Nancy-Université
Université Henri Poincaré

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

➢ Contact SCD Nancy 1 : theses.sciences@scd.uhp-nancy.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

Université Henri Poincaré – Nancy I
U.F.R. Sciences et techniques de la matière et des procédés
École doctorale lorraine de chimie et physique moléculaires

Nancy-Université

THESE

Présentée en vue de l'obtention du grade de

Docteur de l'Université Henri Poincaré

en Chimie Informatique et Théorique

Par

Yasmine ASSES

Conception par modélisation et criblage in silico d'inhibiteurs du récepteur c-Met

sous la direction de Bernard MAIGRET

Équipe Orpailleur LORIA, UMR / CNRS 7503

Thèse présentée et soutenue publiquement le 03 octobre 2011 devant le jury :

Membres du jury :

Rapporteurs :

M. Nicolas FLOQUET	CR CNRS, IBMM, Montpellier
M. Luc MORIN-ALLORY	Professeur, Université d'Orléans, Orléans

Examinateurs :

M. Daniel CANET	Professeur, Université Henri Poincaré, NancyI, Nancy
M. Flavio MAINA	DR INSERM, IBDML, Luminy-Marseille
M. David.W. RITCHIE	DR INRIA, Loria, UHP-Nancy I, Nancy
M. Bernard MAIGRET	DR CNRS, Loria, UHP -Nancy I, (Directeur de thèse)

Invité :

M. Amédéo NAPOLI	DR CNRS, Loria, UHP-Nancy I, Nancy (Invité)

Équipe ORPAILLEUR, Laboratoire Lorrain de Recherche en Informatique et ses Applications (LORIA),
UMR 7503 – Campus Scientifique- BP 239- 54506, Vandoeuvre-Lès-Nancy

Loria

Remerciements

Ce travail de thèse a été effectué au sein du Laboratoire Lorrain de Recherche en Informatique et ses Applications (LORIA), sous la direction du Docteur Bernard MAIGRET. Il a été successivement financé par l'Institut National du Cancer (INCa), la Ligue Nationale Contre le Cancer (LNCC) et la Fondation pour la Recherche Médicale en France (FRM) que je remercie pour leur soutien.

Je tiens aussi à remercier chaleureusement, en tout premier lieu, mon directeur de thèse le Dr Bernard MAIGRET pour m'avoir initiée au monde de la bioinformatique. Travailler à ses côtés fut une expérience très enrichissante aussi bien scientifiquement que humainement. Il m'a permis de m'épanouir en me laissant la liberté d'agir, tout en gardant un œil bienveillant sur ma progression.

Je remercie tout particulièrement le Dr Amédéo NAPOLI pour m'avoir ouvert les portes de l'équipe Orpailleur et avoir mis à ma disposition un cadre de travail idéal à tous points de vue. Je le remercie également pour avoir accepté de juger mes travaux.

Je remercie vivement le Professeur Luc MORIN-ALLORY et le Docteur Nicolas FLOQUET pour avoir accepté de rapporter mon mémoire de thèse. Je suis très honorée de les compter parmi les membres de ce jury.

Je remercie le Professeur Daniel CANET de me faire l'honneur de présider ma soutenance de thèse malgré ses nombreuses occupations.

J'associe à ces remerciements les Docteurs David.W. RITCHIE et Flavio MAINA pour avoir accepté de faire partie de mon jury de thèse. Qu'ils reçoivent l'expression de ma sincère gratitude.

Je remercie l'ensemble des membres de l'équipe Orpailleur ainsi que le personnel du LORIA pour les bons moments passés avec eux, trop nombreux pour être listés ! Je remercie également mes « acolytes » de toujours : Léo, Alex, Matthieu, Vincent, Lionel et Jean-Paul, pour leurs encouragements constants, nos discussions scientifiques et les moments inoubliables passés ensemble.

Je remercie très affectueusement mon amie Hanen pour toute l'aide et le soutien qu'elle a su m'apporter, surtout au cours des derniers mois de mon travail de thèse. Merci Hanen !

Je ne saurais conclure ces remerciements sans exprimer ma gratitude pour le soutien constant de ma famille et mes amis depuis l'autre rive de la Méditerranée. Un merci tout particulier à mon papa, ma maman, ma sœur, mon frère et ma Chichette adorés !

Enfin, j'adresse un remerciement incommensurable à celui qui est mon premier lecteur et mon premier supporter : Florent. Pour ton soutien sans faille, pour tous mes moments de doute et de spleen où tu as toujours su me réconforter et pour tous les Kleenex que tu m'as tendus, je te dis « Grazie mille » !

Sommaire

Partie 1 ..... **9**

I- Introduction : cancer, médicaments et objectifs 11
1. Le cancer en quelques chiffres 11
2. Que cache le mot « Cancer » ? 11
3. Les différents types de cancer et différentes classifications 12
4. Les étapes de la transformation cellulaire tumorale 13
4.1. La cancérogenèse 14
4.2. La tumorogenèse 14
5. Traitements et avancées thérapeutiques de nos jours 16
6. Cheminement du progrès thérapeutique 18
6.1. Les étapes de développement du médicament 18
6.2. Coût de l'innovation thérapeutique 21
7. Méthodes *in silico*, méthodes alternatives 23
8. Descriptif et organisation des travaux réalisés 25

Partie 2 ..... **27**

I- Les récepteurs Tyrosine Kinase (RTK) 29
1. Classification des PK 29
1.1. Classe des « Serine/Thréonine » protéines kinases 29
1.2. Classe des « Serine/Thréonine/Tyrosine » protéines kinases 29
1.3. Classe des « Tyrosine » protéines kinases 29
1.3.2. La famille des « Autres protéines kinases » 30
2. Lumière sur la famille des Tyrosine Kinase TK 31
2.1. Les PTK cytosoliques (CTK) 31
2.2. Les PK membranaires ou Récepteur Tyrosine Kinase (RTK) 32
2.3. Constituants du domaine catalytique des Kinases 33
2.4. Les mécanismes d'activation des kinases 35
2.5. Les voies de signalisation médiées par les RTKs 36

II- c-Met : Enzyme d'intérêt 38
1. La structure de c-Met 38
1.1. Le ligand endogène de c-Met 39
2. Action du complexe HGF/c-Met 41
3. c-Met et le développement du cancer 42
4. c-Met et ses inhibiteurs 43
4.1. Les inhibiteurs kinases et les mutations de c-Met 44
4.2. Détails des interactions des complexes c-Met cristallisés disponibles sur la Protein DataBank 49

Partie 3 ..... **55**

I- La conception de médicaments ou Drug Design 57
1. Introduction 57
2. Approches de la conception de médicaments 58
2.1. Approche « Ligand-based » 58
2.2. Approche « Structure-based » 58
3. Découverte rationnelle de médicaments 61
4. La place de l'ordinateur dans le processus de conception de médicaments 62

II- La Mécanique moléculaire 65
1. Fonction d'énergie potentielle et champ de force 66
1.1. Potentiel entre atomes liés 67
1.2. Potentiel entre atomes non-liés 68

2. L'application du champ de force aux biomolécules (paramétrisation) 69
3. Méthodes d'exploration de l'hypersurface d'énergie d'un système moléculaire 70
3.1. Minimisation de l'énergie 70
3.2. Méthodes de recherche conformationnelle 72
4. Les ensembles thermodynamiques 73
5. La dynamique moléculaire 74
5.1. Principe 74
5.2. Intégration des trajectoires 75
6. Méthodes basées sur les modes normaux de vibration 76
6.1. Définition 76
6.2. L'approximation harmonique 77
6.3. Signification et pertinence des modes normaux 79
6.4. Relation des modes normaux avec les mouvements fonctionnels 80
6.5. Obstacles reliées au calcul des modes normaux 80
6.6. Restriction aux modes de basse fréquence 81
6.7. Inclure les modes normaux dans une procédure de docking 81
7. Méthodes de description de l'environnement 82
7.1. Représentation du solvant 82
7.2. Conditions périodiques aux limites 82
7.3. Prise en compte des interactions entre atomes non-liés 83
7.4. Calcul des interactions électrostatiques 84

III- Les autres outils *in silico* utilisés dans ce travail 85
1. Analyse en groupes ou clustering 85
1.1. Les types de clustering 86
1.2. Les domaines d'application du clustering 89
2. Le docking ou amarrage moléculaire 90
2.1. Principe 90
2.2. Représentation du système 91
2.3. La recherche de la pose optimale 92
2.4. L'évaluation du score des poses 96
2.5. Pertinence des résultats de docking et perspectives 98

Partie 4 101

Article 1 102

I- Contexte 103

II- Identification of new aminoacid amides containing the imidazo[2,1-b]benzothiazol-2-ylphenyl moiety as inhibitors of tumorigenesis by oncogenic Met signaling 104

III- Commentaires 105

Article 2 106

I- Contexte 107

II- Analysis of c-Met Kinase Domain Complexes : A new specific Catalytic Site Receptor Model for Defining Binding Modes of ATP-Competitive Ligands 107

III- Commentaires 109

Article 3 110

I- Contexte 111

II- Exploring c-Met kinase flexibility by sampling and clustering its conformational space 111

III- Commentaires 113

Article 4 114

I- Contexte 115

II- Use of ensemble docking procedure to identify new putative c-Met ligands..... 115
III- Commentaires..... 117

Partie 5 119

Bilan et conclusion..... 121
Perspectives..... 122
Références bibliographiques..... 123
Annexes..... 137

Listes des figures et tableaux

FIG. 1: L'évolution des causes de décès au cours des prochaines décennies. 11
FIG. 2: Progression et invasion des cellules cancéreuses (Schéma représentatif). 15
FIG. 3: Structure chimique de l'Imatinib ST571 ou Glivec. 17
FIG. 4: L'aventure du médicament. 21
FIG. 5: Évolution et comparaison internationale des coûts en R&D. 22
FIG. 6: Intervention des méthodes *in silico* dans le processus de découvertes de médicaments. 23
FIG. 7: Représentation du kinome humain. 30
FIG. 8: Représentation des familles de PTK cytosolique ou nRTK. 32
FIG. 9: Représentation des familles RTKs et de leurs membres. 33
FIG. 10: Représentation d'une Kinase (complexe 1R0P/KSA) sous sa conformation active (motif DFG-in). 34
FIG. 11: Représentation du mécanisme d'activation des RTKs. 35
FIG. 12: Phosphorylation d'une tyrosine (liée à une protéine). 36
FIG. 13: Représentation du domaine de la protéine Met. 39
FIG. 14: Représentation schématique de l'homodimère HGF/SF. Représentation de la structure secondaire 3D de l'homodimère HGF/SF. 40
FIG. 15: Les voies de signalisations enclenchées par l'activation du complexe HGF/c-Met 42
FIG. 16 : Évolution des biotechnologies. 58
FIG. 17: Processus itératif de la conception de médicaments et domaines potentiels d'intervention de l'approche *in silico*. 62
Tableau 1: Illustration des 4 plus importantes contributions d'un champ de force dans le cas d'atomes liés. 67
Tableau 2: Illustration des contributions d'un champ de force dans le cas d'atomes non-liés. 68
FIG. 18 : Représentation de conformations initiales et finales sur une hypersuface d'énergie potentielle. 71
FIG. 19 : Conditions périodiques illustrées en 2 dimensions avec une boîte cubique. 83
FIG. 20 : Arborescence des types de classification. 86
FIG. 21: Exemple de dendrogramme (fréquence des mots dans le "Brown corpus"). 87
FIG. 22 : Complémentarité de surface entre un site de liaison et son ligand (représentation basée sur les harmoniques sphériques). 91

Abréviations

ADN : Acide désoxyribonucléique
Afssaps : Agence Française de Sécurité SAnitaire des Produits de Santé
AMM: Autorisation de Mise sur le Marché
APC : Adenomatous Polyposis Coli
ATP : Adénosine Triphosphate
CRHD : Criblage Réel à Haut Débit
CVHD : Criblage Virtuel à Haut Débit
DM: Dynamique moléculaire
FAK: Focal Adhesion Kinase
FRM : Fondation de la Recherche Médicale
Grb2: Growth factor receptor-bound protein 2
HGF/SF : *Hepatocyte Growth Factor/*Scatter Factor
INCa : Institut National du Cancer
LMC : Leucémie Myéloïde Chronique
LNCC : Ligue Nationale Contre le Cancer
MAPK: Mitogen Activated Protein Kinase
MEK: Mitogen ERK Kinase
MET : Mesenchymal Epithelial Transition
MM/PBSA: Molecular Mechanics / Poisson-Boltzmann Surface Area
MN : Modes Normaux de vibration
OMS : Organisation Mondiale de la Santé
P53 : protéine de 53 kDa
PDB: Protein Data Bank
PK: Protéine(s) Kinase
PTK: Protéine(s) Tyrosine Kinase
R&D: Recherche et Développement
Ras : Ratsarcoma virus
RMN : Résonance Magnétique Nucléaire
RMSD : Root Mean Square Deviation (écart quadratique moyen)
RX : Diffraction des Rayons-X
SH2, SH3 : Domaine d'homologie à Src de type 2, de type 3
SVM : Support Vector Machines
VEGF: Vascular Endothelial Growth Factor
VEGFR: Vascular Endothelial Growth Factor Receptor

PARTIE 1

INTRODUCTION ET ÉTAT DE L'ART

I- Introduction : cancer, médicaments et objectifs

1. Le cancer en quelques chiffres

De nos jours, les pathologies liées aux cancers connaissent une croissance qui va de pair avec celle de la population mondiale. Ainsi, d'après l'OMS (Organisation Mondiale de la Santé), le nombre de cancers diagnostiqués a atteint 12 millions à la fin de l'année 2008. 7 millions de décès proviennent directement de cette maladie. La menace est telle, selon le rapport de l'OMS, que le nombre de cas de cancers devrait doubler entre 2000 et 2020 et tripler d'ici 2030 par rapport au début de ce siècle.

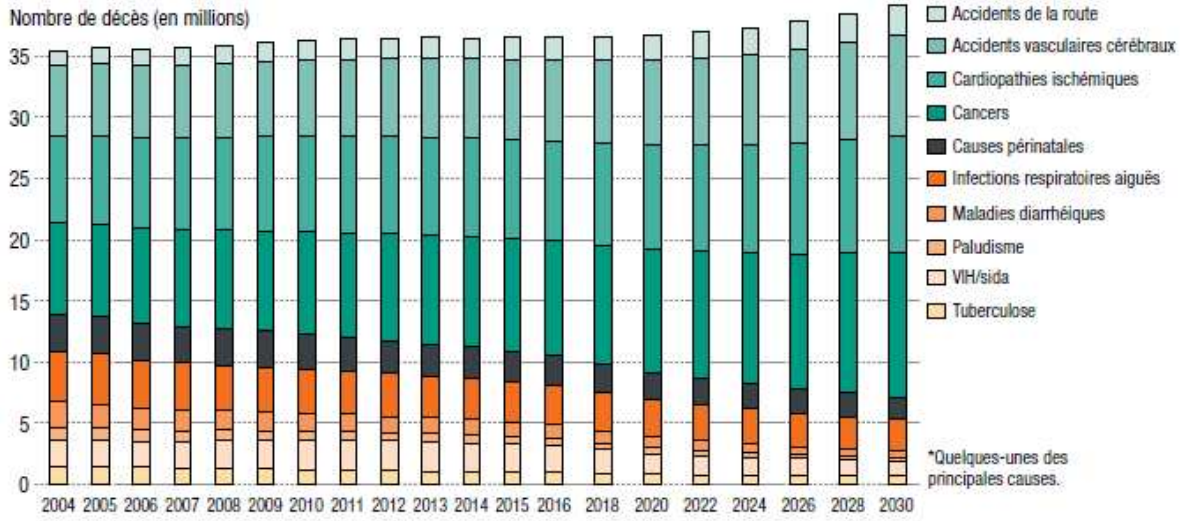


FIG. 1: L'évolution des causes de décès au cours des prochaines décennies[1]

La figure 1 représente l'évolution des causes de mortalité selon le rapport daté de 2008 de l'OMS [1] sur la population mondiale, à titre informatif ainsi que prédictif. Nous constatons l'accroissement du taux de mortalité causée par les maladies cancéreuses. D'où la nécessité immédiate d'effectuer des recherches dans l'axe des pathologies en relation directe avec le cancer.

2. Que cache le mot « Cancer » ?

Angoissant, quand il est prononcé au milieu d'une discussion, le mot « cancer » est habituellement synonyme de fatalité. Cependant, le décès d'un cancer ne signifie pas grand-chose en soit. Une personne atteinte d'un cancer, connu pour être une dégénérescence des cellules de l'organisme, décède en fait des conséquences de sa progression dans l'organisme

comme par exemple : la dénutrition, l'insuffisance respiratoire ou l'empoisonnement cellulaire (accumulation de substances toxiques normalement évacuées par des organes, tels que le foie ou les reins).

Plus explicitement, le cancer, s'il n'est pas traité à temps, provoque le dérèglement des fonctions vitales telles que la respiration, la digestion et l'excrétion.

Lorsqu'une division et/ou prolifération cellulaire anarchique atteint des cellules du tissu de l'organisme, on dit que ce dernier est atteint de cancer [2, 3]. De nouvelles cellules résultant de ce trouble des mécanismes du renouvellement cellulaire se forment. Elles sont connues sous le nom de tumeurs, néoplasmes, carcinomes ou encore métastases (tumeurs filles).

Comme c'est une pathologie liée aux troubles de la reproduction cellulaire, le cancer est une maladie qui se caractérise par des modifications, causées par des facteurs externes ou internes, des gènes.

L'altération constitutionnelle d'un certain type de gène constitue la cause des cancers dits familiaux qui peuvent se transmettre à la descendance.

Les gènes altérés en question se distinguent en 4 catégories :

- Les oncogènes (état muté des proto-oncogènes) : considérés comme des catalyseurs, ils régulent positivement la prolifération ou croissance cellulaire anarchique.
- Les gènes suppresseurs de tumeurs : au contraire de la première catégorie, ils inhibent la prolifération ou croissance cellulaire anarchique.
- Les gènes angiogéniques : la cellule cancéreuse se divise rapidement et a besoin d'un grand apport en oxygène véhiculé par l'hémoglobine. Ce type de gène contrôle l'apport sanguin des cellules. Son altération favorise donc l'angiogénèse des cellules, dont les cancéreuses.
- Les gènes métastasiques : la plupart des cellules sont assignées à des places bien précises dans les tissus s'entourant de cellules voisines qui aident à la définition de la localisation. Une fois les cellules délocalisées de leurs tissus, elles s'arrêtent de proliférer et meurent. Les cellules cancéreuses peuvent acquérir, par le biais de mutations, la faculté de se déplacer sans mourir permettant ainsi leur diffusion dans l'organisme. Les gènes métastasiques permettent le contrôle de la diffusion des cellules cancéreuses (métastases).

3. Les différents types de cancer et différentes classifications

Les pathologies cancéreuses sont d'abord classées selon la localisation du début de leur prolifération. Une fois la première cellule maligne formée et divisée, formant une tumeur primaire constituée de cellules clonales, l'organe est attaqué. La classification se fait selon le

type de tissu attaqué ou selon l'organe affecté. Nous pouvons distinguer dans la première catégorie les types de cancers nommés comme suit :

- Les cancers hématopoïétiques : qui touchent les cellules sanguines
- Les sarcomes : qui évoluent dans les tissus conjonctifs (les os)
- Les carcinomes : qui prolifèrent dans l'épithélium (surface composée de cellules uniquement)

La classification des cancers selon l'organe touché est la plus connue. Les plus fréquents sont :

- Le cancer du poumon
- Le cancer du sein
- Le cancer du colon
- Le cancer du foie
- Le cancer du pancréas
- Le cancer du sang ou leucémie
- Le cancer de la moelle osseuse ou myélome multiple
- Le cancer des vaisseaux sanguins ou sarcome de Kaposi
- Lymphomes maladie de Hodgkin et lymphomes non-hodgkinien ….

4. Les étapes de la transformation cellulaire tumorale

Comme expliqué plus haut, la maladie est la conséquence d'un ou plusieurs dysfonctionnement(s) cellulaire(s) tel(s) que :

- Une perte de contrôle du cycle cellulaire
- Insensibilité à l'apoptose
- Irrégularité dans la réparation de l'ADN

Le cancer est considéré comme une pathologie multigénique (plusieurs gènes sont altérés), contrairement aux maladies monogéniques comme la mucoviscidose, certaines hémophilies et les myopathies. Cette altération peut toucher de 10 à 20 gènes de manière successive, non-aléatoire et pouvant s'étaler sur une longue période (5 à 20 ans).

Chaque type de cancer possède son propre mécanisme d'évolution de la maladie. A titre d'exemple, dans le cancer du colon, le phénomène débute par l'inactivation du gène suppresseur de tumeur APC, la mutation de l'oncogène Ha-ras et l'inactivation du gène suppresseur de tumeur P53. Cette spécificité et cette chronologie bien précise de chaque type de cancer, montre que la diversité génétique est la base de l'hétérogénéité de la pathologie cancéreuse.

L'évolution de la maladie du cancer comporte deux étapes successives : la cancérogenèse et la tumorogenèse [3].

4.1. La cancérogenèse

Il existe des agents carcinogènes entraînant des altérations dans la séquence de l'ADN. À ce stade là, le cancer n'altère pas encore les fonctions de l'organisme et n'est cliniquement pas observable. Ce phénomène précurseur est appelé carcinogenèse, généralement observé dans les cancers dérivant des tissus épithéliaux. De manière générale, l'exposition à un agent carcinogène n'est pas un phénomène irréversible, en terme d'évolution de la maladie. Les agents carcinogènes peuvent être désactivés et/ou éliminés par le biais des agents enzymatiques qui assurent leur métabolisme. S'ils ne sont pas expulsés, s'ensuit alors l'étape de l'**initiation** du processus tumoral. Conséquence directe de la carcinogenèse, cette étape est causée par l'accumulation des anomalies génétiques, en particulier les mutations des gènes maintenant l'intégrité et les caractéristiques des différents types cellulaires. Il en résulte l'activation des oncogènes, responsables de la multiplication cellulaire, et l'inhibition des gènes suppresseurs de tumeurs (P53) responsables de la mort cellulaire programmée. Les cellules résultant de l'ensemble de ces altérations génétiques sont dites « cellules initiées » et ne sont pas encore cancéreuses [4]. Leur prolifération, via d'autres agents carcinogènes, plus ou moins contrôlée, marque le début de la seconde étape de la cancérogenèse, la **promotion**. Au cours de cette multiplication, d'autres mutations sont engendrées et transmises aux cellules filles des cellules initiées. Le passage de la maladie au stade clinique est déclenché par l'instabilité accrue et l'accumulation des anomalies génétiques au niveau des composantes de la cellule qui apportent nutriments et soutien à leur développement (le *stroma*).

4.2. La tumorogenèse

Il s'agit de la progression des cellules précancéreuses d'abord localement, puis allant en grossissant. Le stade clinique de tumeur est alors atteint. Leur délocalisation par voies lymphatiques ou veineuses au niveau des tissus avoisinants donne lieu à des métastases. La circulation sanguine joue un rôle catalyseur dans l'invasion cellulaire ou l'angiogénèse. En effet, les cellules tumorales ayant besoin d'un milieu favorisant leur croissance et leur déplacement, acquièrent un pouvoir d'angiogénèse (système sanguin nourricier) par stimulation des cellules endothéliales des capillaires voisins. L'acquisition de cette capacité à susciter l'angiogénèse des cellules cancéreuses est la réaction cellulaire normale face à l'hypoxie ou le résultat d'une mutation génétique.

Certaines tumeurs primaires peuvent progresser vers un envahissement plus global de l'organisme par échappement de cellules tumorales issues de cette tumeur primaire : on parle alors de métastase.

Les métastases, qui sont l'évolution et la mobilité des cellules cancéreuses vers un ou plusieurs lieux, sont de 2 types :

- Métastases ganglionnaires ou carcinomes : se déplacent par les canaux lymphatiques, arrivées au niveau du ganglion, elles peuvent :
  - Être détruites suite à une inflammation

- Rester latentes sans activité directe
- Se fixer aux ganglions et s'y multiplier, entraînant une grosseur cliniquement palpable
- Traverser le ganglion se dirigeant vers d'autres canaux lymphatiques
- Se faufiler dans tout le circuit lymphatique, signe évident de l'agressivité de la tumeur

➢ Métastases sanguines ou sarcomes : pénètrent les vaisseaux sanguins soit de manière directe ou par le biais des canaux lymphatiques. Elles infiltrent préférentiellement les organes où passe un gros volume de sang (foie, poumons... etc.) ainsi que les os (métastases osseuses).

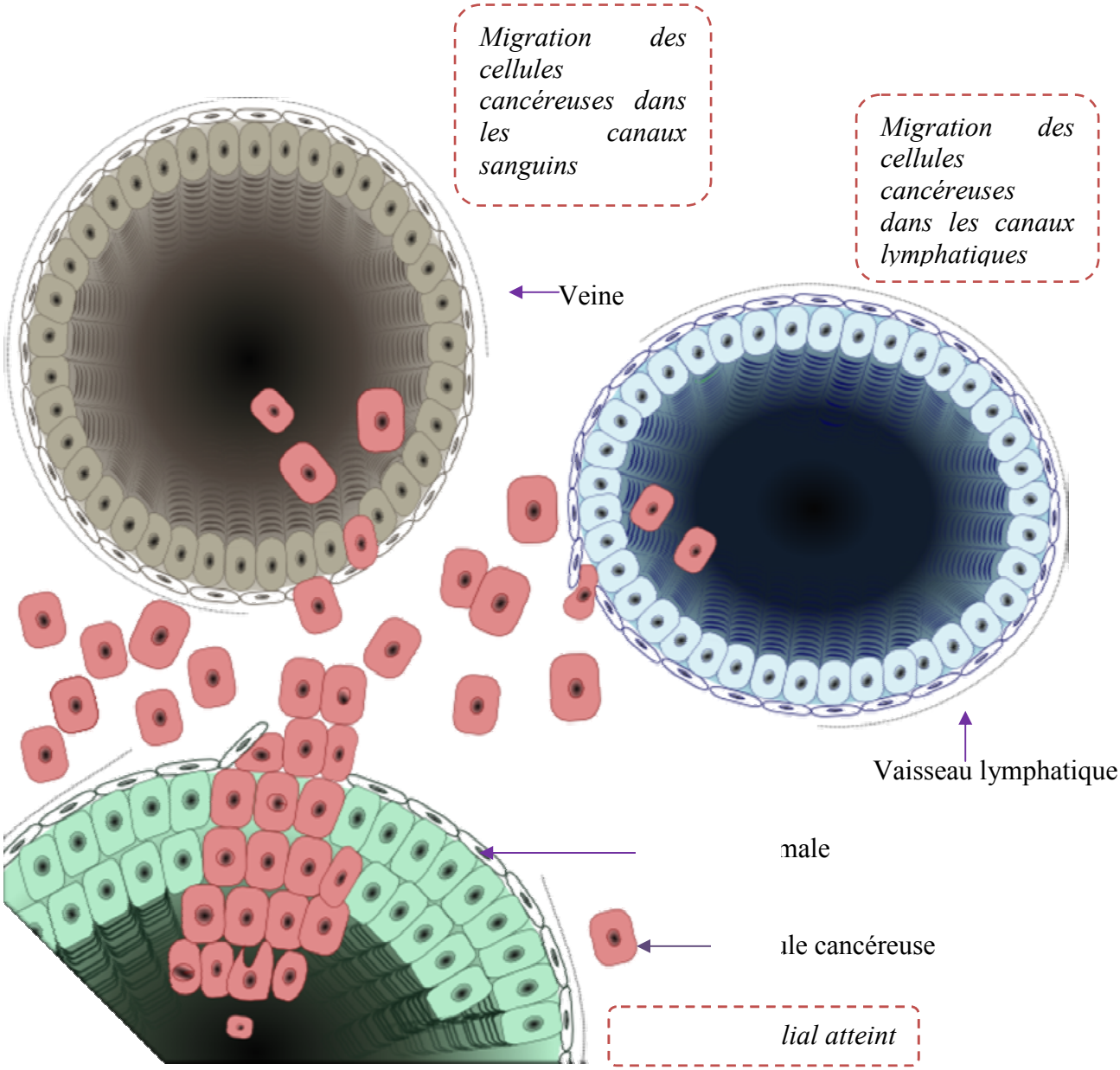


FIG. 2: Progression et invasion des cellules cancéreuses (Schéma représentatif)

Le cancer humain n'étant pas un corps étranger, son évolution ainsi que la prolifération en métastases dépendent essentiellement du type de cancer et de sa prise en charge. Certaines métastases dépistées au début de leur prolifération sont très sensibles aux traitements et permettent une rémission complète. D'autres sont, par contre, très difficilement maîtrisables et entraînent le décès à court terme.

5. Traitements et avancées thérapeutiques de nos jours

Les éléments cancérigènes (carcinogènes ou encore cancérogènes) sont des facteurs prédisposant à la survenue de la maladie dans un organisme. Ils peuvent être aussi bien endogènes (risque génétique) qu'exogènes (risque environnemental). La meilleure prévention reste la non-exposition à ces facteurs à risques ainsi que le contrôle régulier à partir d'un certain âge afin de dépister la maladie avant qu'elle ne s'installe.

Plusieurs nouvelles disciplines, telle que l'oncologie, ont vu le jour avec l'évolution et les découvertes liées à cette pathologie. Selon la gravité du cancer, il existe différents traitements. L'ablation de la tumeur reste la plus radicale [5]. S'ensuivent des séries de chimiothérapie et radiothérapie. L'une utilise les médicaments afin d'anéantir les métastases résiduelles, l'autre les radiations visant à provoquer d'importantes lésions dans l'ADN des cellules cancéreuses.

La biothérapie, regroupant l'hormonothérapie et l'immunothérapie, connaît un essor depuis le début des années 1980 et se trouve en évolution constante. En effet, l'hormonothérapie est utilisée depuis près d'un siècle dans les cas de cancers hormonodépendants comme le cancer des ovaires ou de la prostate. Les progrès issus de l'évolution de la biologie moléculaire et de l'immunologie ont contribué massivement au développement de l'immunothérapie [3]. Basée sur l'idée d'augmenter l'action du système immunitaire, les développeurs de cette thérapie se penchent de moins en moins sur la recherche d'un vaccin préparé à partir de cellules cancéreuses, par rapport aux débuts de l'immunothérapie. Les lymphocytes (cellules immunitaires obtenues par greffe de moelle osseuse) sont également employés comme stimulants non spécifiques de même que d'autres substances produites par la cellule, telles que les interleukines. A la fin des années 1980, les avancées en thérapie génique ont permis la manipulation des gènes afin de mieux équiper les lymphocytes contre les cellules cancéreuses. Il a été également possible de manipuler les cellules immunitaires purifiées, pour leurs interactions avec les cellules tumorales.

D'autres sous domaines de traitements ont vu le jour depuis et, plus les mystères des mécanismes de voies de signalisation s'éclaircissent, plus les nouvelles thérapies apparaissent. C'est ainsi que les nouvelles médications cherchent à rééduquer les cellules de manière à corriger leur comportement anarchique. D'abord prescrits pour les problèmes de peau, les rétinoïdes sont considérés comme un nouveau traitement contre la récidive du cancer. Cette

classe thérapeutique a pour particularité d'atteindre facilement ses récepteurs dans le noyau cellulaire, traversant ainsi la membrane cellulaire, grâce à leur propriété de solubilité dans les graisses (fonction caractéristique de la famille des alcaloïdes). La fixation de la molécule rétinoïde sur son récepteur induit la fixation de ce dernier sur un bout d'ADN permettant la transcription de gènes impliqués dans les mécanismes de croissance cellulaire. Des résultats ont été obtenus dans les cas de leucémie impliquant directement les rétinoïdes avec le blocage du processus de maturation de la cellule leucémique.

Nous porterons une attention toute particulière aux traitements visant à inhiber la transduction du signal cellulaire. En effet, une cascade de signaux de transduction est provoquée par l'activation de certains récepteurs cellulaires aux facteurs de croissance. Ces signaux consistent essentiellement en la croissance des cellules tumorales mettant en route le processus métastasique, activant ainsi l'angiogénèse. Les enzymes de types Tyrosine Kinase sont celles qui interviennent dans ces signaux de transduction. L'un des médicaments ayant fait ses preuves contre la Leucémie Myéloïde Chronique (LMC), en inhibant la protéine activée par le gène Bcr-Abl, est le STI 571, commercialement connu sous le nom du GLEEVEC® (en Amérique du nord) ou GLIVEC® (en Europe) depuis 2001. Issu de la modélisation moléculaire *in silico*, cet inhibiteur agit en tant que compétiteur intracellulaire de l'ATP au sein de la poche de liaison du récepteur de facteur de croissance [6, 7].

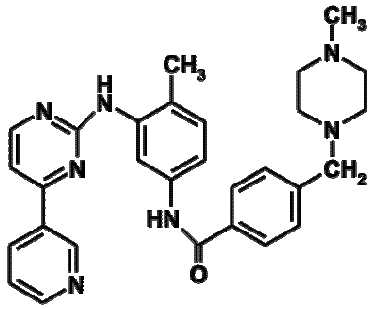


FIG. 3: Structure chimique de l'Imatinib ST571 ou Glivec

Dû à la complexité du mécanisme et des agents responsables des proliférations cancéreuses, l'aboutissement à la découverte d'un médicament se trouve être une tâche ardue. Encore une fois, la pluridisciplinarité des recherches dans le domaine reste la clé pour avancer et obtenir des résultats.

6. Cheminement du progrès thérapeutique

6.1. Les étapes de développement du médicament

La recherche d'une molécule médicamenteuse ressemble en beaucoup de points à la quête du Saint Graal tant le chemin à parcourir est souvent long, fastidieux et incertain. En effet, le processus de recherche et développement, dans l'industrie pharmaceutique, n'est couronné de succès que pour environ une molécule sur 10 000 testées. Ce processus dure en moyenne une dizaine, voire une quinzaine d'années.

Intéressons-nous donc au parcours du médicament depuis la potentielle molécule thérapeutique jusqu'à sa mise sur le marché.

Les 2 volets principaux de développement du médicament sont les études, ou encore essais, précliniques et cliniques [3, 8].

6.1.1. Les études précliniques

Cette phase concerne toutes les recherches d'une molécule originale répondant à une demande curative contre une ou plusieurs pathologies. Issue des milieux académiques ou de départements de recherche et développement industriels (R&D), la molécule est sensée avoir le moins d'effets secondaires. Pour trouver la molécule thérapeutique, les chercheurs peuvent faire appel aussi bien aux outils de la modélisation informatique (modélisation moléculaire, criblage virtuel) qu'aux tests expérimentaux (criblage réel, tests *in vivo* et *in vitro*). Ces tests de type chimique et biologique sont réalisés dans le but de s'assurer de la sécurité, la sélectivité et l'efficacité des molécules synthétisées. Les tests sur les animaux, *in vivo*, permettent une analyse du métabolisme, de l'absorption et de la distribution du médicament sur l'ensemble des constituants de l'organisme. Ce sont des tests, essentiellement toxicologiques, effectués pour la prédiction d'effets secondaires et qui vont permettre la détermination des dosages adéquats pour les essais cliniques, a posteriori, sur l'Homme. D'un point de vue légal, les tests *in vivo* doivent être effectués sur trois espèces animales distinctes avant de passer au stade clinique. Cependant, des méthodes alternatives ont vu le jour et les modèles animaux sont remplacés par des cellules humaines dérivant de différents tissus et organes. Ces essais *in vitro* ont permis, de nos jours, d'épargner des vies animales.

6.1.2. Les études cliniques

Après avoir passé les tests précliniques, un ensemble d'évaluations s'ensuit, ayant pour but de démontrer le caractère inoffensif ainsi que l'efficacité du futur médicament sur l'Homme.

6.1.2.1. *Phase I*

Les tests sont effectués sur des volontaires sains, sauf pour des molécules anticancéreuses et antibiotiques[1]. Cette première phase est aussi appelée phase d'innocuité. Elle permet de prouver la sécurité de l'emploi de cette molécule sur l'Homme et de déterminer la dose minimale active et la dose maximale tolérable. En d'autres termes, l'intérêt se porte sur les propriétés pharmacocinétiques ADME[2] de la molécule.

6.1.2.2. *Phase II*

Appelée également « tests de biodisponibilité ou d'efficacité ». Réalisée sur des groupes de patients malades, cette phase permet l'ajustement des dosages à prescrire en fonction des réponses obtenues des patients. C'est effectivement à partir des résultats des tests effectués au cours de cette phase, que sera déterminée la posologie du traitement pour telle ou telle indication. Cette phase permet également l'évaluation des premiers effets secondaires, qui, une fois confirmés ultérieurement, deviendront les effets secondaires principaux du futur médicament. Dans le cas où les effets secondaires sont trop importants, en comparaison de l'effet thérapeutique du médicament, le développement du traitement est définitivement suspendu.

6.1.2.3. *Phase III*

Ce sont des tests servant à évaluer l'intérêt clinique réel du médicament. Le but de cette étape étant d'établir le rapport bénéfices/risques. Au cours de cette phase, un grand nombre de patients est soumis, soit à la molécule contenant le principe actif, soit à un placebo. L'expérimentation est qualifiée de « double aveugle ». Ni l'administrateur (médecin), ni le patient ne connait le type de traitement donné. Les résultats sont ensuite comparés avec ceux de traitements déjà existants, ce qui devrait confirmer l'efficacité du médicament.

6.1.2.4. *Phase IV*

Il s'agit d'essais cliniques après commercialisation de la substance thérapeutique. Un dossier rapportant les résultats des 3 premières phases est déposé pour l'obtention d'une Autorisation de Mise sur le Marché (AMM). Si l'autorité chargée d'évaluer le dossier (l'Afssaps en France) estime qu'il est suffisant, le médicament est autorisé à la commercialisation. Le cas

[1] Pour des raisons d'éthique, l'exposition des patients sains aux traitements anticancéreux ou aux antibiotiques n'est pas autorisée car les futurs médicaments sont très actifs. Dans ces cas là, les volontaires passent directement à la phase II.

[2] A = vitesse d'Absorption ou de passage dans le sang à partir d'une solution orale ; D = la vitesse de Distribution et de répartition dans les différents tissus à partir du compartiment plasmatique ; M = la vitesse de Métabolisation (transformation biologique par le foie et d'autres organes) ; E = la vitesse d'Elimination de la molécule par l'organisme (« clearance »)

échéant, un complément d'information est demandé dans un délai assez court, 1 à 2 ans, sans empêcher sa mise sur le marché. En résumé, cette dernière phase consiste à relever les éventuels effets secondaires apparaissant plus tard dans la prise du médicament. C'est une phase connue comme étant celle de surveillance « post marketing » ou de « pharmacovigilance ».

6.1.3. Recherche des formes galéniques adaptées au traitement

Effectuées en parallèle aux tests cliniques, des recherches sont effectuées afin de découvrir la forme la plus appropriée de la molécule pour la préservation de l'activité ainsi que pour sa conservation (meilleures conditions de stockage)

L'administration du médicament peut se faire par différentes voies. Ainsi, les formes galéniques les plus connues sont : les injections, les comprimés effervescents ou pas, les crèmes, les sirops, les suppositoires, les inhalateurs, les sprays, les patchs, etc.

Depuis près de 200 ans, les médicaments apportent continuellement leur lot de progrès thérapeutiques. Des innovations majeures permettent la guérison de maladies jusqu'alors incurables, faisant gagner aux médicaments en simplicité d'utilisation, en sécurité et en efficacité. Cette inventivité croissante et progressive est le résultat de travaux pluridisciplinaires très rigoureux et d'investissements colossaux. Ces efforts ne sont pas tout le temps récompensés (en moyenne, sur 10 molécules, 3 pourront rentabiliser les étapes nécessaires à leur développement). Néanmoins, les résultats de ces recherches soutenues se traduisent en bénéfices, souvent majeurs, pour les patients, ouvrant constamment les voies vers de nouvelles perspectives thérapeutiques.

D’une manière plus synthétique, nous présentons un schéma sur le parcours de développement du médicament, du début de sa conception jusqu’à sa disponibilité dans les étalages des pharmacies de quartier.

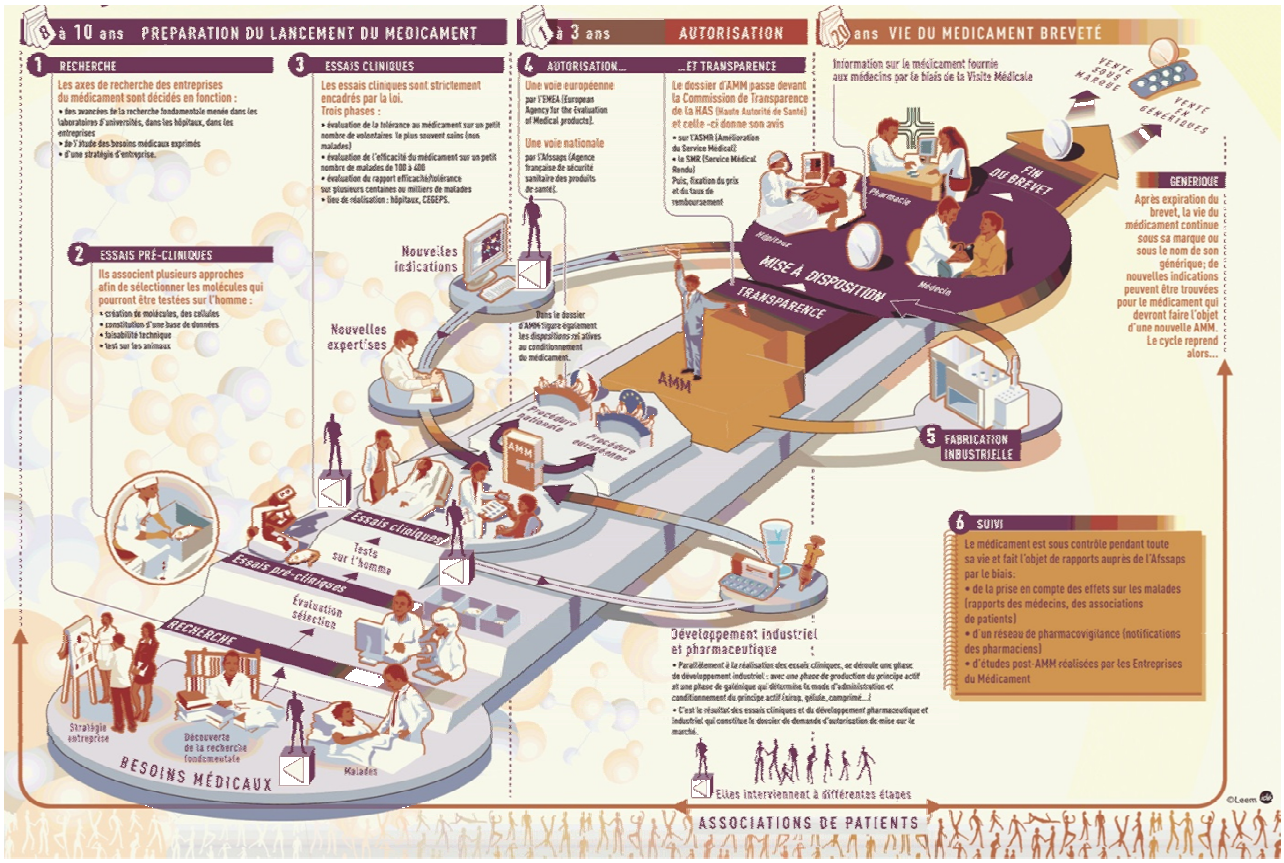


FIG. 4: L'aventure du médicament [9]

6.2.Coût de l'innovation thérapeutique

Un autre aspect non négligeable du développement de solutions thérapeutiques est l'argent mis en jeu. En effet, d'après le dernier rapport des comptes nationaux de la santé datant de septembre 2010, le montant des dépenses courantes de santé a atteint 223,1 milliards d'euros, pour l'ensemble des Français, dont les quatre cinquièmes représentent la consommation de soins et de biens médicaux. Cette dernière consommation de 175,7 milliards d'euros se traduit par une dépense moyenne en soins et biens médicaux équivalente à 2724 euros par habitant en 2009, dont plus de 500 euros consacrés aux médicaments.

En amont de tout cela viennent les investissements de plus en plus importants dans les secteurs de la recherche et du développement du médicament. En 2008, c'est environ 26,6 milliards d'euros qui ont été investis en Europe contre 24,2 milliards d'euros aux Etats-Unis. Une étude comparative a été effectuée sur les dépenses des trois plus grands ensembles de pays investisseurs dans le domaine de la recherche pharmaceutique [9].

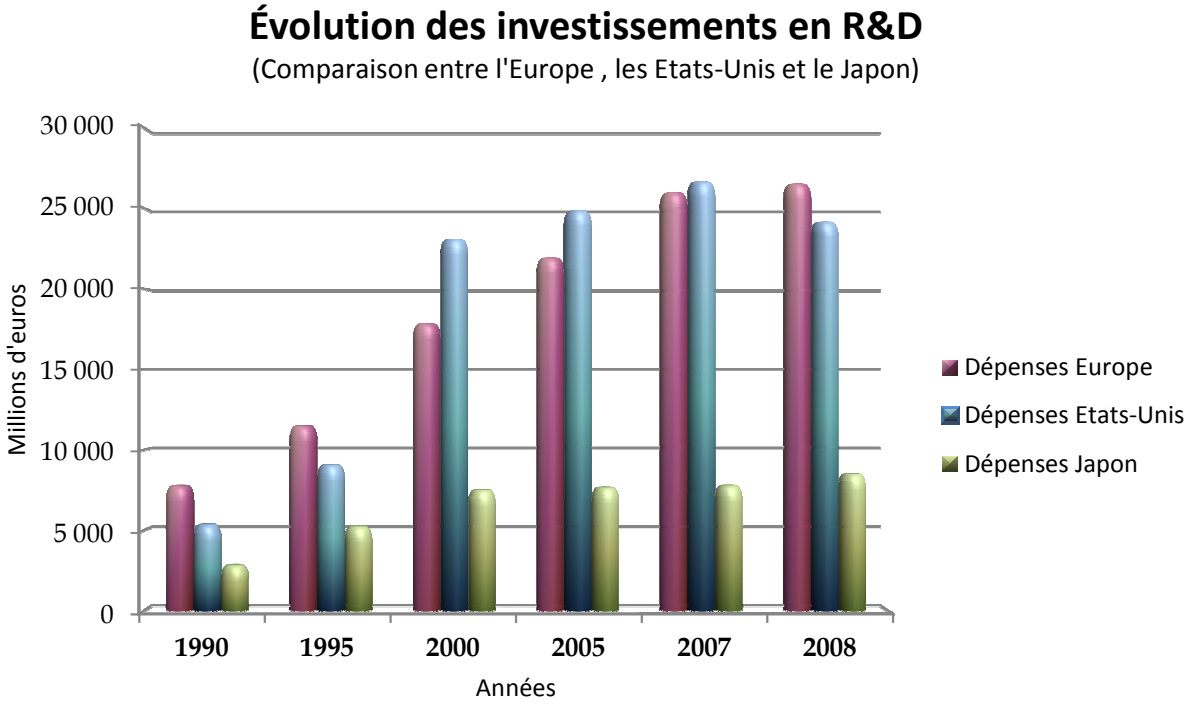


FIG. 5: Évolution et Comparaison internationale des coûts en R&D (d'après données [9])

Le graphique montre bien une croissance mondiale d'investissements dans les institutions de recherche et développement (R&D) des firmes pharmaceutiques qui a explosé durant la dernière décennie. À titre illustratif, durant l'année 2006, 956 autorisations de mise sur le marché ont été délivrées en France [10]. Cependant, malgré le fort taux d'investissement dans la recherche et développement des médicaments, le nombre de molécules innovantes, quant à lui, tend à diminuer. Aux Etats-Unis, par exemple, depuis 1994 une moyenne de 20 nouvelles molécules voient le jour par an [11]. Le fort taux de croissance des dépenses en R&D est justifié par le développement de molécules dites « me too » (moi-aussi). C'est une nouvelle classe thérapeutique où les molécules dérivent de médicaments existant déjà sur le marché ou présentées sous une nouvelle forme. Les modifications sont alors minimes et visent à améliorer la tolérance ou l'efficacité du médicament ou encore à changer le mode d'administration, la forme ou le dosage, ce qui demande une part d'innovation dans les traitements thérapeutiques.

Les avancées dans le domaine des découvertes des structures tridimensionnelles des protéines, de la résolution du séquençage du génome humain ainsi que dans le développement des connaissances des maladies et tout leur système de signalisation ont contribué à la croissance des bases de données des molécules à vertus thérapeutiques. Ce progrès constant des moyens utilisés en R&D du secteur pharmaceutique, contribue à rendre ce domaine vorace en argent et en temps d'investissement. Pour toutes ces raisons, l'avènement des approches *in silico* dans le domaine de la R&D apparaît comme une option de plus en plus prisée, dans le milieu pharmaceutique en particulier.

7. Méthodes *in silico*, méthodes alternatives

L'engouement connu pour les approches basées sur l'informatique depuis les années 1990 ne fait que croître et plus spécialement dans le milieu scientifique. Généralement utilisée comme un outil conduisant à l'élaboration de modèles, l'approche *in silico* permet un gain en temps et en argent quand elle est utilisée en amont du processus de recherche et développement de traitements thérapeutiques.

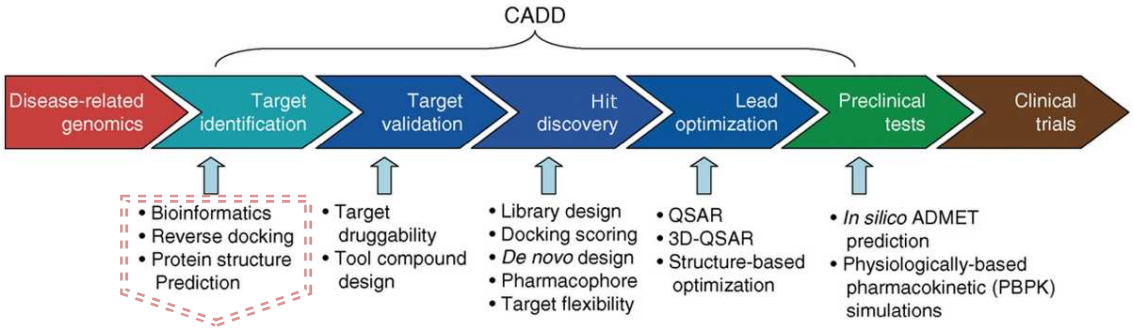


FIG. 6: Intervention des méthodes *in silico* dans le processus de découvertes de médicaments.([12] modifié)

Les méthodes dites « structure-based drug design », sont des associations d'approches visant la production de résultats scientifiques, qui ne sont pas nécessairement basées sur l'expérimental. Elles sont plutôt basées sur les simulations numériques représentant au mieux la réalité de systèmes biologiques complexes. Profitant des avancées matérielles de l'informatique telles que les puissances des machines ou la pertinence des programmes sans cesse mis à jour, les modélisateurs tentent de résoudre les problèmes liés à l'étude de systèmes biomoléculaires difficilement accessibles expérimentalement. Qu'il s'agisse de conditions de température ou de pression laborieusement voire péniblement atteintes expérimentalement, les méthodes de simulations numériques permettent d'outrepasser les obstacles et d'écarter (repousser) virtuellement des contraintes physiques. La contrainte de cette discipline réside dans le temps de calcul nécessaire, qui dépend forcément de la nature et de la complexité du système étudié.

La surabondance des données dans le domaine de la recherche scientifique a induit son association à l'informatique. Les disciplines découlant de la bioinformatique aident énormément dans l'organisation, la gestion et la sauvegarde du déluge de données qui ne fait que croître. C'est en effet dans ce goulot d'étranglement, provoqué par la profusion des données, qu'interviennent des approches telles que le Criblage Virtuel à Haut Débit (CVHD). Venu alléger son homologue réel (Criblage Réel à Haut Débit, CRHD) des contraintes temporelles et budgétaires, il repose néanmoins sur les mêmes principes. En effet, les deux approches ont pour but la recherche de cibles thérapeutiques. A l'aide de programmes et d'outils informatiques, l'approche du CVHD parcourt un ensemble de molécules regroupées dans des chimiothèques ou bases de données chimiques en les testant sur une cible donnée de

façon à sélectionner à la fin un ensemble restreint de molécules, appelées « hits » ou touches. Les outils informatiques utilisés pour cela découlent du regroupement des connaissances en biologie moléculaire, en chimie combinatoire et en modélisation moléculaire. L'ordinateur prend donc une place non négligeable dans le processus de l'innovation thérapeutique, faisant des méthodes *in silico* un substitut partiel de la partie expérience du processus. En restreignant l'espace chimique des molécules susceptibles de devenir des médicaments, le CVHD s'est fait une place dans les premières étapes du processus de R&D de molécules innovantes [13-16]. Le développement de têtes de séries dans les anticancéreux n'est pas en manque d'exemples [17-23], le Glivec® étant l'exemple le plus illustrant et le plus connu [7].

L'apport des méthodes *in silico* consiste à diminuer de façon notable la durée nécessaire à la mise sur le marché d'un médicament. En effet, les phases précliniques et cliniques ont tendance à devenir de plus en plus rigoureuses et la compression du temps accordé n'est pas envisageable. D'où l'optimisation possible en amont de ces tests, ie, en introduisant les approches *in silico* en général, et le CVHD en particulier.

8. Descriptif et organisation des travaux réalisés

Mes travaux de thèse ont débuté dans le cadre d'un projet, soutenu par l'INCa, intitulé « Synthetic and molecular modelling chemistry intercalate with biology for treatment of c-Met-triggered tumour growth and metastasis" et piloté par Flavio Maina de l'IBDML (Developmental Biology Institute of Marseille-Luminy). Ces travaux ont également été soutenus par la Ligue Nationale de la lutte Contre le Cancer (LNCC) et la Fondation de la Recherche Médicale (FRM).

Cette thèse s'articule directement autour de la cible biologique de nos recherches, c-Met. Ainsi, les travaux de recherche présentés dans ce manuscrit de thèse visent essentiellement à :

- Approfondir les connaissances du proto-oncogène MET et plus particulièrement de sa partie tyrosine kinase c-Met. L'expression inappropriée du couple ligand endogène-récepteur membranaire HGF/c-Met a été décrite dans de nombreux types de tumeurs humaines, la suractivation de c-Met aboutissant à un processus tumoral associé à des métastases. En d'autres termes, l'expression de c-Met est souvent associée à un mauvais pronostic.
- Rechercher des molécules originales et inhibitrices du domaine kinase de c-Met en utilisant les méthodes de modélisation moléculaire, de simulation de dynamique moléculaire, d'exploration de la flexibilité de la protéine et de criblage virtuel.

La première partie de ce manuscrit présente un descriptif détaillé de la cible de notre intérêt, c-Met. Nous présenterons également dans ce premier chapitre un état de l'art sur les différents inhibiteurs c-Met déjà connus ou encore en phase d'essais cliniques. En effet, au début de cette étude, une seule structure tridimensionnelle expérimentale était disponible pour ce domaine kinase. Cette structure présente le complexe d'un site actif impliqué dans la propagation du signal Met, couplé à un inhibiteur. Elle est obtenue par cristallographie aux Rayons X et son code PDB est : 1R0P. À ce jour, nous en sommes à 24 structures cristallographiques présentées sur le site de Protein Data Base (PDB). Nous conclurons cette partie par la nécessité de modéliser ces structures et d'explorer leur flexibilité par les approches *in silico* proposées, afin de réaliser un criblage virtuel le plus pertinent possible.

Dans la deuxième partie de ce manuscrit nous détaillerons les méthodes et matériels utilisés pour effectuer nos recherches. Notre exploration de l'espace conformationnel du récepteur c- Met s'appuie sur les techniques de Dynamique Moléculaire et de Modes Normaux de vibration. Ces méthodes ont comme support l'outil informatique, et ce chapitre passera en revue les notions théoriques de base de ces approches *in silico* utilisées.

Le troisième chapitre est consacré à la présentation des résultats de nos travaux de recherche rapportés sous forme d'articles publiés ou soumis. Nous développons les points essentiels dans une introduction et conclusion avant et après chaque article.

Nous terminons ce manuscrit de thèse par une conclusion de l'ensemble des travaux et résultats obtenus ainsi que la présentation des perspectives issues du travail achevé. Le bilan démontrera la nécessité d'opter pour un consensus entre différentes méthodes afin d'explorer exhaustivement l'espace conformationnel de notre système biologique, surtout si ce dernier est doté d'une grande flexibilité moléculaire.

PARTIE 2

PROTEINES TYROSINE KINASES ET C-MET

I- Les récepteurs Tyrosine Kinase (RTK)

Dans le règne animal, la plupart des phosphorylations de protéines eucaryotes sont régies par la superfamille des « Protéines Kinases » (PK) qui partage un domaine catalytique conservé. En effet, de l'ensemble des gènes eucaryote, il existe 1,5 à 2,5 % codant à la famille des protéines kinases [24]. Il s'agit ainsi de l'une des plus larges familles de protéines.

1. Classification des PK

De manière générale, la classification des protéines kinases peut s'effectuer selon leurs structures (leur séquence) ou en fonction de leur emplacement dans la cellule. La classification la plus connue est celle de Hank & Hunter [25] qui a été mise à jour avec la découverte d'autres PK. En effet, le nombre de classes est passé de 6 à 9. Classées en fonction des résidus phosphorylés au cours de la catalyse, les 9 familles appartiennent à 3 grandes classes. À cette classification s'ajoute une autre classe regroupant le reste des kinases :

1.1. **Classe des « Serine/Thréonine » protéines kinases** : met en jeu 2 amino-acides Serine (S) et Thréonine (T) au cours de la catalyse. Comporte les familles suivantes :

1.1.1. ***La famille AGC*** : composée de 63 membres, elle englobe des protéines cycliques dépendantes de l'AMPc et du GMPc.

1.1.2. ***La famille CaMK*** : composée de 74 membres dépendants de Ca^{2+} et de la calmoduline.

1.1.3. ***La famille CK1*** : formée de 12 membres dépendants de la Protéine Kinase C (PKC).

1.1.4. ***La famille CMGC*** : composée de 61 membres comprenant, entre autres, les PK activées par des mitogènes (MAP Kinase), les PK dépendants de la cycline, les PK « Cdk-like ».

1.1.5. ***La famille des RCG*** : possède 5 membres dont la Rhodopsine Kinase ainsi que les Kinases du récepteur β-adrénergique.

1.1.6. ***La famille des TKL*** : composée de 43 membres comportant les kinases tyrosine kinases-like.

1.2. **Classe des « Serine/Thréonine/Tyrosine » protéines kinases** : met en jeu 3 amino-acides Serine (S), Thréonine (T) et Tyrosine (Y) au cours de la catalyse (phosphorylation de ces 3 AA)

1.2.1. ***La famille STE*** : formée de 47 membres, elle inclut les Ste20 kinase et MAPK kinase

1.3. **Classe des « Tyrosine » protéines kinases** : met en jeu 2 amino-acides de Tyrosine (Y) au cours de la catalyse.

1.3.1. ***La famille TK*** : composée de 90 membres. Elle comporte aussi bien les Tyrosine Kinases cytosoliques, telles que, src, fgr, abl ainsi que les Récepteurs à activité Tyrosine Kinase tels que le facteur de croissance épidermique (EGF) et le facteur de croissance plaquettaire (PDGF)

1.3.2. **La famille des « Autres protéines kinases »** comprend 83 membres non classés qui n'appartiennent pas aux familles précédentes

L'illustration des 7 familles essentielles, d'après les travaux de Manning & al. [26], au sein du groupe des PKs, est représentée ci-dessous par ce qui est connu comme le kinome humain.

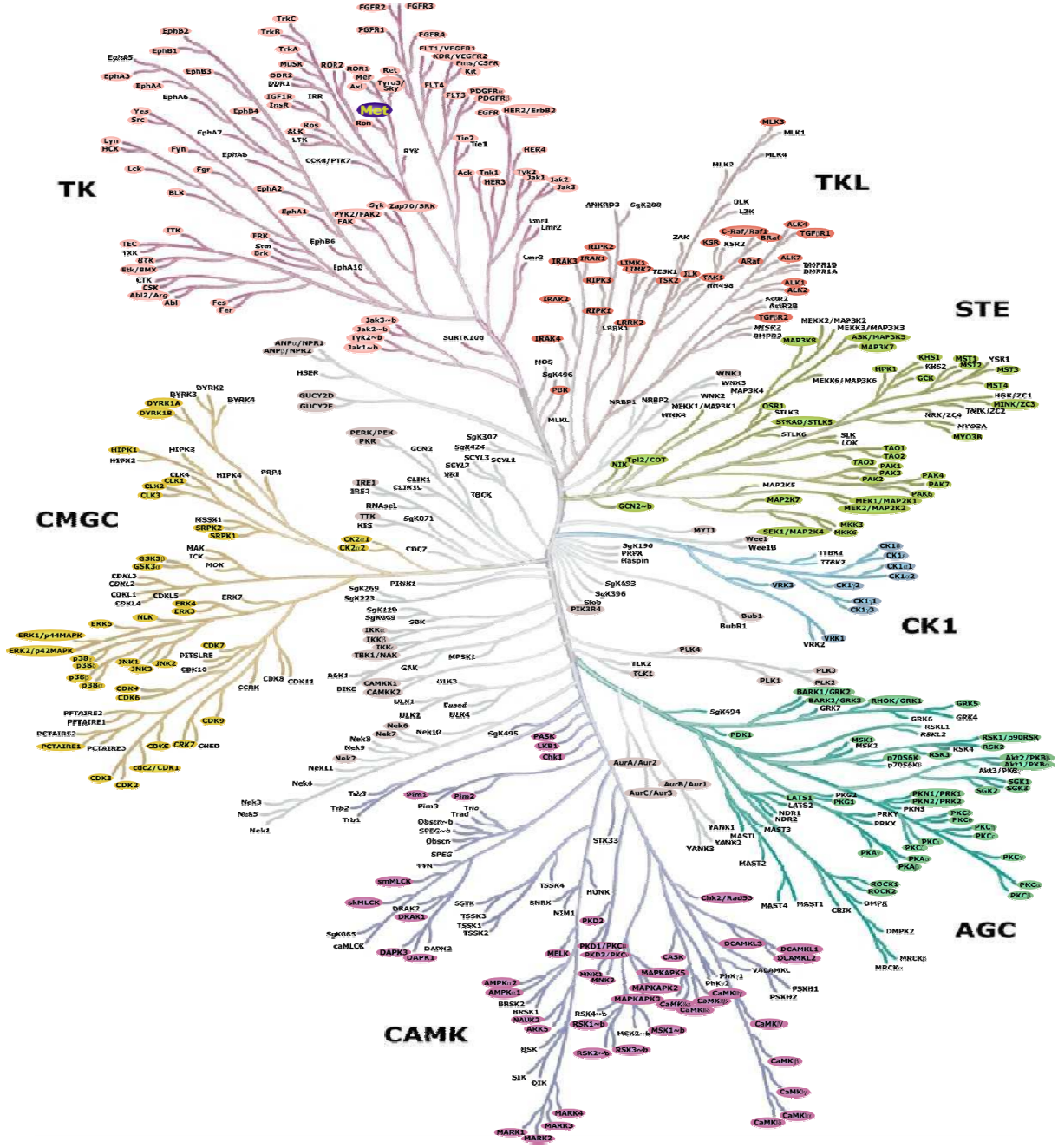


FIG. 7: Représentation du kinome humain [26] (Modifié de www.cellsignal.com)

2. Lumière sur la famille des Tyrosine Kinase TK

Notre intérêt se porte essentiellement sur la famille des Tyrosine Kinases « TK ». Cette famille tire son nom du fait que la phosphorylation des protéines cibles se fait sur des résidus de type Tyrosine grâce à la catalyse d'une protéine membre de cette famille. Les protéines appartenant à la famille des TK se subdivisent en 2 groupes par rapport à leur localisation dans la cellule :

2.1. Les PTK cytosoliques (CTK)

Appelées également les Non-Récepteurs Tyrosine Kinase (nRTK). Cette classification concerne les protéines évoluant dans le cytosol de la cellule (phase liquide du cytoplasme), plus précisément, au niveau de la surface interne de la membrane plasmique. Cette subdivision qui fut découverte avant les Récepteurs Tyrosine Kinases (RTK) comporte les 10 sous-familles suivantes : ABL, ACK, CSK, FAK, FES, FRK, JAK, SRC, SYK et TEC. Ces tyrosines sont impliquées dans les interactions de type protéine-protéine, protéine-lipide et protéine-ADN. Elles jouent un rôle majeur dans les voies de signalisation et doivent se trouver capables de détecter la phosphorylation d'une autre protéine en amont ainsi que son changement conformationnel. Les domaines de ces protéines activant les cascades signalétiques intercellulaires se présentent comme :

2.1.1. ***Le domaine SH2*** : constitué d'une centaine d'amino-acides, il possède 2 poches de fixation édifiées en structure secondaire de feuillets β. La première poche est profonde et possède une arginine conservée formant l'interaction avec la tyrosine phosphorylée. L'autre poche en aval de la première se trouve être spécifique de la protéine phosphorylée. En d'autres termes, ce domaine est capable d'identifier les séquences peptidiques possédant une tyrosine phosphorylée.

2.1.2. ***Le domaine SH3*** : constitué de 3 sous-sites, S0, S1 et S2, comprenant près de 60 acides aminés, ce domaine affiche une spécificité pour des peptides possédant des prolines. Afin d'établir une interaction, le sous-site S2 fixe une arginine près de la région contenant les prolines.

2.1.3. ***Le domaine PH (Pleckstrin Homology)*** : afin d'encadrer les protéines vers la membrane plasmique, ce domaine permet la fixation aux lipides phosphatidylinositol.

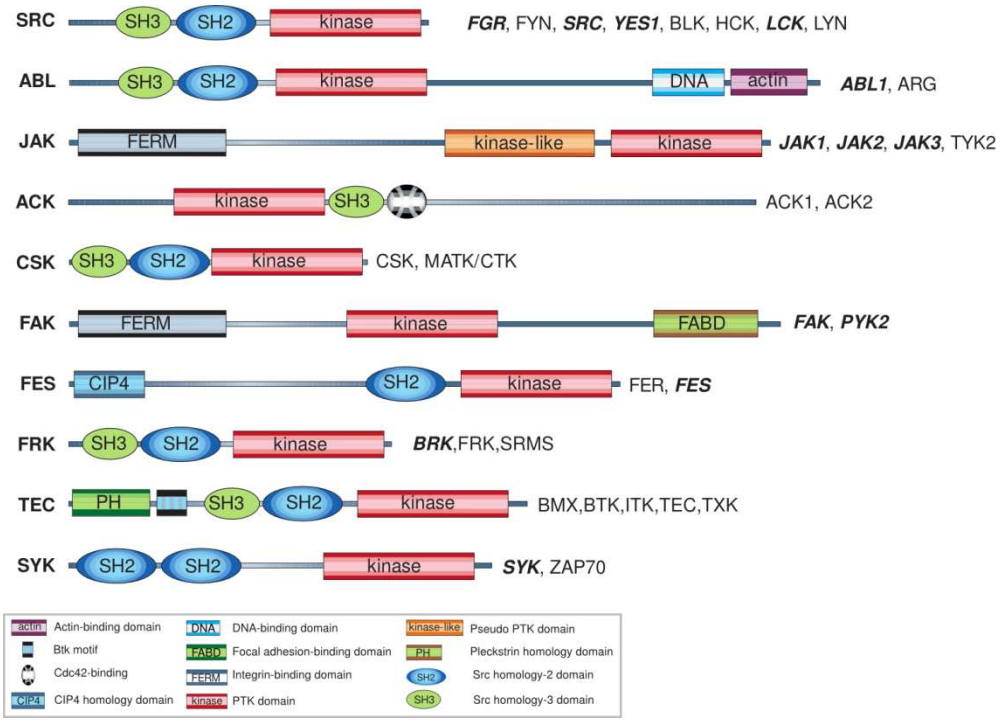


FIG. 8: Représentation des familles de PTK cytosolique ou nRTK. Les membres « en gras » sont impliqués dans le développement de tumeurs malignes (adapté de Blume-Jensen & Hunter [27])

2.2. Les PK membranaires ou Récepteur Tyrosine Kinase (RTK)

Cette dénomination concerne les protéines se trouvant dans la membrane cellulaire. Egalement connues pour leurs propriétés de récepteurs-enzymes, elles jouent un rôle majeur dans les phénomènes de signalisation et transduction cellulaire. [28]. Pour plus de précisions, les RTKs sont des glycoprotéines qui se divisent en 3 sous domaines :

2.2.1. ***Domaine extracellulaire*** : lieu de fixation de l'endogène ou du ligand

2.2.2. ***Domaine transmembranaire*** : essentiellement formé d'hélices α et complètement hydrophobe, il permet au récepteur de s'ancrer dans la membrane plasmique

2.2.3. ***Domaine intracellulaire*** : qui regroupe une région juxta-membranaire, la région catalytique Tyrosine Kinase (TK) et un domaine C-terminal. En plus d'être nécessaire au renouvellement des récepteurs via le phénomène de l'endocytose[3], le domaine C-terminal peut renfermer des sites de liaisons pour des substrats.

Les différentes sous-familles des RTKs sont représentées dans le schéma qui suit.

[3] Phénomène de transport de ligands ou autres de l'extérieur vers l'intérieur de la cellule.

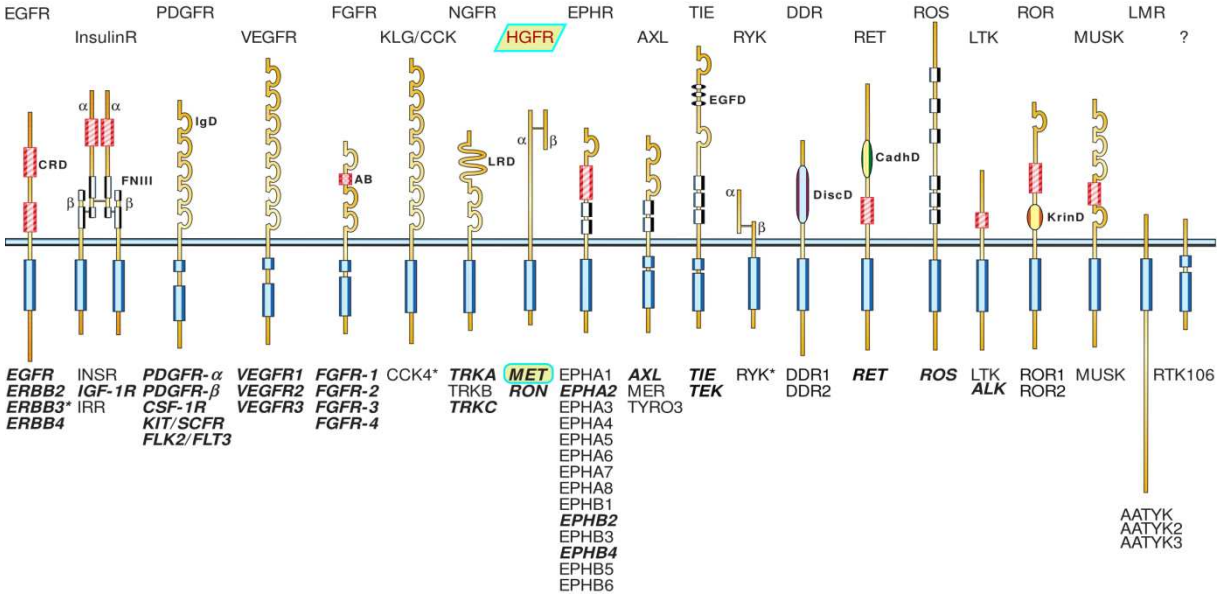


FIG. 9: Représentation des familles RTKs et de leurs membres. Les membres « en gras italique » sont impliqués dans le développement de tumeurs malignes. (Adapté de Blume-Jensen & Hunter [27])

Les Récepteurs Tyrosine Kinases (RTK) sont une des plus grandes familles de protéines transmembranaires qui se lient avec des molécules à travers l'espace extracellulaire, tout en possédant un domaine d'activité tyrosine kinase hautement conservé. Considérés comme des régulateurs des fonctions cellulaires, les 518 RTKs humains dénombrés contrôlent l'activité des protéines en catalysant leur liaison au groupement phosphate (chargé négativement).

D'une manière très générale, nous pouvons dire que, structuralement, les Protéines Tyrosine Kinases (PTK) sont constituées de 3 domaines liés mais bien distincts. Le domaine extra cellulaire et transmembranaire donne la caractéristique de « récepteur » à la protéine en question. Le domaine SH2 et SH3, qui peut interagir avec les différents composants de la cellule, régule l'activité catalytique de la kinase. Enfin le domaine catalytique conservé est caractéristique des membres des différentes familles de protéines kinases.

2.3. Constituants du domaine catalytique des Kinases

Dans les études portant sur la connaissance de la flexibilité ou plasticité des protéines réceptrices, l'intérêt se porte essentiellement sur le domaine catalytique de ces dernières. Ce domaine, lieu de la catalyse, est composé de 250 à 300 acides aminés partagés entre 12 domaines assez conservés [29] (I à XII). Commun à l'ensemble des membres d'une même famille, le domaine est généralement composé des constituants suivants :

- Le lobe N-terminal ou « petit lobe » : contenant les sous domaines I à IV, il est essentiellement impliqué dans le positionnement et la liaison du substrat nucléotidique (ATP) dans le site catalytique. Il est majoritairement constitué des feuillets β ainsi que de l'hélice Cα.
- Le lobe C-terminal ou « petit lobe » : contenant les sous-domaines VI à XI, sièges du site de liaison à l'ion métal cofacteur (Mg^{2+}) et de la boucle kinase catalytique avec le motif DFG. Il est majoritairement constitué des hélices α de la protéine et des boucles d'activation. L'assemblage des acides aminés contenus dans ce lobe permet la reconnaissance du substrat et du positionnement du groupement hydroxyle des résidus tyrosine, sérine et thréonine du substrat.
- La région charnière : contenant le sous-domaine V, il s'agit de la boucle qui relie les 2 lobes constituant la kinase et se trouve être l'un des sites de liaison du complexe protéine/ligand. En effet, le site de liaison en question est connu comme le site de liaison ATP ou phosphoryl donneur. Il est situé dans la région charnière.

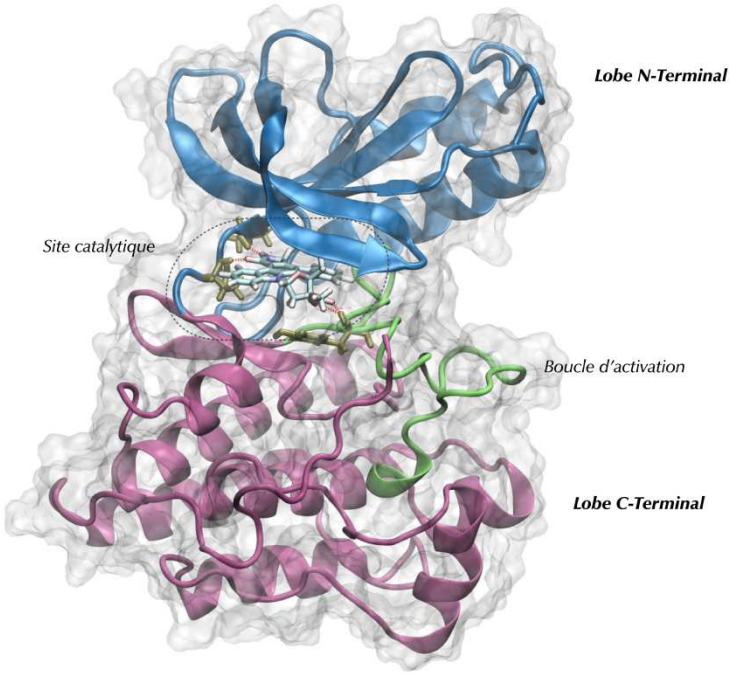


FIG. 10: Représentation d'une Kinase (complexe 1R0P/KSA) sous sa conformation active (motif DFG-in). Le ligand se niche dans la poche catalytique située entre les 2 lobes composant la kinase. Le petit lobe (N-Ter) est représenté en bleu, le grand (C-Ter) en mauve et la boucle d'activation (boucle A) en vert.

2.4. Les mécanismes d'activation des kinases

L'activité de la majorité des kinases dépend de la phosphorylation de leur boucle d'activation par le biais d'une succession d'événements bien définis. Il n'y a que l'origine du déclenchement de cette activation, conduisant à un changement conformationnel de la protéine, qui varie d'une kinase à une autre. Parmi les événements déclencheurs, on peut citer l'interaction de la kinase avec un ligand (exemple de fixation de l'insuline par son récepteur) ou l'interaction de la kinase avec un autre partenaire protéique.

Suite à la phosphorylation de la boucle d'activation et du changement conformationnel de la protéine, les 2 lobes (N et C terminaux) se rapprochent. Cette conformation permet alors un bon alignement d'acides aminés essentiels à la catalyse, un positionnement approprié de l'ATP et une exposition du site de reconnaissance du substrat. L'ATP maintenu dans la poche du petit lobe adjacente à la boucle d'activation, se trouve stabilisé par l'ion magnésium (Mg^{2+}, appelé également agent cofacteur). Il aide à la libération du γ-P induisant la phosphorylation spécifique de quelques résidus de la boucle catalytique. Cette boucle adopte ainsi une conformation tournée vers l'extérieur de la protéine, prête à accueillir le substrat, appelé DFG-in, et mettant ainsi la kinase sous sa forme active. On dit, selon l'appartenance de la PK à une famille ou une autre, que la kinase est capable soit de phosphoryler des substrats ou ligands spécifiques (phosphorylation en *trans* ou intermoléculaire) ou bien de s'autophosphoryler (phosphorylation en *cis* ou intramoléculaire)

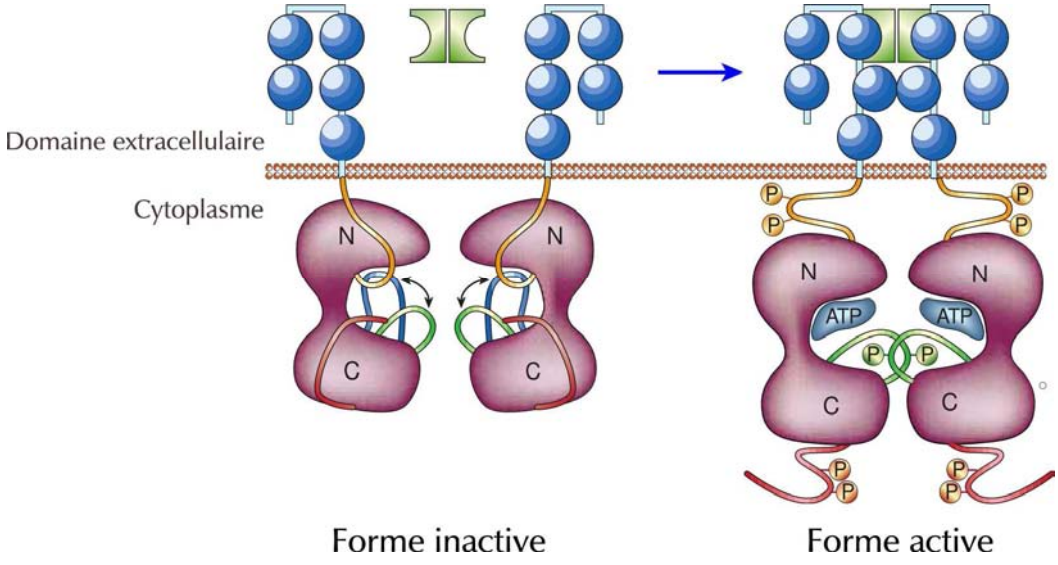


FIG. 11: Représentation du mécanisme d'activation des RTKs.

Le changement de conformation de la kinase induit par la fixation du ligand (vert) favorise le positionnement de l'ATP dans le site catalytique. L'ATP permet la phosphorylation aussi bien de la boucle d'activation (en vert) rendant la kinase active, que la phosphorylation de la boucle coté C-terminal (en rouge) favorisant ainsi l'interaction avec d'autres partenaires (protéines) impliqués dans les voies de signalisation. (adapté de Blume-Jensen et Hunter [27])

Les RTKs catalysent le transfert d'un groupe phosphate de l'ATP à un substrat donné. En effet, l'activité enzymatique de ces protéines membranaires est effective lors de la fixation du ligand conduisant à la formation de dimères, provoquant ainsi la phosphorylation d'une tyrosine (autophosphorylation). Cette activité peut être aussi exercée sur des substrats, de type protéine dans le cytoplasme, qui transmettent à leur tour le signal de phosphorylation en cascade au sein de la cellule.

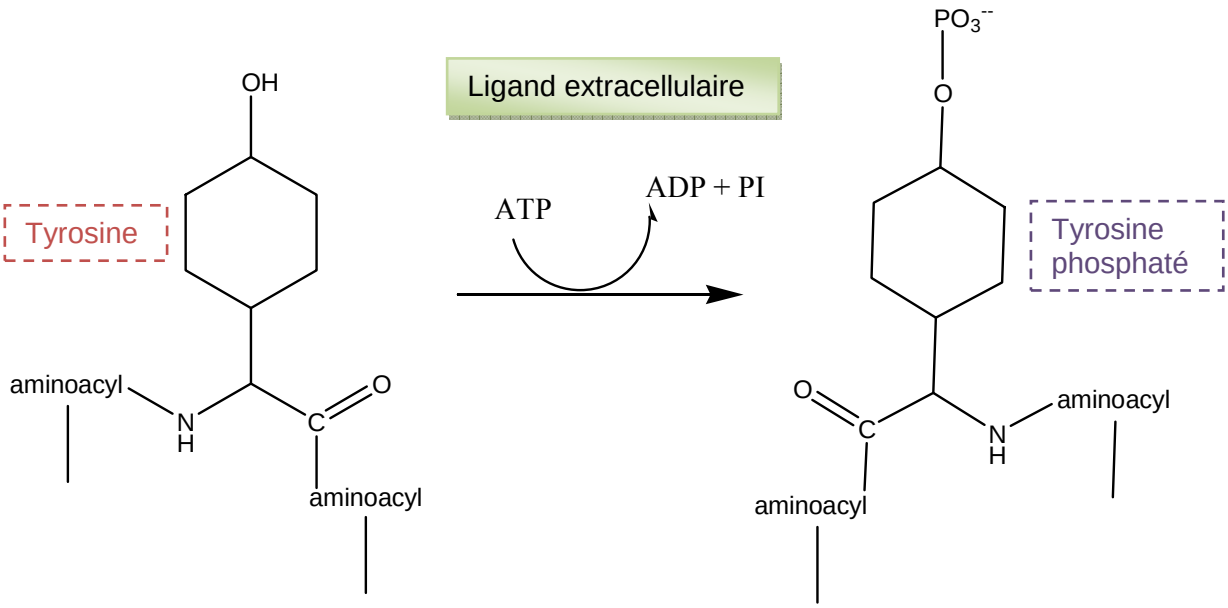


FIG. 12: Phosphorylation d'une tyrosine (liée à une protéine)

Les Tyrosine Kinases intéressantes pour les traitements anticancéreux ont des inhibiteurs qui interagissent essentiellement avec le site de liaison ATP. Notre intérêt se portera essentiellement sur ces couples de complexes. Tel est le cas entre autres du Glivec®.

2.5.Les voies de signalisation médiées par les RTKs

Les RTKs modulent une large variété de processus biologiques, plus spécialement les signaux envoyés de la membrane cellulaire vers des cibles intracellulaires. Ils entrent également en jeu dans la coordination de fonctions biologiques complexes telle que la croissance cellulaire.

Les signaux intracellulaires sont ciblés par l'activation du domaine tyrosine kinase à travers la phosphorylation de multiples substrats. Les sous-familles des RTKs sont catégorisées par la similarité des amino-acides constituant leur séquence et leur structure extracellulaire. C'est pour cette raison que les membres d'une même sous-famille se lient de la même manière à certains ligands et induisent donc une cascade de signalisations complexes

dont la compréhension se simplifie au gré de l'évolution des nouvelles méthodes de modélisation [30].

D'une manière générale, nous pouvons décrire sommairement les deux principales voies de signalisation transduites par la majorité des RTKs. Une description plus détaillée est effectuée dans la partie appropriée dans les paragraphes consacrés à c-Met qui suivent.

- Voie RAS/MAPK : c'est une voie qui concerne directement la croissance et le développement cellulaire. La prolifération cellulaire qui en découle rend cette voie de signalisation capitale dans la compréhension des phénomènes liés au processus cancéreux dans le cas où son activation devient incontrôlable.
- Voie PI3K/PKB (ou AKT) : cette voie régit la plupart des phénomènes liés au métabolisme[4] tels que la synthèse de glycogène, d'acides gras ainsi que la translocation des agents transporteurs de glucose [31].

[4] Ensemble de processus de transformation biochimique qui se produisent dans les tissus des organismes vivants

II- c-Met : Enzyme d'intérêt

Après nous être intéressés aux mécanismes de base conduisant à la prolifération cancéreuse au sein de l'organisme vivant, nous allons détailler les connaissances au sujet de la cible de nos recherches, à savoir l'enzyme « c-Met ». Les premières découvertes, au cours des années quatre-vingts, ont d'abord mis en lumière l'existence d'un agent mitogène, facteur de dispersion dans les cellules épithéliales, connu sous le nom de « HGF » (High Growth Factor). Ce facteur de croissance se retrouve vite lié aux pathologies cancéreuses car il stimule la perte d'adhésion cellulaire, augmente la mobilité et l'invasion des cellules cancéreuses. Ainsi, il s'agit d'un puissant agent angiogénique et lymphangio-génique. Les études issues de la littérature se sont également portées sur le récepteur de ce ligand endogène, qui est c-Met, au vu de la diversité fonctionnelle formée par ce complexe ligand/récepteur.

c-Met est donc considérée comme une cible médicamenteuse émergente à cause de l'inhibition de multiples mécanismes biologiques à travers sa liaison avec des ligands.

