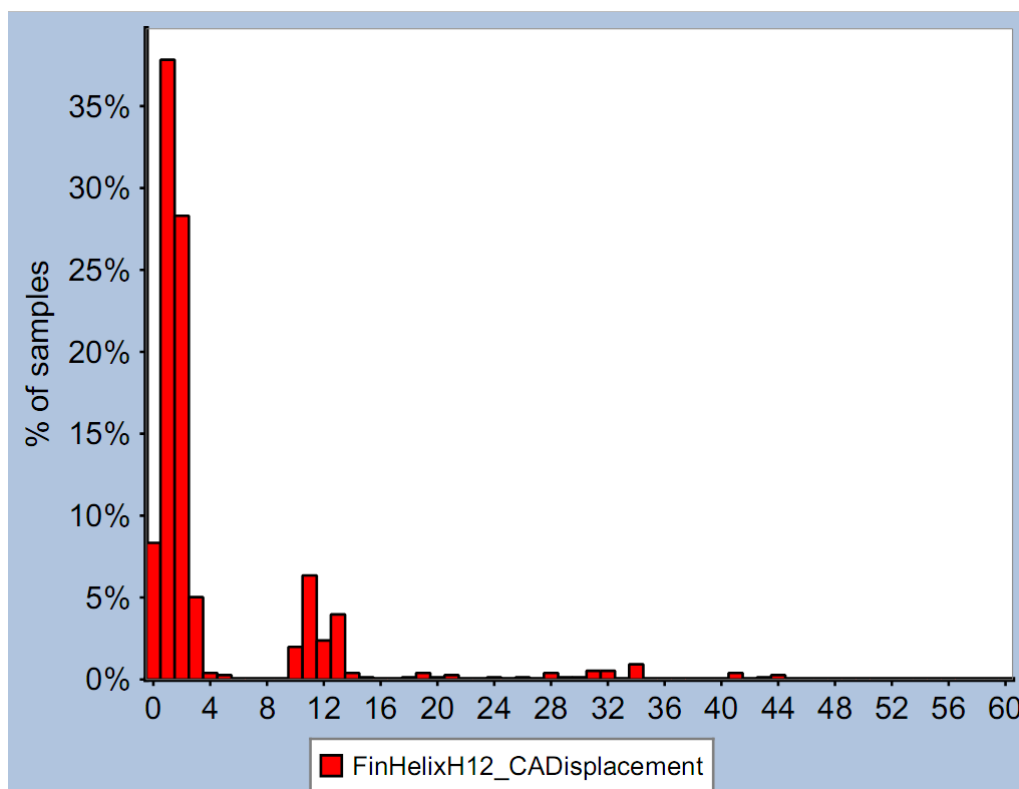


Figure 72 : Distribution du déplacement  $\Delta \text{Ile}_{\text{H12}}\text{-C}_\alpha$  parmi toutes les chaînes de LBD.

La majorité des chaînes présentent un déplacement  $\Delta \text{Ile}_{\text{H12}}\text{-C}_\alpha$  inférieur à  $6.5\text{\AA}$  et se révèlent être en conformation agoniste. Après visualisation des chaînes restantes, toutes ne sont pas en conformation antagoniste. Cela m'a amenée à intégrer un autre critère où le déplacement de l'hélice H12 n'est pas calculé par rapport à un résidu de la référence mais par rapport à la zone où l'hélice H12 devrait se mettre si elle était en conformation antagoniste, soit un point 3D situé dans l'environnement des hélices H3 et H4. Pour cela, j'ai choisi les coordonnées du carbone  $\text{C}_\alpha$  du résidu équivalent à  $\text{Ile}_{\text{H12}}$  dans la structure 1DKF de  $\text{RXR}\alpha$  en conformation antagoniste, après superposition sur le LBD de référence (Figure 73 ; Figure 74).

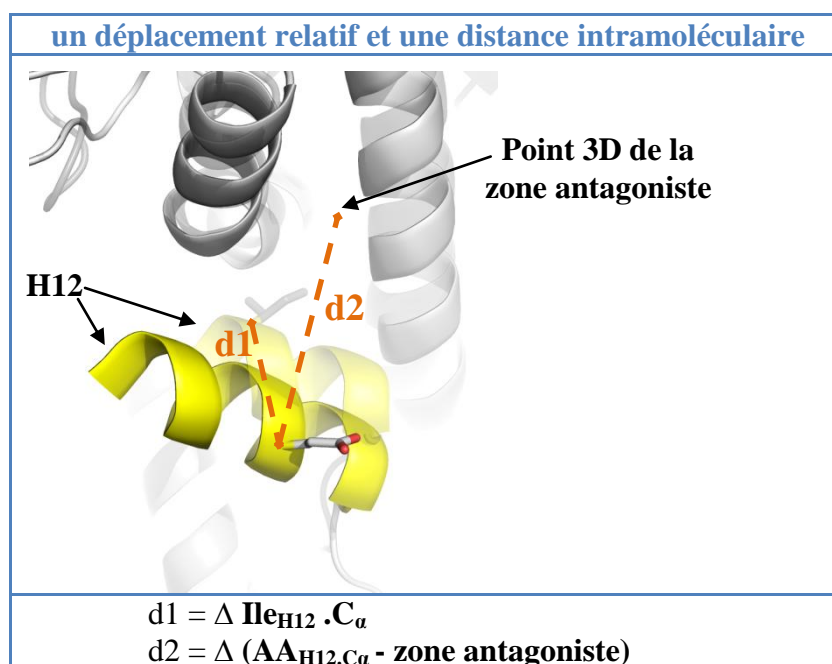


Figure 73 : Un déplacement ( $d1$ ) et une distance ( $d2$ ) sont mesurés pour assigner la conformation. L'hélice H12 du récepteur de référence en conformation agoniste est représentée en jaune transparent. L'hélice H12 d'un autre récepteur préalablement superposé à la référence est représenté en opaque.

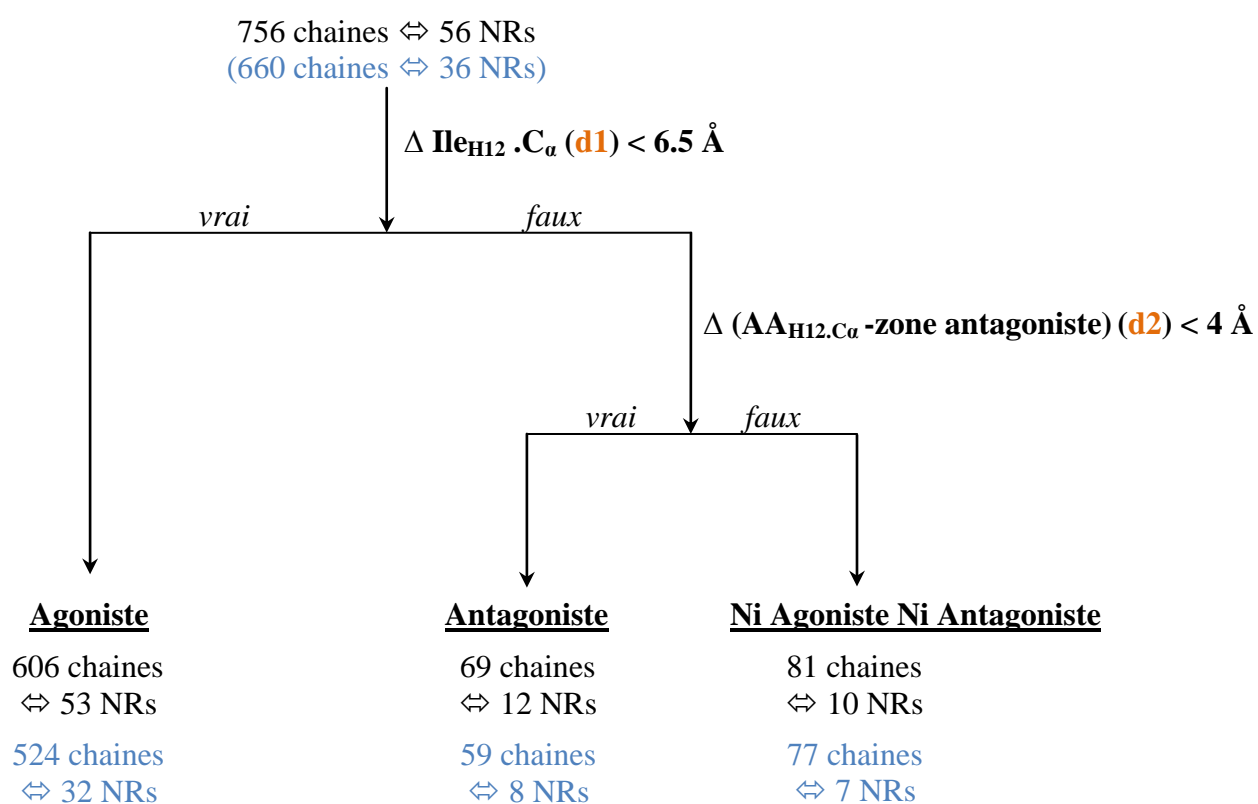


Figure 74 : Arbre de décision pour l'assignation de la conformation des 756 chaînes de LBD. Les chiffres présentés en bleu clair ne concernent que les LBDs humains.