1. La structure de c-Met

Met est un proto-oncogène localisé sur le chromosome 7q31 qui encode pour un récepteur transmembranaire hétérodimère de type Tyrosine Kinase (RTK). Le proto-oncogène Met est issu du clivage protéolytique d'un précurseur de 170 kDa, et se compose d'une chaîne extracellulaire α de 50 kDa reliée par un pont disulfure à la chaîne β qui recouvre la membrane. Son poids moléculaire est égal à 145 kDa. Le segment extracellulaire du récepteur comporte un domaine semaphorine appelé « SEMA » ainsi qu'un domaine riche en cystéine nommé « Cys » caractéristique de Met. Le récepteur contient également 4 structures de type immunoglobuline-like nommées « domaine IPT », ainsi qu'une région d'interaction protéine-protéine typique.

Le domaine intracellulaire du récepteur Met est formé par une section juxta-membranaire, suivie du site catalytique et de la chaîne C-terminale essentielle pour la signalisation en cascade.

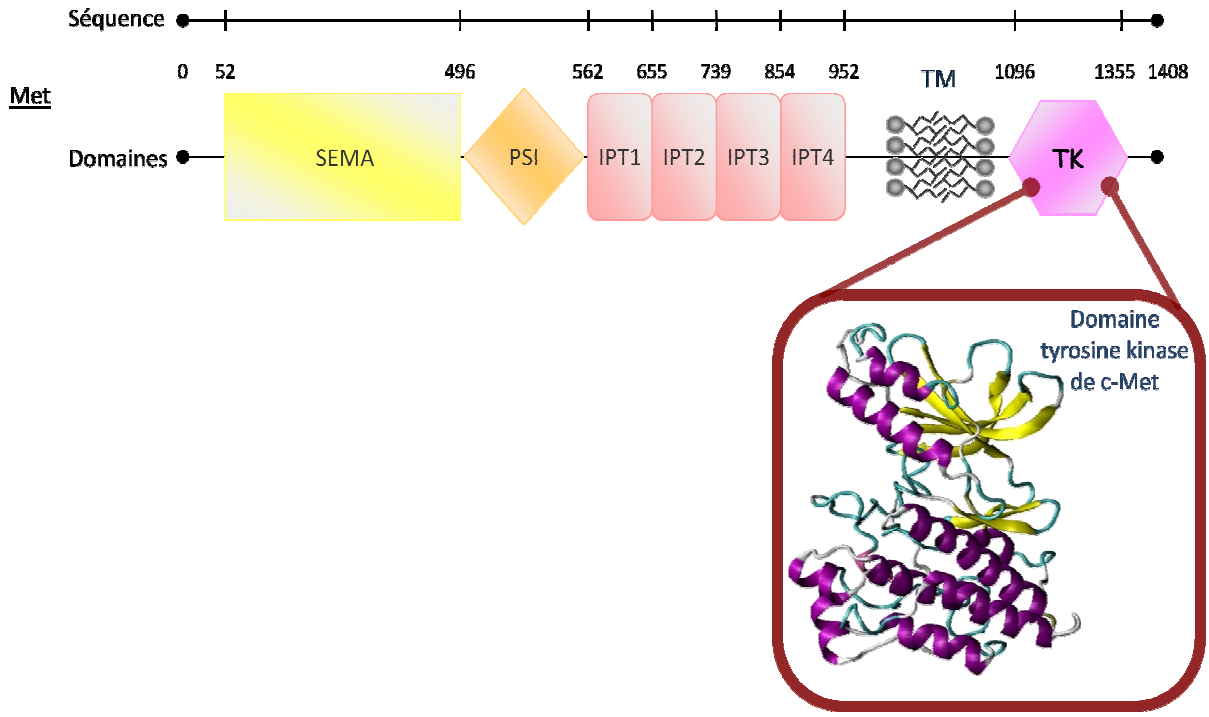


FIG. 13: Représentation du domaine de la protéine Met

En terme d'homologie de séquence, la protéine c-Met se trouve étroitement liée aux récepteurs : Mer, Ron, Axl (Ufo), RYK et Tyro3 (Sky). A ce jour, 26 structures cristallographiques de c-Met ont été publiées [32]. Ces structures ont été obtenues par les méthodes cristallographiques aux rayons X aussi bien sur la forme *apo* que sur la forme complexée avec des molécules inhibitrices du domaine kinase de c-Met.

Comme pour l'ensemble des RTKs, le segment juxta-membranaire du récepteur Met joue un rôle primordial dans la régulation de l'activité de ce dernier. Ce segment de la protéine contient une sérine, (Ser 975) qui inhibe l'activité de la tyrosine suite à la phosphorylation, ainsi qu'une Tyrosine (Tyr 1003) apte à se lier avec le proto oncogène CBL responsable de la dégradation de Met. Le site catalytique, quant à lui, contient bon nombre de Tyrosines qui sont phosphorylées suite à l'interaction du récepteur avec son ligand endogène HGF.

1.1. Le ligand endogène de c-Met

Le facteur de croissance des hépatocytes HGF (Hepatocyte Growth Factor) et le facteur de dispersion SF (Scatter Factor) ont été identifiés, indépendamment l'un de l'autre, respectivement comme inducteur de croissance potentielle des hépatocytes primaires dans les milieux de culture et comme facteur induisant la dispersion des cellules épithéliales. L'équipe de Naldini a démontré que HGF et SF sont des molécules identiques [33]. HGF appartient à la

famille des plasminogènes. Il est sécrété sous la forme d'une chaîne unique comme un précurseur biologiquement inerte qui atteint sa forme mature par protéases extracellulaires. La forme active de HGF est constituée de deux chaînes α et β reliées entre elles par un pont disulfure. La chaîne α contient 80 acides aminés sous forme d'une boucle N-terminale suivie de 4 domaines reliés par 3 ponts disulfure internes. La chaîne β, quant à elle, est homologue à la protéase responsable de la cascade de la coagulation sanguine mais manque d'activité protéolytique à cause des substitutions d'acides aminés dans le site catalytique.

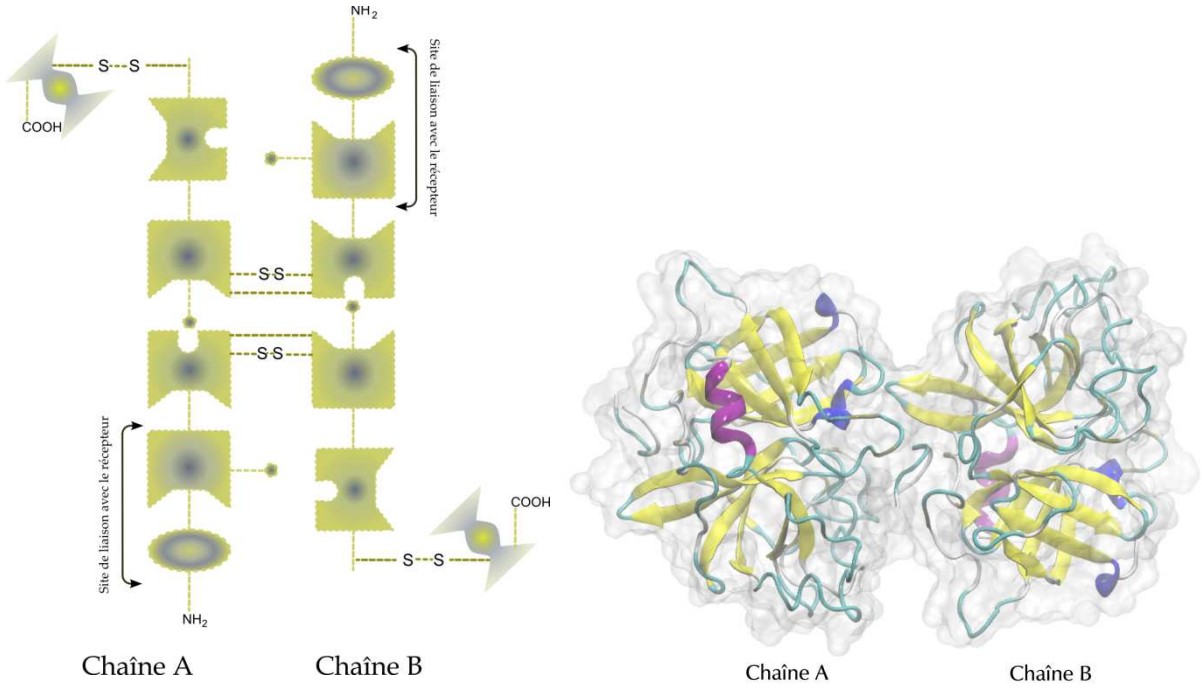


FIG. 14: Représentation schématique de l'homodimère HGF/SF (à gauche). Représentation de la structure secondaire 3D de l'homodimère HGF/SF (à droite). [Code pdb : 1YBW]

L'enzyme c-Met est d'abord exprimée dans les cellules épithéliales puis dans les cellules neuronales, mélanocytaires, hématopoïétiques, microgliales et endothéliales. Le récepteur c-Met et son ligand HGF jouent un rôle essentiel lors de la formation embryonnaire, le développement de la coordination musculaire, l'organisation du système nerveux, le remodelage osseux et l'angiogénèse des cellules. c-Met possède également une importante activité dans la prolifération et la migration cellulaire durant les réparations des lésions aiguës.

Notre intérêt se porte donc sur la partie de la protéine Met évoluant dans le cytoplasme de la cellule d'où elle tire son nom composé « c-Met »

L'activation de c-Met se fait à travers l'autophosphorylation des tyrosines Tyr1234 et Tyr1235 localisées dans la boucle d'activation (boucle A).

2. Action du complexe HGF/c-Met

HGF se lie au récepteur c-Met induisant l'homodimerisation du récepteur ainsi que la phosphorylation du domaine cytoplasmique tyrosine kinase. Ces étapes marquent le début de signalisations activant des réponses biomédicales. Celles-ci font directement appel à différentes protéines (Grb2, Gab1, ERK…) entrant dans les voies de signalisation essentiellement relatives à la prolifération et à l'apoptose cellulaire.

L'activation de c-Met par son ligand endogène HGF provoque la mutation du récepteur ainsi que l'avènement de phénomènes cellulaires métastasiques tels que la mobilité cellulaire, la résistance à l'apoptose et le changement dans la forme de la cellule. Tous ces changements se produisent via une cascade de modifications des voies de signalisation cellulaire.

La liaison HGF/c-Met conduit à l'homodimérisation du récepteur qui est un dimère sous sa forme non complexée. Cette liaison induit l'autophosphorylation des 2 tyrosines 1234 et 1235 dans la partie cytoplasmique de Met, qui à son tour enclenche une succession de réponses dans les voies de signalisation, détaillée comme suit :

- La protéine Gab1 interagit par 2 voies avec Crk et CrkL, 2 protéines qui vont à leur tour provoquer une cascade de signalisations conduisant à des modifications pendant l'étape de la transcription des protéines dans le noyau ainsi qu'à des changements du cytosquelette.
- Fak représente également une autre cible à double fonction de la liaison HGF/c-Met. L'activation de cette kinase lui procure une fonction de transmetteur conduisant à la formation d'adhésions focales, une étape préliminaire à l'invasion tissulaire et à la motilité des cellules transformées.
- L'activation de Grb2 à travers la formation du complexe HGF/c-Met conduit à la transmission du signal vers la protéine Ras. En effet, la mutation de ce transmetteur se trouve impliquée dans les voies de signalisation menant à la régulation et prolifération cellulaire ainsi que dans l'inhibition de l'apoptose [34].
- L'activité de Pi3k, suite à l'amarrage de HGF, agit sur les voies de signalisation dites AKT (dérivée des kinases AKT) qui bloquent l'apoptose.

Pour plus de clarté, les réponses en cascade ainsi que les phénomènes provoqués par la liaison HGF/c-Met sont représentés dans le schéma ci-dessous.

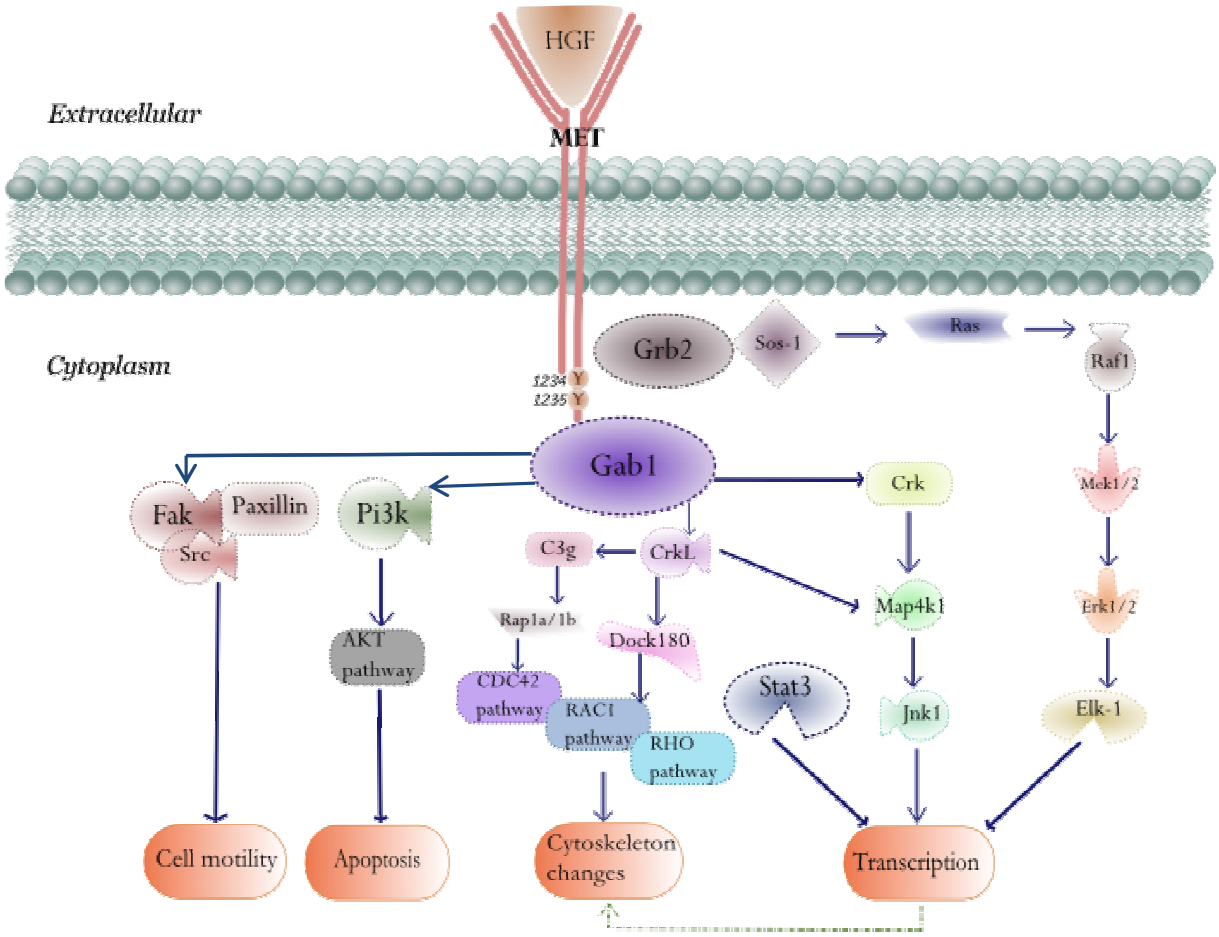


FIG. 15: Les voies de signalisations enclenchées par l'activation du complexe HGF/c-Met

3. c-Met et le développement du cancer

Le dérèglement des voies de signalisation incluant le récepteur c-Met a été observé dans bon nombre de développements de tumeurs malignes humaines. Les effets de l'activation soutenue de c-Met sont largement caractéristiques des modèles précliniques. Le récepteur peut être activé dans plusieurs types de cancers comme le résultat d'une mutation génétique avec ou sans l'amplification génique. Une promotion de l'invasion cellulaire est ainsi provoquée, de même que le déclenchement des métastases via la participation directe des voies de signalisation angiogénique.

Initialement, c-Met a été identifiée comme une protéine résultant d'un oncogène transformé, nommé TPR-Met (Translocated Promoter Region Met), et dérivant d'un réarrangement chromosomique dans une lignée de cellules d'ostéosarcome traitées avec un cancérogène chimique. Dans les modèles animaux, l'expression transgénique de TPR-Met conduit au développement de tumeurs mammaires et autres tumeurs malignes. Chez l'humain, le réarrangement de TPR-Met a été constaté dans les cancers gastriques.

L'un des mécanismes de dérèglement de c-Met les mieux identifiés est l'amplification génique de c-Met responsable de la surexpression de la protéine et de l'activation de son

domaine Kinase. L'amplification génique de c-Met a été observée dans les tumeurs primaires ou comme événement secondaire altérant la réceptivité des cellules cancéreuses aux traitements, même si ce n'est pas le cas pour la majorité des cancers. Cette amplification de l'expression de c-Met a été reportée dans plusieurs tumeurs primaires humaines telles que les carcinomes gastriques et ceux de l'œsophage ainsi que dans les cancers colorectaux [35]. De nombreuses études ont démontré aussi que la croissance du nombre de copies du gène Met se trouve très fortement associée à la présence de métastases au niveau du foie chez des patients atteints du cancer colorectal [36]. De ce fait, les dérèglements dans l'activation de c-Met conduisent à de nombreuses réponses biologiques associées à l'acquisition du phénotype malin (ou transformation cellulaire). Parmi les réponses dont il est question, nous pouvons citer : l'augmentation de la motilité, de l'invasion, de la tumorigénicité et de l'angiogénèse, ainsi que la protection des cellules carcinomateuses de l'apoptose induite par des agents endommageant[5] de l'ADN et l'augmentation de la réparation des ruptures de brins d'ADN. En d'autres termes, toutes ces réponses biologiques favorisent la survie des cellules tumorales déjà génétiquement endommagées, conduisant à une résistance aux traitements chimio- et radiothérapiques.

Les pathologies causées par les altérations génétiques de c-Met font de cette protéine une cible potentiellement pertinente, bien que la majorité des patients atteints par le cancer présentent rarement ce type d'altération de c-Met. La cause d'activation de c-Met la plus fréquente dans les tumeurs humaines est l'augmentation de l'expression de la protéine due à la régulation positive de la transcription en l'absence de l'aberration génétique. A titre d'exemple, l'identification de différents néoplasmes dus à des surexpressions de c-Met est en constante augmentation. Concernant les carcinomes, les cas de niveaux élevés d'expression de c-Met ont été détectés dans tous les types de cancers qui leur sont associés. Des études, dans le cas du cancer du poumon, ont démontré que la surexpression de c-Met apparaît dans plus de 60% des cas [37]. Il en est de même pour le mésothéliome pleural malin (cancer de la plèvre qui entoure les poumons et tapisse la cavité thoracique) et les carcinomes des cellules rénales, où des études ont montré que, dans plus de 80% des cas, la surexpression de c-Met est observée. De manière similaire, un taux élevé de c-Met a été décelé dans les cas de cancers du sein [38] et des ovaires [39].

4. c-Met et ses inhibiteurs

Cibler l'inhibition des protéines kinases est devenu une piste établie dans les thérapies anticancéreuses. Suivant le lancement fructueux de l'imatinib (Glivec ®), développé par Novartis dans la première partie de la décennie, plusieurs autres inhibiteurs de kinases ont été mis sur le marché, alors que d'autres sont encore en phase d'essais cliniques.

[5] Rayons UV, rayons ionisants ou l'adriamycine, médicaments anticancéreux souvent utilisés lors de la chimiothérapie, favorisant la formation des radicaux libres coupant l'ADN

La molécule imatinib a été développée, à l'origine, comme un traitement contre une certaine forme de leucémie chronique à travers l'inhibition des kinases Bcr-Abl. Il se trouve que l'imatinib est un inhibiteur potentiel de plusieurs structures kinases comme c-Kit et PDGFR par des mesures cliniques de son efficacité contre des tumeurs malignes dépendantes de ces kinases.

D'autres inhibiteurs de kinases commercialisés ciblent plusieurs kinases, tels que les molécules, sunitinib (Sutent®) de Pfizer, dasatinib (Spyrel®) de Bristol-Myers Squibb sorafenib (Nexavar®) de Bayer.

Les inhibiteurs ciblant plusieurs kinases à la fois semblent être une valeur thérapeutique, à cause des multiples anormalités dans les voix de signalisation provoquées par l'établissement, la croissance et les métastases tumorales. Cependant, ces inhibiteurs présentent un risque potentiel de toxicité, une réduction de l'inhibition ainsi qu'une émergence de résistance aux médicaments vis à vis d'autres cibles. L'optimisation des inhibiteurs multi-cibles reste un défi médico-chimique considérable, du au fait que ces inhibiteurs ne fournissent pas de validation pour des cibles uniques. Avec la progression de la compréhension des structures et modes de liaison des kinases, le développement d'inhibiteurs de kinases sélectif et à cible unique devient un objectif tout à fait réalisable.

Les approches de découverte des inhibiteurs à cible unique contre inhibiteurs multi-cible existent plus particulièrement pour la cible c-Met. Les tests cliniques des inhibiteurs sélectifs tels que SGX-523, ARQ-197 et PF-2341066, nécessitent une sélection minutieuse de profils de patients possédant des tumeurs suffisamment sensibles aux perturbations des voies de signalisation de c-Met. Les inhibiteurs qui ciblent plusieurs récepteurs reliés aux kinases, tels que XL-880 et MGCD-265, peuvent entraîner des complications associées au dosage limitant les effets secondaires dûs à l'inhibition du récepteur tyrosine kinase VEGFR.

4.1. Les inhibiteurs kinases et les mutations de c-Met

A ce jour, plus d'une vingtaine de mutations ont été effectuées sur c-Met [40]. Les mutations ponctuelles dans le site de liaison des kinases sont responsables de la résistance des molécules thérapeutiques, imatinib (Glivec ®) et gefitinib dans, respectivement, plusieurs cas de leucémies et de cancers du poumon non à petites cellules (CPNPC). Il a été démontré que l'inhibiteur SU-11274 affectait différemment l'activité des kinases et la signalisation des différentes formes mutantes de c-Met [41], à l'instar de AM7 qui inhibe l'ensemble des 5 formes mutantes de c-Met testées [42].

La divergence des modes de liaison de ces 2 inhibiteurs est la cause de la différence de leur profil d'activité, paramètre qui devrait être pris en compte pour la conception d'inhibiteurs. De plus, certains patients peuvent se révéler sensibles à d'autres agents ciblant la

voie de signalisation HGF/c-Met et montrer de la résistance vis-à-vis de certaines formes de traitements c-Met.

La majorité des inhibiteurs c-Met sont décrits comme des compétiteurs de l'ATP, à l'exception de ARQ-197. Ils se lient selon un mode compact ou dit en forme de « U » appelé Type I ou selon un mode plus ou moins étendu appelé Type II. Ces inhibiteurs sont susceptibles d'être plus sélectifs et montrent une résistance à la mutation du résidu Y1230.

L'accroissement des informations structurales sur c-Met, en fonction des publications des structures cristallographiques, peut conduire à l'exploitation d'autres poches de liaisons et autres types d'interactions générant ainsi d'autres inhibiteurs avec des profils optimisés et des structures alternatives. De même, le ciblage des conformations kinases inactives, des sites allostériques et des résidus de la région charnière fut une approche fructueuse dans la conception d'inhibiteurs sélectifs aux kinases. Bon nombre de ces inhibiteurs issus des stratégies discutées font encore l'objet d'essais cliniques et il serait intéressant de voir leur comportement dans le corps humain. Une table récapitulative de l'ensemble des petites molécules inhibitrices de c-Met publiées à ce jour est donnée en annexes (A).

Il existe différentes stratégies pour concevoir des inhibiteurs c-Met, par le biais de l'inhibition du récepteur ou par le biais de l'inhibition de la cascade de signalisation HGF/c-Met. Nous allons détailler ces 2 approches en citant les molécules thérapeutiques conçues pour.

4.1.1. Les antagonistes biologiques de HGF et c-Met

Ce sont des molécules créées pour prévenir les interactions entre le ligand et le récepteur mais incapables d'activer les voies de signalisation. HGF contient 2 sites de liaisons avec c-Met présentant différentes affinités pour le récepteur. Le site de liaisons à forte affinité reconnaît c-Met indépendamment du processus de maturation et transformation de HGF. Celui qui présente une faible affinité devient un site accessible suite à la conversion de HGF, devenant essentiel pour la dimérisation et l'activation de c-Met. En conséquence, un précursseur inactif de HGF, ou un fragment de HGF conservant uniquement le site de forte affinité, peut interagir avec le récepteur c-Met sans induire ni la signalisation ni la dimérisation du récepteur. D'un autre côté, le domaine SEMA de c-Met, qui est une portion extracellulaire, peut déplacer et remplacer HGF et atteindre sa dimérisation [43].

4.1.1.1. Analogues compétiteurs de HGF

Ces agonistes, répertoriés ci-dessous, sont en compétition avec le ligand (HGF) pour liaison avec le récepteur (c-Met) sans l'activer.

- NK2: un fragment naturel de HGF (contenant le domaine N et 2 domaines K) qui inhibe la mitogénèse épithéliale de HGF et la morphogénèse *in vitro*.
- NK4: forme tronquée de HGF contenant la chaîne α entière mais dépourvue de chaîne β. Son activité inhibitrice a été démontrée, *in vivo* et *in vitro*, sur plusieurs modèles. [44, 45].
- Une forme non transformable de HGF, bloquant la molécule dans sa conformation inactive, entre en compétition avec la forme active pour se lier à c-Met et l'activation de HGF. Elle a donc pour effet de bloquer la catalysation et l'activation de la maturation protéolytique de HGF.

4.1.1.2. Divers compétiteurs c-Met

- « Decoy » (ou leurre) de c-Met : molécule enzymatiquement inactive qui correspond à tout le domaine extracellulaire de c-Met. Elle interagit avec le ligand endogène HGF et avec c-Met en capturant le ligand, compromettant ainsi la dimérisation du récepteur natif.
- Le domaine isolé SEMA: inhibe de manière compétitive la liaison du ligand, la dimérisation du récepteur nuisant, de ce fait, à l'activation des signaux de transduction cellulaires dépendants de c-Met et réduisant la migration cellulaire provoquée par HGF.
- Les Anticorps contre c-Met: essentiellement des anticorps monoclonaux (Les anticorps monoclonaux sont des clones d'anticorps naturellement présents dans l'organisme, qui reconnaissent et ciblent des antigènes précis sans toutefois endommager les cellules saines; ils sont reproduits en laboratoire pour être utilisés contre des maladies)
  - DN30: cet anticorps possède la faculté d'induire le clivage protéolytique du domaine extracellulaire de c-Met. Ce mécanisme d'action diminue le nombre de molécules à la surface cellulaire se liant au récepteur, inhibe la liaison HGF-c-Met et la dimérisation de c-Met en conséquence.
  - OA-5D5: cet anticorps inhibe la croissance intracérébrale de HGF lorsqu'il est localement fourni.
- Les Anticorps contre HGF:
  - L2G7 : développé par la firme Galaxy Biotech
  - AMG102 : développé par la firme Amgen

4.1.2. Les petites molécules inhibitrices de c-Met

Elles sont capables d'entrer en compétition avec l'ATP dans le site actif du domaine tyrosine kinase de c-Met et possèdent un petit poids moléculaire. Leur liaison évite la

transactivation du récepteur et la mobilisation des protéines dans la cascade des voies de signalisation.

- K-252a: premier inhibiteur c-Met co-cristallisé, analogue de la staurosporine. Il dispose d'un large champ d'activité inhibitrice de la kinase à des concentrations submicromolaires. [46].
- SU11274/PHA-665752 (Pfizer): caractérisé par le groupement indolin-2-one dans sa structure, cet inhibiteur est très sélectif. Il bloque les réponses biochimiques et biologiques reliées à c-Met dans nombreuses lignées de cellules cancéreuses. SU11274 affiche également une inhibition sélective envers les mutants c-Met identifiés dans le carcinome papillaire rénal [41] alors que PHA-665752 est efficace dans l'inhibition des lignées cellulaires tumorales et les xénogreffes responsables de l'amplification du gène de c-Met [47]. Plusieurs molécules de cette série ont été testées *in vitro* et *in vivo* mais n'ont pas été considérées cliniquement viables à cause de la limite de leurs propriétés pharmacocinétiques, plus particulièrement leur faible biodisponibilité orale.
- PF-02341066 ou Crizotinib (Pfizer): composé oralement disponible, il cible les 2 kinases c-Met et ALK. Il se montre efficace à des doses bien tolérées. Il possède une activité anti-angiogénique, cytoréductrice et antitumorale dans plusieurs modèles néoplasiques comportant les formes activées de c-Met et ALK. L'inhibiteur est actuellement en phase d'essais cliniques I, II et III [48].
- ARQ-197 (ArQule): médicament qui n'est pas en compétition avec l'ATP, il a été identifié comme un inhibiteur sélectif de c-Met dans un criblage évaluant le profil biochimique de 230 kinases. Il est actuellement en phase d'essais cliniques II sur des patients avec des tumeurs solides avancées.
- XL880 (Exelixis): molécule qui inhibe plusieurs kinases dont : c-Met, VEGFR-2, PDGFR, Ron, Kit et Tie2 [49]. L'inhibiteur fait l'objet d'essais cliniques en phase II sur des patients atteints de cancers gastriques diffus peu différenciés, de carcinomes papillaires rénaux et de carcinomes épidermoïdes de la tête et du cou.
- XL-184 (Exelixis): inhibiteur de Tyrosine kinase potentiellement administré pour le traitement du cancer médullaire de la thyroïde, glioblastome multiforme et le cancer du poumon non à petites cellules ou *NSCLC* (cancer du poumon le plus courant). L'inhibiteur a pour principales cibles c-Met, VEGFR-2 et Ret il a été cependant noté à son actif, une activité inhibitrice envers les kinases Kit, FLT3 et Tek. L'inhibiteur présente une biodisponibilité orale et une perméabilité à la barrière hémato-encéphalique significative. XL-184 est en cours d'essai cliniques [50]:
    - phase I sur des patients atteints de tumeurs solides avancées

    - phase II sur des patients atteints de glioblastome récurant ou progressif et de cancer du poumon non à petites cellules
    - phase III sur des patients atteints du cancer médullaire de la thyroïde
- MGCD265 (MethyGene) : inhibiteur qui a pour cibles c-Met, VEGFR-1, VEGFR-2, VEGFR-3, Ron et Tie2. Il joue un rôle important dans le développement et la formation des vaisseaux sanguins (angiogénèse) et dans la survie des tumeurs, par conséquent. L'inhibiteur est actuellement en phase d'essais cliniques I pour le traitement des cancers à tumeurs solides et en phase II pour le traitement du cancer du poumon non à petites cellules et des tumeurs solides.
- MK2461 (Merck): cet inhibiteur est de type multicible et ATP compétiteur de la forme active de c-Met. Il est en cours d'essais cliniques en phase I/ II pour le traitement de patients avec des tumeurs solides avancées.
- MP470 (SuperGen) : petite molécule inhibitrice de c-Met , c-Kit et PDGFR à doses nanomolaires. Il a été démontré que MP470 sensibilise à la radio thérapie les lignées cellulaires de plusieurs glioblastomes multiformes *in vivo* et *in vitro* [51], améliorant ainsi les résultats des patients atteints. Cet inhibiteur est actuellement en phase d'essais cliniques I/II sur des tumeurs solides adultes indéterminées.
- SGX523 (SGX): inhibiteur de c-Met sélectif et considéré comme compétiteur de l'ATP, il possède une activité antitumorale *in vivo* [52]. En cours d'essais cliniques en phase I, son action se fait sur des tumeurs solides adultes indéterminées.
- JNJ-38877605 (Johnson & Johnson): petite molécule inhibitrice de c-Met, oralement biodisponible, elle possède une activité antinéoplasique (anticancéreuse) potentielle. Cet inhibiteur est en cours d'essais cliniques phase I sur des néoplasmes.

Pour conclure ce paragraphe, il est clair que l'activation aberrante du récepteur c-Met est retrouvée dans plusieurs types de cancers humains, et définit un comportement clinique singulier pour la population de patients atteints. De plus, les données pré-cliniques ainsi que les premières données cliniques soutiennent le caractère thérapeutique prometteur de la cible c-Met pour les patients ayant une activité tumorale liée à c-Met. Il est, de ce fait, important de continuer l'exploration des stratégies menant à l'inhibition du récepteur tyrosine kinase c-Met. Suite à l'accroissement du développement des inhibiteurs c-Met, des profils de patients et des scénarios cliniques seront définis par des données à venir afin de faire bénéficier les stratégies thérapeutiques élaborées pour l'inhibition de c-Met.

4.2. Détails des interactions des complexes c-Met cristallisés disponibles sur la Protein DataBank

La littérature des brevets concernant les petites molécules inhibitrices de c-Met a été examinée en 2007. Depuis, l'intérêt porté à ce domaine n'a fait que croître jusqu'à atteindre plus de 70 brevets publiés à la mi 2009, conduisant aussi à l'évolution des publications d'essais cliniques référant aux brevets.

La majorité des inhibiteurs de kinase développés forme des liaisons hydrogène avec les résidus se situant dans la région charnière du récepteur. Les molécules inhibitrices se lient au domaine kinase de c-Met aussi bien dans sa conformation DFG-in que DFG-out. Les ligands formant des liaisons DFG-in sont appelés ligands de Type I et définis comme des inhibiteurs compétiteurs de l'ATP. A l'inverse, les ligands qui se lient avec la conformation DFG-out de la protéine sont appelés ligands de Type II. En plus d'être en compétition avec l'ATP, ces inhibiteurs de Type II engagent la partie charnière de c-Met par des liaisons hydrogène et peuvent interagir avec des protéines situées au voisinage de c-Met durant le changement conformationnel. [53-55]

L'inhibiteur « staurosporine-like » (analogue à la staurosporine), K-252a avec le domaine kinase de c-Met sous sa forme complexée, fut la première structure cristallographique découverte avec une résolution de 1,8Å par l'équipe de Schiering [56]. Il en est de même de la découverte au cours de la même année de la forme *apo* ou libre de c-Met (1R1W). Depuis, 26 structures cristallographiques sont répertoriées sur le site officiel de la « Protein Data Base » (PDB) au moment de la rédaction de cette thèse [32].

Les structures cristallographiques provenant de la PDB ont toutes des éléments communs :

- Une structure générale formée de 2 lobes reliés par la partie charnière de la protéine. Le premier lobe, côté N-terminal, est composé majoritairement de feuillets β, alors que le deuxième qui se trouve du côté C-terminal est majoritairement formé d'hélices α.
- Un résidu « gardien » plus communément connu sous le nom « Gate keeper » (gk). Il s'agit de L1157 qui sera suivi par les résidus spécifiques à la région charnière : P1158-Y1159-M1160 (symbolisés dans la littérature par : gk+1 à gk+3)
- Une boucle d'activation qui spécifiera le type auquel appartient l'inhibiteur selon la conformation de son motif DFG.

L'exemple du complexe 1R0P/K-252a peut servir comme référence commune à la compréhension du mode de liaison c-Met/inhibiteurs.

La boucle d'activation du complexe K-252a/1R0P est ordonnée. Cependant elle ne se présente pas sous la conformation active car elle se positionne loin du site actif de l'ATP. Elle adopte plutôt une conformation auto-inhibitrice lorsqu'elle chevauche le site de liaison du

ligand. De plus, une conformation inhabituelle est observée sur la partie N-terminale de la boucle d'activation : le résidu D1222 du motif DFG n'est pas impliqué dans la formation d'un pont salin avec K1110 (résidu hautement conservé). C'est le résidu D1228 qui se trouve engagé dans cette interaction, provoquant ainsi le déplacement de la boucle d'activation vers l'hélice C et vers la boucle P.

K-252a forme 2 liaisons hydrogène avec la région charnière, et plus précisément avec le résidu M1160 (gk+3), à travers son groupement pyrrolocarbazole. L'indolocarbazole (K-252a) se trouve pris dans l'étroite fente formée par la partie N-terminale et C-terminale de la boucle P.

Cette inhabituelle conformation de la boucle d'activation n'est pas uniquement induite par la liaison du récepteur au ligand indolocarbazole ; elle est également présente dans la structure *apo* de la protéine (1R1W) et dans d'autres complexes avec différents ligands. Toutes les structures cristallographiques de c-Met, découvertes à ce jour, se présentent sous leur forme complexée ou libre. Il n'y a pas de structure de complexe connue avec les résidus Y1234 et Y1235 phosphorylés et la boucle d'activation se replie loin du site de liaison de l'ATP (l'état conformationnel « typiquement » actif) dans tous les cas reportés.

4.2.1. c-Met et la flexibilité ou détail des interactions c-Met/inhibiteurs cristallisés

Le domaine kinase de c-Met affiche une considérable plasticité. Celui-ci adopte différentes conformations selon les diverses molécules avec lesquelles s'effectue la liaison. Tel est le cas pour les ligands dit Amgen (ligands respectifs des structures 2RFN, 2RFS, 3EFJ et 3EFK), pour lesquels un nouveau mode de liaison est décrit. Ainsi, le ligand AM7 de la structure 2RFN se lie à c-Met dans une conformation étendue : le groupement quinoline substitué forme une liaison hydrogène avec le groupement NH du résidu M1160 (gk+3) ; un groupement phenoxy para-substitué se lie en quatrième position et dirige un pyrimidinone benzylé vers l'hélice C. La même description peut être générée pour le complexe 2RFS/SU11274. La partie oxindole se lie à la région charnière via le groupement CO du résidu P1158 (gk+1) et le groupement NH du résidu M1160 (gk+3). Le groupement pyrrole est co-planaire avec le groupement oxindole et se trouve sur le point de former une liaison hydrogène avec le groupement CO du résidu M1160. La géométrie des 3 liaisons hydrogène formées avec la charnière paraît sous-optimale du point de vue des angles d'interaction. Le groupement sulfone forme aussi une liaison hydrogène avec le groupement NH du résidu D1222 et le groupement chlorophenyl s'emboîte dans une interaction Π avec le groupement phenol du résidu Y1230 (Un mode de liaison similaire est observé dans la structure cristallographique de c-Kit).

Les 2 autres inhibiteurs rapportés par le même groupe (Amgen) sont directement reliés à l'inhibiteur AM7. Ce sont les ligands des structures respectives de 3EFJ et 3EFK. Pour passer d'AM7 vers le ligand de 3EFJ, le groupement ethyl morpholine a été tronqué, car il servait apparemment de groupe solubilisant et ne rentrait pas dans les interactions ligand-protéine. L'inhibiteur du récepteur c-Met, 3EFK, est relié à celui du récepteur 3EFJ par la greffe (le remplacement) du groupement 4-fluoroaniline à la place du groupement benzyl. Ces 2 derniers inhibiteurs sont similaires et partagent donc les mêmes caractéristiques de liaisons, à une exception près. La partie la plus hydrophobe de ces 2 inhibiteurs se superpose bien : ce qui est le cas pour le groupement dimethoxyquinoline qui se lie au résidu M1160 (gk+3), le groupement 2-fluorophenoxy et les groupements terminaux 4-fluoroaniline (ligand de 3EFK) / benzyl (ligand de 3EFJ) qui sont orientés vers l'hélice C. Le groupement pyrimidone change d'orientation de façon concertée lorsque le groupement terminal phenyl passe du méthylène à l'aniline nitrogène. Ceci est accompagné par un changement conformationnel au niveau des résidus D1222 et F1223. Pour résumer, les 2 inhibiteurs coordonnés des récepteurs 3EFJ et 3EFK adoptent une conformation DFG-out classique même si leurs ligands n'occupent pas la poche hydrophobique vacante.

D'autres exemples d'inhibiteurs c-Met se liant à la conformation DFG-out de manière plus conventionnelle ont été publiés par les chercheurs du groupe BMS. Les codes PDB de ces structures cristallographiques complexées sont les suivants : 3C1X/ckk, 3CE3/1fn, 3CTH/319, 3CTJ/320 et 3F82/353. L'inhibiteur « ckk » se lie à 3C1X avec la participation de son groupement pyrolotriazine via le résidu M1160 (gk+3) de la région charnière, auquel est attaché le groupement 4-anilino-2-fluorophenoxy en position 5. Le groupement fluorophenyl, quant à lui, occupe la poche du motif DFG.

Il existe des points communs entre les zones d'interaction des 2complexes 3EFK/mt4 et 3C1X/ckk ; cependant, la différence entre ces 2 modes de liaison réside essentiellement dans l'orientation et l'interaction des zones terminales du groupement 4-fluorophenyl. Le groupe fonctionnel « quinole » s'aligne très fortement avec la fonction « pyrrolotriazine » comme le fait le groupement 2-fluorophenoxy proche la région charnière.

La substitution du noyau pyrrolotriazine du ligand ckk par le groupement 2-aminopyridine dans le ligand 319, ou par le groupement pyrrolopyridine provenant du ligand 320, ne donne ni de variations sensibles du mode de liaison des ligands, ni d'importantes réorganisations des chaînes latérales de la protéine. Cependant, l'arrangement des fonctionnalités des liaisons hydrogène dans les ligands ckk, 319 et 320, suggère la capacité des analogues contraints de moduler les propriétés physicochimiques et d'améliorer, potentiellement, l'affinité du ligand pour sa cible en réduisant le nombre de rotations libres. Ce fait est accompli par l'insertion d'une pyridinone donnant lieu à la création (synthèse) de l'inhibiteur « 1fn », cristallisé avec la protéine 3CE3. Cet inhibiteur possède une grande

affinité pour son récepteur et a amélioré l'activité mesurée *in vitro* tout en maintenant l'essentiel du mode de liaison décrit plus haut.

Un analogue légèrement différent existe dans la structure cristallographique complexée 3F82, il s'agit du ligand nommé « 353 » ou encore « BMS-777607 ». Pour cet inhibiteur, le groupement ethoxy a été ajouté au pyridinone et une chlorine remplace la partie pyrrole ethylène, ce qui ne cause pas de changement dans le mode de liaison du complexe.

Plusieurs autres inhibiteurs qui se lient avec différentes conformations de la kinase c-Met inactivée ont été décrits par les chercheurs de l'équipe Amgen. L'origine de l'accroissement de la sélectivité pour c-Met vient des interactions clés qu'elle forme avec les résidus D1222 et Y1230, ainsi que de la conformation inhabituelle de la partie N-terminale de la boucle d'activation. Dans les 2 structures cristallographiques des complexes 3CCN/lkg et 3CD8/l5g, communiquées par Amgen, le groupement triazolopyridazine est positionné entre les résidus M1211 et Y1230 dans une interaction co-planaire de type Π-stacking. De plus, un des nitrogènes du triazole forme une liaison hydrogène avec le groupement NH du résidu D1222. La région charnière est engagée par le résidu M1160 (gk+3 avec le groupement hydroxyl du phenol pour le ligand « lkg », et avec le groupement quinoline nitrogène pour l'inhibiteur l5g. Dans le cas d'une autre structure complexée 2WD1/zzy, des interactions similaires aux cas précédents ont été reportées pour le groupement azaindole de l'inhibiteur qui se lie au groupement nitrophenyl. Dans ce cas, le groupement nitrophenyl forme une liaison hydrogène avec le groupement NH du résidu D1222, alors qu'il est pris entre les résidus M1211 et Y1230. Cette liaison hydrogène semble conférer à l'ensemble du système une importante affinité de liaison alors que les autres analogues s'en trouvent dépourvus et possèdent une affinité de liaison inférieure avec leurs récepteurs respectifs.

Parmi les composants moléculaires de la structure cristallographique de 3F66, le groupement butyrolactone, qui provient du solvant du complexe 3F66/ihx, forme une liaison hydrogène avec le groupement NH du résidu D1222 établissant des contacts lipophiliques ou hydrophobes avec ihx (principal inhibiteur du récepteur). L'inhibiteur ihx est une quinaxoline substituée au niveau de son groupement phenyl. Lorsque le groupement m-phenol est substitué par le groupement nitrophenyl, l'activité inhibitrice augmente de 50 fois. Le groupement nitrophenyl déplace la molécule du solvant, la butyrolactone, de manière à former une liaison hydrogène avec le groupement NH du résidu D1222 [57].

De récentes structures cristallographiques de c-Met ont été publiées sur la PDB : 2WGJ/vgh (2009), 3DKC/ATP, 3DKF/sx8, 3DKG/sx8 (2010). Les deux dernières protéines portent le même ligand dans le site de liaison de l'ATP. Il s'agit d'une molécule inhibitrice similaire aux ligands des récepteurs 3CCN et 3CD8 avec une substitution au niveau du groupement triazolopyridazine. La protéine 3DKC se lie à l'ATP et possède 2 Tyrosines Y1234 et Y1235 mutées en F1234 et E1235. Cette mutation a pour conséquence un

changement de structure de la boucle d'activation en hélice, ce qui tend à la déplacer loin du site de liaison de l'ATP. Le ligand de la protéine 2WGJ, vgh, engage son groupement aminopyridine dans une liaison hydrogène avec les résidus P1158 (gk+1) et M1160 (gk+2) de la région charnière. L'inhibiteur forme également une interaction de type Π-stacking entre son noyau benzylique halogéné et le résidu Y1230.

Pour résumer, les structures des complexes ligand/protéine de c-Met publiées peuvent se diviser en 2 principales catégories selon leurs modes de liaison. En effet, il y a les inhibiteurs dit de Type I qui adoptent une géométrie en forme de U dans le site de liaison de c-Met. Dans ce cas, les interactions se font essentiellement avec la partie charnière et la Tyrosine Y1230 de la boucle d'activation. En revanche, les inhibiteurs dits de Type II adoptent une position plus ou moins étendue. Certains adoptent le mode de liaison conventionnel DFG-out, d'autres un mode plus biaisé de la liaison DFG-out dans lequel la partie hydrophobique du ligand occupe une nouvelle poche.

L'alignement ou superposition des inhibiteurs indique clairement la démarcation entre les 2 types de liaison qui existent dans les interactions ligands/c-Met. Cet état de fait dénote clairement la flexibilité de la poche de liaison de c-Met au vu de la diversité moléculaire de ses inhibiteurs les plus connus.

Conclusion générale

Depuis 2010, plus de 70 brevets concernant les ligands de la kinase c-Met ont été publiés. Ce qui représente bien l'intérêt croissant pour les inhibiteurs c-Met. Bon nombre d'inhibiteurs c-Met sont en phase d'essais cliniques ou de développements pré-cliniques. De ce fait, les informations générées des études conduisant au développement de ces brevets et la compréhension croissante des modes de liaison ainsi que des changements conformationnels de la kinase, conduisent à la progression constante dans le domaine de la recherche de molécules thérapeutiques anticancéreuses, en particulier.

Cependant, il reste encore d'importants défis à relever concernant les inhibiteurs des kinases. Ces interrogations non résolues se rapportent essentiellement aux fonctions et interactions des kinases dans les systèmes biologiques complexes. C'est pour cela que la compréhension de la biologie structurale des kinases semble être un bon moyen pour permettre une amélioration de la spécificité et de l'activité des inhibiteurs et pour développer des stratégies afin de surmonter la résistance aux médicaments.

PARTIE 3

METHODES ET METHODOLOGIES

I- La conception de médicaments ou Drug Design

1. Introduction

Le domaine du « drug design » ou « conception de médicaments » peut être exploré par la modélisation moléculaire et la conception de nouvelles molécules à l'aide des outils informatiques.

Il est aussi appelé drug design rationnel et peut être défini par un ensemble de processus inventifs capables de trouver de nouveaux médicaments basées sur les connaissances de cibles biologiques. Le médicament ou la molécule thérapeutique est généralement une petite molécule activant ou inhibant la fonction d'une biomolécule telle que les protéines et entraînant un bénéfice thérapeutique pour le patient. Fondamentalement, la conception de médicament inclut la conception de petites molécules de formes et charges complémentaires à la cible moléculaire avec laquelle elles interagiront. La conception de médicaments repose fréquemment sur les techniques de modélisation sur ordinateur dites aussi « *in silico* »[58]. Ce type de modélisation est connu sous le nom de conception de médicaments assistée par ordinateur.

Les techniques de modélisation tentent de prédire l'affinité entre le ligand et sa cible et, à ce jour, les résultats sont globalement satisfaisants. Cependant, il existe d'autres propriétés qui doivent être optimisées pour que le ligand devienne sûr et efficace autant que puisse l'être un médicament. Ces caractéristiques moléculaires que sont l'absence d'effets secondaires, la biodisponibilité, la demi vie métabolique, la toxicité etc, sont par ailleurs assez difficiles à optimiser avec les techniques de conception de médicaments rationnelle.

Une cible médicamenteuse est une clé moléculaire qui se trouve dans un métabolisme particulier et évolue dans un cheminement de signalisation spécifique à une pathologie ou une infection, ou à la survie de pathogènes microbiens. Il existe des approches qui tentent d'inhiber le fonctionnement du cheminement de la maladie par l'intervention d'une molécule clé qui arrêterait ce fonctionnement en question. Des médicaments se liant à la région active et inhibant cette partie clé de la molécule peuvent être alors conçus. Une autre approche vise à améliorer le cheminement normal par la promotion de molécules spécifiques, qui affectent ce cheminement en question, dans un état pathologique. En plus de tout cela, les médicaments conçus doivent répondre à la propriété de sélectivité. Cette sélectivité est un terme relatif qui correspond au rapport de l'affinité du médicament pour la cible visée, dans un objectif thérapeutique, sur son affinité pour d'autres cibles. En d'autres termes, si le médicament ne répond pas à la caractéristique de la sélectivité, cela provoquerait de lourds effets secondaires.

Plus communément, les médicaments sont de petites molécules organiques produites à partir de synthèses chimiques, ou à partir de processus biologiques (les médicaments ou nanoparticules basées sur les polymères [59] ou encore de technologies basées sur l'extinction de gènes via l'ARN messager ou « gene silencing[6] ».

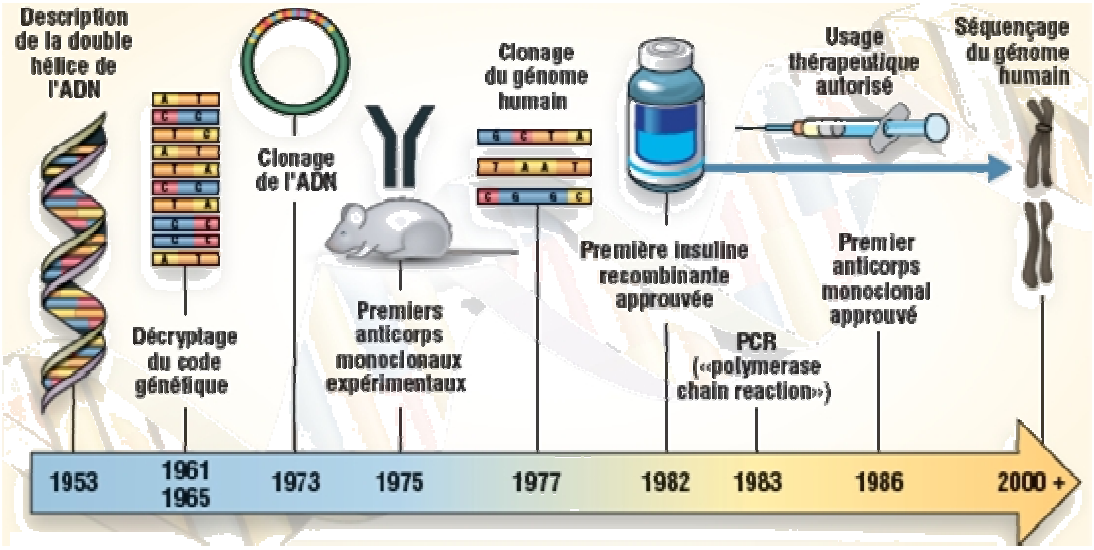


FIG. 16 : Évolution des biotechnologies. (Source Amgen [59])

2. Approches de la conception de médicaments

Deux types d'approches de drug design sont connus. La première fait référence au ligand et se nomme « ligand-based drug design » et la deuxième se réfère à la cible moléculaire et se nomme « stucture-based drug design ».

2.1. Approche « Ligand-based »

Cette démarche repose sur les connaissances acquises concernant d'autres molécules interagissant avec la cible d'intérêt biologique. Ces autres composés peuvent être utilisés afin d'en dériver un pharmacophore qui définit le minimum de caractéristiques que doit posséder la molécule afin d'intéragir avec le système biologique d'intérêt [60]. Ici, le manipulateur détermine une relation de structure-activité quantitative (QSAR) dans laquelle une corrélation entre les propriétés des molécules calculées et leur activité biologique, expérimentalement déterminée, est établie. À leur tour, ces relations QSAR peuvent être utilisées pour la prédiction de nouveaux analogues.

2.2. Approche « Structure-based »

Elle repose sur les connaissances de la structure tridimensionnelle de la cible biologique. Ces connaissances sont obtenues par des méthodes expérimentales telles que la cristallographie par rayons X et la spectroscopie par Résonnance Magnétique Nucléaire (RMN). Ceci est

[6] L'extinction de gènes est un processus épigénétique empêchant la production d'une protéine à partir d'un gène en stoppant la transcription ou la traduction.

possible si la structure de la cible est connue. Par contre, pour toutes les structures non déterminées expérimentalement il est possible d'obtenir un modèle, construit avec la modélisation par homologie, basé sur la structure expérimentale d'une protéine proche (même classe voire même famille). En utilisant la structure de la cible biologique, la conception de molécules ayant une grande affinité sélective à la cible se fait à l'aide d'outils graphiques et interactifs, de procédures de calculs automatisés et en comptant aussi sur l'intuition d'un chimiste médicinal. Avec le développement des méthodes expérimentales (Cristallographie aux Rayons X, spectroscopie RMN), la quantité d'informations des structures 3D des cibles biomoléculaires connaît elle aussi un grand essor. En même temps, les informations relatives aux propriétés électroniques et à la dynamique structurale des ligands sont en continuelle expansion. Tous ces développements se répercutent positivement sur l'évolution des approches « structure-based » de la conception des médicaments. D'où la division de ce type d'approche en deux :

a. La première catégorie s'intéresse à « trouver » un ligand pour une cible ou un récepteur donné dans une base de données servant aux recherches Cette méthode consiste à cribler un grand nombre de ligands pour extraire ceux qui s'adaptent à la poche de liaison du récepteur. L'avantage de cette méthode consiste en l'économie d'efforts de synthèse dans la recherche de nouveaux « composés de tête » (lead compounds).
b. La deuxième catégorie s'intéresse, quant à elle, à la « construction » du ligand, généralement désignée comme « receptor-based » de la conception des médicaments. Les ligands sont édifiés sans tenir compte des contraintes structurales dues à la poche de liaison, en assemblant de petites parties de molécules de manière graduelle. Les différentes parties en question peuvent être des fragments moléculaires ou des atomes. L'avantage de cette méthode, qui connaît de plus en plus d'engouement [61], réside dans la construction de molécules originales qui ne se trouvent pas dans les bases de données

Généralement, la méthodologie utilisée dans la démarche « structure-based » consiste en trois étapes essentielles :

2.2.1. L'identification du site actif

Cette première étape consiste à analyser la protéine en trouvant les coordonnées du site actif ou site de liaison et les résidus essentiels à la formation des liaisons ligand-récepteur. Pour cela, il est nécessaire d'avoir la structure 3D de la protéine avec un ligand pré-docké (pré-arrimé), dans le format de fichier PDB par exemple, et de disposer aussi des propriétés atomiques de l'ensemble. Ligand et récepteur doivent être classés et leurs propriétés atomiques doivent être définies généralement comme suit :

- Les atomes hydrophobes : tous les carbones sont dans des chaînes carbonées ou dans des groupements aromatiques.

- Les liaisons hydrogène donneur : cela concerne les atomes d'hydrogène liés à des atomes d'oxygène et d'azote.
- Les liaisons hydrogène accepteur : cela concerne les liaisons des atomes d'oxygène et d'azote (hybridés sp ou sp2) avec des paires d'électrons libres.

L'espace interne de la région de liaison est étudié avec des atomes sonde virtuels selon les quatre types de liaisons, de manière à étendre les connaissances sur l'ensemble des recoins de cette région. Cela nous donne l'information nécessaire pour faire correspondre les groupements d'interactions du ligand avec les bons points d'ancrage dans la poche de liaison du récepteur.

2.2.2. Les liaisons des fragments du ligand

Le terme « fragment » est utilisé pour désigner les « blocs » d'atomes entrant dans la construction d'une structure organique. Le premier fragment sera introduit dans les différentes régions du site d'activation déjà défini auquel l'expérimentateur combinera d'autres fragments moléculaires un par un. Pour cela, il faudra tenir compte de paramètres tels qu'une énergie de liaison minime extraite de la surface d'énergie potentielle entre les fragments introduits et la poche de liaison. D'où la nécessité de faire des calculs de fonctions de score pour chaque changement de conformation des fragments dérivant des multiples combinaisons. Au vu de l'étendue des combinaisons possibles, il est préférable de prendre comme base de construction la conformation du premier fragment qui forme des interactions significatives. Cette première conformation, assurant une bonne affinité d'interaction, décidera de la manière d'évolution du ligand au cours d'un processus optimisant l'énergie du complexe protéine/ligand. Cette stratégie réduit efficacement la lourdeur des calculs de construction de fragments en diminuant, certes , le nombre de combinaisons possible mais en fiabilisant les ligands construits [61, 62]

2.2.3. Approximation de l'énergie libre de liaison

L'hypothèse de base sous jacente de la conception de médicament « structure-based » est qu'un bon ligand doit se lier à sa cible avec une bonne affinité. Pour cela, le paramètre le plus important dans l'obtention potentielle des nouveaux ligands réside dans la prédiction des affinités de liaisons entre un ligand et sa cible et son utilisation comme critère de sélection. Cette prédiction passe par la fonction empirique développée par Böhm :

$$\Delta G_{bind} = - RT \ln K_d \qquad (1)$$

Avec $K_d = \frac{[Récepteur][Accepteur]}{[Complexe]}$ (Gibbs)

En sachant que $\Delta G_{bind} = \Delta G_{mouvement} + \Delta G_{interaction} + \Delta G_{desolvation} + \Delta G_{configuration}$

L'idée de base est que toute énergie libre de liaison peut se décomposer en éléments indépendants et connus pour être importants dans le processus de liaison. Chaque partie de

l'équation représente une sorte d'énergie libre d'altération durant le processus de liaison entre le ligand et son récepteur. L'établissement et l'étalonnage de sous-modèles de ces équations représentent de vrais challenges pour les chercheurs. Ces sous-modèles dépendent des considérations accordées aux systèmes étudiés et sont continuellement améliorés [63].

Par rapport aux fonctions de score habituellement utilisées dans les algorithmes de docking, ces méthodes plus précises sont plus coûteuses en temps de calcul et ne sont pas encore utilisées couramment en criblage virtuel.

3. Découverte rationnelle de médicaments

Les méthodes traditionnelles de découverte de médicaments s'appuient beaucoup sur la base « essai-erreur » effectués via les tests de substances chimiques sur des cellules de cultures ou sur les animaux. En regroupant les effets apparents des traitements, la conception rationnelle de médicaments se fonde essentiellement sur deux évidences. La première est que les modulations d'une cible spécifique peuvent présenter un potentiel thérapeutique. Cette hypothèse découle de nombreuses études montrant l'association entre les mutations de cibles biologiques et certains états pathologiques. La deuxième évidence est que toute cible thérapeutique doit présenter des propriétés dites « druggable ». Cette propriété décrit la capacité d'une cible à se lier à de petites molécules ayant le pouvoir de moduler son activité. À l'aide de ces deux principaux critères, l'identification de la cible est effectuée pour la cloner ensuite et l'exprimer en plus de la détermination de sa structure tridimensionnelle. L'expression de la cible sert à établir des essais de criblage.

La recherche de petites molécules se liant à la cible commence par un criblage de chimiothèques (librairies de molécules). Ceci peut être fait aussi bien avec le criblage expérimental que le criblage virtuel à condition de disposer de la structure cristallographique 3D ou d'un modèle de la cible. Idéalement, les petites molécules criblées devraient avoir des propriétés dites « drug-like », i.e, des propriétés qui induisent la biodisponibilité orale, une stabilité chimique et métabolique suffisante et un minimum d'effets toxiques. Ces propriétés peuvent être déterminées par de nombreuses méthodes dont la plus connue reste « La règle des cinq » de Lipinski. La prolifération de larges chimiothèques décrivant de potentielles molécules thérapeutiques sans effets secondaires a renforcé les défis de la conception de nouveaux médicaments. En effet, le mélange savant de rationalisme borné et de hasards heureux reste encore une modalité employée dans le processus de conception de médicaments.

Un exemple d'anticancéreux célèbre et issu des méthodes de conception rationnelle de médicament est la molécule de l'imatinib (plus connue sous le nom commercial du Glivec®). Cet inhibiteur de Tyrosine Kinase est spécifique de la protéine Bcr-Abl caractéristique de la Leucémie Myéloïde Chronique (LMC) et de la Leucémie Myéloïde Aigüe (LMA). Son

importance réside dans son caractère sélectif qui le différencie des autres molécules anticancéreuses. En effet, en tant qu'agent de la chimiothérapie, l'imatinib cible les cellules malades qui se divisent rapidement, contrairement à la plupart des anticancéreux qui ne font pas de distinction entre les cellules saines et les cellules malades de l'organisme.

4. La place de l'ordinateur dans le processus de conception de médicaments

La conception de médicaments assistée par ordinateur emploie la chimie computationnelle pour la découverte, l'amélioration et l'étude de médicaments et molécules biologiquement actives. En effet, l'outil informatique aide la conception de médicaments à des étapes spécifiques du processus :

- Dans l'identification des composés potentiellement thérapeutiques, en utilisant le criblage virtuel (CVHD) aussi bien dans la conception basée sur la structure de la cible « structure-based » que sur les propriétés des ligands « ligand-based »
- Dans le processus d'optimisation de l'affinité et de la sélectivité des molécules potentielles vers les têtes de série « lead » ou appelés encore prototypes.
- Dans le processus d'optimisation de la tête de série par rapport aux propriétés pharmacologiques recherchées tout en maintenant une bonne affinité de cette molécule.

Toutes ces étapes d'intervention de l'outil informatique peuvent être synthétisées dans le schéma récapitulatif suivant.

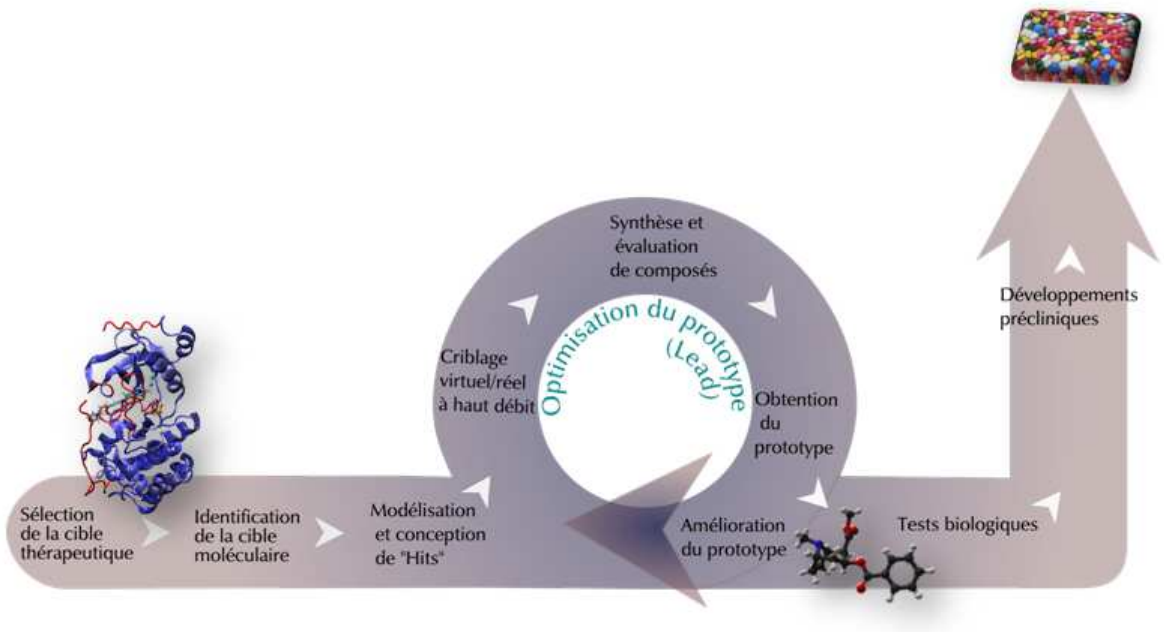


FIG. 17: Processus itératif de la conception de médicaments et domaines potentiels d'intervention de l'approche *in silico*

Le but de cette approche reste la prédiction de la conformation de la molécule qui établira une liaison avec la cible ainsi que celle de la force de la liaison en question. Pour cela, des outils de la mécanique moléculaire tels que la Dynamique Moléculaire (DM) sont utilisés afin de modéliser le changement conformationnel avant, pendant et après l'établissement de la liaison. Les méthodes *ab-initio*, les méthodes semi-empiriques de chimie quantique ou la théorie de la fonctionnelle de la densité (DFT), sont également utilisées dans l'optique du calcul des paramètres des champs de force de la mécanique moléculaire et pour l'estimation des propriétés électroniques de la molécule candidate. Ces propriétés, telles que la polarisabilité et le potentiel électrostatique, influencent positivement l'affinité de liaison.

Les méthodes de mécanique moléculaire peuvent aussi être utilisées pour prédire qualitativement les affinités de liaison protéine/ligand. Les connaissances basées sur les fonctions de scores peuvent fournir des estimations sur les affinités de liaisons. Ces méthodes reposent sur les approches de régression linéaire, les réseaux de neurones, l'apprentissage automatique (les techniques d'apprentissage de type SVM) et tous les autres outils statistiques afin d'obtenir les équations d'affinité de liaisons en ajustant les affinités expérimentales aux énergies d'interactions obtenues par calculs entre les petites molécules et leur cible [64].

Dans le cas idéal, ces méthodes computationnelles nous permettraient de faire sortir du lot une seule molécule, la meilleure, afin de la synthétiser par la suite. En pratique, ces méthodes font des approximations sur les énergies d'interaction entre une molécule et sa cible en filtrant celles qui semblent les plus pertinentes. Ceci nécessite plusieurs itérations de conception, synthèse et tests avant qu'une molécule optimale ne soit repérée. D'un autre côté les avancées continuelles dans le développement des méthodes computationnelles contribuent à la réduction des itérations faisant accélérer les découvertes de nouvelles structures de petites molécules thérapeutiques.

De nombreuses informations issues des représentations 3D des structures des cibles et des complexes sont utilisées comme supports pour pallier à l'insuffisance des prédictions des affinités de liaisons calculées par des fonctions de scores issues des méthodes computationnelles. En effet, dans le processus de conception de médicaments « stucture-based », plusieurs analyses post-criblage concernant les interactions protéine-ligand, ont été développées afin d'améliorer l'enrichissement et l'extraction effective de candidats potentiels. Ces analyses consistent en :

- Les consensus de score : sélection des candidats par l'adoption de multiples fonctions de scores. Cependant, il y a, ici, risque de perte de relation entre l'information structurale du couple protéine-ligand et le critère de score [65].
- L'analyse géométrique : analyse et comparaison des interactions protéine-ligand en inspectant visuellement leur structure individuelle. Cependant, cela devient vite ingérable lorsque le nombre de complexes à analyser croît.

- L'analyse topologique ou analyse de « cluster » : représentation et regroupement des candidats par similitude d'informations structurales des protéine-ligand. Pour cela, il faut impérativement disposer de représentations significatives des couples protéine-ligand [66].

L'ensemble des méthodes théoriques citées plus haut est détaillé dans le reste du contenu de ce chapitre.

II- La Mécanique moléculaire

La modélisation moléculaire regroupe des méthodes permettant de simuler le comportement d'un système allant d'une simple molécule diatomique à des macromolécules de plusieurs milliers d'atomes. Avec les outils de la modélisation moléculaire, l'utilisateur peut décrire d'une part les comportements électroniques des atomes et des molécules dans le but d'expliquer leur réactivité, d'autre part l'importance de certains acides aminés par rapport à d'autres[7] ou encore de comprendre le processus de repliement d'une protéine.

Ce genre d'étude vise à déterminer l'expression de l'énergie d'interaction des atomes du système moléculaire en fonction de leurs positions relatives. Pour effectuer cela, les méthodes de la modélisation moléculaire reposent sur les lois de deux approches : la Mécanique Quantique (MQ) et la Mécanique Moléculaire (MM). La mécanique quantique permet de fournir une description satisfaisante, sur le plan théorique, de l'énergie et de l'état d'un système.

Apparue en 1930 [67], la MM basée sur la MQ s'est développée avec l'évolution des moyens informatiques dans les années soixante. Les techniques des MQ correspondantes sont basées sur une résolution mathématique approchée[8] de l'équation de Schrödinger. Toutefois, la complexité associée à de tels calculs est si importante que ces simulations dites *ab initio* sont limitées à de petits systèmes ne comportant qu'un faible nombre d'atomes. En revanche, ces calculs traitent explicitement les électrons d'un système moléculaire et permettent ainsi l'étude précise de sa réactivité chimique (création ou destruction de liaisons).

L'approximation de Born-Oppenheimer est utilisée pour la modélisation de systèmes moléculaires [68] en MM. C'est une approximation qui considère les noyaux atomiques fixes par rapport aux électrons[9]. Le mouvement des noyaux n'est plus décrit par un Hamiltonien quantique contrairement à la Mécanique Quantique. Il est décrit selon les lois de la mécanique classique où les atomes sont assimilés à des points ou sphères et les liaisons chimiques à des ressorts mécaniques. Les techniques *ab initio,* quant à elles, se concentrent sur la caractérisation des seuls électrons. La simulation de systèmes de grande taille tels que les biomolécules nécessite d'avoir recours à des modèles bien plus simplifiés.

À condition d'être correctement paramétrisées, les méthodes de la MM mènent à des résultats similaires à ce qui serait obtenu par les méthodes de la MQ avec une consommation moindre de temps de calcul. Cependant, les méthodes de la MM sont inadaptées à la description des propriétés électroniques des systèmes étudiés. Elles permettent néanmoins, sous certaines conditions, de déterminer de façon appropriée leurs propriétés structurales et thermodynamiques. Les méthodes de la MM offrent des techniques de choix[10] pour simuler

[7] Les acides aminés d'un site catalytique enzymatique par exemple.
[8] L'équation de Schrödinger ne peut pas mathématiquement être résolue de façon exacte pour les systèmes moléculaires.

[9] En effet, la masse des noyaux étant ~ 1 000 fois plus importante que celle des électrons, on peut supposer que les électrons se déplaceront beaucoup plus vite que les noyaux et donc considérer ces derniers comme fixes.
[10] La Dynamique Moléculaire par exemple

des systèmes biologiques de taille importante[11] permettant ainsi d'obtenir des informations détaillées qui relient la structure d'un système biologique à sa fonction.

1. Fonction d'énergie potentielle et champ de force

Comme spécifié dans l'introduction, la MM assimile les atomes d'un système à un ensemble de masses ponctuelles chargées interagissant entre elles à travers une fonction d'énergie potentielle prédéfinie. Elle s'appuie, pour cela, sur les lois de la mécanique Newtonienne afin de modéliser le comportement des systèmes moléculaires. La fonction d'énergie potentielle d'un système est représentée comme une hypersurface à 3 dimensions. Elle est définie empiriquement comme étant la somme des différentes contributions traduisant les interactions physiques inter et intramoléculaires du système étudié. En d'autres termes, la fonction d'énergie potentielle, associée à un jeu de paramètres qui lui est propre, constitue ce qu'on appelle un champ de force.

Le champ de force est l'assemblage d'une équation qui régit l'énergie d'un système et de ses paramètres associés [69]. Il définit le lien entre la position des atomes du système et leur énergie potentielle au cours de la simulation [70]. En effet, il contient tous les paramètres qui entrent en jeu au cours de la simulation, que ce soit la dynamique moléculaire ou la minimisation de l'énergie du système étudié. En d'autres termes, le champ de force est un des ingrédients les plus délicats à mettre en œuvre [71].

Plusieurs champs de force ont été créés en accord avec le système étudié. Nous dressons la liste des champs de force les plus utilisés en concordance avec le domaine d'application :

- CHARMM [70, 72-76] : utilisé pour les protéines
- AMBER [77-83] : utilisé pour les protéines et les acides nucléiques
- MM2, MM3 et MM4 [84] : utilisés pour les petites molécules organiques
- ECEPP [85] : utilisé pour les protéines
- BIOSYM (MSI) : utilisé pour les protéines
- CVFF [86] : utilisé dans la recherche de fréquence de vibration
- Sybyl (de Tripos) [87] : utilisé pour les molécules organiques

La fonction d'énergie potentielle peut être décomposée en la somme des potentiels qui régissent l'interaction entre atomes liés et les atomes non-liés. Elle a, par exemple[12], pour expression générale :

$$V_T = \underbrace{\sum V_{liaisons} + \sum V_{angles} + \sum V_{dièdres} + \sum V_{dièdresimpropres}}_{V_{\text{covalente}}} + \underbrace{\sum V_{vdW} + \sum V_{électrostatique}}_{V_{\text{non covalente}}} \qquad (2)$$

Le détail de ces deux types de potentiels est donné dans les paragraphes qui suivent.

[11] De 10^5 à 10^6 atomes

[12] Selon le type de champ de force associé. Dans le cas de notre étude, il est question du champ de force CHARMM

1.1. Potentiel entre atomes liés

Le potentiel entre atomes liés concerne les liaisons covalentes, les angles de valence et les angles dièdres entre autres. Au sein d'une structure moléculaire, ce type de potentiel peut se décomposer en interactions entre séries de deux, trois, quatre, ... atomes liés. Si l'on suppose que ces différents termes sont découplés[13] et que l'on néglige les termes d'ordres supérieurs à 5, on obtient l'expression générale la plus courante du potentiel entre atomes liés:

$$V_T = \sum V_{liaisons} + \sum V_{angles} + \sum V_{dièdres} + \sum V_{dièdresimpropres} + \sum V_{vdW} + \sum V_{électrostatique}$$

Les variables de calcul sont les coordonnées internes du système. Nous pouvons les décrire comme suit. $\mathbf{V}_{liaisons}$ décrit l'interaction formée à partir d'une liaison covalente entre 2 atomes ; elle est caractérisée au niveau de la topologie par les longueurs des liaisons r. $\mathbf{V}_{angles}$ est l'interaction pour une suite de 3 atomes liés consécutivement ; il s'agit de l'angle de valence θ. $\mathbf{V}_{dièdres}$ et $\mathbf{V}_{dièdres\ impropres}$ s'appliquent à une suite de 4 atomes et correspondent aux angles dièdres ϕ et ω respectivement. Le formalisme le plus fréquent emploie des fonctions harmoniques :

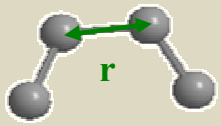
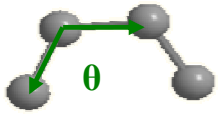
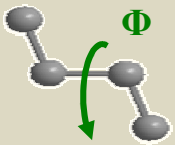
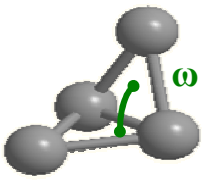
Équations	Représentations
$V_{liaisons} = \sum_{liaisons} K_r (r - r_0)^2$	
$V_{angles} = \sum_{angles} K_\theta (\theta - \theta_0)^2$	
$V_{dièdres} = \sum_{dièdres} K_\phi^n (1 + \cos(n\phi - \delta))$	
$V_{dièdresimpropres} = K_\omega (\omega - \omega_0)^2$	

Tableau 1: Illustration des 4 plus importantes contributions d'un champ de force dans le cas d'atomes liés

[13] Cas des champs de force de classe I tels que CHARMM, AMBER

Pour les groupes d'atomes auxquels ils s'appliquent, chacun des termes est minimal lorsque les valeurs des paramètres géométriques associés correspondent aux valeurs d'équilibre thermodynamique. L'éloignement par rapport à ces valeurs de référence correspond à une différence de potentiel qui peut être assimilée à une pénalité.

1.2. Potentiel entre atomes non-liés

$$V_T = \sum V_{liaisons} + \sum V_{angles} + \sum V_{dièdres} + \sum V_{dièdresimpropres} + \sum V_{vdW} + \sum V_{électrostatique}$$

La deuxième partie de la fonction d'énergie potentielle $\mathbf{V}_{non\text{-}liés}$ peut se développer en somme de termes à n corps (n > 1). On se limite en général aux termes à deux corps, d'une part parce que les suivants ont une amplitude bien plus faible, d'autre part pour des raisons de complexité[14].

Ce qui est sous entendu par le terme « non-liés », c'est aussi bien les atomes de deux molécules distinctes que ceux appartenant à une même molécule mais séparés par au moins trois liaisons covalentes. Tel est le cas dans le champ de force CHARMM. La justification de ce choix vient du fait que les interactions entre atomes séparés par au moins trois liaisons sont suffisamment bien décrites par les fonctions de déformation des longueurs de liaisons et d'angles formés par deux liaisons covalentes.

Dans sa formulation, cette deuxième partie de l'expression de l'énergie potentielle est distinguée par des équations de potentielles entre atomes *i* et *j* non-liés. Ces équations représentent les interactions électrostatiques et interactions dites de van der Waals résumées dans le tableau suivant:

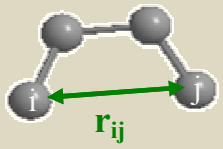
Équations	Représentations
$V_{vdW} = \sum_{\substack{ij,atomes \\ non-liés}} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]$ **(Lennard-Jones)**	
$V_{électrostatique} = \sum_{\substack{ij,atomes \\ non-liés}} \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon_r r_{ij}}$	

Tableau 2: Illustration des contributions d'un champ de force dans le cas d'atomes non-liés.

Le premier terme concernant la partie des atomes non-liés de la fonction de l'énergie potentielle est l'équation caractérisant les interactions de van der Waals. Ces interactions sont habituellement représentées par le potentiel de Lennard-Jones [88]. Il est constitué d'un terme répulsif en « r^{-12} » qui représente le principe d'exclusion de Pauli et la répulsion internucléaire

[14] Contrairement aux termes $V_{liés}$ qui ne s'appliquent qu'aux atomes liés (ce qui en limite le nombre), les termes $V_{non\text{-}liés}$ s'appliquent au moins à toute paire d'atomes du système, d'où la complexité en l'absence d'optimisation.

à courte distance. L'autre terme que contient le potentiel en « r^{-6} » est un terme attractif représentant les forces de London[15].

Le potentiel de van der Waals peut également être représenté par le potentiel de Buckingham avec l'équation suivante :

$$V_{vdW} = A'_{ij} e^{-B'_{ij} r_{ij}} - \frac{C_{ij}}{r_{ij}^6} \qquad (3)$$

Le dernier terme représenté par l'équation de Coulomb définit les interactions électrostatiques entre différentes paires d'atomes distantes de « $r_{i,j}$ » portant des charges ponctuelles « q_i » et « q_j ». Ce potentiel dépend aussi de la constante diélectrique du milieu « ε ».

Notons que certains champs de force incluent, dans l'expression du potentiel entre atomes non-liés, un terme décrivant les liaisons hydrogène qui, dans sa forme la plus simple, est du type :

$$V_{HB} = \sum_{paires} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{10}} \qquad (4)$$

Certains contiennent également des termes dus au couplage entre les variations de différentes coordonnées internes. C'est le cas des champs de force dits de classe II tels que, par exemple, UFF [89], CFF [90] et MM3 [91].

2. L'application du champ de force aux biomolécules (paramétrisation)

Le choix du champ de force, avant toute simulation de MM, s'établit sur la base des résultats déjà obtenus dans la littérature concernant son application aux systèmes moléculaires analogues au système étudié. En effet, l'emploi d'un champ de force paramétrisé pour décrire une certaine catégorie de molécules (p. ex. les protéines) se limite à une classe de molécules ayant des ressemblances structurales et fonctionnelles. Les champs de force couramment utilisés pour la modélisation de biomolécules [92], tels que CHARMM [72, 74-76], AMBER [78, 93-98], GROMOS [99, 100] et OPLS [101, 102], fruits de nombreuses années de recherches spécialisées, peuvent être considérés comme matures. Leur paramétrisation est spécifiquement orientée vers la reproduction correcte du comportement des acides aminés et des acides nucléiques. Ils permettent donc, couplés à des programmes appropriés, tels que NAMD [103-105], GROMOS [106], CHARMM [73, 107] et AMBER [108, 109], de conduire des simulations de dynamique moléculaire considérées comme fidèles, a priori,

[15] Autrement dit, les interactions d'origine dipolaire dues aux fluctuations des nuages électroniques de chacun des atomes

lorsqu'elles sont appliquées à des systèmes biomoléculaires (comme les complexes protéiques).

Comme expliqué dans les précédentes sections, le processus de la MM relie la structure du système (liste des atomes, de leur nature, de leurs coordonnées, des liaisons qu'ils forment et de la nature de celles-ci) à sa topologie (liste des groupes d'atomes que chaque terme du potentiel doit considérer). Celle-ci est ensuite associée aux paramètres prédéfinis dans le champ de force. Les paramètres sont, au cours d'une simulation de MM, des constantes qui doivent être définies rigoureusement avant toute simulation. La détermination des paramètres du champ de force s'effectue sur la base de résultats expérimentaux (données spectroscopiques, calorimétriques, structurales,…), ou lorsque ceux-ci ne sont pas disponibles (ou trop peu précis), à partir de résultats de calculs de MQ *ab initio* [110]. Cela correspond à différentes procédures itératives complexes [93, 111-113].

Le fait de transférer des paramètres d'une molécule vers un groupement structuralement similaire d'une autre molécule est l'une des hypothèses fondamentales de la MM. Bien que cette approche puisse paraître rudimentaire, les multiples applications effectuées depuis son introduction ont permis de démontrer son efficacité. La validation d'un champ de force correspond à la reproduction, sur la base de celui-ci, de résultats expérimentaux. Gardons en tête que chaque champ de force constitue un compromis entre simplicité conceptuelle et précision. L'amélioration d'un champ de force est délicate, par exemple à cause de la détermination de la corrélation entre paramètres et résultats expérimentaux, ou encore du fait que différents paramètres peuvent être couplés [114]. Toutefois, l'augmentation des performances informatiques permet d'y incorporer des termes de plus en plus sophistiqués sans que cela n'alourdisse les temps de calculs.

3. Méthodes d'exploration de l'hypersurface d'énergie d'un système moléculaire

3.1. Minimisation de l'énergie

La minimisation de l'énergie potentielle est l'une des méthodes les plus couramment employées pour optimiser la structure des biomolécules. Cette opération peut être utile afin de relaxer la structure initiale de la molécule (en éliminant les mauvais contacts atomiques) et de rechercher une géométrie de basse énergie, correspondant à un état stable du système. La minimisation de l'énergie potentielle trouve ainsi son application dans les protocoles de raffinement de structures moléculaires obtenues expérimentalement (RMN ou RX) et dans l'optimisation de la structure de biomolécules en préalable à une simulation de dynamique moléculaire.

Le nom d'"hypersurface" est employé pour représenter l'énergie potentielle du système étudié par un espace dans lequel les positions atomiques sont des variables. Le problème consiste à trouver les coordonnées des atomes qui permettent de minimiser la fonction d'énergie potentielle (1) du système étudié. La solution à ce problème est loin d'être triviale car le nombre important de degrés de liberté fait fluctuer le système autour d'un ensemble de conformations stables correspondant à différents minima sur l'hypersurface d'énergie potentielle. L'état de plus basse énergie, sur l'ensemble de l'hypersurface de potentiel, est appelé minimum *global*, par opposition aux minima *locaux*.

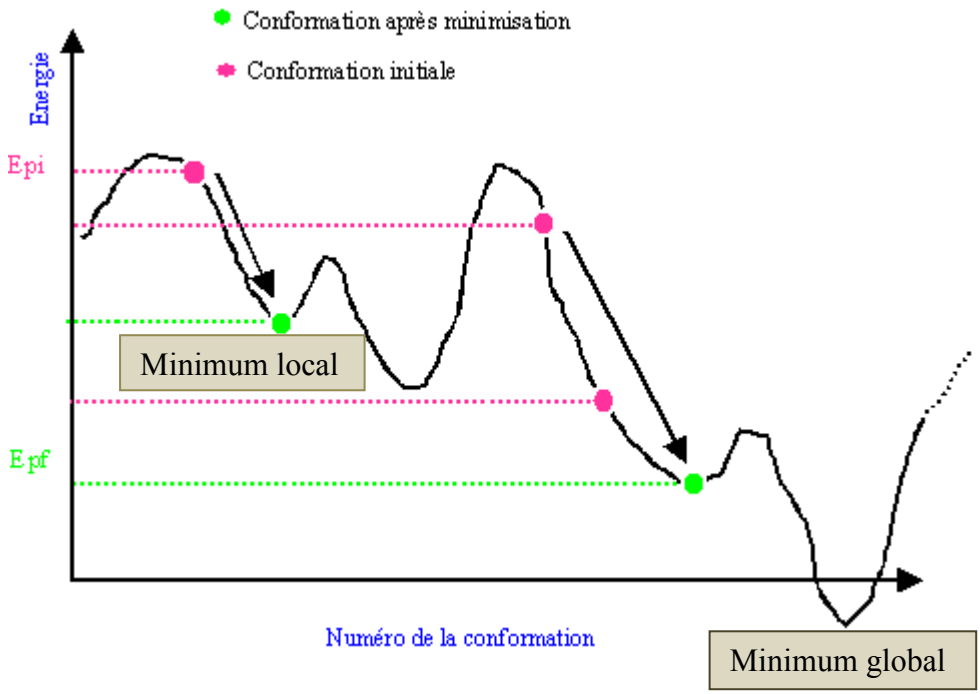


FIG. 18 : Représentation de conformations initiales et finales sur une hypersurface d'énergie potentielle

Il n'existe pas de méthode mathématique générale qui permette de trouver ce minimum global. Cependant, des méthodes d'analyse numérique, telles que les algorithmes de minimisation, peuvent être employées afin de trouver des minima locaux. L'approche la plus souvent utilisée consiste à descendre de manière itérative le long de l'hypersurface dans la direction correspondant à une diminution de l'énergie ; ce qui peut se résumer en :

$$r_{n+1} = r_n + \delta_n \qquad (5)$$

où r_i correspond aux coordonnées des atomes du système étudié, n est le numéro de l'itération et δ_n le énième déplacement dans l'espace des configurations.

Les algorithmes de minimisation les plus courants [115] diffèrent tous dans la manière de se déplacer sur l'hypersurface, c'est-à-dire dans l'évaluation des δ_n. Celles de la plus grande pente et des gradients conjugués utilisent uniquement les dérivées premières de la fonction d'énergie potentielle, alors que les méthodes dites de Newton-Raphson requièrent le calcul des dérivées secondes de cette fonction.

La méthode de la plus grande pente s'avère efficace lorsque la conformation de départ est loin du minimum ; cependant elle converge lentement et peut adopter un comportement oscillatoire au voisinage de celui-ci. La méthode du gradient conjugué converge mieux, car elle présente l'intérêt d'échapper aux comportements oscillatoires autour du minimum tout en accélérant la convergence, mais elle n'est pas adaptée aux structures présentant de mauvais contacts ou pas très bien relaxées. Au vu des avantages et inconvénients de chacune des deux méthodes, on a généralement recours à la combinaison des deux algorithmes de minimisation l'un après l'autre.

De leur côté, les méthodes de Newton-Raphson utilisent la dérivée seconde de l'énergie en calculant de manière explicite le hessien H_i de l'énergie potentielle (matrice des dérivées secondes) pour rechercher l'extremum le plus proche. Ces méthodes permettent de converger très rapidement vers le minimum local, à condition d'en être déjà assez proche au préalable. Le temps de calcul requis par les méthodes de Newton-Raphson, ainsi que l'espace mémoire nécessaire, les rendent non adaptées à l'étude de gros systèmes comme les biomolécules.

3.2. Méthodes de recherche conformationnelle

Les algorithmes de minimisation permettent de rechercher une estimation de la stabilité énergétique du système étudié. Cependant, pour éviter d'être piégé dans le minimum le plus proche, il est conseillé d'effectuer un échantillonnage de l'espace conformationnel. Cette recherche a pour but de déterminer par la simulation un ensemble représentatif de conformations stables de la molécule à l'équilibre thermodynamique, c'est-à-dire susceptibles d'exister dans les conditions expérimentales.

Outre la recherche systématique, les méthodes de recherche conformationnelle peuvent être divisées en deux grandes classes : les méthodes stochastiques et les méthodes déterministes. Nous évoquerons brièvement les méthodes stochastiques à travers la méthode Monte Carlo. Les méthodes déterministes seront illustrées en décrivant la dynamique moléculaire, technique dont l'utilisation fut centrale pour ce travail. Nous consacrerons par la suite, une section décrivant la méthode de recherche conformationnelle à l'aide des Modes Normaux de vibrations (NM) qui ont été largement utilisés dans nos travaux.

3.2.1. Recherche stochastique – exemple du Monte Carlo

Contrairement à la recherche systématique trop coûteuse en temps de calcul (voir annexes B), la recherche stochastique fait varier les degrés de liberté du système de manière aléatoire. La méthode de Monte Carlo [115] est basée sur la génération aléatoire d'un ensemble de conformations et utilise certains critères de sélection pour conserver ou non les nouvelles conformations créées. Ces critères assurent que la probabilité d'obtenir une conformation donnée est égale à son facteur de Boltzmann[16] $e^{-\frac{V(r_i)}{k_B T}}$, où k_B est la constante de Boltzmann, T la température et $V(r_i)$ est calculé par la fonction d'énergie potentielle. La plupart des simulations de Monte Carlo se réfèrent à l'utilisation de l'algorithme de Metropolis [116].

Dans une simulation de Monte Carlo, chaque nouvelle conformation générée résulte du mouvement aléatoire d'un ou plusieurs atome(s) ou de la variation d'un ou plusieurs angle(s) dièdre(s). L'énergie de la conformation obtenue est alors évaluée. Si la nouvelle conformation est de plus basse énergie que la précédente, elle est acceptée. En revanche, si elle est plus haute en énergie, alors le facteur de Boltzmann de la différence d'énergie entre les deux conformations est calculé et constitue la probabilité d'acceptation de la nouvelle conformation. Cette approche permet de visiter des états de plus haute énergie que la configuration actuelle ; cependant, les conformations de plus basse énergie restent préférentiellement échantillonnées. Les méthodes stochastiques ne garantissent pas l'obtention du minimum global de l'énergie, mais elles permettent d'obtenir un ensemble de conformères de faible énergie constituant un échantillonnage supposé être représentatif.

3.2.2. Recherche déterministe – exemple de la dynamique moléculaire

Alors qu'au cours d'une simulation de Monte Carlo les conformations sont générées de manière aléatoire et discrète (dépendant de la probabilité de transition entre l'état courant et le suivant), celles issues d'une simulation par dynamique moléculaire sont, elles, connectées dans le temps et forment une trajectoire continue. La dynamique moléculaire est dite déterministe puisque tous les états[17] futurs du système peuvent être prédits à partir de l'état actuel. Cette technique de simulation, que nous avons utilisée pour ce travail, est détaillée dans une prochaine section.

4. Les ensembles thermodynamiques

Lors d'une simulation, il est possible de travailler dans différents ensembles thermodynamiques. Les plus communément utilisés sont : l'ensemble microcanonique NVE

[16] Le facteur de Boltzmann caractérise la probabilité relative d'un état *i* dans une situation d'équilibre thermique à la température *T*.

[17] Pour la dynamique moléculaire, un état du système ne contient pas seulement l'ensemble des coordonnées spatiales de chaque atome mais aussi les valeurs des vecteurs quantité de mouvement.

(nombre d'atomes, volume et énergie constants), l'ensemble canonique NVT (nombre d'atomes, volume et température constants), et l'ensemble isobare-isotherme NPT (nombre d'atomes, pression et température constants).

Les simulations de Monte Carlo s'effectuent traditionnellement dans l'ensemble NVT alors que les simulations de dynamique moléculaire sont généralement associées à l'ensemble NVE. Cependant, par l'insertion de contraintes dans leurs implémentations respectives, ces deux techniques peuvent s'appliquer à d'autres ensembles thermodynamiques. Dans le cadre de ce travail, toutes nos simulations de dynamique moléculaire ont été réalisées dans l'ensemble isobare-isotherme NPT, où la pression et la température du système sont maintenues constantes par couplage à un barostat et à un thermostat respectivement [117, 118]. Cet ensemble est, en effet, celui qui se rapproche le plus des conditions réelles d'expérience de laboratoire.

5. La dynamique moléculaire

5.1. Principe

La dynamique moléculaire [114, 115] (DM) est une technique couramment utilisée pour la simulation de biomolécules [119]. Son but est d'étudier l'évolution d'un système moléculaire (défini par mécanique moléculaire, voir section I.1) au cours du temps en intégrant les équations de Newton relatives au système :

$$m_i \vec{a}_i = \vec{F}_i \qquad (6)$$

où m_i est la masse d'un atome i, $\vec{a}_i$ son accélération et $\vec{F}_i$ la somme des forces qui lui sont appliquées du fait de son interaction avec les autres atomes et l'environnement.

Lors de la simulation de DM, le système subit des changements conformationnels et cinétiques qui permettent d'explorer l'espace des phases espace-temps accessible par le système. À chaque particule, en tout temps t, on associe un couple (position $\vec{r}_i(t)$, vitesse $\vec{v}_i(t)$). L'ensemble des coordonnées sur la totalité de l'espace temporel exploré constitue ce qu'on appelle une trajectoire. Suivant l'hypothèse ergodique, l'étude d'une trajectoire infiniment longue d'un système par DM revient à échantillonner tout l'espace des phases de ce système. Il est alors possible d'accéder à des grandeurs thermodynamiques (coefficients de diffusion, fonctions de distributions radiales, énergie libre, *etc.*) afin de relier la simulation à l'échelle microscopique aux expérimentations à l'échelle macroscopique.

Les équations du mouvement de Newton peuvent s'écrire de la manière suivante :

$$m_i \frac{d\vec{v}_i(t)}{dt} = -\frac{d\vec{V}(\vec{r}_1,\vec{r}_2,...,\vec{r}_N)}{d\vec{r}_i} = \vec{F}_i(t) \qquad (7)$$

où m_i est la masse d'un atome i, $\vec{v}_i(t)$ sa vitesse à l'instant t, et $\vec{r}_i(t)$ sa position dans l'espace. $\vec{V}(\vec{r}_1,\vec{r}_2,...,\vec{r}_N)$ est le potentiel d'interaction (champ de force) entre les atomes du système.

À partir d'un système dont les conditions initiales ont été fixées, la DM consiste en la répétition de deux opérations : tout d'abord évaluer la force agissant sur chaque atome au temps t, puis déterminer les coordonnées et les vitesses des atomes au temps $t+\Delta t$ en fonction des forces subies par chacun d'entre eux où Δt correspond au pas d'intégration.

5.2. Intégration des trajectoires

Il n'existe pas de solution analytique exacte aux équations du mouvement[18]. Différents algorithmes ont été développés afin de les résoudre numériquement. Pour cela, une approximation usuelle consiste à diviser l'évolution du système en intervalles de temps (discrétisation temporelle). A l'issue de chacun de ces pas de temps, le potentiel de chaque particule est recalculé. L'erreur induite par cette approximation est négligeable quand les intervalles de temps utilisés sont suffisamment petits. Dans la pratique, le pas d'intégration Δt , adapté pour la simulation de systèmes biologiques, est de l'ordre de la femtoseconde (10^{-15}s).

Différents intégrateurs sont disponibles, chacun se caractérisant par un rapport spécifique entre précision et efficacité. L'algorithme de Verlet [120] est parmi les plus utilisés. Dans cet algorithme, les coordonnées sont développées en séries de Taylor au troisième ordre, aux temps $t+\Delta t$ et $t-\Delta t$:

$$\vec{r}_i(t+\Delta t) = \vec{r}_i(t) + \vec{v}_i(t)\Delta t + \frac{\vec{F}_i(t)}{m_i}\frac{\Delta t^2}{2!} + \frac{d\vec{F}_i(t)}{m_i dt}\frac{\Delta t^3}{3!} \qquad (8)$$

$$\vec{r}_i(t-\Delta t) = \vec{r}_i(t) - \vec{v}_i(t)\Delta t + \frac{\vec{F}_i(t)}{m_i}\frac{\Delta t^2}{2!} - \frac{d\vec{F}_i(t)}{m_i dt}\frac{\Delta t^3}{3!} \qquad (9)$$

La différence de ces deux séries permet d'obtenir les vitesses des atomes :

$$\vec{v}_i(t) = \frac{\vec{r}_i(t+\Delta t) - \vec{r}_i(t-\Delta t)}{2\Delta t} \qquad (10)$$

[18] Leur résolution analytique peut être fastidieuse, voire impossible, en particulier si les mouvements des particules sont couplés (problème à N-corps).

Des variantes optimisées de l'algorithme de Verlet existent. On peut citer par exemple : les algorithmes *leap frog* [121], de Beeman [122] et, la plus utilisée, le *velocity Verlet* [123].

La modélisation d'un système macromoléculaire dans des conditions de solvatation explicite requiert la gestion d'un nombre de degrés de liberté tel que souvent il s'avère nécessaire de réduire artificiellement ce nombre. Pour accélérer le calcul, on peut également avoir recours à l'algorithme SHAKE [124-126] qui permet d'augmenter modérément le pas d'intégration (p.ex. $\Delta t = 2$ fs au lieu de 1 fs) en corrigeant les oscillations trop importantes qui peuvent alors apparaître sur les degrés de liberté les plus "rapides" du système.

Pour améliorer l'efficacité de l'intégration, des intégrateurs à pas multiples, tels que l'algorithme r-RESPA (*reversible Reference System Propagator Algorithm*) [127-129] peuvent être employés afin de réduire substantiellement le temps de la simulation (d'environ un ordre de grandeur), sans perte de précision notable. Une telle optimisation diffère ici de la contrainte de type SHAKE indiquée précédemment car ce sont les degrés de liberté les plus lents qui sont périodiquement figés[19], tandis que le pas de calcul reste invariant.

6. Méthodes basées sur les modes normaux de vibration

6.1. Définition

La prise en compte de la flexibilité des protéines ainsi que leurs mouvements d'origine peuvent être étudiés grâce aux simulations de dynamique moléculaire. Cependant, de nos jours, dépasser la microseconde de simulation est difficilement envisageable lorsqu'on s'intéresse à des protéines de plusieurs centaines d'acides aminés. Cette constatation s'applique même en ayant accès à des moyens de calculs importants et en utilisant une représentation simplifiée des effets du solvant.

Pour palier à cette limite, des solutions ont été élaborées par l'étude de ces mouvements dans le cadre de deux approximations :

a) Par une représentation du système dans laquelle les amino acides constitutifs de la protéine sont représentés par un ou deux pseudo-atomes. Cette représentation est dite « gros grain » (Coarse grain).

b) Par une approximation harmonique (ou approximation des petits déplacements). Cette dernière permet la description du mouvement de chacun des atomes d'un système telle une combinaison de modes de vibration indépendants les uns des autres. Dérivant de cette approximation et adoptée depuis les années quatre-vingts, l'approche par Modes Normaux est actuellement très utilisée [130-135]. Elle consiste à modéliser la

[19] Les forces qui varient le plus lentement (p. ex. l'électrostatique à longue distance) sont recalculées moins fréquemment que les plus rapides (p. ex. interactions entre atomes liés).

dynamique d'une protéine, ou d'un complexe macromoléculaire, comme la somme de mouvements élémentaires. Elle est d'ores et déjà utilisée pour le raffinement de structures expérimentales de basse résolution[20]. Des structures ainsi résolues ont d'ailleurs été déposées à la Protein Data Bank.

En 1976, l'équipe de Martin Karplus s'est essayée à la prédiction du mouvement fonctionnel du lysozyme humain, en émettant la supposition que ce dernier se rapproche de son mode de vibration de plus basse fréquence et qu'il s'agit d'un mouvement « de charnière » (hinge-bending) dans lequel deux domaines quasi-rigides se déplacent l'un par rapport à l'autre. Au début des années quatre-vingts, les modes de vibration de peptides, puis de petites protéines, furent calculés dans les équipes de Martin Karplus, Nobuhiro Go et Michael Levitt [136, 137]. Il s'agit de systèmes comme: le BPTI, la crambine, la ribonucléase et le lysozyme humain. À la même époque, Robert W. Harrison calcula les modes de vibration de l'hexokinase (protéine d'environ 450 acides aminés) en utilisant la méthode de Ritz-Raleigh [138]. Son approche consista à réduire la dimension du système étudié. En effet dans la majorité des cas pour décrire avec une bonne précision le changement de conformation fonctionnel des protéines, un ou deux modes de vibration est suffisant.

Cependant, avec l'avènement des simulations de dynamique moléculaire de protéines en solvant explicite, l'intérêt pour les calculs de modes de vibration s'est affaibli. Il semblerait que cette diminution a lieu car les premiers calculs étaient effectués *in vacuo* et qu'une représentation suffisamment concrète de l'environnement paraissait utopique. La proposition des modèles de solvant « implicites » a rendu la description de l'environnement d'un système plus réaliste. Ces modèles tiennent aussi bien en compte de l'énergie libre de solvatation que ceux utilisant l'approche de type Poisson-Boltzman. Il est désormais possible de calculer les modes de vibration de basse fréquence de complexes protéiques de très grande taille [139](ribosome), en utilisant l'approximation harmonique. En effet c'est une approche qui permet de recueillir des informations dynamiques pertinentes [134, 135, 140, 141], en dépit de certaines limites expliquées par la suite.

6.2.L'approximation harmonique

Il a été démontré dans nombreuses études que les mouvements de basse fréquence et de grande amplitude des biomolécules sont nettement anharmoniques [142-144]. Au début d'un changement conformationnel, sur un temps très court, l'approximation des modes normaux est permise par la déformation de faible amplitude. Cependant, la question de connaître l'amplitude maximum utilisée pour « déformer » le système lors des modes normaux se pose.

Pour un mode normal de basse fréquence calculé selon les méthodes de mécanique moléculaire typiquement utilisées, la mécanique statistique prévoit qu'à 300K, le carré de

[20]Pour des données obtenues par SASX ou par cryo-microscopie électronique et des structures cristallographiques.

l'amplitude moyenne de la fluctuation thermique soit très petit (de l'ordre de 1Å[2] [21]). Il s'avère que les amplitudes observées peuvent être beaucoup plus importantes, comme lors d'une déformation induite par le couplage avec un ligand. Nous abordons ici les fluctuations induites par une perturbation extérieure en sortant du cadre des fluctuations thermiques spontanées de la molécule. Ces fluctuations induites apportent un gain d'énergie qui permet de gagner de l'amplitude.

Ainsi, pour de petits déplacements des coordonnées de ses N atomes, l'énergie potentielle d'une protéine, peut être donnée par un développement en série limitée :

$$V = V_0 + \sum_{i=1}^{3N} \left(\frac{\partial V}{\partial r_i}\right)_0 (r_i - r_i^0) + \frac{1}{2}\sum_{i=1}^{3N}\sum_{i=j}^{3N} \left(\frac{\partial^2 V}{\partial r_i \partial r_j}\right)_0 (r_i - r_i^0)(r_j - r_j^0) + \cdots \quad (11)$$

Où « V » est défini à une constante additive près et les autres termes comme suit :
r_i : la i[ème] coordonnée
V_0 : l'énergie potentielle de la structure dont on veut calculer les modes de vibration ; l'indice « 0 » est associé aux coordonnées de la conformation considérée. On peut poser : $V_0 = 0$.
Si V_0 est un minimum de la fonction V, alors on a pour tout « i » :

$$\left(\frac{\partial V}{\partial r_i}\right)_0 = 0 \quad (12)$$

Pour de petits déplacements atomiques, les termes d'ordre supérieur du développement en série peuvent être négligés pour obtenir:

$$V = \frac{1}{2}\sum_{i=1}^{3N}\sum_{i=j}^{3N} \left(\frac{\partial^2 V}{\partial r_i \partial r_j}\right)_0 (r_i - r_i^0)(r_j - r_j^0) \quad (13)$$

L'énergie potentielle d'une protéine au voisinage d'une configuration d'équilibre possède une forme quadratique si l'on ne s'intéresse qu'aux petits mouvements des atomes de ce système. L'énergie cinétique de tout système peut aussi être mise sous une forme quadratique, sauf en présence de contraintes dépendant distinctement du temps. La solution analytique des équations du mouvement s'écrit alors comme suit :

$$r_i(t) = r_i^0 + \frac{1}{\sqrt{m_i}}\sum_{k=1}^{3N} C_k a_{ik} \cos(2\pi\nu_k t + \phi_k) \quad (14)$$

Où :
m_i : la masse associée à la coordonnée « i » et o`u
C_k : l'amplitude du mode de vibration « k »
k : la phase du mode de vibration[22].

C_k est fonction de E_k, l'énergie totale du mode k. Notamment, quand l'équipartition de l'énergie est respectée, on a :

$$C_k = \frac{\sqrt{2k_B T}}{2\pi\nu_k} \quad (15)$$

Où :

[21] Selon la relation *kBTlk*, *kB* étant la constante de Boltzman, *T* la température et *k* la constante de raideur
[22] C_k et k dépendent des conditions initiales, i .e, des positions et des vitesses des atomes à t = 0.

T : la température

k_B : la constante de Boltzmann.

L'amplitude du mode k est inversement proportionnelle à sa fréquence. En effet, les modes de fréquence inférieure à 30-100 cm^{-1} sont responsables de l'essentiel des déplacements des atomes d'une protéine [145], en utilisant l'approximation harmonique. Les 3N valeurs des fréquences sont obtenues en diagonalisant H, le Hessien de l'énergie potentielle pondéré par les masses atomiques. Autrement dit, c'est la matrice dont l'élément « ij » s'écrit:

$$\left(\frac{\partial^2 V}{\sqrt{m_i m_j}\partial r_i \partial r_j}\right)_0 \qquad (16)$$

La diagonalisation de H implique son expression dans un nouveau repère tel que, pour tout « i » différent de « j », nous obtenons:

$$\left(\frac{\partial^2 V}{\partial q_i \partial q_j}\right)_0 = 0 \qquad (17)$$

Les éléments diagonaux de la matrice transformée, ou les valeurs propres de la matrice, permettent d'obtenir les fréquences de vibration du système étudié, nous avons alors:

$$\left(\frac{\partial^2 V}{\partial^2 q_i}\right)_0 = (2\pi\nu_i)^2 \qquad (18)$$

Les nouvelles coordonnées, dites coordonnées normales, sont alors des combinaisons linéaires des coordonnées initiales, telles que, pour la coordonnée « k » :

$$q_k = \sum_i^{3N} a_{ik}\sqrt{m_i}\,(r_i - r_i^0) \qquad (19)$$

En pratique, il est toujours possible de diagonaliser « H » car c'est une matrice réelle symétrique. Dans le cas des protéines, « H » est d'ordinaire une matrice creuse, puisque les interactions entre atomes lointains ne sont en général pas calculées.

6.3. Signification et pertinence des modes normaux

Les renseignements obtenus par l'application de l'approche des modes normaux (MN) concernent surtout la direction des mouvements de grande amplitude et non leur durée. Les mouvements de grande amplitude sont généralement beaucoup plus lents que l'évaluation faite par les valeurs propres des MN. L'explication que l'on pourrait donner à cela est que la présence du solvant ralentit les mouvements, or les calculs sont souvent effectués *in vacuo*. De même, les changements de grande amplitude suivent des trajectoires harmoniques, mais pas les temps d'un oscillateur harmonique libre.

De nos jours, on considère uniquement les vecteurs propres lors de l'analyse vibrationnelle. Ceci revient à s'interroger sur le degré d'exactitude des informations fournies par le moyen de la diagonalisation de la matrice « H » (le Hessien), à savoir si la diagonalisation peut donner des vecteurs propres et des valeurs propres corrects.

Il faut inclure les molécules d'eau dans la matrice Hessienne afin d'avoir les bonnes fréquences. Or la détermination des fréquences vibrationnelles dans les milieux liquides, observées par spectroscopies Raman et infrarouge (IR), n'est possible qu'avec l'utilisation de la transformée de Fourier de la fonction de corrélation des vitesses, calculées par la dynamique moléculaire, et non par la diagonalisation de la matrice Hessienne. De plus, il faut savoir distinguer et identifier, parmi les modes calculés, ceux qui peuvent avoir un sens et une existence biologique. Pour cela, la description des mouvements fonctionnels est obtenue à l'aide d'un ou quelques modes de basse fréquence qui sollicitent peu d'énergie. Donc, la solution demeure dans la comparaison avec les données expérimentales. L'utilisation de cette hypothèse ne nous permet donc, plus l'usage du terme de modélisation prédictive.

6.4. Relation des modes normaux avec les mouvements fonctionnels

Afin de mettre en évidence ce lien, le recouvrement entre chacun des modes de vibration et le vecteur « $\overrightarrow{\Delta r}$ » sont calculés. Ce dernier est relié aux déplacements des atomes lors de changements conformationnels :

$$\overrightarrow{\Delta r} = \overrightarrow{r_b} - \overrightarrow{r_a} \qquad (20)$$

Après superposition de ces deux conformations par la méthode des moindres carrés, les deux termes correspondent à :

$\overrightarrow{\Delta r_a}$: Le vecteur position des atomes dans les conformations "a"

$\overrightarrow{\Delta r_b}$: Le vecteur position des atomes dans les conformations "b".

Il est alors supposé par cette équation que les atomes se déplacent de manière rectiligne pour aller de la conformation "a" à la conformation "b", ce qui n'est pas toujours correct. Cependant, dans de nombreux cas observés, il existe au moins un mode de vibration qui ressemble au changement de conformation obtenu expérimentalement, habituellement par cristallographie des rayons X[23].

6.5. Obstacles reliées au calcul des modes normaux

La diagonalisation de la matrice des dérivées secondes de l'énergie est la base de calcul des MN. Pénalisés par leur lourdeur, ces calculs se sont longtemps limités à des systèmes d'assez petite taille (quelques milliers d'atomes). De nos jours, l'application des calculs de MN sur des systèmes de très grande taille est désormais possible grâce aux développements algorithmiques notamment des modèles dits « gros grain » [146] (qui proposent une représentation simplifiée des protéines) ou des modèles dits « elastic network » [147]. Une autre difficulté reliée au calcul des MN se pose par rapport à la minimisation préalable de la structure étudiée. En effet, la conformation qui correspond au minimum d'énergie doit être

[23]Des comparaisons plus quantitatives furent effectuées pour les cas de la citrate synthase dimérique et de l'hémoglobine.

déterminée avec une grande précision afin d'aboutir à des solutions correctes. Ceci est particulièrement requis dans le cas de systèmes complexes. Le cas échéant entraîne des valeurs propres négatives de la matrice Hessienne, et donc des fréquences fantaisistes. Ceci tout en sachant que l'étape de la minimisation est elle-même très gourmande en temps de calcul. Toutefois, comme les méthodes de calcul des MN d'un système découlent de l'emploi de forces induites par le mouvement de deux molécules en évolution et en interaction, il n'est pas nécessaire d'atteindre l'équilibre lors de calculs de MN.

6.6. Restriction aux modes de basse fréquence

Comme rapporté dans plusieurs travaux, les modes de basse fréquence (généralement inferieure à 10 cm^{-1}) ont la capacité de décrire les mouvements de grande amplitude des systèmes de grande taille [148-150]. Ceci voudrait dire que la flexibilité des biomolécules est une propriété intrinsèque liée à leur structure, aidant les changements conformationnels nécessaires à leur fonction. Cependant, la question de la justification de l'utilisation unique des modes de basse fréquence, hormis la comparaison avec l'expérience, se pose. Des éléments théoriques peuvent y répondre. À température ambiante, ne sont bien définis que les modes normaux de fréquence supérieure à environ 10^{13} Hz, les autres, à cause de l'agitation thermique se trouvent entremêlés. De plus lors des amarrages de macromolécules, les temps de réaction sont plus lents. Ce qui conduit à des mouvements de grande amplitude, donc incompatibles avec les vibrations thermiques. Cependant, lors de calculs au moyen des MN, ces vibrations induites par l'agitation thermique sont remplacées par des déformations permanentes, portées par des forces extérieures exercées sur la molécule

6.7. Inclure les modes normaux dans une procédure de docking

À l'état libre, une macromolécule vibre autour de son état d'équilibre avec un mouvement qui peut être décrit par les modes normaux. Son énergie augmente à mesure que la macromolécule s'éloigne de cet équilibre en suivant la direction du mode normal pour atteindre une conformation instable sans la présence d'un ligand. En d'autres termes, les modes normaux ne conviendraient pas à un calcul d'interaction impliquant de grandes modifications conformationnelles car ils ne sont valides qu'autour de la position d'équilibre.

En dépit de cela, la raison avancée pour utiliser l'analyse vibratoire est que les mouvements de grande amplitude suivent des trajectoires harmoniques décrites par les modes normaux. Cependant, l'anharmonicité, souvent liée aux nombreuses faibles barrières d'énergie rencontrées lors des transitions conformationnelles (dues notamment aux mouvements rapides des chaînes latérales), n'est pas décrite par les modes normaux. Au contraire, le problème de l'anharmonicité ne se pose pas dans l'approche des MN qui permet la description de tous les mouvements. D'où l'utilisation des modes normaux de vibrations dans bons nombres d'études de docking [130, 131, 133, 151].

7. Méthodes de description de l'environnement

7.1. Représentation du solvant

Les premières applications de simulation de protéines, il y a 30 ans, ont été réalisées dans le vide [152]. Bien que les résultats obtenus aient permis de donner un aperçu de la flexibilité de la macromolécule, ils ne pouvaient pas rendre compte précisément des propriétés dynamiques de celle-ci en milieu solvaté [137]. Il est pourtant primordial de tenir compte des effets de solvant lors de l'étude des biomolécules car ils jouent un rôle essentiel dans la stabilisation de ces dernières. Pour considérer les effets de solvant lors d'une simulation, deux approches sont possibles : la solvatation implicite ou la solvatation explicite.

Le traitement implicite du solvant repose sur une forme de potentiel qui traite le solvant comme un continuum diélectrique (p. ex. via le modèle de Born généralisé) [76]. Les divers modèles de solvatation implicite se révèlent utiles pour réduire de façon significative le temps de calcul des simulations. Toutefois, bien qu'ils reproduisent fidèlement l'effet global du solvant, ils ne peuvent pas, par définition, reproduire les interactions des molécules de solvant au niveau local (p. ex. à l'interface eau/biomolécule) [70].

Le traitement explicite, lui, correspond à la modélisation d'un nombre suffisant de molécules d'eau autour des systèmes biomoléculaires d'étude au sein d'une cellule qui constitue une boîte d'eau placée dans les conditions périodiques. Bien que ce traitement s'avère très coûteux en temps de calcul (car il augmente considérablement le nombre total d'atomes à considérer), il permet de modéliser de façon plus réaliste le milieu physiologique dans lequel les biomolécules évoluent. Dans le cadre de ce travail, les simulations ont été conduites dans les conditions de solvatation explicite.

7.2. Conditions périodiques aux limites

Une méthode particulièrement adaptée pour conduire une simulation dans les conditions de solvatation explicite, tout en réduisant les effets de bord, est celle des conditions de limites périodiques [116]. Cela consiste à répliquer implicitement l'ensemble fini de particules du système d'étude, réparties dans une boîte centrale (en général cubique ou parallélépipédique) selon les trois directions de l'espace.

Les atomes dans les cellules images reproduisent les mouvements des atomes correspondants dans la cellule centrale (voir figure 19). La simulation par DM s'effectue pour les atomes de la cellule centrale uniquement en tenant compte de la présence des cellules images lors du calcul du potentiel. Le caractère pseudo-infini du système ainsi simulé contraint à effectuer certaines approximations concernant le traitement des interactions entre molécules, en particulier celle dite de "l'image minimale" qui suppose que chaque particule

« *i* » de la cellule centrale n'interagit qu'avec l'image la plus proche de toutes les autres particules « *j* ».

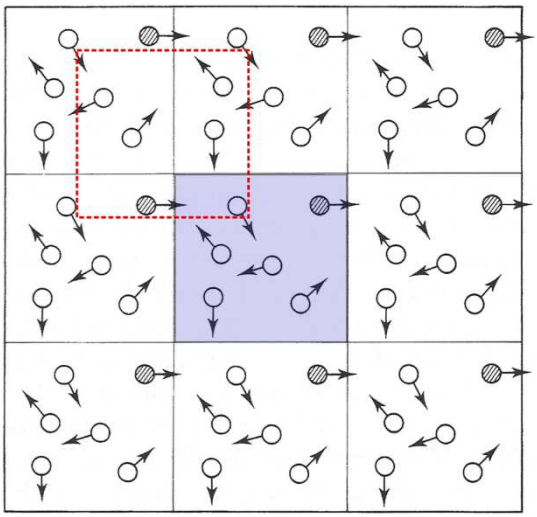


FIG. 19 : Conditions périodiques illustrées en 2 dimensions avec une boîte cubique. La boîte en bleu représente la cellule centrale répliquée ; la boîte en pointillés rouges symbolise l'image minimale.

7.3. Prise en compte des interactions entre atomes non-liés

La complexité des algorithmes de DM, en l'absence d'optimisation, est en $O(n^2)$; cela limite rapidement l'intervalle d'espace et de temps accessible des simulations. En effet, alors qu'à un ensemble de particules liées ne correspond qu'un nombre limité d'interactions à prendre en compte dans le calcul (d'où une complexité en n du terme $V_{liés}$), le nombre de paires de particules non-liées croît, lui, en $O(n^2)$. L'optimisation du calcul des interactions entre atomes non-liés est par conséquent centrale afin d'améliorer l'efficacité des algorithmes de DM.

Afin de limiter le temps de calcul nécessaire au calcul des potentiels entre les paires d'atomes non-liés s'exerçant à longue distance, il est courant de limiter la prise en compte de ces interactions pour un atome donné à ses voisins les plus proches, c'est-à-dire à ceux inclus dans une sphère dite de troncature ("*cut-off*"), centrée sur l'atome en question.

Dans le cas de l'utilisation des conditions périodiques aux limites, le rayon de troncature doit être inférieur ou égal à la moitié du plus petit côté de la cellule centrale, afin qu'il n'y ait pas plus d'une image de chaque atome prise en compte. Ainsi, la complexité du calcul n'est plus que de $O(n)$ car le nombre total d'atomes au sein de la cellule centrale est borné. Mais, si l'emploi d'un cut-off de l'ordre de 8-10 Å est acceptable pour les interactions de van der Waals, ce n'est pas le cas de l'interaction Coulombienne en 1/r, même dans le cas de faibles charges atomiques partielles. Ainsi, le problème des interactions entre atomes non-liés se réduit principalement à celui du traitement des interactions électrostatiques [153].

7.4. Calcul des interactions électrostatiques

L'emploi d'un cut-off abrupt sur les interactions non-liées provoque des artefacts importants dans les calculs de DM [154]. Ceci peut être atténué par l'emploi d'un cut-off progressif, défini en réalité par deux seuils de cut-off, délimitant ainsi l'intervalle de prise en charge des interactions électrostatiques de façon décroissante de 100% à 0%. Dans les conditions périodiques, la technique de sommation d'Ewald [155] permet un calcul exact, mais au prix d'une complexité en $O(n^{3/2})$[24]. Pour pallier à cela, une des solutions les plus utilisées de nos jours est la variante SMPE (*Smooth Particle Mesh Ewald*) [156] (complexité en $O(n\log(n))$), qui est d'ailleurs implémentée dans le programme NAMD [105] utilisé pour conduire les simulations de ce travail.

[24] Cette technique introduit comme contrainte supplémentaire la neutralité électrique de l'ensemble du système. Si nécessaire, cette neutralité est obtenue artificiellement par l'emploi de contre-ions placés sur les bords de la cellule centrale.

III- Les autres outils *in silico* utilisés dans ce travail

Comme nous l'avons développé dans la première section de ce chapitre, les approches de conception de médicaments aussi bien « structure-based » que « ligand-based » s'appuient sur des techniques de criblage virtuel à haut débit. Nous allons aborder dans cette section l'aspect théorique des techniques utilisées au cours de nos travaux

1. Analyse en groupes ou clustering

L'analyse en groupes ou clustering se dit d'un assignement d'un ensemble d'observations similaires en sous-ensembles appelés groupes ou clusters. L'analyse en groupes est une méthode d'apprentissage non-supervisé et une technique couramment utilisée pour l'analyse des données statistiques dans plusieurs domaines tels que l'exploration de données, l'apprentissage machine, l'analyse d'images, la reconnaissance de formes et la bioinformatique.

La classification automatique peut être considérée comme synonyme du clustering en termes de similitude de sens, tout comme la taxonomie numérique, l'analyse typologique et la botryologie[25]. Mais la classification ou l'apprentissage supervisé, dans son acceptation générale, diffère du clustering dans le sens où, dans la première approche, il est nécessaire d'avoir un ensemble de cas d'apprentissage pour chaque groupe ou classe [157]. Au contraire, le clustering ou l'apprentissage non-supervisé assume l'absence de schémas de catégorisation préexistante puisqu'aucun modèle ne fournit d'exemples d'ensembles d'apprentissage avec étiquettes de classes. En d'autres termes, les clusters sont induits à partir des données.

Une des étapes les plus importantes dans les clusterings est la sélection d'une mesure de distance qui déterminera le degré de similitude calculé entre les éléments. Cette mesure influencera la forme des clusters à cause de l'éloignement ou du rapprochement des points, les uns par rapport aux autres, qui les constituent.

[25] Du grecque « botrus » signifiant grappe, est une méthode de séparation en groupes d'objets interconnectés.

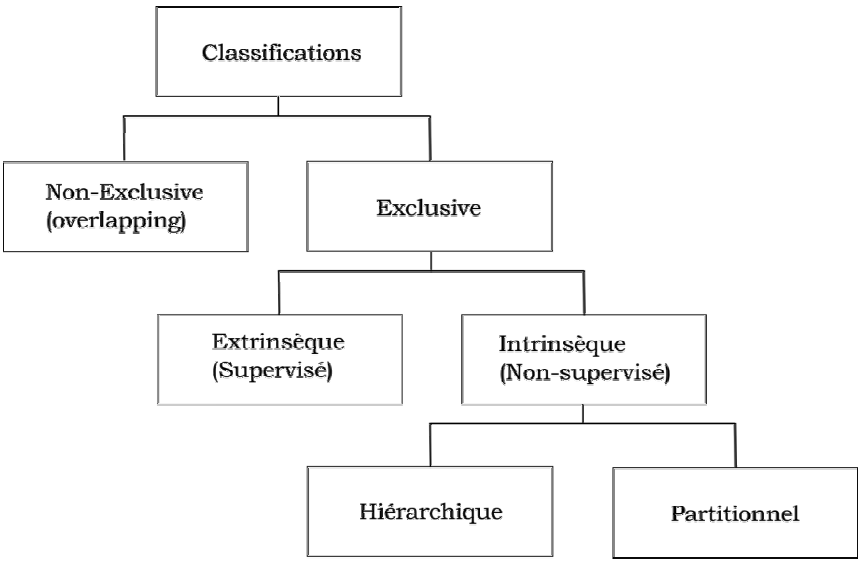


FIG. 20 : Arborescence des types de classification

Les fonctions de distances les plus connues sont :

- La distance Euclidienne appelée aussi distance à vol d'oiseau : c'est la mesure la plus utilisée.
- Distance de Manhattan ou encore taxi-distance : c'est la distance parcourue d'un nœud de réseau à un autre en utilisant les déplacements horizontaux et verticaux du réseau.
- La norme maximale ou norme infinie
- La distance de Mahalanobis : c'est une métrique basée sur la corrélation entre des variables par lesquelles différents modèles peuvent être identifiés et analysés. elle corrige les données pour différentes échelles et corrélations dans les variables.
- L'angle entre 2 vecteurs peut être utilisé comme une mesure de distance dans le regroupement de données à grandes dimensions.
- La distance de Hamming : elle mesure le nombre minimum de substitutions demandées pour changer un membre dans un autre. Elle permet de quantifier la différence entre deux séquences de symboles. À deux suites de symboles de même longueur, elle associe le nombre de positions où les deux suites diffèrent. C'est une métrique très utilisée en informatique, en traitement du signal et dans les télécommunications.

1.1. Les types de clustering

1.1.1. Clustering hiérarchique

1.1.1.1. Définition

Ce type de clustering est en général le plus utilisé. Il crée une hiérarchie de clusters qui peuvent être représentés en arborescence appelée aussi dendrogramme (en grec *dendron* : arbre, *gramma* : dessiner). L'application de l'image de l'arbre consiste à ce que la racine soit considérée comme le cluster contenant tous les objets, donc l'ensemble des propriétés

observées, et que les feuilles correspondent aux propriétés individuelles observées. Plus les feuilles se ramifient, plus les propriétés observées et calculées se personnalisent et s'individualisent.

Différentes métriques peuvent être utilisées comme mesure de similarité entre plusieurs observations. Le choix déterminant le regroupement des données en clusters est défini par un critère de liaison considéré comme une fonction des distances par couple entre les observations données.

Le nombre de clusters voulus dépend de la hauteur à laquelle on veut couper la hauteur de l'arborescence. En d'autres termes, plus le clustering est précis, plus le nombre de clusters est grand avec un minimum d'individus.

Le clustering hiérarchique peut se baser sur deux sortes d'algorithmes : celui dit d'agglomération, qui part des ramifications pour regrouper successivement les clusters ensemble, et celui appelé algorithme de division, qui part de la racine et divise alternativement les clusters.

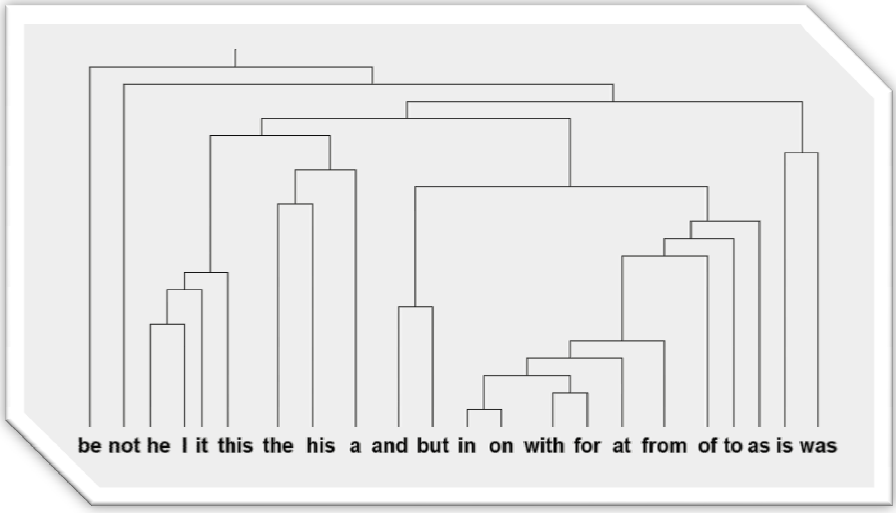


FIG. 21: Exemple de dendrogramme (fréquence des mots dans le "Brown corpus")

1.1.2. Mode de fonctionnement

Pour un set de données à clusteriser (à regrouper), un processus basique de clustering hiérarchique fut défini déjà en 1967 par Johnson comme suit :

- Commencer par attribuer à chaque objet un cluster, pour un clustering d'agglomération. La métrique de similarité entre les différents objets d'un cluster est représentée par une distance.
- Trouver les paires de clusters similaires ou à distances proches et les fusionner de manière à avoir moins de clusters.
- Calculer les distances (similarités) entre les nouveaux clusters issus de la précédente étape et les comparer aux anciennes distances.
- Répéter les deux précédentes étapes et fusionner les clusters ayant les mêmes distances jusqu'à arriver à un cluster unique (regroupement de tous les objets)

Le clustering hiérarchique de division s'effectue de la même manière. Il faut attribuer initialement à l'ensemble des objets un seul cluster qui se divisera en plusieurs tant que des dissimilarités existeront entre les objets.

Comme le clustering hiérarchique fournit plus d'informations, il est préférable de l'utiliser dans le cas d'analyses de données détaillées. Cependant, l'algorithme devant calculer n×n coefficients de similarité à chaque itération du processus est considéré comme moins rapide

1.1.3. Clustering non-hiérarchique (clustering plat ou partitionnel)

1.1.3.1. Définition

La plupart des algorithmes des clustering non-hiérarchiques sont itératifs. Ils commencent avec un ensemble de clusters initial et les améliorent de façon itérative en utilisant un schéma de réaffectation.

Le centre initial des clusters peut être déterminé par application d'algorithmes de clustering hiérarchique sur un sous-ensemble d'objets à regrouper surtout dans des cas de mauvais comportement d'ensembles d'objets. Le clustering peut être arrêté si l'un de ces critères se réalise : une similarité de moyennes de groupes, la vraisemblance des données par rapport aux clusters, la satisfaction du principe de la longueur de description minimale [158], l'existence d'informations mutuelles entre les clusters adjacents.

L'utilisation des algorithmes de clustering non-hiérarchique se fait préférentiellement quand l'ensemble des données est grand et l'efficacité est nécessaire. Dans la majorité des études rencontrées, l'algorithme le plus utilisé dans le clustering non-hiérachique reste « *k-means* ». En effet, l'algorithme est conceptuellement le plus simple. Il utilise par exemple les distances Euclidienne comme métrique.

1.1.3.2. Mode de fonctionnement

L'algorithme de « *k-means* » attribue chaque point (objet) à un centre de cluster qui se trouve dans le voisinage immédiat de ce dernier (proche du centroïde). Le centroïde est le milieu de tous les points se trouvant dans un cluster. Ses coordonnées sont la moyenne arithmétique des coordonnées des points dans le cluster. Les différentes étapes de l'algorithme peuvent être résumées comme suit :

- Commencer par choisir le nombre de clusters à générer
- Générer aléatoirement le nombre de clusters n et déterminer les centroïdes, ou générer directement n points aléatoires considérés comme des centres de clusters.
- Affecter chaque point de l'ensemble des données au cluster dont le centre est le plus proche.
- Recalculer les nouveaux centroïdes
- Répéter les deux précédentes étapes jusqu'à satisfaction d'un critère de convergence (généralement jusqu'à ce que les affectations des points ne changent pas)

1.1.3.3. Autres types de clustering non hiérarchique

Dans la catégorie des clusterings non-hiérarchique il existe également le clustering flou. Comme dans la logique floue, au lieu d'appartenir complètement à un seul cluster, chaque point (objet) possède un degré de dépendance à différents clusters. Donc les points se situant aux bords d'un cluster y sont contenus à un degré moindre que les points se situant au centre de ce cluster. L'algorithme de minimisation minimise les divergences internes des clusters mais le minimum reste (un minimum) local. Ainsi, tout comme pour les problèmes rencontrés avec « *k-means* », les résultats de l'algorithme de clustering flou dépendent du choix (ou affectation) initial. Pour cela, l'algorithme espérance-maximisation, qui a de meilleures propriétés de convergence, est en général préféré à l'algorithme flou.

Le Quality-Threshold (QT) clustering (clustering seuil de qualité) [159] est une méthode alternative au partitionnement de données créé pour le regroupement de gènes. Il nécessite plus de puissance de calcul que « *k-means* », mais ne demande pas de spécification de nombres de cluster, donnant ainsi les mêmes résultats à chaque exécution.

La classification qui utilise l'Analyse en Composante Principale (ACP) est également utilisée dans la distinction entre différents groupes parmi des objets donnés. Cette méthode d'analyse de données repose à la fois sur l'approche géométrique et statistique.

1.2. Les domaines d'application du clustering

Parmi les larges applications de la biologie, différentes disciplines utilisent les approches du clustering. Nous énonçons les plus courants :

- Dans le domaine de l'imagerie, la classification des données peut prendre différentes formes basées sur la dimensionnalité des données (clustering par points, régions ou volumes)
- Dans le domaine des plantes et de l'écologie animale, le clustering est utilisé pour faire des comparaisons de communautés (ou assemblages) spatiales et temporelles d'organismes dans des environnements hétérogènes. Il est également utilisé dans la systématique des plantes pour générer des phylogénies artificielles ou des clusters d'organismes (individus) par rapport aux espèces, genres ou différents niveaux qui partagent certains nombres d'attributs.
- Dans le domaine de la bioinformatique et de la biologie computationnelle :
    - La transcriptomique[26] utilise le clustering pour construire des groupes de gènes avec les profils d'expression liés. Souvent, de tels groupes contiennent des protéines fonctionnellement liées telles que les enzymes pour certaines voix de signalisation ou des gènes co-régulés.
    - L'analyse de séquences utilise également les méthodes de clustering pour grouper les séquences homologues dans des familles de gènes. C'est un concept très important en bioinformatique et en biologie évolutive en général.

[26] Il s'agit del'étude de l'ensemble des ARN messagers produits lors du processus de transcription d'un génome

  - Les plateformes de génotypage à haut débit utilisent le clustering pour affecter automatiquement des génotypes.
- Le clustering de données est également utilisé en chemo-informatique dans les études basées sur le QSAR ainsi que la modélisation moléculaire.

2. Le docking ou amarrage moléculaire

Nous nous limiterons à une description assez générale des techniques de docking afin de situer le contexte de ce travail. Cette section n'a donc pas vocation de s'ajouter aux nombreuses revues publiées sur les caractéristiques détaillées des différents programmes de docking ou sur l'évaluation de ceux-ci [160-171].

2.1. Principe

Le docking est utilisé pour prédire la structure du complexe intermoléculaire résultant de l'association entre au moins deux molécules. Quand il s'agit de deux protéines, on parle de docking protéine-protéine [172], par opposition au docking protéine-ligand que nous avons utilisé pour ce travail.

Le processus du docking est itératif et chaque passe de calcul s'articule en deux étapes. Tout d'abord, une portion limitée de l'espace des conformations du complexe protéine-ligand est explorée afin d'améliorer la *pose* (orientation et/ou conformation) du ligand dans le site actif de la protéine. L'étape suivante fait intervenir une fonction de score qui évalue la qualité de la pose générée [173]. Cette estimation *in silico* de l'affinité du ligand pour la cible est basée sur un examen simplifié des interactions entre les deux partenaires. La répétition du cycle de recherche associé à l'estimation du score guidant l'exploration de l'espace, doit assurer la convergence de l'algorithme vers un état représentatif du minimum global de l'énergie libre d'association.

Avant de décrire plus en détail les deux composantes du principe de docking protéine-ligand évoquées ici, nous présenterons les principaux types de représentation du système employés par les diverses techniques de docking. Nous verrons ensuite en quoi la nature de l'espace de recherche permet de classer les algorithmes de docking en plusieurs familles distinctes. Enfin, nous discuterons de la signification des résultats obtenus dans le contexte du criblage virtuel et des perspectives générales du docking.

2.2. Représentation du système

La plupart des méthodes de docking reposent sur une représentation simplifiée qui peut réduire la protéine aux atomes du site actif et du voisinage de celui-ci et dans laquelle les molécules du solvant ne sont pas représentées explicitement. Les modèles utilisés sont généralement analogues à ceux de la mécanique moléculaire (cf. Partie 2, Section I), dans lesquels chaque atome des structures moléculaires est représenté comme un point unique sur lequel ses propriétés sont projetées. La majorité des méthodes de docking utilise des types atomiques simplifiés par rapport aux définitions d'un champ de force courant. Dans certains cas, des pseudo-atomes supplémentaires sont ajoutés afin de simplifier le traitement de tous les types d'interaction entre protéine et ligand, par exemple les paires électroniques libres afin de mieux détecter la formation d'éventuelles liaisons hydrogène.

Moins fréquemment, les ligands et/ou les sites actifs peuvent être modélisés par l'intermédiaire de surfaces. Différents types de représentation peuvent être envisagés, par exemple un ensemble de points dans l'espace ou bien une série de fonctions harmoniques sphériques [174, 175].

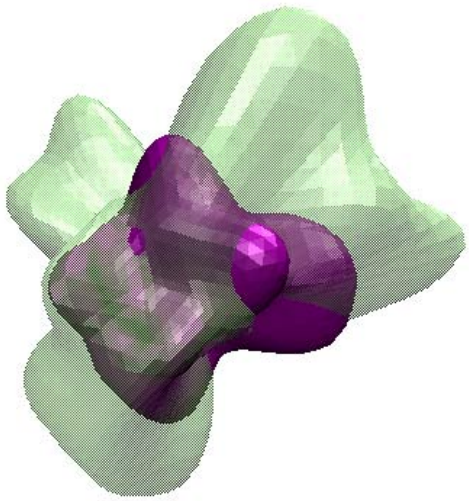


FIG. 22 : Complémentarité de surface entre un site de liaison et son ligand (représentation basée sur les harmoniques sphériques)

Les propriétés des ligands et/ou des sites actifs réparties dans l'espace peuvent également être représentées à travers leur projection sur une grille. Ce type de représentation est souvent employé quand un ensemble de conformations de la protéine cible (pouvant être obtenu expérimentalement ou par simulation de dynamique moléculaire par exemple) est utilisé pour considérer sa flexibilité sans l'implémenter explicitement au niveau de l'algorithme de recherche. Ces conformations peuvent être combinées en une seule pseudo-conformation de diverses façons (moyennées, unifiées, *etc.*). Les programmes DOCK [176], AutoDock [177] et FlexE [178] permettent l'emploi de ce type d'approche.

Il est aussi possible d'effectuer un docking sur chacune des multiples conformations [179, 180]. Cette approche systématique peut s'avérer coûteuse mais a l'avantage de ne pas dépendre des paramètres et de la pertinence d'un programme spécifique.

2.3. La recherche de la pose optimale

Un système biologique réduit à la zone d'interaction protéine-ligand fait intervenir trois acteurs dans le phénomène de reconnaissance moléculaire : la protéine, le ligand et le solvant. Un algorithme de docking idéal prendrait en compte tous les paramètres associés à ces trois composantes, tels que la complète flexibilité du ligand et de la protéine, les effets de solvant, *etc.* Toutefois, une telle approche implique l'exploration d'un espace composé d'un nombre très important de degrés de liberté. La résolution du docking par une recherche exhaustive nécessiterait alors des temps de calcul considérables. Pour cette raison, la complexité du système est généralement réduite au niveau de la modélisation structurale. Les approximations les plus courantes à ce niveau consistent à considérer la protéine comme une entité rigide et à ne pas représenter explicitement les molécules du solvant.

On peut distinguer plusieurs niveaux de représentation du système dont les approximations évoluent conjointement avec le progrès des moyens informatiques au fil des ans. Au début des années 1980, quand l'approche de modélisation moléculaire par docking a été pour la première fois étudiée [181], Kuntz et ses collaborateurs ont proposé la stratégie du tout rigide dans laquelle l'exploration de l'espace des positions (discrétisé par l'emploi d'une grille) se limite au positionnement du ligand et n'exploite donc que six degrés de liberté élémentaires (rotations et translations). Depuis, l'essor des moyens de calcul a permis de considérer le système d'étude comme semi-flexible : la protéine est traitée de façon rigide, contrairement au ligand dont l'espace conformationnel est pris en compte. Cette approche reste la plus employée aujourd'hui, les algorithmes tenant compte explicitement de l'espace conformationnel du site actif étant encore en cours de développement.

Les différentes approches du docking que nous allons à présent détailler se distinguent au niveau de leurs conditions d'application et de la nature des informations qu'elles peuvent fournir. La pertinence du choix d'un programme de docking donné repose en premier lieu sur l'adéquation entre ces caractéristiques et celles du système étudié. L'efficacité de l'algorithme choisi sera par ailleurs un compromis entre rapidité d'exécution et précision des résultats.

2.3.1. Docking rigide

Dans le cas des méthodes de docking rigide, la recherche de la pose optimale se limite au positionnement. Cette opération consiste en la recherche exhaustive dans l'espace discrétisé des 6 degrés de liberté. Certains programmes, s'ils n'appartiennent pas à la famille

des techniques de docking rigide, utilisent plusieurs étapes successives d'optimisation dont les premières peuvent s'apparenter à du docking rigide. Par exemple, le programme Glide [182] utilise initialement, dans son approche multi-étapes, une recherche systématique pour positionner le ligand de façon approchée au sein du site actif de la protéine.

Il est possible de considérer indirectement la flexibilité des ligands en utilisant des programmes de docking rigide. Pour cela, un jeu de conformères de basse énergie pour chaque molécule à tester sur le site actif ciblé peut être généré efficacement par les programmes tels qu'OMEGA [183] ou Catalyst [184]. Ces données sont calculées une fois pour toutes, réutilisables pour d'autres criblages une fois stockées. Certains programmes tels que FLOG [185], FRED [183] et EUDOC [186] travaillent sur un ensemble de conformères pré-calculés par un programme tiers ou générés à la volée par leurs algorithmes. Une telle approche est nécessaire dans le contexte du docking rigide car il est admis que la conformation complexée d'un ligand peut varier considérablement par rapport à sa conformation « libre » [187].

2.3.2. Docking semi-flexible

Lorsque l'espace conformationnel des ligands est exploré, le nombre de degrés de liberté de l'espace de recherche peut être conséquent dans le cas de molécules flexibles. Dans un tel contexte, l'emploi de méthodes de recherche exhaustives apparaît souvent inapproprié car nécessitant des simplifications importantes au niveau de l'échantillonnage. D'autres algorithmes, dits de fragmentation, sont employés pour construire de façon incrémentielle le ligand au sein du site actif de la protéine. L'espace des conformations du ligand est alors restreint au voisinage d'un ensemble initial d'états simplifiés. Cette stratégie de recherche par construction, qui se présente sous diverses variantes [161], est notamment adoptée par les programmes DOCK [181, 188], FLExX [189] et Hammerhead [190].

Les programmes de docking semi-flexible considérés comme les plus efficaces emploient des méthodes de recherche aléatoires ou stochastiques. L'exploration de l'espace de recherche se fait de façon plus ou moins aléatoire et les états générés sont soit acceptés, soit rejetés, suivant des règles spécifiques. On distingue trois principales classes de méthodes aléatoires : les méthodes de Monte Carlo (cf. Partie 2, Section I), les méthodes évolutionnaires basées sur des algorithmes génétiques [191] et les méthodes de recherche Tabou [192].

Les algorithmes basés sur la méthode de Monte Carlo génèrent aléatoirement des états du système acceptés ou rejetés sur la base probabiliste de la fonction de Boltzmann. Les principaux programmes utilisant cette méthode sont ICM [193], QXP [194] et MCDOCK [195].

Les algorithmes génétiques [191] s'inspirent des théories de l'évolution pour sélectionner les états. Une population initiale aléatoire de confirmations du ligand dans le site actif de la protéine est définie et les degrés de liberté à explorer sont assimilés à un jeu de *gènes*. L'échantillonnage de l'espace se fait ensuite par des opérations génétiques (mutations, croisements et migrations) sur la population. La sélection des individus-conformères générés est basée sur leur capacité d'adaptation à l'environnement (la fonction de score). Malgré leur efficacité algorithmique, les algorithmes génétiques appliqués au docking ont parfois tendance à sélectionner des minima locaux. Pour pallier cela, une solution consiste à répéter un même calcul plusieurs fois afin de maximiser les chances d'obtenir, à l'issue de cette procédure, au moins un résultat satisfaisant (structure proche du minimum global). Les programmes de docking les plus connus implémentant un algorithme génétique sont AutoDock [196], GOLD [197] et DARWIN [198].

Le principe de base des méthodes de recherche Tabou, utilisées par exemple dans PRO_LEADS [199], est de pouvoir prendre en compte les régions de l'espace ayant déjà été visitées (par des calculs de similarité entre ligands, p. ex. calcul de RMSD). La recherche dans les régions inexplorées est privilégiée, réduisant considérablement la taille de l'espace à explorer.

2.3.3. Docking flexible

De nombreuses études utilisant le docking par l'approche ligand flexible/protéine rigide ont montré que cette stratégie semi-flexible conduisait à des résultats concluants [200, 201]. Toutefois, les réussites issues de son utilisation concernent surtout l'étude de protéines relativement rigides. À la suite de leur liaison avec un ligand, de nombreuses protéines peuvent pourtant subir des réarrangements structuraux de plus ou moins grande amplitude. Négliger cet aspect, désigné sous le terme d'"*induced fit*" [202], peut avoir des conséquences fâcheuses sur la pertinence des résultats issus d'un docking [203]. Même concernant des changements conformationnels mineurs, la flexibilité du site actif de la protéine peut avoir une grande influence dans le phénomène de reconnaissance moléculaire avec un ligand [204].

La mise au point d'algorithmes de docking prenant en compte explicitement la flexibilité de la protéine est toutefois une tâche délicate car le nombre de degrés de liberté associé à une telle représentation peut s'avérer très important (>> 50). Les programmes incorporant, au moins partiellement, la flexibilité du site actif adoptent des stratégies assez diversifiées dont certaines d'entre elles sont évoquées ici à titre illustratif.

Le docking rigide permissif (dit "*soft docking*") considère la flexibilité de la protéine de manière indirecte, en atténuant certains termes de répulsion. Cela peut permettre au ligand de pénétrer légèrement la surface de la protéine en prévision des réarrangements qui auraient eu lieu lors de l'association de partenaires flexibles [205]. Une telle approche indirecte de la

flexibilité de la protéine est cependant limitée pour couvrir le spectre des effets d'*induced fit* qui peuvent être observés, par exemple par dynamique moléculaire.

Une librairie de rotamères pour un ensemble de chaînes latérales de la protéine autorisées à être flexibles peut être utilisée dans une approche plus explicite [206, 207]. ROSETTALIGAND [208] et Glide [182] sont deux exemples de programmes utilisant cette stratégie.

Enfin, de nouveaux programmes de docking qui traitent la flexibilité du site actif en employant la même stratégie que celle des ligands (chaque rotamère constitue un degré de liberté), sont actuellement en développement [209, 210]. Certains reposent sur des algorithmes innovants tels que la simulation de colonies de fourmis ou l'optimisation par essaims de particules [211].

2.3.4. L'ensemble Docking

Le nombre de structures de protéines disponibles a explosé au cours de ces deux dernières décennies, plus particulièrement celles qui possèdent une pertinence clinique. En conséquence, les besoins des modélisateurs ont eux aussi évolué du stade de la recherche d'un modèle disponible pour la cible vers le stade d'une meilleure utilisation de l'ensemble des données disponibles pour la cible d'intérêt. De plus, dans certaines familles de protéines la modélisation des boucles reste un grand défi, plus particulièrement dans le cas des kinases. La solution apparente aux problèmes cités est donc d'effectuer le docking sur plusieurs modèles de protéines et c'est ce qui définit « l'ensemble docking ».

D'autre part, il est admis que les protéines sont des objets flexibles pouvant adopter plusieurs conformations en équilibre dynamique. Ceci est particulièrement vrai pour les kinases telles que c-Met [212]. Il est donc désormais nécessaire de prendre en compte cette flexibilité dans les calculs de docking moléculaire. Mais cela rend ces calculs beaucoup plus complexes du fait de considérer, en plus des possibilités conformationnelles du ligand, celles de la protéine. Une technique intéressante pour surmonter ce problème est de considérer la flexibilité de la protéine cible par l'intermédiaire d'un ensemble de conformations exprimant cette flexibilité.

L'ensemble docking traite donc à la fois d'un ensemble de protéines, chacune représentée par un ensemble de conformères. Cette méthode d' « ensemble docking » est de plus en plus employée [213, 214]

Le logiciel « GOLD » que nous avons utilisé permet dans sa dernière version de traiter de tels problèmes d'« ensemble docking ». Il résout l'énorme problème combinatoire posé[27] par un algorithme génétique astucieux.

[27] Par exemple : docking de 100 000 molécules sur 10 protéines cibles, chacune décrite par 20 conformations

La méthode employée pour l'« ensemble docking » nécessite au préalable la préparation des protéines cibles en superposant les structures à docker par rapport à une référence structurale (centre de gravité ou liste d'atomes définissant un centre actif commun). Pour cela, une analyse structurale[28] sur les conformations en question peut s'avérer utile, en particulier celle concernant le site de liaison.

2.4. L'évaluation du score des poses

L'évaluation et le classement des différentes conformations du ligand positionné dans le site actif de la protéine constituent le second aspect crucial du docking, après la nature de la recherche conformationnelle. Une fonction de score doit théoriquement pouvoir distinguer, parmi les différentes poses générées pour un ligand, celles qui correspondent aux modes de liaison les plus représentatifs de la reconnaissance moléculaire. Ainsi, elle doit permettre, sur la base d'une pose optimale proposée, de différencier les molécules bioactives des autres composés inactifs vis-à-vis de la protéine.

Des techniques basées sur le calcul d'énergie libre ont été développées et constituent un moyen quantitatif rigoureux pour estimer l'affinité de liaison d'un ligand pour une protéine [215, 216]. Bien qu'elles soient très précises, leur coût (en temps, en contraintes et en expertise) ne les rend pas appropriées pour une application de docking dans le cadre d'un criblage virtuel. Pour représenter le meilleur compromis entre vitesse et précision, les fonctions de score implémentées dans les programmes de docking sont donc basées sur la simplification des phénomènes impliqués dans la reconnaissance moléculaire, en particulier de ceux qui sont délicats à évaluer en dehors des calculs d'énergie libre (p. ex. l'entropie).

Les différentes fonctions de score implémentées dans les programmes de docking ont fait l'objet de nombreuses publications [173, 217-219]. Elles sont généralement classées suivant trois catégories : les fonctions de score basées sur un champ de force, les fonctions de score empiriques et celles basées sur des connaissances statistiques. Enfin, plusieurs fonctions de score de nature différente peuvent être combinées pour former des fonctions de score dites de *consensus*.

2.4.1. Les fonctions de score basées sur un champ de force

Les champs de force dans leur forme standard (cf. Partie 1, II. Mécanique moléculaire) évaluent la somme de deux énergies : l'énergie entre atomes liés au sein d'une molécule donnée (énergie interne) et l'énergie entre atomes non-liés. Ce deuxième terme correspond au terme principal de l'énergie d'interaction protéine-ligand dans le cas d'une application à un

[28] Analyse basée sur la comparaison de RMSD et les mouvements des plus importantes parties des conformations.

problème de docking. La plupart du temps, les fonctions de score basées sur un champ de force ne considèrent qu'une conformation donnée de la protéine. Ainsi, si l'on compare l'activité de deux ligands par rapport à cette même conformation, cela permet de faire abstraction du terme d'énergie interne de la protéine qui s'annule dans l'expression de la différence d'énergie libre d'interaction entre les deux ligands.

L'affinité d'un ligand donné pour le site actif, mesuré en tant qu'énergie d'interaction, correspond à la somme des énergies d'interaction de van der Waals (souvent représentée par un potentiel de Lennard-Jones) et électrostatiques (potentiel de Coulomb). On peut ajouter à la fonction de score un terme d'énergie interne du ligand. Tous ces termes peuvent s'exprimer à partir des paramètres du champ de force considéré.

De telles fonctions de score présentent certaines limitations qui s'ajoutent à celles induites par la représentation du système (p. ex. la non représentation explicite du solvant). En particulier, les effets d'entropie, qui peuvent varier d'un ligand à l'autre pour un site actif donné, aussi bien que pour un ligand donné d'un site actif à l'autre, ne sont pas pris en compte. Seule la contribution enthalpique de l'énergie libre d'interaction est ainsi prise en compte.

Les fonctions G-Score [87] (basée sur le champ de force de Tripos [87]) et celle implémentée dans AutoDock [220] (basée sur le champ de force AMBER [94]) sont des exemples de ce type de fonction de score.

2.4.2. Les fonctions de score empiriques

Ce type de fonction de score approxime l'énergie libre de liaison en sommant de façon pondérée différents termes d'interaction dérivés de paramètres structuraux. Les différents poids de la fonction de score sont ajustés pour reproduire en priorité des données expérimentales, telles que les constantes de liaison tirées d'un jeu d'entraînement de complexes protéine-ligand.

La plupart des programmes de docking implémentent ce type de fonction de score témoignant de leur efficacité (en terme de rapport précision/rapidité). Cependant, le principal inconvénient de ces fonctions empiriques est leur forte dépendance aux données utilisées pour les calibrer qui, en cas de mauvaise paramétrisation, peut limiter leur transférabilité sur des systèmes différents. Parmi les principales fonctions de score empiriques, on peut citer : ChemScore [221], PLP [222], et LigScore [223].

2.4.3. Les fonctions de score basées sur des connaissances statistiques

Ces fonctions de score sont construites à partir de règles fondées sur une analyse statistique des complexes protéine-ligand résolus expérimentalement. Elles partent du principe que les distances interatomiques les plus représentées statistiquement dans les complexes constituent des contacts énergétiques favorables et, qu'à l'inverse, les plus rares représentent des interactions moins stables. Ainsi, leur paramétrisation dépend de la quantité d'informations expérimentales disponibles et on doit leur apparition à la profusion de données structurales accessibles dans des bases de données telles que la *Protein Data Bank*. Les exemples populaires de ces fonctions de score sont PMF [224], SMoG [225] et DrugScore [226].

2.4.4. Les fonctions consensus

Ces fonctions hybrides combinent les résultats issus de diverses fonctions de score. On estime qu'il est possible de compenser partiellement les faiblesses intrinsèques de chacune des fonctions de score employées, évitant leurs erreurs individuelles et ainsi d'augmenter la probabilité d'identifier des composés actifs [227]. Cependant, si les termes des différentes fonctions de score sont fortement corrélés, l'intérêt du consensus devient limité car il peut entraîner une amplification des erreurs, au lieu de les atténuer. Ces fonctions consensus ont récemment fait l'objet d'une revue exhaustive [228].

2.4.5. Autres types de fonctions de score

Les méthodes de docking reposant sur des surfaces utilisent des fonctions de score qui sont adaptées et principalement centrées sur la complémentarité géométrique surface-surface et auxquelles il peut être adjoint une estimation d'interactions sur le modèle des fonctions de score plus conventionnelles. Par exemple, LigandFit [229] génère, par Monte Carlo, les conformations du ligand dont les formes sont ensuite comparées à celles du site actif. D'autres programmes, tels que FRED [183, 230] ou SHEF [231], comparent la forme de chacun des conformères, générés au préalable (p. ex. avec OMEGA [183]), à la forme du site actif de la protéine. Cette approche est celle utilisée dans la première étape de la stratégie de la plateforme de criblage VSM-G décrite dans ce travail.

2.5. Pertinence des résultats de docking et perspectives

La façon la plus simple et la plus courante d'exploiter les résultats d'un criblage virtuel par docking consiste à se limiter au classement des molécules candidates suivant la valeur de la fonction de score pour la structure optimale générée. Dans ce contexte, une fonction de score est considérée comme efficace si elle est capable de discriminer les composés actifs des non-actifs vis-à-vis de la cible et, mieux, de classer les molécules par ordre d'affinité. On

considère alors qu'afin d'améliorer la pertinence des programmes de docking, il convient en priorité d'améliorer celle des fonctions de score, sans entraîner un surcoût de calculs trop important.

Une autre approche plus qualitative consiste à considérer que le but du docking est la production d'une structure géométrique pertinente du complexe protéine-ligand, laquelle peut servir de base pour des calculs ultérieurs d'optimisation. Dans ce cas, la corrélation du score avec l'énergie libre d'interaction pourrait s'avérer secondaire, du moment que la fonction de score permet à l'algorithme de recherche de converger avec une forte probabilité vers une structure pertinente. Dans une optique de criblage, il s'avèrera alors nécessaire de mettre en œuvre une étape d'optimisation *post-docking* (p. ex. via un minimiseur basé sur un potentiel MM/PBSA [232]) afin de discriminer les molécules candidates de façon automatisée (et éventuellement fournir une valeur de score mieux corrélée à l'énergie libre d'interaction).

Il convient aussi de tenir compte des interactions entre fonction de score et recherche conformationnelle au sein du processus de docking. Une amélioration du score doit en effet être évitée si elle est susceptible, par exemple, de faire converger trop vite la recherche conformationnelle vers un minimum local non pertinent du point de vue de l'interaction protéine-ligand. De telles considérations peuvent rendre le développement de techniques de docking efficaces particulièrement délicat.

Même si, du point de vue de la précision, certains programmes actuels se distinguent [168], il peut être difficile de déterminer dans l'absolu qu'un programme de docking donné « fonctionne mieux » qu'un autre ou de tirer des conclusions générales sur leur pouvoir prédictif. Ceci pour plusieurs raisons [233, 234]. En effet, chaque programme de docking est unique par la représentation du système, la méthode de recherche et la fonction de score employée. Cette spécificité se répercute sur sa pertinence en termes de rapport entre précision et efficacité de calcul, rapport qui dépend aussi des paramètres employés pour les calculs et de la nature de la cible étudiée. C'est pourquoi on assiste, par exemple, à l'émergence de fonctions de score spécifiques à certaines classes de protéines [235]. Enfin, juger de la qualité d'un programme peut se faire à plusieurs niveaux, notamment sur la qualité des poses des ligands les mieux classés, sur sa prédiction de l'affinité de liaison ou encore sur son efficacité à cribler des chimiothèques.

De façon plus générale, le choix d'un programme donné pour un criblage virtuel par docking doit répondre à la question suivante : « Que veut-on obtenir, en combien de temps, et avec quelle précision ? ». Il semblerait qu'une approche combinant plusieurs programmes s'avère souvent plus pertinente que l'emploi d'un seul programme, particulièrement dans le cadre d'un criblage à haut débit [236-238].

L'utilisation des programmes de docking a conduit à de nombreux succès dans le domaine de la découverte de nouvelles molécules bioactives [200, 201]. Néanmoins, leurs algorithmes sont toujours en maturation. Les axes principaux de leur amélioration méthodologique visent essentiellement à considérer la totale flexibilité de la protéine pendant le processus de docking et à prendre en compte les effets d'entropie et de solvant dans l'évaluation de l'affinité du complexe protéine-ligand par la fonction de score. Certains travaux concernant le développement de programmes de docking innovants [209, 239, 240] sont la preuve que les limitations actuelles des algorithmes de docking et des fonctions de score pourraient être surmontées dans un avenir relativement proche. Le potentiel actuel et futur des techniques de docking est incontestablement source d'innovation thérapeutique.

PARTIE 4

RESULTATS

ARTICLE 1

I- Contexte

A cette époque (année universitaire 2007-2008), seules deux structures cristallographiques de la kinase c-Met étaient disponibles dans la Protein Data Bank (PDB). Il s'agissait des structures de codes PDB 1R0P et 1R1W correspondant à un mutant de la kinase sauvage (mutation Y1235D décrite initialement dans des métastases du carcinome Head and Neck Squamous Cell [241]). Ces structures tridimensionnelles [56] avaient été établies d'une part pour un complexe entre la kinase et l'alcaloïde microbien K-252-a (molécule de la famille des staurosporines) et d'autre part pour la protéine libre. Ces structures 3D étaient incomplètes car des morceaux importants de c-Met n'avaient pas pu être résolus par cristallographie, en particulier la boucle d'activation.

Mon premier travail a donc consisté à obtenir des modèles 3D complets en utilisant la technique dite de modélisation par homologie, afin d'obtenir des structures réalistes des parties manquantes. J'avais donc recherché dans la PDB des kinases présentant à la fois une bonne homologie de séquence avec c-Met et une structure complète. Ce travail une fois terminé, je disposais donc des modèles complets de c-Met sous forme *apo* et *holo*.

L'étape suivante concernait l'établissement de la chimiothèque dont les molécules seraient utilisées pour lancer une campagne de criblage virtuel par docking moléculaire ciblant les deux modèles obtenus ci-dessus. A cette époque, peu de chimiothèques ciblant des kinases étaient disponibles et nous avons donc préféré établir une chimiothèque virtuelle basée sur une analyse des molécules déjà connues pour inhiber des activités kinases (j'avais déjà procédé de cette façon au cours de mon magistère pour définir des molécules susceptibles d'être des ligands du récepteur de la cholecystokinine [242]). Avec les collègues chimistes impliqués dans le projet, j'ai donc participé à l'établissement d'une chimiothèque virtuelle d'environ 2000 molécules susceptibles d'être synthétisées et présentant les caractéristiques physicochimiques des ligands de c-Met connus à cette époque. Ces molécules consistaient en un châssis sur lequel se greffaient des groupes fonctionnels devant pouvoir se lier aux amino acides du site actif de c-Met comme indiqué dans [56].

Une première campagne de criblage virtuel a donc été lancée qui a permis de dégager quelques molécules intéressantes [243]. Ceci a également donné lieu à un dépôt de brevet en cours de validation (« Novel aminoacids derivatives, their process of preparation and their therapeutical uses as Met inhibitors », Europeen patent EP09305902, déposé le 18 septembre 2009).

Suite à cette première partie du travail, deux nouvelles structure cristallographiques de c-Met sont apparues (codes PDB 2RFS et 2RFN [42]), ce qui nous a donné de nouvelles idées concernant le site de liaison de la protéine et la structure chimique de ligands potentiels. Nous avons donc étendu notre chimiothèque et refait une nouvelle campagne de criblage utilisant cette fois les modèles issus des deux structures RX nouvellement disponibles. Un va et vient entre les équipes de chimie et de modélisation a abouti à l'élaboration des composés décrits dans la publication qui suit.

II- Identification of new aminoacid amides containing the imidazo[2,1-b]benzothiazol-2-ylphenyl moiety as inhibitors of tumorigenesis by oncogenic Met signaling

European Journal of Medicinal Chemistry xxx (2011) 1–16

Contents lists available at SciVerse ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Original article

Identification of new aminoacid amides containing the imidazo[2,1-*b*]benzothiazol-2-ylphenyl moiety as inhibitors of tumorigenesis by oncogenic Met signaling

Alessandro Furlan [a,1], Francesco Colombo [b,1], Andrea Kover [c,1], Nathalie Issaly [a,1], Cristina Tintori [d], Lucilla Angeli [d], Vincent Leroux [e,f], Sébastien Letard [g], Mercedes Amat [c], Yasmine Asses [e], Bernard Maigret [e], Patrice Dubreuil [g], Maurizio Botta [d], Rosanna Dono [a,*****], Joan Bosch [c,****], Oreste Piccolo [h,***], Daniele Passarella [b,**], Flavio Maina [a,*]

[a] *Aix-Marseille Univ, IBDML, CNRS UMR 6216, Parc Scientifique de Luminy, Case 907, 13288 Marseille, France*
[b] *Dipartimento di Chimica Organica e Industriale, Università degli Studi di Milano, Via Venezian 21, 20133 Milano, Italy*
[c] *Laboratory of Organic Chemistry, Faculty of Pharmacy and Institute of Biomedicine (IBUB), University of Barcelona, 08028 Barcelona, Spain*
[d] *Dipartimento Farmaco Chimico Tecnologico, Università degli Studi di Siena, Via Alcide de Gasperi 2, I-53100 Siena, Italy*
[e] *Nancy Université, LORIA-UMR 7503, Equipe projet Orpailleur, Campus scientifique, BP 239, 54506 Vandœuvre-lès-Nancy Cedex, France*
[f] *University of Oslo, Chemistry Department, P.O. Box 1033 Blindern, 0315 Oslo, Norway*
[g] *Centre de Recherche en Cancérologie de Marseille – U891 INSERM, Institut Paoli-Calmettes, Université de la Méditerranée 27, boulevard Lei Roure BP 30059, 13273 Marseille Cedex 09, France*
[h] *Studio di consulenza scientifica, Via Bornò 5, 23896 Sirtori (LC), Italy*

ARTICLE INFO

Article history:
Received 9 June 2011
Received in revised form
30 September 2011
Accepted 27 October 2011
Available online xxx

Keywords:
Met RTK
Met signaling inhibitor
Inhibitor binding modes
Aminoacid-amide derivatives
Anti-tumor agents
Imidazo[2,1-*b*]benzothiazol-2-ylphenyl moiety

ABSTRACT

The Met receptor tyrosine kinase is a promising target in anticancer therapies for its role during tumor evolution and resistance to treatment. It is characterized by an unusual structural plasticity as its active site accepts different inhibitor binding modes. Such feature can be exploited to identify distinct agents targeting tumor dependence and/or resistance by oncogenic Met. Here we report the identification of bioactive agents, featuring a new 4-(imidazo[2,1-*b*]benzothiazol-2-yl)phenyl moiety, targeting cancer cells dependent on oncogenic Met. One of these compounds (**7c; Trifloreas**) impairs survival, anchorage-independent growth, and in vivo tumorigenesis, without showing side effects. Our medicinal chemistry strategy was based on an in-house Met-focused library of aminoacid-amide derivatives enriched through structure-based computer modeling, taking into account the Met multiple-binding-mode feature. Altogether, our findings show how a rational structure-based drug design approach coupled to cell-based drug evaluation strategies can be applied in medicinal chemistry to identify new agents targeting a given oncogenic-dependency setting.

© 2011 Elsevier Masson SAS. All rights reserved.

1. Introduction

Receptor tyrosine kinase (RTK) signaling is one core pathway altered frequently in cancer. Therapeutic approaches based on compounds targeting selectively oncogenic RTKs, to which cells are dependent, are constantly developed [1–3]. Among those investigated anticancer drug targets, the Met RTK has drawn significant attention for the ability of its oncogenic forms to confer growth advantage, protection from apoptosis, invasive properties, and resistance to chemotherapies [4–6]. Met is activated upon binding of its natural ligand hepatocyte growth factor, also known as scatter factor (HGF/SF). Following its activation, the Met RTK triggers a number of signaling pathways that regulate specific biological

Abbreviations: RTK, Receptor tyrosine kinase; HGF, Hepatocyte growth factor; SF, Scatter factor; i.p., intraperitoneal; ATP, Adenosine triphosphate; IL-3, Interleukin-3; TPR, Translocated promoter region; MS, Mass spectrometry; DCC, Dicyclohexylcarbodiimide; DCI, Diisopropylcarbodiimide; HATU, 2-(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate; HOBt, 1-hydroxybenzotriazole; DIPEA, Diisopropylethylamine; DMAP, 4-N,N-dimethylaminopyridine; BOPCl, Bis(2-oxo-3-oxazolidinyl)phosphinic chloride; TFA, Trifluoroacetic acid; DMSO, Dimethyl sulfoxide; RMSD, Root mean square deviation.

* Corresponding author. Tel.: +33 4 91 26 97 69.
** Corresponding author. Tel.: +39 02 50 31 40 81.
*** Corresponding author. Tel.: +39 39 39 47 98 70.
**** Corresponding author. Tel.: +34 93 402 45 38.
***** Corresponding author. Tel.: +33 4 91 26 92 66.
E-mail addresses: rosanna.dono@ibdml.univmed.fr (R. Dono), joanbosch@ub.edu (J. Bosch), orestepiccolo@tin.it (O. Piccolo), Daniele.passarella@unimi.it (D. Passarella), flavio.maina@ibdml.univmed.fr (F. Maina).
URL: http://www.maina-ibdml.eu
[1] These authors contributed equally to this work.

0223-5234/$ – see front matter © 2011 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2011.10.051

Please cite this article in press as: A. Furlan, et al., Identification of new aminoacid amides containing the imidazo[2,1-*b*]benzothiazol-2-ylphenyl moiety as inhibitors of tumorigenesis by oncogenic Met signaling, European Journal of Medicinal Chemistry (2011), doi:10.1016/j.ejmech.2011.10.051

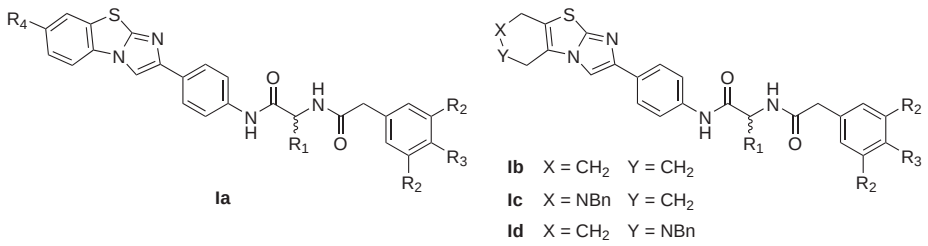


Fig. 1. Compounds of general formula **Ia**–**d**.

events during development [7–12,50] and regenerative processes [13,14], but also tumorigenesis [4,6,15]. Moreover, during tumor evolution oncogenic Met confers oncogene addiction, drug and/or radiotherapy resistance [16–20]. Therefore, agents able to antagonize oncogenic Met are expected to have a strong impact in molecularly targeted therapies for different types of human cancers. The great importance given to anti-Met agents for anticancer treatment as well as for fighting against resistance mechanisms caused by activated Met is mirrored by several research attempts to discover small-molecule inhibitors (see Supporting Information) [21,22]. However, the relatively high number of Met inhibitors already available does not deter academic labs and biopharma companies from investing persistent efforts aimed at uncovering novel compounds. This is shown also by the constant stream of released crystallographic complexes on the Protein Data Bank (PDB) database [23].

The Imatinib successful story has demonstrated that kinase inhibitors can effectively target specific inactive kinase conformations, allowing for the design of selective drugs [24,25]. This fostered tremendous research efforts in the industry for designing a new generation of anticancer drugs [26]. It is now acknowledged that upon the binding of ATP-competitive inhibitors, RTKs may adapt three distinct binding modes. However, RTKs currently known as able to accommodate all three kinds of inhibitor binding (to the RTK active state, DFG-out inactive state, and C-helix-out inactive state) are rare [27]. By applying molecular modeling and molecular mechanics to analyze the distribution of ligand interactions on Met residues as shown on available X-ray complexes, we previously demonstrated that Met is among the few RTKs accepting the three possible binding modes [28]. It has already been reported that the striking Met RTK plasticity, at the time with two known binding modes [29], is a clear opportunity for designing new Met-targeting compounds. We present the first study, to our knowledge, of a medicinal chemistry attempt aided by rational structure-oriented modeling, taking explicitly into account the triple-binding-mode feature of Met.

We report here in silico modeling studies, the synthesis, and the biological characterization of a new class of aminoacid amide derivatives, containing a 4-(imidazo[2,1-*b*]benzothiazol-2-yl) phenyl moiety (**Ia**, Fig. 1). Compounds were first biologically evaluated for their ability to impair Met-triggered cell scattering in vitro. Detailed in vitro biological characterization showed that distinct members of this new class of compounds impair survival and anchorage-independent growth of cancer cells addicted to oncogenic Met. A cell-based screen revealed selectivity toward the Met family members. These agents did not cause any major toxic effect neither on primary neuron nor on hepatocyte cultures. Finally, in vivo studies proved anti-tumor effects of the most active compound toward Met-addicted cancer cells, without causing side effects. Thus, our studies show how multidisciplinary strategies including in silico structure-based drug design and a cell-based focused drug screen can lead to the generation of agents that contrast tumorigenesis triggered by oncogenic signals like those emanating from the Met RTK.