### 7.3.4 Cartographie des cystéines

En analogie avec la recherche des cystéines chez les kinases humaines, la cartographie des cystéines dans le site actif des LBDs de NRs humains s'est faite en distinguant la conformation agoniste de la conformation antagoniste. Au final, ce sont 42 des 48 récepteurs nucléaires humains qui ont au moins une cystéine dans leur site actif, répartie parmi 16 positions de cystéines qui ont été répertoriées dans la conformation agoniste (Figure 76 ; Tableau 18). Comme pour les kinases, on remarque que certaines positions sont beaucoup moins conservées que d'autres donc intéressantes pour la conception futur d'un ligand covalent.

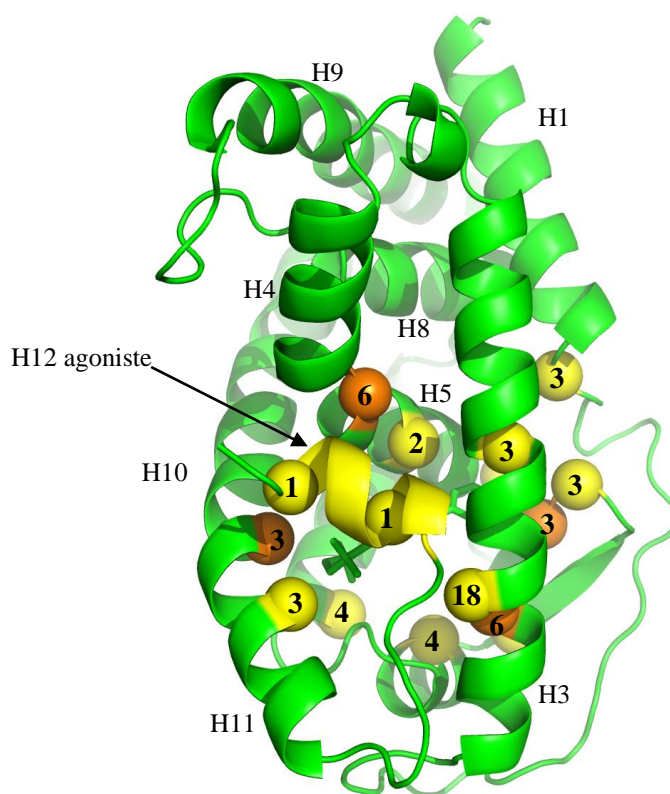


Figure 76 : Cartographie des positions de cystéines trouvées dans le site de liaison du ligand des 48 LBDs de NRs humains, représentée sur une structure du LBD de PPAR $\gamma$  en conformation agoniste (code PDB : 1ZGY). Chaque boule correspond à une position de cystéine et le chiffre indique le nombre de NRs humains partageant cette cystéine. Les boules orange indiquent que ces positions sont déjà ciblées par des ligands covalents rapportés dans la littérature.

Éléments de structure secondaire dans le LBD	Conformation	Résidus correspondant aux positions de cystéine dans PPARG (code PDB : 1ZGY)	Liste des NRs	Nombre de NRs
H1	Agoniste	PHE226	RORG, RORA, RO RB	3
H3	Agoniste Antagoniste	CYS285	PPARA*, PPARD*, PPARG* RORA, RO RB, RORG	6
H3	Agoniste Antagoniste	GLN286	RXRA*, RXRB*, RXRG* HNF4A, HNF4G NR2C2 NR2E1 COT1, COT2, NR2F6 STF1, NR5A2 NR6A1 ERR1*, ERR2, ERR3* NR0B1, NR0B2	18
H3	Agoniste Antagoniste	ALA292	RARA*, RARB*, RARG*	3
H5	Agoniste Antagoniste	VAL322	THA*, THB* RARA*, RARB*, RARG* RO RB	6
H5	Agoniste Antagoniste	ILE326	NR1I2, NR1I3	2
H5	Agoniste Antagoniste	LEU333	NR2C1, NR2C2 COT1, COT2	4
B1	Agoniste Antagoniste	SER342	VDR* NR1I2, NR1I3	3
Boucle B1-B2	Agoniste	GLU343	NR4A1, NR4A2*, NR4A3	3
B2	Agoniste	MET348	NR2C1	1
H6	Agoniste	LEU353	NR4A1, NR4A2*, NR4A3 HNF4G	4
H7	Agoniste Antagoniste	MET364	ANDR*, GCR*, PRGR*, MCR*	4
H10	Agoniste Antagoniste	HIS449	RXRA*, RXRB*, RXRG*	3
H11	Agoniste Antagoniste	LEU453	GCR*, PRGR*, MCR*	3
Boucle H11-H12	Antagoniste	LYS458	ESR1*, ESR2*	2
H12	Agoniste	LEU469	NR6A1	1
H12	Agoniste	TYR473	NR1I3	1

Tableau 18 : Liste exhaustive des NRs humains ayant une cystéine dans le site de liaison du ligand dans le LBD. Les NRs marqués d'une étoile (\*) sont des cibles thérapeutiques. H=hélice ; B=brin.

Dans la conformation antagoniste, seule une position de cystéine se distingue de la conformation agoniste (Tableau 18). Elle se situe sur la boucle entre les hélices H11 et H12, et est commune à seulement deux récepteurs : ESR1 et ESR2 (Figure 77).

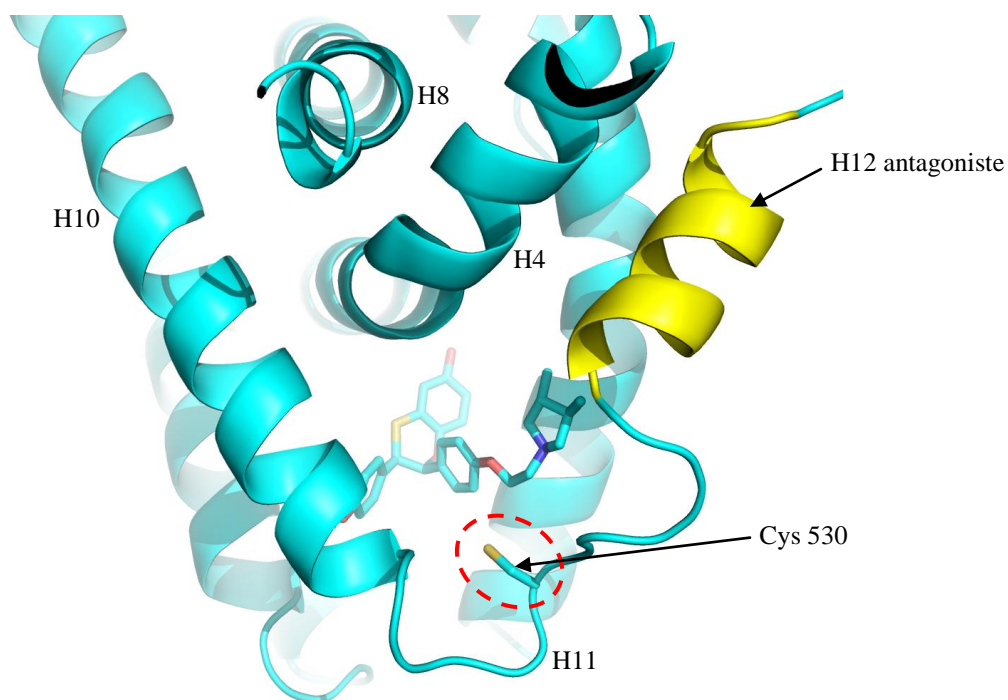


Figure 77 : Structure du LBD d'ESR1 humain qui est le seul récepteur nucléaire, avec ESR2, à disposer d'une cystéine (Cys 530, entourée en pointillés) qui ne participe au site de liaison du LBD que dans la conformation antagoniste (code PDB : 1XP1).

### 7.3.5 Perspectives

Les récepteurs nucléaires constituent la famille de protéines la plus étudiée à l'IGBMC. Parmi ceux étudiés, certains présentent une cystéine peu conservée dans le site de fixation du ligand : VDR, PPAR $\gamma$ , ESR1 et ESR2. Les intérêts apportés par une inhibition (ou une activation) covalente de ces récepteurs devraient être prochainement discutés, avant que des travaux de conception rationnelle de petites molécules covalentes ne commencent.





## Publications et Communications

---

**Liste des publications :**

- \* Leproult, E., Barluenga, S., Moras, D., Wurtz, JM. & Winssinger, N. Cysteine Mapping in Conformationally Distinct Kinase Nucleotide Binding Sites: Application to the Design of Selective Covalent Inhibitors. *Journal of Medicinal Chemistry* 54, 1347-1355 (2011)
- \* Cysteine Vista, "Research Highlights" in *Nature Chemical Biology* 7, 190-191 (2011)
- \* Leproult E., Moras D., Poch O., Wurtz JM. Structural Analysis of the Conformational Diversity inside the Kinase Family: A Drug Design Perspective (*soumis*).
- \* Leproult E. and Wurtz JM. Kinatown: A Web Interface for Kinase Structure Exploration and Subsequent Design of Inhibitors (*soumis*).