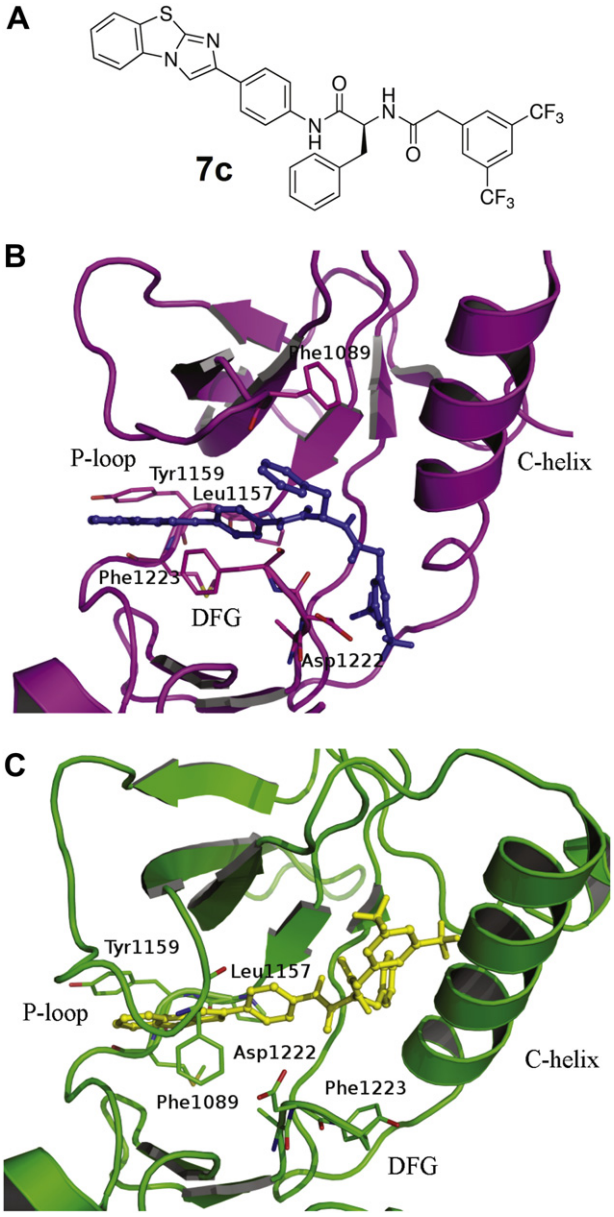
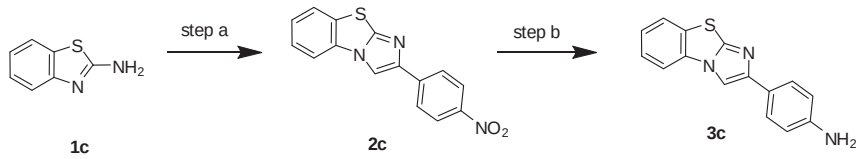
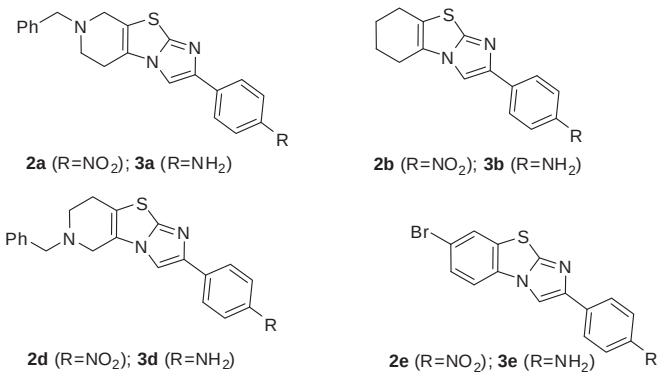


Fig. 2. Computer modeling shows that compound **7c** interacts with the Met ATP binding pocket. (A) 2D structure of compound **7c** (CF052). (B) Binding mode of compound **7c** (blue) within the 3CTJ structure (purple). (C) Binding mode of compound **7c** (yellow) within the 3EFK X-ray structure (green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: A. Furlan, et al., Identification of new aminoacid amides containing the imidazo[2,1-*b*]benzothiazol-2-ylphenyl moiety as inhibitors of tumorigenesis by oncogenic Met signaling, European Journal of Medicinal Chemistry (2011), doi:10.1016/j.ejmech.2011.10.051



Scheme 1. Synthesis of compound **3c**. Step a: 2-bromo-4′-nitroacetophenone, EtOH or i-PrOH, reflux, 90 min. Step b: Fe in HCl 12N, H_2O, EtOH, reflux, 90 min or $SnCl_2$, HCl 12N, MeOH, reflux, 90 min. A similar strategy was adopted for the synthesis of other compounds (**3a**, **3b**, **3d**, and **3e**).



Scheme 2. Intermediates **2** and **3** prepared.

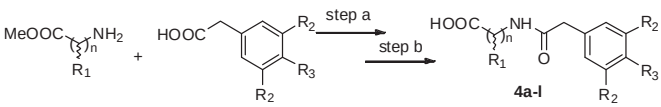
2. Results

2.1. Computer-aided rational design

2.1.1. Rational design of a Met-focused virtual library and docking calculations to select potential inhibitors for experimental investigation

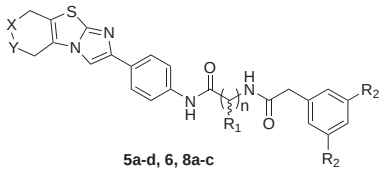
A starting series of compounds was modeled from in-house organic chemistry knowledge and analysis of known anticancer compounds reported in the IMS Pioneer (http://www.imshealth.com) and Becker's Pharma Market (http://www.beckerpharmaceuticals.com) databases. The small molecules were first subjected to docking calculations, using as target models the experimentally-derived Met RTK conformation available in the PDB: 1R0P [30]. Results were analyzed by both molecular modelers and organic chemists, leading to an incremental virtual library design process that was followed as long as clear improvements of average docking performance were achieved. Subsequent docking studies were performed (see Experimental Protocol section) and a subset of best molecules was identified on the basis of scoring function values. Analysis of this subset suggested optimal functional groups and/or regions where improvements could possibly be further made. Based on this information, devised probable chemically-feasible optimizations were further made, resulting in a new family of candidate agents. It should be noted that docking program scoring functions were applied to identify probable structural optimizations from a set of compounds with similar chemical features rather than for hit prediction, as we do not consider it accurate enough for such a purpose. Through a repeated process merging organic chemistry and molecular model knowledge, a total of 2275 structures (2559, counting stereoisomers) were analyzed through 7 generation/optimization steps. The aforementioned process started with 2146 compounds, with a subsequent selection process leading progressively to 22, 15, 35, 28, and 18 molecules, resulting to a final group of 11 compounds (Supplementary Table 1).

Docking studies were then performed using two other experimentally-derived Met conformations from the PDB: 2RFN and 2RFS [31]. In 2RFS, Met is bound to the SU11274 kinase inhibitor, with a binding mode similar to the 1R0P's K-252a staurosporine analog (kinase active state). The 2RFN conformation corresponds to the Met-compound (AM7) binding in an inactive C-helix-out type state [28,29]. These analyses showed that we indeed selected optimal compounds for targeting the 1R0P conformation, as both the average and best scoring functional values were improved through the 7 generations (Supplementary Table 1). Surprisingly, for the 57 compounds corresponding to generation 5–7, we found that their scores were optimal for both 1R0P and 2RFN conformations, and, to a lesser extent, for 2RFS. These



Compound	n	R_1	R_2	R_3
4a	1	(*S*)-Bn	CH_3	H
4b	1	(*R*)-Bn	CH_3	H
4c	1	(*S*)-Bn	CF_3	H
4d	1	(*R*)-Bn	CF_3	H
4e	1	(*S*)-Bn	F	H
4f	1	H	CH_3	H
4g	1	(*S*)-*p*OH-$C_6H_4CH_2$	CH_3	H
4h	2	H	F	H
4i	1	(*R*)-CH_2S-Tr*	F	H
4j	1	(*S*)-Me	F	H
4k	1	(*R*)-Me	F	H
4l	1	(*S*)-Bn	H	CH_3

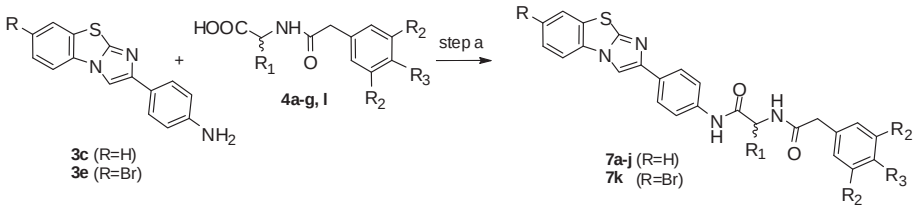
Scheme 3. Synthesis of compounds **4a**–**l**. Step a: DCC or DCI, THF, room temperature (rt), 48 h. Step b: NaOH 1 M, MeOH, rt, 1 h *Tr = trityl (triphenylmethyl group).



Compound	n	X	Y	R_1	R_2
5a	1	NCH_2Ph	CH_2	(*S*)-Bn	Me
5b	1	NCH_2Ph	CH_2	(*R*)-Bn	Me
5c	2	NCH_2Ph	CH_2	H	F
5d	1	NCH_2Ph	CH_2	(*R*)-CH_2SH^*	F
6	1	CH_2	CH_2	(*S*)-Bn	Me
8a	2	CH_2	NCH_2Ph	H	F
8b	1	CH_2	NCH_2Ph	(*S*)-Me	F
8c	1	CH_2	NCH_2Ph	(*R*)-Me	F

Scheme 4. Compounds **5**, **6**, and **8** prepared. *The trityl group was removed, after condensation, as reported in the experimental section.

Please cite this article in press as: A. Furlan, et al., Identification of new aminoacid amides containing the imidazo[2,1-*b*]benzothiazol-2-ylphenyl moiety as inhibitors of tumorigenesis by oncogenic Met signaling, European Journal of Medicinal Chemistry (2011), doi:10.1016/j.ejmech.2011.10.051



Compound	Code	R	R_1	R_2	R_3
7a	CF022	H	(*S*)-Bn	CH_3	H
7b	CF056	H	(*R*)-Bn	CH_3	H
7c	CF052 (Triflorcas)	H	(*S*)-Bn	CF_3	H
7d	CF081	H	(*R*)-Bn	CF_3	H
7e	CF082	H	(*S*)-Bn	F	H
7f	CF083	H	H	CH_3	H
7g	CF023	H	(*S*)-*p*OH-$C_6H_4CH_2$	CH_3	H
7h	CF201	H	(*S*)-Bn	H	CH_3
7i	CF207	H	(*S*)-*p*CH_3-$C_6H_4CH_2$	CH_3	H
7j	CF209	H	(*S*)-*p*CH_3-$C_6H_4CH_2$	CF_3	H
7k	CF203	Br	(*S*)-Bn	CF_3	H

Scheme 5. Synthesis of compounds **7a**–**h**,**k**. Step a: HATU, DIPEA, THF, rt, 16h or BOPCl, HOBt, DIPEA, CH_2Cl_2, rt, 48 h. Compounds **7i** and **7j** were prepared with a different procedure as reported in the experimental section.

outcomes may originate from the virtual screening protocol and the design choices, and suggest that compounds highlighted from our virtual library could be either inhibitor targeting an inactive Met RTK conformation, or dual-specificity inhibitors that could be further optimized. Summarizing all gathered data, two distinct classes of compounds, characterized by a kinase inhibitor profile, were highlighted. One of these two is defined by the imidazo[2,1-*b*] benzothiazol-2-ylphenyl moiety (**Ia**, Fig. 1), with some chemical features shared with another series of Met inhibitors that we were investigating (**Ib**–**d**, Fig. 1) [32]. **Ia**-based compounds were selected for chemical synthesis and biological evaluation, resulting in the present publication.

2.1.2. Molecular modeling of *Ia* compounds in complex with the Met RTK

In parallel to chemistry and biology experimental efforts dedicated to the **Ia** class (described in the next sections), computer models of the corresponding protein–ligand complexes were generated. Additional Met crystallographic complexes had been released in the PDB, allowing us to derive a precise molecular mechanics-based model for characterizing Met RTK interactions with inhibitors. It appeared that Met inhibitors could be clustered into three distinct families, corresponding to the active state, the DFG-out inactive state, and the C-helix-out inactive state [28]. Being capable of accepting these three ligand binding modes is an uncommon feature for kinases [27]. In order to optimally exploit this Met RTK bound conformational space, the **Ia** compounds were docked against each of those three possible states.

2.1.2.1. Docking on the active state conformation. Docking **Ia**-derived compounds on the conformation from the 2RFS complex [31] representative of the Met active state, resulted in significantly lower binding scores than with the two other inactive state conformations described below. This is not a surprising result, considering that the compounds we have designed possess structural properties typical of inactive-state-binding kinase inhibitors. Indeed, only Met inactivation may provide additional hydrophobic area (either by displacement of the A-loop or the C-helix) required to accommodate small molecules possessing an extended scaffold, such as **Ia**–**d**.

2.1.2.2. Docking on the DFG-out conformation. The 3CTJ X-ray complex [33] was selected as representative of the DFG-out inactive state (Fig. 2 is referred to **7c**). Fig. 2B shows compound **7c** (Triflorcas) bound to this model. The docking calculations showed a common binding mode for the left part of all inhibitors. The benzothiazole (A-ring) is bound to the hinge region and partially superimposed to the pyrrolopyridine found in the crystallographic complex. The central aromatic ring (B-ring) π-stacks with Phe1223 (DFG motif) and is flanked on the opposite face by hinge residue Leu1157. The –NH–CO–CH–NH chain is involved in H-bonds with Asp1222, Glu1127, and Nε of Lys1110. Interactions with these residues are consistently found among ligands bound to Met inactive states [28]. The **7c** bimethylfluorobenzyl fragment occupies the DFG pocket. Other ligands reported in this paper, upon docking to this conformation, share the same general orientation, with either their C- or D-ring occupying the DFG pocket.

2.1.2.3. Docking on the C-helix-out conformation. The 3EFK X-ray complex [34] was selected as representative of the Met C-helix-out inactive state. Fig. 2C shows the result with **7c**. Similarly to DFG-out docking results, the A-ring is found in the hinge region. The B-ring makes lipophilic interactions with the side chains of Leu1157 and Phe1089 (stacking) rather than with Phe1223 of the DFG triad. As a result, the C- and D-rings are oriented toward the C-helix and away from the A-loop. The **7c** C-ring binds to a small hydrophobic pocket formed around Ile1145 by C-helix displacement. At this stage, docking results suggest that **7c** and analogs may equally bind to either a DFG-out or C-helix-out inactive state Met conformation.

Please cite this article in press as: A. Furlan, et al., Identification of new aminoacid amides containing the imidazo[2,1-*b*]benzothiazol-2-ylphenyl moiety as inhibitors of tumorigenesis by oncogenic Met signaling, European Journal of Medicinal Chemistry (2011), doi:10.1016/j.ejmech.2011.10.051

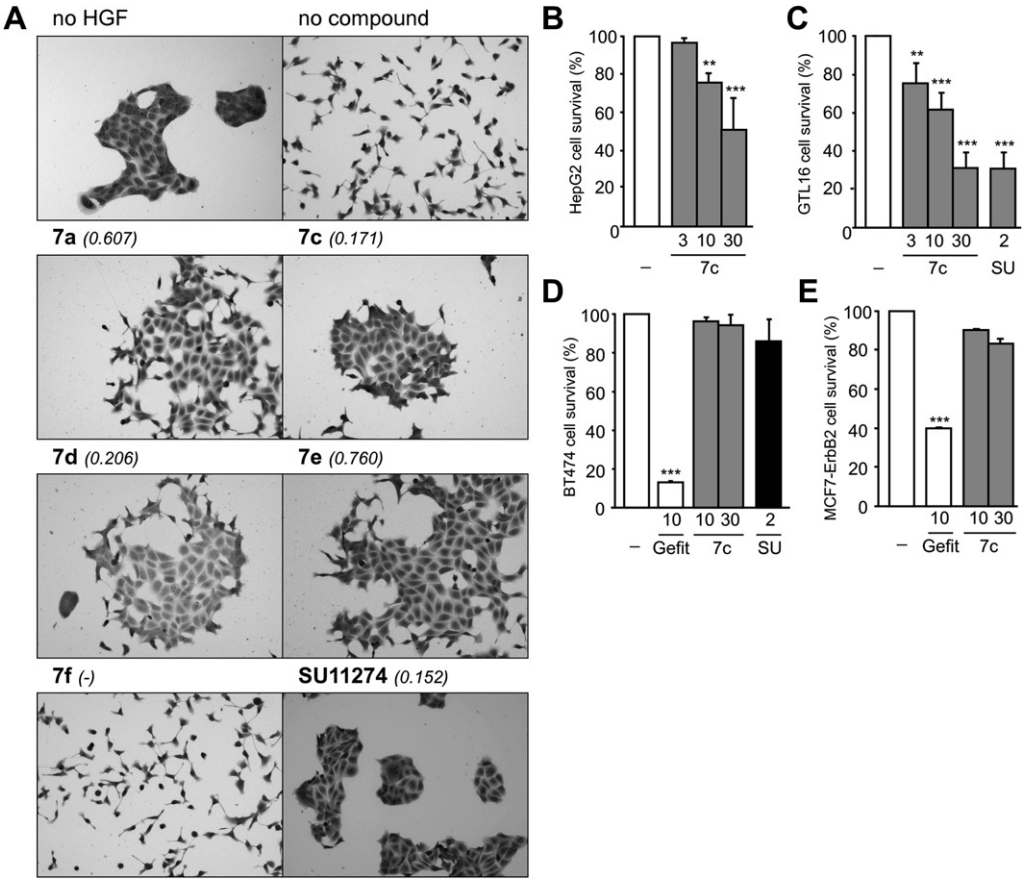


Fig. 3. Imidazo[2,1-*b*]benzothiazol-2-ylphenyl compounds block Met-triggered cell scattering and cell survival. (A) HGF-induced scattering of MDCK cells is blocked by **7a**, **7c**, **7d**, **7e**, and not by **7f**. SU11274 was used as positive control. The IC_{50} is indicated in μM. (B and C) Survival of HepG2 (B) and GTL-16 (C) cells was reduced by **7c**, in a dose dependent manner (μM). SU11274 (SU) was used at 2 μM (D and E) Survival of BT474 (D) and MCF7-ErbB2 (E) cells, which are addicted to the ErbB oncogenes, was impaired by the ErbB inhibitor Gefitinib (Gefit), but not by **7c** or by SU11274 (μM). For survival assays ($n = 3$), cells were serum-starved for 24 h and then incubated with **7c**, SU11274, or Gefitinib for 48 h. Values are expressed as means ± s.e.m. $^{**}P < 0.01$; $^{***}P < 0.001$; Student-*t* test.

2.2. Chemistry

To synthesize compounds **2** (Schemes 1 and 2), we applied a similar approach using 4 different 2-aminothiazole derivatives. The starting 2-aminotetrahydrobenzothiazole **1b** was obtained by reaction of cyclohexanone with thiourea in the presence of iodine, whereas the compounds 2-amino-5-benzyl-4,5,6,7-tetrahydrothiazolo[5,4-c]pyridine **1a**, 2-amino-benzothiazole **1c**, and 2-amino-5-benzyl-4,5,6,7-tetrahydrothiazolo[4,5-c]pyridine **1d** were commercially available. Reaction of compound **1** with 2-bromo-4′-nitroacetophenone in ethanol provided the corresponding nitro derivatives **2** by alkylation of the ring nitrogen and concomitant dehydration using EtOH or *i*-PrOH [35]. The use of a polar solvent secured the concomitant second step, whereas the use of toluene induces only the first step of the reaction. The reduction of the nitro group was secured with $SnCl_2$ in MeOH-HCl or with Fe-powder in the presence of H_2SO_4 to give the corresponding anilines **3** (synthesis of compound **3c** is exemplified in Scheme 1). The use of $SnCl_2$ induces the secondary introduction of a chlorine atom at position 3 of the tricyclic core. In particular, this was the case for the preparation of **3b** and **3c,** whose yield was lower than 50%.

Glycine, *R*- and *S*-alanine, β-alanine, *R*- and *S*-phenylalanine, *R*-cysteine and *S*-tyrosine were selected for the generation of compounds **4** (Scheme 3), and subsequently condensed with the tricyclic scaffolds. The methyl esters of these aminoacids were reacted with 4-methylphenyl-, 3,5-difluorophenyl-, 3,5-dimethylphenyl-, and 3,5-bis(trifluoromethyl)phenyl-acetic acids in the presence of DCC or DCI. In the case of cysteine the corresponding methyl ester and S-trityl derivative was used. The hydrolysis of the methyl esters (NaOH, MeOH, 1 h) took place without any significant racemization (^{1}H NMR by shift reagent) and permitted the subsequent condensation reaction to complete the structure of the designed compounds (Schemes 4 and 5).

The building block **3a** (Scheme 2) was used for the formation of four different compounds (**5a**–**d**) (Scheme 4) using HATU (yield 70–90%) or BOPCl (yield 45–70%) as dehydrating agents. The formation of **5d** required removal of the trityl group with TFA and Et_3SiH in CH_2Cl_2. The use of building blocks **3b**, **3c**, **3d**, and **3e** (Schemes 1 and 2) permitted the production of **6, 7a**–**j**, **8a**–**c**, and **7k**, respectively (Schemes 4 and 5). All the obtained compounds were stable as solids and in solution (DMSO, $CHCl_3$, MeOH). Their structures were confirmed on the basis of the NMR and MS spectra.

2.3. Biology

2.3.1. Identification of imidazo[2,1-b]benzothiazol-2-ylphenyl compounds as a new class of inhibitors blocking Met-triggered cell scattering and cell survival

We have previously shown that compounds can be efficiently screened and/or biologically validated for their inhibitory properties toward Met-triggered biological responses by using cell scattering assays [32]. In particular, MDCK epithelial cells acquire a “scattered phenotype” following stimulation by HGF, the Met ligand. The scattering response to HGF is impaired in the presence

Please cite this article in press as: A. Furlan, et al., Identification of new aminoacid amides containing the imidazo[2,1-*b*]benzothiazol-2-ylphenyl moiety as inhibitors of tumorigenesis by oncogenic Met signaling, European Journal of Medicinal Chemistry (2011), doi:10.1016/j.ejmech.2011.10.051

of inhibitors targeting Met [32,36]. For these studies, SU11274 was used as a reference Met inhibitor [37]. Among 20 newly synthesized imidazo[2,1-*b*]benzothiazol-2-ylphenyl compounds (**Ia**), we found that **7a**, **7b, 7c**, **7d**, **7e**, **7h**, and **7k** (Scheme 5) elicited inhibitory activity on Met-triggered cell scattering. By evaluating the IC_{50} of the most active compounds, we found that **7c** and **7d** impaired Met-triggered cell scattering at concentrations comparable to SU11274 (**7c**: 0.171 μM; **7d**: 0.206 μM; SU11274: 0.152 μM; Fig. 3A and Supplementary Fig. 1). No toxic effects were observed at biologically active concentrations and only found when compounds were applied at doses 50–100 fold higher (data not shown). Further validations were obtained by testing **Ia** compounds on human MCF10A breast cells and similar inhibitory properties on Met-triggered scattering were observed (data not shown).

We next investigated the ability of **Ia** compounds to prevent Met-triggered survival of cancer cells. These studies were performed on human HepG2 hepatocellular carcinoma cells, whose tumorigenesis is Met-dependent, and on human GTL-16 gastric carcinoma cells, in which *c-met* gene amplification results in high Met protein levels leading to its over-activation in a ligand independent manner. As a consequence, GTL-16 cells are extremely aggressive, "Met-addicted" for survival, anchorage-independent growth, and tumor formation when injected into nude mice [38]. We found that compounds **7a**, **7c**, **7d**, and **7e** reduced survival of both HepG2 and GTL-16 cells in a dose-dependent manner (Fig. 3B and C, and Supplementary Fig. 2A). Notably, **Ia** compounds did not restrain survival of cancer cell lines, such as BT474 and MCF7-ErbB2 (Fig. 3D and E), which are addicted to ErbB oncogenes [39]. Altogether, these studies identified a new class of biologically active compounds characterized by a new moiety, namely imidazo[2,1-*b*]benzothiazol-2-ylphenyl, and displaying inhibitory effects on cells with activated Met.

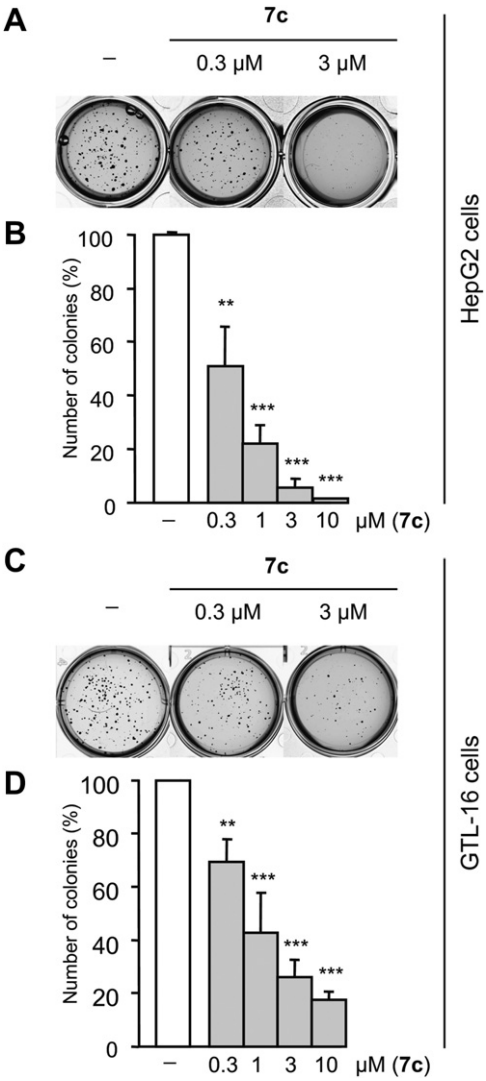


Fig. 4. Compound **7c** blocks in vitro tumorigenesis triggered by Met. **7c** impairs anchorage-independent growth of HepG2 (A and B) and GTL-16 (C and D) cells, in a dose dependent manner (μM; $n = 3$). Values are expressed as means ± s.e.m. $^{**}P < 0.01$; $^{***}P < 0.001$; Student-*t* test.

2.3.2. *Imidazo[2,1-b]benzothiazol-2-ylphenyl compounds interfere with Met-triggered in vitro tumorigenesis*

We next ascertained whether the identified **Ia** class of compounds also prevented Met-triggered anchorage-independent growth, a hallmark of oncogenic transformation. Soft-agar growth of HepG2 cells requires intact Met as it is restrained by the Met inhibitor SU11274 [38]. We found that compounds **7a**, **7c**, **7d**, and **7e** impaired in vitro tumorigenesis of HepG2 cells, in a dose dependent manner (Fig. 4A and B, and Supplementary Fig. 2B). By evaluating the IC_{50} of the most active compounds, we found that **7c** and **7d** interfered with Met-triggered anchorage-independent growth at concentrations 3–5 folds lower than those required for SU11274 (**7c**: 0.321 μM; **7d**: 0.620 μM; SU11274: 1.561 μM; Fig. 4A and B, Supplementary Figs. 2B and 3). Tumorigenesis of GTL-16 cells addicted to the Met oncogene was also blocked by **Ia** compounds, in a dose dependent manner (Fig. 4C and D, and Supplementary Fig. 2C). **7c** and **7d** were again the most active compounds (**7c**: 0.811 μM; **7d**: 1.194 μM; SU11274: 0.228 μM; Supplementary Fig. 3). Altogether, these findings demonstrate that **Ia** compounds restrain in vitro tumorigenesis of cells dependent or addicted to the Met oncogene. As **7c** appeared among the most effective compounds we have identified, its properties were further investigated through a series of in vitro and in vivo assays.

2.3.3. *Selectivity of imidazo[2,1-b]benzothiazol-2-ylphenyl compounds*

As several Met inhibitors additionally target other kinases, we tested **Ia** compound selectivity toward a panel of 16 Ba/F3 cells expressing the active form of different kinases. These studies were performed by following the proliferation of cells in the presence or absence of **7c** at concentrations ranging from 0.1 to 40 μM. As shown in Fig. 5, the **7c** compound at 1 μM interfered predominantly

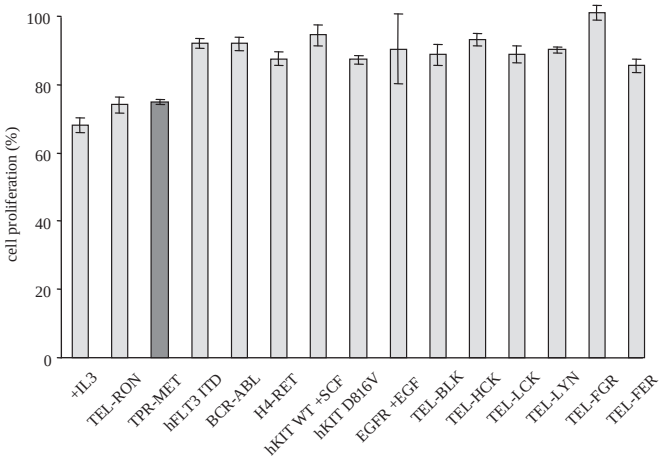


Fig. 5. Inhibition of BaF/3 cells transfected with active forms of the indicated kinases by compound **7c**. Cells were incubated in the presence of **7c** (1 μM) for 48 h before analyzing cell proliferation using Cell Titer Glo. Percentage represents proliferation of each cell line in the presence of **7c** compared to untreated cells.

Please cite this article in press as: A. Furlan, et al., Identification of new aminoacid amides containing the imidazo[2,1-*b*]benzothiazol-2-ylphenyl moiety as inhibitors of tumorigenesis by oncogenic Met signaling, European Journal of Medicinal Chemistry (2011), doi:10.1016/j.ejmech.2011.10.051

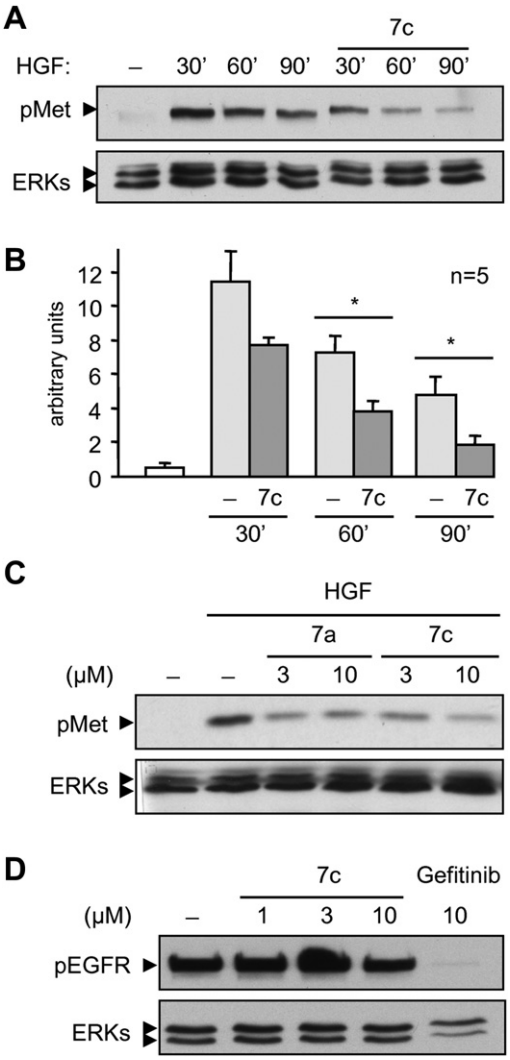


Fig. 6. Compound **7c** interferes with Met phosphorylation in living cells. (A) Met phosphorylation following HGF stimulation (20 ng/ml) in HepG2 cells was reduced in presence of **7c**. For western blot analyses, cells were treated with HGF for 30, 60, and 90 min in presence or not of **7c** (10 μM). Total cell lysates were analyzed using anti-phospho-Tyr1234/1235-Met (pMet) (upper panel). (B) Quantification of Met phosphorylation in presence or not of **7c** ($n = 5$). Values are expressed as means ± s.e.m. $^{*}P < 0.05$; Student-t test. (C) Met phosphorylation was also reduced by **7a** and **7c** applied at 3 and 10 μM in human MDA-MB231 breast cancer cells exposed to HGF stimulation for 15 min (D) **7c** does not affect EGFR phosphorylation in human BT-474 cells at 1, 3, and 10 μM concentrations. Gefitinib was used as positive control (10 μM). ERK protein levels were used as loading controls (lower panels in A, C, and D).

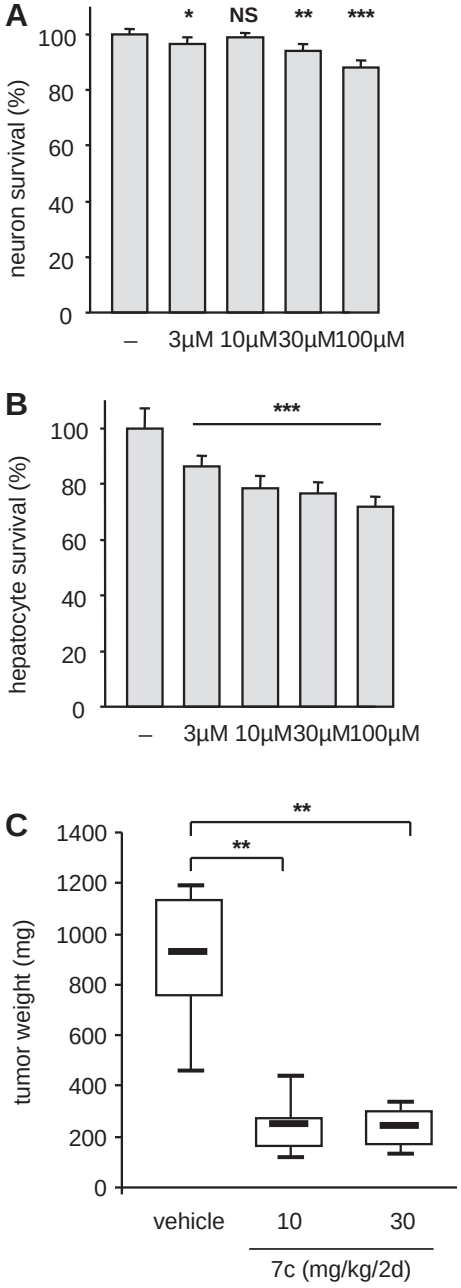


Fig. 7. Compound **7c** impairs in vivo tumor growth of cancer cells dependent on the Met oncogenic signaling, without eliciting major side effects either on primary neuron or hepatocyte cultures. (A and B) Survival of cortical neurons (A) and hepatocytes (B) was not drastically affected by **7c** (μM). For survival assays, cells were incubated with 7c for 24 h. (C) Compound **7c** interferes with growth of tumors triggered by cancer cells dependent on the Met oncogene. **7c** treatment (i.p. 10 or 30 mg kg^{-1} every 2 days) reduces tumor weight in nude mice injected intra-peritoneally with GTL-16 cells. Values are reported as boxplots. Two independent experiments were performed and 8 mice per group were used. Values are expressed as means ± s.e.m. $^{*}P < 0.05$; $^{**}P < 0.01$; Student-t test.

with proliferation of Ba/F3 cells expressing the oncogenic form of Met (TPR-Met), its family member Ron, or IL3. In this cellular system, the IC_{50} value of **7c** with regard to Met was 4 μM, which is higher than what was found when IC_{50} value was evaluated in HepG2 and GTL-16 cells. This difference could be due to the diversity in the biological assays and to the artifactual TPR-Met oncogenic form. Altogether, these studies establish that **Ia** compounds have a certain degree of selectivity toward inhibition of the Met-associated pathways in cells. Moreover, the inhibitory effects on Ba/F3 cells with IL3 suggest that they may also act on other RTK-distinct signaling molecules.

*2.3.4. Compound **7c** interferes with Met phosphorylation in living cells and with Met activation in vitro*

We next biochemically evaluated the ability of **Ia** compounds to interfere with Met activation by following the phosphorylation levels of two Tyrosine residues located in its kinase domain, namely Tyr1234 and Tyr1235. High levels of phospho-Met were observed upon HGF stimulation in HepG2 cells, which progressively decreased over time (Fig. 6A and B). Met phosphorylation was reduced by **7c** by 30–50% over time when compared to controls (Fig. 6A and B). Reduced Met phosphorylation by **7c** was also observed in other cell lines, such as in human MDA-MB231 breast cancer cells exposed to HGF stimulation (Fig. 6C). In contrast, **7c** did not interfere with phosphorylation of ErbB1 when applied to BT474 cells (Fig. 6D).

Please cite this article in press as: A. Furlan, et al., Identification of new aminoacid amides containing the imidazo[2,1-*b*]benzothiazol-2-ylphenyl moiety as inhibitors of tumorigenesis by oncogenic Met signaling, European Journal of Medicinal Chemistry (2011), doi:10.1016/j.ejmech.2011.10.051

We next assessed the effects of **7c** on Met activation by using the Kinexus compound profiling service. This strategy determines the profile inhibition of a compound against protein kinases and protein phosphatase targets. The profiling data against Met revealed that **7c** restrains Met activation by 21 and 53% at 1 and 10 μM, respectively. As suggested by computer modeling and biological assays, these findings identify the new **Ia** scaffold as inhibitors of the Met RTK.

*2.3.5. Compound **7c** impairs in vivo tumor growth of cancer cells dependent on the Met oncogene, without eliciting in vitro and in vivo side effects*

To have a first indication of tolerance to compound **7c** for therapeutic use, we assessed its side effects on primary cultures of neurons and hepatocytes. Interestingly, compound **7c** was very well tolerated by both primary cells, leading to only a modest loss of cell viability even at high doses of 100 μM (neuron: 12% reduction; hepatocytes: 28% reduction; Fig. 7A and B).

Finally, the ability of **7c** to impair tumor growth of cancer cells dependent on Met oncogenic signaling was evaluated in vivo using xenograft nude mice as tumor initiation models. For these studies, GTL-16 cells were chosen for their dependency to the oncogenic Met signaling, leading to aggressiveness and resistance to chemotherapeutic agents. GTL-16 cells (10^6) were intra-peritoneally injected into nude mice and tumor growth was examined in mice treated with **7c** or vehicle alone. Intra-peritoneal injection of cancer cells leads to development of several nodules in the peritoneal cavity, offering the possibility to evaluate compound efficacy on tumor weight. Notably, we found that the tumor weight was drastically reduced in mice injected with **7c** tested at different doses (Fig. 7C). In particular, when **7c** was administered to mice at a dose of 10 or 30 mg kg^{-1} every other day, we found 70% reduction in tumor weight (control: 865.4 ± 348.9; **7c** injected 10 mg kg^{-1}: 268.8 ± 159.7, $P = 0.002$; **7c** injected 30 mg kg^{-1}: 234.8 ± 88.5, $P = 0.001$; Fig. 7C). No major side effects were observed, evaluated by following the weight of mice treated with **7c** (data not shown). Taken together, these findings demonstrate that in vivo **7c** elicits tumor growth inhibition of cancer cells dependent on the Met oncogene. Moreover, the absence of side effects neither in neurons and hepatocytes nor in vivo indicates that **7c** is well tolerated when injected into mice at doses required to elicit its anti-tumor effects.

3. Conclusion

The discovery that RTKs might be targeted by chemical agents at different sites in their conformational space has generated great hope for molecularly targeted anticancer therapies. Previous studies have shown that Met is a particularly attractive target in this regard [28]. Indeed, the triple-binding-mode capacity of the Met RTK can be exploited to identify new agents with inhibitory properties toward cancer cells addicted to oncogenic Met signaling. Our findings show how support from computer-aided drug design could be useful to medicinal chemistry investigations. An integrated interdisciplinary screening strategy, based on rational drug design coupled to a cell-based focused screen and supported by molecular modeling from existing crystallographic complexes, led us to identify aminoacid amides containing the imidazo[2,1-*b*]benzothiazol-2-ylphenyl moiety **Ia** as new effective agents targeting Met-driven tumorigenesis. In particular, the biological evaluation of these agents in living cells allowed us to directly assess their inhibitory properties toward oncogenic Met signaling while excluding compounds causing side effects, leading to the identification of those exerting anti-tumorigenic activity in xenograft models.

Notably, as the identified imidazo[2,1-*b*]benzothiazol-2-ylphenyl derivatives have been screened in cultured cancer cells addicted to the Met oncogene, they may act on signals required to execute the Met-driven oncogenic program, in addition to targeting the Met RTK directly. This possibility is well supported by our previous studies showing that the tetrahydrobenzothiazole scaffold impairs Met signaling and also prevents Met restoration after degradation [32]. Our cell-based selectivity assays show that the identified **Ia** compounds predominantly target Met family members rather than other RTKs, which are impaired by Met inhibitors previously discovered [31,37,38,40–42]. Intriguingly, the most active imidazo [2,1-*b*]benzothiazol-2-ylphenyl derivative interferes with anchorage-independent growth of HepG2 cells at lower concentrations than those required for SU11274. Instead, GTL-16 cells are slightly more sensitive to SU11274 inhibition rather than to imidazo [2,1-*b*]benzothiazol-2-ylphenyl derivatives. This might reflect their inhibitory properties according to the level of activated Met. Alternatively, their efficacy may depend on the sets of signals activated by oncogenic Met, which vary between cancer cells and is paralleled by their differential sensitiveness to the action of compounds. Future studies will further establish the drug action of **Ia** compounds and their effectiveness toward a panel of different cancer cells to fully exploit their anticancer properties for single or combined therapies.

4. Experimental protocols

4.1. Computational part

4.1.1. Representation of molecular structures

ChemDraw 6.0 was used to represent the 2D structures. Pymol (DeLano Scientific LLC) and GNU Image Manipulation Program (GIMP 2.6) were used for generating the 3D representations.

4.1.2. Docking simulations

Docking calculations were done with GOLD v4.0 (CCDC) [43–45]. The Kinase Scoring function was chosen. This ChemScore variant aims at better assessing weak H-bonds by adding specific terms relevant to the kinase inhibitors' signature N-heterocycle/CH…O interactions. The automatic generic algorithm settings of GOLD were employed using the search efficiency set to 100%, and the number of docking runs per ligand set to 30. The binding site was defined within 5 Å of the cocrystallized ligands coordinates. The Met receptor structures were prepared (protonation, modeling of missing elements) from the original PDB files using Maestro v8.5 (Schrödinger). Starting structures of inhibitors were built using Maestro 3D-sketcher then minimized with the OPLS_2005 force field [46] using the Polak-Ribiere conjugated gradient method (0.001 kJ/Å·mol convergence). The reliability of the docking protocol was tested by retrieving the binding mode of Met inhibitors as cocrystallized in the 2RFS, 3CTJ, and 3EFK PDB X-ray structures, giving a negligible deviation from the crystallographic-resolved positions (RMSD < 1 Å) in all cases.

4.2. Chemistry

4.2.1. General

Melting points were determined in a capillary tube on a Büchi apparatus and are uncorrected. NMR spectra were recorded at 200, 300, 400 or 500 MHz (^{1}H) and 75.4, 100.6 or 125.9 MHz (^{13}C). Chemical shifts are reported in δ values downfield from TMS. IR spectra were recorded on a Perkin–Elmer 1600 spectrophotometer and only noteworthy IR absorptions are listed. Optical rotations were measured on a Perkin–Elmer 241 polarimeter using a 1 dm cell with a total volume of 1 ml. EI mass spectra were recorded at an ionizing voltage of 6 KeV on a VG 70-70 EQ. ESI mass spectra were recorded on FT-ICR APEX^{II} (Bruker Daltonics). High resolution mass spectra (HMRS; LC/MSD TOF Agilent Technologies) were performed by *Serveis Científico-Tècnics*, Barcelona. Thin-layer chromatography

Please cite this article in press as: A. Furlan, et al., Identification of new aminoacid amides containing the imidazo[2,1-*b*]benzothiazol-2-ylphenyl moiety as inhibitors of tumorigenesis by oncogenic Met signaling, European Journal of Medicinal Chemistry (2011), doi:10.1016/j.ejmech.2011.10.051

was done on SiO_2 (silica gel 60 F_{254}, Merck), and the spots were located with UV light or aqueous potassium permanganate solution. Flash chromatoghraphy was carried out using SiO_2 (silica gel 60, SDS, 35–70 μ). All nonaqueous reactions were performed under an inert atmosphere. Solvents for chromatography were distilled at atmospheric pressure prior to use and dried using standard procedures. Drying of the organic extracts during the workup of reactions was performed over anhydrous Na_2SO_4 or $MgSO_4$. Evaporation of solvents was accomplished with a rotatory evaporator. Elemental analyses were performed by *Centre d'Investigació i Desenvolupament* (CSIC). All test compounds showed >95% purity as determined by combustion analysis.

4.2.2. 2-Amino-4,5,6,7-tetrahydrobenzothiazole (**1b**)

A mixture of cyclohexanone (3.0 g, 30.6 mmol), thiourea (4.65 g, 61.1 mmol) and iodine (7.76 g, 30.6 mmol) was stirred at 110 °C for 12 h. The reaction mixture was cooled to room temperature. Hot water was then added and the resulting solution was stirred for 30 min. The aqueous solution was washed with diethyl ether and then neutralized by the addition of solid $NaHCO_3$. The pale yellow crystals were collected by filtration. The hydroiodide salt was dissolved in a hot saturated aqueous solution of Na_2CO_3. After cooling, the aqueous solution was extracted with CH_2Cl_2. The organic layer was dried over anhydrous Na_2SO_4, filtered, and concentrated under reduced pressure. The compound **1b** was obtained without any further purification as pale yellow solid (3.39 g, 72%). ^{1}H NMR (DMSO-d_6 300 MHz) δ 6.61 (2H, s), 2.52–2.48 (2H, m), 2.38–2.35 (2H, m), 1.72–1.68 (4H, m). MS-EI *m*/*z* $[M]^+$ 154.

4.2.3. 2-Amino-6-bromo-2-benzothiazole (**1e**)

A solution of 4-bromoaniline (0.500 g, 2.91 mmol) and potassium thiocyanate (1.13 g, 11.6 mmol) in AcOH (10 ml) was stirred at 20 °C for 10 min. Bromine (150 μl, 2.91 mmol) was added over 20 min to the above solution. The reaction mixture was stirred further at room temperature for 60 min. On completion of reaction following a TLC examination, the reaction mixture was poured into a solution of NH_3 5M and extracted with AcOEt. The organic layer was dried over anhydrous Na_2SO_4 and concentrated under reduced pressure. The crude product obtained was purified by flash chromatography (Hex:AcOEt 1:1) to afford **1e** as white solid (2.67 g, yield 79%). ^{1}H NMR (DMSO-d_6, 400 MHz): δ 7.91 (1H, d, *J* = 2.5 Hz), 7.63 (2H, s), 7.35 (1H, dd, *J* = 11.6, 2.5 Hz), 7.26 (1H, d, *J* = 11.6 Hz). MS-EI *m*/*z* $[M]^+$ 228.

4.2.4. General procedure for synthesis of (**2**)

A mixture of **1** (4 mmol) and 2-bromo-4′-nitroacetophenone (4.4 mmol) was refluxed in EtOH or *i*-PrOH (6–10 ml) for 50–90 min, the mixture was cooled to 0 °C. The solid was collected by filtration and washed with ethanol to afford **2** without any further purification.

4.2.5. 7-Benzyl-2-(4-nitrophenyl)-5,6,7,8-tetrahydroimidazo [2′,1′:2,3]thiazolo[5,4-c]pyridine (**2a**)

2-amino-5-benzyl-4,5,6,7-tetrahydrothiazolo[5,4-*c*]pyridine (**1a**, 1.0 g, 4.08 mmol) and 2-bromo-4′-nitroacetophenone (1.09 g, 4.49 mmol) were dissolved in *i*-PrOH (6 ml). **2a** (1.67 g, 41%), pale yellow solid. ^{1}H NMR ($CDCl_3$, 400 MHz): δ 8.22 (2H, d, *J* = 8.8 Hz), 7.94 (2H, d, *J* = 8.8 Hz), 7.70 (1H, s), 7.27–7.37 (5H, m), 3.78 (2H, s), 3.61 (2H, s), 3.01 (2H, t, *J* = 6.0 Hz), 2.81 (2H, t, *J* = 6.0 Hz). MS-EI *m*/*z* $[M]^+$ 390.

4.2.6. 2-(4-Nitrophenyl)-5,6,7,8-tetrahydroimidazo[2,1-b] benzothiazole (**2b**) [47]

1b (3.00 g, 19.5 mmol) and 2-bromo-4′-nitroacetophenone (5.24 g, 21.4 mmol). **2b** (2.31 g, yield 40%), yellow solid. ^{1}H NMR (DMSO-d_6, 300 MHz): δ 8.52 (1H, s), 8.27 (2H, d, *J* = 8.8 Hz), 8.09 (2H, d, *J* = 8.8 Hz), 2.73–2.69 (4H, m), 1.92–1.88 (4H, m); MS-EI *m*/*z* $[M]^+$ 299.

4.2.7. 2-(4-Nitrophenyl)imidazo[2,1-b]benzothiazole (**2c**) [48]

2-Aminobenzothiazole (**1c**, 1.00 g, 6.4 mmol) and 2-bromo-4′-nitroacetophenone (1.83 g, 7.1 mmol). **2c** (495 mg, yield 26%) yellow solid. ^{1}H NMR (DMSO-d_6, 300 MHz): δ 9.10 (1H, s), 8.34 (2H, d, *J* = 9.0 Hz), 8.12 (2H, d, *J* = 9.0 Hz), 8.07 (1H, d, *J* = 8.0 Hz), 8.02 (1H, d, *J* = 8.0 Hz), 7.64 (1H, t, *J* = 8.0 Hz), 7.49 (1H, t, *J* = 8.0 Hz). MS-EI *m*/*z* $[M]^+$ 295.

4.2.8. 6-Benzyl-2-(4-nitrophenyl)-5,6,7,8-tetrahydroimidazo [2′,1′:2,3]thiazolo[4,5-c]pyridine (**2d**)

2-Amino-5-benzyl-4,5,6,7-tetrahydrothiazolo[4,5-*c*]pyridine (**1d**, 3.9 g, 16.1 mmol) and 2-bromo-4′-nitroacetophenone (4.7 g, 19.4 mmol) in *i*-PrOH (80 ml). **2d** (2.28 g, 45%) was directly submitted to the next step without further purification.

4.2.9. 7-Bromo-2-(4-nitrophenyl)imidazo[2,1-b]bromobenzole (**2e**)

A mixture of **21** (1.20 g, 5.24 mmol) and 2-bromo-4′-nitroacetophenone (1.41 g, 5.76 mmol) was refluxed in ethanol for 90 min, the mixture was cooled to 0 °C. The solid was collected by filtration and washed with ethanol to afford **2e** without any further purification. Yellow solid (0.665 g, yield 34%). ^{1}H NMR (DMSO-d_6, 400 MHz, 80 °C): δ 8.92 (1H, s), 8.32 (1H, d, *J* = 1.9 Hz), 8.28 (2H, d, *J* = 8.8 Hz), 8.12 (2H, d. *J* = 8.8 Hz), 7.96 (1H, d, *J* = 8.6 Hz), 7.76 (1H, dd, *J* = 8.6, 1.9 Hz). ^{13}C NMR (DMSO-d_6 100.6 MHz) 80 °C: δ155.67, 146.92, 145.02, 140.78, 132.15, 131.47, 130.15, 127.93, 125.93 (2C), 124.62 (2C), 117.76, 115.53, 112.47. MS-EI $[M]^+$ 373.

4.2.10. General procedure for synthesis of (**3**)

Procedure A: Fe (powder, 7 g) and H_2SO_4 (12 N, 0.7 ml) were added to a solution of **2** (6 mmol) in ethanol/water (5:1, 50–100 ml) and the mixture was stirred at 90 °C for 1 h. The hot solution was filtered through celite and washed with hot ethanol. The solvent was removed in vacuo and the crude material was extracted by satured solution of $NaHCO_3$ and AcOEt. The organic phase was dried with Na_2SO_4 anhydrous. Compound **3** was obtained without any further purification. *Procedure B:* HCl (37%, 100 ml) was slowly added to a mixture of **2** (3.4 mmol) and $SnCl_2$ (13 mmol) in MeOH (100 ml). The crude mixture was refluxed for 30 min. A solution of K_2CO_3 (2M) was slowly added to reach pH 9. The precipitate was filtered and purified by flash chromatography.

4.2.11. 2-(4-Aminophenyl)-7-benzyl-5,6,7,8-tetrahydroimidazo [2′,1′:2,3]thiazolo[5,4-c]pyridine (**3a**)

From crude **2a** using the procedure B. Flash chromatography (AcOEt + 0.1% NEt_3). **3a** (26%), pale yellow solid. ^{1}H NMR ($CDCl_3$, 400 MHz): δ 7.60 (2H, d, *J* = 8.4 Hz), 7.39 (1H, s), 7.25–7.36 (5H, m), 6.69 (2H, d, *J* = 8.4 Hz), 3.73 (2H, s), 3.56 (2H, s), 2.95 (2H, t, *J* = 6.0 Hz), 2.71 (2H, t, *J* = 6.0 Hz). MS-EI *m*/*z* $[M]^+$ 360.

4.2.12. 2-(4-Aminophenyl)-5,6,7,8-tetrahydroimidazo[2,1-b] benzothiazole (**3b**)

From crude **2b** using the procedure A. **3b** (89%), yellow solid. ^{1}H NMR (DMSO-d_6, 300 MHz): δ 7.87 (1H, s), 7.50 (2H, d, *J* = 8.2 Hz), 6.58 (2H, d, *J* = 8.2 Hz), 5.13 (2H, s), 2.69–2.64 (4H, m), 1.89–1.85 (4H, m). MS-EI *m*/*z* $[M]^+$ 269.

4.2.13. 2-(4-Aminophenyl)imidazo[2,1-b]benzothiazole (**3c**)

From crude **2c** using the procedure A. **3c** (87%), pale yellow amorphous solid. ^{1}H NMR (DMSO-d_6, 300 MHz): δ 8.48 (1H, s), 8.05 (1H, d, *J* = 8.0 Hz), 7.93 (1H, d, *J* = 8.0 Hz), 7.61–7.51 (3H, m), 7.42 (1H, t, *J* = 8.0 Hz), 6.64 (2H, d, *J* = 9.0 Hz), 5.23 (2H, s). MS-EI *m*/*z* $[M]^+$ 265.

4.2.14. 2-(4-Aminophenyl)-6-benzyl-5,6,7,8-tetrahydroimidazo [2′,1′:2,3]thiazolo[4,5-c]pyridine (**3d**)

From **2d** using the procedure B. Flash chromatography (CH_2Cl_2 to CH_2Cl_2/MeOH = 98.5:1.5). **3d** (23%). R_f (CH_2Cl_2/MeOH = 95/5):

Please cite this article in press as: A. Furlan, et al., Identification of new aminoacid amides containing the imidazo[2,1-*b*]benzothiazol-2-ylphenyl moiety as inhibitors of tumorigenesis by oncogenic Met signaling, European Journal of Medicinal Chemistry (2011), doi:10.1016/j.ejmech.2011.10.051

0.18 .^{1}H NMR ($CDCl_3$ with drops of CD_3OD, 400 MHz) δ 7.52 (2H, d, $J = 8.8$ Hz), 7.33–7.39 (5H, m), 6.73 (2H, d, $J = 8.8$ Hz), 3.80 (2H, s), 3.57 (2H, t(dd), $J = 1.5$ Hz), 2.95 (2H, t(dd), $J = 5.2$ Hz), 2.82 (2H, t(dd), $J = 5.2$ Hz). ^{13}C NMR ($CDCl_3$ with drops of CD_3OD, 100.6 MHz) δ 148.2, 146.6, 145.7, 136.8, 128.8 (2C), 128.3 (2C), 127.4, 125.9 (2C), 124.2, 124.0, 119.3, 115.2 (2C), 104.1, 61.4, 49.9, 48.8, 24.2. HRMS calcd for $C_{21}H_{21}N_4S$ $[M + H]^+$: 361.1487; Found: 361.1480.

4.2.15. 2-(4-Aminophenyl)-7-bromoimidazo[2,1-b]benzothiazole (**3e**)

From **2e** using the procedure A. **3e** (92%), amorphous brown solid. ^{1}H NMR (DMSO-d_6, 400 MHz): δ 8.46 (1H, s), 8.31 (1H, d, $J = 1.9$ Hz), 7.89 (1H, d, $J = 8.6$ Hz), 7.73 (1H, dd, $J = 8.6, 1.9$ Hz), 7.53 (2H, d, $J = 8.5$ Hz), 6.62 (2H, d, $J = 8.5$ Hz), 5.21 (2H, s). ^{13}C NMR (DMSO-d_6, 100.6 MHz): δ 149.41, 149.07, 147.41, 132.39, 132.35, 130.52, 128.41, 126.94 (2C), 122.78, 117.31, 115.67, 115.08 (2C), 107.58. HR-EIMS calcd for $[C_{15}H_{10}BrN_3S]^+$ 342.9779; Found 342.9784.

4.2.16. General procedure for synthesis of (**4**)

Procedure A: A solution of a phenylacetic acid (3.7 mmol), DMAP (5.55 mmol) and DCC (5.55 mmol) in THF (40–70 ml) was stirred at room temperature (rt) for 30 min. The hydrochloride salt of the carboxy protected aminoacid (3.7 mmol) was added and the solution was stirred for 48 h at rt. The crude mixture was filtered over celite and the solvent was removed in vacuo. The product was purified by flash chromatography. The methyl ester (1eq) was dissolved in methanol. A solution of NaOH (1N, 17 ml) was added and the reaction mixture was stirred for 1 h at rt. H_2O was added and methanol only was evaporated under reduced pressure. The pH was reduced to 2 using HCl solution 1N. The crude material was extracted with ethyl acetate and the organic phase was dried with Na_2SO_4 anhydrous. The solvent was removed in vacuo to afford **4** without any further purification. *Procedure B*: A solution of phenylacetic acid (6 mmol), the hydrochloride salt of the carboxy protected aminoacid (6 mmol), DMAP (3 mmol) and DCI (10 mmol) in (50–80 ml) was stirred at r.t. for 14 h. The crude mixture was filtered over celite and the solvent was removed in vacuo. The product was purified by flash chromatography. Treatment with NaOH as reported in procedure A.

4.2.17. (S)-2-[2-(3,5-Dimethylphenyl)acetamido]-3-phenylpropanoic acid (**4a**)

From L-phenylalanine methyl ester hydrochloride and 3,5-dimethylphenylacetic acid with procedure A. Flash chromatography (Hex:AcOEt 2:1) to afford the methyl ester as white solid (70%). ^{1}H NMR ($CDCl_3$, 400 MHz): δ 7.24–7.18 (3H, m), 6.94–6.91 (3H, m), 6.82 (2H, s), 5.83 (1H, d, $J = 7.6$ Hz), 4.91–4.83 (1H, m), 3.66 (3H, s), 3.52 (1H, d, $J = 15.8$ Hz), 3.48 (1H, d, $J = 15.8$ Hz), 3.09 (1H, dd, $J = 13.8, 5.5$ Hz), 3.04 (1H, dd, $J = 13.8, 5.5$ Hz), 2.19 (6H, s). MS-EI m/z $[M]^+$ 325. The NaOH treatment gave **4a** (95%), white solid. mp 85–88 °C, $[\alpha]_D^{20} = +35.12$ (c 0.80, $CHCl_3$). ^{1}H NMR ($CDCl_3$, 400 MHz): δ 10 (1H, s), 7.24–7.19 (3H, m), 6.97 (2H, dd, $J = 7.2, 1.6$ Hz), 6.94 (1H, s), 6.78 (2H, s), 6.01 (1H, d, $J = 7.6$ Hz), 4.87–4.83 (1H, m), 3.52 (1H, d, $J = 16.4$ Hz), 3.48 (1H, d, $J = 16.4$ Hz), 3.16 (1H, dd, $J = 14.0, 5.6$ Hz), 3.06 (1H, dd, $J = 14.0, 5.6$ Hz), 2.3 (6H, s). ^{13}C NMR ($CDCl_3$, 100.6 MHz): δ 174.4, 172.2, 138.6 (2C), 135.4, 133.7, 129.7, 129.2 (2C), 128.5 (2C), 127.3 (2C), 127.1, 53.2, 43.2, 36.9, 21.2 (2C). HR-EIMS calcd for $C_{19}H_{21}NO_3$ 311.1521. Found 311.1523.

4.2.18. (R)-2-[2-(3,5-Dimethylphenyl)acetamido]-3-phenylpropanoic acid (**4b**)

From D-phenylalanine methyl ester hydrochloride and 3,5-dimethylphenylacetic acid with procedure A. Flash chromatography (Hex:AcOEt 2:1) to afford the methyl ester (68%) as white solid. ^{1}H NMR ($CDCl_3$, 400 MHz): δ 7.24–7.18 (3H, m), 6.94–6.91 (3H, m), 6.82 (2H, s), 5.83 (1H, d, $J = 7.6$ Hz), 4.91–4.83 (1H, m), 3.66 (3H, s), 3.52 (1H, d, $J = 15.8$ Hz), 3.48 (1H, d, $J = 15.8$ Hz), 3.09 (1H, dd, $J = 13.8, 5.5$ Hz), 3.04 (1H, dd, $J = 13.8, 5.5$ Hz), 2.19 (6H, s). MS-EI m/z $[M]^+$ 325. The NaOH treatment gave **4b** (98%), white solid. mp 85–88 °C, $[\alpha]_D^{20} = -38.52$ (c 1.28, $CHCl_3$). ^{1}H NMR ($CDCl_3$, 400 MHz): δ 10 (1H, s), 7.24–7.19 (3H, m), 6.97 (2H, dd, $J = 7.2, 1.6$ Hz), 6.94 (1H, s), 6.78 (2H, s), 6.01 (1H, d, $J = 7.6$ Hz), 4.87–4.83 (1H, m), 3.52 (1H, d, $J = 16.4$ Hz), 3.48 (1H, d, $J = 16.4$ Hz), 3.16 (1H, dd, $J = 14.0, 5.6$ Hz), 3.06 (1H, dd, $J = 14.0, 5.6$ Hz), 2.3 (6H, s). ^{13}C NMR ($CDCl_3$, 100.6 MHz): δ 174.4, 172.2, 138.6 (2C), 135.4, 133.7, 129.7, 129.2 (2C), 128.5 (2C), 127.2 (2C), 127.1, 53.2, 43.2, 36.9, 21.2 (2C). HR-EIMS calcd for $C_{19}H_{21}NO_3$ 311.1521. Found 311.1526.

4.2.19. (S)-2-[2-(3,5-Bis(trifluoromethyl)phenyl)acetamido]-3-phenylpropanoic acid (**4c**)

From L-phenylalanine methyl ester hydrochloride and 3,5-bis(trifluoromethyl)phenylacetic acid with procedure A. Flash chromatography (Hex:AcOEt 2:1) to afford the methyl ester (63%), white solid. ^{1}H NMR ($CDCl_3$, 400 MHz): δ 7.82 (1H, s), 7.72 (2H, s), 7.27–7.23 (3H, m), 7.01–6.98 (2H, m), 5.97 (1H, d, $J = 7.2$ Hz), 4.93–4.88 (1H, m), 3.79 (3H, s), 3.66 (1H, d, $J = 15.6$ Hz), 3.61 (1H, d, $J = 15.6$ Hz), 3.17 (1H, dd, $J = 13.6, 5.6$ Hz), 3.09 (1H, dd, $J = 13.6, 5.4$ Hz). MS-EI m/z 433 $[M]^+$. The treatment with NaOH gave **4c** (99%) as white solid. mp 138–140 °C, $[\alpha]_D^{20} = +10.17$ (c 1.28, MeOH). ^{1}H NMR (DMSO-d_6, 400 MHz): δ 8.52 (1H, d, $J = 8.2$ Hz), 7.96 (1H, s), 7.88 (2H, s), 7.22–7.13 (5H, m), 4.49–4.43 (1H, m), 3.65 (1H, d, $J = 15.3$ Hz), 3.57 (1H, d, $J = 15.3$ Hz), 3.09 (1H, dd, $J = 13.8, 4.6$ Hz), 2.86 (1H, dd, $J = 13.8, 9.8$ Hz).^{13}C NMR (DMSO-d_6, 100.6 MHz): δ 173.9, 169.9, 140.7, 138.5, 131.1 (2C, q, $J = 32.8$ Hz), 130.9 (2C), 130.1 (2C), 129.1 (2C), 127.4, 124.5 (2C, q, $J = 272.6$ Hz), 121.2, 54.6, 42.2, 37.9. HR-EIMS calcd. for $C_{19}H_{15}F_6NO_3$ 419.0956. Found 491.0952.

4.2.20. (R)-2-[2-(3,5-Bis(trifluoromethyl)phenyl)acetamido]-3-phenylpropanoic acid (**4d**)

From D-phenylalanine methyl ester hydrochloride and 3,5-bis(trifluoromethyl)phenylacetic acid with procedure A. Flash chromatography (Hex:AcOEt 2:1) to afford the methyl ester as white solid (62%). ^{1}H NMR ($CDCl_3$, 400 MHz): δ 7.82 (1H, s), 7.72 (2H, s), 7.27–7.23 (3H, m), 7.01–6.98 (2H, m), 5.97 (1H, d, $J = 7.2$ Hz), 4.93–4.88 (1H, m), 3.79 (3H, s), 3.66 (1H, d, $J = 15.6$ Hz), 3.61 (1H, d, $J = 15.6$ Hz), 3.17 (1H, dd, $J = 13.6, 5.6$ Hz), 3.09 (1H, dd, $J = 13.6, 5.4$ Hz). MS-EI m/z $[M]^+$ 433. The treatment with NaOH gave **4d** (99%), white solid. mp 138–140 °C, $[\alpha]_D^{20} = +10.5$ (c 0.21, MeOH). ^{1}H NMR (DMSO-d_6, 400 MHz): δ 8.52 (1H, d, $J = 8.2$ Hz), 7.96 (1H, s), 7.88 (2H, s), 7.22–7.13 (5H, m), 4.49–4.43 (1H, m), 3.65 (1H, d, $J = 15.3$ Hz), 3.57 (1H, d, $J = 15.3$ Hz), 3.09 (1H, dd, $J = 13.8, 4.6$ Hz), 2.86 (1H, dd, $J = 13.8, 9.8$ Hz).^{13}C NMR (DMSO-d_6, 100.6 MHz): δ 173.8, 169.9, 140.7, 138.5, 131.1 (2C, q, $J = 32.8$ Hz), 130.9 (2C), 130.1 (2C), 129.1 (2C), 127.4, 124.5 (2C, q, $J = 272.6$ Hz), 121.2, 54.6, 42.1, 37.9. HR-EIMS calcd for $C_{19}H_{15}F_6NO_3$ 419.0955. Found 491.0952.

4.2.21. (S)-2-[2-(3,5-Difluorophenyl)acetamido]-3-phenylpropanoic acid (**4e**)

From L-phenylalanine methyl ester hydrochloride and 3,5-difluorophenylacetic acid with procedure A. Flash chromatography (Hex:AcOEt 2:1) to afford the methyl ester as white solid (75%). ^{1}H NMR ($CDCl_3$, 300 MHz): δ 7.25–7.23 (3H, m), 6.96–6.92 (2H, m), 6.74–6.70 (3H, m), 5.83 (1H, d, $J = 6.8$ Hz), 4.88–4.82 (1H, m), 3.73 (3H, s), 3.50 (1H, d, $J = 14.2$ Hz), 3.47 (1H, d, $J = 14.2$ Hz), 3.12 (1H, dd, $J = 13.8, 5.6$ Hz), 3.02 (1H, dd, $J = 13.8, 5.9$ Hz). MS-EI m/z 333 $[M]^+$. The treatment with NaOH gave **4e** (91%) as white solid. mp 182–184 °C, $[\alpha]_D^{20} = +33.63$ (c 0.96, MeOH:$CHCl_3$ = 1:1). ^{1}H NMR (DMSO-d_6, 400 MHz): δ 8.48 (1H, d, $J = 8.2$ Hz), 7.26–7.18 (5H, m), 7.09–7.03 (1H, m), 6.87–6.83 (2H, m), 4.48–4.42 (1H, m),

Please cite this article in press as: A. Furlan, et al., Identification of new aminoacid amides containing the imidazo[2,1-*b*]benzothiazol-2-ylphenyl moiety as inhibitors of tumorigenesis by oncogenic Met signaling, European Journal of Medicinal Chemistry (2011), doi:10.1016/j.ejmech.2011.10.051

3.49 (1H, d, J = 14.2 Hz), 3.43 (1H, d, J = 14.2 Hz), 3.09 (1H, dd, J = 13.8, 4.7 Hz), 2.86 (1H, dd J = 13.8, 9.8 Hz). ^{13}C NMR (DMSO-d_6, 100.6 MHz): δ 173.4, 169.4, 162.5 (2C, dd, J = 245.5, 14.1 Hz), 140.9 (t, J = 10.0 Hz), 137.9, 129.5 (2C), 128.5 (2C), 126.8, 112.6 (2C, d, J = 25.2 Hz), 102.2 (2C, t, J = 25.2 Hz), 53.9, 41.8, 37.1. HR-EIMS calcd for $C_{17}H_{15}F_2NO_3$ 319.1020. Found 319.1019.

*4.2.22. 2-[2-(3,5-Dimethylphenyl)acetamido]acetic acid (**4f**)*

From glycine methyl ester hydrochloride and 3,5-dimethylphenylacetic acid with procedure A. Flash chromatography (Hex:AcOEt 2:3) gave the methyl ester as white solid (52%). ^{1}H NMR ($CDCl_3$, 300 MHz): δ 6.92 (1H, s), 6.88 (2H, s), 5.92 (1H, d, J = 5.1 Hz), 3.99 (2H, d, J = 5.1 Hz), 3.72 (3H, s), 3.54 (2H, s), 2.30 (6H, s). MS-EI m/z 235 $[M]^+$. Treatment with NaOH gave **4f** (95%) as white solid. mp 129–131 °C ^{1}H NMR (DMSO-d_6, 400 MHz): δ 8.30 (1H, t, J = 5.7 Hz), 6.88, (2H, s), 6.85 (1H, s), 3.75 (2H, d, J = 5.7 Hz), 3.39 (2H, s), 2.23 (6H, s). ^{13}C NMR (DMSO-d_6, 100.6 MHz): δ 171.7, 171.0, 137.5 (2C), 136.4, 128.2, 127.3 (2C), 42.3, 41.2, 21.3 (2C). HR-EIMS calcd for $C_{12}H_{15}NO_3$ 221.1051. Found 221.1053.

*4.2.23. (S)-2-[2-(3,5-Dimethylphenyl)acetamido]-3-(4-hydroxyphenyl)propanoic acid (**4g**)*

From L-tyrosine methyl ester hydrochloride and 3,5-dimethylphenylacetic acid with procedure A. Flash chromatography (Hex:AcOEt 2:3) gave the methyl ester as white solid (43%). ^{1}H NMR ($CDCl_3$, 400 MHz): δ 7.12–7.05 (3H, m), 6.80–6.78 (2H, m), 6.69 (2H, d, J = 9.0 Hz), 5.71 (1H, d, J = 7.6 Hz), 4.85–4.79 (1H, m), 3.53 (3H, s), 3.41 (1H, d, J = 15.7 Hz), 3.34 (1H, d, J = 15.7 Hz), 2.98 (1H, dd, J = 13.8, 5.8 Hz), 2.96 (1H, dd, J = 13.8, 5.8 Hz), 2.07 (6H, s). MS-EI m/z 341 $[M]^+$. The treatment with NaOH gave **4g** (97%) as white solid. mp 145–148 °C, $[\alpha]_D^{20}$ = +22.0 (c 1.35, MeOH). ^{1}H NMR ($CDCl_3$, 400 MHz): δ 7.12–7.07 (3H, m), 6.91 (2H, dd, J = 7.1, 1.7 Hz), 6.68 (2H, d, J = 9.01 Hz), 5.86 (1H, d, J = 7.6 Hz), 4.91–4.85 (1H, m), 3.59 (1H, d, J = 16.4 Hz), 3.51 (1H, d, J = 16.4 Hz), 3.23 (1H, dd, J = 14.0, 5.6 Hz), 3.06 (1H, dd, J = 14.0, 5.6 Hz), 2.3 (6H, s). ^{13}C NMR ($CDCl_3$, 100.6 MHz): δ 174.3, 172.2, 155.6, 138.6 (2C), 133.7, 130.3, 129.7, 129.4 (2C), 127.3 (2C), 115.7 (2C), 53.2, 43.2, 36.9, 21.2 (2C). EIMS m/z 327 (M^+).

*4.2.24. 3-[2-(3,5-Difluorophenyl)acetamido]propanoic acid (**4h**)*

From β-alanine methyl ester hydrochloride and 3,5-difluorophenylacetic acid with procedure B. Flash chromatography (hexane/ethyl acetate 2:3) gave the methyl ester (75%). ^{1}H NMR ($CDCl_3$, 400 MHz): δ 6.73 (2H, tt, J = 6.8, 2.4 Hz), 6.73 (1H, tt, J = 11.2, 2.4 Hz), 6.22 (1H, brs), 3.67 (3H, s), 3.50 (2H, t, J = 6.0 Hz), 2.53 (2H, t, J = 6.0 Hz). HR-EIMS calcd for $C_{12}H_{13}F_2NO_3$ 258.0941. Found 258.0946. The treatment with NaOH gave **4h** (81%) as amorphous solid. ^{1}H NMR ($CDCl_3$, 400 MHz): δ 7.02 (1H, brs), 6.74 (2H, tt, J = 6.6, 2.4 Hz), 6.63 (1H, tt, J = 11.2, 2.4 Hz), 3.40 (3H, s), 3.73 (2H, t, J = 6.0 Hz), 2.42 (2H, t, J = 6.0 Hz). ^{13}C NMR ($CDCl_3$, 100.6 MHz): δ 172.4, 170.4, 164.1 (d, J = 124.7 Hz), 161.6 (d, J = 124.7 Hz); 138.4 (t, J = 92.6 Hz), 112.1 (d, J = 70.4 Hz), 111.8 (d, J = 70.4 Hz), 102.4 (t, J = 245.8 Hz), 35.1, 33.4. HR-ESIMS calcd for $C_{11}H_{12}F_2NO_3$ 244.0780 ($M + H^+$). Found 244.0785.

*4.2.25. (R)-2-[2-(3,5-Difluorophenyl)acetamido]-3-(tritylthio) propanoic acid (**4i**)*

L-Cysteine methyl ester hydrochloride (1.00 g, 5.83 mmol) was dissolved in 6 ml of CH_2Cl_2, then trityl chloride (3.57 g, 12.8 mmol) was added and the reaction mixture was stirred at r.t. for 42 h and was quenched with the addition of 0.5 M HCl solution. The not reacted trityl chloride was eliminated with ethyl acetate extraction, then the aqueous layer was basified to pH 9 with saturated solution of $NaHCO_3$ and the product was extracted with AcOEt. The organic layer was washed with water, brine, filtered and concentrated that furnished the desired product. Viscous oil (1.19 g, 60%). ^{1}H NMR ($CDCl_3$, 300 MHz): δ 7.41–7.45 (6H, m), 7.18–7.28 (9H, m), 3.65 (3H, s), 3.20 (1H, dd, J = 7.5, 4.8 Hz), 2.59 (1H, dd, J = 12.6, 4.8 Hz), 2.47 (1H, dd, J = 12.6, 7.5 Hz). HR-FABMS calcd for $C_{30}H_{25}F_2NO_3S$: 400.1347 $[M + Na]^+$. Found 400.1345. S-trityl-L-cysteine methyl ester hydrochloride and 3,5-difluorophenylacetic acid were reacted according the general procedure B. The crude product was directly treated with NaOH to give **4i** (83%) as amorphous solid. ^{1}H NMR ($CDCl_3$, 300 MHz): δ 7.34–7.37 (6H, m), 7.18–7.28 (9H, m), 6.81 (2H, tt, J = 6.6, 2.4 Hz), 6.68 (1H, tt, J = 11.2, 2.4 Hz), 6.58 (1H, d, J = 9.6 Hz), 4.47 (1H, m), 3.69 (2H, s), 2.67 (2H, d, J = 7.6 Hz). HR-FABMS calcd for $C_{30}H_{25}F_2NO_3SNa$ 540.1421 $[M + Na]^+$. Found 540.1422.

*4.2.26. (S)-2-[(3,5-Difluorophenyl)acetamido]propanoic acid (**4j**)*

From L-alanine methyl ester hydrochloride and 3,5-difluorophenylacetic with the procedure B. Flash chromatography (Ethyl acetate/Hexane = 1/1) gave the methyl ester (82%). ^{1}H NMR ($CDCl_3$, 400 MHz) δ 6.83 (2H, tt(dddd), J = 6.6, 2.4 Hz), 6.73 (1H, tt(dddd), J = 11.2, 2.4 Hz), 6.26 (1H, d, J = 5.2 Hz), 4.58 (1H, qd (dddd), J = 7.2 Hz), 3.74 (3H, s), 3.55 (2H, s), 1.39 (3H, d, J = 7.2 Hz). ^{13}C NMR ($CDCl_3$, 100.6 MHz) δ 173.3, 168.9, 164.3 (d, J_{C-F} = 131.8 Hz), 161.8 (d, J_{C-F} = 131.8 Hz), 138.2 (t(dd), J_{C-F} = 93.1 Hz), 112.3 (d, J_{C-F} = 70.4 Hz), 112.1 (d, J_{C-F} = 70.4 Hz), 102.8 (t(dd), J_{C-F} = 253 Hz), 52.5, 48.2, 42.8, 18.2. The treatment with NaOH gave **4j** (68%) as amorphous solid. ^{1}H NMR ($CDCl_3$, 300 MHz) δ 6.82 (2H, tt(dddd), J = 6.6, 2.1 Hz), 6.75 (1H, tt(dddd), J = 9.3, 2.1 Hz), 6.03 (1H, d, J = 6.0 Hz), 4.60 (1H, qd (dddd), J = 7.2 Hz), 3.57 (2H, s), 1.45 (3H, d, J = 7.2 Hz). ^{13}C NMR (100.6 MHz, CD_3OD) δ 175.9, 172.4, 165.6 (d, J_{C-F} = 49.9 Hz), 163.2 (d, J_{C-F} = 49.5 Hz), 141.0 (t, J_{C-F} = 40.0 Hz), 113.2 (d, J_{C-F} = 28.0 Hz), 113.0 (d, J_{C-F} = 28.0 Hz), 102.9 (t, J_{C-F} = 102.0 Hz), 48.2, 42.7, 17.5.

*4.2.27. (R)-2-[(3,5-Difluorophenyl)acetamido]propanoic acid (**4k**)*

The application of the procedure described for the preparation of **4j** and the use of D-alanine methyl ester hydrochloride as starting material resulted in the obtainment of **4k**.

*4.2.28. (S)-2-[2-(4-Methylphenyl)acetamido]-3-phenylpropanoic acid (**4l**)*

From L-phenylalanine methyl ester hydrochloride and *p*-tolylacetic acid with procedure B. Flash chromatography (MeOH:CH_2Cl_2 = 95:5) gave the methyl ester. ^{1}H NMR ($CDCl_3$, 300 MHz) δ 7.18–7.20 (3H, m), 7.13 (2H, d, J = 7.8 Hz), 7.06 (2H, d, J = 7.8 Hz), 6.89 (2H, dd, J = 2.4, 7.2 Hz), 5.78 (1H, d, J = 6.6 Hz), 4.81–4.88 (1H, m), 3.70 (3H, s), 3.51 (2H, s), 3.06 (1H, dd, J = 6.0, 13.6 Hz), 2.98 (1H, dd, J = 6.0, 13.6 Hz), 2.35 (3H, s). The treatment with NaOH gave **4l** (95%) as amorphous solid. ^{1}H NMR ($CDCl_3$, 300 MHz) δ 7.19–7.21 (3H, m), 7.11 (2H, d, J = 7.8 Hz), 7.01 (2H, d, J = 7.8 Hz), 6.94 (2H, dd, J = 2.2, 7.6 Hz), 5.80 (1H, d, J = 6.8 Hz), 4.75–4.82 (1H, m), 3.51 (2H, s), 3.14 (1H, dd, J = 5.4, 14.0 Hz), 3.02 (1H, dd, J = 6.8, 14.0 Hz), 2.35 (3H, s).

*4.2.29. General procedures for the preparation of compounds **5**–**8***

Procedure A: A solution of **4** (0.35 mmol), HOBt (0.38 mmol), BOPCl (0.38 mmol) and DIPEA (1.4 mmol) in CH_2Cl_2 was stirred for 10 min at r.t. A solution of amine **3** (0.35 mmol) in CH_2Cl_2 was added and the mixture was stirred for 48 h at r.t. After treatment with $NaHCO_3$ (sat. sol.) the mixture was washed with AcOEt. The organic layers were washed with brine and dried on $MgSO_4$. *Procedure B:* HATU (0.38 mmol) and DIPEA (0.7 mmol) were added to a solution of **4** (0.38 mmol) in THF. The solution was stirred for 15 min at rt. After the addition of **3** (0.35 mmol) the solution was stirred overnight at r.t. *Procedure C:* A solution of **4** (0.35 mmol), DCI (0.35 mmol), HOBt (0.35 mmol) in CH_2Cl_2 was stirred for 15 min. A solution of amine **3** (0.62 mmol) in CH_2Cl_2 was added and the

Please cite this article in press as: A. Furlan, et al., Identification of new aminoacid amides containing the imidazo[2,1-*b*]benzothiazol-2-ylphenyl moiety as inhibitors of tumorigenesis by oncogenic Met signaling, European Journal of Medicinal Chemistry (2011), doi:10.1016/j.ejmech.2011.10.051

solution war stirred for 17 h at rt. After treatment with $NaHCO_3$ (sat. sol.) the mixture was washed with AcOEt. The organic layers were washed with brine and dried on $MgSO_4$.

*4.2.30. (S)-N-[4-(7-Benzyl-5,6,7,8-tetrahydroimidazo[2′,1′:2,3] thiazolo[5,4-c]pyridin-2-yl)phenyl]-2-[2-(3,5-dimethylphenyl) acetamido]-3-phenylpropanamide (**5a**)*

From **3a** and **4a** using general procedure A. Flash chromatography (MeOH:CH_2Cl_2 1:99) to give **5a** (55%) as amorphous solid; $[\alpha]_D^{20} = +3.60$ (c 0.23, MeOH). ^{1}H NMR ($CDCl_3$ with drops of CD_3OD, 400 MHz): δ 9.37 (1H, brs), 7.69 (2H, d, $J = 8.8$ Hz), 7.56 (1H, s), 7.47 (2H, d, $J = 8.4$ Hz), 7.38 (1H, s), 7.37 (1H, d, $J = 1.6$ Hz), 7.32–7.47 (3H, m), 7.20–7.24 (3H, m), 7.10 (2H, dd, $J = 7.2$, 2.4 Hz), 6.91 (1H, s), 6.82 (1H, d, $J = 8.0$ Hz), 6.78 (2H, s), 4.75 (1H, dt, $J = 14.8$, 7.2 Hz), 3.79 (2H, 2H), 3.62 (2H, s), 3.46 (2H, s), 3.09 (1H, dd, $J = 14.0$, 6.8 Hz), 3.02 (2H, t, $J = 5.6$ Hz), 2.98 (1H, dd, $J = 14.0$, 6.8 Hz), 2.80 (2H, t, $J = 5.6$ Hz), 2.28 (6H, s). ^{13}C NMR ($CDCl_3$ with drops of CD_3OD, 100.6 MHz): δ 172.01, 169.29, 148.34, 145.89, 138.21, 136.84, 136.83, 136.66, 135.95, 133.89, 129.91, 129.04, 128.98, 128.75, 128.37, 128.24, 127.50, 126.82, 126.67, 125.34, 124.53, 120.23, 120.14, 118.88, 105.56, 61.01, 54.72, 50.02, 48.48, 42.95, 38.02, 22.53, 20.90 (2C); HR EIMS calcd for $C_{40}H_{40}N_5O_2S$ m/z $[M + H]^+$ 654.2902. Found 654.2889.

*4.2.31. (R)-N-[4-(7-Benzyl-5,6,7,8-tetrahydroimidazo[2′,1′:2,3] thiazolo[5,4-c]pyridin-2-yl)phenyl]-2-[2-(3,5-dimethylphenyl) acetamido]-3-phenylpropanamide (**5b**)*

From **3a** and **4b** using procedure **A**. Flash chromatography (MeOH:CH_2Cl_2 1:99) to give **5b** (45%) as a white amorphous solid $[\alpha]_D^{20} = -3.40$ (c 0.20, MeOH). ^{1}H NMR ($CDCl_3$ with drops of CD_3OD, 400 MHz): δ 9.37 (1H, brs), 7.69 (2H, d, $J = 8.8$ Hz), 7.56 (1H, s), 7.47 (2H, d, $J = 8.4$ Hz), 7.38 (1H, s), 7.37 (1H, d, $J = 1.6$ Hz), 7.32–7.47 (3H, m), 7.20–7.24 (3H, m), 7.10 (2H, dd, $J = 7.2$, 2.4 Hz), 6.91 (1H, s), 6.82 (1H, d, $J = 8.0$ Hz), 6.78 (2H, s), 4.75 (1H, dt, $J = 14.8$, 7.2 Hz), 3.79 (2H, 2H), 3.62 (2H, s), 3.46 (2H, s), 3.09 (1H, dd, $J = 14.0$, 6.8 Hz), 3.02 (2H, t, $J = 5.6$ Hz), 2.98 (1H, dd, $J = 14.0$, 6.8 Hz), 2.80 (2H, t, $J = 5.6$ Hz), 2.28 (6H, s). ^{13}C NMR ($CDCl_3$ with drops of CD_3OD, 100.6 MHz): δ 172.01, 169.29, 148.34, 145.89, 138.21, 136.84, 136.83, 136.66, 135.95, 133.89, 129.91, 129.04, 128.98, 128.75, 128.37, 128.24, 127.50, 126.82, 126.67, 125.34, 124.53, 120.23, 120.14, 118.88, 105.56, 61.01, 54.72, 50.02, 48.48, 42.95, 38.02, 22.53, 20.90 (2C). HR EIMS calcd for $C_{40}H_{40}N_5O_2S$ m/z $[M + H]^+$ 654.2902. Found 654.2891.

*4.2.32. (S)-N-[4-(7-Benzyl-5,6,7,8-tetrahydroimidazo[2′,1′:2,3] thiazolo[5,4-c]pyridin-2-yl)phenyl]-3-[2-(3,5-difluorophenyl) acetamido]propanamide (**5c**)*

From **3a** and **4h** using procedure **A**. Flash chromatography (MeOH:CH_2Cl_2 3:97) to give **5c** (70%) as an amorphous solid. ^{1}H NMR ($CDCl_3$ with drops of CD_3OD, 400 MHz): δ 9.48 (1H brs), 7.72 (2H, dt, $J = 9.2$, 1.6 Hz), 7.58 (1H, s), 7.55 (1H, dt, $J = 9.2$, 1.6 Hz), 7.30–7.38 (5H, m), 6.82 (1H, tt, $J = 6.4$, 2.4 Hz), 6.69 (1H, tt, $J = 11.2$, 2.4 Hz), 3.80 (2H, s), 3.64 (2H, s), 3.52 (2H, dd, $J = 12.0$; 6.4 Hz), 3.48 (2H, s), 3.03 (2H, t, $J = 5.6$ Hz), 2.83 (2H, t, $J = 5.6$ Hz), 2.57 (2H, t, $J = 6.4$ Hz). ^{13}C NMR ($CDCl_3$ with drops of CD_3OD, 100.6 MHz): δ 172.96, 170.27, 163.97 (d, $J = 123.7$ Hz), 161.52 (d, $J = 132.8$ Hz), 148.32, 145.90, 138.48 (t, $J = 92.6$ Hz), 137.13, 136.84, 129.61, 128.94, 128.32, 127.45, 124.58, 119.93, 118.92, 111.90 (d, $J = 70.4$ Hz), 111.72 (d, $J = 69.4$ Hz), 105.56, 102.21 (t, $J = 253.0$ Hz), 61.00, 50.00, 48.46, 42.41, 35.96, 33.74, 22.51. HR EIMS calcd for $C_{32}H_{29}F_2N_5O_2S$ m/z $[M + H]^+$ 586.2088. Found 586.2077.

*4.2.33. (R)-N-[4-(7-Benzyl-5,6,7,8-tetrahydroimidazo[2′,1′:2,3] thiazolo[5,4-c]pyridin-2-yl)phenyl]-2-[2-(3,5-difluorophenyl) acetamido]-3-sulfanylpropanamide (**5d**)*

From **3a** and **4i** using procedure C. Flash chromatography (MeOH:CH_2Cl_2 1:99) to give **5d** (70%) as white solid. ^{1}H NMR ($CDCl_3$, 300 MHz): δ 8.36 (1H, brs), 7.67 (1H, dt, $J = 8.1$, 1.5 Hz), 7.47 (1H, s), 7.15–7.40 (22H, m), 6.87 (1H, brs), 6.73 (2H, dt, $J = 6.4$, 2.4 Hz), 6.66 (tt, $J = 6.4$, 2.4 Hz), 4.24 (1H, tt, $J = 11.2$, 2.4 Hz), 3.75 (2H, s), 3.57 (2H, s), 3.41 (2H, s), 2.97 (2H, t, $J = 5.4$ Hz), 2.73 (2H, t, $J = 5.4$ Hz), 2.67 (2H, d, $J = 6.3$ Hz). S-trityl protected **5d** (0.27 g, 0.29 mmol) was dissolved in anhydrous CH_2Cl_2, then were added TFA (2.3 ml, 2.92 mmol) and Et_3SiH (0.11 g, 0.96 mmol). The reaction was stirred at rt for 1 h. The mixture was concentrated under vacuum and dissolved in AcOEt, washed with $NaHCO_3$ satured solution, water and brine. The crude was filtered, concentrated under vacuum and purified by flash chromatography (AcOEt:Hex 1:1) to afford **5d** as pale yellow amorphous solid (0.08 g, yield 40%). $[\alpha]_D^{20} = +2.30$ (c 0.17, CH_2Cl_2). ^{1}H NMR ($CDCl_3$ with drops of CD_3OD, 400 MHz): δ 9.48 (1H, brs), 7.78 (1H, d, $J = 8.0$ Hz), 7.73 (1H, dt, $J = 9.2$, 1.6 Hz), 7.58 (1H, s), 7.56 (1H, dt, $J = 9.2$, 1.6 Hz), 7.30–7.38 (5H, m), 6.88 (2H, tt, $J = 6.4$, 2.4 Hz), 6.73 (1H, tt, $J = 11.2$, 2.4 Hz), 3.80 (2H, s), 3.66 (2H, s), 3.63 (1H, s), 3.60 (2H, s), 3.37 (1H, qd, $J = 5.6$ Hz), 3.03 (2H, t, $J = 5.6$ Hz), 2.95 (1H, dd, $J = 9.6$, 6.0 Hz), 2.87 (1H, dd, $J = 9.6$, 6.0 Hz), 2.82 (2H, t, $J = 5.6$ Hz,). ^{13}C NMR ($CDCl_3$ with drops of CD_3OD, 100.6 MHz): δ 170.61, 168.12, 164.00 (d, $J = 132.8$ Hz), 161.52 (d, $J = 123.7$ Hz), 148.32, 145.73, 138.18 (t, $J = 98.6$ Hz), 136.74, 136.55, 130.24, 128.93, 128.29, 127.43, 125.33, 124.55, 120.16, 118.91, 112.00 (d, $J = 70.4$ Hz), 111.81 (d, $J = 62.4$ Hz), 105.67, 102.30 (t, $J = 253.0$ Hz), 60.95, 55.39, 49.98, 48.34, 42.03, 26.06, 22.45. HR EIMS calcd for $C_{32}H_{29}F_2N_5O_2S_2$ m/z $[M + H]^+$ 618.1809. Found 618.1792.

*4.2.34. (S)-N-[4-(5,6,7,8-Tetrahydroimidazo[2,1-b]benzothiazol-2-yl)phenyl]-2-[2-(3,5-dimethylphenyl)acetamido]-3-phenylpropanamide (**6**)*

From **3b** and **4a** using procedure B. Flash chromatography (EtOH:CH_2Cl_2 1:30) to give **6** (91%) as a white solid. mp 225–228 °C, $[\alpha]_D^{20} = +5.79$ (c 0.96, $CHCl_3$). ^{1}H NMR (DMSO-d_6, 400 MHz): δ 10.14 (1H, s), 8.44 (1H, d, $J = 8.2$ Hz), 8.10 (1H, s), 7.77 (2H, d, $J = 8.6$ Hz), 7.60 (2H, d, $J = 8.6$ Hz), 7.31–7.19 (5H, m), 6.81 (1H, s), 6.75 (2H, s), 4.70–4.67 (1H, m), 3.39 (1H, d, $J = 14.1$ Hz), 3.32 (1H, d, $J = 14.1$ Hz), 3.07 (1H, dd, $J = 14.2$, 4.9 Hz), 2.90 (1H, dd, $J = 14.2$, 9.6 Hz), 2.72–2.68 (4H, m), 2.20 (6H, bs), 1.89–1.85 (4H, m). ^{13}C NMR (DMSO-d_6, 100.6 MHz): δ 171.31, 171.13, 148.12, 146.22, 138.74, 138.63, 138.05, 137.11 (2C), 131.12, 131.97, 130.32 (2C), 129.15 (2C), 128.78, 127.88 (2C), 127.45, 126.03 (2C), 121.57, 120.69 (2C), 107.97, 55.97, 43.09, 38.90, 24.91, 23.83, 23.28, 22.30, 21.97(2C). HR EIMS calcd for $C_{34}H_{34}N_4O_2S$ m/z $[M + H]^+$ 562.2402. Found 562.2408.

*4.2.35. (S)-N-[4-(Imidazo[2,1-b]benzothiazol-2-yl)phenyl]-2-[2-(3,5-dimethylphenyl)acetamido]-3-phenylpropanamide (**7a**)*

From **3c** and **4a** (0.70 g, 2.25 mmol) using procedure B. Flash chromatography (EtOH:CH_2Cl_2 1:30) and crystallization with $CHCl_3$ to give **7a** (75%) as white solid. Mp 224–226 °C, $[\alpha]_D^{20} = +4.52$ (c 1.28, $CHCl_3$). ^{1}H NMR ($CDCl_3$, 400 MHz): δ 8.60 (1H, s), 7.92 (1H, s), 7.79 (2H, d, $J = 8.4$ Hz), 7.70 (1H, d, $J = 7.8$ Hz), 7.61 (1H, d, $J = 7.8$ Hz), 7.48–7.44 (3H, m), 7.37 (1H, t, $J = 7.8$ Hz), 7.29–7.21 (3H, m), 7.12–7.10 (2H, m), 6.95 (1H, s), 6.78 (2H, s), 6.29 (1H, d, $J = 7.6$ Hz), 4.92–4.86 (1H, m), 3.51–3.47 (2H, m), 3.14 (1H, dd, $J = 14.0$, 6.4 Hz), 3.05 (1H, dd, $J = 14.0$, 7.8 Hz), 2.30 (6H, s); ^{13}C NMR ($CDCl_3$, 100.6 MHz): δ 172.60, 170.02, 148.34, 146.43, 139.19, 138.16, 137.08, 134.73 (2C), 132.56, 131.01, 130.79, 129.91 (2C), 129.76, 129.22 (2C), 127.78 (2C), 127.56, 127.16, 126.44 (2C), 125.91, 124.80, 121.00 (2C), 113.68, 107.62, 55.99, 44.11, 38.48, 21.84 (2C). HR EIMS calcd for $C_{34}H_{30}N_4O_2S$ m/z $[M + H]^+$ 558.2089. Found 558.2086.

*4.2.36. (R)-N-[4-(Imidazo[2,1-b]benzothiazol-2-yl)phenyl]-2-[2-(3,5-dimethylphenyl)acetamido]-3-phenylpropanamide (**7b**)*

From **3c** and **4b** using procedure B. Flash chromatography (EtOH:CH_2Cl_2 1:30) and crystallization with $CHCl_3$ to give **7b** (78%)

Please cite this article in press as: A. Furlan, et al., Identification of new aminoacid amides containing the imidazo[2,1-*b*]benzothiazol-2-ylphenyl moiety as inhibitors of tumorigenesis by oncogenic Met signaling, European Journal of Medicinal Chemistry (2011), doi:10.1016/j.ejmech.2011.10.051

as white solid. Mp 224–226 °C, $[\alpha]_D^{20} = -4.28$ (c 0.83, $CHCl_3$). ^1H NMR ($CDCl_3$, 400 MHz): δ 8.60 (1H, s), 7.92 (1H, s), 7.79 (2H, d, $J = 8.4$ Hz), 7.70 (1H, d, $J = 7.8$ Hz), 7.61 (1H, d, $J = 7.8$ Hz), 7.48–7.44 (3H, m), 7.37 (1H, t, $J = 7.8$ Hz), 7.29–7.21 (3H, m), 7.12–7.10 (2H, m), 6.95 (1H, s), 6.78 (2H, s), 6.29 (1H, d, $J = 7.6$ Hz), 4.92–4.86 (1H, m), 3.51–3.47 (2H, m), 3.14 (1H, dd, $J = 14.0$, 6.4 Hz), 3.05 (1H, dd, $J = 14.0$, 7.8 Hz), 2.30 (6H, s); ^{13}C NMR ($CDCl_3$, 100.6 MHz): δ 172.60, 170.02, 148.34, 146.43, 139.19, 138.16, 137.08, 134.73 (2C), 132.56, 131.01, 130.79, 129.91 (2C), 129.76, 129.22 (2C), 127.78 (2C), 127.56, 127.16, 126.44 (2C), 125.91, 124.80, 121.00 (2C), 113.68, 107.62, 55.99, 44.11, 38.48, 21.84 (2C). HR EIMS calcd for $C_{34}H_{30}N_4O_2S$ *m/z* $[M + H]^+$ 558.2089. Found 558.2088.

4.2.37. (S)-N-[4-(Imidazo[2,1-b]benzothiazol-2-yl)phenyl]-2-[2-(3,5-bis(trifluoromethyl)phenyl) acetamido]-3-phenylpropanamide (**7c**)

From **3c** and **4c** using procedure B. Flash chromatography (Hex:AcOEt 1:1) and crystallization with CH_2Cl_2 to afford **7c** (72%) as white solid. Mp 243–245 °C, $[\alpha]_D^{20} = +1.54$ (c 0.31, DMSO). ^1H NMR (DMSO-d_6, 400 MHz): δ 10.25 (1H, s), 8.71–8.69 (2H, m), 8.04 (1H, d, $J = 8.0$ Hz), 7.98 (1H, d, $J = 8.0$ Hz), 7.96 (1H, s), 7.90 (2H, s), 7.82 (2H, d, $J = 8.6$ Hz), 7.66 (2H, d, $J = 8.6$ Hz), 7.58 (1H, t, $J = 8.0$ Hz), 7.44 (1H, t, $J = 8.0$ Hz), 7.29–7.13 (5H, m), 4.75–4.70 (1H, m), 3.74 (1H, d, $J = 15.1$ Hz), 3.70 (1H, d, $J = 15.1$ Hz), 3.10 (1H, dd, $J = 13.8$, 4.8 Hz), 2.89 (1H, dd, $J = 13.8$, 9.7 Hz). ^{13}C NMR (DMSO-d_6, 100.6 MHz): δ 172.13, 169.45, 146.82, 145.91, 141.12, 139.44, 138.87, 133.21, 131.98 (2C, q, $J = 33.2$ Hz) 131.80, 131.01 (2C), 130.51, 130.23 (2C), 129.03 (2C), 127.79, 127.41, 126.23 (2C), 126.18, 126.10, 123.42 (2C, q, $J = 271.0$ Hz), 121.79, 120.84 (2C), 114.37, 109.56, 56.04, 42.60, 38.65. HR EIMS calcd for $C_{34}H_{24}F_6N_4O_2S$ *m/z* $[M + H]^+$ 666.1524. Found 666.1526.

4.2.38. (R)-N-[4-(Imidazo[2,1-b]benzothiazol-2-yl)phenyl]-2-[2-(3,5-bis(trifluoromethyl)phenyl) acetamido]-3-phenylpropanamide (**7d**)

From **3c** and **4d** using procedure B. Flash chromatography (Hex:AcOEt 1:1) and crystallization with CH_2Cl_2 to afford **7d** (74%) as white solid. Mp 243–245 °C, $[\alpha]_D^{20} = -1.46$ (c 0.34, DMSO). ^1H NMR (DMSO-d_6, 400 MHz): δ 10.25 (1H, s), 8.71–8.69 (2H, m), 8.04 (1H, d, $J = 8.0$ Hz), 7.98 (1H, d, $J = 8.0$ Hz), 7.96 (1H, s), 7.90 (2H, s), 7.82 (2H, d, $J = 8.6$ Hz), 7.66 (2H, d, $J = 8.6$ Hz), 7.58 (1H, t, $J = 8.0$ Hz), 7.44 (1H, t, $J = 8.0$ Hz), 7.29–7.13 (5H, m), 4.75–4.70 (1H, m), 3.74 (1H, d, $J = 15.1$ Hz), 3.70 (1H, d, $J = 15.1$ Hz), 3.10 (1H, dd, $J = 13.8$, 4.8 Hz), 2.89 (1H, dd, $J = 13.8$, 9.7 Hz). ^{13}C NMR (DMSO-d_6, 100.6 MHz): δ 172.13, 169.45, 146.82, 145.91, 141.12, 139.44, 138.87, 133.21, 131.98 (2C, q, $J = 33.2$ Hz) 131.80, 131.01 (2C), 130.51, 130.23 (2C), 129.03 (2C), 127.79, 127.41, 126.23 (2C), 126.18, 126.10, 123.42 (2C, q, $J = 271.0$ Hz), 121.79, 120.84 (2C), 114.37, 109.56, 56.04, 42.60, 38.65. HR EIMS calcd for $C_{34}H_{24}F_6N_4O_2S$ *m/z* $[M + H]^+$ 666.1524. Found 666.1528.

4.2.39. (S)-N-[4-(Imidazo[2,1-b]benzothiazol-2-yl)phenyl]-2-[2-(3,5-difluorophenyl)acetamido]-3-phenylpropanamide (**7e**)

From **3c** and **4e** using procedure B. Flash chromatography (Hex:AcOEt 1:1) and crystallization with isopropanol to afford **7e** (62%) as white solid. Mp 242–243 °C, $[\alpha]_D^{20} = -4.34$ (c 0.27, CH_3OH:$CHCl_3$ = 1:1). ^1H NMR (DMSO-d_6, 400 MHz): δ 10.29 (1H, s), 8.73 (1H, s), 8.62 (1H, d, $J = 8.3$ Hz), 8.05 (1H, d, $J = 7.9$ Hz), 7.99 (1H, d, $J = 7.9$ Hz), 7.83 (2H, d, $J = 8.6$ Hz), 7.67 (2H, d, $J = 8.6$ Hz), 7.59 (1H, t, $J = 8.0$ Hz), 7.45 (1H, t, $J = 8.0$ Hz), 7.32–7.20 (5H, m), 7.05 (1H, tt, $J = 9.5$, 2.3 Hz), 6.86 (2H, dd, $J = 8.4$, 2.3 Hz), 4.75–4.70 (1H, m), 3.53 (1H, d, $J = 14.1$ Hz), 3.49 (1H, d, $J = 14.1$ Hz), 3.11 (1H, dd, $J = 13.7$, 4.8 Hz) 2.90 (1H, dd, $J = 13.7$, 9.8 Hz). ^{13}C NMR (DMSO-d_6, 100.6 MHz): δ 170.56, 169.51, 162.53 (2C, dd, $J = 244.5$, 14.1 Hz), 147.29, 146.32, 141.07 (t, $J = 10.1$ Hz), 138.46, 137.94, 132.27, 129.67 (2C), 129.40, 129.32, 128.48 (2C), 127.21, 126.87, 125.61 (2C), 125.57, 125.52, 120.10 (2C), 113.80, 112.62 (2C, d, $J = 24.1$ Hz), 109.01, 102.28 (t, $J = 26.2$ Hz), 55.39, 41.90, 38.31. HR EIMS calcd for $C_{32}H_{24}F_2N_4O_2S$ *m/z* $[M + H]^+$ 566.1588. Found 566.1584.

4.2.40. N-[4-(Imidazo[2,1-b]benzothiazol-2-yl)phenyl]-2-[2-(3,5-dimethylphenyl)acetamido] acetamide (**7f**)

From **3c** and **4f** using procedure B. Flash chromatography ($MeOH/CH_2Cl_2$ 1:30) and crystallization with CH_3CN to afford **7f** (72%) as white solid. Mp 250–252 °C ^1H NMR (DMSO-d_6, 400 MHz): δ 10.04 (1H, s), 8.71 (1H, s), 8.37–8.33 (1H, m), 8.05 (1H, d, $J = 7.9$ Hz), 7.98 (1H, d, $J = 7.9$ Hz), 7.82 (2H, d, $J = 8.6$ Hz), 7.66 (2H, d, $J = 8.6$ Hz), 7.59 (1H, t, $J = 7.9$ Hz), 7.44 (1H, t, $J = 7.9$ Hz), 6.92 (2H, s), 6.87 (1H, s), 3.92 (2H, d, $J = 5.6$ Hz), 3.34 (2H, s), 2.26 (6H, s). ^{13}C NMR (DMSO-d_6, 100.6 MHz): δ 171.83, 168.80, 147.93, 147.42, 139.14, 138.16 (2C), 137.09, 132.99, 130.29, 130.21, 128.88, 128.02 (2C), 127.78, 126.29 (2C), 126.17, 126.08, 120.53 (2C), 114.36, 109.48, 44.10, 43.19, 21.99 (2C). EIMS 468 $[M]^+$.

4.2.41. (S)-N-[4-(Imidazo[2,1-b]benzothiazol-2-yl)phenyl]-2-[2-(3,5-dimethylphenyl)acetamido]-3-(4-hydroxyphenyl)propanamide (**7g**)

From **3c** and **4g** using procedure B. Flash chromatography (*i*-$PrOH/CH_2Cl_2$ 95:5) to afford **7g** (52%) as pale yellow solid. Mp 265–268 °C, $[\alpha]_D^{20} = +2.76$ (c 0.26, DMSO). ^1H NMR (DMSO-d_6, 400 MHz): δ 10.25 (1H, s), 8.73 (1H, s), 8.39 (1H, d, $J = 7.8$ Hz), 8.08 (1H, d, $J = 8.1$ Hz), 7.98 (1H, d, $J = 8.1$ Hz), 7.81 (2H, d, $J = 8.6$ Hz), 7.67 (2H, d, $J = 8.6$ Hz), 7.58 (1H, t, $J = 8.1$ Hz), 7.43 (1H, t, $J = 8.1$ Hz), 7.11 (2H, d, $J = 7.5$ Hz), 6.84 (1H, s), 6.78 (2H, s), 6.67 (2H, d, $J = 7.5$ Hz),5.89 (1H, s), 4.65–4.54 (1H, m), 3.48–3.30 (2H, m), 2.98 (1H, dd, $J = 14.2$, 4.5 Hz), 2.81 (1H, dd, $J = 14.2$, 9.3 Hz), 2.53 (6H, s). ^{13}C NMR (DMSO-d_6, 100.6 MHz): δ 171.64, 169.12, 154.75, 147.78, 146.05, 138.85, 138.06, 134.65 (2C), 132.84 (2C), 132.24, 130.77, 130.61, 130.19, 129.12, 127.13 (2C), 116.78 (2C), 127.02, 126.08 (2C), 125.34, 124.63, 120.53 (2C), 112.89, 107.41, 54.11, 44.01, 35.21, 21.81 (2C). EIMS *m/z* $[M]^+$ 574.

4.2.42. (S)-N-[4-(Imidazo[2,1-b]benzothiazol-2-yl)phenyl]-2-[2-(4-methylphenyl)acetamido)-3-phenylpropanamide (**7h**)

From **3c** and **4l** using procedure B. Flash chromatography (CH_2Cl_2/MeOH 5%) to afford **7h** (57%). ^1H NMR ($CDCl_3$ with drops of CD_3OD-d_4, 500 MHz) δ 7.89 (s, 1H-Ar), 7.68 (d, $J = 8.0$ Hz, 2H-Ar), 7.65 (d, $J = 8.0$ Hz, 1H-Ar), 7.57 (d, $J = 8.0$ Hz, 1H-Ar), 7.39–7.43 (m, 3H-Ar), 7.29 (dd, $J = 7.5$, 7.5 Hz, 1H-Ar), 7.16–7.17 (m, 3H-Ar), 7.04–7.06 (m, 4H-Ar), 6.98 (d, $J = 7.5$ Hz, 2H-Ar), 4.70 (dd, $J = 7.5$, 7.5 Hz, 1H, $CHCH_2Ph$), 3.43 (s, 2H, CH_2CO), 3.05 (dd, $J = 14.0$, 7.5 Hz, 1H, $CHCH_2Ph$), 2.92 (dd, $J = 14.0$, 7.5 Hz, 1H, $CHCH_2Ph$), 2.27 (s, 3H, CH_3). ^{13}C NMR ($CDCl_3$ with drops of CD_3OD-d_4, 125.7 MHz) δ 172.0, 169.4, 148.0, 146.7, 136.9, 136.7, 135.9, 131.9, 131.0, 129.8, 129.5, 129.4, 129.1, 128.9, 128.3, 126.7, 126.3, 125.5, 124.9, 124.2, 120.2, 112.6, 106.7, 54.7, 42.6, 38.1, 20.8. EIMS *m/z* $[M + H]+$ 545.2.

4.2.43. (S)-N-[4-(Imidazo[2,1-b]benzothiazol-2-yl)phenyl]-2-[2-(3,5-dimethylphenyl)acetamido]-3-(4-methylphenyl)propanamide (**7i**)

HATU (258 mg, 0.68 mmol) and DIPEA (197 μl, 1.13 mmol) were added to a solution of **3c** (150 mg, 0.56 mmol) and (*S*)-*N*-Boc-4-methylphenylalanine (158 mg, 0.56 mmol) in dry THF (11 ml). After 48 h at r.t. the solvent was removed under reduced pressure and the crude mixture was purified by flash chromatography (CH_2Cl_2:MeOH 1%) to afford the intermediate that was directly used for the next step. ESIMS *m/z* $[M + H]^+$ 527.2. The obtained compound was dissolved in dry CH_2Cl_2 (5 ml). TFA (5 ml) was then added and the reaction mixture was stirred at r.t. for 4 h. Then the solvent was removed under reduced pressure and the crude mixture was purified by flash chromatography (CH_2Cl_2:EtOH 10%) to afford the corresponding free amine (227 mg, 95%).

Please cite this article in press as: A. Furlan, et al., Identification of new aminoacid amides containing the imidazo[2,1-*b*]benzothiazol-2-ylphenyl moiety as inhibitors of tumorigenesis by oncogenic Met signaling, European Journal of Medicinal Chemistry (2011), doi:10.1016/j.ejmech.2011.10.051

^{1}H NMR (CD_3OD-d_4, 400 MHz) δ 8.31 (1H, s), 7.79 (2H, dd, J = 8.0, 8.0 Hz), 7.73 (2H, d, J = 6.8 Hz), 7.54 (2H, d, J = 6.8 Hz), 7.47 (1H, dd, J = 8.0, 8.0 Hz), 7.35 (1H, dd, J = 8.0, 8.0 Hz), 7.16 (2H, dd, J = 16.2, 7.8 Hz), 4.20 (1H, dd, J = 7.4, 7.4 Hz), 3.34 (2H, s, 2H), 3.24 (1H, dd, J = 14.0, 7.4 Hz), 3.11 (1H, dd, J = 14.0, 7.4 Hz), 2.28 (3H, s). Anal. Calcd. for $C_{25}H_{22}N_4OS$: C, 70.40; H, 5.2; N, 13.14. Found: C, 70.77; H, 5.18; N, 13.09. 3,5-Dimethylphenylacetic acid (60.6 mg, 0.37 mmol) and DMAP (21.5 mg, 0.17 mmol) were added to a suspension of the previously obtained free amine (150 mg, 0.35 mmol) in dry CH_2Cl_2 (4 ml). Once the product was solved, DCI (83 μl, 0.53 mmol) was added. After 4.5 h at r.t. the solvent was removed under reduced pressure and the crude mixture was purified by flash chromatography (CH_2Cl_2:EtOH 2%) to afford **7i** contaminated with DCI. The product was finally purified by crystallization in $CHCl_3$ (65 mg, 32%). ^{1}H NMR ($CDCl_3$, 400 MHz) δ 8.36 (1H, s), 7.87 (1H, s), 7.74–7.77 (2H, m), 7.63 (2H, d, J = 7.6 Hz), 7.56 (1H, d, J = 7.6 Hz), 7.41–7.45 (3H, m), 7.32 (1H, dd, J = 7.6, 7.6 Hz), 7.05 (2H, d, J = 7.6 Hz), 6.98 (2H, d, J = 7.6 Hz), 6.92 (1H, s), 6.75 (2H, s), 6.14 (1H, d, J = 7.5 Hz), 4.77 (1H, dd, J = 14.4, 7.6 Hz), 3.49 (2H, s), 3.02–3.08 (2H, m), 2.32 (3H, s), 2.27 (6H, s). ^{13}C NMR ($CDCl_3$, 100.4 MHz) δ 172.1, 169.0, 148.0, 147.2, 138.7, 136.9, 136.5 (2C), 133.8, 133.1, 132.1 (2C), 130.2, 130.0, 129.4, 129.2, 129.1 (2C), 127.1, 126.2 (2C), 125.7 (2C), 124.8, 124.3, 120.2 (2C), 112.6, 106.5, 55.3, 43.4, 36.7, 29.7, 21.2, 21.1. ESIMS m/z $[M + H]^+$ 573.2.

4.2.44. (S)-N-[4-(Imidazo[2,1-b]benzothiazol-2-yl)phenyl]-2-[2-(3,5-bis(trifluoromethyl)phenyl)acetamido]-3-(4-methylphenyl) propanamide (**7j**)

3,5-Bis(trifluoromethyl)phenylacetic acid (111 mg, 0.40 mmol) and DMAP (24 mg, 0.19 mmol) were added to a suspension of the free amine used for the preparation of compound **7i** (164 mg, 0.38 mmol) in dry CH_2Cl_2 (4 ml). Once the product was solved, DCI (90 μl, 0.57 mmol) was added. After 4.5 h at r.t. the solvent was removed under reduced pressure and the crude mixture was purified by flash chromatography (CH_2Cl_2:EtOH 2%) to afford **7j** contaminated with DCI. The product was finally purified by crystallization in $CHCl_3$ (115 mg, 44%). ^{1}H NMR ($CDCl_3$, 400 MHz) δ 7.92 (1H, s), 7.73–7.75 (3H, m), 7.68–7.70 (3H, m), 7.61 (1H, d, J = 7.2 Hz), 7.48 (1H, s), 7.46 (1H, s), 7.42 (1H, d, J = 8.2 Hz), 7.33 (1H, dd, J = 8.2, 8.2 Hz), 6.99 (4H, s), 4.68 (1H, dd, J = 7.2, 7.2 Hz), 3.58 (2H, s), 3.06 (1H, dd, J = 14.0, 7.5 Hz), 2.94 (1H, dd, J = 14.0, 7.2 Hz), 2.23 (3H, s). ^{13}C NMR ($CDCl_3$ with drops of CD_3OD-d_4, 125.5 MHz) δ 169.7, 166.0, 147.7, 145.3, 137.4 (2C), 137.2 (2C), 136.5 (2C), 132.7, 131.7 (2C), 131.4, 129.9, 129.3, 129.4, 129.1 (2C), 128.9, 126.6 (2C), 125.7 (2C), 125.5, 124.4, 120.9, 120.4 (2C), 113.0, 107.0, 55.0, 42.0, 37.9, 20.6. ESIMS m/z $[M + H]^+$ 681.2.

4.2.45. (S)-N-[4-(7-Bromoimidazo[2,1-b]benzothiazol-2-yl) phenyl]-2-[2-(3,5-bis(trifluoromethyl) phenyl)acetamido]-3-phenylpropanamide (**7k**)

From **3e** and **4c** using procedure B. Flash chromatography (Hex:AcOEt = 1:1) to afford **7k** (48%) as a white solid. $[\alpha]_D^{20}$ = 2.95 (Na, c = 0.28 in DMSO). ^{1}H NMR (DMSO-d_6, 400 MHz): δ 10.26 (1H, s), 8.71 (1H, d, J = 8.7 Hz), 8.70 (1H, s), 8.34 (1H, d, J = 1.8 Hz), 7.95 (1H, s), 7.93 (1H, d, J = 8.5 Hz), 7.89 (2H, s), 7.81 (2H, d, J = 8.7 Hz), 7.75 (1H, dd, J = 8.5, 1.8 Hz), 7.66 (2H, d, J = 8.7 Hz), 7.27–7.13 (5H, m), 4.75–4.69 (1H, m), 3.75 (1H, d, J = 16.4 Hz), 3.70 (1H, d, J = 16.4 Hz), 3.10 (1H, dd, J = 13.8, 4.8 Hz), 2.89 (1H, dd, J = 13.8, 9.8 Hz). ^{13}C NMR (DMSO-d_6, 100.6 MHz): δ 170.48, 169.49, 147.42, 146.85, 140.10, 138.44, 137.86, 131.82, 131.59, 130.58 (2C, q, J = 37 Hz), 130.37 (2C), 129.99, 129.59, 129.50, 128.40 (2C), 127.86, 126.80 (2C), 125.60 (2C), 123.85 (2C, q, J = 272 Hz), 120.68, 120.11, 117.12, 115.29, 109.14, 55.42, 41.48, 38.30. HR-EIMS Calcd for $[C_{34}H_{23}BrF_6N_4O_2S]^+$ 744.0629; Found 744.06134.

4.2.46. (S)-N-[4-(6-Benzyl-5,6,7,8-tetrahydroimidazo[2′,1′:2,3] thiazolo[4,5-c]pyridin-2-yl)phenyl]-3-[2-(3,5-difluorophenyl) acetamido]propanamide (**8a**)

From **3d** and **4h** using procedure B. Flash chromatography (CH_2Cl_2/MeOH = 99/1) to give **8a** (65%). ^{1}H NMR ($CDCl_3$ with drops of CD_3OD, 400 MHz) δ 9.40 (1H, bs), 7.67 (1H, dt(ddd), J = 8.8, 1.6, 1.6 Hz), 7.52 (1H, dt(ddd), J = 8.8, 1.6, 1.6 Hz), 7.47 (1H, t(dd), J = 5.6 Hz), 7.43 (1H, s), 7.30–7.41 (6H, m), 6.80 (2H, tt(dddd), J = 6.4, 2.4 Hz), 6.67 (1H, tt(dddd), J = 11.2, 2.4 Hz), 3.81 (2H, s), 3.59 (2H, s), 3.51 (2H, dt, J = 12.4; 6.4 Hz), 3.47 (2H, s), 2.96 (2H, t(dd), J = 5.2 Hz), 2.83 (2H, t(dd), J = 5.2 Hz), 2.55 (2H, t(dd), J = 6.0 Hz). ^{13}C NMR ($CDCl_3$ with drops of CD_3OD-d_4, 100.6 MHz) δ 170.84 (d, J = 90.5 Hz), 170.27 (d, J = 93.6 Hz), 164.05 (d, J_{C-F} = 123.7 Hz), 161.57 (d, J_{C-F} = 132.8 Hz), 145.92, 148.68, 138.47 (t(dd), J_{C-F} = 93.6 Hz), 137.12 (t(dd), J_{C-F} = 92.6 Hz), 137.11, 136.98, 129.69, 128.89, 128.44, 127.52, 125.36, 124.11, 119.95, 119.89, 112.01(d, J_{C-F} = 69.4 Hz), 111.83 (d, J_{C-F} = 70.4 Hz), 105.28, 102.35 (t(dd), J_{C-F} = 253.0 Hz), 61.55, 50.04, 48.94, 42.61, 36.06, 35.92, 24.47. HRMS calculated for $C_{32}H_{30}F_2N_5O_2S$ $[M + H]^+$: 586.2088; Found: 586.2075.

4.2.47. (S)-N-[4-(6-Benzyl-5,6,7,8-tetrahydroimidazo[2′,1′:2,3] thiazolo[4,5-c]pyridin-2-yl)phenyl]-2-[2-(3,5-difluorophenyl) acetamido]propanamide (**8b**)

From **3d** and **4j** using procedure B. Flash chromatography (CH_2Cl_2/MeOH = 99/1) to give **8b** (48%). $[\alpha]_D^{20}$ = +3.8 (c 0.17, MeOH); ^{1}H NMR ($CDCl_3$ with drops of CD_3OD, 400 MHz) δ 9.54 (brs, 1H, NH), 7.67 (dt(ddd), J = 8.8, 1.6, 1.6 Hz, 1H, CH_{Ar}), 7.63 (d, J = 7.6 Hz, 1H, NH), 7.52 (dt(ddd), J = 8.8, 1.6, 1.6 Hz, 1H, CH_{Ar}), 7.32–7.42 (m, 6H, 5CH_{Ar}, H-3), 6.84 (tt(dddd), J = 6.4, 2.4 Hz, 2H, CH_{Ar}), 6.70 (tt(dddd), J = 11.2, 2.4 Hz, 1H, CH_{Ar}), 4.57 (qd, J = 7.2 Hz, 1H, CH), 3.81 (s, 2H, CH_2Ph), 3.58 (s, 2H, H-8), 3.54 (s, 2H, CH_2), 2.95 (t(dd), J = 5.6 Hz, 2H, H-6), 2.82 (t(dd), J = 5.6 Hz, 2H, H-5), 1.41 (d, J = 6.8 Hz, 3H, CH_3). ^{13}C NMR ($CDCl_3$ with drops of CD_3OD-d_4, 100.6 MHz) δ 170.80 (d, J = 101.6 Hz), 170.24 (d, J = 70.4 Hz), 164.05 (d, J_{C-F} = 132.8 Hz), 161.58 (d, J_{C-F} = 124.7 Hz), 145.84, 148.67, 138.30 (t(dd), J_{C-F} = 93.1 Hz), 136.97, 136.90 (t(dd), J_{C-F} = 93.1 Hz), 129.90, 128.87, 128.42, 127.51, 125.38, 124.09, 120.14, 120.05, 119.89, 112.07 (d, J_{C-F} = 69.4 Hz), 111.89 (d, J_{C-F} = 69.4 Hz), 105.32, 102.40 (t(dd), J_{C-F} = 253 Hz), 61.53, 50.02, 49.46, 48.93, 42.25, 24.45, 17.95. HRMS calcd. for $C_{32}H_{30}F_2N_5O_2S$ $[M + H]^+$: 586.2088; Found: 586.2075.

4.2.48. (R)-N-[4-(6-Benzyl-5,6,7,8-tetrahydroimidazo[2′,1′:2,3] thiazolo[4,5-c]pyridin-2-yl)phenyl]-2-[2-(3,5-difluorophenyl) acetamido]propanamide (**8c**)

From **3d** and **4k** using procedure B. Flash chromatography (CH_2Cl_2/MeOH = 99/1) to give **8c** (48%). $[\alpha]_D^{20}$ = −3.6 (c 0.23, MeOH); ^{1}H NMR and ^{13}C NMR see **8b**. HRMS calcd. for $C_{32}H_{30}F_2N_5O_2S$ $[M + H]^+$: 586.2088; Found: 586.2077.

4.3. Biology

4.3.1. Cell culture

MDCK cells were grown in Dulbecco's modified Eagle's medium (DMEM) (Gibco BRL) containing 4 mM L-glutamine and supplemented with 2% (v/v) fetal bovine serum (Gibco BRL), 100 U/ml penicillin, and 100 μg/ml streptomycin. Cells were plated at a density of 1000 cells per well in 24-well microplates and allowed settling overnight at 37 °C in a humidified atmosphere of 5% CO_2 prior to treatments. Human MCF10A breast cells, human MDA-MB231 breast cancer cells, human GTL-16 gastric carcinoma cells, human HepG2 hepatocellular carcinoma, and human BT474 breast cancer cells were grown in RPMI medium (Gibco BRL) containing 4 mM L-glutamine and supplemented with 10% (v/v) fetal bovine

Please cite this article in press as: A. Furlan, et al., Identification of new aminoacid amides containing the imidazo[2,1-*b*]benzothiazol-2-ylphenyl moiety as inhibitors of tumorigenesis by oncogenic Met signaling, European Journal of Medicinal Chemistry (2011), doi:10.1016/j.ejmech.2011.10.051

serum (Gibco BRL), 100 U/ml penicillin, and 100 μg/ml streptomycin and kept at 37 °C in a humidified atmosphere of 5% CO_2.

4.3.2. Primary cultures of embryonic cortical neurons and hepatocyte

Primary embryonic cortical and hepatocyte cultures and survival assays were performed as previously described [11,12,49]. For toxicity assays, primary cells were exposed to compound **7c** for 24 h. Viability was assessed with the Cell Titer Glo Luminescent Assay (Promega), according to the manufacturer's instructions. Data on biological assays were performed in quadruplicate.

4.3.3. Compound treatments

For scattering assays, MDCK cells were pre-incubated with compounds overnight at 0.1–100 μM concentrations at 37 °C in a humidified atmosphere of 5% CO_2, followed by a 24 h stimulation with 20 ng/ml HGF (R&D Systems), as previously described [32]. Cells were further incubated at 37 °C in an atmosphere of 5% CO_2 for 24–48 h, washed with phosphate-buffered saline (PBS; Gibco BRL), and fixed with 4% PFA (Sigma). The quantification of scattering response was performed by counting the number of cells with scattered morphology in 30 independent colonies. The IC_{50} corresponds to the concentration of compounds leading to a 50% inhibition of Met-triggered cell scattering. Active compounds were assayed to impair scattering response also on MCF10A cells following the same procedure. For survival assays, cells were cultured in serum-free media for 24 h prior to compound addition for 48 h. Viability was assessed with the Cell Titer Glo Luminescent Assay (Promega). For in vitro tumorigenesis, soft agar growth assays were performed using GTL-16 and HepG2 cells, as previously described [32]. Data on biological assays are representative of three independent experiments performed in duplicate. For the Ba/F3 cell assay, a total of 2.10^4 cells/well/50 μl were seeded in a 96-wells plate. Treatment was initiated by addition of a 2× drug solution of ½ serial dilutions ranging from 0 to 10 μM. Cells were grown for 48 h at 37 °C and then incubated with 10 μl/well of Cell Titer Blue (G8080; Promega). Proliferation was quantified by measuring the absorbance at 490 nm using a scanning multiwell spectrophotometer (MultiSkan MS, Thermo-LabSystems, France). A blank well without cells was used as a background control for the spectrophotometer and all assays were performed in triplicate.

4.3.4. Biochemical evaluation of Met inhibitors

After starvation, cells were pretreated overnight with inhibitors, then stimulated with HGF (20 ng/ml), and total extracts were analyzed as described [10]. Antibodies used were anti-tubulin (Sigma), anti-phospho Tyr1234/1235-Met, anti phospho-ErbBs, anti-ERKs (Cell Signaling), anti-ErbBs, anti-Met (Santa Cruz).

4.3.5. Protein kinase assay

The inhibitory profile of 7c toward Met RTK was performed by applying the Kinexus compound profiling service.

4.3.6. In vivo tumorigenesis assays

Xenografts of GTL-16 cells were established by intraperitoneal (i.p.) injection of cells (10^6) in nude mice (S/SOPF SWISS NU/NU; Charles River). Mice were treated with **7c** (i.p. 10 or 30 mg kg^{-1}) or vehicle at day 1 and treatment was repeated every other day. Mice were then sacrificed after 21 days of treatment. Tumor nodules present in the peritoneal cavity were isolated and quantified according to their total weight. For **7c** in vivo treatment, the drug was formulated in Cremophor EL-DMSO (1:1, v/v) and diluted in sterile 0.9% (w/v) sodium chloride for administration at dose levels of 10 and 30 mg kg^{-1}. Two independent tumorigenesis assays were performed (8 mice per group were used). Mice were kept at the IBDML animal facilities. All experimental procedures were performed according to the European Union and institutional guidelines for the care and use of laboratory animals, and were approved by an official committee.

4.3.7. Statistical analysis

Results were expressed as the mean ± s.e.m. Quantification of biological assays and tumor growth in mice was analyzed by the student-*t* test. Statistical significance was defined as ns: $P > 0.05$; *: $P < 0.05$; **: $P < 0.01$; ***: $P < 0.001$. Statistic analysis was carried out using the Analyze-it software for Microsoft Excel (Analyse-it Software, Ltd).

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

We are particularly grateful to: Patrick Faure, Romain Rauly, Keith Dudley, Sandro De Falco, Fabienne Lamballe, Filippo Conti, and all members of our labs for helpful discussions and comments; Annalisa Fico for her help with in vivo studies and valuable inputs. We thank: Sylvie Richelme for technical contribution; staff at the IBDML animal house facility and technical support; the imaging platform at the IBDML. This work was supported by funds from the INCa (Insitut National du Cancer), ARC (Association pour la recherche contre le cancer), FRM (Fondation pour la recherche médicale), FdF (Fondation de France), Fondation Bettencourt-Schueller, Marie Curie Host Fellowship for the Transfer of Knowledge (MTKD-CT-2004-509804), Protisvalor, Valorpaca, and Oséo to F.M; by the Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) PRIN 2007 – Program "Sviluppo e caratterizzazione di nuovi inibitori di tirosine chinasi cellulari con attività antiproliferativa e antiangiogenica nei confronti di differenti tumori" to D.P.; and by the Spanish Ministry of Science and Innovation (Project CTQ2009-07021/BQU) and the AGAUR, Generalitat de Catalunya (Grant 2009-SGR-111) to M.A. and J.B. This research has been developed under the umbrella of CM0602 COST Action "Inhibitors of Angiogenesis: design, synthesis and biological exploitation". A.F. was supported by the INCa and ARC; N.I. by Valorpaca.

Appendix. Supplementary information

Supplementary data related to this article can be found online at doi:10.1016/j.ejmech.2011.10.051.

References

[1] E. Zwick, J. Bange, A. Ullrich, Receptor tyrosine kinases as targets for anticancer drugs, Trends Mol. Med. 8 (2002) 17–23.
[2] M.A. Lemmon, J. Schlessinger, Cell signaling by receptor tyrosine kinases, Cell 141 (2010) 1117–1134.
[3] T. Hunter, Tyrosine phosphorylation: thirty years and counting, Curr. Opin. Cell Biol. 21 (2009) 140–146.
[4] B.S. Knudsen, G. Vande Woude, Showering c-MET-dependent cancers with drugs, Curr. Opin. Genet. Dev. 18 (2008) 87–96.
[5] P. Kaposi-Novak, J.S. Lee, L. Gomez-Quiroz, C. Coulouarn, V.M. Factor, S.S. Thorgeirsson, Met-regulated expression signature defines a subset of human hepatocellular carcinomas with poor prognosis and aggressive phenotype, J. Clin. Invest. 116 (2006) 1582–1595.
[6] L. Trusolino, A. Bertotti, P.M. Comoglio, MET signalling: principles and functions in development, organ regeneration and cancer, Nat. Rev. Mol. Cell Biol. 11 (2010) 834–848.
[7] F. Maina, F. Casagranda, E. Audero, A. Simeone, P. Comoglio, R. Klein, C. Ponzetto, Uncoupling of Grb2 from the Met receptor in vivo reveals complex roles in muscle development, Cell 87 (1996) 531–542.
[8] F. Maina, M.C. Hilton, R. Andres, S. Wyatt, R. Klein, A.M. Davies, Multiple roles for hepatocyte growth factor in sympathetic neuron development, Neuron 20 (1998) 835–846.

[9] F. Maina, M.C. Hilton, C. Ponzetto, A.M. Davies, R. Klein, Met receptor signaling is required for sensory nerve development and HGF promotes axonal growth and survival of sensory neurons, Genes Dev. 11 (1997) 3341–3350.
[10] F. Maina, G. Pante, F. Helmbacher, R. Andres, A. Porthin, A.M. Davies, C. Ponzetto, R. Klein, Coupling Met to specific pathways results in distinct developmental outcomes, Mol. Cell 7 (2001) 1293–1306.
[11] A. Moumen, A. Ieraci, S. Patane, C. Sole, J.X. Comella, R. Dono, F. Maina, Met signals hepatocyte survival by preventing Fas-triggered FLIP degradation in a PI3k-Akt-dependent manner, Hepatology 45 (2007) 1210–1217.
[12] A. Moumen, S. Patane, A. Porras, R. Dono, F. Maina, Met acts on Mdm2 via mTOR to signal cell survival during development, Development 134 (2007) 1443–1451.
[13] M. Genestine, E. Caricati, A. Fico, S. Richelme, H. Hassani, C. Sunyach, F. Lamballe, G.C. Panzica, B. Pettmann, F. Helmbacher, C. Raoul, F. Maina, R. Dono, Enhanced neuronal Met signalling levels in ALS mice delay disease onset, Cell Death Dis. 2 (2011) e130.
[14] L. Tonges, T. Ostendorf, F. Lamballe, M. Genestine, R. Dono, J.C. Koch, M. Bahr, F. Maina, P. Lingor, Hepatocyte growth factor protects retinal ganglion cells by increasing neuronal survival and axonal regeneration in vitro and in vivo, J. Neurochem. 117 (2011) 892–903.
[15] A. Furlan, V. Stagni, A. Hussain, S. Richelme, F. Conti, A. Prodosmo, A. Destro, M. Roncalli, D. Barila, F. Maina, Abl interconnects oncogenic Met and p53 core pathways in cancer cells, Cell Death Differ. 18 (2011) 1608–1616.
[16] C.L. Arteaga, HER3 and mutant EGFR meet MET, Nat. Med. 13 (2007) 675–677.
[17] J.M. Stommel, A.C. Kimmelman, H. Ying, R. Nabioullin, A.H. Ponugoti, R. Wiedemeyer, A.H. Stegh, J.E. Bradner, K.L. Ligon, C. Brennan, L. Chin, R.A. DePinho, Coactivation of receptor tyrosine kinases affects the response of tumor cells to targeted therapies, Science 318 (2007) 287–290.
[18] J.A. Engelman, K. Zejnullahu, T. Mitsudomi, Y. Song, C. Hyland, J.O. Park, N. Lindeman, C.M. Gale, X. Zhao, J. Christensen, T. Kosaka, A.J. Holmes, A.M. Rogers, F. Cappuzzo, T. Mok, C. Lee, B.E. Johnson, L.C. Cantley, P.A. Janne, MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling, Science 316 (2007) 1039–1043.
[19] D.L. Shattuck, J.K. Miller, K.L. Carraway 3rd, C. Sweeney, Met receptor contributes to trastuzumab resistance of Her2-overexpressing breast cancer cells, Cancer Res. 68 (2008) 1471–1477.
[20] P.M. Comoglio, S. Giordano, L. Trusolino, Drug development of MET inhibitors: targeting oncogene addiction and expedience, Nat. Rev. Drug Discov. 7 (2008) 504–516.
[21] T.L. Underiner, T. Herbertz, S.J. Miknyoczki, Discovery of small molecule c-Met inhibitors: evolution and profiles of clinical candidates, Anticancer Agents Med. Chem. 10 (2010) 7–27.
[22] J.J. Cui, Inhibitors targeting hepatocyte growth factor receptor and their potential therapeutic applications, Expert Opin. Ther. Patents 17 (2007) 1035–1045.
[23] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, Nucleic Acids Res. 28 (2000) 235–242.
[24] R. Capdeville, E. Buchdunger, J. Zimmermann, A. Matter, Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug, Nat. Rev. Drug Discov. 1 (2002) 493–502.
[25] T. Schindler, W. Bornmann, P. Pellicena, W.T. Miller, B. Clarkson, J. Kuriyan, Structural mechanism for STI-571 inhibition of abelson tyrosine kinase, Science 289 (2000) 1938–1942.
[26] Y. Liu, N.S. Gray, Rational design of inhibitors that bind to inactive kinase conformations, Nat. Chem. Biol. 2 (2006) 358–364.
[27] M.D. Jacobs, P.R. Caron, B.J. Hare, Classifying protein kinase structures guides use of ligand-selectivity profiles to predict inactive conformations: structure of lck/imatinib complex, Proteins 70 (2008) 1451–1460.
[28] Y. Asses, V. Leroux, S. Tairi-Kellou, R. Dono, F. Maina, B. Maigret, Analysis of c-Met kinase domain complexes: a new specific catalytic site receptor model for defining binding modes of ATP-competitive ligands, Chem. Biol. Drug Des 74 (2009) 560–570.
[29] I. Dussault, S.F. Bellon, c-Met inhibitors with different binding modes: two is better than one, Cell Cycle 7 (2008) 1157–1160.
[30] N. Schiering, S. Knapp, M. Marconi, M.M. Flocco, J. Cui, R. Perego, L. Rusconi, C. Cristiani, Crystal structure of the tyrosine kinase domain of the hepatocyte growth factor receptor c-Met and its complex with the microbial alkaloid K-252a, Proc. Natl. Acad. Sci. U S A 100 (2003) 12654–12659.
[31] S.F. Bellon, P. Kaplan-Lefko, Y. Yang, Y. Zhang, J. Moriguchi, K. Rex, C.W. Johnson, P.E. Rose, A.M. Long, A.B. O'Connor, Y. Gu, A. Coxon, T.S. Kim, A. Tasker, T.L. Burgess, I. Dussault, c-Met inhibitors with novel binding mode show activity against several hereditary papillary renal cell carcinoma-related mutations, J. Biol. Chem. 283 (2008) 2675–2683.
[32] S. Patane, N. Pietrancosta, H. Hassani, V. Leroux, B. Maigret, J.L. Kraus, R. Dono, F. Maina, A new Met inhibitory-scaffold identified by a focused forward chemical biological screen, Biochem. Biophys. Res. Commun. 375 (2008) 184–189.
[33] Z.W. Cai, D. Wei, G.M. Schroeder, L.A. Cornelius, K. Kim, X.T. Chen, R.J. Schmidt, D.K. Williams, J.S. Tokarski, Y. An, J.S. Sack, V. Manne, A. Kamath, Y. Zhang, P. Marathe, J.T. Hunt, L.J. Lombardo, J. Fargnoli, R.M. Borzilleri, Discovery of orally active pyrrolopyridine- and aminopyridine-based Met kinase inhibitors, Bioorg. Med. Chem. Lett. 18 (2008) 3224–3229.
[34] N.D. D'Angelo, S.F. Bellon, S.K. Booker, Y. Cheng, A. Coxon, C. Dominguez, I. Fellows, D. Hoffman, R. Hungate, P. Kaplan-Lefko, M.R. Lee, C. Li, L. Liu, E. Rainbeau, P.J. Reider, K. Rex, A. Siegmund, Y. Sun, A.S. Tasker, N. Xi, S. Xu, Y. Yang, Y. Zhang, T.L. Burgess, I. Dussault, T.S. Kim, Design, synthesis, and biological evaluation of potent c-Met inhibitors, J. Med. Chem. 51 (2008) 5766–5779.
[35] M. Amat, A. Kövér, D. Jokic, O. Lozano, M. Pérez, N. Landoni, F. Subrizi, J. Bautista, J. Bosch, Synthesis of a tetrahydroimidazo[2′,1′:2,3]thiazolo[5,4-c] pyridine derivative with Met inhibitory activity, ARKIVOC 3 (2010) 145–151.
[36] G.K. Chan, B.A. Lutterbach, B.S. Pan, I. Kariv, A.A. Szewczak, High-throughput analysis of HGF-stimulated cell scattering, J. Biomol. Screen. 13 (2008) 847–854.
[37] M. Sattler, Y.B. Pride, P. Ma, J.L. Gramlich, S.C. Chu, L.A. Quinnan, S. Shirazian, C. Liang, K. Podar, J.G. Christensen, R. Salgia, A novel small molecule met inhibitor induces apoptosis in cells transformed by the oncogenic TPR-MET tyrosine kinase, Cancer Res. 63 (2003) 5462–5469.
[38] H.Y. Zou, Q. Li, J.H. Lee, M.E. Arango, S.R. McDonnell, S. Yamazaki, T.B. Koudriakova, G. Alton, J.J. Cui, P.P. Kung, M.D. Nambu, G. Los, S.L. Bender, B. Mroczkowski, J.G. Christensen, An orally available small-molecule inhibitor of c-Met, PF-2341066, exhibits cytoreductive antitumor efficacy through anti-proliferative and antiangiogenic mechanisms, Cancer Res. 67 (2007) 4408–4417.
[39] K.M. Nicholson, C.H. Streuli, N.G. Anderson, Autocrine signalling through erbB receptors promotes constitutive activation of protein kinase B/Akt in breast cancer cell lines, Breast Cancer Res. Treat. 81 (2003) 117–128.
[40] S.Y. Wang, B. Chen, Y.Q. Zhan, W.X. Xu, C.Y. Li, R.F. Yang, H. Zheng, P.B. Yue, S.H. Larsen, H.B. Sun, X. Yang, SU5416 is a potent inhibitor of hepatocyte growth factor receptor (c-Met) and blocks HGF-induced invasiveness of human HepG2 hepatoma cells, J. Hepatol. 41 (2004) 267–273.
[41] F. Qian, S. Engst, K. Yamaguchi, P. Yu, K.A. Won, L. Mock, T. Lou, J. Tan, C. Li, D. Tam, J. Lougheed, F.M. Yakes, F. Bentzien, W. Xu, T. Zaks, R. Wooster, J. Greshock, A.H. Joly, Inhibition of tumor cell growth, invasion, and metastasis by EXEL-2880 (XL880, GSK1363089), a novel inhibitor of HGF and VEGF receptor tyrosine kinases, Cancer Res. 69 (2009) 8009–8016.
[42] N. Munshi, S. Jeay, Y. Li, C.R. Chen, D.S. France, M.A. Ashwell, J. Hill, M.M. Moussa, D.S. Leggett, C.J. Li, ARQ 197, a novel and selective inhibitor of the human c-Met receptor tyrosine kinase with antitumor activity, Mol. Cancer Ther. 9 (2010) 1544–1553.
[43] G. Jones, P. Willett, R.C. Glen, Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation, J. Mol. Biol. 245 (1995) 43–53.
[44] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, Development and validation of a genetic algorithm for flexible docking, J. Mol. Biol. 267 (1997) 727–748.
[45] M.L. Verdonk, J.C. Cole, M.J. Hartshorn, C.W. Murray, R.D. Taylor, Improved protein-ligand docking using GOLD, Proteins 52 (2003) 609–623.
[46] W.L. Jorgensen, D.S. Maxwell, J. Tirado-Rives, Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, J. Am. Chem. Soc. 118 (1996) 11225–11236.
[47] N. Pietrancosta, F. Maina, R. Dono, A. Moumen, C. Garino, Y. Laras, S. Burlet, G. Quelever, J.L. Kraus, Novel cyclized Pifithrin-alpha p53 inactivators: synthesis and biological studies, Bioorg. Med. Chem. Lett. 15 (2005) 1561–1564.
[48] Q. Chao, K.G. Sprankle, R.M. Grotzfeld, A.G. Lai, T.A. Carter, A.M. Velasco, R.N. Gunawardane, M.D. Cramer, M.F. Gardner, J. James, P.P. Zarrinkar, H.K. Patel, S.S. Bhagwat, Identification of N-(5-tert-butyl-isoxazol-3-yl)-N′-{4-[7-(2-morpholin-4-yl-ethoxy)imidazo[2,1-b][1,3]benzothiazol-2-yl]phenyl}urea dihydrochloride (AC220), a uniquely potent, selective, and efficacious FMS-like tyrosine kinase-3 (FLT3) inhibitor, J. Med. Chem. 52 (2009) 7808–7816.
[49] J. Segarra, L. Balenci, T. Drenth, F. Maina, F. Lamballe, Combined signaling through ERK, PI3K/AKT, and RAC1/p38 is required for Met-triggered cortical neuron migration, J. Biol. Chem. 281 (2006) 4771–4778.
[50] F. Lamballe, M. Genestine, N. Caruso, V. Arce, S. Richelme, F. Helmbacher, F. Maina, Pool-specific regulation of motor neuron survival by neurotrophic support, J Neurosci. 31 (2011) 11144–11158.

Please cite this article in press as: A. Furlan, et al., Identification of new aminoacid amides containing the imidazo[2,1-*b*]benzothiazol-2-ylphenyl moiety as inhibitors of tumorigenesis by oncogenic Met signaling, European Journal of Medicinal Chemistry (2011), doi:10.1016/j.ejmech.2011.10.051

III- Commentaires

Ce travail de synthèse long et délicat a demandé plusieurs années après proposition des molécules de notre part. À l'issue de tests biologiques, une molécule s'est révélée particulièrement intéressante. Cet article rapporte les principales étapes de synthèse, de modélisation et de caractérisation biologique d'une nouvelle classe de dérivés amides d'acides aminés contenant le groupement 4-(imidazo[2,1-b]benzothiazol-2-yl) phenyl. Les composés ont d'abord été évalués biologiquement pour leur capacité de nuisance à la diffusion cellulaire déclenchée par Met, *in vitro*. Cette étude biologique a montré que des membres de cette nouvelle classe de molécules nuit à la survie et à la croissance des cellules cancéreuses sur-exprimant l'oncogène c-Met.

Pour cette molécule, un criblage sur cellules a également été effectué. Il a révélé une sélectivité pour les membres de la famille de c-Met. De plus, les études *in vivo* réalisées ont prouvé l'effet anti-tumoral du composé le plus actif de la série envers les cellules cancéreuses où c-Met est impliquée. Ce composé n'a montré aucun effet secondaire ni *in vivo* ni sur les cultures primaires ou les cultures hépatocyte.

En perspective de ces travaux et afin d'exploiter pleinement les propriétés anti cancéreuses des composés synthétisés, d'autres études seront poursuivies ayant pour but d'établir l'action médicamenteuse ainsi que l'efficacité de ces molécules à l'égard de différents types de cellules cancéreuses.

ARTICLE 2

I- Contexte

Pendant que le travail de criblage se poursuivait en collaboration avec les collègues chimistes et biologistes (année universitaire 2008-2009), de très nombreuses structures cristallographiques de c-Met complexée à divers ligands sont apparues (en 2008 les structures de codes PDB : 3F66, 3EFJ, 3EFK, 3CE3, 3CTH, 3CTJ, 3CCN, 3CD8, 3C1X ; en 2009 celles de codes PDB : 3GQL, 2WGJ, 3F82, 3DKC, 3DKF, 3DKG). Cette « explosion » confirmait l'intérêt de la communauté internationale pour trouver des molécules originales ciblant c-Met et à cette époque de nombreuses revues allaient en ce sens (voir par exemple [244-248]).

Nous avons donc décidé d'entreprendre une étude exhaustive de comparaison entre toutes les structures 3D de c-Met disponibles alors. Notre but était d'identifier les éléments structuraux spécifiques de certains types de ligand et d'essayer de classer ces structures selon le mode de liaison de leurs ligands respectifs (voir aussi annexe C).

II- Analysis of c-Met Kinase Domain Complexes : A new specific Catalytic Site Receptor Model for Defining Binding Modes of ATP-Competitive Ligands

Chem Biol Drug Des 2009; **74:** 560–570

Research Article

© 2009 John Wiley & Sons A/S

doi: 10.1111/j.1747-0285.2009.00895.x

Analysis of c-Met Kinase Domain Complexes: A New Specific Catalytic Site Receptor Model for Defining Binding Modes of ATP-Competitive Ligands

Yasmine Asses[1†], Vincent Leroux[1†], Safia Tairi-Kellou[2], Rosanna Dono[3], Flavio Maina[3] and Bernard Maigret[1*]

[1]*Nancy Université, LORIA–UMR 7503, Équipe-projet Orpailleur, Campus scientifique, BP 239, 54506 Vandœuvre-lès-Nancy Cedex, France*

[2]*Laboratory of Theoretical Chemical Physics and Computer Chemistry, USTHB, Algiers, Algeria*

[3]*Developmental Biology Institute of Marseille-Luminy (IBDML), UMR 6216 CNRS–Université de la Méditerrannée, Campus de Luminy, Case 907, 13288 Marseille Cedex 09, France*

*Corresponding author: Bernard Maigret, Bernard.Maigret@loria.fr

[†]*These two authors contributed equally to this work.*

The receptor tyrosine kinase c-Met have multiple roles during cancer development and is currently considered as an important target for molecularly targeted therapies. Structural knowledge of how compounds interact on c-Met catalytic site could guide structure-based drug design strategies towards more effective and selective anticancer drug candidates. However, although 17 crystal structures of c-Met complexed with adenosine triphosphate (ATP)-competitive kinase inhibitors are publicly available (August 2009), there are still open questions regarding the prediction of ligand binding modes. We have applied molecular modeling and molecular mechanics to analyze the distribution of ligands interaction energy on c-Met residues, and deduced a new model of the active site allowing for an unambiguous identification of ligand binding modes. We demonstrate that the binding of known ligands on the c-Met catalytic site involves seven identified structurally-distinct areas. Five of these match the generic kinase ATP binding site model built by Novartis scientists in the 1990s, while the two others are distinct allosteric regions that can be exploited by second generation kinase inhibitors such as Gleevec. We show here that c-Met can accept both such kinds of allosteric inhibitors, a very unusual feature in the kinase family that opens new grounds for highly specific drug design.

Key words: allosteric interactions, ATP binding site, c-Met, kinase inhibitors, ligand binding modes, molecular modeling, protein-ligand complexes, protein receptor model, structure-based drug design

Received 19 May 2009, revised 28 August 2009 and accepted for publication 10 September 2009

c-Met, also known as HGFR (Hepatocyte Growth Factor Receptor), is a 1390 amino-acid long protein encoded by the MET (Mesenchymal-Epithelial Transition factor) proto-oncogene. c-Met is found exclusively in vertebrates and belongs to the receptor tyrosine kinase class of the protein kinase superfamily (1,2). It is a key membrane receptor that is essential for embryological development, physiological repair and defense mechanisms during life (3–6).

c-Met and its mutants have a well-established pivotal role in tumorigenesis, and excessive MET levels are directly related to the poor prognosis of a variety of cancers (1,2,7,8). Amongst the deregulated mechanisms, c-Met overexpression allows altered cells to trigger invasive growth and turn it into an oncogenic process with a wide spectrum of cancer progression-related events leading to metastasis (9–11). c-Met is notoriously involved in resistance phenomena developed towards existing therapies (12,13), and has been recently located at the core of the proteomics network responsible for cancer resistance development to kinase inhibitors administration (14). MET signaling is also of great pharmaceutical interest outside anticancer research, because of its implication in pathogenesis such as listeriosis (15) and malaria (16).

c-Met biological activation is a complex molecular process that can be subdivided into distinct mechanisms, the most well-known being: (i) dimerization, mediated by the extracellular binding of HGF to the c-Met SEMA (semaphorin homolog) domain, (ii) intracellular adenosine triphosphate (ATP) binding to the kinase domain initiating autophosphorylation, and (iii) the binding of SH2 domain-containing partners on the phosphorylated c-Met kinase to allow signal transduction inside the cells. Each of those is able to provide one or several possible targets as the basis for structure-based drug design strategies (2,17).

In this work, we focus on ATP-competitive inhibitors, targeting the corresponding intracellular binding site on c-Met kinase domain.

We used as starting points the 17 publicly available X-ray structures of c-Met kinase complexed with ATP-competitive ligands.

Numerous compounds are also reported with activity and specificity for MET but missing an experimental structure bound to the c-Met protein (18). Amongst these, there is a good number of second-generation kinase inhibitors designed to exploit specific allosteric interactions in addition to those performed by ATP on all kinases. Imatinib/Gleevec (19,20) is a living proof that this kind of compounds can turn into highly successful drugs, while Gefitinib/Iressa (12,21) reminds that specificity is still a crucial issue with those. A detailed study of ligand binding based on the available c-Met kinase complexes crystal structures is a prerequisite for future structure-based drug design efforts. We have to take into account the fact that kinase catalytic sites are known to be quite flexible so that induced fit effects (22) are prone to occur upon ligand binding. In the case of c-Met, Dussault reported that at least two distinct binding modes are available to ligands (23). We are also aware that a kinase drug design strategy exploiting target specificity might identify highly potent and specific compounds (24). In this paper, our aim is to precisely describe c-Met kinase ATP receptor features in order to facilitate ligand binding modes analysis, as to provide a solid background for future lead optimization and virtual screening approaches.

The conserved binding mode of ATP to kinases is well known since the publication of the cAMP-ATP complex crystal structure in 1991 Protein Data Bank (PDB: 1ATP) (25,26). In 1995, Furet and coworkers at Novartis proposed a simplified model of the EGFR kinase ATP binding site (27). This model was used successfully as a reference for the rational design of ATP-competitive protein tyrosine kinase inhibitors (28). In 2000, X-ray experiments demonstrated that a second class of inhibitors could exploit additional allosteric regions that are reachable in kinases inactive form (29). This feature is best described by the crystal structure of Imatinib bound to Bcl/Abl (PDB: 2HYY) (30). Liu and Gray updated the classical Novartis model in 2006 in order to characterize the general binding mode of Imatinib-like, inactive state-targeting (type II) kinase inhibitors (31).

Applying directly this model on c-Met would have led us to several issues. In the first stages of this study, there was no available experimental structure of ATP-bound c-Met. Modeling the complex in order to fill this void was not attempted for several reasons. First, predicting induced fit effects upon docking a large ligand like ATP on a flexible receptor such as a kinase catalytic site might constitute a complete study by itself. Moreover, comparative manual modeling using available ATP-bound structures of other kinases would most probably have been too approximate, as c-Met is known to be quite specific amongst tyrosine kinases, both sequentially and structurally (32). More importantly, some c-Met ligands were known to exhibit binding specificities by targeting unusual allosteric regions (23), and those features outside the classical kinase binding site model coverage are of prime interest to us. In such a situation, and considering the fair number of available 3D structures of c-Met kinase complexes, we decided to develop a new c-Met specific kinase catalytic site model able to encompass the reported structural diversity of Met inhibitor binding.

Methods

Input structural data

As of August 2009 as much as 16 X-ray structures of c-Met kinase complexed with ATP-competitive ligands are publicly available in the PDB (33). Corresponding PDB IDs are: 1R0P, 2RFN, 2RFS, 2WD1, 2WGJ, 3C1X, 3CCN, 3CD8, 3CE3, 3CTH, 3CTJ, 3DKF/3DKG, 3EFJ, 3EFK, 3F66 and 3F82. The structure of c-Met in complex with ATP was released recently as 3DKC. One additional X-ray structure was found in an Agouron pharmaceuticals patent and will be referred to in this work as AGOU. The present study is entirely based on the analysis of this set of 18 experimental structures. Additional details about these can be found in Supporting Information.

Secondary structure prediction

Secondary structure elements of c-Met as shown in the crystal structures were defined by consensus using the DSSP (34) and Stride (35) tools available from the PDB website interface ('Sequence details' tab). Secondary structure annotations in the PDB file itself are chosen wherever DSSP and Stride predictions are not in agreement. This information is absent in AGOU: in this case, to best match the other structures results, Stride was used to define helices and DSSP to define sheets.

Representation of 3D molecular structures

We used VMD (36) for visualizing the molecular structures. For clarity, we represented the N-side and C-side lobes separately, the frontier between them being the double backbone H-bonding residue at the core of kinases catalytic site on which ATP binding is centered (37), identified as M_{1160} for c-Met.

Molecular modeling from the crystal structures

As none of the crystal structures was defect-free (see Supporting Information), curative molecular modeling was required. First, the PDB files were treated as to retain a single chain (chain A chosen for dimers 2RFN, 3EFJ, 3EFK and 3F66) and a single coordinate set for all atoms (first proposal retained if none with better occupancy factor). Homology modeling (38) using InsightII (Accelrys) was then conducted between residues 1048 and 1365 using a superposition of all available c-Met structures, including those without a bound ligand (PDB codes: 1R1W and 2G15). Missing residues and atoms (including hydrogens according to pH 7 amino acid protonation) were added, and mutations reversed to wild-type. Solvent molecules and cofactors were removed. In two X-rays (2RFS and AGOU), some ligand parts were not resolved: those missing parts were modeled back.

The structures completed as described above were minimized using NAMD (39) and the CHARMM22 forcefield (40) in which parameters for the ligands were added (41). The added parts were first submitted to 6400 steepest descent followed by 6400 conjugate gradient minimization steps. The whole protein structure was then slightly relaxed using 1000 steepest descent steps. It was verified that this process does not temper with the ligands binding as observed in the crystals.

An interaction energy-based approach to the binding site description

In previous studies, we successfully described protein–ligand (42,43) or protein–protein (44) interfaces on the basis of interaction energies distribution. When a molecular system is mapped into a classical molecular mechanics forcefield, non-covalent interactions are simply the sum of the van der Waals and Coulomb terms for each atom pair. Therefore, in the case of a protein–ligand system, the interaction energy can be trivially decomposed on different protein and ligand parts. In this work, we compute the distribution of ligand interaction on c-Met residues in order to define qualitatively distinct clusters of residues involved in ligand binding.

Calculations and graphical representation of ligands interaction distribution on c-Met residues

The NAMD 'pair interactions' feature allows to redefine the energy output during a simulation from the total energy of the system (default) to the interaction energy (van der Waals + Coulomb terms computed from the associated forcefield parameters) between two previously-defined non-bonded groups of atoms (referred to as a *pair*). NAMD is primarily a program for performing molecular dynamics but can be turned into simply calculating interaction energies between a protein residue and the ligand on a single structure by (i) converting the system coordinates (in PDB format) into a 1-frame DCD-format trajectory using the catdcd tool, (ii) launching a single-timestep dynamics reprocessing run at 0 K with the pair interaction option activated [a], and (iii) parsing the output text. With the help of in-house shell scripts, the aforementioned steps are fully automated so that a single batch of NAMD runs is launched for the analysis of each of the 18 crystal-derived models.

Using an in-house program, we processed the results of the aforementioned molecular mechanics calculations to a graphical representation using a color scale (up to -10 kcal mol^{-1} here) to represent interaction energy fluctuations, and the residues being put on a vertical axis. The energy scale extends to -10 kcal mol^{-1} as no amino acid–ligand (excluding ATP) interaction was measured significantly higher in one of the complexes. It should be noted that all ligands (except ATP) studied here are neutral electronically, should we have had charged ligands, it would have probably resulted in a larger scale requirement (42).

Results and Discussion

General description of ligand binding on the c-Met kinase domain

A nomenclature for kinase family sequences was defined by Hanks in 1988 (45,46), identifying 12 conserved subdomains noted in roman numerals (I, II, III, IV, V, VI_A, VI_B, VII, VIII, IX, X and XI). Those regions were identified on c-Met (see Table 1) by performing a sequence alignment with other kinases (see Supporting Information). Ligands interactions primarily involve the II, V and VII domains and negligible after VII. Residues outside Hanks' domains are secondarily targeted (see Figure 1), prompting for special attention be given to them upon inhibitor optimization. It should be noted that despite c-Met sequence being typical of receptor tyrosine kinases in this area (47), the flexibility of the VI_B and VII domains residues could confer conformational specificity to specific classes of inhibitors.

Geometrically, as most receptor tyrosine kinases (48), the c-Met kinase domain is constituted of two large subdomains, the N-terminal lobe mostly composed of strands and the C-lobe mostly composed of helices: a $\beta\beta\beta\alpha\beta\beta$-$\alpha\alpha\beta\beta\alpha\alpha\alpha$ secondary structure is observed in all X-ray complexes. Some residues are poorly defined, especially those of the large activation loop between β7 and α4, but the number of crystal structures allows for defining a clear consensus on c-Met secondary structure (see Table 2 and Supporting Information).

Table 1: c-Met kinase conserved regions (also see Figure 1)

Region	c-Met residues	Sequence
I	1084 → 1092	9 : I**G**R**G**H**FG**C**V**
II	1107 → 1112	6 : *C***AVK**SL
III	1127 → 1131	5 : **E**GI*I*M
IV	1136 → 1146	11 : **H**PNVLSLLGIC
V	1153 → 1169	17 : PLVVLPYMKHGD**L**RNF/
VI_A	1176 → 1196	21 : PTVKDLIGFGLQVAKGMK**YL***A*
VI_B	1200 → 1212	13 : FV**HRDL***AAR***N***CML*
VII	1218 → 1226	9 : V**K**VA**DFGL**A
VIII	1245 → 1255	11 : *L*PVKWM**A***L***E**SL
IX	1260 → 1284	25 : FTTKS**D**VWSF**G**VVLWELMTRGAP**P**Y
X	1292 → 1296	5 : ITVYL
XI	1308 → 1340	33 : CPDPLYEVMLKCWHPKAEM**R**PSFSE LVSRISAI

Residues in boldface/boldface-underlined are those found as observed in most/all kinases, while on the contrary, residues in italic differ in c-Met compared to most kinases.

Table 2: Secondary structure elements of the c-Met kinase domain (also see Figure 2)

Element	c-Met residues	Sequence
β1	1076 → 1084	9 : LIVHFNEVI
P-loop	1085 → 1090	6 : GRGHFG
β2	1091 → 1100	10 : CVYHGTLLDN
β3	1104 → 1110	7 : KIHCAVK
α1/C-helix	1118 → 1133	16 : IGEVSQFLTEGIIMKD
β4	1144 → 1146	3 : GIC
β5	1154 → 1158	5 : LVVLP
Hinge	1159 → 1164	6 : YMKHGD
α2	1165 → 1170	6 : LRNFIR
α3	1178 → 1198	21 : VKDLIGFGLQVAKGMKYLASK
β6	1210 → 1212	3 : CML
β7	1218 → 1220	3 : VKV
A-loop	1221 → 1251	31 : ADFGLARDMYDKEFDSVHNKT GAKLPVKWMA
α4	1252 → 1257	6 : LESLQT
α5	1262 → 1278	17 : TKSDVWSFGVVLWELMT
α6	1292 → 1298	7 : ITVYLLQ
α7	1310 → 1320	11 : DPLYEVMLKCW

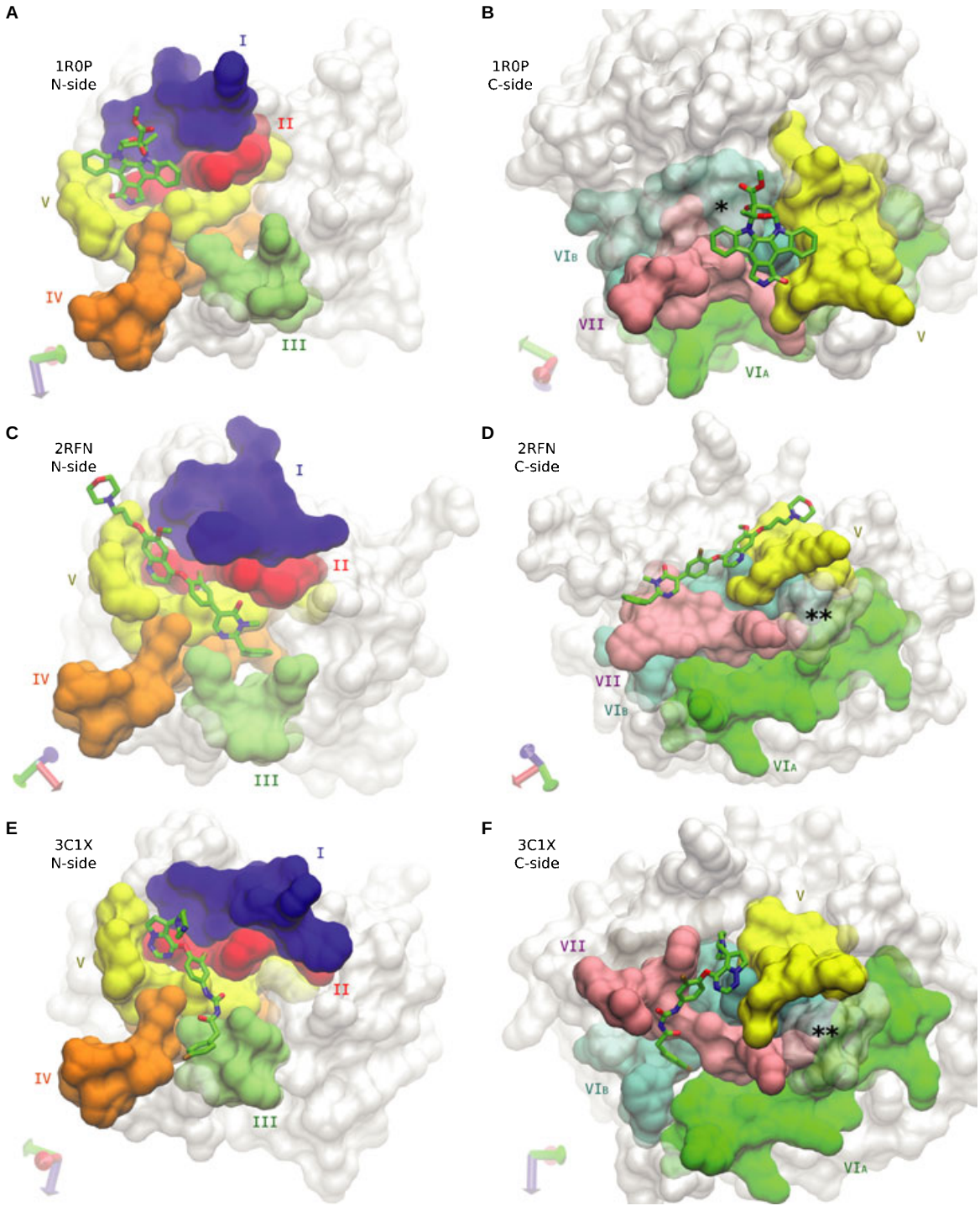


Figure 1: Kinase conserved elements (see Table 1) in c-Met structure (surface representation) and position of the ATP binding site, as shown in the 1R0P, 2RFN and 3C1X complexes. The N-side and C-side domains in the reconstructed model are represented separately for clarity. The ligands (C in green, O red, N blue, H hidden) are shown as to locate the binding site position. (A) 1R0P complex, N-side domain. Conserved domain I in blue, II in red, III in light green, IV in orange, V (1153–1160) in yellow. (B) 1R0P, C-side domain. V(1160–1169): yellow, VI_A: green, VI_B: light blue, VII: pink. Others: white. Domains VIII, IX, X and XI are not involved in the binding of any of the ligands studied here. Other c-Met residues in transparent white. *: amongst those, M_{1229} and Y_{1230} are mostly interacting with the ligand. (C) 2RFN, N-side. Same colors as in (A). (D) 2RFN, C-side. Same colors as in (B). **: $_{1213}$DEFKT$_{1217}$ linker between VI_B and VII. (E) 3C1X, N-side. Same colors as in (A). (F) 3C1X, C-side. Same colors as in (B). ** as in (D).

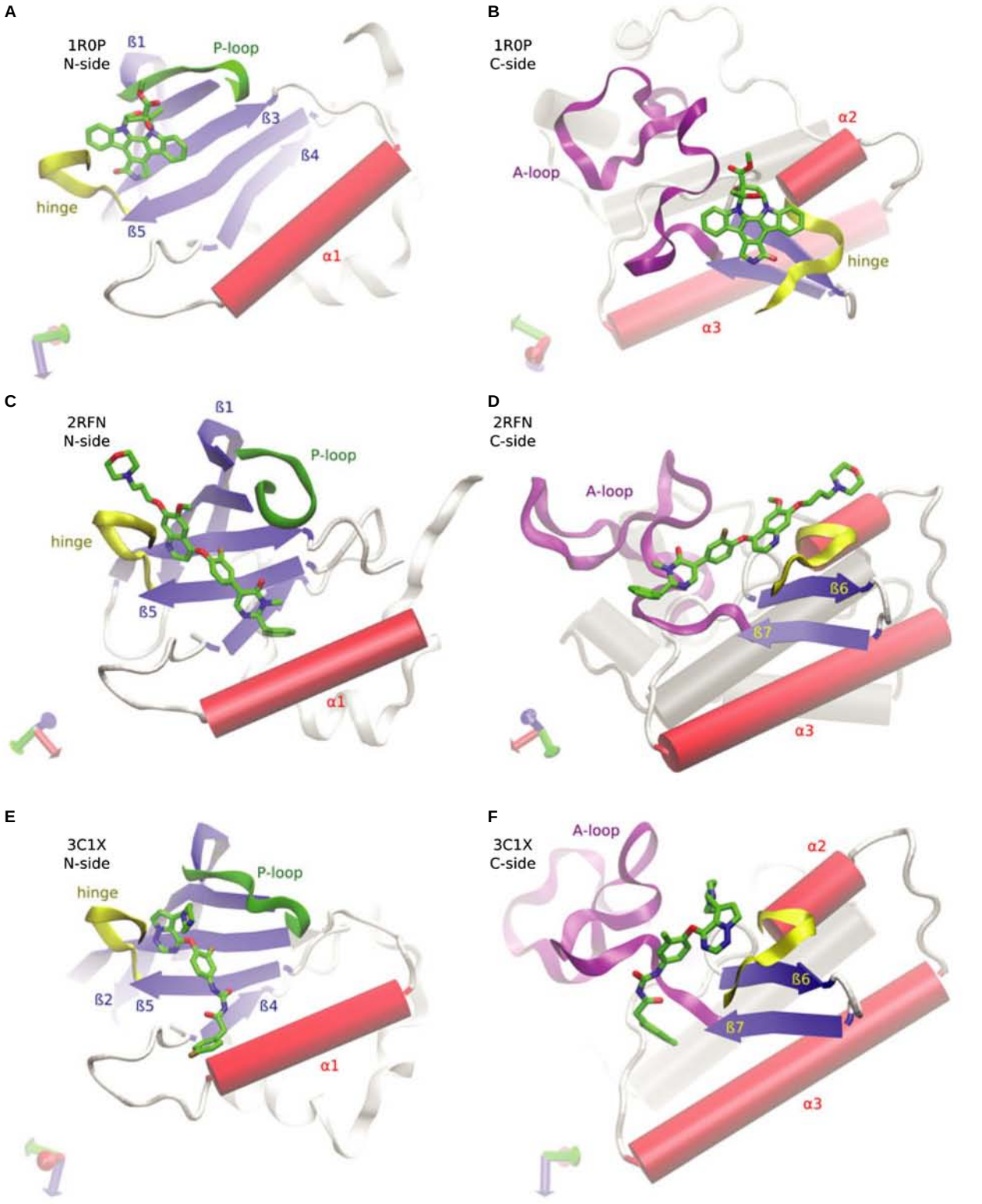


Figure 2: c-Met kinase domain secondary structure (see Table 2) in cartoon representation and position of the ATP binding site, as shown in the 1R0P, 2RFN and 3C1X complexes. Orientation is as in Figure 1. (A) 1R0P, N-side (residues 1049–1160). β sheets are colored in blue, C-helix in red, P-loop in green, hinge region in yellow. (B) 1R0P, C-side (residues 1160–1320). Hinge colored in yellow, β sheets in blue, helices α2 and α3 in red, A-loop in purple, helices α4, α5, α6 and α7 in grey. Residues after α7 (1321–1365) are not represented for clarity. (C) 2RFN, N-side, similar to (A). (D) 2RFN, C-side, similar to (B). (E) 3C1X, N-side, similar to (A). (F) 3C1X, C-side, similar to (B). The β6-β7 linker is the same as VI_B-VII (Figure 4D,F).

As with kinase conserved domains, a limited number of secondary structure elements are involved in ligand binding, the most notable being β3, the C-helix, the hinge between the two lobes and the A-loop. Some of those may adopt distinct geometries depending on ligands. α4, α5, α6 and α7 are clearly not part of the receptor (see Figure 2). Largest involved secondary structure elements, especially the A-loop, may adopt several binding geometries, depending on the ligands.

Ligand binding areas: desired specifications

The desired properties for the distinct receptor regions forming our model were defined as follows, from the most important to the least:

- Binding areas are related to contiguous and non-overlapping sequences of residues so that they are easily referred to the conserved kinase regions (Table 1; Figure 1) and secondary structure elements (Table 2; Figure 2). This is in our opinion particularly important as not to compromise the use of the model as part of a chemogenomics strategy.

- Each region is associated to a single possible mode of binding for inhibitors, so that the binding mode of a given ligand is implicitly and unambiguously given by the list of the regions in which it is mostly interacting.

- Ligand interaction energy should not fluctuate much between the areas so that those have comparable weight energetically.

- Two major interactions must be grouped in a single area if they are related in all reference complexes, or assigned in two distinct areas if not. This feature is also desired for secondary interactions, whenever possible regarding previous requirements.

- Each region should be extended until the corresponding gain of its total energy weight is minimal in all cases, or until the next residue to add corresponds to a single ligand-specific secondary interaction. As a result, it should be verified in all reference complexes that the ligand interactions outside the model coverage is minimal.

Ligand binding areas: identification

A graphical representation of the ligands interactions distribution on c-Met residues is obtained by applying the previously-described pair interaction molecular mechanics-based methodology (see Figure 3). By using this data and the guidelines presented above, we identified seven distinct ligand binding areas, listed in Table 3. These will be further referred to as 1, 2, 3, 4, 5, 6 and 7. Three representative structures are shown on Figure 4 with the 7 binding areas highlighted. Tables 1, 2 and 3 are related to Figures 1, 2 and 4, respectively. Figure 5 projects the groups definitions in Tables 1, 2 and 3 on the c-Met sequence.