**Communications orales :**

- \* Leproult E., Barluenga S., Poch O., Moras D., Winssinger N., Wurtz JM. Kinase Conformational Diversity and Design of Selective Covalent Inhibitors. *Séminaire interne des Laboratoires Pierre Fabre*, Octobre 2010
- \* Leproult E., Barluenga S., Poch O., Moras D., Winssinger N., Wurtz JM. Kinase Conformational Diversity and Design of Selective Covalent Inhibitors. *Séminaire interne de l'IGBMC*, Septembre 2010

**Posters :**

- \* Leproult E., Barluenga S., Moras D., Wurtz JM, Winssinger N. Design of Highly Selective Covalent Kinase Inhibitors (version 2). *18<sup>ème</sup> Journée Des Jeunes Chercheurs En Chimie Thérapeutique*, Paris, Février 2011
- \* Leproult E., Barluenga S., Moras D., Wurtz JM, Winssinger N. Design of Highly Selective Covalent Kinase Inhibitors (version 2). *Congrès EuroQSAR*, Grèce, Septembre 2010
- \* Leproult E., Barluenga S., Moras D., Wurtz JM, Winssinger N. Design of Highly Selective Covalent Kinase Inhibitors (version 1). *2<sup>nd</sup> Summer School of Chemoinformatics*, Obernai, Juin 2010
- \* Leproult E., Barluenga S., Moras D., Wurtz JM, Winssinger N. Design of Highly Selective Covalent Kinase Inhibitors (version 1). *Journées Campus Illkirch*, Illkirch, Mai 2010, **Prix de la meilleure affiche**

## Bibliographie

---

- Adams, J. A. (2001). Kinetic and Catalytic Mechanisms of Protein Kinases. *Chemical Reviews* 101, 2271-2290.
- Akin, C., Brockow, K., D'Ambrosio, C., Kirshenbaum, A. S., Ma, Y., Longley, B. J., et Metcalfe, D. D. (2003). Effects of tyrosine kinase inhibitor STI571 on human mast cells bearing wild-type or mutated c-kit. *Experimental Hematology* 31, 686-692.
- Albou, L., Poch, O., et Moras, D. (2011). M-ORBIS: Mapping of mOlecular Binding sites and Surfaces. *Nucleic Acids Research* 39, 30-43.
- Albou, L., Schwarz, B., Poch, O., Wurtz, J. M., et Moras, D. (2009). Defining and characterizing protein surface using alpha shapes. *Proteins: Structure, Function, and Bioinformatics* 76, 1-12.
- Allen, F. H. (2002). The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr B Struct Sci* 58, 380-388.
- Allen, L. F., Eiseman, I. A., Fry, D. W., et Lenehan, P. F. (2003). CI-1033, an irreversible pan-erbB receptor inhibitor and its potential application for the treatment of breast cancer. *Seminars in Oncology* 30, 65-78.
- Alton, G. R., et Lunney, E. A. (2008). Targeting the unactivated conformations of protein kinases for small molecule drug discovery. *Expert Opin. Drug Discov.* 3, 595-605.
- Andzelm, E. R., Lew, J., et Taylor, S. (1995). Bound to activate: conformational consequences of cyclin binding to CDK2. *Structure* 3, 1135-1141.
- Arrowsmith, J. (2011). Trial watch: Phase III and submission failures: 2007–2010. *Nat Rev Drug Discov* 10, 87.
- Asses, Y., Leroux, V., Tairi-Kellou, S., Dono, R., Maina, F., et Maigret, B. (2009). Analysis of c-Met Kinase Domain Complexes: A New Specific Catalytic Site Receptor Model for Defining Binding Modes of ATP-Competitive Ligands. *Chemical Biology & Drug Design* 74, 560-570.
- Aureus Pharma (2010). Available at: <http://www.aureus-pharma.com/>.
- Azam, M., Seeliger, M. A., Gray, N. S., Kuriyan, J., et Daley, G. Q. (2008). Activation of tyrosine kinases by mutation of the gatekeeper threonine. *Nat. Struct. Mol. Biol* 15, 1109-1118.
- Bard, N., Bolze, R., Caron, E., Desprez, F., Heymann, M., Friedrich, A., Moulinier, L., Nguyen, N. H., Poch, O., et Toursel, T. (2010). Décryphon grid - grid resources dedicated to neuromuscular disorders. *Stud Health Technol Inform* 159, 124-133.
- Barluenga, S., Jogireddy, R., Koripelly, G. K., et Winssinger, N. (2010). In Vivo Efficacy of Natural Product-Inspired Irreversible Kinase Inhibitors. *ChemBioChem* 11, 1692-1699.

- Bar-Sagi, D., Rotin, D., Batzer, A., Mandiyan, V., et Schlessinger, J. (1993). SH3 domains direct cellular localization of signaling molecules. *Cell* 74, 83-91.
- Bateman, A. (2004). The Pfam protein families database. *Nucleic Acids Research* 32, 138D-141.
- Bennett, B. L. (2006). c-Jun N-terminal kinase-dependent mechanisms in respiratory disease. *Eur Respir J* 28, 651-661.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., et Bourne, P. E. (2000). The Protein Data Bank. *Nucl. Acids Res.* 28, 235-242.
- BIRD (2010). Available at: <http://alnitak.u-strasbg.fr/wikili/index.php/BIRD>.
- Blair, J. A., Rauh, D., Kung, C., Yun, C., Fan, Q., Rode, H., Zhang, C., Eck, M. J., Weiss, W. A., et Shokat, K. M. (2007). Structure-guided development of affinity probes for tyrosine kinases using chemical genetics. *Nat Chem Biol* 3, 229-238.
- Blay, J., Le Cesne, A., Alberti, L., et Ray-Coquart, I. (2005). Targeted cancer therapies. *Bull Cancer* 92, E13-18.
- Bogoyevitch, M. A., et Fairlie, D. P. (2007). A new paradigm for protein kinase inhibition: blocking phosphorylation without directly targeting ATP binding. *Drug Discovery Today* 12, 622-633.
- Boudeau, J., Miranda-Saavedra, D., Barton, G. J., et Alessi, D. R. (2006). Emerging roles of pseudokinases. *Trends in Cell Biology* 16, 443-452.
- Britto, P. J., Knipling, L., et Wolff, J. (2002). The Local Electrostatic Environment Determines Cysteine Reactivity of Tubulin. *Journal of Biological Chemistry* 277, 29018-29027.
- Brooijmans, N., Chang, Y., Mobilio, D., Denny, R. A., et Humblet, C. (2010). An enriched structural kinase database to enable kinome-wide structure-based analyses and drug discovery. *Protein Sci* 19, 763-774.
- Burgi, H. B., Dunitz, J. D., et Shefter, E. (1973). Geometrical reaction coordinates. II. Nucleophilic addition to a carbonyl group. *Journal of the American Chemical Society* 95, 5065-5067.
- Burstein, H. J., Sun, Y., Dirix, L. Y., Jiang, Z., Paridaens, R., Tan, A. R., Awada, A., Ranade, A., Jiao, S., Schwartz, G., et al. (2010). Neratinib, an Irreversible ErbB Receptor Tyrosine Kinase Inhibitor, in Patients With Advanced ErbB2-Positive Breast Cancer. *J Clin Oncol*.
- Carlson, S. M., et White, F. M. (2011). Using Small Molecules and Chemical Genetics To Interrogate Signaling Networks. *ACS Chemical Biology* 6, 75-85.
- Carmi, C., Cavazzoni, A., Vezzosi, S., Bordi, F., Vacondio, F., Silva, C., Rivara, S., Lodola, A., Alfieri, R. R., La Monica, S., et al. (2010). Novel Irreversible Epidermal Growth

- Factor Receptor Inhibitors by Chemical Modulation of the Cysteine-Trap Portion. *Journal of Medicinal Chemistry* 53, 2038-2050.
- Celis, J. E., Madsen, P., Celis, A., Nielsen, H. V., et Gesser, B. (1987). Cyclin (PCNA, auxiliary protein of DNA polymerase [ $\delta$ ]) is a central component of the pathway(s) leading to DNA replication and cell division. *FEBS Letters* 220, 1-7.
- Chen, G., Heim, A., Riether, D., Yee, D., Milgrom, Y., Gawinowicz, M. A., et Sames, D. (2003). Reactivity of Functional Groups on the Protein Surface: Development of Epoxide Probes for Protein Labeling. *Journal of the American Chemical Society* 125, 8130-8133.
- Choi, J., Chen, J., Schreiber, S. L., et Clardy, J. (1996). Structure of the FKBP12-Rapamycin Complex Interacting with Binding Domain of Human FRAP. *Science* 273, 239 -242.
- Classement des dix Leaders de l'industrie pharmaceutique au niveau mondial en 2010 | Pharmactua (2010). Available at: <http://www.pharmactua.com/2010/classement-des-dix-leaders-de-l'industrie-pharmaceutique-au-niveau-mondial-en-2010/>.
- Cohen, M. S., Zhang, C., Shokat, K. M., et Taunton, J. (2005). Structural Bioinformatics-Based Design of Selective, Irreversible Kinase Inhibitors. *Science* 308, 1318-1321.
- Cohen, P. (2002a). Protein kinases - the major drug targets of the twenty-first century? *Nat Rev Drug Discov* 1, 309-315.
- Cohen, P. (2002b). The origins of protein phosphorylation. *Nat Cell Biol* 4, E127-E130.
- Copeland, R. A., Pompliano, D. L., et Meek, T. D. (2006). Drug-target residence time and its implications for lead optimization. *Nat Rev Drug Discov* 5, 730-739.
- Corneillie, T. M., Whetstone, P. A., Lee, K. C., Wong, J. P., et Meares, C. F. (2004). Converting Weak Binders into Infinite Binders. *Bioconjugate Chemistry* 15, 1389-1391.
- Cowan-Jacob, S. W. (2006). Structural biology of protein tyrosine kinases. *Cell. Mol. Life Sci.* 63, 2608-2625.
- Cowan-Jacob, S. W., Möbitz, H., et Fabbro, D. (2009). Structural biology contributions to tyrosine kinase drug discovery. *Curr. Opin. Cell Biol* 21, 280-287.
- Cowan-Jacob, S. W., Fendrich, G., Manley, P. W., Jahnke, W., Fabbro, D., Liebetanz, J., et Meyer, T. (2005). The Crystal Structure of a c-Src Complex in an Active Conformation Suggests Possible Steps in c-Src Activation. *Structure* 13, 861-871.
- Cuerrier, D., Moldoveanu, T., Inoue, J., Davies, P. L., et Campbell, R. L. (2006). Calpain Inhibition by  $\alpha$ -Ketoamide and Cyclic Hemiacetal Inhibitors Revealed by X-ray Crystallography<sup>†,‡</sup>. *Biochemistry* 45, 7446-7452.
- Dakas, P., Barluenga, S., Totzke, F., Zirrgiebel, U., et Winssinger, N. (2007). Modular Synthesis of Radicol A and Related Resorcylic Acid Lactones, Potent Kinase

- Inhibitors 13. *Angewandte Chemie International Edition* 46, 6899-6902.
- Dakas, P., Jogireddy, R., Valot, G., Barluenga, S., et Winssinger, N. (2009). Divergent Syntheses of Resorcylic Acid Lactones: L-783277, LL-Z1640-2, and Hypothemycin. *Chemistry - A European Journal* 15, 11490-11497.
- Dalton, G. D., et Dewey, W. L. (2006). Protein kinase inhibitor peptide (PKI): A family of endogenous neuropeptides that modulate neuronal cAMP-dependent protein kinase function. *Neuropeptides* 40, 23-34.
- De Bondt, H. L., Rosenblatt, J., Jancarik, J., Jones, H. D., Morgant, D. O., et Kim, S. (1993). Crystal structure of cyclin-dependent kinase 2. *Nature* 363, 595-602.
- Del Rio, A., Sgobba, M., Parenti, M., Degliesposti, G., Forestiero, R., Percivalle, C., Conte, P., Freccero, M., et Rastelli, G. (2010). A computational workflow for the design of irreversible inhibitors of protein kinases. *Journal of Computer-Aided Molecular Design* 24, 183-94.
- Dhillon, A. S., Hagan, S., Rath, O., et Kolch, W. (2007). MAP kinase signalling pathways in cancer. *Oncogene* 26, 3279-3290.
- Dumont, F. J., et Su, Q. (1995). Mechanism of action of the immunosuppressant rapamycin. *Life Sciences* 58, 373-395.
- Eder, J. P., Shapiro, G. I., Appleman, L. J., Zhu, A. X., Miles, D., Keer, H., Cancilla, B., Chu, F., Hitchcock-Bryan, S., Sherman, L., et al. (2010). A Phase I Study of Foretinib, a Multi-Targeted Inhibitor of c-Met and Vascular Endothelial Growth Factor Receptor 2. *Clinical Cancer Research* 16, 3507 -3516.
- Engh, R. A., et Bossemeyer, D. (2002). Structural aspects of protein kinase control-role of conformational flexibility. *Pharmacol. Ther* 93, 99-111.
- Erlanson, D. A., Braisted, A. C., Raphael, D. R., Randal, M., Stroud, R. M., Gordon, E. M., et Wells, J. A. (2000). Site-directed ligand discovery. *Proc Natl Acad Sci U S A* 97, 9367-9372.
- Erlanson, D. A., Lam, J. W., Wiesmann, C., Luong, T. N., Simmons, R. L., DeLano, W. L., Choong, I. C., Burdett, M. T., Flanagan, W. M., Lee, D., et al. (2003). In situ assembly of enzyme inhibitors using extended tethering. *Nat Biotech* 21, 308-314.
- Fabian, M. A., Biggs, W. H., Treiber, D. K., Atteridge, C. E., Azimioara, M. D., Benedetti, M. G., Carter, T. A., Ciceri, P., Edeen, P. T., Floyd, M., et al. (2005). A small molecule-kinase interaction map for clinical kinase inhibitors. *Nat Biotech* 23, 329-336.
- Fiser, A., et Sali, A. (2003). Modeller: Generation and Refinement of Homology-Based Protein Structure Models. Dans *Macromolecular Crystallography, Part D* (Academic Press), p. 461-491.

- Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., et al. (2004). Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry* 47, 1739-1749.
- Furitsu, T., Tsujimura, T., Tono, T., Ikeda, H., Kitayama, H., Koshimizu, U., Sugahara, H., Butterfield, J. H., Ashman, L. K., et Kanayama, Y. (1993). Identification of mutations in the coding sequence of the proto-oncogene c-kit in a human mast cell leukemia cell line causing ligand-independent activation of c-kit product. *J. Clin. Invest.* 92, 1736-1744.
- Gazdar, A. F. (2009). Activating and resistance mutations of EGFR in non-small-cell lung cancer: role in clinical response to EGFR tyrosine kinase inhibitors. *Oncogene* 28, S24-S31.
- Giles, N. M., Giles, G. I., et Jacob, C. (2003). Multiple roles of cysteine in biocatalysis. *Biochemical and Biophysical Research Communications* 300, 1-4.
- Graves, J. D., et Krebs, E. G. (1999). Protein Phosphorylation and Signal Transduction. *Pharmacology & Therapeutics* 82, 111-121.
- Greenman, C., Stephens, P., Smith, R., Dalgliesh, G. L., Hunter, C., Bignell, G., Davies, H., Teague, J., Butler, A., Stevens, C., et al. (2007). Patterns of somatic mutation in human cancer genomes. *Nature* 446, 153-158.
- Gumireddy, K., Reddy, M. R., Cosenza, S. C., Nathan, R. B., Baker, S. J., Papathi, N., Jiang, J., Holland, J., et Reddy, E. P. (2005). ON01910, a non-ATP-competitive small molecule inhibitor of Plk1, is a potent anticancer agent. *Cancer Cell* 7, 275-286.
- Hanks, S. K., et Hunter, T. (1995). Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J* 9, 576-596.
- Harlow, K., Smith, D., Katzenellenbogen, J., Greene, G., et Katzenellenbogen, B. (1989). Identification of cysteine 530 as the covalent attachment site of an affinity-labeling estrogen (ketononestrol aziridine) and antiestrogen (tamoxifen aziridine) in the human estrogen receptor. *J. Biol. Chem.* 264, 17476-17485.
- Heal, W. P., Dang, T. H. T., et Tate, E. W. (2011). Activity-based probes: discovering new biology and new drug targets. *Chem. Soc. Rev.* 40, 246-257.
- Hearn, B. R., Sundermann, K., Cannoy, J., et Santi, D. (2007). Semisynthesis and Cytotoxicity of Hypothemycin Analogues. *ChemMedChem* 2, 1598-1600.
- Hofmann (2008). Thèse: Resorcylic lactone L-783277 as a new lead structure for kinase inhibition. Available at: <http://e-collection.ethbib.ethz.ch/view/eth:41868>.
- Honigberg, L. A., Smith, A. M., Sirisawad, M., Verner, E., Louny, D., Chang, B., Li, S., Pan, Z., Thamm, D. H., Miller, R. A., et al. (2010). The Bruton tyrosine kinase inhibitor PCI-32765 blocks B-cell activation and is efficacious in models of autoimmune



- disease and B-cell malignancy. *Proceedings of the National Academy of Sciences* 107, 13075-13080.
- Hornbeck, P. V., Chabra, I., Kornhauser, J. M., Skrzypek, E., et Zhang, B. (2004). PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics* 4, 1551-1561.
- Huang, L., He, Z., Guo, L., et Wang, H. (2007). Improvement of Cognitive Deficit and Neuronal Damage in Rats with Chronic Cerebral Ischemia via Relative Long-term Inhibition of Rho-kinase. *Cell Mol Neurobiol* 28, 757-768.
- Hubbard, S. R. (2002). Autoinhibitory mechanisms in receptor tyrosine kinases. *Front. Biosci* 7, d330-340.
- Hubbard, S. R. (1997). Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog. *EMBO J* 16, 5572-5581.
- Hubbard, S. R. (2004). Juxtamembrane autoinhibition in receptor tyrosine kinases. *Nat Rev Mol Cell Biol* 5, 464-471.
- Hubbard, S. R., Wei, L., et Hendrickson, W. A. (1994). Crystal structure of the tyrosine kinase domain of the human insulin receptor. *Nature* 372, 746-754.
- Hughes, B. (2009). 2008 FDA drug approvals. *Nat Rev Drug Discov* 8, 93-96.
- Hunter, T. (1994). 1001 Protein Kinases Redux--Towards 2000. *Seminars in Cell Biology* 5, 367-376.
- Hur, W., Velentza, A., Kim, S., Flatauer, L., Jiang, X., Valente, D., Mason, D. E., Suzuki, M., Larson, B., Zhang, J., et al. (2008). Clinical stage EGFR inhibitors irreversibly alkylate Bmx kinase. *Bioorganic & Medicinal Chemistry Letters* 18, 5916-5919.
- Hussain, H., Krohn, K., Flörke, U., Schulz, B., Draeger, S., Pescitelli, G., Salvadori, P., Antus, S., et Kurtán, T. (2007). Absolute configuration of hypothemycin and 5'-O-methylhypothemycin from *Phoma* sp.--a test case for solid state CD/TDDFT approach. *Tetrahedron: Asymmetry* 18, 925-930.
- Hwang, J. Y., Arnold, L. A., Zhu, F., Kosinski, A., Mangano, T. J., Setola, V., Roth, B. L., et Guy, R. K. (2009). Improvement of Pharmacological Properties of Irreversible Thyroid Receptor Coactivator Binding Inhibitors. *Journal of Medicinal Chemistry* 52, 3892-3901.
- Itoh, T., Fairall, L., Amin, K., Inaba, Y., Szanto, A., Balint, B. L., Nagy, L., Yamamoto, K., et Schwabe, J. W. (2008). Structural basis for the activation of PPAR $\gamma$  by oxidized fatty acids. *Nat Struct Mol Biol* 15, 924-931.
- Jacobs, M. D., Caron, P. R., et Hare, B. J. (2008). Classifying protein kinase structures guides use of ligand-selectivity profiles to predict inactive conformations: Structure of lck/imatinib complex. *Proteins: Structure, Function, and Bioinformatics* 70, 1451-1460.