Ligand binding areas: structural description

According to the Novartis model (27,31), the kinases catalytic site is a well-defined cavity which is 'floored' by the hinge region on which the ATP adenine group is H-bonded, whose 'walls' are hydrophobic and with a half-open 'ceiling' around the P-loop. The 'sky' extends to the A-loop in which the ATP phosphates bind. In our model, the adenine core region is clearly area 4, the P-loop ceiling area 1, and the phosphate-binding region area 7. These

Table 3: Definition of the ligand binding areas in the c-Met-specific model described in this paper (also see Figure 4)

Group	c-Met residues	Sequence
1	1083 → 1094	12 : VIGRGHFGCVYH
2	1105 → 1112	8 : IHCAVKSL
3	1123 → 1146	24 : QFLTEGIIMKDFSHPNVLSLLGIC
4	1154 → 1164	11 : LVVLPYMKHGD
5	1194 → 1212	19 : YLASKKFVHFDLAARNCML
6	1218 → 1225	8 : VKVADFGL
7	1226 → 1233	8 : ARDMYDKE

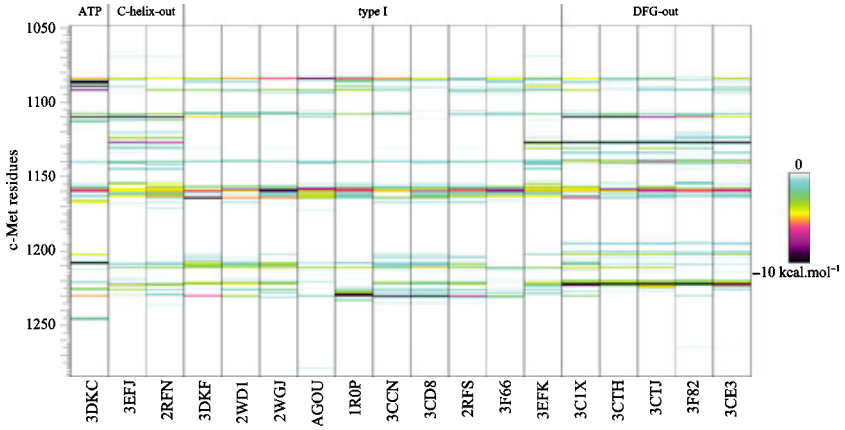


Figure 3: Repartition of protein-ligand interaction energies on c-Met kinase residues as calculated from the crystal-derived structures of c-Met complexes. The complexes were ordered as to show more clearly the different interaction patterns clusters (identified on top).

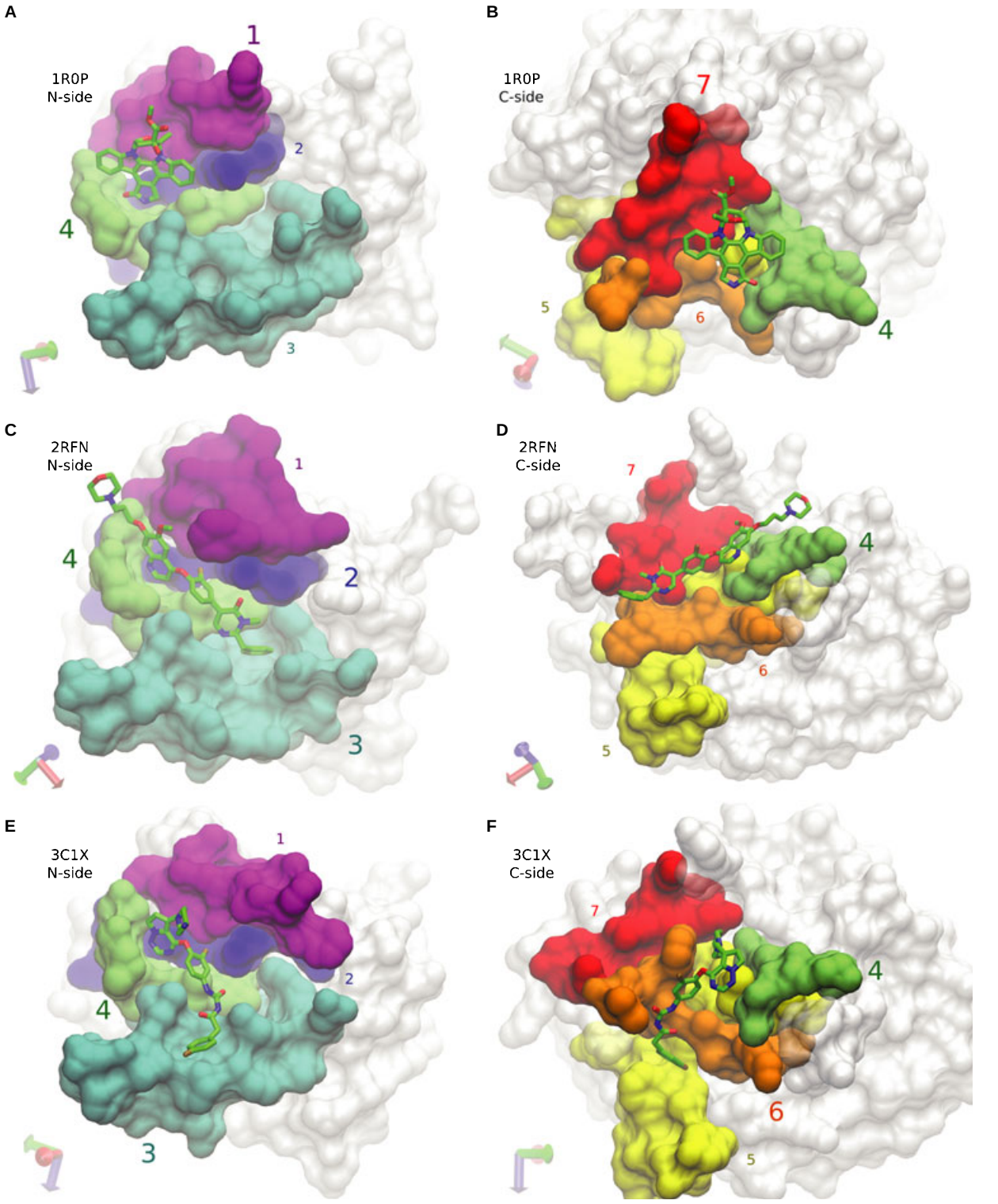


Figure 4: Ligand binding to c-Met kinase domain subdivided using the model defined in this paper (see Table 3). The distinct ligand binding modes are: 1-4-7 for 1R0P (ATP binding mode), 2-3-4 for 2RFN (C-helix-out), and 3-4-6 for 3C1X (DFG-out, similar to Imatinib). Orientation is as in Figures 1 and 2. (A) 1R0P, N-side (residues 1049–1160). 1: purple, 2: blue, 3: light blue, 4 (1154–1160): green. (B) 1R0P, C-side (residues 1160–1365). 4 (1160–1164): green, 5; yellow, 6: orange, 7: red. (C) 2RFN, N-side, similar to (A). (D) 2RFN, C-side, similar to (B). (E) 3C1X, N-side, similar to (A). (F) 3C1X, C-side, similar to (B). Note the 5-6 linker, identical to VI_B-VII and $\beta6$-$\beta7$ (Figures 1D,F and 2D,F).

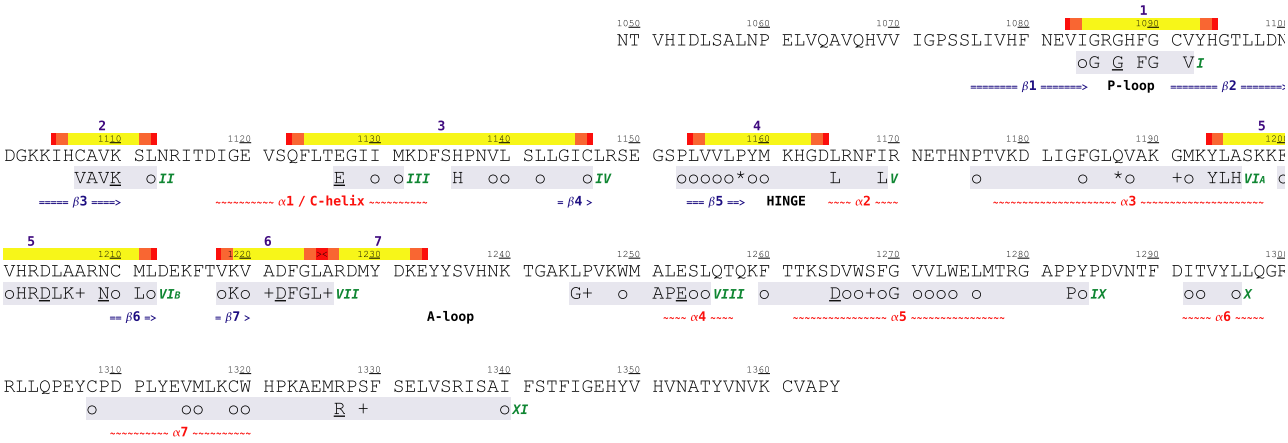


Figure 5: Classification of c-Met sequence residues. The 7 ligand binding areas defined in this work (depicted as numerals 1 to 7) are shown on the first line. Areas borders are colored in orange-red: note that 6 ends at L_{1225} while 7 begins at A_{1226}. Kinase conserved regions (roman numerals I to XI) are represented in light blue background immediately below the sequence. Residues conserved in most kinases are represented in those regions, and residues conserved in all known kinases are underlined. o denotes an hydrophobic residue, + any small and neutral residue, and * any polar residue. c-Met substructure elements are represented in the last line, α helices in red and β strands in blue.

relationships can be seen on the structure of c-Met bound to the K-252a ligand (1R0P complex, see Figures 2A,B and 4A,B) whose binding mode, as a staurosporine analogue, is close to ATP; they are also confirmed by visual analysis of the X-ray structure of c-Met in complex with ATP (3DKC).

Comparing with the determined kinase conserved regions (Table 1) and c-Met secondary structure elements (Table 2), the ligand binding areas classification (Table 3) delineates the central hinge interacting residues (area 4), 3 residues groups C-side to it (1, 2 and 3) and 3 groups N-side (5, 6 and 7). On the C-side, 1 corresponds to the P-loop and conserved sequence region I, 2 is the union of β3 and II, and 3 is a large subsequence that encompasses half the C-helix, β4, III and IV. 4 is the union of β5 and the hinge (or V minus α2). On the N-side, 5 encompasses the large hydrophobic region from the end of α3 up to β6, 6 corresponds to VII (conserved region including β7 and the first A-loop residues) and 7 is the interacting part of the A-loop that is more c-Met specific (see Figure 5).

Validation of the model

We will now review the c-Met model in agreement with the previously-defined specifications of its ligand binding areas components.

- Seven continuous and non-overlapping sequences constitute the model.

- Structurally, each ligand binding area is either a well-defined pocket or a large hydrophobic region, so that only one major local binding mode is allowed for targeting ligands

- The energetical weights are on the same order of magnitude.

- Four distinct polar residues involved in ligand binding are most easily seen on Figure 5. Each corresponds to at least one H-bond and is located on a specific ligand binding area: K_{1110} on 2, E_{1127} on 3, D_{1222} on 6 and Y_{1230} on 7. Except for area 5 (secondary interactions), there is no notable difference in binding between two interacting ligands in a given region.

- The amount of interaction energy outside the defined regions in X-ray resolved residues is negligible in all cases (<5%), so that our binding site model has an effective coverage of the protein-ligand interactions observed in all reference complexes.

The model defined on Table 3 therefore respects our initially-defined specifications for c-Met experimental structures. There has been a continuous stream of c-Met crystals published in the PDB since 2 years, consequently delaying the completion of the present study, but also allowing us to provide a well-refined model. It should be noted that the release of a crystal structure featuring a new binding mode in the future, extending the range of the current ligand binding areas or revealing new interactions according to our methodology, would require us to revise our model. Interestingly, the ligand binding areas definition presented in Table 3 did not have to undergo such a revision since June 2008 (8 structures out of 17), suggesting that the binding modes of newest inhibitors do not differ much from those of the previous ones, and that the model is mature.

The release of the ATP-bound c-Met complex confirmed *a posteriori* the relationships that were proposed between 5 of the 7 regions in our model and the regions constituting the Novartis kinase-ATP generic model. In the 3DKC crystal, ATP does not significantly target c-Met further on the A-loop as compared to the other co-crystallized ligands. This is caused by the mutations that were made to the crystallized c-Met sequence (especially Y_{1234}F and Y_{1235}D) in order to prevent the phosphorylation process: ATP is not interacting in its native binding mode in this region. However, the additional A-loop residues targeted by ATP phosphates

(mostly Y_{1234} and Y_{1235}) are probably out of reach for drug-like hinge-bound inhibitors so their precise identification is not needed for the purpose of our model.

Description of ligand binding on c-Met using the new model

Jacobs *et al.* demonstrated recently using two simple geometrical metrics that kinase complexes can be generally classified in three distinct conformational clusters, the first being the active state, and the two others two structurally different inactive states (49). Interestingly, two corresponding allosteric pockets are identified. The first allosteric region is only accessible when the well-conserved DFG triad (found at positions 1222–1224 on c-Met sequence) bordering the A-loop N-side are rotated away from the active site. Imatinib binds there in addition to the catalytic core, locking its target kinases A-loop in an inactive state. The second allosteric site is opened by C-helix displacement and is not taken into account in the generic kinase model as updated (31) to describe Imatinib binding.

The 'DFG-out' inactive kinase conformation is revealed in our c-Met model by area 6 and is characterized by a specific strong H-bonding interaction of the corresponding ligands with D_{1222} of the DFG motif, keeping c-Met in an inactive state. The 'C-helix-out' inactive conformation is also clearly defined by our model. Indeed, some c-Met ligands feature a C-helix displacement (area 3) that allows for ligand access to the corresponding allosteric pocket defined by area 2. A strong H-bond with K_{1110} can be formed in such a case, presumably stabilizing c-Met inactivation similarly to the D_{1222} H-bond lock of DFG-out inhibitors. Therefore, a very interesting finding resulting from this study is that like Cdk6 and Abl1 only (as far as current knowledge on ATP-competitive kinase inhibitors goes) (49), c-Met allows ligands to bind using any of the three reported major kinase inhibitor binding modes.

The relationship between ligand binding mode and the specific targeting of areas 2 and 6 is further demonstrated by plotting the complexes distribution on the basis of the protein/ligand interaction energy in these two regions (Figure 6). The type I, DFG-out and C-helix-out binding modes identified by visual analysis are clearly forming three distinct clusters. There is only one complex, 3CTH, which is not clearly located, as its interaction energy in area 2 is higher than what is found in the DFG-out cluster. Figure 6 should be put in parallel with Figure 3 in Jacobs' paper (49) as the purpose, identifying kinase ligand binding modes, is exactly the same. Jacobs' approach is applicable to all kinases while ours is c-Met specific. Our interaction energy-based methodology measures directly whether a ligand binds or not to the kinase allosteric regions, while Jacobs' geometrical metrics are more indirect, as they are related to those regions accessibility. This leads to approximations in both cases, as a ligand can either significantly interact with allosteric pockets residues without the pocket being made accessible (*e.g.* 3CTH), or, on the opposite, provide accessibility by induced fit effects without actually binding there (*e.g.* 1R0P). In this context, it is clear that the two binding mode identification qualitative models are perfectly complementary, so that a consensus approach would certainly provide better results.

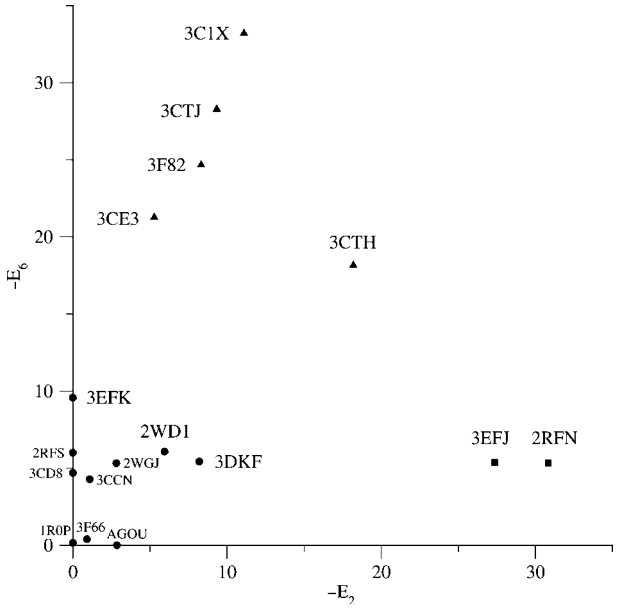


Figure 6: Distribution of the c-Met complexes ligands interaction energy (in kcal mol^{-1}) on the 2 and 6 regions of the receptor model described in this paper. Circles represent complexes in which the ligand targets the Met kinase in its activated state. Triangles and squares represent those in which the ligand is structurally bound to the 'DFG-out' and 'C-helix-out' pockets respectively, in addition to interactions also found in activated state/type I ligands.

According to the X-ray structures, K-252a (1R0P) and SU11274 (2RFS) are typical type I ATP-like inhibitors (defining the binding mode as the list of targeted pharmacophores in our model: 1-4-7, see Figure 4A,B), while the 2RFN/3EFJ complexes feature type II C-helix-out hit structures (2-3-4, Figure 4C,D), and 3C1X/3CTJ obvious type II DFG-out Imatinib analogues (3-4-6, Figure 4E,F). Interestingly, the distribution of primary interactions shows that the two allosteric binding modes might be closer than what the drastically different inhibitor conformations insinuate, suggesting possible optimization strategies. Indeed, it is observed that without actually binding to the C-helix-out pocket, several inhibitors bound to inactive c-Met in the DFG-out conformation (especially 3C1X and 3CTH, see Figure 3) can possess the K_{1110} H-bond that presumably favors the C-helix-out state. Some variations are also seen amongst the binding patterns of the active state-targeting (type I) ligands. Those inhibitors are either 'oriented' towards C-side area 1 (2WD1, 2WGJ, AGOU) or N-side area 7 (2RFS, 3CCN, 3CD8, 3DKF). With the K-252a large staurosporine analog (1R0P), both areas 1 and 7 are reached, while the small ligand in the 3F66 complex can target neither of these two regions. 3EFK has a specific interaction pattern in the type I group, as it is targeting area 3, a feature shared by both kind of 'allosteric' inhibitors.

Conclusions and Future Directions

Imatinib demonstrated that drugs could efficiently target inactive kinases and is often cited as a typical success story in modern drug design (20). Interestingly, initial structure-based lead optimization attempts failed in this case as long as modeling was done

on the unsuited kinase active state model. *In silico* structure-based drug design can only be successful when based upon an appropriate reference model. As the number and diversity of kinase complexes resolved crystal structures constantly increases, the usefulness of more precise and specific kinases catalytic sites descriptions for drug design is clear. This appears especially true in the case of c-Met.

The new c-Met kinase-specific binding site model defined in this paper will now constitute the backbone of future structure-based drug design strategies targeting c-Met in our group. It might be interesting to extend the ligand binding mode identification approach described here for c-Met to other members of the kinase family, with the help of Jacobs *et al.* findings (49). This could constitute the basis for a more extended structure-based kinomics (50) investigation.

This work highlights c-Met catalytic site plasticity and as a kinase target, the fact that three distinct binding modes can be adopted by ligands could open new routes towards the discovery of novel, highly potent and specific classes of inhibitors. This work also points out that the generic kinase active site model, which currently relies on ATP and Imatinib binding features (31), should be further improved in order to describe the interactions of inhibitors bound in the C-helix-out mode. The choice of a reference compound for doing so might be debated amongst kinase drug design specialists.

Acknowledgments

The present work was funded by INCa (Institut National du Cancer) and ARC (Association pour la Recherche sur le Cancer) grants. Vincent Leroux and Yasmine Asses received financial support from INCa.

References

1. Birchmeier B., Birchmeier W., Gherardi E., Vande Woude G.F. (2003) MET, metastasis, motility and more. Nat Rev Mol Cell Biol;4:915–925.
2. Ma P.C., Maulik G., Christensen J., Salgia R. (2003) c-Met: structure, functions and potential for therapeutic intervention. Cancer Met Rev;22:309–325.
3. Michalopoulos G.K., DeFrances M.C. (1997) Liver regeneration. Science;276:60–66.
4. Birchmeier C., Gherardi E. (1998) Developmental roles of HGF/SF and its receptor, the c-Met tyrosine kinase. Trends Cell Biol;8:404–410.
5. Maina F., Panté G., Helmbacher F., Andres R., Porthin A., Davies A.M., Ponzetto C., Klein R. (2001) Coupling Met to specific pathways results in distinct developmental outcomes. Mol Cell;7:1293–1306.
6. Chmielowiec J., Borowiak M., Morkel M., Stradal T., Munz B., Werner S., Wehland J., Birchmeier C., Birchmeier W. (2007) c-Met is essential for would healing in the skin. J Cell Biol;177:151–162.
7. Maulik G., Shrikhande A., Kijima T., Ma P.C., Morrison P.T., Salgia R. (2002) Role of the hepatocyte growth factor receptor, c-Met, in oncogenesis and potential for therapeutic inhibition. Cytokine Growth Factor Rev;13:41–59.
8. Christensen J.G., Burrows J., Salgia R. (2005) c-Met as a target for human cancer and characterization of inhibitors for therapeutic intervention. Cancer Lett;225:1–26.
9. Zhang Y.-W., Vande Woude G.F. (2003) HGF/SF-Met signaling in the control of branching morphogenesis and invasion. J Cell Biochem;88:408–417.
10. Mazzone M., Comoglio P.M. (2006) The Met pathway: master switch and drug target in cancer progression. FASEB J;20:1611–1621.
11. Gentile A., Trusolino L., Comoglio P.M. (2008) The Met tyrosine kinase receptor in development and cancer. Cancer Met Rev;27:85–94.
12. Engelman J.A., Zejnullahu K., Mitsudomi T., Song Y., Hyland C., Park J.O., Lindeman N. *et al.* (2007) MET amplification leads to Gefitinib resistance in lung cancer by activating ERBB3 signaling. Science;316:1039–1043.
13. Shattuck D.L., Miller J.K., Carraway K.L., Sweeney C. (2008) Met receptor contributes to trastuzumab resistance of Her2-overexpressing breast cancer cells. Cancer Res;68:1471–1477.
14. Guo A., Villén J., Kornhauser J., Lee K.A., Stokes M.P., Rikova K., Possemato A. *et al.* (2008) Signaling networks assembled by oncogenic EGFR and c-Met. PNAS;105:692–697.
15. Shen Y., Naujokas M., Park M., Ireton K. (2000) InlB-dependent internalization of listeria is mediated by the Met receptor tyrosine kinase. Cell;103:501–510.
16. Cunha-Rodrigues M., Prudêncio M., Mota M.M., Haas W. (2006) Antimalarial drugs - host targets (re)visited. Biotechnol J;1:321–332.
17. Knudsen B.S., Vande Woude G.F. (2008) Showering c-MET-dependent cancers with drugs. Curr Opin Genet Dev;18:87–96.
18. Cui J.J. (2007) Inhibitors targeting hepatocyte growth factor receptor and their potential therapeutic applications. Expert Opin Ther Pat;17:1035–1045.
19. Druker B.J., Talpaz M., Resta D.J., Peng B., Buchdunger E., Ford J.M., Lydon N.B., Kantarjian H., Capdeville R., Ohno-Jones S., Sawyers C.L. (2001) Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. N Engl J Med;344:1031–1037.
20. Capdeville R., Buchdunger E., Zimmermann J., Matter A. (2002) Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug. Nat Rev Drug Discov;1:493–502.
21. Brehmer D., Greff Z., Godl K., Blencke S., Kurtenbach A., Weber M., Müller S., Klebl B., Cotten M., Kéri G., Wissing J., Daub H. (2005) Cellular targets of Gefitinib. Cancer Res;65:379–382.
22. Koshland D.E. (1994) The key-lock theory and the induced fit theory. Angew Chem;33:2375–2378.
23. Dussault I., Bellon S.F. (2008) c-Met inhibitors with different binding modes. Two is better than one. Cell Cycle;7:1157–1160.
24. Miller J.R., Dunham S., Mochalkin I., Banotai C., Bowman M., Buist S., Dunkle B. *et al.* (2009) A class of selective antibacterials derived from a protein kinase inhibitor pharmacophore. PNAS;106:1737–1742.
25. Knighton D.R., Zheng J.H., Ten Eyck L.F., Ashford V.A., Xuong N.H., Taylor S.S., Sowadski J.M. (1991) Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. Science;253:407–414.

26. Knighton D.R., Zheng J.H., Ten Eyck L.F., Xuong N.H., Taylor S.S., Sowadski J.M. (1991) Structure of a peptide inhibitor bound to the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. Science;253:414–420.
27. Furet P., Caravatti G., Lydon N., Priestle J.P., Sowadski J.M., Trinks U., Traxler P. (1995) Modelling study of protein kinase inhibitors: binding mode of staurosporine and origin of the selectivity of CGP 52411. J Comput Aided Mol Des;9:465–472.
28. Traxler P., Furet P. (1999) Strategies toward the design of novel and selective protein tyrosine kinase inhibitors. Pharmacol Ther;82:195–206.
29. Schindler T., Bornmann W., Pellicena P., Miller W.T., Clarkson B., Kuriyan J. (2000) Structural mechanism for STI-571 inhibition of Abelson tyrosine kinase. Science;289:1938–1942.
30. Cowan-Jacob S.W., Fendrich G., Floersheimer A., Furet P., Liebetanz J., Rummel G., Rheinberger P., Centeleghe M., Fabbro D., Manley P.W. (2007) Structural biology contributions to the discovery of drugs to treat chronic myelogenous leukaemia. Acta Crystallogr;63:80–93.
31. Liu Y., Gray N.S. (2006) Rational design of inhibitors that bind to inactive kinase conformations. Nat Chem Biol;2:358–364.
32. Kinnings S.L., Jackson R.M. (2009) Binding site similarity analysis for the functional classification of the protein kinase family. J Chem Inf Model;49:318–329.
33. Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E. (2000) The Protein Data Bank. Nucleic Acids Res;28:235–242.
34. Kabsch W., Sander C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers;22:2577–2637.
35. Frishman D., Argos P. (1995) Knowledge-based protein secondary structure assignment. Proteins;23:566–579.
36. Humphrey W., Dalke A., Schulten K. (1996) VMD: visual molecular dynamics. J Mol Graph;14:33–38.
37. Ghose A.K., Herbertz T., Pippin D.A., Salvino J.M., Mallamo J.P. (2008) Knowledge based prediction of ligand binding modes and rational inhibitor design for kinase drug discovery. J Med Chem;51:5149–5171.
38. Sánchez R., Sali A. (1997) Advances in comparative protein-structure modelling. Curr Opin Struc Biol;7:206–214.
39. Phillips J.C., Braun R., Wang W., Gumbart J., Tajkhorshid E., Villa E., Chipot C., Skeel R.D., Kalé L., Schulten K. (2005) Scalable molecular dynamics with NAMD. J Comp Chem;26:1781–1802.
40. MacKerell A.D. Jr, Bashford D., Bellott M., Dunbrack R.L., Evanseck J.D., Field M.J., Fischer S. *et al.* (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B;102:3586–3616.
41. Momany F.A., Rone R. (1992) Validation of the general purpose QUANTA 3.2/CHARMm force field. J Comp Chem;13:888–900.
42. Leroux V., Gresh N., Liu W.-Q., Garbay C., Maigret B. (2007) Role of water molecules for binding inhibitors in the SH2 domain of Grb2: a molecular dynamics study. J Mol Struc-Theochem;806:51–66.
43. Patané S., Pietrancosta N., Hassani H., Leroux V., Maigret M., Kraus J.-L., Dono R., Maina F. (2008) A new Met inhibitory-scaffold identified by a focused forward chemical biological screen. Biochem Biophys Res Comm;375:184–189.
44. Déliot N., Chavent M., Nourry C., Lécine P., Arnaud C., Hermant A., Maigret B., Borg J.-P. (2009) Biochemical studies and molecular dynamics simulations of Smad3–Erbin interaction identify a non-classical Erbin PDZ binding. Biochem Biophys Res Comm;378:360–365.
45. Hanks S.K., Quinn A.M., Hunter T. (1988) The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. Science;241:42–52.
46. Hanks S.K., Hunter T. (1995) The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. FASEB J;9:576–596.
47. Hanks S.K. (2003) Genomic analysis of the eukaryotic protein kinase superfamily: a perspective. Genome Biol;4:111.
48. Moretti L. (2007) Exploring Structure and Plasticity of Tyrosine Kinase Domains for Drug Discovery. PhD thesis. Université de Genève.
49. Jacobs M.D., Caron P.R., Hare B.J. (2008) Classifying protein kinase structures guides use of ligand-selectivity profiles to predict inactive conformations: structure of lck/imatinib complex. Proteins;70:1451–1460.
50. Vieth M., Higgs R.E., Robertson D.H., Shapiro M., Gragg E.A., Hemmerle H. (2004) Kinomics - structural biology and chemogenomics of kinase inhibitors and targets. Biochim Biophys Acta;1697:243–257.

Note

[a]An appropriate example configuration file is given in the NAMD manual. http://www.ks.uiuc.edu/Research/namd/current/ug.pdf

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Supporting Information

Figure S1. Ligands featured in the available X-ray structures of c-Met complexes. On the right: X-ray PDB code, ligand name (if any), resolution and release date of structure, origin, patent(s) for compound (if any).

Figure S2. Alignment of the human c-Met kinase domain sequence with those of other tyrosine kinases whose ATP-complexed structure is known.

Figure S3. Overview of the X-ray resolved c-Met complexes, their secondary structures, and their defects subsequently corrected during the modeling process described in the paper.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

III- Commentaires

Ce travail nous a permis de mettre en évidence 3 types de modes de liaisons basés sur la description structurale de 17 structures cristallographiques de la PDB et d'un brevet de la kinase c-Met. Nous avons défini ces 3 types de mode de liaisons comme étant : Type I (Type I classique connu pour la plupart des kinases), Type II (ou ATP-like) et Type III qui regroupe le type DFG-out et C-helix-out. L'existence de ces types de modes de liaisons montre l'étendue de la plasticité du site catalytique de la cible kinase. De même que le fait que les ligands adoptent 3 modes de liaisons distincts peut élargir l'horizon en ce qui concerne le développement de nouvelles classes d'inhibiteurs hautement actifs et spécifiques. Nous avons également souligné le fait que le modèle du site actif de la kinase générique, qui repose actuellement sur les caractéristiques de liaisons de l'ATP et de l'Imatinib, devrait être amélioré en tenant compte des interactions décrites par les liaisons des inhibiteurs dans le mode C-helix-out, comme nous l'avons montré dans le présent article.

ARTICLE 3

I- Contexte

Compte tenu de l'étude précédente et des nouvelles structures parues en 2010 (les codes PDB correspondants sont 2WKM, 3LQ8, 3L8V, 3A4P, 3I5N), dont certaines étaient indiquées par leurs auteurs comme différentes de tout ce qui avait été publié auparavant, il nous a semblé nécessaire d'explorer de manière intensive l'espace conformationnel de c-Met. Ceci devait nous permettre d'une part de découvrir éventuellement de nouvelles conformations non encore « révélées » par l'expérience, d'autre part de voir comment pouvait être reliées les structures *holo* aux structures *apo*, donnant ainsi une mesure des phénomènes de sélection conformationnelle et d' « induced fit » (effet induit) provoqués par la liaison des ligands et de leur hiérarchie dans les mécanismes de liaison entre les partenaires. Il existe actuellement plusieurs travaux qui vont dans ce sens [249-252]

Pour effectuer ce genre de travail, plusieurs écoles de pensée existent, certaines privilégiant une exploration par dynamique moléculaire, d'autres l'utilisation des modes normaux de vibrations, d'autres encore préconisant l'utilisation de multiples conformations issues des études par RMN et cristallographie. Nous avons choisi de ne privilégier aucune de ces approches et de les utiliser toutes afin de pouvoir comparer leurs mérites et leurs limitations respectives.

II- Exploring c-Met kinase flexibility by sampling and clustering its conformational space

*Manuscript
Click here to view linked References

Exploring c-Met kinase flexibility by sampling and clustering its conformational space

Yasmine Asses[1,*], Vishwesh Venkatraman[1*], Vincent Leroux[1],
David W. Ritchie[1] and Bernard Maigret[1+]

1. Nancy Université, LORIA/UMR 7503, Équipe-projet Orpailleur, Campus Scientifique, BP 239, 54506 Vandœuvre-lès-Nancy Cedex, France

+: Corresponding author: bernard.maigret@loria.fr / Tel: +33 354 958 593 Fax: +33 383 278 319.

* : These two authors contributed equally to this work

Abstract

It is now widely recognized that the flexibility of both partners has to be considered in molecular docking studies. However, the question how best to handle the huge computational complexity of exploring the protein binding site landscape is still a matter of debate. Here we investigate the flexibility of c-Met kinase as a test case for comparing several simulation methods. The c-Met kinase catalytic site is an interesting target for anticancer drug design. In particular, it harbors an unusual plasticity compared to other kinases ATP binding sites. Exploiting this feature may eventually lead to the discovery of new anticancer agents with exquisite specificity.

We present in this paper an extensive investigation of c-Met kinase conformational space using large-scale computational simulations in order to extend the knowledge already gathered from available X-ray structures. In the process, we compare the relevance of different strategies for modeling and injecting receptor flexibility information into early stage *in silico* structure-based drug discovery pipeline. The results presented here are currently being exploited in on-going virtual screening investigations on c-Met.

Keywords; c-Met protein-ligand complexes; protein flexibility; conformational sampling; conformational preselection; induced fit.

INTRODUCTION

Studying molecular flexibility is now considered as fundamental to the understanding of the biological functioning of biological polymers [1-3] and has already been highlighted in many works (see for example ref. [4-6]). It is an important aspect of molecular recognition between a ligand and its receptor and its implication in drug discovery has already been pointed out in several reviews [7-11]. Protein flexibility has a significant role that should be taken into account in computer-aided drug design because many systems undergo rearrangements upon ligand binding, ranging from local motions of side-chains to large domain movements. To tackle efficiently full flexibility of both partners in docking and virtual screening approaches is therefore still challenging.

It has been proposed that molecular dynamics (MD) approaches can overcome most difficulties and several papers discuss the use of MD simulations for docking [12,13]. Nevertheless, an explicit and direct consideration of flexibility of the whole receptor and solvent effects in docking procedures using all-atoms MD is still computationally costly, except for researchers having dedicated MD machines possibilities such as the Anton [14]. Consequently it is difficult to use such MD simulations in an intensive virtual screening process. To limit the computing time, several other methods have been introduced, ranging from limiting protein flexibility to normal mode variations [15,16] to coarse-graining reduction of the protein/ligand/solvent system [17].

Because of the known limits of each of these exploration methods, it has been pointed out recently that the use of several receptor conformers can be a valuable strategy of incorporating flexibility into the docking problem with only a moderate computational cost [18-22]. But in this case, how should the conformers be generated and how many are needed? Furthermore, how should they be used or results be combined? Different approaches on how to use ensembles of receptor conformations in ligand docking have already been presented, ranging from the use of several existing X-ray data to normal modes and MD simulations to identify principal key motions of the protein of interest. But what are the merits and drawbacks of these different choices to describe extensively the protein conformational landscape? This question is central to the use of ensemble conformations [23].

This paper attempts to answer some of these issues by using and comparing different ensembles sampled for a particular case, namely the kinase domain of the c-Met receptor. c-Met kinase plays an important role in several human cancers [24-26] and a large body of work has been devoted to this protein for which several X-ray structures are available in complex with a wide variety of ligands (see **supplementary material**). Taken together, the large body of published 3D structures highlights the large plasticity of the c-Met kinase which must be taken into account during the *in*

silico design of new ligands [27].

In our study, we have used several current approaches simultaneously to investigate the kinase domain flexibility with three interests: (i) comparing the merits (and drawbacks) of different sampling procedures, (ii) detecting the most relevant conformational states of the kinase to be used subsequently in a virtual screening campaign aimed to discover putative c-Met blockers and (iii) proposing a possible role for conformational selection and induced-fit effects to be considered as a consequence of ligand binding.

For this purpose, we started to explore extensively the conformational landscape of the c-Met *apo* structure for which we successively performed a long 1 μs all-atom MD simulation in explicit solvent and also normal mode calculations followed by short all-atom MD simulations in explicit solvent. Shorter molecular dynamics runs were performed in order to describe the dynamic behaviour of the 26 *holo* c-Met crystallographic structures presently available in the complexed ligand bound form. The goal was to try to relate the *holo* kinase conformers to some of those found for the *apo* form of the protein. For that, all the produced samples were analysed in order to classify them into conformational clusters and to compare the representative conformers extracted from these families.

RESULTS

Sampling c-Met conformational space

The general flowchart of our sampling procedure is presented in Figure 1. In order to focus our analysis on the protein backbone conformational changes, we decided to perform the clustering calculations only on (i) the backbone part of the protein structures and (ii) the main core of the protein, namely residues 1,060 to 1,344. This would help avoid producing "noise" in our analysis due to the side chains variations and those due to the two highly flexible N- and C-terminal parts. Consequently, all the samples described in the next section concern only truncated structures as described above. A comparison of RMSD conformational maps of both of the complete and truncated protein had been made initially to justify our choice of considering only residues 1,060 to 1,344 from the backbone part.

Firstly, all the MD frames were grouped according to the sampling method used, giving several large conformational samples to be analyzed. **Table I** summarizes the number of frames collected from each sampling procedure.

The 1μs MD simulation of the *apo* structure produced around 1 million frames. We call this simulation the Long Molecular Dynamic (LMD). However, because it was not possible in the

present study to handle such a huge sample, this dataset was strided so that only 8,702 frames were finally retained for the cluster analysis.

From the 106 modes provided by the NOMAD-REF server for the *apo* structure, we retained only the three lowest modes 7 to 9 as the relevant ones for global movements of the protein. We call this simulation the Normal Mode Molecular Dynamic (NM_MD). In these modes, the RMSD of the protein backbone varied from 0 to 4 Å as shown on **figure 2** for mode 7 for example. From each normal mode RMSD curve, 4 conformations were selected as a starting point for the MD (**figure 2**) simulations so that a total of 12x10 ns MD runs were performed. The molecular dynamics simulations performed on these selected NM conformations produced a total of 10,500 conformational frames per mode so that a total of 126,000 frames was recorded and stride to give the 8,406 frames used for the clustering analysis.

10 ns of MD was performed on each of the 26 X-Ray *holo* PDB structures, here called X-Rays Diffractions (XRD) runs. The corresponding trajectory files produced a total of 273,000 frames which were strided and gathered to obtain a final file containing 9,117 frames.

Samples analysis

- *RMSD maps*

The next step in our approach was to visualize the conformational changes occurring within each sample using RMSD maps.

In order to get a preliminary idea about the conformational space of c-Met during the LMD simulation, a RMSD map is presented on **figure 4**. This map was obtained using a RMSD threshold between different conformations greatest than or equal to 4Å in order to give a clearer representation. This RMSD map shows that the conformational space of the *apo* model as explored by the long 1µs MD trajectory can be roughly grouped into two main conformational families. From each family, several representative conformations were then chosen according to the differences of their RMSDs.

From a preliminary RMSD analysis of the NM_MD sample runs, each mode was first considered separately and next the three modes were gathered to avoid cluster duplicate when going from one mode to the other. The final clustering scheme for each mode and for the three modes gathered are shown in **figure $5_{a,b}$** respectively. Comparing the representative conformations extracted from the global NM_MD, the sample RMSD map shows that here the explored conformational flexibility of c-Met is much more limited compared to the one obtained from the LMD simulation.

From the *holo* structures sampling, according to RMSD map results, 2 large families are also distinguished and the final clustering scheme for the 26 XRD structures is represented in **figure 6**.

- *MDS plots*

In order to visualize the distribution of the conformations in 3D space, **figure 3$_{a-c}$** presents the multi-dimensional scaling (MDS) plots obtained from each sampling. It appears from these figures that the conformational space covered by each method presents a common region while supplementary extended areas, typical of each sampling method, are also found. Therefore, some regions are specific to the XRD sampling, while another dense area is specific of the LMD sampling and some conformations are related to the NM_MD sampling only.

- *Wordom Cluster analysis*

In the case of the LMD simulation of the *apo* protein, the Wordom clustering also produced two main families, in accordance with the RMSD map representation above. The hierarchical method also gave 2 major families for the NM_MD sample. In the case of the XRD *holo* sampling, 6 families were found by this software and for each resulting family a representative conformation was selected.

- *k-means clustering analysis*

We analysed the different trajectories using k-means clustering for a range of cluster radii. The results of the clustering for radii values of 2-5Å are shown in **Table II$_a$**. For small radii values (~2-2.5Å) a large number of clusters is produced in all the three ensembles, and gradually declines as the threshold increases. We retained the partition obtained for a 3 Å threshold which produced 4 clusters for the LMD sampling to be compared to the 14 and 13 clusters found for the NM_MD and RXD samplings, respectively. This behaviour shows that for the same threshold, three times as many clusters are found for the NM_MD and RXD samplings, suggesting a larger deviation between the conformations within these ensembles.

We further extended the analysis by comparing the quality of the clusters obtained by other techniques based on agglomerative and partitioned algorithms. For this, the number of clusters identified by the *k*-means approach for different clustering radii was given as inputs to the three clustering schemes. The results are presented in table **II$_{c-d}$**. From these results, in order to select the most appropriate number of clusters, we retained the partitions showing a cluster average radius around 3 Å. This choice was motivated by the compactness of the clouds of conformations within the clusters. With such a selection criteria, 2 or 4 clusters were retained for the LMD sampling, 6 or 8 for the NM_MD one and 6 or 13 for the RXD. After comparing the different proposed partitions, it appeared that the most well-separated and compact ones corresponded to 4, 6 and 13 clusters for the LMD, NM_MD and RXD samplings respectively. From the above analysis, we initially retained a set of 23 c-Met conformers representative of these 23 identified clusters.

Representative c-Met conformations retained for virtual screening

Analysing the above results with the partitions suggested from the RMSD maps, the MDS plots and of the Wordom clusterings led us to consider the value of using a consensus between all these different clustering methods. We decided therefore to retain only 10 conformers (see **Table III**) to be representative of the c-Met conformational space, namely 5 *apo* structures (2 from LMD and 3 from NM_MD) and 5 *holo* ones (from XRD). For instance, conformation number 1,069 extracted among the 8,702 frames of the LMD sample clearly appeared as a pertinent representative (labeled LMD1). This conformation was found in 3 of the 5 clustering methods applied, and the selection of this conformation was also justified by its RMSD difference with the other selected conformations. A similar pattern was seen for conformations number 1,378 and 8,737 (labelled as WD1s_MD and RFN_MD, respectively) that belong to the XRD *holo* structure sample containg 9,117 frames. Another XRD *holo* structure conformation (number 22, labeled DKC_MD) appeared also to be a relevant representative conformer. Despite the fact that this conformation appeared in only one of the 5 applied clustering methods, its RMSD comparison with the other conformations shows a high and significant difference giving this conformation a place in the retained conformation list. On the contrary, there were some common conformations to 3 on the 5 clustering methods applied that show insignificant RMSD values after the cross comparison: that was the case, for example, for conformation number 1,227 belonging to the NM set which was not retained.

Those conformers that were finally retained are presented for comparison on MDS plots in **figure 3$_c$**. Their RMSD differences are graphically depicted as a spider plot in **figure 7$_a$**. To illustrate the hierarchy between these parent conformers, a dendrogram showing their linkage is presented in **figure 7$_b$**. From this dendrogram it appears that our *apo* LMD and NM representative conformers are linked to most of the *holo* XRD conformations, with the exception of one LMD conformer (LMD2) and one NM (NM1_MD) which presents in the dendrogram the largest link with all the others.

In order to check the representativity of our selection, we also analysed a recently available X-ray structure (**3Q6W**) [28] for an active form of c-Met which is considered by the authors as different from all the other X-ray structures already published. It already appears from the dendrogram of **figure 7$_b$** that this new active form of c-Met is linked to the LMD2 representative and could therefore be described as an induced-fit form resulting from the uncomplexed NM1_MD conformer.

From the analysis of the structural elements usually considered in kinase (C-helix, DFG motif, P- and A-loops, hinge region: see **figure 8**) and of the pieces of secondary structures, it appears that weak distortions are found, at least at the individual level and that most of these pieces

are fairly stable as individual elements. These results are summarized in **figure 9**, and can be presented according to different elements of c-Met secondary structure.

As expected with the most conserved region in kinases catalytic site, the hinge region is very stable, with RMSD variations ranging from 0.8 to 1.3 Å.

The C-helix is always quite stable in the 10 representatives. The exceptions being RFS_MD and CTH_MD were significant distortions within the helix are observed.

All our representative conformers present the typically observed DFG-in and DFG-out backbone conformations. This motif is weakly distorted for the LMD2 and RFN_MD conformers.

No major conformational changes are observed for the P-loop with the exception of RFN_MD loop conformer which appears as uncommon with the position of Arg1086 and His1088 side chains diametrically opposed relative to their counterparts from other representative conformers.

A large number of A-loop conformations is obtained, showing the high flexibility of this essential moiety for c-Met action, with RMSD ranging from 6.1 to 9.7 Å. Each A-loop conformation is different between the representatives. The A-loop conformation observed for the newly resolved **3Q6W** X-ray structure is quite similar to the one obtained for the NM2 representative.

The pieces of secondary structures are also quite stable as presented below.

For all residues constituting the β-sheet in the N-terminal domain, the RMSD varies between 1.9 to 3.8 Å. But this RMSD is reduced to only values ranging from 0.7 to 1.3Å if we consider only the pieces (defined in **table IV$_a$**) of the β_1 to β_5 sheets that are considered as a common stable motif in all kinases. It is the RFN_MD representative which aligns the least with the other ones: residues Asn_{1081} to Ile_{1084} are pointing in an almost perpendicular direction relative to their counterparts from others representatives. This deformation should be due to the influence of the ligand on this part of the sheets close to the ATP binding site. In this case the last sheet consisting of residues Val_{1218}, Lys_{1219} and Val_{1220} is shifted and do not align on the corresponding residues in all other representatives.

The RMSD values for the helical bundle within the C-terminal domain range from 1.6 to 3.0 Å. If we consider only the α_E, α_F, α_H and α_I helices (see **Table IV$_a$** for definition), the RMSD values do not exceed 1.5 Å. Only the third small α_G helix (Leu_{1255}-Thr_{1257}) from the LMD1 and 2 representatives is not well aligned with others.

In order to analyse the relative inter-domains and intra-domains movements within the 10 conformations chosen as representative of c-Met conformational space, we used the position of the two most stable pieces of secondary structures, namely pieces of the β_1-β_5 sheets belonging

to the N-terminal domain and the α_E, α_F, α_H, α_I four helix-bundle of C-terminal domain (see **table IV$_a$** for definition), as the reference to check the relative domain-domain displacements as well as the ones of others structural moieties. For that, all our 10 representative conformers were aligned according to the β_1-β_5 sheets of the N-terminal domain used as a reference system. Then, the relative positions of the other structural pieces were calculated using the inertia tensor of each of them. From this analysis, it appears (**Table IV$_b$**) that there are quite large variations in the relative position of the N- and C-terminal domains with a combined rotation/torsion movement between the two domains considered as individual rigid bodies. **Figure 9$_a$** presents examples of such movements. If we now consider intra-domain movements, the most significant one if found in the N-terminal domain and concerns the C-helix position relative to the β_1-β_5 sheets (**figure 9$_b$**). Two groups of C-helix position are clearly identified and correspond to the in- and out- movement of this helix reported previously in some papers [29]. Here again, if we plot the C-helix position found in the most recently available PDB structures **3Q6W** and also in the **3RHK** [30], we can see that they both fit very well within our representative conformational space. This is particularly remarkable for the **3RHK** structure as the authors said that "the structure of ARQ-197 in complex with the c-Met kinase domain shows that the inhibitor binds a conformation that is distinct from published kinase structures".

DISCUSSION-CONCLUSION

Because of the complexity of its function and considering the large body of data available, c-Met kinase represents a typical test case for establishing a possible strategy to handle protein flexibility by using an ensemble of representative protein conformations. To answer one of the key questions concerning the quality of the ensemble sampling emerging from our study, we have checked the links between the latest *holo* XRD structures recently available (**3Q6W**, **2WKM**) and which were not considered in our sampling and our 10 representative conformers. As shown in the dendrogram of **figure 7$_b$**, the new *holo* 3Q6W X-ray, which has been considered by its authors as an unconventional conformation [28], is nicely connected to one of our *apo* representatives, namely NM1_MD, and consequently fits into our classification. The new **2WKM** conformation shows no new specificity and this conformation can be assimilated to the CTH_MD representative of the XRD *holo* sampling. These results suggest that our sampling and our choice for selecting representative conformations is a valuable one.

Most of our representative structures coming from the uncomplexed c-Met *apo* conformations are linked to a complexed *holo* one, namely LMD1 with RFS_MD and WD1s_MD,

NM2_MD with CTH_MD, NM3_MD with DKC_MD, NM1_MD with Q6W. This could be related to a possible conformational selection process in which a given ligand would stabilize a particular equilibrium state of the protein conformational landscape. Next, this could be followed by an induced fit mechanism occurring within c-Met after ligand binding. Only one of our *apo* representatives from the long MD simulation (namely LMD2) is not yet associated to an existing X-Ray *holo* structure, suggesting a new c-Met binding mode might exist. Such a conformational selection mechanism which could drive the ligand binding preferences toward a particular protein conformation has already been considered by molecular dynamics simulations as a valuable anchoring mechanism for several tyrosine kinases [31]. The hypothesis that a conformational equilibrium exists within the kinases is reinforced by several experiments (for example Voggtherr [32] demonstrates using NMR experiments that an equilibrium between the DFG-in and DFG-out must be considered to understand kinases behaviour). In fact to dissect the binding events between a conformational selection and induced fit mechanism is a matter of debate [22,33-35]. According to our results, we would propose that both events co-exist and that the binding of a particularly adapted ligand would displace the protein conformational equilibrium toward a preferred conformation which could be next modified by an induced-fit mechanism. Such a hierarchy of events has already been considered by Silva et al. [36].

On the methodology side, it appears that the association of normal modes selection followed by short all-atom solvated MD simulation of the protein presents several advantages in term of quality of the sampling versus the CPU time needed.

Concerning clustering of the samples, a useful starting point in the analysis is the calculation of RMSD maps representing all-by-all pairwise deviations between specified frames. They not only provide a convenient visualization tool for extracting information in a trajectory but also provide further options for multidimensional data analysis. MDS methods focus on providing a low-dimensional mapping (a geometrical picture typically shown in 2 or 3 dimensions) of high-dimensional data. Fixing the dimension in MDS allows it produce a picture that resembles clusters on a plot. Both methods, however, are linear and hence cannot identify non-linear patterns. On the other hand, cluster analysis techniques group objects or points into small units whose members are more like each other. These methods also attempt to reduce the number of dimensions in which MD trajectories reside, but the output is influenced by the number of dimensions or clusters that are required to be chosen a priori. It seems clear that the construction of an exhaustive list of conformations passes by the establishment of a consensus between the various parameters of several used methods.

The next step of this study will be the use of our representative set of c-Met conformers in an ensemble docking study in order to find new putative ligands for blocking the proteins aberrant behavior in tumor cells (work in progress).

Materials and Methods

Preparation of the 3D starting structures

The list of all the X-ray c-Met kinase structures presently available is given in supplementary material. As most of the kinase activation loop residues as well as some other parts were missing in the X-ray structures, it was necessary to add these components to achieve a complete 3D structure. To address this issue, we used a superposition of all available c-Met structure, using, where possible, parts existing in some of the other structures to build those with missing ones. Residues that were found missing or mutated from the c-Met sequence were added in their wild-type form. Dimer structures were cut to retain only their chain A. After all the structures were processed as described above, hydrogen atoms were added according to the amino acid protonation states at pH = 7. The structures were then minimized using the Insight package from Accelrys [37]. It was verified that the minimization process did not disrupt the crystal structures and also did not modify the ligand placement relative to the binding site.

Molecular dynamics

All the MD simulations were carried out using as starting points the *apo* or *holo* c-Met X-ray structures prepared as described above. The systems were solvated with an 80$Å^3$ cubic box of TIP3P explicit water molecules, and Na^+ ions added on the edges to ensure electrostatic neutrality, giving a total of ~51,000 atoms per system. Missing parameters for the ligands were added using the CHARMM22 [38] forcefield. The dynamics were computed with NAMD [39] using the particle-mesh Ewald approach to treat long-range electrostatics. The simulations were carried out in the NPT ensemble using Langevin dynamics and piston to fix temperature (300 K) and pressure (1 atm). Hydrogen-heavy atom bonds were constrained to their equilibrium value. Each system was first energy minimized (6,400 conjugate gradients steps) and then equilibrated (500 ps) prior to the recorded trajectories. All MD trajectory frames were recorded at 1 ps intervals.

Normal modes calculations

Normal mode analysis was obtained for the *apo* model using the NOMAD-REF web server [40] using default parameters and asking to the maximum calculated modes available on this server (106 modes). From the results obtained, we extracted the most relevant lowest modes and the corresponding protein movements were sampled into several conformations, each of them being next used as the starting point for a short MD simulation using the procedure described above.

Cluster analysis of the samples

An important aspect of analyzing molecular dynamics trajectories is to reveal the inherent relationships between the different conformational states present in the data. To this end, clustering techniques have been very popular for MD trajectory analysis [41-46], followed by other schemes based on principal component analysis (PCA). Because the MD conformational space partition obtained with clustering often depends on the algorithm used and on the choice of some parameters, the separation of data into clusters can often be artificial. We therefore applied different clustering techniques in order to extract a consensus partition for use in a virtual screening experiment. The methods include:

- *ClusteRMSD* is an in-house program that allows similar conformations to be grouped based on their crossed RMSDs. Maps built with these measures are often used to depict the similarity between conformational states obtained from MD trajectories [46].
- The program *Wordom* [47], is also based on RMSD comparisons and provides three different clustering schemes, namely the hierarchical, quality threshold, and lead like. After conducting several trials, we adopted the hierarchical algorithm which respects the order of the starting conformations order and also performs a more precise calculation compared with the others. However, because of the high precision calculations required by this algorithm, the number of conformations analysed is limited to 20,000.
- *k*-means clustering as implemented in the MMTSB software [48]: The method starts with a preselected number of *k* random clusters and iteratively reassigns members to different groups while maximizing the inter-cluster variability. Because the number of clusters is not always initially known, in many cases a clustering radius is usually applied that is geared towards identifying spherical clusters of a given value. In this study, we used a range of clustering radii (2Å-5Å) with increments of 0.5. The number of clusters identified for each radius was then used as an input to other clustering algorithms.

We further extended the analysis by comparing the quality of the obtained clusters by other techniques based on agglomerative and partition-based algorithms as implemented in the software package CLUTO [49]. For this, the different number of clusters identified by the k-means approach for different clustering radii was provided as inputs to the three clustering schemes discussed below:

- Clustering based on repeated bisections (further referred to as Rb): The method proceeds by iteratively bisecting the data until the desired number of clusters is obtained. Each bisection produces two clusters that are chosen by optimizing a clustering criterion function that maximizes the similarity between each data point and the centroid of the cluster that it is

assigned to.

- Direct clustering (further referred to as Direct): This approach follows a similar scheme to that of *k*-means by choosing as set of *k* seeds as cluster centres which are repeatedly refined, subject to a clustering criterion function.
- Biased Agglomerative clustering (further referred to as Bagglo): This scheme uses a partitioned clustering approach to bias the agglomeration process.

High Dimensional Data visualization

To visualize the conformational spaces for the trajectories produced from our three different sources, we plotted the data points from all recorded MD frames in 3D space. For this, we used multidimensional scaling (MDS) because this statistical technique provides a geometric (visual) representation of the underlying relations (similarities/distances) between the data points. MDS projects a set of *n*-dimensional vectors into a *p*-dimensional space such that the greater the dissimilarity between two points, the further apart they are in the map or space. In this study, the MDS was performed using R [50].

Computational Details

All the calculations were conducted on a local 128-node cluster composed of 16 Infiniband-linked units, each composed of two Intel Xeon L5320 1.86 GHz quad-core processors and 8 GB memory. The pictures were produced using the VMD molecular graphics software [51].

Acknowledgements

Y. Asses was funded by grants from the Institut National du Cancer (INCa), the Ligue Nationale Contre le Cancer (LNCC) and the Fondation pour la Recherche Médicale (FRM). We thank Openeye and Chemaxon for allowing free access to their software according to an academic license. This work was supported by Region Lorraine within the framework of the PRST MISN (MBI project).

Bibliography

1. Mittag, T.; Kay, L. E.; Forman-Kay, J. D., Protein dynamics and conformational disorder in molecular recognition. *J Mol Recognit* **2010,** 23, 105-116.
2. Teilum, K.; Olsen, J. G.; Kragelund, B. B., Functional aspects of protein flexibility. *Cell Mol Life Sci* **2009,** 66, 2231-2247.
3. Ohkura, K., Exploring unique structures: flexibility is a significant factor in biological activity. *Biol Pharm Bull* **2007,** 30, 1025-1036.
4. Durrant, J. D.; de Oliveira, C. A.; McCammon, J. A., Including receptor flexibility and induced fit effects into the design of MMP-2 inhibitors. *J Mol Recognit* **2010,** 23, 173-182.
5. Mishra, N.; Basu, A.; Jayaprakash, V.; Sharon, A.; Basu, M.; Patnaik, K. K., Structure based virtual screening of GSK-3beta: importance of protein flexibility and induced fit. *Bioorg Med Chem Lett* **2009,** 19, 5582-5585.
6. Miller, M., The importance of being flexible: the case of basic region leucine zipper transcriptional regulators. *Curr Protein Pept Sci* **2009,** 10, 244-269.
7. Wong, C. F.; McCammon, J. A., Protein flexibility and computer-aided drug design. *Annu Rev Pharmacol Toxicol* **2003,** 43, 31-45.
8. Teague, S. J., Implications of protein flexibility for drug discovery. *Nat Rev Drug Discov* **2003,** 2, 527-541.
9. Carlson, H. A., Protein flexibility is an important component of structure-based drug discovery. *Curr Pharm Des* **2002,** 8, 1571-1578.
10. B-Rao, C.; Subramanian, J.; Sharma, S. D., Managing protein flexibility in docking and its applications. *Drug Discov Today* **2009,** 14, 394-400.
11. Fischer, B.; Merlitz, H.; Wenzel, W., Receptor flexibility for large-scale in silico ligand screens: chances and challenges. *Methods Mol Biol* **2008,** 443, 353-564.
12. Alonso, H.; Bliznyuk, A. A.; Gready, J. E., Combining docking and molecular dynamic simulations in drug design. *Med Res Rev* **2006,** 26, 531-568.
13. Armen, R. S.; Chen, J.; Brooks, C. L., An Evaluation of Explicit Receptor Flexibility in Molecular Docking Using Molecular Dynamics and Torsion Angle Molecular Dynamics. *J Chem Theory Comput* **2009,** 5, 2909-2923.
14. Shaw, D.E.; Dror, R.O.; Salmon, J.K.: Grossman, J.P.; Mackenzie, K.M.: Bank, J.A.; Young, C.; Deneroff, M.M.; Batson, B.; Bowers, K.J.; Chow, E.; Eastwood, M.P.; Ierardi, D.J.; Klepeis, J.L.; Kuskin, J.S.; Larson, R.H. Kresten; L.L.; Maragakis, P.; Moraes, M.A.; Piana, S.; Shan, Y.; Towles, B. Millisecond-scale molecular dynamics simulations on Anton. Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis 2009, 1-11

15. May, A.; Zacharias, M., Protein-ligand docking accounting for receptor side chain and global flexibility in normal modes: evaluation on kinase inhibitor cross docking. *J Med Chem* **2008,** 51, 3499-3506.
16. Cavasotto, C. N.; Kovacs, J. A.; Abagyan, R. A., Representing receptor flexibility in ligand docking through relevant normal modes. *J Am Chem Soc* **2005,** 127, 9632-960.
17. Heath, A. P.; Kavraki, L. E.; Clementi, C., From coarse-grain to all-atom: toward multiscale analysis of protein landscapes. *Proteins* **2007,** 68, 646-661.
18. Craig, I. R.; Essex, J. W.; Spiegel, K., Ensemble docking into multiple crystallographically derived protein structures: an evaluation based on the statistical analysis of enrichments. *J Chem Inf Model* **2010,** 50, 511-524.
19. Paulsen, J. L.; Anderson, A. C., Scoring ensembles of docked protein:ligand interactions for virtual lead optimization. *J Chem Inf Model* **2009,** 49, 2813-2189.
20. Bolstad, E. S.; Anderson, A. C., In pursuit of virtual lead optimization: pruning ensembles of receptor structures for increased efficiency and accuracy during docking. *Proteins* **2009,** 75, 62-74.
21. Totrov, M.; Abagyan, R., Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr Opin Struct Biol* **2008,** 18, 178-184.
22. Huang, S. Y.; Zou, X., Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking. *Proteins* **2007,** 66, 399-421.
23. Sperandio, O.; Mouawad, L.; Pinto, E.; Villoutreix, B. O.; Perahia, D.; Miteva, M. A., How to choose relevant multiple receptor conformations for virtual screening: a test case of Cdk2 and normal mode analysis. *Eur Biophys J* **2010,** 39, 1365-1372.
24. Yap, T. A.; de Bono, J. S., Targeting the HGF/c-Met axis: state of play. *Mol Cancer Ther* **2010,** 9, 1077-1079.
25. Liu, X.; Newton, R. C.; Scherle, P. A., Developing c-Met pathway inhibitors for cancer therapy: progress and challenges. *Trends Mol Med* **2010,** 16, 37-45.
26. Cañadas, I.; Rojo, F.; Arumi-Uria, M.; Rovira, A.; Albanell, J.; Arriola, E., C--Met as a new therapeutic target for the development of novel anticancer drugs. *Clin Transl Oncol* **2010,** 12, 253-260
27. Asses, Y.; Leroux, V.; Tairi-Kellou, S.; Dono, R.; Maina, F.; Maigret, B., Analysis of c-Met kinase domain complexes: a new specific catalytic site receptor model for defining binding modes of ATP-competitive ligands. *Chem Biol Drug Des* **2009,** 74, 560-570.
28. Rickert, K. W., Patel, S. B., Allison, T. J., Byrne, N. J., Darke, P. L., Ford, R.E., Guerin, D.J., Hall, D. L., Kornienko, M., Lu, J., Munshi, S K., Reid, J C., Shipman, J. M., Stanton, E. F., Wilson, K. J., Young, J. R., Soisson, S. M., Lumb, K.J., Structural basis selective small-

molecule kinase inhibition of activated c-Met, *J. Biol.Chem* **2011**, 286, 11218-11225.

29. Bellon, S.F.; Kaplan-Lefko, P.; Yang, Y.; Zhang, Y.; Moriguchi, J.; Rex, K.; Johnson, C.W.; Rose, P.E.; Long, A.M.; O'Connor, A.B.; Gu, Y.; Coxon, A.; Kim, T.S.; Tasker, A.; Burgess, T.L.; Dussault, I. c-Met inhibitors with novel binding mode show activity against several hereditary papillary renal cell carcinoma-related mutations. J Biol Chem. 2008, 283, 2675-2683.
30. Eathiraj, S.; Palma, R.; Volckova, E.; Hirschi, M.; France, D.S.; Ashwell, M.A.; Chan, T.C. Discovery of a Novel Mode of Protein Kinase Inhibition Characterized by the Mechanism of Inhibition of Human Mesenchymal-epithelial Transition Factor (c-Met) Protein Autophosphorylation by ARQ-197. J Biol Chem. 2011, 286, 20666-20676
31. Aleksandronov, A.; Simonson, T., Molecular dynamics simulations show that conformational selection governs the binding preferences of imatinib for several tyrosine kinases. *J Biol Sci* **2010,** 285, 13807-13815
32. Vogtherr, M.; Saxena, K.. Hoelder, S.; Grimme, S.; Betz, M.; Schieborr, U.; Pescatore, B.; Robin, M.; Delarbre, L.; Langer, T.; Wendt, K.U.; Schwalbe, H. NMR characterization of kinase p38 dynamics in free and ligand-bound forms. Angew Chem Int Ed Engl. 2006, 45, 993-997.
33. Csermely, P.; Palotai, R.; Nussinov, R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem Sci*. **2010**, 35, 539-546
34. Zhou, H.X. From induced fit to conformational selection: a continuum of binding mechanism controlled by the timescale of conformational transitions. *Biophys J*. **2010**: 98, L15-7
35. Hammes, G.G.; Chang, Y.C. ; Oas, T.G. Conformational selection or induced fit: a flux description of reaction mechanism. *Proc Natl Acad Sci U S A*. **2009**, 106, 13737-13741
36. Silva, D.A.; Bowman, G.R.; Sosa-Peinado, A.; Huang, X. A Role for Both Conformational Selection and Induced Fit in Ligand Binding by the LAO Protein. *PLoS Comput Biol*. 2011, in the press
37. http://accelrys.com/
38. Buck, M.; Bouguet-Bonnet, S.; Pastor, R. W.; MacKerell, A. D., Jr., Importance of the CMAP correction to the CHARMM22 protein force field: dynamics of hen lysozyme. *Biophys J* **2006,** 90, L36-38.
39. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable molecular dynamics with NAMD. *J Comput Chem* **2005,** 26, 1781-1802.
40. http://lorentz.dynstr.pasteur.fr/nomad-ref.php

41. Bouvier, G.; Evrard-Todeschi, N.; Girault, J. P.; Bertho, G., Automatic clustering of docking poses in virtual screening process using self-organizing map. *Bioinformatics* **2010,** 26, (1), 53-60.
42. Priestle, J. P., 3-D clustering: a tool for high throughput docking. *J Mol Model* **2009,** 15, 551-560.
43. Frickenhaus, S.; Kannan, S.; Zacharias, M., Efficient evaluation of sampling quality of molecular dynamics simulations by clustering of dihedral torsion angles and Sammon mapping. *J Comput Chem* **2009,** 30, 479-492.
44. Steindl, T. M.; Crump, C. E.; Hayden, F. G.; Langer, T., Pharmacophore modeling, docking, and principal component analysis based clustering: combined computer-assisted approaches to identify new inhibitors of the human rhinovirus coat protein. *J Med Chem* **2005,** 48, 6250-6260.
45. Kozakov, D.; Clodfelter, K. H.; Vajda, S.; Camacho, C. J., Optimal clustering for detecting near-native conformations in protein docking. *Biophys J* **2005,** 89, 867-875.
46. Laboulais, C.; Ouali, M.; Le Bret, M.; Gabarro-Arpa, J., Hamming distance geometry of a protein conformational space: application to the clustering of a 4-ns molecular dynamics trajectory of the HIV-1 integrase catalytic core. *Proteins* **2002,** 47, 169-179.
47. Seeber, M.; Cecchini, M.; Rao, F.; Settanni, G.; Caflisch, A., Wordom: a program for efficient analysis of molecular dynamics simulations. *Bioinformatics* **2007,** 23, 2625-2627.
48. Feig, M., J Karanicolas and C.L. Brooks III, MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology**,** *J Mol Graph Model* **2004**, 22 (5), 377-395.
49. Zhao, Y. and G. Karypis, Empirical and Theoretical Comparisons of Selected Criterion Functions for Document Clustering. *Machine Learning,* **2004**, 55, 311-331
50. www.r-project.org
51. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J Mol Graph* **1996,** 14, 33-8, 27-8.

Legends

Figure 1: General flowchart of the present sampling

Figure 2: Selection of the NM frames as starting points for short 10ns MD simulations. Example of mode 7 and frames at intervals 2,4,6,8 (red points).

Figure 3_a: Individual sample representations. The color coding (red for the 1µs apo MD simulation, blue for the normal mode MD simulation, and green dor the X-ray derived holo 10ns MD simulation) is further employed in other figures.
Figure 3_b: Global samples representation
Figure 3_c: Common frames to the 3 samples

Figure 4: RMSD Clustering results from the 1µs MD sampling. Preliminary selected conformations are circled in blue on the map.

Figure 5_a: RMSD Clustering results from each NM sampling during 40 ns MD simulation
Figure 5_b: RMSD Clustering results from the whole NM sampling. Preliminary selected conformations are circled in blue on the map.

Figure 6: RMSD Clustering results for 26 *holo* X-Ray structures. Preliminary selected conformations are circled in blue on the map.

Figure 7_a: Crossed RMSD of the 11 final representative conformations: The chart represents a RMSD comparison for each conformation toward the others. Every circle (ring) is a RMSD graduation beginning from the inner ring (which corresponds to 0Å) to the most external one (which corresponds to 9Å). Cross-sections are separated by 1Å.
Figure 7_b: Dendrogram for the final retained conformations

Figure 8: Structural elements usually considered in c-Met kinase. The overall c-Met structure is represented in transparent. The secondary structural elements are highlighted according to the assigned color code.

Figure 9: Superpositions of the 10 retained conformation (in grey) with **3Q6W** (in red) according to their secondary structures elements

Figure 10_a: Variation of the position of the C-terminal helix bundle relative to the β_1-β_5 sheets of the N-terminal domain. The colours are the ones used to identify the LMD (red), NM_MD (blue) and RX_MD (green) conformations.

Figure 10_b: Variation of the C-helix relative to the β_1-β_5 sheets. The position of the recently available **3Q6W** and **3RHK** PDB c-Met are shown.

Table I: Number of frames stored and retained in each sampling

Table II_a: K-means clustering results. The retained parameters are written in red. This colour is similarly employed in the II_b – II_d tables.

Table II_b: Clustering based on repeated bisections

Table II_c: Direct clustering

Table II_d: Biased Agglomerative clustering

Table III: List of retained conformers taken from different sources

Table IV: Domain/domain relative positions in our representatives

Figures

Figure 1

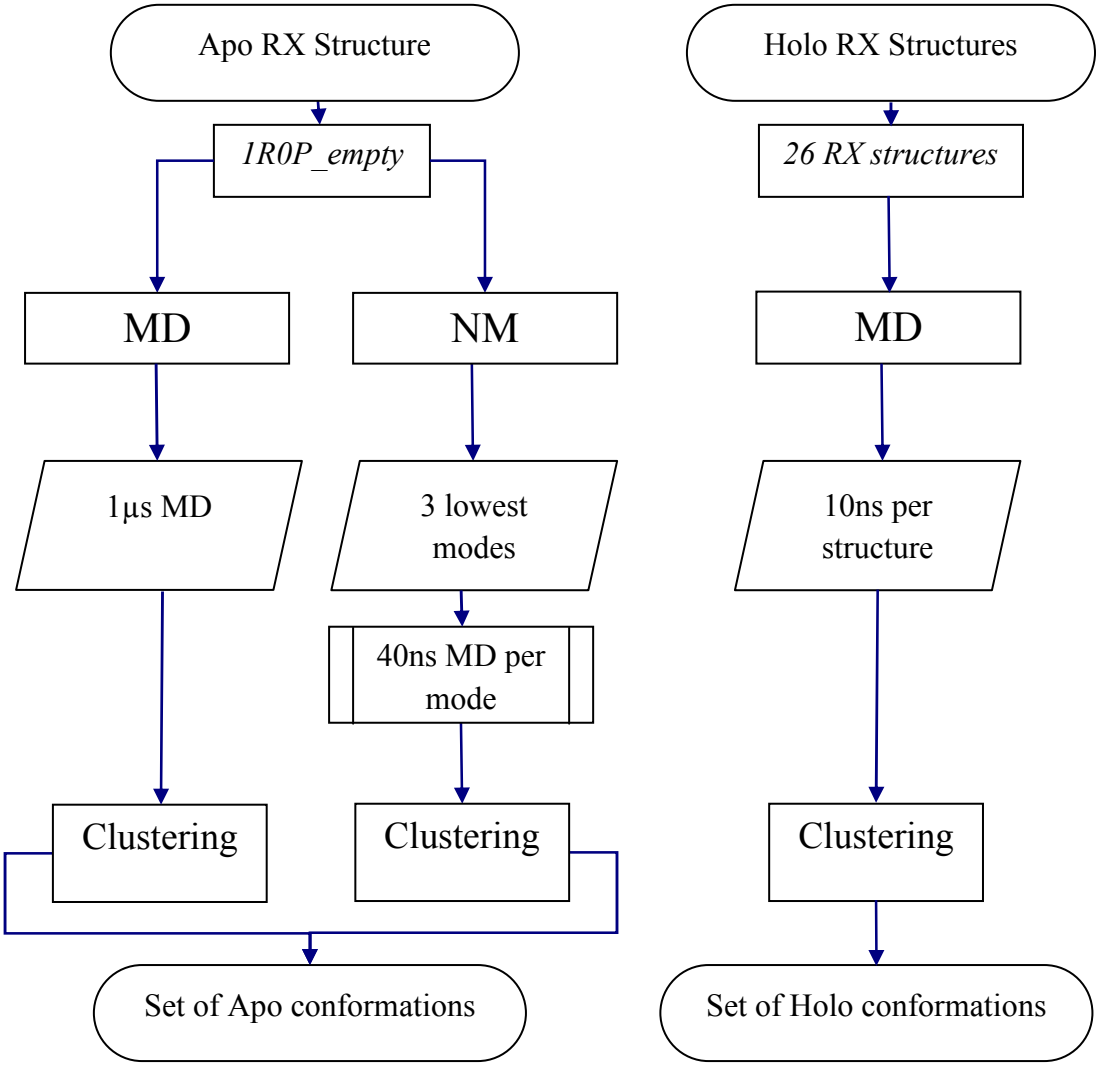


Figure 2

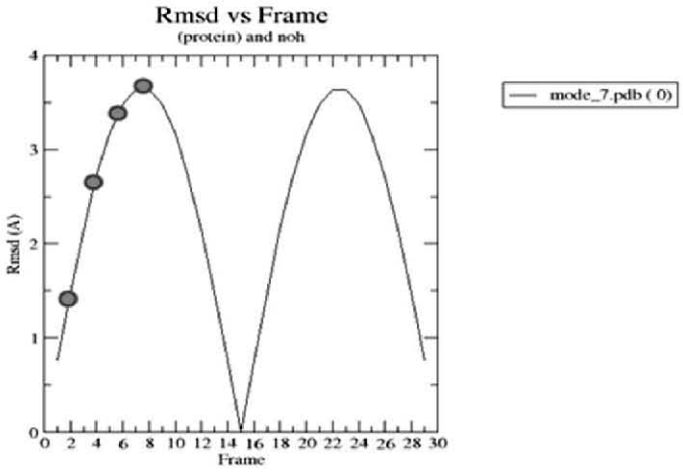

Rmsd vs Frame
(protein) and noh
Rmsd (A)
Frame
0
1
2
3
4
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30
mode_7.pdb (0)

Figure 3a

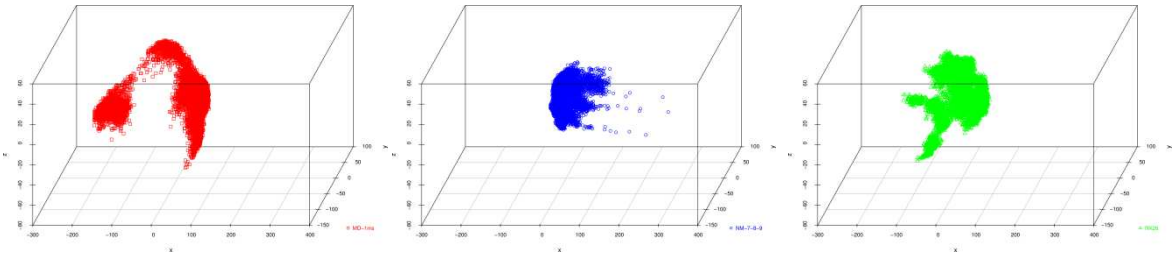


Figure 3b

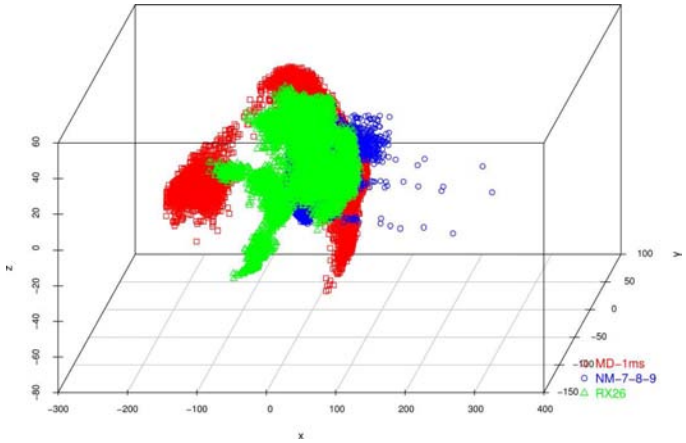


Figure 3c

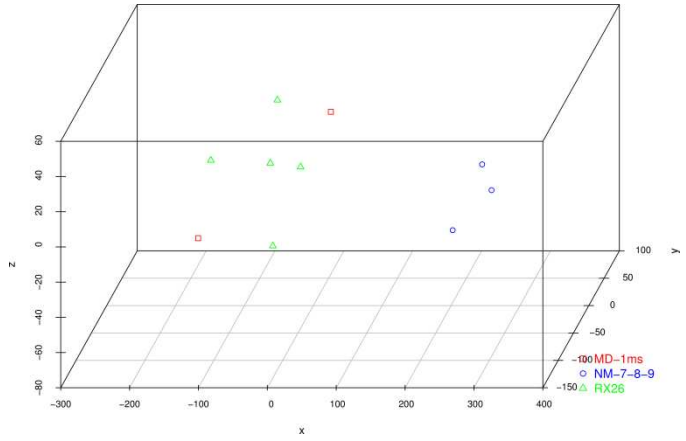


Figure 4

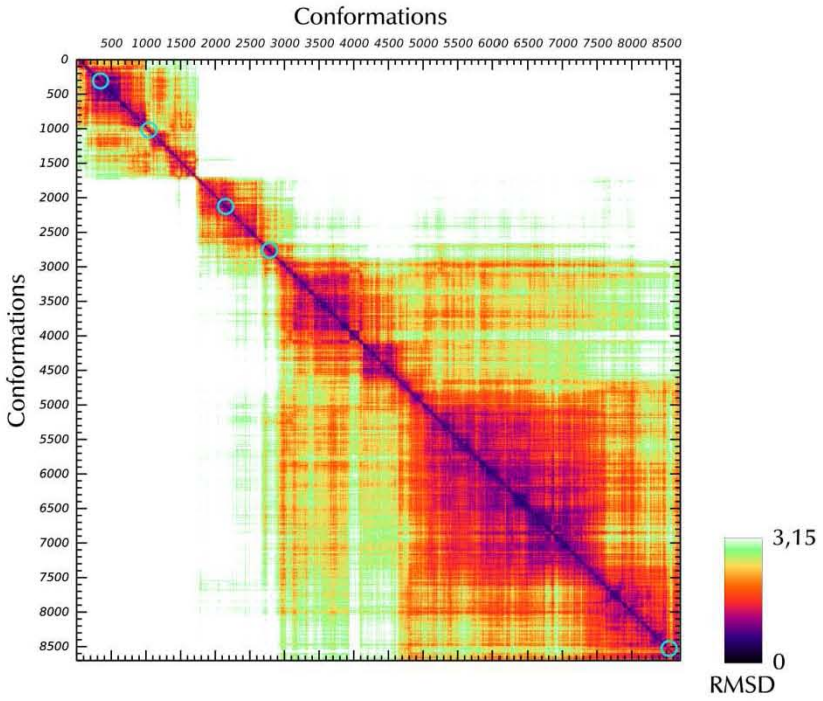
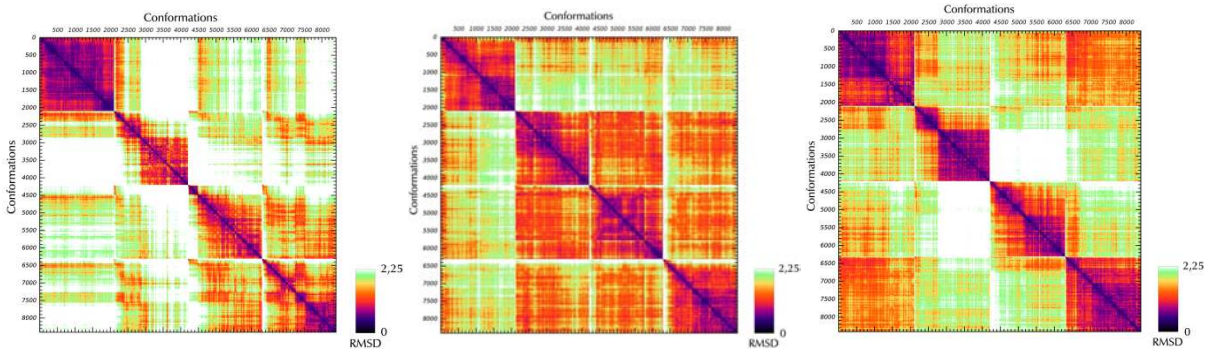
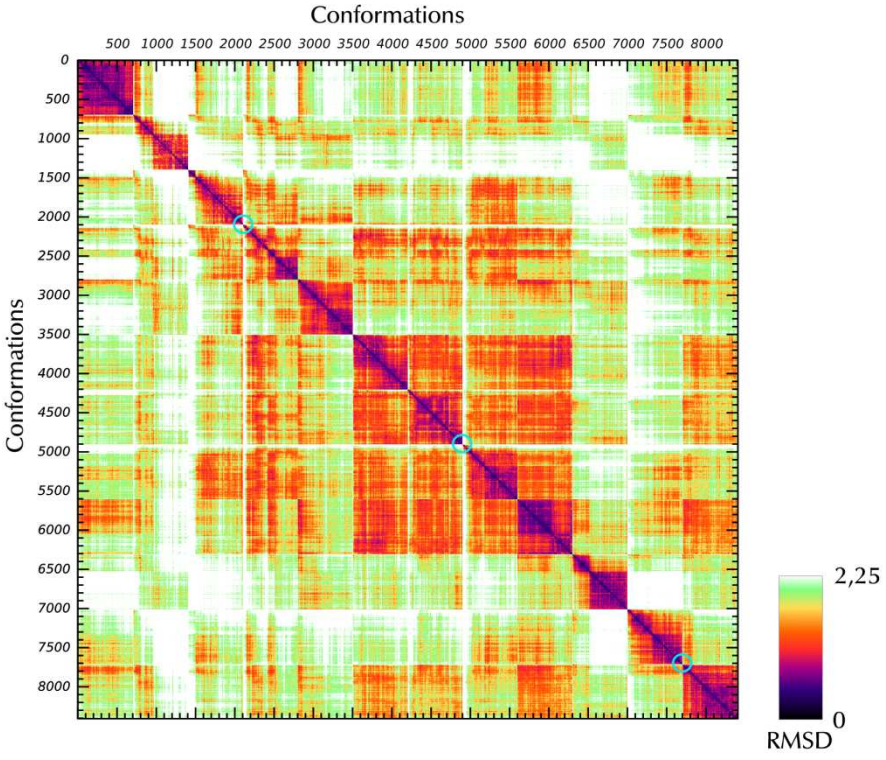

Conformations
500 1000 1500 2000 2500 3000 3500 4000 4500 5000 5500 6000 6500 7000 7500 8000 8500
Conformations
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
5500
6000
6500
7000
7500
8000
8500
3,15
0
RMSD

Figure 5a



Mode 7 Mode 8 Mode 9

Figure 5b



Mode 7 + 8 + 9

Figure 6

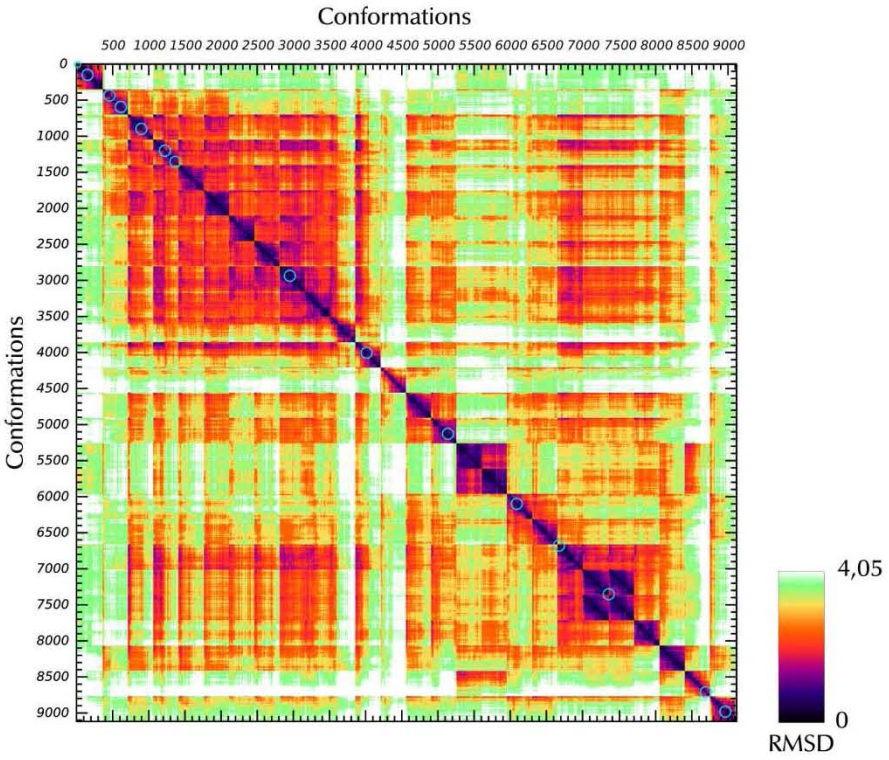

Conformations
500 1000 1500 2000 2500 3000 3500 4000 4500 5000 5500 6000 6500 7000 7500 8000 8500 9000
Conformations
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
5500
6000
6500
7000
7500
8000
8500
9000
4,05
0
RMSD