- Janne, P., Boss, D. S., Camidge, D. R., Britten, C. D., Engelman, J. A., Garon, E. B., Guo, F., Wong, S. G., Liang, J. Q., Letrent, S. P., et al. (2011). Phase I Dose-escalation Study of the Pan-HER Inhibitor, PF299804, in Patients with Advanced Malignant Solid Tumors. *Clin Cancer Res.*
- Jeffrey, P. D., Russo, A. A., Polyak, K., Gibbs, E., Hurwitz, J., Massague, J., et Pavletich, N. P. (1995). Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. *Nature* 376, 313-320.
- Jessani, N., et Cravatt, B. F. (2004). The development and application of methods for activity-based protein profiling. *Curr Opin Chem Biol* 8, 54-59.
- Jogireddy, R., Dakas, P., Valot, G., Barluenga, S., et Winssinger, N. (2009). Synthesis of a Resorcylic Acid Lactone (RAL) Library Using Fluorous-Mixture Synthesis and Profile of its Selectivity Against a Panel of Kinases. *Chemistry - A European Journal* 15, 11498-11506.
- Johnson, L. N. (2009). Protein Kinase Inhibitors: Contributions from Structure to Clinical Compounds. *Quarterly Reviews of Biophysics* 42, 1-40.
- Johnson, L. N., Noble, M. E. M., et Owen, D. J. (1996). Active and Inactive Protein Kinases: Structural Basis for Regulation. *Cell* 85, 149-158.
- Karaman, M. W., Herrgard, S., Treiber, D. K., Gallant, P., Atteridge, C. E., Campbell, B. T., Chan, K. W., Ciceri, P., Davis, M. I., Edeen, P. T., et al. (2008). A quantitative analysis of kinase inhibitor selectivity. *Nat Biotech* 26, 127-132.
- Katritch, V., Byrd, C. M., Tseitin, V., Dai, D., Raush, E., Totrov, M., Abagyan, R., Jordan, R., et Hruby, D. E. (2007). Discovery of small molecule inhibitors of ubiquitin-like poxvirus proteinase I7L using homology modeling and covalent docking approaches. *J. Comput. Aided Mol. Des* 21, 549-558.
- Kaya, T., Swamy, N., Persons, K. S., Ray, S., Mohr, S. C., et Ray, R. (2009). Covalent labeling of nuclear vitamin D receptor with affinity labeling reagents containing a cross-linking probe at three different positions of the parent ligand: Structural and biochemical implications. *Bioorganic Chemistry* 37, 57-63.
- Kemmer, K., Corless, C. L., Fletcher, J. A., McGreevey, L., Haley, A., Griffith, D., Cummings, O. W., Wait, C., Town, A., et Heinrich, M. C. (2004). KIT Mutations Are Common in Testicular Seminomas. *The American Journal of Pathology* 164, 305-313.
- Kennelly, P. J., et Krebs, E. G. (1991). Consensus sequences as substrate specificity determinants for protein kinases and protein phosphatases. *Journal of Biological Chemistry* 266, 15555 -15558.
- Kim, M. J., Yamamoto, D., Matsumoto, K., Inoue, M., Ishida, T., Mizuno, H., Sumiya, S., et Kitamura, K. (1992). Crystal structure of papain-E64-c complex. Binding diversity of E64-c to papain S2 and S3 subsites. *Biochem J* 287, 797-803.
- Kim, Y. J., Pannell, L. K., et Sackett, D. L. (2004). Mass spectrometric measurement of

- differential reactivity of cysteine to localize protein-ligand binding sites.: Application to tubulin-binding drugs. *Analytical Biochemistry* 332, 376-383.
- Kirkland, L. O., et McInnes, C. (2009). Non-ATP competitive protein kinase inhibitors as anti-tumor therapeutics. *Biochem. Pharmacol* 77, 1561-1571.
- Kleywegt, G. J. (1996). Use of Non-crystallographic Symmetry in Protein Structure Refinement. *Acta Crystallogr D Biol Crystallogr* 52, 842-857.
- Knight, Z. A., et Shokat, K. M. (2005). Features of Selective Kinase Inhibitors. *Chemistry & Biology* 12, 621-637.
- Knighton, D. R., Zheng, J. H., Ten Eyck, L. F., Ashford, V. A., Xuong, N. H., Taylor, S. S., et Sowadski, J. M. (1991). Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science* 253, 407-414.
- Kola, I., et Landis, J. (2004). Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov* 3, 711-716.
- Kramer, J. A., Sagartz, J. E., et Morris, D. L. (2007). The application of discovery toxicology and pathology towards the design of safer pharmaceutical lead candidates. *Nat Rev Drug Discov* 6, 636-649.
- Krebs, E. G. (1998). AN ACCIDENTAL BIOCHEMIST. *Biochemistry* 67, xiii-xxxii.
- Kutach, A. K., Villaseñor, A. G., Lam, D., Belunis, C., Janson, C., Lok, S., Hong, L., Liu, C., Deval, J., Novak, T. J., et al. (2010). Crystal Structures of IL-2-inducible T cell Kinase Complexed with Inhibitors: Insights into Rational Drug Design and Activity Regulation. *Chemical Biology & Drug Design* 76, 154-163.
- LaFevre-Bernt, M., Sicheri, F., Pico, A., Porter, M., Kuriyan, J., et Miller, W. T. (1998). Intramolecular Regulatory Interactions in the Src Family Kinase Hck Probed by Mutagenesis of a Conserved Tryptophan Residue. *Journal of Biological Chemistry* 273, 32129-32134.
- Laheru, D., Croghan, G., Bukowski, R., Rudek, M., Messersmith, W., Erlichman, C., Pelley, R., Jimeno, A., Donehower, R., Boni, J., et al. (2008). A phase I study of EKB-569 in combination with capecitabine in patients with advanced colorectal cancer. *Clin. Cancer Res* 14, 5602-5609.
- Lee, C., Jia, Y., Li, N., Sun, X., Ng, K., Ambing, E., Gao, M., Hua, S., Chen, C., Kim, S., et al. (2010). Crystal structure of the ALK (anaplastic lymphoma kinase) catalytic domain. *Biochem. J.* 430, 425-437.
- Leproult, E., Barluenga, S., Moras, D., Wurtz, J., et Winssinger, N. (2011). Cysteine Mapping in Conformationally Distinct Kinase Nucleotide Binding Sites: Application to the Design of Selective Covalent Inhibitors. *Journal of Medicinal Chemistry* 54, 1347-1355.

- Letunic, I., et Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23, 127-128.
- Levinson, N. M., Kuchment, O., Shen, K., Young, M. A., Koldobskiy, M., Karplus, M., Cole, P. A., et Kuriyan, J. (2006). A Src-Like Inactive Conformation in the Abl Tyrosine Kinase Domain. *PLoS Biol* 4, e144.
- Levitsky, K., Ciolli, C. J., et Belshaw, P. J. (2003). Selective Inhibition of Engineered Receptors via Proximity-Accelerated Alkylation. *Organic Letters* 5, 693-696.
- Liao, J. J. (2007). Molecular Recognition of Protein Kinase Binding Pockets for Design of Potent and Selective Kinase Inhibitors. *Journal of Medicinal Chemistry* 50, 409-424.
- Liegl-Atzwanger, B., Fletcher, J. A., et Fletcher, C. D. M. (2010). Gastrointestinal stromal tumors. *Virchows Arch* 456, 111-127.
- Linger, R. M., Keating, A. K., Earp, H. S., et Graham, D. K. (2008). TAM Receptor Tyrosine Kinases: Biologic Functions, Signaling, and Potential Therapeutic Targeting in Human Cancer. Dans (Academic Press), p. 35-83.
- Liu, Y., et Gray, N. S. (2006). Rational design of inhibitors that bind to inactive kinase conformations. *Nat Chem Biol* 2, 358-364.
- Lugo, T., Pendergast, A., Muller, A., et Witte, O. (1990). Tyrosine kinase activity and transformation potency of bcr-abl oncogene products. *Science* 247, 1079 -1082.
- le Maire, A., Grimaldi, M., Roecklin, D., Dagnino, S., Vivat-Hannah, V., Balaguer, P., et Bourguet, W. (2009). Activation of RXR-PPAR heterodimers by organotin environmental endocrine disruptors. *EMBO Rep* 10, 367-373.
- Maly, D. J., Allen, J. A., et Shokat, K. M. (2004). A Mechanism-Based Cross-Linker for the Identification of Kinase-Substrate Pairs. *Journal of the American Chemical Society* 126, 9160-9161.
- Manning, G., Whyte, D. B., Martinez, R., Hunter, T., et Sudarsanam, S. (2002). The Protein Kinase Complement of the Human Genome. *Science* 298, 1912-1934.
- Mayans, O., van der Ven, P. F. M., Wilm, M., Mues, A., Young, P., Furst, D. O., Wilmanns, M., et Gautel, M. (1998). Structural basis for activation of the titin kinase domain during myofibrillogenesis. *Nature* 395, 863-869.
- Meares, C. F. (2008). The chemistry of irreversible capture. *Advanced Drug Delivery Reviews* 60, 1383-1388.
- Mediator : Combien de morts? | France Soir (2011). Available at: <http://www.francesoir.fr/actualite/sante/mediator-combien-morts-72304.html>.
- Mian, A. A., Schull, M., Zhao, Z., Oancea, C., Hundertmark, A., Beissert, T., Ottmann, O. G., et Ruthardt, M. (2009). The gatekeeper mutation T315I confers resistance against small molecules by increasing or restoring the ABL-kinase activity accompanied by

- aberrant transphosphorylation of endogenous BCR, even in loss-of-function mutants of BCR/ABL. *Leukemia* 23, 1614-1621.
- Michalczyk, A., Klüter, S., Rode, H. B., Simard, J. R., Grütter, C., Rabiller, M., et Rauh, D. (2008). Structural insights into how irreversible inhibitors can overcome drug resistance in EGFR. *Bioorganic & Medicinal Chemistry* 16, 3482-3488.
- Min, X., Lee, B., Cobb, M. H., et Goldsmith, E. J. (2004). Crystal Structure of the Kinase Domain of WNK1, a Kinase that Causes a Hereditary Form of Hypertension. *Structure* 12, 1303-1311.
- Mol, C. D., Dougan, D. R., Schneider, T. R., Skene, R. J., Kraus, M. L., Scheibe, D. N., Snell, G. P., Zou, H., Sang, B., et Wilson, K. P. (2004). Structural basis for the autoinhibition and STI-571 inhibition of c-Kit tyrosine kinase. *J. Biol. Chem* 279, 31655-31663.
- Moulin, E., Zoete, V., Barluenga, S., Karplus, M., et Winssinger, N. (2005). Design, Synthesis, and Biological Evaluation of HSP90 Inhibitors Based on Conformational Analysis of Radicol and Its Analogues. *Journal of the American Chemical Society* 127, 6999-7004.
- Mullard, A. (2011). 2010 FDA drug approvals. *Nat Rev Drug Discov* 10, 82-85.
- Neganova, I., Vilella, F., Atkinson, S. P., Lloret, M., Passos, J. F., von Zglinicki, T., O'Connor, J., Burks, D., Jones, R., Armstrong, L., et al. (2011). An Important Role for CDK2 in G1 to S Checkpoint Activation and DNA Damage Response in Human Embryonic Stem Cells. *STEM CELLS*, N/A.
- Ninomiya-Tsuji, J., Kajino, T., Ono, K., Ohtomo, T., Matsumoto, M., Shiina, M., Mihara, M., Tsuchiya, M., et Matsumoto, K. (2003). A Resorcylic Acid Lactone, 5Z-7-Oxozeaenol, Prevents Inflammation by Inhibiting the Catalytic Activity of TAK1 MAPK Kinase Kinase. *Journal of Biological Chemistry* 278, 18485 -18490.
- Nolen, B., Taylor, S., et Ghosh, G. (2004). Regulation of protein kinases; controlling activity through activation segment conformation. *Mol. Cell* 15, 661-675.
- Ohori, M., Kinoshita, T., Yoshimura, S., Warizaya, M., Nakajima, H., et Miyake, H. (2007). Role of a cysteine residue in the active site of ERK and the MAPKK family. *Biochemical and Biophysical Research Communications* 353, 633-637.
- Ohren, J. F., Chen, H., Pavlovsky, A., Whitehead, C., Zhang, E., Kuffa, P., Yan, C., McConnell, P., Spessard, C., Banotai, C., et al. (2004). Structures of human MAP kinase kinase 1 (MEK1) and MEK2 describe novel noncompetitive kinase inhibition. *Nat Struct Mol Biol* 11, 1192-1197.
- Orfao, A., Garcia-Montero, A. C., Sanchez, L., et Escribano, L. (2007). Recent advances in the understanding of mastocytosis: the role of KIT mutations. *British Journal of Haematology* 138, 12-30.
- Our choices from the recent literature (2011). *Nat Chem Biol* 7, 190-191.

- Overington, J. P., Al-Lazikani, B., et Hopkins, A. L. (2006). How many drug targets are there? *Nat Rev Drug Discov* 5, 993-996.
- Pan, Z., Scheerens, H., Li, S., Schultz, B., Sprengeler, P., Burrill, L., Mendonca, R., Sweeney, M., Scott, K., Grothaus, P., et al. (2007). Discovery of Selective Irreversible Inhibitors for Bruton's Tyrosine Kinase. *ChemMedChem* 2, 58-61.
- Pandya, N., Santani, D., et Jain, S. (2005). Role of mitogen-activated protein (MAP) kinases in cardiovascular diseases. *Cardiovasc Drug Rev* 23, 247-254.
- Pao, L. I., Badour, K., Siminovitch, K. A., et Neel, B. G. (2007). Nonreceptor Protein-Tyrosine Phosphatases in Immune Cell Signaling. *Annu. Rev. Immunol.* 25, 473-523.
- Pargellis, C., Tong, L., Churchill, L., Cirillo, P. F., Gilmore, T., Graham, A. G., Grob, P. M., Hickey, E. R., Moss, N., Pav, S., et al. (2002). Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. *Nat Struct Mol Biol* 9, 268-272.
- Paul, S. M., Mytelka, D. S., Dunwiddie, C. T., Persinger, C. C., Munos, B. H., Lindborg, S. R., et Schacht, A. L. (2010). How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov* 9, 203-214.
- Perez, D. I., Conde, S., Pérez, C., Gil, C., Simon, D., Wandosell, F., Moreno, F. J., Gelpí, J. L., Luque, F. J., et Martínez, A. (2009). Thienylhalomethylketones: Irreversible glycogen synthase kinase 3 inhibitors as useful pharmacological tools. *Bioorg. Med. Chem* 17, 6914-6925.
- Pfizer-Objectif de BPA 2012 confirmé en sabrant les coûts Available at: <http://www.investir.fr/infos-conseils-boursiers/infos-conseils-valeurs/infos/pfizer-objectif-de-bpa-2012-confirme-en-sabrant-les-couts-317991.php>.
- Pharmaceutical Research News & Drug Development Pipeline Information (2010). Available at: <http://www.pharmaprojects.com/>.
- Pipeline Pilot is Accelrys' scientific informatics platform (2010). Available at: <http://accelrys.com/products/pipeline-pilot/>.
- Pompella, A., Visvikis, A., Paolicchi, A., Tata, V. D., et Casini, A. F. (2003). The changing faces of glutathione, a cellular protagonist. *Biochemical Pharmacology* 66, 1499-1503.
- Porter, J., Lumb, S., Franklin, R. J., Gascon-Simorte, J. M., Calmiano, M., Riche, K. L., Lallemand, B., Keyaerts, J., Edwards, H., Maloney, A., et al. (2009). Discovery of 4-azaindoles as novel inhibitors of c-Met kinase. *Bioorganic & Medicinal Chemistry Letters* 19, 2780-2784.
- Potashman, M. H., et Duggan, M. E. (2009). Covalent Modifiers: An Orthogonal Approach to Drug Design. *Journal of Medicinal Chemistry* 52, 1231-1246.
- Potter, L. R., et Hunter, T. (1998). Phosphorylation of the Kinase Homology Domain Is Essential for Activation of the A-Type Natriuretic Peptide Receptor. *Mol. Cell. Biol.* 18, 2164-2172.

Powers, J. C., Asgian, J. L., Ekici, Ö. D., et James, K. E. (2002). Irreversible Inhibitors of Serine, Cysteine, and Threonine Proteases. *Chemical Reviews* 102, 4639-4750.

Protein Kinase Substrate Recognition Available at:  
[http://www.neb.com/nebecomm/tech\\_reference/protein\\_tools/protein\\_kinase\\_substrate\\_recognition.asp](http://www.neb.com/nebecomm/tech_reference/protein_tools/protein_kinase_substrate_recognition.asp).

Raaf, J., Issinger, O., et Niefind, K. (2009). First Inactive Conformation of CK2[alpha], the Catalytic Subunit of Protein Kinase CK2. *Journal of Molecular Biology* 386, 1212-1221.