Figure 7a

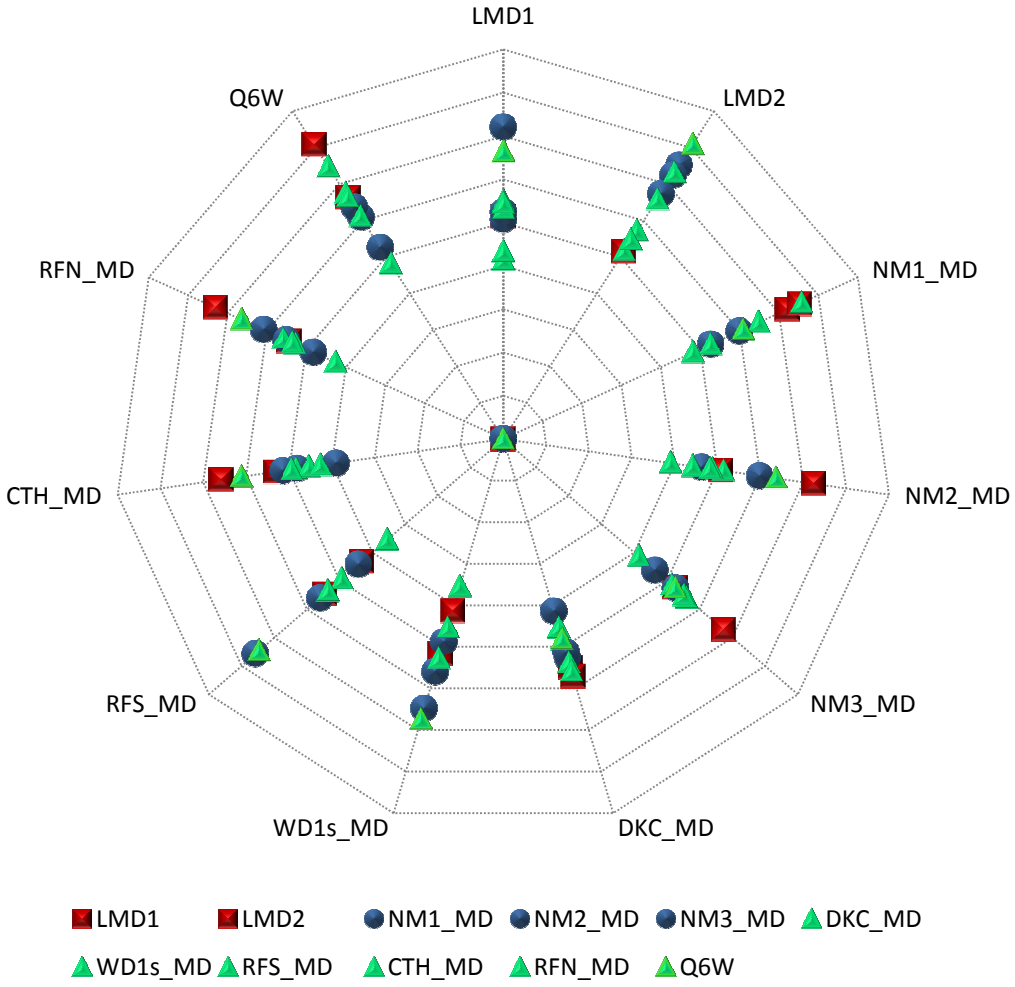


Figure 7b

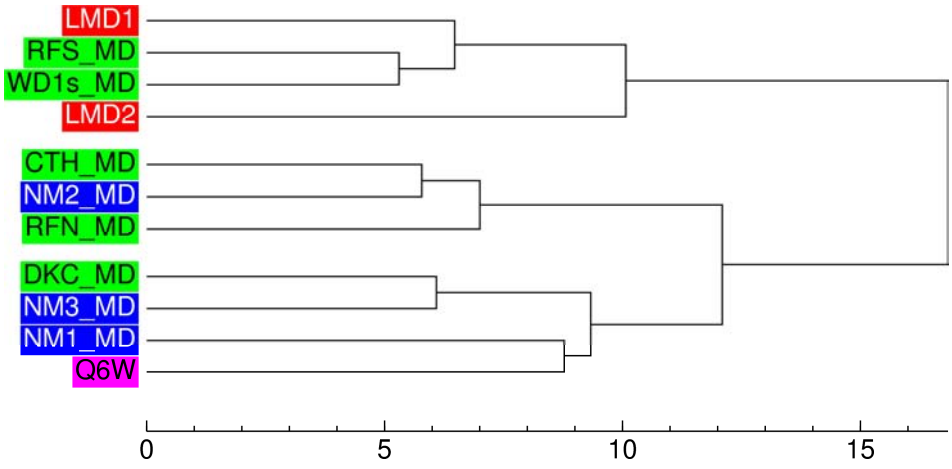


Figure 8

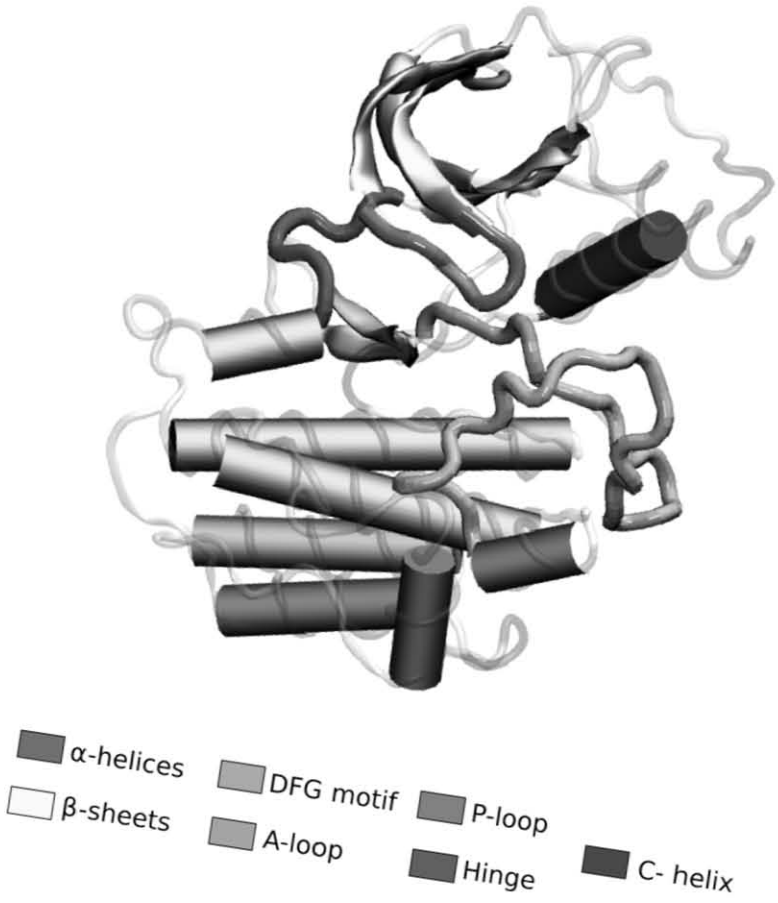


Figure 9

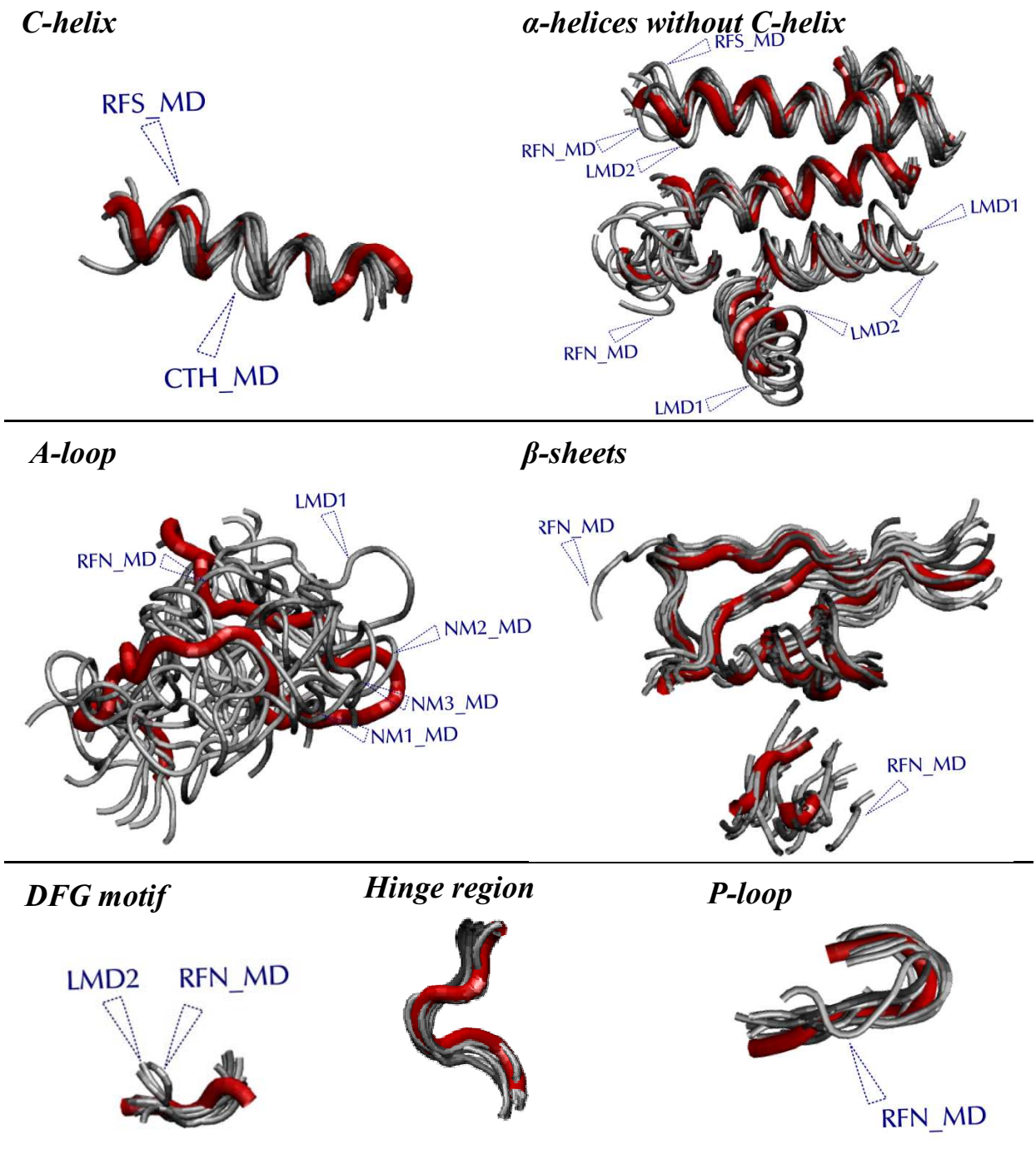
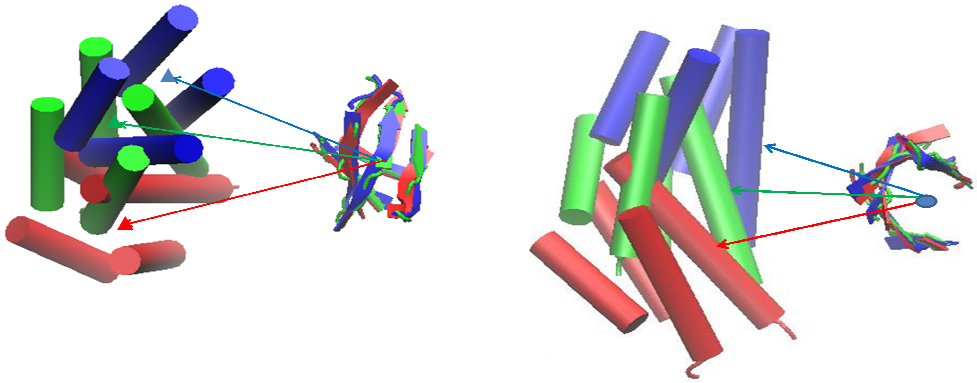
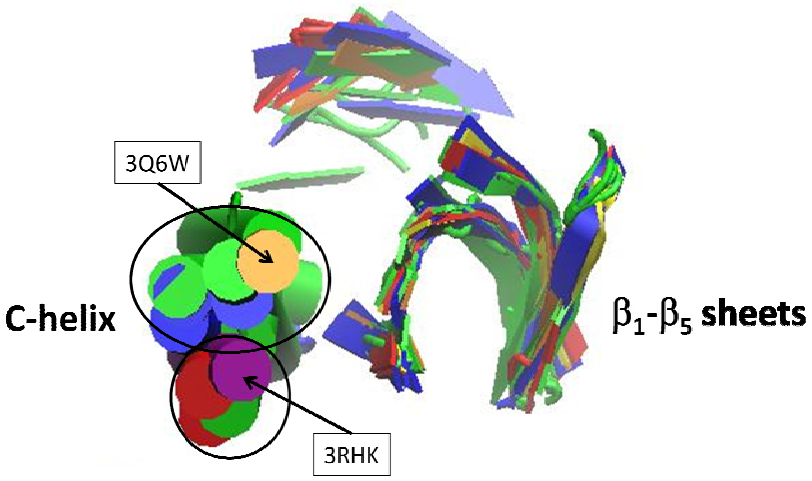

C-helix
RFS_MD
CTH_MD
α-helices without C-helix
RFS_MD
RFN_MD
LMD2
LMD1
LMD2
RFN_MD
LMD1
A-loop
LMD1
RFN_MD
NM2_MD
NM3_MD
NM1_MD
β-sheets
RFN_MD
RFN_MD
DFG motif
LMD2
RFN_MD
Hinge region
P-loop
RFN_MD

Figure 10_a



Torsion movement Rotation movement

Figure 10_b



Tables

Table I

	LMD	NM	RX
Initial number of frames	1 000 505	126 000	273 000
Final number of frames	8702	8406	9117
Trajectory $RMSD_{max}$	6,571	5,419	5,285
Number of retained frames	2	3	5

Table II_a

	R=2.0Å	R=2.5Å	R=3.0Å	R=3.5Å	R=4.0Å	R=4.5Å	R=5.0Å
MD1μs	29	14	4	3	2	1	1
NMA-7-8-9	21	17	14	13	8	6	4
RX-26	41	20	13	6	3	1	1

Table II_b

MD1μs		NMA-7-8-9		RX-26	
#clusters	Average Radius	#clusters	Average Radius	#clusters	Average Radius
29	2.30828	21	2.24381	41	2.52585
14	2.62357	17	2.38235	20	2.864
4	3.2925	14	2.57857	13	3.18
3	3.41667	13	2.62846	6	3.18333
2	3.985	8	2.88875	3	4.04333
-		6	3.18333	-	
-		4	3.7825	-	

Table II_c

MD1μs		NMA-7-8-9		RX-26	
#clusters	Average Radius	#clusters	Average Radius	#clusters	Average Radius
29	2.24241	21	2.11	41	2.19707
14	2.47071	17	2.26412	20	2.592
4	3.175	14	2.485	13	2.85846
3	3.2766	13	2.64462	6	3.69333
2	3.985	8	3.1425	3	3.92333
-		6	3.76333	-	
-		4	4.4	-	

Table II_d

MD1μs		NMA-7-8-9		RX-26	
#clusters	Average Radius	#clusters	Average Radius	#clusters	Average Radius
29	2.34	21	2.21	41	
14	2.56	17	2.36	20	
4	3.77	14	2.49	13	3.1
3	3.32	13	2.56	6	
2	3.2	8	3.1	3	
-	-	6	3.47	-	-
-	-	4	4.27	-	-

Table III

Source	Conformations N°	Method	Origin	Name
LMD	1069	Bagglo + Direct + Rb		LMD1
	2782	ClusteRMSD + Wordom		LMD2
NM	2103	Bagglo + Direct + Rb	Mode 7	NM1_MD
	4905	Bagglo + Direct + Rb	Mode 8	NM2_MD
	7007	Bagglo + Direct	Mode 9	NM3_MD
RX-26	138	ClusteRMSD + Wordom	DKC	DKC_MD
	1378	Bagglo + Direct	WD1solv	WD1s_MD
	3828	Bagglo + Direct	RFS	RFS_MD
	6155	ClusteRMSD + Wordom	CTH	CTH_MD
	8737	Bagglo + Direct	RFN	RFN_MD

Table IV$_a$

domain	Secondary structure name	Residues concerned
N-terminal	β_1	1076 to 1081
	β_2	1092 to 1098
	β_3	1104 to 1110
	β_4	1144 to 1146
	β_5	1154 to 1158
C-terminal	α_E	1178 to 1198
	α_F	1263 to 1278
	α_H	1310 to 1320
	α_I	1330 to 1343

Table IV$_b$

	helix bundle, β sheets		C-helix, β sheets	
	distance	**twist**	**distance**	**twist**
LMD1	35.5 Å	111°	18,9 Å	35°
LMD2	32.8	105	18,8	43
NM1_MD	31.0	169	18.2	45
NM2_MD	32.4	36	17,5	85
NM3_MD	31.6	29	18.6	44
CTH_MD	32.1	146	17,2	78
DKC_MD	30.3	123	18,0	44
RFN_MD	35.6	153	17,3	75
WD1S_MD	31.3	107	18,7	69
RFS_MD	34.0	73	17.2	88
Q6W	29.6	136	16.9	85

III- Commentaires

Afin de réaliser ce travail d'échantillonnage sur différentes trajectoires de simulations de DM, nous avons choisi l'approche de l'analyse en groupes ou clustering. L'analyse en question, ou le clustering, est basé sur la comparaison de RMSD (Root Mean Square Deviation). Il a été établit comme étant une métrique qui peut être utilisée afin de réduire un ensemble vers un groupe restreint mais significatif de conformations avant de procéder aux expériences de criblage virtuel [253, 254]. Le clustering, basé sur la comparaison de RMSD, est intéressant car il fournit des informations sur la conformation la plus dominante échantillonnée lors des simulations de DM. Sur la base de statistiques nous pouvons argumenter que les clusters les plus peuplés par des structures de récepteurs sont les plus rencontrés par les ligands, les plus stables énergétiquement et les plus significatifs de l'équilibre thermodynamique de la protéine.

Cette étude a montré clairement tout l'intérêt d'utiliser une combinaison des modes normaux et de courtes simulations de dynamique moléculaire. Ce travail nous a également permis de mettre en œuvre des techniques d'analyses variées, particulièrement des méthodes de clustering. Nous avons pu ainsi montrer qu'un consensus de méthodes était préférable à l'utilisation d'une seule technique incapable d'analyser dans sa globalité la complexité de l'hyperespace conformationnel exploré.

ARTICLE 4

I- Contexte

La notion d' « ensemble docking » est apparue en 2007 [255] dans un souci de considérer la flexibilité de la protéine cible dans les calculs de « docking » moléculaire. La flexibilité des ligands était, elle, traitée depuis quelques années dans les programmes de docking semi-flexible tels que GOLD [256] et AUTODOCK [257] qui restreignaient la prise en compte des modifications conformationnelles de la cible en se limitant à faire varier les positions de quelques chaînes latérales des résidus amino acides (flexibilité « locale »). La prise en compte d'une flexibilité « globale » de la protéine cible est au cœur des développements actuels des méthodes de docking (voir par exemple ces deux revues récentes [258, 259]). Certains auteurs préconisent d'aborder le problème en traitant directement la flexibilité des deux partenaires dans l'algorithme de docking [260] ; mais cela conduit à des problèmes d'optimisation traitant d'un très grand nombre de variables et à la difficulté de sortir ces algorithmes du « piège » des minima locaux.

Compte tenu de nos résultats antérieurs concernant la description des conformations représentatives de c-Met et de l'introduction d'une possibilité d' « ensemble docking » dans le logiciel GOLD que nous avions utilisé jusqu'alors dans nos campagnes de criblage virtuel, nous avons donc lancé une nouvelle campagne de criblage. Pour cela, nous avons considéré un ensemble de 45 conformations de c-Met (19 extraites de notre échantillonnage conformationnel de la kinase et issues de l'article précédent + les 26 structures cristallographiques de la PDB).

Les chimiothèques considérées ici ne comportent que des molécules disponibles. Nous avons ainsi utilisé la Chimiothèque Nationale [261], une chimiothèque issue de la faculté de pharmacie de Strasbourg [262], des chimiothèques spécialisées des companies Life Chemical [263], Timtec [264] et Chembridge [265]. En tout 100.000 molécules environ ont été testées.

Sur ces 100.000 molécules, une cinquantaine d'entre elles ont été choisies sur la base de leurs scores et envoyées à nos collègues biologistes pour procéder à des tests de validation.

II- Investigation of the ensemble docking strategy to improve docking relevance in virtual screening against the c-Met kinase

Investigation of the ensemble docking strategy to improve docking relevance in virtual screening against the c-Met kinase

Yasmine Asses [1*], Vincent Leroux [1,2*], Alessandro Furlan [3], Flavio Maina [3], Rosanna Dono [3], Bernard Maigret [1+].

1. Nancy Université, LORIA/UMR 7503, Équipe-projet Orpailleur, Campus Scientifique, BP 239, 54506 Vandœuvre-lès-Nancy Cedex, France
2. University of Oslo, Chemistry departement, P.O.Box 1033 Blindern, 0315 Oslo, Norway
3. IBDML (Developmental Biology Institute of Marseille-Luminy), Campus de Luminy - case 907, 13288 Marseille cedex 09, France

+: Corresponding author: bernard.maigret@loria.fr / +33 688 726 110.
* : These two authors contributed equally to this work

Keywords: virtual screening, ensemble docking, structure-based drug design, cross-docking validation, induced fit, conformational selection, c-Met kinase

Abstract

Using an ensemble docking strategy, a large scale virtual screening campaign was undertaken in order to identify new putative hits against the c-Met kinase target. A set of 45 conformers of the kinase, obtained previously from a large sampling of its conformational space, has been used as docking targets. About100, 000 chemical compounds were extracted from several chemical libraries to be tested *in silico* for their potential affinities towards the kinase binding site. The ensemble docking facility available in the last release of the GOLD.5 software was used. Thanks to the large scale computer facilities provided by the french Grid5000 network, a set of 1,000 possible hits was extracted from this virtual screening campaign. The top 25 molecules from this list were tested experimentally against the c-Met kinase in order to validate our approach and to give new insights concerning the design of innovative leads. This work re-inforces the hypothesis that, concerning c-Met kinase, induced fit effects should be modelled when simulating ligand binding and that the flexibility of both partners must be considered in the docking calculations.

INTRODUCTION

Reliably considering protein conformational flexibility in protein-ligand docking calculations is an unsolved challenge whose importance is widely acknowledged by computational drug design specialists [1-9]. This can be attempted using two different approaches: on-the-fly optimization of the receptor conformation while exploring ligand binding possibilities, using a model that considers the protein and ligand conformational spaces as one (fully-flexible docking) [10], or using established semi-flexible docking on a pre-defined conformer library of receptors (ensemble docking) [11-17]. One fundamental question is whether the ligand-bound conformation of a protein is enforced by the incoming ligand (induced fit) [18-21] or, on the contrary, if it is selected from a pre-existing set of favorable ligand-independent conformations (conformational selection) [22, 23]. Both fully-flexible docking and ensemble docking are currently being actively investigated, the former being in direct connection with the induced fit hypothesis, and the latter with the conformational selection one. Fully-flexible docking needs improved algorithms due to the large increase of variables (additional receptor rotamers) compared to the fixed protein scheme. Some recent programs may be up to this [24]. However, it also requires a complete re-thinking of scoring functions for accommodating energy fluctuations correlated to receptor flexibility, a daunting task considering scoring function relevance is the Achilles' heel of docking programs already [25-27]. Therefore, ensemble docking currently seems more accessible, as only little adaptations from well-established semi-flexible computational protocols have to be made. The conceptual difficulty of that approach appears prior to the calculation itself: how to generate a relevant ensemble for a given receptor? We have given an initial answer to this question in one of our recent works [28].

The c-Met kinase is a pharmaceutically-relevant target in anticancer drug design (for reviews on the matter, see [29, 30]). We already demonstrated [31] that it is particularly interesting regarding its plasticity upon ligand binding. Indeed, it can accommodate three distinct ligand binding modes and associated receptor conformations, a feature that is only rarely seen in the kinase family [32] and that may allow for designing drugs with improved efficiency and selectivity profiles [33, 34]. Here, we now provide the results of an investigation using a large-scale ensemble docking on c-Met.

METHOD

Screened chemical libraries

We have used several general and focused libraries (see list in Table I). These libraries were firstly filtered according to general drug-like definitions [35] so that around 70,000 compounds were finally retained for our screening campaign.

Our choice for the selected chemical libraries was motivated by trying:

- to use the highest possible chemical diversity (Chimiothèque Nationale and Life Chemical diversity sets) in order to cover a large spectrum of chemical structures [36].
- To include compounds likely to be active on a broad range of kinase targets already (ChemBridge kinase set) as a possible source of c-Met inhibitors. Such a choice has already been described as a valuable one [37].
- To explore some natural product library (TimTec natural product set) as natural compounds could be a promising source of anti cancer drug discovery [38].
- To take into account the repositioning of antiviral compounds (Life Chemicals antiviral set) as drug repurposing presents an increasing interest in cancer research [39] [40]. A typical example of such viral-to-cancer drug repositioning is gemcitabine with US patent 4,808,614, which claims the chemical compound and a method of using it for treating viral infections, and the later-issued US 5,464,826, which claims a method of using it to treat cancer. It might be postulated that as compounds in the drug shelves with chemical profiles significantly more diverse, and activities larger than average, antivirals are a promising source for repurposing strategies.
- to keep the possibility of chemical modifications of the found hits, if any, by explicitly including a set of molecules (JJB set) for which chemical modifications can be made proficiently.

Selected c-Met conformational ensemble

We have retained an extended set of 21 representative conformations obtained previously from a large sampling of c-Met conformational space [28] to which the 24 X-ray PDB [41] structures were added (c-Met complexes available on the PDB as of January 2011), provided a working conformational set of 45 c-Met conformers. As all used proteins must be superimposed before they are used in an ensemble docking calculation, all these

conformations were structurally aligned according to their conserved and most rigid secondary structure patterns as described previously [28]. The ensemble docking target data is constituted by the resulting set of aligned protein conformations in 45 PDB format files, plus the Cartesian coordinates of the shared active site center. The calculation of this position is described in the next section.

Description of the ensemble docking protocol

We have used ensemble docking facility recently proposed in the last GOLD5 release of the well-known GOLD docking program. The new ensemble docking GOLD feature evaluates different receptor conformations concurrently during docking exploration.

When docking an ensemble of conformations of a given protein, the binding site definition must describe as accurately as possible the binding sites of all loaded and superimposed conformers. The binding site must therefore be defined using a method that isn't conformer specific i.e. a point and a radius defining the sphere of the common binding region, a list of common residues or a common reference ligand. As it is not possible to define the active site using a list of atoms or residues in the present GOLD5 version and considering the c-Met ligand diversity we have chosen to define the active site from a point and a radius around this point. This choice is certainly the most sensible part of the present docking process as it governed the quality of the docking results. To define this point, we have used the LIGSITE program [42] which provides identification of protein cavities as well as their center of mass. For each of the 45 aligned conformers used here we have therefore calculated the center of mass of their individual binding site. From these data, an average center point was calculated to be used next as the binding site definition for the GOLD program. Figure 1 provides an example of the center of mass dispersion among some conformers and figure 2 presents the position of the average used center point within the 45 protein conformers. A radius of 20 Å was associated to this average binding point to define the binding cavity of each conformer. 20 Å is above the GOLD default parameter (10 Å) in order to encompass correctly the receptor for all the conformations in the ensemble, including conformational variations around the center (see Figure 1). We also verified that the resulting sphere was encompassing all groups of residues previously identified as potential interaction areas for c-Met ligands [31] (see Figure 2). This results in a loss of precision of the docking procedure, because the size of the modelled space is greatly raised compared to GOLD defaults – this is being compensated by the choice of a greater number of maximum docking runs per ligand compared to defaults parameters (50 instead of 10, see docking procedure validation section).

For each docking run, 50 starting poses/molecule were considered for the GOLD generic algorithm. We used the gold standard scoring function to rank the tested compounds.

Grid facilities

Because of the massive calculations needed (~70,000 molecules x 45 protein ensemble conformers x max. 50 poses/molecule), and considering the computing time to achieve only one run, we used the Grid5000 facility [43] providing the required computer power. We used around 1,300 cpus with 4GB RAM/core and infiniband connections for each run. The docking performances run around 300 docked ligands/ensemble/hour.

Biochemical assays

MDCK cells were grown in Dulbecco's modified Eagle's medium (DMEM) (Gibco BRL) containing 4mM L-glutamine and supplemented with 2 % (v/v) foetal bovine serum (Gibco BRL), 100 U/mL penicillin, and 100 µg/mL streptomycin. Cells were plated at a density of 1,000 cells per well in 24-well microplates and allowed settling overnight at 37°C in a humidified atmosphere of 5 % CO_2 prior to treatments. Human MCF10A breast cells, human MDA-MB231 breast cancer cells, human GTL-16 gastric carcinoma cells, human HepG2 hepatocellular carcinoma cells, human U-87 glioblastoma-astrocytoma cells, and human LLC Lewis lung carcinoma cells were grown in RPMI medium (Gibco BRL) containing 4mM L-glutamine and supplemented with 10 % (v/v) foetal bovine serum (Gibco BRL), 100 U/mL penicillin, and 100 µg/mL streptomycin and kept at 37°C in a humidified atmosphere of 5 % CO_2.

Primary embryonic cortical and hepatocyte cultures and survival assays were performed as previously described. For toxicity assays, primary cells were exposed to compound **7c** for 24 hours. Viability was assessed with the Cell Titer Glo Luminescent Assay (Promega), according to the manufacturer's instructions. Data on biological assays were performed in quadruplicate.

For scattering assays, MDCK cells were pre-incubated with compounds overnight at 0.1-100 µM concentrations at 37°C in a humidified atmosphere of 5 % CO_2, followed by a 24 h stimulation with 20ng/ml HGF (R&D Systems), as previously described. Cells were further incubated at 37°C in an atmosphere of 5 % CO_2 for 24-48 h, washed with phosphate-buffered saline (PBS; Gibco BRL), and fixed with 4 % PFA (Sigma). The quantification of scattering response was performed by counting the number of cells with scattered morphology in 30

independent colonies. The IC_{50} corresponds to the concentration of compounds leading to a 50 % inhibition of Met-triggered cell scattering. Active compounds were assayed to impair scattering response also on MCF10A cells following the same procedure. For survival assays, GTL-16 and HepG2 cells were cultured in serum-free media for 24 h prior to compound addition for 48 hrs. Viability was assessed with the Cell Titer Glo Luminescent Assay (Promega). For in vitro tumorigenesis, soft agar growth assays were performed using GTL-16 and HepG2 cells, as previously described. Data on biological assays are representative of three independent experiments performed in duplicate. For the Ba/F3 cell assay, a total of 2.10^4 cells/well/50 µl were seeded in a 96-wells plate. Treatment was initiated by addition of a 2X drug solution of ½ serial dilutions ranging from 0 to 10 µM. Cells were grown for 48 h at 37°C and then incubated with 10 µl/well of Cell Titer Blue (G8080; Promega). Proliferation was quantified by measuring the absorbance at 490 nm using a scanning multiwell spectrophotometer (MultiSkan MS, Thermo-LabSystems, France). A blank well without cells was used as a background control for the spectrophotometer and all assays were performed in triplicate.

After starvation, cells were pretreated overnight with inhibitors, then stimulated with HGF (20 µg/ml), and total extracts were analyzed as described. Antibodies used were anti-tubulin (Sigma), anti-phospho Tyr1234/1235-Met, anti-Akt, anti-phospho-Akt, anti-phospho-Gab1, anti phosphor-ErbBs, anti-phospho-PDGFR, anti-PDGFRs, anti-ERKs (Cell Signaling), anti-ErbBs, anti-Met (Santa Cruz).

RESULTS

Validation of the GOLD5 procedure

The first step of our campaign consisted in comparing the docking results for the 24 XRD structures using 3 binding site definitions (residue list, existing XRD ligand and center point respectively) and both target ensemble docking versus individual target docking. We present in table III an example of the results obtained for individual dockings using the three binding site definitions showing that the center point average definition as defined above and used latter for the ensemble docking screening campaign was a valuable choice: for each individual docking target, the three definitions provided the same rank for the XRD 2RFN/AM7 ligand and almost the same values for the docking scores. Moreover the poses of this ligand found using the binding site definitions, the average center point and the ligand XRD pose were almost the same and similar to the pose found in the crystal structures as exampled on figure

3. The pose associated with the best score using a residue list was opposite to the XRD one, suggesting that the residue list we used for this test was not exhaustive enough as this situation was found in 80% of the poses corresponding to the best scores obtained using this residue list definition.

The preliminary ensemble docking calculations concerned the targets set constituted by the 24 XRD PDB structures and a molecules set build from their 24 associated ligands. The question there concerned the possibility to retrieve in the targets ensemble the right target for a given ligand, with similar rank, score and pose compared to the ones found for the individual target dockings. We present in table IV the results for the ligands already presented in the previous table III. It is seen that only the KSA ligand is able to retrieve its original partner in the ensemble of the 24 PDB conformers of the protein target. The obtained scores are almost the same in the individual docking and in the ensemble docking, especially when considering a larger number of starting poses. A much slower convergence in the genetic algorithm could explain the difficulty of retrieving the right partner of a given ligand: increasing the number of tested poses and changing the number of steps in the optimization process would improve the results. Another point to be considered here is the limited possibilities of GOLD to handle entirely the protein target flexibility.

From these results it appears nevertheless that the ensemble docking procedure, using a center point to define the binding region, and with appropriate parameters is a valuable method to tackle multiple conformers docking.

Analysis of the docking campaign

After the protocol validation, the second step consisted in performing a virtual screening campaign with a set of 45 conformations and ligands derived from the sixth cited chemical libraries. The docking results analysis can be presented in two different ways:

Analysis per docked conformations:

We analyzed all the docking results first per screened conformations and presented it by histogram in figure 4.

The first observation is that molecules preferentially docked in the *holo* forms of the protein c-Met (here represented by the original XRD structures and XRD structures after MD simulations). The docked ligands show no affinity for the representatives of the *apo* form. The same statements can be made concerning the statistics made on the top 1000 docked ligands

on the set of conformations (see Figure 5). In addition, the two histograms show us the best potential targets (C1X, RFN and EFJ for the XRD structure after MD simulation and EFK, EFJ, CTJ, LQ8, CE3, F82, and CTH C1X).

Analysis per database

We also conducted an analysis of the various docked chemical library. For this we took the best 20 molecules from each docking run on each docked compound libraries. We divided the results as follows:

- For conformations extracted from the preliminary sampling (ref paper 3), we distinguish three groups: the conformations of the long MD, NM conformations of the conformations and structures from X-ray followed by an MD simulation.
- We have distinguished the original XRD structures by type of binding modes (paper ref 1). Another group that gathered unclassified structures is also expressed.

The resulting sectors clearly confirm the tendency of ligands to preferentially dock on the *holo* conformations. Indeed, for all the chemical libraries, there are an almost non-existent percentage of bonds between molecules and c-Met *apo* conformations, represented here by the conformations LMD and NM_MD.

Selection of the putative ligands

After docking calculations, Gold results were ranked to obtain different sets from screened libraries. The retained molecules are presented in the table:

To visualize the position of the best docked ligand, we have superposed the F0725-0356 called molecule with the original 3EFK ligand (MT4). The figure 7a shows that F0725 is docked in the same binding pocket as MT4, they also have the same general alignment despite their structural differences (see Figure 7b). Regarding the interactions between this selected molecule and the c-Met conformation 3EFK we performed calculations that give 2D schematic diagrams of protein-ligand interactions. A comparison between the XRD complex interaction (3EFK/mt4) and the screened complex interactions (3EFK/F025-0356) shows almost same interactions. Indeed, the most important residues known in c-Met interactions (Met1160, Asp1222, Tyr1159) are present in the docked complex (3EFK/MT4). (Figure 8)

In vitro tests

DISCUSSION

CONCLUSION

Bibliography:

1. Fuentes, G., et al., *Role of protein flexibility in the discovery of new drugs.* Drug development research, 2011. **72**: p. 26-35.
2. Cozzini, P., et al., *Target flexibility: an emerging consideration in drug discovery and design.* Journal of medicinal chemistry, 2008. **51**: p. 6237-6255.
3. Henzler-Wildman, K. and D. Kern, *Dynamic personalities of proteins.* Nature, 2007. **450**: p. 964-972.
4. Barril, X. and X. Fradera, *Incorporating protein flexibility into docking and structure-based drug design.* Expert opinion on drug discovery, 2006. **1**: p. 335-349.
5. Cavasotto, C.N. and R.A. Abagyan, *Protein flexibility in ligand docking and virtual screening to protein kinases.* Journal of molecular biology, 2004. **337**: p. 209-225.
6. Erickson, J.A., et al., *Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy.* Journal of medicinal chemistry, 2004. **47**: p. 45-55.
7. Carlson, H.A., *Protein flexibility is an important component of structure-based drug discovery.* Current pharmaceutical design, 2002. **8**: p. 1571-1578.
8. Carlson, H.A., *Protein flexibility and drug design: how to hit a moving target.* Current opinion in chemical biology, 2002. **6**: p. 447-452.
9. Teague, S.J., *Implications of protein flexibility for drug discovery.* Nature reviews drug discovery, 2003. **2**: p. 527-541.
10. Lill, M.A., *Efficient incorporation of protein flexibility and dynamics into molecular docking simulations.* Biochemistry, 2011.
11. Amaro, R.E. and W.W. Li, *Emerging methods for ensemble-based virtual screening.* Current topics in medicinal chemistry, 2010. **10**: p. 3-13.
12. Craig, I.R., J.W. Essex, and K. Spiegel, *Ensemble docking into multiple crystallographically derived protein structures: an evaluation based on the statistical analysis of enrichments.* Journal of chemical information and modeling, 2010. **50**: p. 511-524.
13. Sperandio, O., et al., *How to choose relevant multiple receptor conformations for virtual screening: a test case of Cdk2 and normal mode analysis.* European biophysics journal, 2010. **39**: p. 1365-1372.
14. Boehr, D.D., R. Nussinov, and P.E. Wright, *The role of dynamic conformational ensembles in biomolecular recognition.* Nature chemical biology, 2009. **5**: p. 789-796.
15. Cheng, L.S., et al., *Ensemble-based virtual screening reveals potential novel antiviral compounds for asian influenza neuraminidase.* Journal of medicinal chemistry, 2008. **51**: p. 3878-3894.
16. Totrov, M. and R.A. Abagyan, *Flexible ligand docking to multiple receptor conformations: a practical alternative.* Current opinion in structural biology, 2008. **18**: p. 178-184.
17. Barril, X. and S.D. Morley, *Unveiling the full potential of flexible receptor docking using multiple crystallographic structures.* Journal of medicinal chemistry, 2005. **48**: p. 4432-4443.
18. Park, I.-H. and C. Li, *Dynamic ligand-induced-fit simulation via enhanced conformational samplings and ensemble dockings: a survivin example.* Journal of physical chemistry B, 2010. **114**: p. 5144-5153.
19. Seeliger, D. and B.L. de Groot, *Conformational transitions upon ligand binding: holo-structure prediction from apo conformations.* PLoS computational biology, 2010. **6**: p. e1000634.

20. Bosshard, H.R., *Molecular recognition by induced fit: how fit is the concept?* News in physiological sciences, 2001. **16**: p. 171-3.
21. Koshland, D.E., *The key–lock theory and the induced fit theory.* Angewandte Chemie international edition in English, 1994. **33**: p. 2375-2378.
22. Aleksandrov, A. and T. Simonson, *Molecular dynamics simulations show that conformational selection governs the binding preferences of imatinib for several tyrosine kinases.* Journal of biological chemistry, 2010. **285**: p. 13807-13815.
23. Bakan, A. and I. Bahar, *The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding.* Proceedings of the National academy of sciences of the USA, 2009. **106**: p. 14349-14354.
24. Korb, O., T. Stützle, and T.E. Exner, *PLANTS: Application of Ant Colony Optimization to Structure-Based Drug Design.* Lecture Notes in Computer Science, 2006. **4150**: p. 247-258.
25. Moitessier, N., et al., *Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go.* British journal of pharmacology, 2008. **153 Suppl**: p. S7-S26.
26. Spyrakis, F., et al., *Scoring functions for virtual screening.* Frontiers in drug design & discovery, 2007. **3**: p. 317-379.
27. Coupez, B. and R.A. Lewis, *Docking and scoring - theoretically easy, practically impossible?* Current medicinal chemistry, 2006. **13**: p. 2995-3003.
28. Asses, Y., et al., *Exploring c-Met kinase flexibility by sampling and clustering its conformational space.* Proteins (under submission), 2011.
29. Cui, J.J., *Inhibitors targeting hepatocyte growth factor receptor and their potential therapeutic applications.* Expert opinion on therapeutic patents, 2007. **17**: p. 1035-1045.
30. Underiner, T.L., T. Herbertz, and S.J. Miknyoczki, *Discovery of small molecule c-Met inhibitors: Evolution and profiles of clinical candidates.* Anti-cancer agents in medicinal chemistry, 2010. **10**: p. 7-27.
31. Asses, Y., et al., *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, 2009. **74**: p. 560-570.
32. Jacobs, M.D., P.R. Caron, and B.J. Hare, *Classifying protein kinase structures guides use of ligand-selectivity profiles to predict inactive conformations: structure of lck/imatinib complex.* Proteins, 2008. **70**: p. 1451-1460.
33. Druker, B.J., et al., *Efficacy and Safety of a Specific Inhibitor of the BCR-ABL Tyrosine Kinase in Chronic Myeloid Leukemia.* New England Journal of Medicine, 2001. **344**(14): p. 1031-1037.
34. Capdeville, R., et al., *Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug.* Nat Rev Drug Discov, 2002. **1**(7): p. 493-502.
35. Lipinski, C.A., *Lead- and drug-like compounds: the rule-of-five revolution.* Drug Discovery Today: Technologies, 2004. **1**(4): p. 337-341.
36. Huggins, D.J., A.R. Venkitaraman, and D.R. Spring, *Rational Methods for the Selection of Diverse Screening Compounds.* ACS Chemical Biology, 2011. **6**(3): p. 208-217.
37. Xi, H. and E.A. Lunney, *The design, annotation, and application of a kinase-targeted library.* Methods Mol Biol., 2011. **685**: p. 279-291.
38. Mondal, S., et al., *Natural Products: Promising Resources For Cancer Drug Discovery.* Anticancer Agents Med Chem, 2011. **Under press**.
39. Keiser, M.J., et al., *Predicting new molecular targets for known drugs.* Nature, 2009. **462**: p. 175-181.

40. Sukhai, M.A., et al., *New sources of drugs for hematologic malignancies.* Blood, 2011. **117**(25): p. 6747-6755.
41. Berman, H.M., et al., *The Protein Data Bank.* Acta crystallographica D, 2002. **58**: p. 899-907.
42. Huang, B. and M. Schroeder, *LIGSITEcsc: predicting ligand binding sites using the Connolly surface and degree of conservation.* BMC Struct Biol, 2006. **6**: p. 19.
43. www.grid5000.fr/mediawiki/index.php/Grid5000:Home. *Grid5000*.

Legends

Figure 1: Pocket centers for individual conformers (in orange) versus the average center (in green). In black the sphere centered on the average point and used in the center point definition of the binding site

Figure 2: Position of the average center point (in green) in the 45 used conformers

Figure 3: Poses of the AM7 ligand in the X-ray 2RFN structure: in black the original pose of the ligand in the PDB protein conformation; in orange, green and purple the best docking poses obtained by GOLD on the 2RFN target using a definition of the binding site from a list of residues, from the original ligand and from a center point respectively

Figure 4: Docking results on all chemical libraries

Figure 5: Top 1000 extracted from whole docking results on all chemical libraries

Figure 6: Distribution and Rate of best ligands docked from each chemical library. These ligands are docked on conformations classified by their original affiliation.

Figure 7: Supersposition of F0725-0356 and MT4 in 3EFK PDB structure

Figure 8: Interaction plots for 3EFK original (left) and screened (right) complex comparison

Table I: List of the selected chemical libraries used in the present virtual screening campaign

Table II: List of the secondary structure elements used for aligning all the conformers

Table III: Comparison of the docking results using the 3 binding site definitions

Table IV: Comparison between the individual dockings versus the ensemble one. The active site definition is the same in both cases, namely the average center point as defined above.

Table V: Docked ligands retained from all the screened chemical libraries.

TABLES

Table I: List of the selected chemical libraries used in the present virtual screening campaign

Supplier	Database name: description	population
French laboratories http://chimiotheque-nationale.enscm.fr	Chimiothèque Nationale [refs]: collection of diverse molecules	11,644
Collaboration http://medchem.u-strasbg.fr	JJB: highly-specific molecules from a chemical synthesis perspective	2,566
ChemBridge http://www.chembridge.com/index.php	ChemBridge/Kinase: subset with kinase drug properties from ChemBridge database	15,129
Life Chemicals http://www.lifechemicals.com	LifeChem/Antivirals: Life Chemicals antivirals subset	36,582
	LifeChem/Diversity: small subset from Life Chemicals with maximal chemical diversity	3,918
TimTec http://www.timtec.net	TimTec/Naturals: natural product subset from TimTec database	3,032

Table II: List of the secondary structure elements used for aligning all the conformers

domain	Secondary structure name	Residues concerned
N-terminal	β_1	1076 to 1081
	β_2	1092 to 1098
	β_3	1104 to 1110
	β_4	1144 to 1146
	β_5	1154 to 1158
C-terminal	α_E	1178 to 1198
	α_F	1263 to 1278
	α_H	1310 to 1320
	α_I	1330 to 1343

Table III: Comparison of the docking results using the 3 binding site definitions

Definition of the binding site	Target PDB name	Ligand PDB name	Rank number	Score value
Center + radius 20 Å				
	3DKC	ATP	1	105.5
	2RFN	AM7	1	100.8
	1R0P	KSA	2	81.6
	2WKM	PFY	4	77.7
Residues list				
	3DKC	ATP	1	102.8
	2RFN	AM7	1	98.0
	1R0P	KSA	2	81.3
	2WKM	PFY	4	78.5
From its XRD ligand				
	3DKC	ATP	1	107.1
	2RFN	AM7	1	106.6
	1R0P	KSA	2	78.1
	2WKM	PFY	3	80.4

Table IV: Comparison between the individual dockings versus the ensemble one. The active site definition is the same in both cases, namely the average center point as defined above.

Ligand PDB name	Original PDB target	Rank and Score for the docking in the original PDB target	Rank and score in the ensemble docking 50 poses	Preferred PDB target in the ensemble docking	Rank and score in the ensemble docking 100 poses	Preferred PDB target in the ensemble docking 100 poses
ATP	3DKC	1 105.5	2 99.2	3F66	3 103.3	3F82
AM7	2RFN	1 100.8	3 98.8	3EFJ	1 106.4	3L8V
KSA	**1R0P**	2 81.6	11 81.2	**1R0P**	12 81.2	**1R0P**
PFY	2WKM	4 77.7	14 78.5	2WGJ	13 80.1	2WGJ

Table V: Docked ligands retained from all the screened chemical libraries.

Database name	Retained number	Molecules name	Protein conformer
Chimiothèque Nationale	01	CN2010/AB-00034754	39 (md_l8v)
Bourguignon (JJB set)	01	JJB_focused_library/1085	33 (md_c1x)
Kinase set	02	ChemBridge/7746657 ChemBridge/7923448	11 (rx_efj) 14 (rx_f82)
Antiviral set	25 (top 10)	LifeChemicals/F0725-0356 LifeChemicals/F0772-0607 LifeChemicals/F0375-0311 LifeChemicals/F0816-0342 LifeChemicals/F0737-0405 LifeChemicals/F0737-0393 LifeChemicals/F0301-0263 LifeChemicals/F0721-0868 LifeChemicals/F0715-0299 LifeChemicals/F0539-1482	12 (rx_efk) 12 (rx_efk) 14 (rx_f82) 11 (rx_efj) 14 (rx_f82) 12 (rx_efk) 12 (rx_efk) 12 (rx_efk) 12 (rx_efk) 38 (md_efj)
Diversity set	0		
Natural product set (Timtec)	06	NDL-3040/ST012878 NDL-3040/ST038610 NDL-3040/ST031278 NDL-3040/ST031410 NDL-3040/ST012881 NDL-3040/ST004333	18 (rx_q6w) 38 (md_efj) 38 (md_efj) 33 (md_c1x) 37 (md_dkc) 18 (rx_q6w)

FIGURES

Figure 1: pocket centers for individual conformers (in orange) versus the average center (in green). In black the sphere centered on the average point and used in the center point definition of the binding site



Figure 2: Position of the average center point (in green) in the 45 used conformers

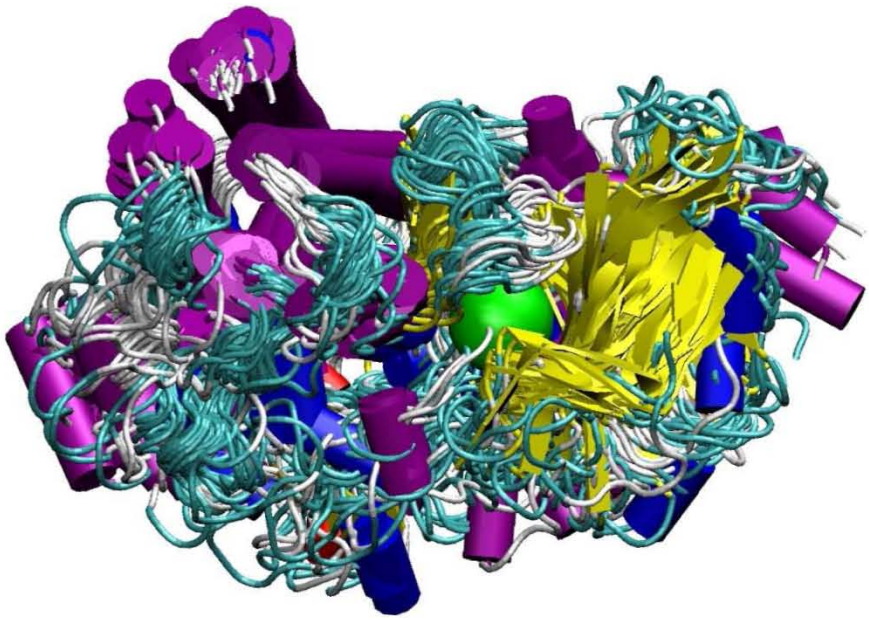


Figure 3: Poses of the AM7 ligand in the X-ray 2RFN structure: in black the original pose of the ligand in the PDB protein conformation; in orange, green and purple the best docking poses obtained by GOLD on the 2RFN target using a definition of the binding site from a list of residues, from the original ligand and from a center point respectively

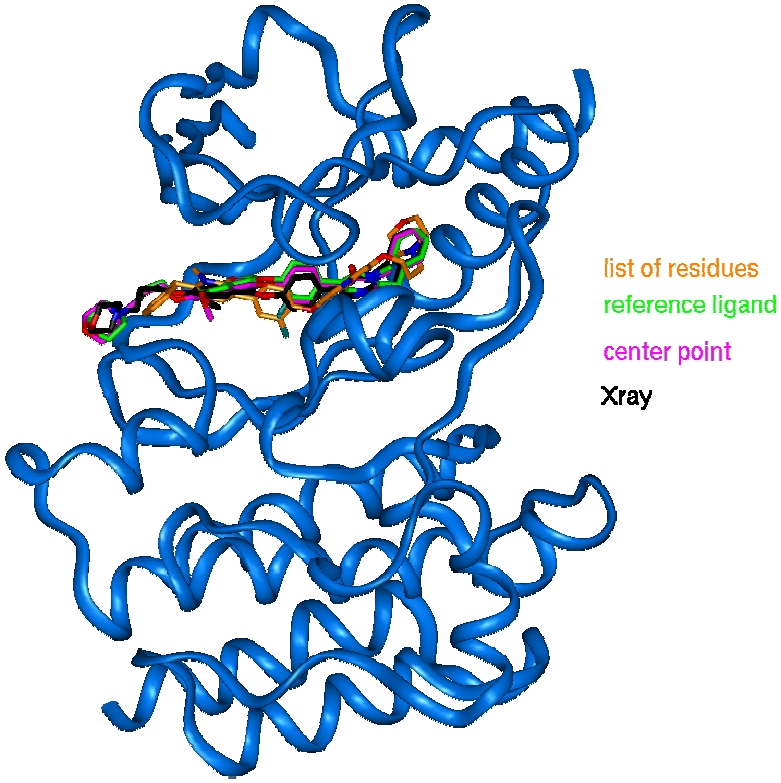


Figure 4: Docking results on all chemical libraries

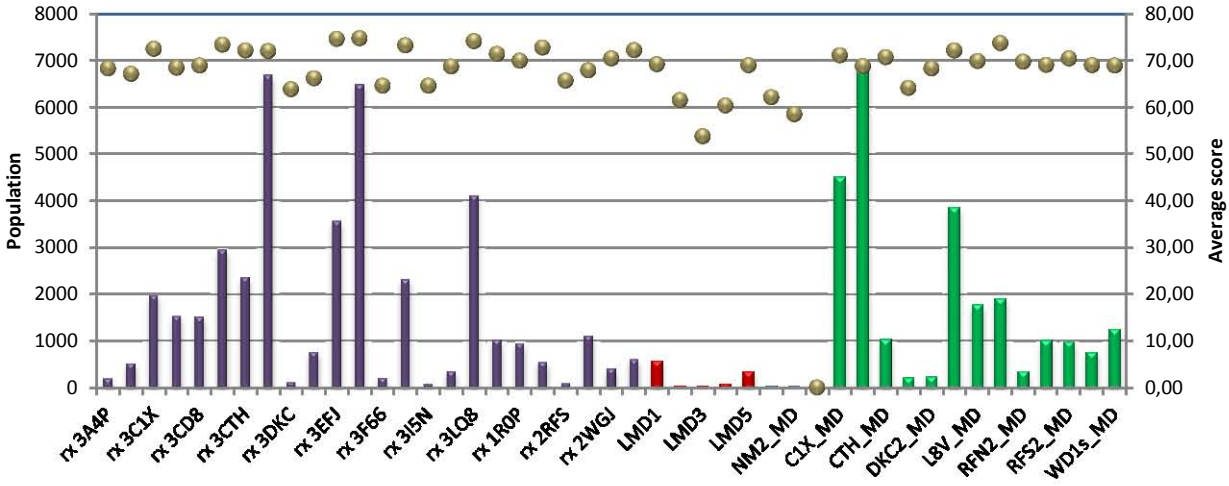


Figure 5: Top 1000 extracted from whole docking results on all chemical libraries

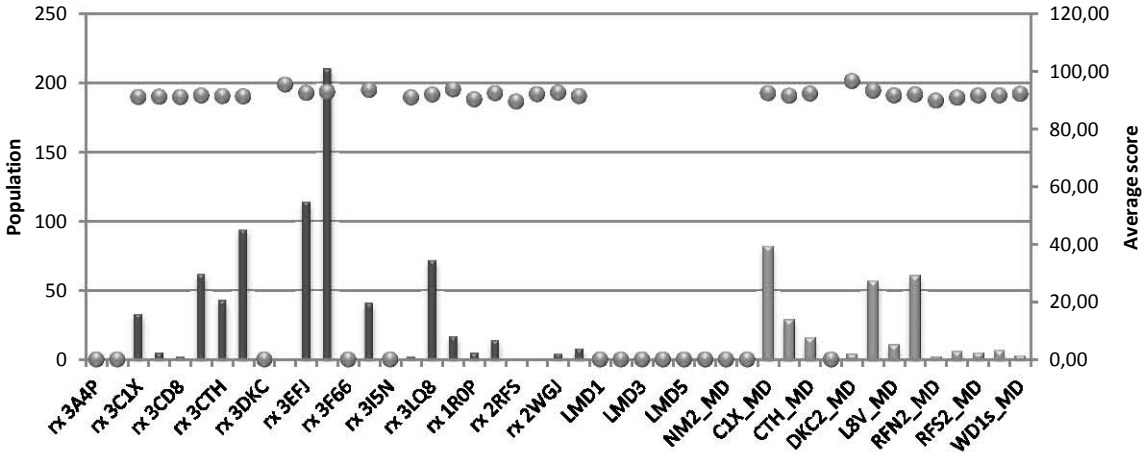


Figure 6: Distribution and Rate of best ligands docked from each chemical library. These ligands are docked on conformations classified by their original affiliation.

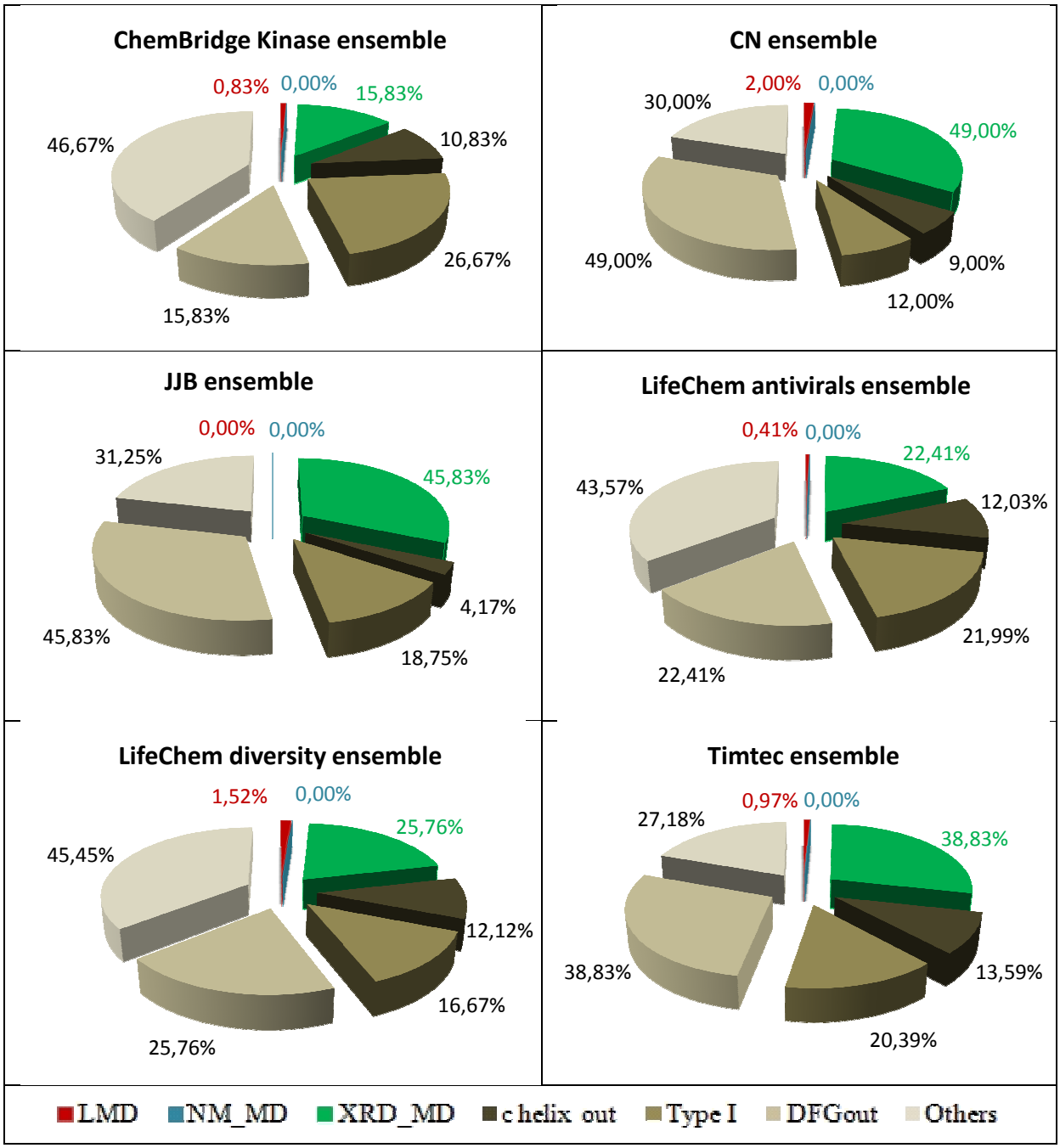


Figure 7: Supersposition of F0725-0356 and MT4 in 3EFK PDB structure

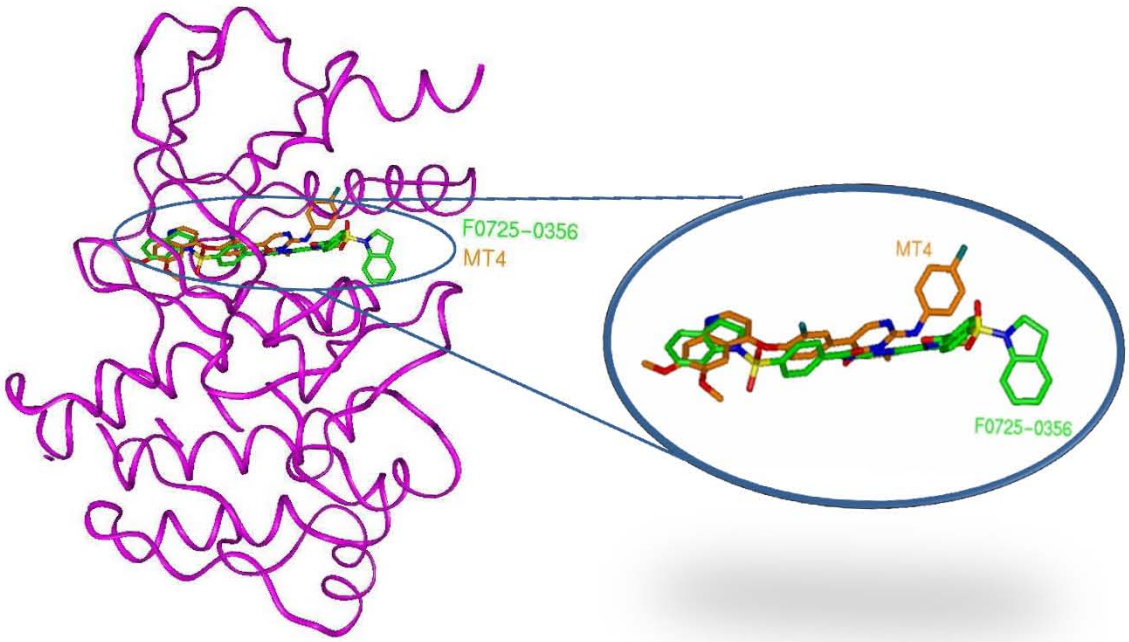
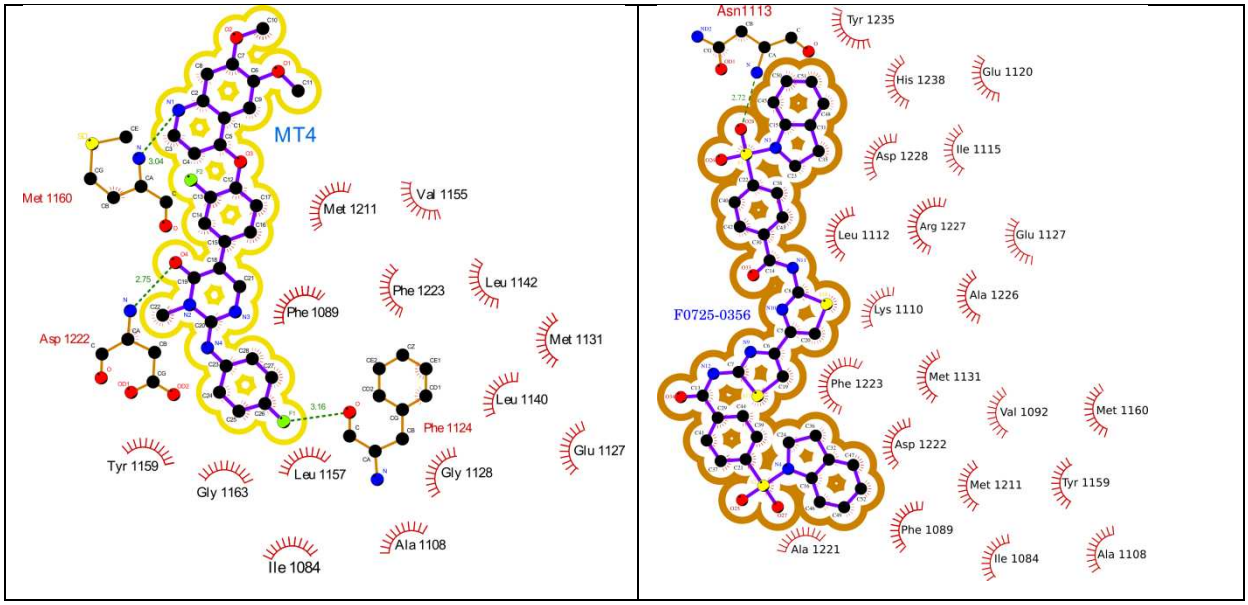
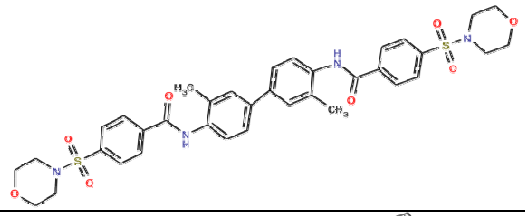
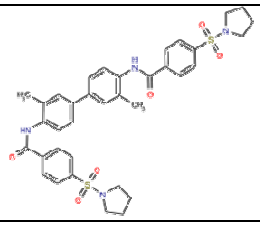
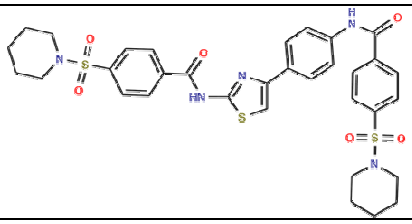
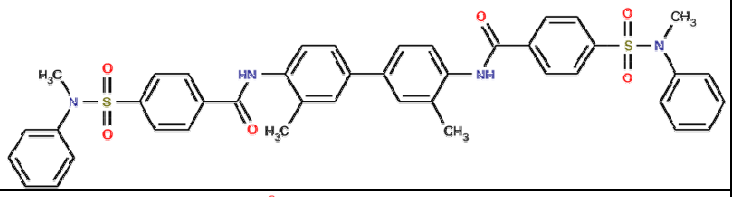
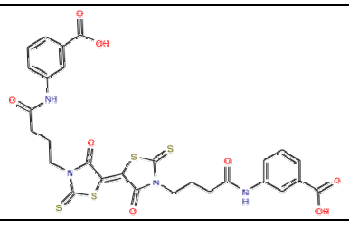
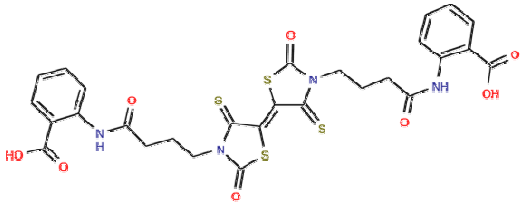
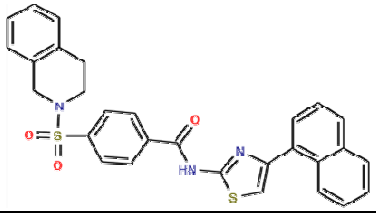


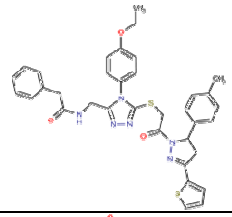
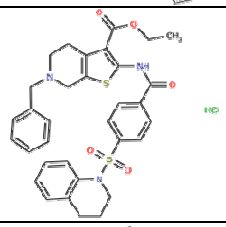
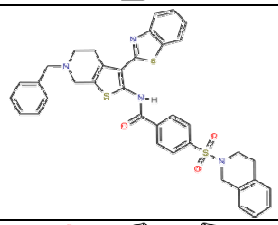
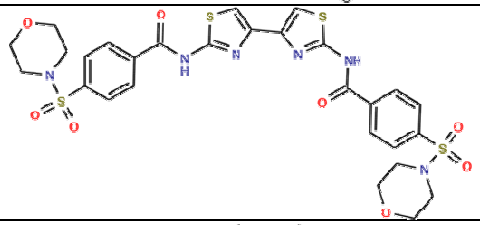
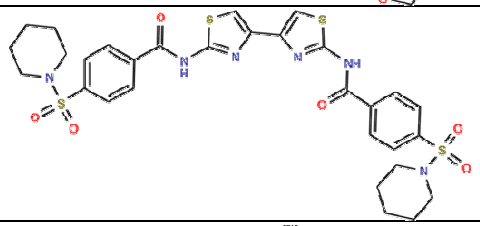
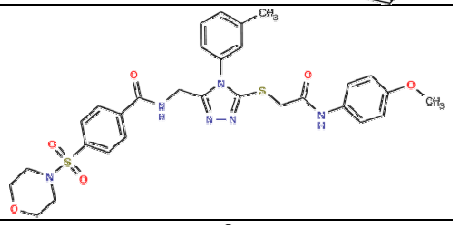
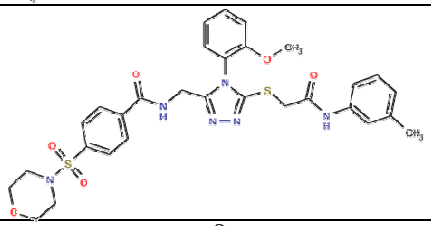
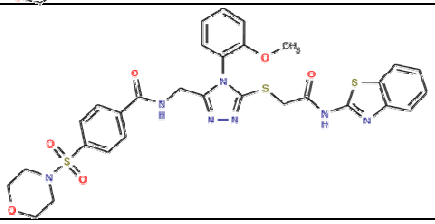
Figure 8: Interaction plots for 3EFK original (left) and screened (right) complexe comparison

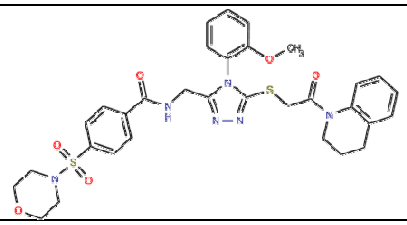
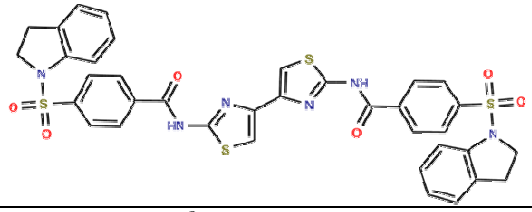
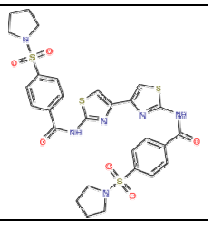
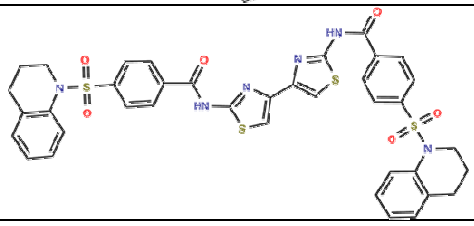
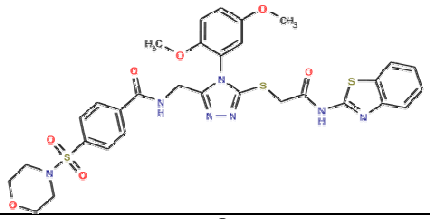
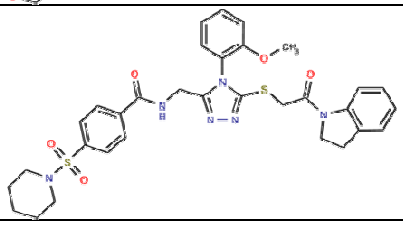
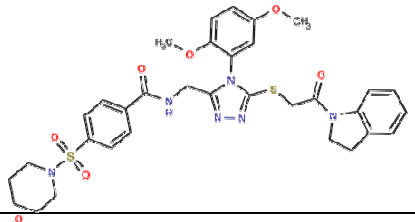
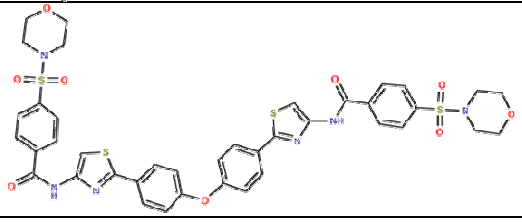


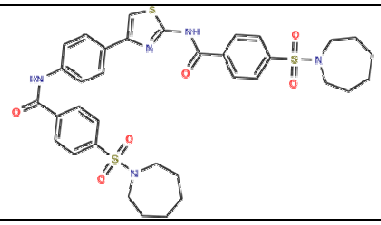
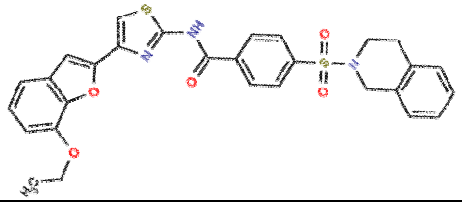
Supplementary material

ID selected molecules from LifeChemical (antiviral) database docking with their chemical formula.

Number	ID	Chemical formula
1	F0011-0324	
2	F0012-0227	
3	F0174-0048	
4	F0301-0263	
5	F0385-0029	
6	F0385-0334	
7	F0473-0261	

8	F0514-4011	
9	F0526-0094	
10	F0539-1482	
11	F0715-0299	
12	F0715-0300	
13	F0721-0868	
14	F0721-0900	
15	F0721-0906	

16	F0721-0911	
17	F0725-0356	
18	F0737-0393	
19	F0737-0405	
20	F0772-0607	
21	F0772-2099	
22	F0772-2147	
23	F0816-0342	

24	F1620-0074	
25	F2252-0240	

III- Commentaires

Cette campagne de criblage a nécessité des ressources informatiques conséquentes qui n'étaient disponibles que sur le réseau GRID5000 : pour effectuer nos calculs, nous avons utilisé une moyenne de 1000 processeurs pendant plusieurs week-ends, seules plages horaires où nous pouvions disposer de 60 heures de calculs d'affilée. Au terme de cette campagne, sur ces 100.000 molécules, un millier d'entre elles ont été retenues sur la base de leurs scores et, parmi ce millier de molécules potentiellement intéressantes, 25 d'entres elles, acquises auprès de Life Chemical, ont été envoyées à nos collègues biologistes pour procéder à des tests de validation.

Cette campagne de criblage nous a apporté plusieurs enseignements intéressants :

- Tout d'abord, elle confirme l'importance du phénomène d' « induced fit » dans l'amarrage des ligands dans le site actif de c-Met. En effet, si l'on compare aux résultats obtenus pour les conformères *holo*, peu de conformères de la forme *apo* ont donné des scores de liaison intéressants (environ 2% sur les 100 meilleures molécules identifiées sur chaque chimiothèque). Ceci montre que le site de liaison *apo* n'a pas les caractéristiques géométriques adéquates pour lier les ligands. Ceci est sans doute dû à l'insuffisante prise en compte de la flexibilité de la protéine par le programme de docking GOLD. De ce fait, cela ne permet pas une bonne adaptation du site de liaisons lors de la reconnaissance protéine-ligand.
- Le logiciel GOLD, dans sa version 5.0.1, apporte certes des options nouvelles et intéressantes, mais son utilisation nécessite des ressources calculatoires particulières : nous avons du demander aux concepteurs de GOLD des adaptations pour traiter notre ensemble de 45 conformères, ce qui, pour chaque machine utilisée, a requis de disposer d'espace mémoire très important. D'autre part, nous avons, grâce à cette campagne de criblage, pu identifier des problèmes dans l'utilisation de cette version du logiciel, comme, par exemple, un pic de transfert de données au début des calculs qui provoquait une saturation des connections entre les nœuds des clusters et, par conséquence, un arrêt intempestif des machines. Finalement, nous avons constaté que seuls des clusters disposant d'un réseau infiniband de dernière génération étaient capables de supporter nos calculs.
- La technique de l' « ensemble docking » est prometteuse, mais nécessite encore de nombreuses améliorations pour être réellement utilisée en routine : si, au cours de nos calculs de validation, nous avons retrouvé les bonnes poses des ligands dans leur structure PDB d'origine, nous avons constaté que la fonction de score utilisée par GOLD demanderait à être optimisée.

Au stade actuel de la rédaction de ce manuscrit, les molécules transmises aux biologistes sont en cours de tests afin d'en mesurer les affinités réelles pour la kinase c-Met. Si nous ne pouvons pas nous prononcer sur ces résultats qui seuls permettront de valider (ou pas) la pertinence de notre stratégie, il est cependant clair que, quel que soit le résultat de ces tests, une optimisation de certaines de ces molécules serait un projet intéressant à réaliser. En effet, les résultats des dockings effectués indiquent, pour ces composés, une affinité théorique suffisamment intéressante pour que, même si leur liaison à c-Met est faible, ils constituent néanmoins une source d'innovation possible par rapport aux composés existants.

PARTIE 5

CONCLUSIONS ET PERSPECTIVES

Bilan et conclusion

Mon sujet de recherche avait pour objectif l'identification et l'optimisation de nouvelles petites molécules inhibitrices du domaine kinase de c-Met, visant à bloquer le processus de prolifération des métastases. Dans ce projet pluridisciplinaire, soutenu par l'INCa, mon rôle consista à réaliser des « expériences » de criblage *in silico* de chimiothèques et à étudier, par le biais de la modélisation moléculaire et de simulations de dynamique moléculaire, le mode d'action des composés actifs connus. Mon but était d'identifier de nouveaux inhibiteurs potentiels du domaine tyrosine kinase de c-Met, de les proposer aux chimistes impliqués dans ce projet lorsque des synthèses étaient nécessaires et enfin aux biologistes pour tester la validité de nos propositions.

Au début de cette thèse, nous avons effectué un criblage préliminaire basé sur des molécules issues de la chimie combinatoire sur la cible c-Met. À cette époque, seule une structure *holo* (1R0P) et une structure *apo* (1R1W) étaient disponibles. Ces 2 structures servirent de base à ce premier criblage qui révéla une molécule intéressante issue d'une série de molécules appartenant à la classe de dérivés amides d'acides aminés. La publication de nouvelles structures cristallographiques peu de temps après, nous a conduits à établir une stratégie permettant la prise en compte de celles-ci dans notre criblage. Ces travaux font l'objet d'une publication (Furlan & al.).

Ensuite nous avons analysé les données concernant c-Met dont nous disposions à ce moment là, en particulier les structures cristallographiques du récepteur tyrosine kinase c-Met publiées sur le site de la PDB. Les analyses en question concernaient essentiellement tous types d'interactions existantes (hydrogènes, aromatiques, électrostatiques…). Ainsi, des calculs d'énergies d'interactions entre le ligand et la cible cristallisés ont été réalisés. Ceci nous a permis de définir les zones d'interactions essentielles à c-Met avec les ligands cristallisés qui, par leurs différences structurales, nous permettent de croire que nous avons envisagé la plupart des types d'interactions c-Met/ligand possibles. Cette exploration nous a permis également de constater un autre mode de liaisons, non connu encore pour les autres tyrosines kinase, qui implique le motif DFG ainsi que l'hélice C de la protéine. Ce travail a donné lieu à une publication (Asses et al.).

La publication des nouvelles structures des complexes tridimensionnels avec c-Met d'une part et les constats issus des analyses que nous avons déjà réalisées d'autre part, nous ont montré qu'il était absolument nécessaire de considérer la flexibilité de c-Met dans nos calculs. Afin de tenir compte au mieux de la flexibilité de notre cible moléculaire pour la modéliser convenablement, nous avons alors décidé de réaliser des simulations plus approfondies de dynamique moléculaire. Ces simulations ont été réalisées dans le but d'échantillonner aussi complètement que possible l'espace conformationnel de c-Met. Pour effectuer une exploration exhaustive, nous avons tenu compte aussi bien de la forme *holo* (représentée par 26 structures cristallographiques) que de la forme *apo* (représentée par une structure cristallographique) de c-Met. Pour ces 2 types de formes nous avons effectué des

simulations de DM égales à 1micro-seconde pour la forme *apo* et de 10 nano-secondes/structures cristallographiques pour la forme *holo*. Nous avons également utilisé l'approche des modes normaux de vibration sur la forme *apo* de la cible. Les trois conformations issues des modes normaux furent soumises à 40 nano-secondes de simulations de DM chacune. L'étape suivante de ce travail fut la détermination des conformations les plus représentatives de l'intégralité de notre ensemble. Pour cela nous avons utilisé l'approche de l'analyse en groupe ou « clustering » en employant aussi bien le clustering hiérarchique que le non-hiérarchique. Les résultats obtenus sont encourageants puisque, sur un total de 1 400 000 conformations, nous avons extrait 10 conformations représentatives. Le caractère représentatif de ces conformations s'est avéré correct car nous avons montré l'appartenance structurale à cet ensemble de deux conformations dites actives de c-Met suite à leur publication récente (code pdb : 3Q6W et 3RHK). Un article concernant cette partie de mon travail est en cours de soumission (Asses et al.).

Enfin, l'ensemble des conformations représentatives retenues de notre étude a fait l'objet d'une campagne de criblage sur un ensemble de chimiothèques. Ces dernières ont été filtrées et traitées dans l'objectif de ne cibler que des molécules dites « druglike ». En tout 10 chimiothèques ont été criblées sur un ensemble de 45 conformations de c-Met. Cet ensemble comportait les 24 conformations issues des études cristallographiques, les 10 conformères représentatifs issus du travail d'échantillonnage et 11 conformères supplémentaires considérés comme secondaires par rapport aux 10 précédents. Cette sélection avait été faite afin d'élargir notre choix de représentants et de créer un équilibre par rapport aux 24 conformations issues des études cristallographiques. Les premiers résultats obtenus semblent très prometteurs. Un article décrivant ce travail est en cours de rédaction.

Perspectives

Le premier constat issu de notre étude est que les protéines tyrosine kinases n'ont pas fini de révéler leurs secrets. L'intérêt porté à c-Met s'est concrétisé ces 3 dernières années par l'explosion des structures cristallographiques, confirmant ainsi son rôle important dans la thérapie anticancéreuse. Parmi les perspectives directes de notre étude, nous pouvons citer la réalisation de tests biologiques sur un grand nombre de molécules issues de notre criblage et l'identification de nouvelles molécules intéressantes. Pour celles-ci, nous pourrons envisager la détermination des structures tridimensionnelles des complexes avec c-Met. Ces études devraient contribuer à apporter une meilleure connaissance du comportement structural de la cible d'intérêt c-Met, des interactions qu'elle peut avoir avec ses ligands et de ses relations structure-fonction.

D'un point de vue technique, on pourrait procéder à l'automatisation des protocoles de calculs utilisés afin de les rendre plus facilement utilisables par d'autres.

D'un point de vue personnel, le caractère pluridisciplinaire du projet dans lequel s'inscrit cette thèse, au carrefour de la chimie, de la biologie et de l'informatique, m'a ouvert de nombreuses perspectives d'avenir.

Références bibliographiques

1. Evans, T. and W.V. Lergerghe, *Rapport sur la santé dans le monde, 2008. Les soins de santé primaires. Mainenant plus que jamais.* Bibliothèque de l'OMS (Organisation mondiale de la Santé), 2008.
2. Gabriel, J.A., *The Biology of Cancer.* Wiley, 2007.
3. Delehedde, M. and F. Denhez, *Que sait-on du cancer ?* EDP Sciences, 2006.
4. Rousselet-Chapeau, M.C., *modèle général de la carcinogénèse multi étapes*. 2007, M1 STOS-UE physiopathologie des maladies cancereuses
5. Morère, J.F., et al., eds. *Thérapeutique du cancer*. 2001, Springer.
6. Druker, B.J., et al., *Chronic Myelogenous Leukemia.* Hematology, 2001. **2001**(1): p. 87-112.
7. Deininger, M.W.N. and B.J. Druker, *Specific Targeted Therapy of Chronic Myelogenous Leukemia with Imatinib.* Pharmacological Reviews, 2003. **55**(3): p. 401-423.
8. Page, C.P., et al., eds. *Pharmacologie intégrée*. De Boeck Université ed. 1999.
9. *Leem*. Available from: http://www.leem.org/.
10. *Afssaps*. Available from: http://www.afssaps.fr/.
11. *FDA*. Available from: http://www.fda.gov.
12. Bharath, E.N., S.N. Manjula, and A. Vijaychand, *In silico drug design-tool for overcoming the innovation deficit in the drug discovery process* International Journal of Pharmacy and Pharmaceutical Sciences, 2011. **3**(2).
13. Rigby, A.C., *Exploring novel chemical space through the use of computational and structural biology.* Comb. Chem. High Throughput Screen., 2009. **12**: p. 927-928.
14. Lipinski, C. and A. Hopkins, *Navigating chemical space for biology and medicine.* Nature, 2004. **432**: p. 855-861.
15. Dobson, C.M., *Chemical space and biology.* Nature, 2004. **432**: p. 824-828.
16. Lipinski, C., et al., *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings.* Adv. Drug Delivery Rev., 2001. **46**: p. 3-26.
17. Manley, P.W., S.W. Cowan-Jacob, and J. Mestan, *Advances in the structural biology, design and clinical development of Bcr-Abl kinase inhibitors for the treatment of chronic myeloid leukaemia.* Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics, 2005. **1754**(1-2): p. 3-13.
18. Olivieri, A. and L. Manzione, *Dasatinib: a new step in molecular target therapy.* Annals of Oncology, 2007. **18**(suppl 6): p. vi42-vi46.
19. Boschelli, F., K. Arndt, and C. Gambacorti-Passerini, *Bosutinib: A review of preclinical studies in chronic myelogenous leukaemia.* European journal of cancer (Oxford, England : 1990), 2010. **46**(10): p. 1781-1789.
20. Kimura, S., et al., *NS-187, a potent and selective dual Bcr-Abl/Lyn tyrosine kinase inhibitor, is a novel agent for imatinib-resistant leukemia.* Blood, 2005. **106**(12): p. 3948-3954.
21. Radi, M., et al., *Discovery and SAR of 1,3,4-thiadiazole derivatives as potent Abl tyrosine kinase inhibitors and cytodifferentiating agents.* Bioorganic & Medicinal Chemistry Letters, 2008. **18**(3): p. 1207-1211.
22. Demetri, G.D., et al., *Efficacy and safety of sunitinib in patients with advanced gastrointestinal stromal tumour after failure of imatinib: a randomised controlled trial.* The Lancet, 2006. **368**(9544): p. 1329-1338.