Ranganathan, S., Wang, Y., Kern, F. G., Qu, Z., et Li, R. (2007). Activation loop phosphorylation-independent kinase activity of human protein kinase C  $\zeta$ . *Proteins* 67, 709-719.

Rastelli, G., Rosenfeld, R., Reid, R., et Santi, D. V. (2008). Molecular modeling and crystal structure of ERK2-hypothemycin complexes. *Journal of Structural Biology* 164, 18-23.

Renaud, J., Moras, D., et Wurtz, J. M. (2006). Nuclear hormone receptors : Insights for drug design from structure and modeling. Dans *Comprehensive Medicinal Chemistry*, 2nd Ed., p. 725-747.

Report by the IMS Institute for Healthcare Informatics (2010). The Use of Medicines in the United States: Review of 2010. Available at:  
[http://www.imshealth.com/deployedfiles/imshealth/Global/Content/IMS\\_Institute/Static\\_File/IHII\\_UseOfMed\\_report.pdf](http://www.imshealth.com/deployedfiles/imshealth/Global/Content/IMS_Institute/Static_File/IHII_UseOfMed_report.pdf).

Reschke, M., Mihic-Probst, D., van der Horst, E. H., Knyazev, P., Wild, P. J., Hutterer, M., Meyer, S., Dummer, R., Moch, H., et Ullrich, A. (2008). HER3 Is a Determinant for Poor Prognosis in Melanoma. *Clinical Cancer Research* 14, 5188 -5197.

Robertson, J. G. (2005). Mechanistic Basis of Enzyme-Targeted Drugs. *Biochemistry* 44, 5561-5571.

Robinson, M. J., Harkins, P. C., Zhang, J., Baer, R., Haycock, J. W., Cobb, M. H., et Goldsmith, E. J. (1996). Mutation of Position 52 in ERK2 Creates a Nonproductive Binding Mode for Adenosine 5'-Triphosphate $\dagger$ , $\ddagger$ . *Biochemistry* 35, 5641-5646.

Roskoski Jr., R. (2005). Structure and regulation of Kit protein-tyrosine kinase--The stem cell factor receptor. *Biochemical and Biophysical Research Communications* 338, 1307-1315.

Rossi, A., Kapahi, P., Natoli, G., Takahashi, T., Chen, Y., Karin, M., et Santoro, M. G. (2000). Anti-inflammatory cyclopentenone prostaglandins are direct inhibitors of IKB kinase. *Nature* 403, 103-118.

Russo, A. A., Jeffrey, P. D., Patten, A. K., Massagué, J., et Pavletich, N. P. (1996a). Crystal structure of the p27Kip1 cyclin-dependent-kinase inhibitor bound to the cyclin A-

- Cdk2 complex. *Nature* 382, 325-331.
- Russo, A. A., Jeffrey, P. D., et Pavletich, N. P. (1996b). Structural basis of cyclin-dependent kinase activation by phosphorylation. *Nat Struct Mol Biol* 3, 696-700.
- Saharinen, P., Takaluoma, K., et Silvennoinen, O. (2000). Regulation of the Jak2 tyrosine kinase by its pseudokinase domain. *Mol. Cell. Biol* 20, 3387-3395.
- Sanofi-aventis signe avec Avila Therapeutics une alliance mondiale de recherche sur des médicaments anticancéreux covalents ciblés - Paperblog Available at: <http://www.paperblog.fr/3973871/sanofi-aventis-signe-avec-avila-therapeutics-une-alliance-mondiale-de-recherche-sur-des-medicaments-anticancereux-covalents-cibles/>.
- Santarius, T., Shipley, J., Brewer, D., Stratton, M. R., et Cooper, C. S. (2010). A census of amplified and overexpressed human cancer genes. *Nat Rev Cancer* 10, 59-64.
- Schaller, M. D., Hildebrand, J. D., et Parsons, J. T. (1999). Complex Formation with Focal Adhesion Kinase: A Mechanism to Regulate Activity and Subcellular Localization of Src Kinases. *Mol Biol Cell* 10, 3489-3505.
- Schindler, T., Bornmann, W., Pellicena, P., Miller, W. T., Clarkson, B., et Kuriyan, J. (2000). Structural Mechanism for STI-571 Inhibition of Abelson Tyrosine Kinase. *Science* 289, 1938 -1942.
- Schirmer, A., Kennedy, J., Murli, S., Reid, R., et Santi, D. V. (2006). Targeted covalent inactivation of protein kinases by resorcylic acid lactone polyketides. *Proceedings of the National Academy of Sciences of the United States of America* 103, 4234-4239.
- Schrödinger Available at: <http://www.schrodinger.com/>.
- Schulze-Gahmen, U., De Bondt, H. L., et Kim, S. (1996). High-Resolution Crystal Structures of Human Cyclin-Dependent Kinase 2 with and without ATP: Bound Waters and Natural Ligand as Guides for Inhibitor Design†. *Journal of Medicinal Chemistry* 39, 4540-4546.
- Sequist, L. V., Besse, B., Lynch, T. J., Miller, V. A., Wong, K. K., Gitlitz, B., Eaton, K., Zacharchuk, C., Freyman, A., Powell, C., et al. (2010). Neratinib, an Irreversible Pan-ErbB Receptor Tyrosine Kinase Inhibitor: Results of a Phase II Trial in Patients With Advanced Non-Small-Cell Lung Cancer. *Journal of Clinical Oncology* 28, 3076 - 3083.
- Shan, Y., Seeliger, M. A., Eastwood, M. P., Frank, F., Xu, H., Jensen, M. Ø., Dror, R. O., Kuriyan, J., et Shaw, D. E. (2009). A conserved protonation-dependent switch controls drug binding in the Abl kinase. *Proceedings of the National Academy of Sciences* 106, 139 -144.
- Shi, F., Telesco, S. E., Liu, Y., Radhakrishnan, R., et Lemmon, M. A. (2010). ErbB3/HER3 intracellular domain is competent to bind ATP and catalyze autophosphorylation. *Proc Natl Acad Sci U S A* 107, 7692-7697.



- Sicheri, F., Moarefi, I., et Kuriyan, J. (1997). Crystal structure of the Src family tyrosine kinase Hck. *Nature* 385, 602-609.
- Simard, J. R., Getlik, M., Grutter, C., Pawar, V., Wulfert, S., Rabiller, M., et Rauh, D. (2009). Development of a Fluorescent-Tagged Kinase Assay System for the Detection and Characterization of Allosteric Kinase Inhibitors. *Journal of the American Chemical Society* 131, 13286-13296.
- Simons, S. S. (1987). Selective covalent labeling of cysteines in bovine serum albumin and in hepatoma tissue culture cell glucocorticoid receptors by dexamethasone 21-mesylate. *Journal of Biological Chemistry* 262, 9669-9675.
- Singh, J., Petter, R. C., et Niu, D. (2010). ALGORITHM FOR DESIGNING IRREVERSIBLE INHIBITORS - Patent application. Available at: <http://www.faqs.org/patents/app/20100185419>.
- Singh, J., Petter, R. C., Baillie, T. A., et Whitty, A. (2011). The resurgence of covalent drugs. *Nat Rev Drug Discov* 10, 307-317.
- Smith, A. J. T., Zhang, X., Leach, A. G., et Houk, K. N. (2009). Beyond Picomolar Affinities: Quantitative Aspects of Noncovalent and Covalent Binding of Drugs to Proteins. *Journal of Medicinal Chemistry* 52, 225-233.
- Somoza, J. R., Palmer, J. T., et Ho, J. D. (2002). The Crystal Structure of Human Cathepsin F and Its Implications for the Development of Novel Immunomodulators. *Journal of Molecular Biology* 322, 559-568.
- Spicer, J., et Rudman, S. (2010). EGFR inhibitors in non-small cell lung cancer (NSCLC): the emerging role of the dual irreversible EGFR/HER2 inhibitor BIBW 2992. *Targeted Oncology* 5, 245-255.
- Stamos, J., Sliwkowski, M. X., et Eigenbrot, C. (2002). Structure of the Epidermal Growth Factor Receptor Kinase Domain Alone and in Complex with a 4-Anilinoquinazoline Inhibitor. *Journal of Biological Chemistry* 277, 46265-46272.
- Steiner, et Bugen (2007). The Continuing Evolution of the Pharmaceutical Industry: Career Challenges and Opportunities. Available at: [http://www.pharmawealthmanager.com/pdf/Pharma\\_Paper\\_ADVANCE\\_PUBLIC\\_COPY\\_Nov-30-2007.pdf](http://www.pharmawealthmanager.com/pdf/Pharma_Paper_ADVANCE_PUBLIC_COPY_Nov-30-2007.pdf).
- Subramaniam, D. S., et Hwang, J. (2011). BIBW 2992 in non-small cell lung cancer. *Expert Opin. Investig. Drugs* 20, 415-422.
- Tanaka, H., Nishida, K., Sugita, K., et Yoshioka, T. (1999). Antitumor Efficacy of Hypothemycin, A New Ras-signaling Inhibitor. *Cancer Science* 90, 1139-1145.
- Tanramluk, D., Schreyer, A., Pitt, W. R., et Blundell, T. L. (2009). On the origins of enzyme inhibitor selectivity and promiscuity: a case study of protein kinase binding to staurosporine. *Chem Biol Drug Des* 74, 16-24.

- Taylor, S. S., et Radzio-Andzelm, E. (1997). Protein kinase inhibition: natural and synthetic variations on a theme. *Current Opinion in Chemical Biology* 1, 219-226.
- Taylor, S. S., et Kornev, A. P. (2011). Protein kinases: evolution of dynamic regulatory proteins. *Trends in Biochemical Sciences* 36, 65-77.
- The Nobel Prize in Physiology or Medicine 1992 Available at: [http://nobelprize.org/nobel\\_prizes/medicine/laureates/1992/index.html](http://nobelprize.org/nobel_prizes/medicine/laureates/1992/index.html).
- The Nobel Prize in Physiology or Medicine 2000 Available at: [http://nobelprize.org/nobel\\_prizes/medicine/laureates/2000/](http://nobelprize.org/nobel_prizes/medicine/laureates/2000/).
- The Nobel Prize in Physiology or Medicine 2001 Available at: [http://nobelprize.org/nobel\\_prizes/medicine/laureates/2001](http://nobelprize.org/nobel_prizes/medicine/laureates/2001).
- The UniProt Consortium (2011). Ongoing and future developments at the Universal Protein Resource. *Nucleic Acids Res* 39, D214-D219.
- Thompson, J. D., Gibson, T. J., et Higgins, D. G. (2002). Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics Chapter 2*, Unit 2.3.
- Vicik, R., Busemann, M., Baumann, K., et Schirmeister, T. (2006). Inhibitors of cysteine proteases. *Curr Top Med Chem* 6, 331-353.
- Vilella-Bach, M., Nuzzi, P., Fang, Y., et Chen, J. (1999). The FKBP12-Rapamycin-binding Domain Is Required for FKBP12-Rapamycin-associated Protein Kinase Activity and G1 Progression. *Journal of Biological Chemistry* 274, 4266 -4272.
- Vistoli, G., Pedretti, A., et Testa, B. (2008). Assessing drug-likeness - what are we missing? *Drug Discovery Today* 13, 285-294.
- Walsh, D. A., Perkins, J. P., et Krebs, E. G. (1968). An Adenosine 3',5'-Monophosphate-dependant Protein Kinase from Rabbit Skeletal Muscle. *Journal of Biological Chemistry* 243, 3763 -3765.
- Wang, Y. (2007). Mitogen-Activated Protein Kinases in Heart Development and Diseases. *Circulation* 116, 1413-1423.
- Weerapana, E., Wang, C., Simon, G. M., Richter, F., Khare, S., Dillon, M. B. D., Bachovchin, D. A., Mowen, K., Baker, D., et Cravatt, B. F. (2010). Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature* 468, 790-795.
- Wickstrom, S. A., Lange, A., Montanez, E., et Fassler, R. (2010). The ILK/PINCH/parvin complex: the kinase is dead, long live the pseudokinase! *EMBO J* 29, 281-291.
- Wiesner, S., Wybenga-Groot, L. E., Warner, N., Lin, H., Pawson, T., Forman-Kay, J. D., et Sicheri, F. (2006). A change in conformational dynamics underlies the activation of Eph receptor tyrosine kinases. *EMBO J* 25, 4686-4696.
- Wodicka, L. M., Ciceri, P., Davis, M. I., Hunt, J. P., Floyd, M., Salerno, S., Hua, X. H., Ford,

- J. M., Armstrong, R. C., Zarrinkar, P. P., et al. (2010). Activation state-dependent binding of small molecule kinase inhibitors: structural insights from biochemistry. *Chem. Biol* 17, 1241-1249.
- Wood, E. R., Shewchuk, L. M., Ellis, B., Brignola, P., Brashear, R. L., Caferro, T. R., Dickerson, S. H., Dickson, H. D., Donaldson, K. H., Gaul, M., et al. (2008). 6-Ethynylthieno[3,2-d]- and 6-ethynylthieno[2,3-d]pyrimidin-4-anilines as tunable covalent modifiers of ErbB kinases. *Proceedings of the National Academy of Sciences* 105, 2773-2778.
- Wood, E. R., Truesdale, A. T., McDonald, O. B., Yuan, D., Hassell, A., Dickerson, S. H., Ellis, B., Pennisi, C., Horne, E., Lackey, K., et al. (2004). A Unique Structure for Epidermal Growth Factor Receptor Bound to GW572016 (Lapatinib). *Cancer Research* 64, 6652 -6659.
- Wybenga-Groot, L. E., Baskin, B., Ong, S. H., Tong, J., Pawson, T., et Sicheri, F. (2001). Structural Basis for Autoinhibition of the EphB2 Receptor Tyrosine Kinase by the Unphosphorylated Juxtamembrane Region. *Cell* 106, 745-757.
- Xu, B., English, J. M., Wilsbacher, J. L., Stippec, S., Goldsmith, E. J., et Cobb, M. H. (2000). WNK1, a Novel Mammalian Serine/Threonine Protein Kinase Lacking the Catalytic Lysine in Subdomain II. *J. Biol. Chem.* 275, 16795-16801.
- Xu, R. M., Carmel, G., Sweet, R. M., Kuret, J., et Cheng, X. (1995). Crystal structure of casein kinase-1, a phosphate-directed protein kinase. *EMBO J* 14, 1015-1023.
- Xu, W., Harrison, S. C., et Eck, M. J. (1997). Three-dimensional structure of the tyrosine kinase c-Src. *Nature* 385, 595-602.
- Yamaguchi, H., et Hendrickson, W. A. (1996). Structural basis for activation of human lymphocyte kinase Lck upon tyrosine phosphorylation. *Nature* 384, 484-489.
- Yarbrough, J. W., et Schultz, T. W. (2007). Abiotic Sulfhydryl Reactivity: A Predictor of Aquatic Toxicity for Carbonyl-Containing  $\alpha,\beta$ -Unsaturated Compounds. *Chemical Research in Toxicology* 20, 558-562.
- Yeh, H. H., Ogawa, K., Balatoni, J., Mukhopadhyay, U., Pal, A., Gonzalez-Lepera, C., Shavrin, A., Soghomonyan, S., Flores, L., Young, D., et al. (2011). Molecular imaging of active mutant L858R EGF receptor (EGFR) kinase-expressing nonsmall cell lung carcinomas using PET/CT. *Proceedings of the National Academy of Sciences* 108, 1603 -1608.
- Yun, C., Mengwasser, K. E., Toms, A. V., Woo, M. S., Greulich, H., Wong, K., Meyerson, M., et Eck, M. J. (2008). The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proceedings of the National Academy of Sciences* 105, 2070-2075.
- Zhang, J., Yang, P. L., et Gray, N. S. (2009). Targeting cancer with small molecule kinase inhibitors. *Nat Rev Cancer* 9, 28-39.

- Zhang, X. D., Gillespie, S. K., et Hersey, P. (2004). Staurosporine induces apoptosis of melanoma by both caspase-dependent and -independent apoptotic pathways. *Molecular Cancer Therapeutics* 3, 187-197.
- Zhao, A., Lee, S. H., Mojena, M., Jenkins, R. G., Patrick, D. R., Huber, H. E., Goetz, M. A., Hensens, O. D., Zink, D. L., Vilella, D., et al. (1999). Resorcylic acid lactones: naturally occurring potent and selective inhibitors of MEK. *J. Antibiot* 52, 1086-1094.
- Zheng, J., Trafny, E. A., Knighton, D. R., Xuong, N. H., Taylor, S. S., Ten Eyck, L. F., et Sowadski, J. M. (1993a). 2.2 A refined crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MnATP and a peptide inhibitor. *Acta Crystallogr. D Biol. Crystallogr* 49, 362-365.
- Zheng, J., Knighton, D. R., Xuong, N. H., Taylor, S. S., Sowadski, J. M., et Ten Eyck, L. F. (1993b). Crystal structures of the myristylated catalytic subunit of cAMP-dependent protein kinase reveal open and closed conformations. *Protein Sci* 2, 1559-1573.
- Zheng, J., Knighton, D. R., Ten Eyck, L. F., Karlsson, R., Xuong, N., Taylor, S. S., et Sowadski, J. M. (1993c). Crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with magnesium-ATP and peptide inhibitor. *Biochemistry* 32, 2154-2161.
- Zhou, W., Ercan, D., Chen, L., Yun, C., Li, D., Capelletti, M., Cortot, A. B., Chirieac, L., Iacob, R. E., Padera, R., et al. (2009). Novel mutant-selective EGFR kinase inhibitors against EGFR T790M. *Nature* 462, 1070-1074.
- Zhou, W., Hur, W., McDermott, U., Dutt, A., Xian, W., Ficarro, S. B., Zhang, J., Sharma, S. V., Brugge, J., Meyerson, M., et al. (2010). A structure-guided approach to creating covalent FGFR inhibitors. *Chem. Biol* 17, 285-295.
- Zhu, F., Han, B., Kumar, P., Liu, X., Ma, X., Wei, X., Huang, L., Guo, Y., Han, L., Zheng, C., et al. (2010). Update of TTD: Therapeutic Target Database. *Nucleic Acids Res* 38, D787-D791.
- Zuccotto, F., Ardini, E., Casale, E., et Angiolini, M. (2010). Through the "Gatekeeper Door": Exploiting the Active Kinase Conformation. *Journal of Medicinal Chemistry* 53, 2681-2694.