23. Li, Z., et al., *Erlotinib Effectively Inhibits JAK2V617F Activity and Polycythemia Vera Cell Growth.* Journal of Biological Chemistry, 2007. **282**(6): p. 3428-3432.
24. Garrett, R.H. and C.M. Grisham, eds. *Biochimie*. 2 ed. 1999, De Boeck Université.
25. Hanks, S. and T. Hunter, *Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification.* The FASEB Journal, 1995. **9**(8): p. 576-596.
26. Manning, G., et al., *The Protein Kinase Complement of the Human Genome.* Science, 2002. **298**(5600): p. 1912-1934.
27. Blume-Jensen, P. and T. Hunter, *Oncogenic kinase signalling.* Nature, 2001. **411**(6835): p. 355-365.
28. King, N. and S.B. Carroll, *A receptor tyrosine kinase from choanoflagellates: Molecular insights into early animal evolution.* Proceedings of the National Academy of Sciences, 2001. **98**(26): p. 15032-15037.
29. Hanks, S.K. and A.M. Quinn, *Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members.* Methods Enzymol., 1991. **200**: p. 38-62.
30. Oda, K., et al., *A comprehensive pathway map of epidermal growth factor receptor signaling.* Mol Syst Biol, 2005. **1**.
31. Matsui, R., et al., *Glucose-6 Phosphate Dehydrogenase Deficiency Decreases the Vascular Response to Angiotensin II.* Circulation, 2005. **112**(2): p. 257-263.
32. *Protein DataBase*. Available from: http://www.pdb.org/pdb/home/home.do.
33. Naldini, L., et al., *Scatter factor and hepatocyte growth factor are indistinguishable ligands for the MET receptor.* EMBO Journal, 1991. **10**: p. 2867-2878.
34. Mitsudomi, T. and Y. Yatabe, *Mutations of the epidermal growth factor receptor gene and related genes as determinants of epidermal growth factor receptor tyrosine kinase inhibitors sensitivity in lung cancer.* Cancer Science, 2007. **98**(12): p. 1817-1824.
35. Umeki, K., G. Shiota, and H. Kawasaki, *Clinical Significance of c-met Oncogene Alterations in Human Colorectal Cancer.* Oncology, 1999. **56**(4): p. 314-321.
36. Zeng, Z.-S., et al., *c-Met gene amplification is associated with advanced stage colorectal cancer and liver metastases.* Cancer Letters, 2008. **265**(2): p. 258-269.
37. Ma, P.C., et al., *Functional Expression and Mutations of c-Met and Its Therapeutic Inhibition with SU11274 and Small Interfering RNA in Non–Small Cell Lung Cancer.* Cancer Research, 2005. **65**(4): p. 1479-1488.
38. Garcia, S., et al., *c-Met overexpression in inflammatory breast carcinomas: automated quantification on tissue microarrays.* Br J Cancer, 2007. **96**(2): p. 329-335.
39. Wong, A.S.T., et al., *Coexpression of hepatocyte growth factor-Met: an early step in ovarian carcinogenesis?* Oncogne, 2001. **20**: p. 1318-1328.
40. Danilkovitch-Miagkova, A. and B. Zbar, *Dysregulation of Met receptor tyrosine kinase activity in invasive tumors.* The Journal of Clinical Investigation, 2002. **109**(7): p. 863-867.
41. Berthou, S., et al., *The Met kinase inhibitor SU11274 exhibits a selective inhibition pattern toward different receptor mutated variants.* Oncogene, 2004. **23**(31): p. 5387-5393.
42. Bellon, S.F., et al., *c-Met Inhibitors with Novel Binding Mode Show Activity against Several Hereditary Papillary Renal Cell Carcinoma-related Mutations.* Journal of Biological Chemistry, 2008. **283**(5): p. 2675-2683.
43. Comoglio, P.M., S. Giordano, and L. Trusolino, *Drug development of MET inhibitors: targeting oncogene addiction and expedience.* Nat Rev Drug Discov, 2008. **7**(6): p. 504-516.

44. Matsumoto, K. and T. Nakamura, *NK4 (HGF-antagonist/angiogenesis inhibitor) in cancer biology and therapeutics.* Cancer Science, 2003. **94**(4): p. 321-327.
45. Matsumoto, K. and T. Nakamura, *NK4 gene therapy targeting HGF-Met and angiogenesis* Frontiers in Bioscience 2008. **13**: p. 1943-1951.
46. Morotti, A., et al., *K252a inhibits the oncogenic properties of Met, the HGF receptor.* Oncogene, 2002. **21**: p. 4885–4893.
47. Soman, N.R., et al., *The TPR-MET oncogenic rearrangement is present and expressed in human gastric carcinoma and precursor lesions.* Proceedings of the National Academy of Sciences, 1991. **88**(11): p. 4892-4896.
48. Zou, H.Y., et al., *An Orally Available Small-Molecule Inhibitor of c-Met, PF-2341066, Exhibits Cytoreductive Antitumor Efficacy through Antiproliferative and Antiangiogenic Mechanisms.* Cancer Research, 2007. **67**(9): p. 4408-4417.
49. Eder, J.P., et al., *Novel Therapeutic Inhibitors of the c-Met Signaling Pathway in Cancer.* Clinical Cancer Research, 2009. **15**(7): p. 2207-2214.
50. Zhang, Y., et al., *XL-184, a MET, VEGFR-2 and RET kinase inhibitor for the treatment of thyroid cancer, glioblastoma multiforme and NSCLC.* IDrugs, 2010. **13**: p. 112-121.
51. Welsh, J., et al., *The c-Met receptor tyrosine kinase inhibitor MP470 radiosensitizes glioblastoma cells.* Radiation Oncology, 2009. **4**(1): p. 69.
52. Buchanan, S.G., et al., *SGX523 is an exquisitely selective, ATP-competitive inhibitor of the MET receptor tyrosine kinase with antitumor activity in vivo.* Molecular Cancer Therapeutics, 2009. **8**(12): p. 3181-3190.
53. Backes, A., et al., *Small-molecule inhibitors binding to protein kinases. Part I: exceptions from the traditional pharmacophore approach of type I inhibition.* Expert Opinion on Drug Discovery, 2008. **3**: p. 1409-1425.
54. Backes, A., et al., *Small-molecule inhibitors binding to protein kinase. Part II: the novel pharmacophore approach of type II and type III inhibition.* Expert Opinion on Drug Discovery, 2008. **3**(12): p. 1427-1449.
55. Liao, J.J.-L., *Molecular Recognition of Protein Kinase Binding Pockets for Design of Potent and Selective Kinase Inhibitors.* ChemInform, 2007. **38**(22): p. no-no.
56. Schiering, N., et al., *Crystal structure of the tyrosine kinase domain of the hepatocyte growth factor receptor c-Met and its complex with the microbial alkaloid K-252a.* Proceedings of the National Academy of Sciences, 2003. **100**(22): p. 12654-12659.
57. Porter, J., et al., *Discovery of a novel series of quinoxalines as inhibitors of c-Met kinase.* Bioorganic & Medicinal Chemistry Letters, 2009. **19**(2): p. 397-400.
58. Cohen, N.C., *Guidebook on Molecular Modeling in Drug Design.* Boston: Academic Press, 1996. **ISBN 012178245x**.
59. Livage, J., *Chimie de la matière condensée.* L'annuaire du Collège de France 2010.
60. Guner, O.F., *Pharmacophore Perception, Developement, and use in Drug Design.* La jolla, Calif: International University Line, 2000. **ISBN 0-9636817-6-1**.
61. Schneider, G. and U. Fechner, *Computer-based de novo design of drug-like molecules.* Nat Rev Drug Discov, 2005. **4**(8): p. 649-663.
62. Wang, R., Y. Gao, and L. Lai, *LigBuilder: a multi-purpose program for structure-based drug design.* J. Mol. Model., 2000. **6**: p. 498-516.
63. Wang, R., L. Lai, and S. Wang, *Further development and validation of empirical scoring functions for structure-based binding affinity prediction.* Journal of Computer-Aided Molecular Design, 2002. **16**(1): p. 11-26.
64. de Azevedo WF, D.R., *Computational Methods for Calculation of Ligand-Binding Affinity.* Current Drug Targets, 2008. **9**: p. 1031-1039.

65. Liang, S., et al., *Consensus scoring for enriching near-native structures from protein–protein docking decoys.* Proteins: Structure, Function, and Bioinformatics, 2009. **75**(2): p. 397-403.
66. Amari, S., et al., *VISCANA: Visualized Cluster Analysis of Protein−Ligand Interaction Based on the ab Initio Fragment Molecular Orbital Method for Virtual Ligand Screening.* Journal of Chemical Information and Modeling, 2005. **46**(1): p. 221-230.
67. Andrews, D.H., *The Relation Between the Raman Spectra and the Structure of Organic Molecules.* Physical Review, 1930. **36**(3): p. 544.
68. Born, M. and R. Oppenheimer, *Zur Quantentheorie der Molekeln.* Annalen der Physik, 1927. **389**(20): p. 457-484.
69. Monard, G. (2003) *Introduction à la Modélisation Moléculaire*. Formation continue CNRS.
70. MacKerell Jr., A.D., *Empirical force fields for biological macromolecules: Overview and issues.* Journal of Computational Chemistry, 2004. **25**(13): p. 1584-1604.
71. Pearlman, D.A. and P.A. Kollman, *Evaluating the Assumptions Underlying Force Field Development and Application, Using Free Energy Conformational Maps for Nucleosides.* J. Am. Chem. Soc., 1991. **113**: p. 7167-7177.
72. MacKerell Jr., A.D., J. Wiórkiewicz-Kuczera, and M. Karplus, *An all-atom empirical energy function for the simulation of nucleic acids.* Journal of the American Chemical Society, 1995. **117**: p. 11946-11975.
73. MacKerell Jr., A.D., et al., *CHARMM: The energy function and its parametrization with an overview of the program*, in *The Encyclopedia of computational chemistry*, P.V.R. Schleyer, et al., Editor. 1998, John Wiley & sons: Chichester. p. 271-277.
74. MacKerell Jr., A.D., et al., *All-atom empirical potential for molecular modeling and dynamics studies of proteins.* Journal of Physical Chemistry, 1998. **102**: p. 3586-3616.
75. MacKerell Jr., A.D. and N.K. Benavali, *All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution.* Journal of Computational Chemistry, 2000. **21**(2): p. 105-120.
76. MacKerell Jr., A.D., M. Feig, and C.L. Brooks III, *Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations.* Journal of Computational Chemistry, 2004. **25**(11): p. 1400-1415.
77. Weiner, S.J., et al., *An all atom force field for simulations of proteins and nucleic acids.* Journal of Computational Chemistry, 1986. **7**(2): p. 230-252.
78. Weiner, S.J., et al., *A new force field for molecular mechanical simulation of nucleic acids and proteins.* Journal of the American Chemical Society, 1984. **106**: p. 765-784.
79. Cornell, W.D., et al., *A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules.* Journal of the American Chemical Society, 1995. **117**(19): p. 5179-5197.
80. Kollman, P., et al., *The development/application of a "minimalist" force field using a combination of ab initio and experimental data. in Computer simulation of biomolecular systems: Theoretical and experimental applications.* Van Gunsteren, W. F.; Weiner, P. K., Eds. Escom, 1997. **3**(83-86).
81. Cheatham III, T.E., P. Cieplak, and P.A. Kollman, *A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat.* Journal of Biomolecular Structure & Dynamics, 1999. **16**(4): p. 845-862.
82. Duan, Y., et al., *A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations.* Journal of Computational Chemistry, 2003. **24**(16): p. 1999-2012.

83. Wang, J., et al., *Development and testing of a general amber force field.* Journal of Computational Chemistry, 2004. **25**(9): p. 1157-1174.
84. Allinger, N.L., *Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms.* Journal of the American Chemical Society, 1977. **99**(25): p. 8127-8134.
85. Nemethy, G., M.S. Pottle, and S.H. A., J. Phys. Chem, 1983. **87**(1883-1887).
86. Hagler.A.T., In Hruby.VJ. and MeienhoferJ. (eds), The Peptides. Academic Press, New York, 1985: p. 213-299.
87. Kramer, B., M. Rarey, and T. Lengauer, *Evaluation of the FlexX incremental construction algorithm for protein-ligand docking.* Proteins: Structure, Function, and Genetics, 1999. **37**(2): p. 228-241.
88. Lennard-Jones, J.E., *Cohesion.* Proceedings of the Physical Society, 1931. **43**(5): p. 461-482.
89. Rappé, A., et al., *UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations.* Journal of the American Chemical Society, 1992. **114**(25): p. 10024-10035.
90. Maple, J., U. Dinur, and A. Hagler, *Derivation of force fields for molecular mechanics and dynamics from ab initio energy surfaces.* Proceedings of the National Academy of Sciences of the United States of America, 1988. **85**(15): p. 5350-5354.
91. Allinger, N., Y. Yuh, and J. Lii, *Molecular mechanics. The MM3 force field for hydrocarbons. 1.* Journal of the American Chemical Society, 1989. **111**(23): p. 8551-8566.
92. Ponder, J.W. and D.A. Case, *Force fields for protein simulations.* Advances in Protein Chemistry, 2003. **66**: p. 27-85.
93. Kollman, P.A., et al., *The development/application of a "minimalist" organic/biochemical molecular mechanic force field using a combination of ab initio calculations and experimental data*, in *Computer simulations of biomolecular systems*, A. Wilkinson, P. Weiner, and W.F. van Gunsteren, Editors. 1997, Elsevier. p. 83-96.
94. Weiner, S.J., et al., *An all-atom force field for simulations of proteins and nucleic acids.* Journal of Computational Chemistry, 1986. **7**: p. 230-252.
95. Cornell, W.D., et al., *A second generation force field for the simulation of proteins, nucleic acids, and organic molecules.* Journal of the American Chemical Society, 1995. **117**: p. 5179-5197.
96. Cheatham III, T.E., P. Cieplak, and P.A. Kollman, *A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat.* Journal of Biomolecular Structure & Dynamics, 1999. **16**: p. 845-862.
97. Duan, Y., et al., *A point-charge force field for molecular mechanics simulations of proteins.* Journal of Computational Chemistry, 2003. **24**: p. 1999-2012.
98. Wang, J., et al., *Development and testing of a general Amber force field.* Journal of Computational Chemistry, 2004. **25**: p. 1157-1174.
99. Daura, X., A. Mark, and W. van Gunsteren, *Parametrization of aliphatic CHn united atoms of GROMOS96 force field.* Journal of computational chemistry, 1998. **19**(5): p. 535-547.
100. Oostenbrink, C., et al., *A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6.* Journal of computational chemistry, 2004. **25**(13): p. 1656-1676.
101. Jorgensen, W. and J. Tirado-Rives, *The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin.* Journal of the American Chemical Society, 1988. **110**(6): p. 1657-1666.

102. Jorgensen, W., D. Maxwell, and J. Tirado-Rives, *Development and testing of the OPLS all-atom force-field on conformational energetics and properties of organic liquids.* Journal of the American Chemical Society, 1996. **118**(45): p. 11225-11236.
103. Nelson, M., et al., *NAMD - A parallel, object-oriented molecular dynamics program.* International Journal of Supercomputer Applications and High Performance Computing, 1996. **10**: p. 251-268.
104. Kalé, L., et al., *NAMD2: Greater scalability for parallel molecular dynamics.* Journal of Computational Physics, 1999. **151**: p. 283-312.
105. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD.* Journal of Computational Chemistry, 2005. **26**(16): p. 1781-1802.
106. Scott, W., et al., *The GROMOS biomolecular simulation program package.* Journal of Physical Chemistry A, 1999. **103**(19): p. 3596-3607.
107. Brooks, B.R., et al., *CHARMM: A program for macromolecular energy, minimization, and dynamics calculations.* Journal of Computational Chemistry, 1983. **4**(2): p. 187-217.
108. Weiner, P.K. and P.A. Kollman, *AMBER: Assisted Model Building with Energy Refinement. A general program for modeling molecules and their interactions.* Journal of Computational Chemistry, 1981. **2**(3): p. 287-303.
109. Pearlman, D.A., et al., *AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules.* Computer Physics Communications, 1995. **91**: p. 1-41.
110. Neumaier, A., *Molecular modeling of proteins and mathematical prediction of protein structure.* SIAM review, 1997. **39**(3): p. 407-460.
111. Momany, F.A., et al., *Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occuring amino acids.* Journal of Physical Chemistry, 1975. **79**(22): p. 2361-2381.
112. Némethy, G., M.S. Pottle, and H.A. Sheraga, *Energy parameters in polypeptides. 9. Updating of geometrical parameters, nonbonded interactions, and hydrogen bond interactions for the naturally occuring amino acids.* Journal of Physical Chemistry, 1983. **87**: p. 1833-1887.
113. Foloppe, N. and A.D. MacKerell Jr., *All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data.* Journal of Computational Chemistry, 2000. **21**(2): p. 86-104.
114. van Gunsteren, W.F. and J.C. Berendsen, *Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry.* Angewandte Chemie, International Edition in English, 1990. **29**: p. 992-1023.
115. Leach, A.R., *Molecular modelling: principles and applications (2nd Edition)*. 2001: Prentice Hall.
116. Metropolis, N., et al., *Equation of state calculations by fast computing machines.* Journal of Chemical Physics, 1953. **21**: p. 1087-1092.
117. Berendsen, H., et al., *Molecular dynamics with coupling to an external bath.* Journal of Chemical Physics, 1984. **81**: p. 3684.
118. Andersen, H., *Molecular dynamics simulations at constant pressure and/or temperature.* Journal of Chemical Physics, 1980. **72**(4): p. 2384-2393.
119. Karplus, M. and J. Kuriyan, *Molecular dynamics and protein function.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(19): p. 6679-85.

120. Verlet, L., *Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules.* Physical Review, 1967. **159**(1): p. 98-103.
121. Hockney, R.W., *The potential calculation and some applications.* Methods in Computational Physics, 1970. **9**: p. 136-211.
122. Beeman, D., *Some multistep methods for use in molecular dynamics simulations.* Journal of Computational Physics, 1976. **20**: p. 130-139.
123. Swope, W.C., et al., *A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules:application to small water clusters.* Journal of Chemical Physics, 1982. **76**: p. 637-649.
124. Ryckaert, J.-P., G. Ciccotti, and H.J.C. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes.* Journal of Computational Physics, 1977. **23**(3): p. 327-341.
125. Kräutler, V., W.F. van Gunsteren, and P.H. Hünerberger, *A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations.* Journal of Computational Chemistry, 2001. **22**(5): p. 501-508.
126. van Gunsteren, W.F. and H.J.C. Berendsen, *Algorithms for macromolecular dynamics and constraint dynamics.* Molecular Physics, 1977. **34**: p. 1311-1327.
127. Grubmüller, H., et al., *Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions.* Molecular Simulation, 1991. **6**: p. 121-142.
128. Tuckerman, M.E., B.J. Berne, and G.J. Martyna, *Reversible multiple time scale molecular dynamics.* Journal of Chemical Physics, 1992. **97**(3): p. 1990-2001.
129. Humphreys, D.D., R.A. Freisner, and B.J. Berne, *A multiple-time-step molecular dynamics algorithm for macromolecules.* Journal of Physical Chemistry, 1994. **98**: p. 6885-6892.
130. Cavasotto, C.N., J.A. Kovacs, and R.A. Abagyan, *Representing Receptor Flexibility in Ligand Docking through Relevant Normal Modes.* Journal of the American Chemical Society, 2005. **127**(26): p. 9632-9640.
131. Dobbins, S.E., V.I. Lesk, and M.J.E. Sternberg, *Insights into protein flexibility: The relationship between normal modes and conformational change upon protein–protein docking.* Proceedings of the National Academy of Sciences, 2008. **105**(30): p. 10390-10395.
132. May, A. and M. Zacharias, *Energy minimization in low-frequency normal modes to efficiently allow for global flexibility during systematic protein–protein docking.* Proteins: Structure, Function, and Bioinformatics, 2008. **70**(3): p. 794-809.
133. May, A. and M. Zacharias, *Protein−Ligand Docking Accounting for Receptor Side Chain and Global Flexibility in Normal Modes: Evaluation on Kinase Inhibitor Cross Docking.* Journal of Medicinal Chemistry, 2008. **51**(12): p. 3499-3506.
134. Maréchal, J.-D. and D. Perahia, *Use of normal modes for structural modeling of proteins: the case study of rat heme oxygenase 1.* European Biophysics Journal, 2008. **37**(7): p. 1157-1165.
135. Floquet, N., et al., *Normal mode analysis as a prerequisite for drug design: Application to matrix metalloproteinases inhibitors.* FEBS Letters, 2006. **580**(22): p. 5130-5136.
136. Levitt, M., C. Sander, and P.S. Stern, *Protein normal-mode dynamics : trypsin inhibitor, crambin, ribonuclease and lysozyme.* J. Mol. Biol., 1985. **181**: p. 423-447.
137. Levitt, M. and R. Sharon, *Accurate simulation of protein dynamics in solution.* Proceedings of the National Academy of Sciences of the United States of America, 1988. **85**(20): p. 7557-61.

138. Harrisson, R.W., *Variational calculation of the normal modes of a large macromolecules : Methods and some initial results.* Biopolymers, 1984. **23**: p. 2943-2949.
139. Schuyler, A.D. and G.S. Chirikjian, *Efficient determination of low-frequency normal modes of large protein structures by cluster-NMA.* Journal of Molecular Graphics and Modelling, 2005. **24**(1): p. 46-58.
140. Batista, P.R., et al., *Consensus modes, a robust description of protein collective motions from multiple-minima normal mode analysis-application to the HIV-1 protease.* Physical Chemistry Chemical Physics, 2010. **12**(12): p. 2850-2859.
141. Rueda, M., P. Chacón, and M. Orozco, *Thorough Validation of Protein Normal Mode Analysis: A Comparative Study with Essential Dynamics.* Structure, 2007. **15**(5): p. 565-575.
142. Balog, E., et al., *Direct Determination of Vibrational Density of States Change on Ligand Binding to a Protein.* Physical Review Letters, 2004. **93**(2): p. 028103.
143. Réat, V., et al., *Solvent dependence of dynamic transitions in protein solutions.* Proceedings of the National Academy of Sciences, 2000. **97**(18): p. 9961-9966.
144. Smith, J.C., *Protein dynamics: comparison of simulations with inelastic neutron scattering experiments.* Quarterly Reviews of Biophysics, 1991. **24**: p. 227-291.
145. Levy, R.M., D. Perahia, and M. Karplus, *Molecular dynamics of an alpha-helical polypeptide : temperature dependence and deviation from harmonic behavior.* Proceedings of the National Academy of Sciences, 1982. **79**: p. 1346-1350.
146. Doruker, P., R.L. Jernigan, and I. Bahar, *Dynamics of large proteins through hierarchical levels of coarse-grained structures.* Journal of Computational Chemistry, 2002. **23**(1): p. 119-127.
147. Atilgan, A.R., et al., *Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model.* Biophysical Journal, 2001. **80**(1): p. 505-515.
148. Cui, Q., et al., *A Normal Mode Analysis of Structural Plasticity in the Biomolecular Motor F1-ATPase.* Journal of Molecular Biology, 2004. **340**(2): p. 345-372.
149. Krebs, W.G., et al., *Normal mode analysis of macromolecular motions in a database framework: Developing mode concentration as a useful classifying statistic.* Proteins: Structure, Function, and Bioinformatics, 2002. **48**(4): p. 682-695.
150. Seno, Y. and N. Go, *Deoxymyoglobin studied by the conformational normal mode analysis: I. Dynamics of globin and the heme-globin interaction.* Journal of Molecular Biology, 1990. **216**(1): p. 95-109.
151. Rueda, M., G. Bottegoni, and R. Abagyan, *Consistent Improvement of Cross-Docking Results Using Binding Site Ensembles Generated with Elastic Network Normal Modes.* Journal of Chemical Information and Modeling, 2009. **49**(3): p. 716-725.
152. McCammon, J.A., B.R. Gelin, and M. Karplus, *Dynamics of folded proteins.* Nature, 1977. **267**(5612): p. 585-90.
153. Koehl, P., *Electrostatics calculations: latest methodological advances.* Current Opinion in Structural Biology, 2006. **16**(2): p. 142-151.
154. Saito, M., *Molecular dynamics simulations of proteins in solutions: Artifacts caused by the cutoff approximation.* Journal of Chemical Physics, 1994. **101**(5): p. 4055-4061.
155. Ewald, P.P., *Die berechnung optischer und elektrostatischer gitterpotentiale.* Annals of Physics, 1921. **64**: p. 253-287.
156. Essmann, U., et al., *A smooth particle mesh Ewald method.* Journal of Chemical Physics, 1995. **103**(19): p. 8577-8593.
157. Manning, C.D., P. Raghavan, and H. Schütze, *An Introduction to Information Retrieval.* Cambridge University Press, 2008(0521865719).

158. Rissanen, J., *Modeling by shortest data description.* Automatica, 1978. **14**(5): p. 465-471.
159. Heyer, L.J., S. Kruglyak, and S. Yooseph, *Exploring Expression Data: Identification and Analysis of Coexpressed Genes.* Genome Research, 1999. **9**(11): p. 1106-1115.
160. Abagyan, R. and M. Totrov, *High-throughput docking and lead generation.* Current Opinion in Chemical Biology, 2001. **5**: p. 375-382.
161. Brooijmans, N. and I.D. Kuntz, *Molecular recognition and docking algorithms.* Annual Review of Biophysics and Biomolecular Structure, 2003. **32**: p. 335-373.
162. Kitchen, D.B., et al., *Docking and scoring in virtual screening for drug discovery: methods and applications.* Nature Reviews Drug Discovery, 2004. **3**(11): p. 935-949.
163. Rester, U., *Dock around the Clock - Current Status of Small Molecule Docking and Scoring.* QSAR & Combinatorial Science, 2006. **25**(7): p. 605-615.
164. Bursulaya, B.D., et al., *Comparative study of several algorithms for flexible ligand docking.* Journal of Computer-Aided Molecular Design, 2003. **17**: p. 755-763.
165. Sousa, S.F., P.A. Fernandes, and M.J. Ramos, *Protein-ligand docking: current status and future challenges.* Proteins: Structure, Function, and Bioinformatics, 2006. **65**(1): p. 15-26.
166. Kontoyianni, M., L.M. McClellan, and G.S. Sokol, *Evaluation of docking performance: comparative data on docking algorithms.* Journal of Medicinal Chemistry, 2004. **47**(3): p. 558-565.
167. Perola, E., W.P. Walters, and P.S. Charifson, *A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance.* Proteins: Structure, Function, and Bioinformatics, 2004. **56**(2): p. 235-49.
168. Kellenberger, E., et al., *Comparative evaluation of eight docking tools for docking and virtual screening accuracy.* Proteins: Structure, Function, and Bioinformatics, 2004. **57**: p. 225-242.
169. Cummings, M.D., et al., *Comparison of automated docking programs as virtual screening tools.* Journal of Medicinal Chemistry, 2005. **48**(4): p. 962-976.
170. Zhou, Z., et al., *Comparative Performance of Several Flexible Docking Programs and Scoring Functions: Enrichment Studies for a Diverse Set of Pharmaceutically Relevant Targets.* J. Chem. Inf. Model., 2007. **47**(4): p. 1599-1608.
171. Leroux, V. and B. Maigret, *Should structure-based virtual screening techniques be used more extensively in modern drug discovery?* Comp. App. Chem., 2007. **24**(1): p. 1-10.
172. Mendez, R., et al., *Assessment of CAPRI predictions in rounds 3-5 shows progress in docking procedures.* Proteins: Structure, Function, and Bioinformatics, 2005. **60**(2): p. 150-169.
173. Gohlke, H. and G. Klebe, *Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors.* Angewandte Chemie, International Edition in English, 2002. **41**(15): p. 2644-2676.
174. Ritchie, D.W. and G.J.L. Kemp, *Fast computation, rotation, and comparison of low resolution spherical harmonic molecular surfaces.* Journal of Computational Chemistry, 1999. **20**(4): p. 383-395.
175. Cai, W., X. Shao, and B. Maigret, *Protein-ligand recognition using spherical harmonic molecular surfaces: towards a fast and efficient filter for large virtual throughput screening.* Journal of Molecular Graphics & Modelling, 2002(4): p. 313-328.
176. Knegtel, R.M., I.D. Kuntz, and C.M. Oshiro, *Molecular docking to ensembles of protein structures.* Journal of Molecular Biology, 1997. **266**(2): p. 424-440.

177. Osterberg, F., et al., *Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock.* PROTEINS: Structure, Function, and Genetics, 2002. **46**(1): p. 34-40.
178. Claussen, H., et al., *FlexE: efficient molecular docking considering protein structure variations.* J Mol Biol, 2001. **308**(2): p. 377-395.
179. Barril, X. and S.D. Morley, *Unveiling the full potential of flexible receptor docking using multiple crystallographic structures.* J Med Chem, 2005. **48**(13): p. 4432-4443.
180. Wong, C.F., et al., *Molecular docking of balanol to dynamics snapshots of protein kinase A.* Proteins: Structure, Function, and Bioinformatics, 2005. **61**(4): p. 850-858.
181. Kuntz, I.D., et al., *A geometric approach to macromolecule-ligand interactions.* Journal of Molecular Biology, 1982. **161**(2): p. 269-288.
182. Friesner, R.A., et al., *Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy.* Journal of Medicinal Chemistry, 2004. **47**(7): p. 1739-1749.
183. (link). *OpenEye Science Software: Santa Fe, NM, USA.*; Available from: http://www.eyesopen.com.
184. (link). *Accelrys Inc., San Diego, CA, USA.*; Available from: http://www.accelrys.com/products/catalyst/.
185. Miller, M.D., et al., *FLOG: a system to select 'quasi-flexible' ligands complementary to a receptor of known three-dimensional structure.* Journal of Computer-Aided Molecular Design, 1994. **8**(2): p. 153-174.
186. Pang, Y.P., et al., *EUDOC: a computer program for identification of drug interaction sites in macromolecules and drug leads from chemical databases.* Journal of Computational Chemistry, 2001. **22**(15): p. 1750-1771.
187. Nicklaus, M.C., et al., *Conformational changes of small molecules binding to proteins.* Bioorg Med Chem, 1995. **3**(4): p. 411-428.
188. Ewing, T.J., et al., *DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases.* Journal of Computer-Aided Molecular Design, 2001. **15**(5): p. 411-428.
189. Rarey, M., et al., *A fast flexible docking method using an incremental construction algorithm.* Journal of Molecular Biology, 1996. **261**(3): p. 470-489.
190. Welch, W., J. Ruppert, and A.N. Jain, *Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites.* Chemistry & Biology, 1996. **3**(6): p. 449-462.
191. Mitchell, M., *An introduction to genetic algorithms*. 1996: MIT Press. 205.
192. Glover, F. and M. Laguna, *Tabu Search*. 1997: Springer.
193. Abagyan, R., M. Totrov, and D. Kuznetsov, *ICM - A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation.* Journal of Computational Chemistry, 1994. **15**(5): p. 488-506.
194. McMartin, C. and R.S. Bohacek, *QXP: powerful, rapid computer algorithms for structure-based drug design.* Journal of Computer-Aided Molecular Design, 1997. **11**(4): p. 333-344.
195. Liu, M. and S. Wang, *MCDOCK: a Monte Carlo simulation approach to the molecular docking problem.* Journal of Computer-Aided Molecular Design, 1999. **13**(5): p. 435-451.
196. Morris, G., et al., *Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function.* Journal of Computational Chemistry, 1998. **19**(14): p. 1639-1662.
197. Jones, G., et al., *Development and validation of a genetic algorithm for flexible docking.* Journal of Molecular Biology, 1997. **267**(3): p. 727-748.

198. Taylor, J.S. and R.M. Burnett, *DARWIN: a program for docking flexible molecules.* Proteins: Structure, Function, and Genetics, 2000. **41**(2): p. 173-191.
199. Baxter, C.A., et al., *Flexible docking using Tabu search and an empirical estimate of binding affinity.* Proteins: Structure, Function, and Genetics, 1998. **33**(3): p. 367-382.
200. Alvarez, J.C., *High-throughput docking as a source of novel drug leads.* Curr Opin Chem Biol, 2004. **8**(4): p. 365-370.
201. Ghosh, S., et al., *Structure-based virtual screening of chemical libraries for drug discovery.* Current Opinion in Chemical Biology, 2006. **10**(3): p. 194-202.
202. Koshland Jr., D., *The key–lock theory and the induced fit theory.* Angewandte Chemie, International Edition in English, 1994. **33**(23-24): p. 2375-2378.
203. Erickson, J.A., et al., *Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy.* J Med Chem, 2004. **47**(1): p. 45-55.
204. Teague, S.J., *Implications of protein flexibility for drug discovery.* Nature Reviews Drug Discovery, 2003. **2**(7): p. 527-541.
205. Ferrari, A.M., et al., *Soft docking and multiple receptor conformations in virtual screening.* Journal of Medicinal Chemistry, 2004. **47**(21): p. 5076-5084.
206. Leach, A.R., *Ligand docking to proteins with discrete side-chain flexibility.* Journal of Molecular Biology, 1994. **235**(1): p. 345-356.
207. Kallblad, P. and P.M. Dean, *Efficient conformational sampling of local side-chain flexibility.* Journal of Molecular Biology, 2003. **326**(5): p. 1651-1665.
208. Meiler, J. and D. Baker, *ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility.* Proteins: Structure, Function, and Bioinformatics, 2006. **65**(3): p. 538-548.
209. Korb, O., T. Stützle, and T. Exner, *PLANTS: Application of Ant Colony Optimization to Structure-Based Drug Design*. Ant Colony Optimization and Swarm Intelligence. 2006: Springer Berlin / Heidelberg. 247-258.
210. Chen, H.M., et al., *SODOCK: swarm optimization for highly flexible protein-ligand docking.* J Comput Chem, 2007. **28**(2): p. 612-623.
211. Banks, A., J. Vincent, and C. Anyakoha, *A review of particle swarm optimization. Part I: background and development.* Natural Computing, 2007. **6**(4): p. 467-484.
212. Huse, M. and J. Kuriyan, *The Conformational Plasticity of Protein Kinases.* Cell, 2002. **109**(3): p. 275-282.
213. Jacobs, D.J., *Ensemble-based methods for describing protein dynamics.* Current Opinion in Pharmacology, 2010. **10**(6): p. 760-769.
214. Amaro, R.E. and W.W. Li, *Emerging Methods for Ensemble-Based Virtual Screening.* Curr. Top. Med. Chem., 2010. **10**(1): p. 3-13.
215. Kollman, P., *Free energy calculations: Applications to chemical and biochemical phenomena.* Chemical Reviews, 1993. **93**(7): p. 2395-2417.
216. Simonson, T., G. Archontis, and M. Karplus, *Free energy simulations come of age: protein-ligand recognition.* Accounts of Chemical Research, 2002. **35**(6): p. 430-437.
217. Ferrara, P., et al., *Assessing scoring functions for protein-ligand interactions.* Journal of Medicinal Chemistry, 2004. **47**: p. 3032-3047.
218. Schulz-Gasch, T. and M. Stahl, *Scoring functions for protein-ligand interactions: a critical perspective.* Drug Discovery Today, 2004. **1**(3): p. 231-239.
219. Wang, R., et al., *An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein-ligand complexes.* J Chem Inf Comput Sci, 2004. **44**(6): p. 2114-2125.
220. Goodsell, D.S., G.M. Morris, and A.J. Olson, *Automated docking of flexible ligands: applications of AutoDock.* Journal of Molecular Recognition, 1996. **9**(1): p. 1-5.

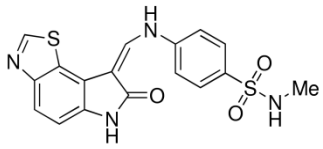
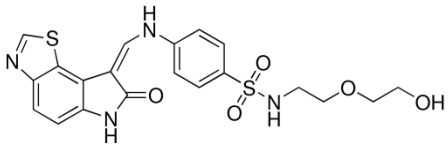
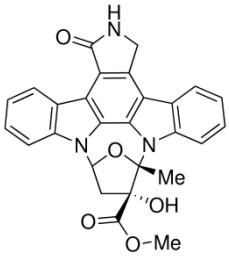
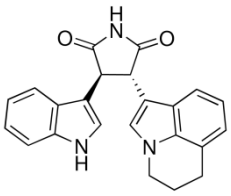
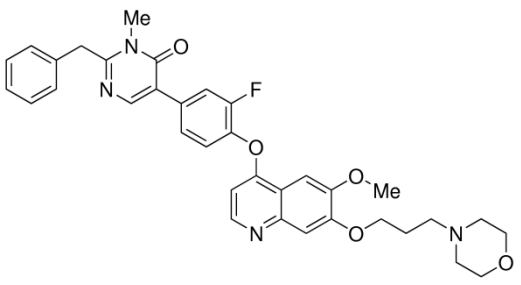
221. Eldridge, M.D., et al., *Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes.* Journal of Computer-Aided Molecular Design, 1997. **11**(5): p. 425-445.
222. Gehlhaar, D.K., et al., *Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming.* Chemical Biology, 1995. **2**(5): p. 317-324.
223. Krammer, A., et al., *LigScore: a novel scoring function for predicting binding affinities.* Journal of Molecular Graphics and Modelling, 2005. **23**(5): p. 395-407.
224. Muegge, I. and Y.C. Martin, *A General and Fast Scoring Function for Protein-Ligand Interactions: A Simplified Potential Approach.* Journal of Medicinal Chemistry, 1999. **42**(5): p. 791-804.
225. DeWitte, R.S. and E.I. Shakhnovich, *SMoG: de Novo Design Method Based on Simple, Fast, and Accurate Free Energy Estimates. 1. Methodology and Supporting Evidence.* Journal of the American Chemical Society, 1996. **118**(47): p. 11733-11744.
226. Gohlke, H., M. Hendlich, and G. Klebe, *Knowledge-based scoring function to predict protein-ligand interactions.* Journal of Molecular Biology, 2000. **295**(2): p. 337-356.
227. Charifson, P.S., et al., *Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins.* Journal of Medicinal Chemistry, 1999. **42**(25): p. 5100-5109.
228. Feher, M., *Consensus scoring for protein-ligand interactions.* Drug Discovery Today, 2006. **11**(9-10): p. 421-428.
229. Venkatachalam, C.M., et al., *LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites.* Journal of Molecular Graphics & Modelling, 2003. **21**(4): p. 289-307.
230. McGann, M.R., et al., *Gaussian docking functions.* Biopolymers, 2003. **68**(1): p. 76-90.
231. Cai, W., et al., *SHEF: a vHTS geometrical filter using coefficients of spherical harmonic molecular surfaces.* Journal of Molecular Modeling, 2008. **14**(5): p. 393-401.
232. Kuhn, B., et al., *Validation and use of the MM-PBSA approach for drug discovery.* J Med Chem, 2005. **48**(12): p. 4040-4048.
233. Cole, J.C., et al., *Comparing protein-ligand docking programs is difficult.* Proteins: Structure, Function, and Bioinformatics, 2005. **60**(3): p. 325-332.
234. Leroux, V. and M. Maigret, *Should structure-based virtual screening techniques be used more extensively in modern drug discovery?* Computers and Applied Chemistry, 2007. **24**(1): p. 1-10.
235. Laederach, A. and P.J. Reilly, *Specific empirical free energy function for automated docking of carbohydrates to proteins.* Journal of Computational Chemistry, 2003. **24**(14): p. 1748-1757.
236. Walters, W.P., M.T. Stahl, and M.A. Murcko, *Virtual screening - an overview.* Drug Discovery Today, 1998. **3**(4): p. 160-178.
237. Bleicher, K.H., et al., *Hit and lead generation: beyond high-throughput screening.* Nature Reviews Drug Discovery, 2003. **2**(5): p. 369-378.
238. Beautrait, A., et al., *Multiple-step virtual screening using VSM-G: overview and validation of fast geometrical matching enrichment.* Journal of Molecular Modeling, 2008. **14**(2): p. 135-148.
239. Corbeil, C.R., P. Englebienne, and N. Moitessier, *Docking ligands into flexible and solvated macromolecules. 1. Development and validation of FITTED 1.0.* J Chem Inf Model, 2007. **47**(2): p. 435-449.

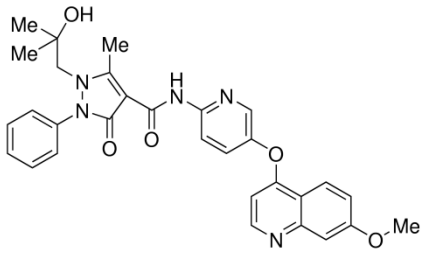
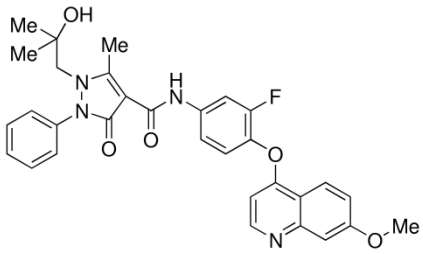
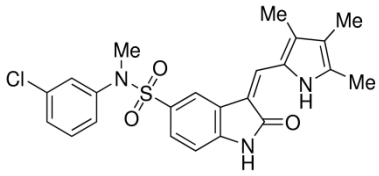
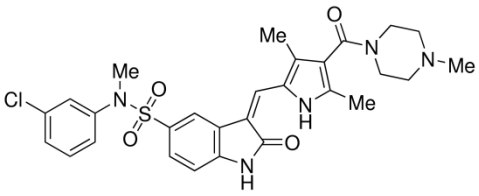
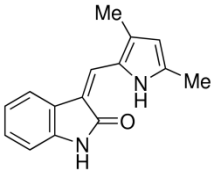
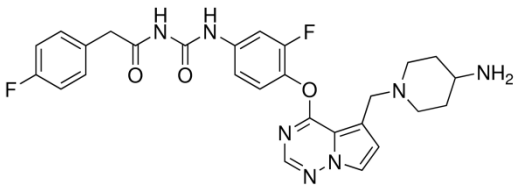
240. Zhao, Y. and M.F. Sanner, *FLIPDock: docking flexible ligands into flexible receptors.* Proteins: Structure, Function, and Bioinformatics, 2007. **68**(3): p. 726-737.
241. M. F. Di Renzo, et al., *Somatic mutations of the MET oncogene are selected during metastatic spread of human HNSC.* Oncogene, 2000. **19**: p. 1547-55.
242. Asses, Y., *Modélisation et Étude Structurale de Ligands du Récepteur périphérique de la Cholécystokinine : Screening Virtuel* in *Chimie Physique et Théorique*. 2006, University of Sciences and Technologies Houari Boumediene USTHB: Algiers. p. 128.
243. Patané, S., et al., *A new Met inhibitory-scaffold identified by a focused forward chemical biological screen.* Biochemical and Biophysical Research Communications, 2008. **375**(2): p. 184-189.
244. Kim E. S. and S. R., *MET pathway as a therapeutic target.* J Thorac Oncol., 2009. **4**: p. 444-447.
245. Stellrecht C. M. and G. V., *MET receptor tyrosine kinase as a therapeutic anticancer target.* Cancer Letters, 2009. **280**: p. 1-14.
246. Liu X., et al., *Targeting the c-MET signaling pathway for cancer therapy.* Expert Opin Investig Drugs, 2008. **17**: p. 997-1011.
247. Dussault I. and Bellon S. F. , *c-Met inhibitors with different binding modes: two is better than one.* Cell Cycle, 2008. **7**: p. 1157-1160.
248. Bellon S. F., et al., *c-Met inhibitors with novel binding mode show activity against several hereditary papillary renal cell carcinoma-related mutations.* J Biol Chem., 2008. **283**: p. 2675-2683.
249. Hammes G. G., Chang Y. C., and Oas T. G., *Conformational selection or induced fit: a flux description of reaction mechanism.* Proc Natl Acad Sci U S A, 2009. **106**
p. 13737-13741.
250. Csermely P., Palotai R., and Nussinov R., *Induced fit, conformational selection and independent dynamic segments: an extended view of binding events.* Trends Biochem Sci, 2010. **35**: p. 539-546.
251. Zhou H. X., *From induced fit to conformational selection: a continuum of binding mechanism controlled by the timescale of conformational transitions.* Biophys. J., 2010. **98**: p. L15-L17.
252. Silva, D.-A., et al., *A Role for Both Conformational Selection and Induced Fit in Ligand Binding by the LAO Protein.* PLoS Comput Biol, 2011. **7**(5): p. e1002054.
253. Cheng, L.S., et al., *Ensemble-Based Virtual Screening Reveals Potential Novel Antiviral Compounds for Avian Influenza Neuraminidase.* Journal of Medicinal Chemistry, 2008. **51**(13): p. 3878-3894.
254. Deng, J., et al., *Dynamic Receptor-Based Pharmacophore Model Development and Its Application in Designing Novel HIV-1 Integrase Inhibitors.* Journal of Medicinal Chemistry, 2005. **48**(5): p. 1496-1505.
255. Huang, S.-Y. and X. Zou, *Ensemble docking of multiple protein structures: Considering protein structural variations in molecular docking.* Proteins: Structure, Function, and Bioinformatics, 2007. **66**(2): p. 399-421.
256. *GOLD*. Available from: http://www.ccdc.cam.ac.uk/products/life_sciences/gold/.
257. *Autodock*. Available from: http://autodock.scripps.edu/.
258. Ivetac A. and McCammon J. A., *Molecular Recognition in the Case of Flexible Targets.* Curr Pharm Des, 2011 May 27. **sous presse**.
259. Huang S. Y. and Zou X., *Advances and challenges in protein-ligand docking.* Int J Mol Sci, 2010. **11**: p. 3016-3034.
260. Lin J. H., *Accommodating protein flexibility for structure-based drug design.* Curr Top Med Chem, 2011. **11**(2): p. 171-178.

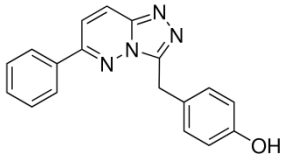
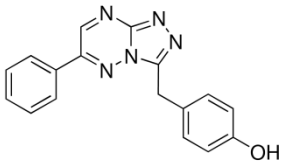
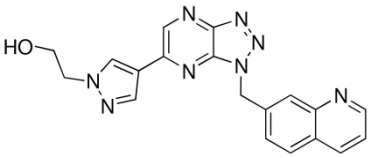
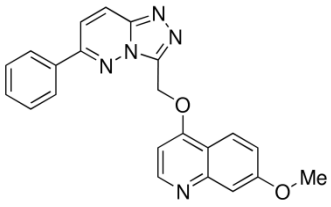
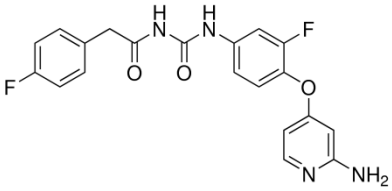
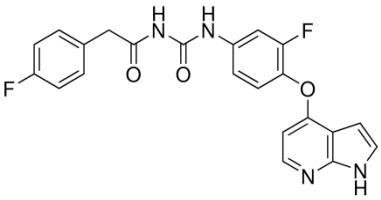
261. *Chimiothèque nationale*. Available from: http://chimiotheque-nationale.enscm.fr/.
262. Bourguignon J. J., *Communication personnelle.*
263. *Life Chemicals*. Available from: http://www.lifechemicals.com/.
264. *Timtec*. Available from: http://www.timtec.net/.
265. *Chembridge*. Available from: http://www.chembridge.com/screening_libraries/.

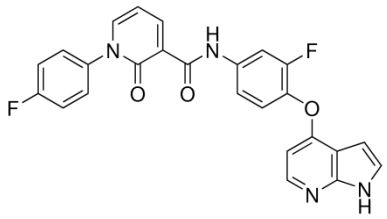
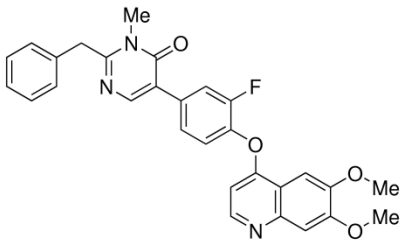
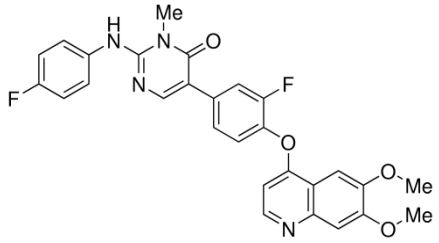
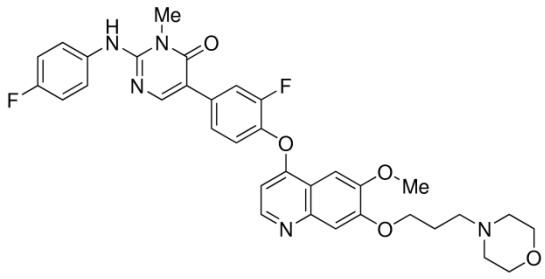
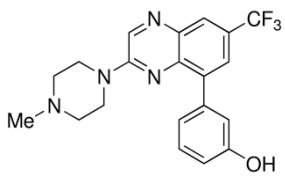
Annexes

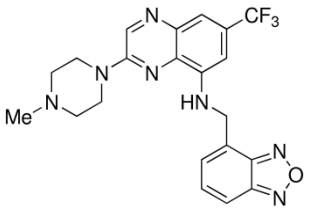
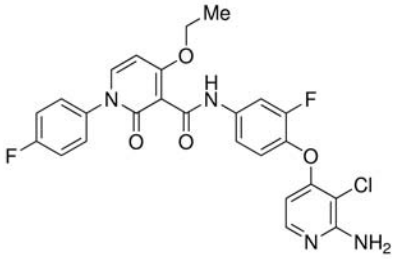
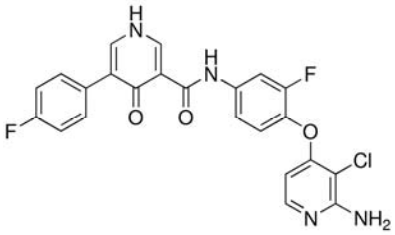
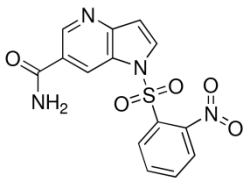
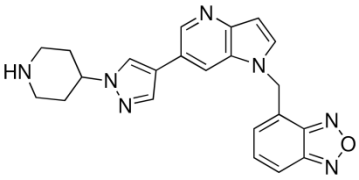
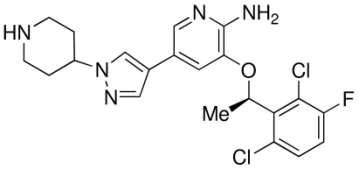
A) Les structures RX expérimentales des complexes de c-Met.
Elles sont classifiées par ordre de publication sur la PDB de la plus ancienne à la plus récente. D'autres ligands c-Met connus sont présentés lorsqu'ils sont reliés au ligand issus du complexe associé. Ce dernier est toujours présenté en premier dans ce tableau.

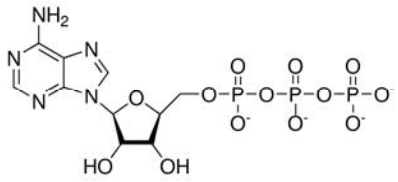
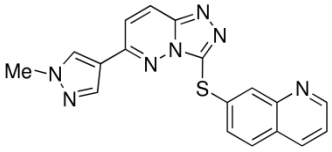
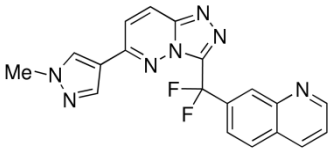
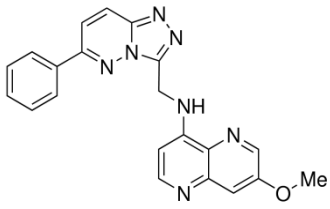
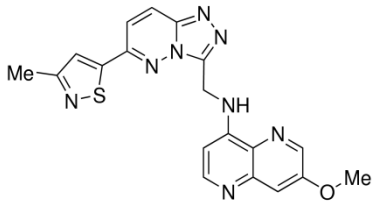
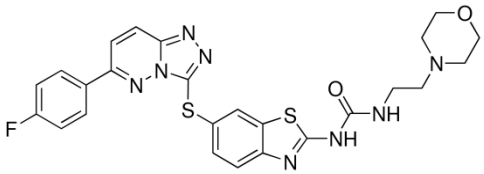
PDB entry Resolution Release date	Ligand formula Compound information, related compounds
PDB in patent 1.22 Å 2002/09	 in X-ray: incomplete Agouron (now Pfizer)-patented compound (EP/2002/1243596): 
1R0P [1] 1.80 Å 2003/10	 K-252a staurosporine analog [2-6], WO/1991/09034  Most promising compound from ArQule patent WO/2006/086484. ARQ-197 Met-targeting anticancer drug (structure not disclosed yet) is currently in phase II.
2RFN [7] 2.50 Å 2007/11	 Amgen, WO/2006/060318 (ex.1)

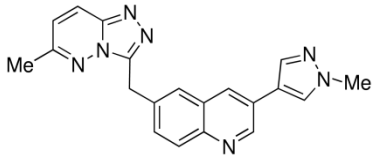
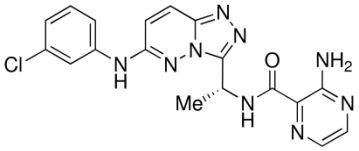
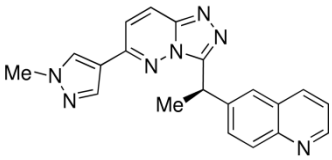
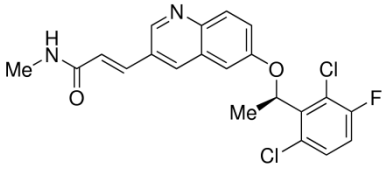
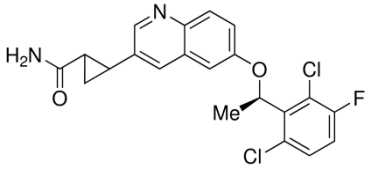
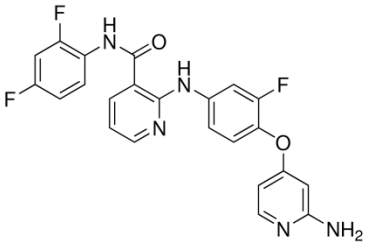
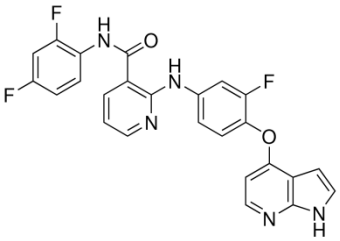
	 AMG 458 [8-9], WO/2006/116713, discontinued [10] (reacts with glutathione)  Potent AMG-458 analog reported with no unwanted *in vivo* reactivity [11]
2RFS [7] 2.20 Å 2007/11	 in X-ray: incomplete SU11274 [12-17], Pfizer (former Sugen), WO/2005/082411:  SU5416 / Semaxanib [18-20], WO/2002/055517, failed phase I:  SU11274 is also related to PHA-665752 (see 2WKM, WO/2002/096361). Sugen patented other sulfonylindolinone-based c-Met inhibitors (see WO/2005/005378).
3C1X [21] 2.17 Å 2008/03	 Bristol-Myers Squibb, WO/2006/004636 (ex.155)

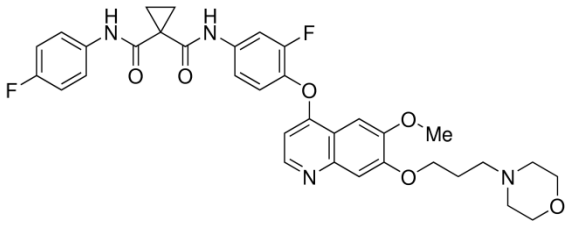
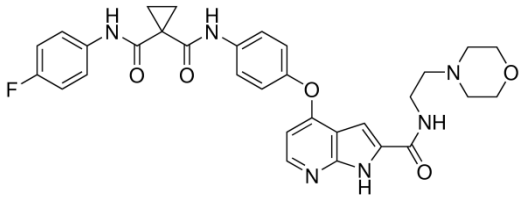
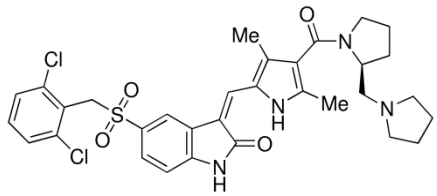
3CCN [22-23] 1.90 Å 2008/04	 Compound based on Sugen patent WO/2005/010005 including:  Development of this series led to Pfizer phase I Met-targeting drug:  PF-4217903 [23], WO/2007/132308, WO/2007/138472, WO/2009/068955
3CD8 [22-23] 2.00 Å 2008/04	 Amgen, WO/2008/008539 (ex.123) This compound was made on the basis of 3CCN unexpected binding mode [22].
3CTH [24] 2.30 Å 2008/06	 Bristol-Myers Squibb, WO/2005/117867 (ex.24)
3CTJ [24] 2.50 Å 2008/06	 Bristol-Myers Squibb, WO/2006/004636 (ex.132) 3C1X, 3CTH and 3CTJ are close analogs. Related patents: WO/2005/116028, WO/2005/117867, WO/2006/004833.

3CE3 [25] 2.40 Å 2008/08	 Bristol-Myers Squibb, WO/2006/004636 (ex.242) This class of compounds (see also WO/2007/035428 and WO/2007/065100) led to BMS-777607 (3F82).
3EFJ [26] 2.60 Å 2008/10	 Amgen, WO/2006/060318 (ex.4)
3EFK [26] 2.20 Å 2008/10	 Amgen, WO/2006/060318 (ex.10) 2RFN, 3EFJ and 3EFK are part of the same series [26], in which the most potent compound (patent ex.17) is: 
3F66 [27] 1.40 Å 2008/12	 rem: co-crystallized with γ-butyrolactone This compound is part of a UCB Celltech quinoxaline series in which the most interesting drug candidate as a c-Met inhibitor is:

	
3F82 [28] 2.50 Å 2009/03	 BMS-777607, Bristol-Myers Squibb, WO/2008/058229, WO/2009/094427, phase I  BMS-817378, Bristol-Myers Squibb, WO/2009/094417, WO/2009/094427
2WD1 [29] 2.00 Å 2009/04	 rem: co-crystallized with γ-butyrolactone The most interesting analog drug candidate [29] using this 4-azaindole scaffold is:  It features a 4-pyrazol-1-ylpiperidine also found on phase II drug PF-2341066 (see 2WGJ), and a 4-methyl-2-benzofurazane as with the best 3F66 analog originating from the same UCB Celltech group.
2WGJ [23] 2.00 Å 2009/06	 PF-2341066 [23, 30-31], Pfizer, WO/2006/02188(1,4,5,6), WO/2007/06618(5,7), phase II

3DKC [32] 1.52 Å 2009/07	 co-crystallized with Mg^{2+} c-Met kinase native partner ATP
3DKF / 3DKG [32] 1.80 / 1.91 Å 2009/07	 3DKF and 3DKG: same ligand SGX-523 [32-33], SGX (now Eli Lilly), WO/2008/051808, failed phase I (metabolite crystallized in patient kidneys)  JNJ-38877605, Johnson & Johnson, WO/2007/075567 (ex.61), still in phase I
3I5N [34] 2.00 Å 2010/01	 Amgen, WO/2008/008539 (ex.161) This molecule is part of a series of 3CD8-based analogues [34], out of which the most promising compound (same patent, ex.4) is:  Triazolopyridazines form a popular c-Met inhibitor class in the pharma industry. Some recent examples (see also 3CCN, 3CD8, 3DKF/G):  Sanofi-Aventis, WO/2009/056692 (ex.9)

	 Eli Lilly (from SGX), WO/2008/051805 (ex.124)  Vertex Pharmaceuticals, WO/2008/036272 (ex.4) – see also WO/2007/064797  Pfizer, WO/2007/138472 (ex.5)
3A4P [35] 2.54 Å 2010/02	 Chugai Pharmaceutical, designed from the c-Met / PF-2341066 complex crystal structure (2WGJ). A more potent analogue is: 
3L8V [36] 2.40 Å 2010/05	 Biarylamine-based compound from Bristol-Myers Squibb. More potent analogue: 

3LQ8 [37] 2.02 Å 2010/05	 XL880 *a.k.a.* GSK1363089 / EXEL-2880 / Foretinib, Exelixis / GlaxoSmithKline, MET/VEGFR2 kinase inhibitor, in phase II  XL-184 *a.k.a.* BMS-907351, Exelixis / Bristol-Myers Squibb, MET/RET/VEGFR2 kinase inhibitor, in phase III
2WKM 2.20 Å 2010/08	 PHA-665752 [38-45], Pfizer, WO/2002/096361 (ex.201)

1. Schiering, N.; Knapp, S.; Marconi, M.; Flocco, M. M.; Cui, J.; Perego, R.; Rusconi, L.; Cristiani, C., Crystal structure of the tyrosine kinase domain of the hepatocyte growth factor receptor c-Met and its complex with the microbial alkaloid K-252a. *PNAS* **2003,** *100* (22), 12654-12659.

2. Delsite, R.; Djakiew, D., Anti-proliferative effect of the kinase inhibitor K252a on human prostatic carcinoma cell lines. *J Androl* **1996,** *17* (5), 481-490.

3. Morotti, A.; Mila, S.; Accornero, P.; Tigliabue, E.; Ponzetto, C., K252a inhibits the oncogenic properties of Met, the HGF receptor. *Oncogene* **2002,** *21* (32), 4885-4893.

4. Cristiani, C.; Rusconi, L.; Perego, R.; Schiering, N.; Kalisz, H. M.; Knapp, S.; Isacchi, A., Regulation of the wild-type and Y1235D mutant met kinase activation. *Biochemistry* **2005,** *44* (43), 14110-14119.

5. Kase, H.; Iwahashi, K.; Matsuda, Y., K-252a, a potent inhibitor of protein kinase C from microbial origin. *J Antibiot* **1986,** *39* (8), 1059-1065.

6. Kino, M.; Shono, K.; Nishimura, T.; Nagamura, S., Practical preparation of K-252a from a fermentation solution. *Biosci Biotech Bioch* **1998,** *62* (8), 1627-1629.

7. Bellon, S. F.; Kaplan-Lefko, P.; Yang, Y.; Zhang, Y.; Moriguchi, J.; Rex, K.; Johnson, C. W.; Rose, P. E.; Long, A. M.; O'Connor, A. B.; Gu, Y.; Coxon, A.; Kim, T.-S.; Tasker, A.; Burgess, T. L.; Dussault, I., c-Met inhibitors with novel binding mode show activity against several hereditary papillary renal cell carcinoma-related mutations. *J Biol Chem* **2008,** *283* (5), 2675-2683.

8. Liu, L.; Siegmund, A.; Xi, N.; Kaplan-Lefko, P.; Rex, K.; Chen, A.; Lin, J.; Moriguchi, J.; Berry, L.; Huang, L.; Teffera, Y.; Yang, Y.; Zhang, Y.; Bellon, S. F.; Lee, M.; Shimanovich, R.; Bak, A.; Dominguez, C.; Norman, M. H.; Harmange, J.-C.; Dussault, I.; Kim, T.-S., Discovery of a potent, selective, and orally bioavailable c-Met inhibitor: 1-(2-hydroxy-2-methylpropyl)-N-(5-(7-methoxyquinolin-4-yloxy)pyridin-2-yl)-5-methyl-3-oxo-2-phenyl-2,3-dihydro-1H-pyrazole-4-carboxamide (AMG 458). *J Med Chem* **2008,** *51* (13), 3688-3691.

9. Siegmund, A.; Retz, D.; Xi, N.; Dominguez, C.; Bürli, R.; Liu, L., Selective β-hydroxyethylation at the N-1 position of a pyrazolone : Synthesis of benzyl 1-(β-hydroxyethyl)-5-methyl-3-oxo-2-phenyl-2,3-dihydro-1H-pyrazole-4-carboxylate. *Synlett* **2008,** *7*, 1005-1008.

10. Teffera, Y.; Colletti, A. E.; Harmange, J. C.; Hollis, L. S.; Albrecht, B. K.; Boezio, A. A.; Liu, J.; Zhao, Z., Chemical reactivity of methoxy 4-O-aryl quinolines: Identification of glutathione displacement products in vitro and in vivo. *Chemical Research in Toxicology* **2008,** *21* (11), 2216-2222.

11. Zhang, Y.; Kaplan-Lefko, P. J.; Rex, K.; Yang, Y.; Moriguchi, J.; Osgood, T.; Mattson, B.; Coxon, A.; Reese, M.; Kim, T.-S.; Lin, J.; Chen, A.; Burgess, T. L.; Dussault, I., Identification of a novel Recepteur d'Origine Nantais/c-Met small-molecule kinase inhibitor with antitumor activity in vivo. *Cancer Res* **2008,** *68* (16), 6680-6687.

12. Sattler, M.; Pride, Y. B.; Ma, P.; Gramlich, J. L.; Chu, S. C.; Quinnan, L. A.; Shirazian, S.; Liang, C.; Polar, K.; Christensen, J. G.; Salgia, R., A novel small molecule Met inhibitor induces apoptosis in cells transformed by the oncogenic TPR-Met tyrosine kinase. *Cancer Res* **2003,** *63*, 5462-5469.

13. Wang, X.; Le, P.; Liang, C.; Chan, J.; Kiewlich, D.; Miller, T.; Harris, C.; Sun, L.; Rice, A.; Vasile, S.; Blake, R. A.; Howlett, A. R.; Patel, N.; McMahon, G.; Lipson, K. E., Potent and selective inhibitors of the Met [hepatocyte growth factor/scatter factor (HGF/SF) receptor] tyrosine kinase block HGF/SF-induced tumor cell growth and invasion. *Mol Cancer Ther* **2003,** *2*, 1085-1092.

14. Berthou, S.; Aebersold, D. M.; Schmidt, L. S.; Stroka, D.; Heigl, C.; Streit, B.; Stalder, D.; Gruber, G.; Liang, C.; Howlett, A. R.; Candinas, D.; Greiner, R. H.; Lipson, K. E.; Zimmer, Y., The Met kinase inhibitor SU11274 exhibits a selective inhibition pattern toward different receptor mutated variants. *Oncogene* **2004,** *23*, 5387-5393.

15. Ma, P. C.; Jagadeeswaran, R.; Jagadeesh, S.; Tretiakova, M. S.; Nullasura, V.; Fox, E. A.; Hansen, M.; Schaefer, E.; Naoki, K.; Lader, A.; Richards, W.; Sugarbaker, D.; Husain, A. N.; Christensen, J. G.; Salgia, R.,

Functional expression and mutations of c-Met and its therapeutic inhibition with SU11274 and small interfering RNA in non-small cell lung cancer. *Cancer Res* **2005,** *65* (4), 1479-1488.

16. Koon, E. C.; Ma, P. C.; Salgia, R.; Welch, W. R.; Christensen, J. G.; Berkowitz, R. S.; Mok, S. C., Effect of a c-Met-specific, ATP-competitive small-molecule inhibitor SU11274 on human ovarian carcinoma cell growth, motility, and invasion. *Int J Gynecol Cancer* **2007,** *18* (5), 976-984.

17. Puri, N.; Ahmed, S.; Janamanchi, V.; Tretiakova, M.; Zumba, O.; Krausz, T.; Jagadeeswaran, R.; Salgia, R., c-Met is a potentially new therapeutic target for treatment of human melanoma. *Clin Cancer Res* **2007,** *13* (7), 2246-2253.

18. Wang, S. Y.; Chen, B.; Zhan, Y. Q.; Wu, W. X.; Li, C. Y.; Yang, R. F.; Zheng, H.; Yue, P. B.; Larsen, S. H.; Sun, H. B.; Yang, X., SU5416 is a potent inhibitor of hepatocyte growth factor receptor (c-Met) and blocks HGF-induced invasiveness of human HepG2 hepatoma cells. *J Hepatol* **2004,** *41* (2), 267-273.

19. Hoff, P. M.; Wolff, R. A.; Bogaard, K.; Waldrum, S.; Abbruzzese, J. L., A phase I study of escalating doses of the tyrosine kinase inhibitor Semaxanib (SU5416) in combination with Ironotecan in patients with advanced colorectal carcinoma. *Jpn J Clin Oncol* **2006,** *36* (2), 100-103.

20. Sun, L.; Tran, N.; Tang, F.; App, H.; Hirth, P.; McMahon, G.; Tang, C., Synthesis and biological evaluations of 3-substituted indolin-2-ones: a novel class of tyrosine kinase inhibitors that exhibit selectivity toward particular receptor tyrosine kinases. *J Med Chem* **1998,** *41* (14), 2588-2603.

21. Schroeder, G. M.; Chen, X.-T.; Williams, D. K.; Nirschl, D. S.; Cai, Z.-W.; Wei, D.; Tokarski, J. S.; An, Y.; Sack, J.; Chen, Z.; Huynh, T.; Vaccaro, W.; Poss, M.; Wautlet, B.; Gullo-Brown, J.; Kellar, K.; Manne, V.; Hunt, J. T.; Wong, T. W.; Lombardo, L. J.; Fargnoli, J.; Borzilleri, R. M., Identification of pyrrolo[2,1-f][1,2,4]triazine-based inhibitors of Met kinase. *Bioorg Med Chem Lett* **2008,** *18* (6), 1945-1951.

22. Albrecht, B. K.; Harmange, J.-C.; Bauer, D.; Berry, L.; Bode, C.; Boezio, A. A.; Chen, A.; Choquette, D.; Dussault, I.; Fridrich, C.; Hirai, S.; Hoffman, D.; Larrow, J. F.; Kaplan-Lefko, P.; Lin, J.; Lohman, J.; Long, A. M.; Moriguchi, J.; O'Connor, A.; Potashman, M. H.; Reese, M.; Rex, K.; Siegmund, A.; Shah, K.; Shimanovich, R.; Springer, S. K.; Teffera, Y.; Yang, Y.; Zhang, Y.; Bellon, S. F., Discovery and optimization of triazolopyridazines as potent and selective inhibitors of the c-Met kinase. *J Med Chem* **2008,** *51* (10), 2879-2882.

23. Timofeevski, S. L.; McTigue, M. A.; Ryan, K.; Cui, J.; Zou, H. Y.; Zhu, J. X.; Chau, F.; Alton, G.; Karlicek, S.; Christensen, J. G.; Murray, B. W., Enzymatic characterization of c-Met receptor tyrosine kinase oncogenic mutants and kinetic studies with aminopyridine and triazolopyrazine inhibitors. *Biochemistry* **2009,** *48* (23), 5339-5349.

24. Cai, Z.-W.; Wei, D.; Schroeder, G. M.; Cornelius, L. A. M.; Kim, K.; Chen, X.-T.; Schmidt, R. J.; Williams, D. K.; Tokarski, J. S.; An, Y.; Sack, J. S.; Manne, V.; Kamath, A.; Zhang, Y.; Marathe, P.; Hunt, J. T.; Lombardo, L. J.; Fargnoli, J.; Borzilleri, R. M., Discovery of orally active pyrrolopyridine- and aminopyridine-based Met kinase inhibitors. *Bioorg Med Chem Lett* **2008,** *18* (11), 3224-3229.

25. Kim, K. S.; Zhang, L.; Schmidt, R.; Cai, Z.-W.; Wei, D.; Williams, D. K.; Lombardo, L. J.; Trainor, G. L.; Xie, D.; Zhang, Y.; An, Y.; Sack, J. S.; Tokarski, J. S.; Darienzo, C.; Kamath, A.; Marathe, P.; Zhang, Y.; Lippy, J.; Jeyaseelan, R. S.; Wautlet, B.; Hentley, B.; Gullo-Brown, J.; Manne, V.; Hunt, J. T.; Fargnoli, J.; Borzilleri, R. M., Discovery of pyrrolopyridine-pyridone based inhibitors of Met kinase: synthesis, X-ray crystallographic analysis, and biological activities. *J Med Chem* **2008,** *51* (17), 5330-5341.

26. D'Angelo, N. D.; Bellon, S. F.; Booker, S. K.; Cheng, Y.; Coxon, A.; Dominguez, C.; Fellows, I.; Hoffman, D.; Hungate, R.; Kaplan-Lefko, P.; Lee, M. R.; Li, C.; Liu, C.; Rainbeau, E.; Reider, P. J.; Rex, K.; Siegmund, A.; Sun, Y.; Tasker, A. S.; Xi, N.; Xu, S.; Yang, Y.; Zhang, Y.; Burgess, T. L.; Dussault, I.; Kim, T.-S., Design, synthesis, and biological evaluation of potent c-Met inhibitors. *J Med Chem* **2008,** *51* (18), 5766-5779.

27. Porter, J.; Lumb, S.; Lecomte, F.; Reuberson, J.; Foley, A.; Calmiano, M.; le Riche, K.; Edwards, H.; Delgado, J.; Franklin, R. J.; Gascon-Simorte, J. M.; Maloney, A.; Meier, C.; Batchelor, M., Discovery of a novel series of quinoxalines as inhibitors of c-Met kinase. *Bioorg Med Chem Lett* **2009,** *19* (2), 397-400.

28. Schroeder, G. M.; An, Y.; Cai, Z.-W.; Chen, X.-T.; Clark, C.; Cornelius, L. A. M.; Dai, J.; Gullo-Brown, J.; Gupta, A.; Henley, B.; Hunt, J. T.; Jeyaseelan, R.; Kamath, A.; Kim, K.; Lippy, J.; Lombardo, L. J.; Manne, V.; Oppenheimer, S.; Sack, J. S.; Schmidt, R. J.; Shen, G.; Stefanski, K.; Tokarski, J. S.; Trainor, G. L.; Wautlet, B. S.; Wei, D.; Williams, D. K.; Zhang, Y.; Zhang, Y.; Fargnoli, J.; Borzilleri, R. M., Discovery of N-(4-(2-Amino-3-chloropyridin-4-yloxy)-3-fluorophenyl)-4-ethoxy-1-(4-fluorophenyl)-2-oxo-1,2-dihydropyridine-3-carboxamide (BMS-777607), a selective and orally efficacious inhibitor of the Met kinase superfamily. *J Med Chem* **2009,** *52* (5), 1251–1254.

29. Porter, J.; Lumb, S.; Franklin, R. J.; Gascon-Simorte, J. M.; Calmiano, M.; Le Riche, K.; Lallemand, B.; Keyaerts, J.; Edwards, H.; Maloney, A.; Delgado, J.; King, L.; Foley, A.; Lecomte, F.; Reuberson, J.; Meier, C.; Batchelor, M., Discovery of 4-azaindoles as novel inhibitors of c-Met kinase. *Bioorg Med Chem Lett* **2009,** *19* (10), 2780-2784.

30. Christensen, J. G.; Zou, H. Y.; Arango, M. E.; Li, Q.; Lee, J. H.; McDonnell, S. R.; Yamazaki, S.; Alton, G. R.; Mroczkowski, B.; Los, G., Cytoreductive antitumor activity of PF-2341066, a novel inhibitor of anaplastic lymphoma kinase and c-Met, in experimental models of anaplastic large-cell lymphoma. *Mol Cancer Ther* **2007,** *6* (12 pt.1), 3314-3322.

31. Zou, H. Y.; Li, Q.; Lee, J. H.; Arango, M. E.; McDonnell, S. R.; Yamazaki, S.; Koudriakova, T. B.; Alton, G.; Cui, J. J.; Kung, P.-P.; Nambu, M. D.; Los, G.; Bender, S. L.; Mroczkowski, B.; Christensen, J. G., An orally available small-molecule inhibitor of c-Met, PF-2341066, exhibits cytoreductive antitumor efficacy through antiproliferative and antiangiogenic mechanisms. *Cancer Res* **2007,** *67* (9), 4408-4417.

32. Buchanan, S. G.; Hendle, J.; Lee, P. S.; Smith, C. R.; Bounaud, P.-Y.; Jessen, K. A.; Tang, C. M.; Huser, N. H.; Felce, J. D.; Froning, K. J.; Peterman, M. C.; Aubol, B. E.; Gessert, S. F.; Sauder, J. M.; Schwinn, K. D.; Russell, M.; Rooney, I. A.; Adams, J.; Leon, B. C.; Do, T. H.; Blaney, J. M.; Sprengeler, P. A.; Thomson, D. A.; Smyth, L.; Pelletier, L. A.; Atwell, S.; Holme, K.; Wasserman, S. R.; Emtage, S.; Burlay, S. K.; Reich, S. H., SGX523 is an exquisitely selective, ATP-competitive inhibitor of the MET receptor tyrosine kinase with antitumor activity in vivo. *Mol Cancer Ther* **2009,** *8* (12), 3181-3190.

33. Guessous, F.; Zhang, Y.; di Pierro, C.; Marcinkiewicz, L.; Sarkaria, J.; Schiff, D.; Buchanan, S.; Abounader, R., An orally bioavailable c-Met kinase inhibitor potently inhibits brain tumor malignancy and growth. *Anti-Cancer Agents in Medicinal Chemistry* **2010,** *10* (1), 28-35.

34. Boezio, A. A.; Berry, L.; Albrecht, B. K.; Bauer, D.; Bellon, S. F.; Bode, C.; Chen, A.; Choquette, D.; Dussault, I.; Hirai, S.; Kaplan-Lefko, P.; Larrow, J. F.; Lin, M. H. J.; Lohman, J.; Potashman, M. H.; Rex, K.; Santostefano, M.; Shah, K.; Shimanovich, R.; Springer, S. K.; Teffera, Y.; Yang, Y. J.; Zhang, Y. H.; Harmange, J. C., Discovery and optimization of potent and selective triazolopyridazine series of c-Met inhibitors. *Bioorg Med Chem Lett* **2009,** *19* (22), 6307-6312.

35. Nishii, H.; Chiba, T.; Morikami, K.; Fukami, T. A.; Sakamoto, H.; Ko, K.; Koyano, H., Discovery of 6-benzyloxyquinolines as c-Met selective kinase inhibitors. *Bioorg Med Chem Lett* **2010,** *20* (4), 1405-1409.

36. Williams, D. K.; Chen, X.-T.; Tarby, C.; Kaltenbach, R.; Cai, Z.-W.; Tokarski, J. S.; An, Y.; Sack, J. S.; Wautlet, B.; Gullo-Brown, J.; Henley, B. J.; Jeyaseelan, R.; Kellar, K.; Manne, V.; Trainor, G. L.; Lombardo, L. J.; Fargnoli, J.; Borzilleri, R. M., Design, synthesis and structure-activity relationships of novel biarylamine-based Met kinase inhitibors. *Bioorg Med Chem Lett* **2010,** *in the press*.

37. Qian, F.; Engst, S.; Yamaguchi, K.; Yu, P.; Won, K. A.; Mock, L.; Lou, T.; Tan, J.; Li, C.; Tam, D.; Lougheed, J.; Yakes, F. M.; Bentzien, F.; Xu, W.; Zaks, T.; Wooster, R.; Greshock, J.; Joly, A. H., Inhibition of tumor cell growth, invasion, and metastasis by EXEL-2880 (XL880, GSK1363089), a novel inhibitor of HGF and VEGF receptor tyrosine kinases. *Cancer Res* **2009,** *69* (20), 8009-8016.

38. Christensen, J. G.; Schreck, R.; Burrows, J.; Kuruganti, P.; Chan, E.; Le, P.; Chen, J.; Wang, X.; Ruslim, L.; Blake, R.; Lipson, K. E.; Ramphal, J.; Do, S.; Cui, J. J.; Cherrington, J. M.; Mendel, D. B., A selective small molecule inhibitor of c-Met kinase inhibits c-Met-dependent phenotypes in vitro and exhibits cytoreductive antitumor activity in vivo. *Cancer Res* **2003,** *63* (21), 7345-7355.

39. Ma, P. C.; Schaefer, E.; Christensen, J. G.; Salgia, R., A selective small molecule c-MET Inhibitor, PHA665752, cooperates with rapamycin. *Clin Cancer Res* **2005,** *11* (6), 2312-2319.

40. Mukohara, T.; Civiello, G.; Davis, I. J.; Taffaro, M. L.; Christensen, J.; Fischer, D. E.; Johnson, B. E.; Jänne, P. A., Inhibition of the met receptor in mesothelioma. *Clin Cancer Res* **2005,** *11* (22), 8122-8130.

41. Smolen, G. A.; Sordella, R.; Muir, B.; Mohapatra, G.; Barmettler, A.; Archibald, H.; Kim, W. J.; Okimoto, R. A.; Bell, D. W.; Sgroi, D. C.; Christensen, J. G.; Settleman, J.; Haber, D. A., Amplification of MET may identify a subset of cancers with extreme sensitivity to the selective tyrosine kinase inhibitor PHA-665752. *PNAS* **2006,** *103* (7), 2316-2321.

42. Watson, G. A.; Zhang, X.; Stang, M. T.; Levy, R. M.; Queiroz de Oliveira, P. E.; Gooding, W. E.; Christensen, J. G.; Hughes, S. J., Inhibition of c-Met as a therapeutic strategy for Esophageal Adenocarcinoma. *Neoplasia* **2006,** *8* (11), 949-955.

43. Puri, N.; Khramtsov, A.; Ahmed, S.; Nallasura, V.; Hetzel, J. T.; Jagadeeswaran, R.; Karczmar, G.; Salgia, R., A selective small molecule inhibitor of c-Met, PHA665752, inhibits tumorigenicity and angiogenesis in mouse lung cancer xenografts. *Cancer Res* **2007,** *67*, 3529-2524.

44. Chattopadhyay, C.; El-Naggar, A. K.; Williams, M. D.; Clayman, G. L., Small molecule c-MET inhibitor PHA665752: effect on cell growth and motility in papillary thyroid carcinoma. *Head Neck* **2008,** *30* (8), 991-1000.

45. Yang, Y.; Wislez, M.; Fujimoto, N.; Prudkin, L.; Izzo, J. G.; Uno, F.; Ji, L.; Hanna, A. E.; Langley, R. R.; Liu, D.; Johnson, F. M.; Wistuba, I.; Kurie, J. M., A selective small molecule inhibitor of c-Met, PHA-665752, reverses lung premalignancy induced by mutant K-ras. *Mol Cancer Ther* **2008,** *7* (4), 952-960.

B)

Méthodes de recherche conformationnelle

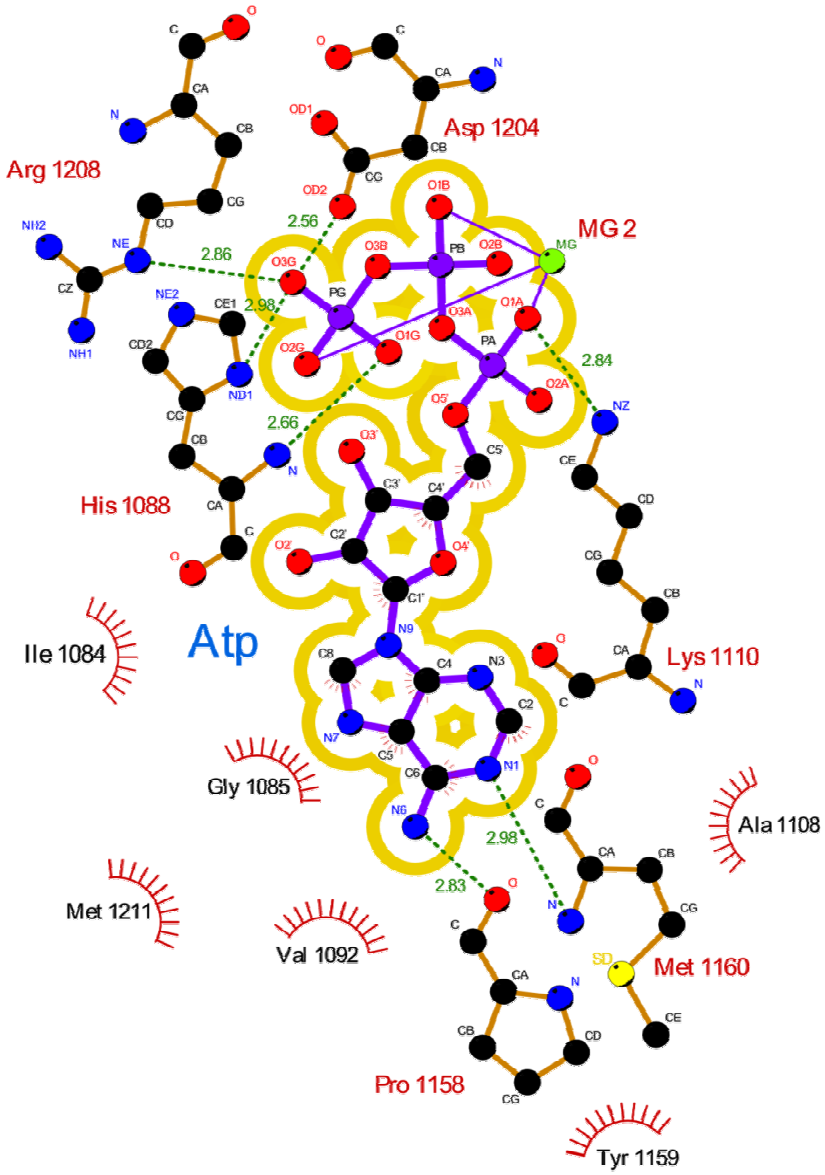
Recherche systématique

L'approche d'exploration d'une hypersurface la plus évidente consiste à en évaluer tous les états possibles. Dans le cas d'un système moléculaire, il s'agit de calculer l'énergie de chaque conformère en faisant varier les angles dièdres considérés par pas de *n* degrés. On suppose ici que les angles de valence et les longueurs de liaison sont fixés aux valeurs d'équilibre. Cette approximation est acceptable car les variations par rapport à l'état d'équilibre des termes correspondants sont bien plus importantes que celles des autres termes du potentiel décrits dans les équations de potentiel de liaisons et d'angles. Le nombre de conformères théoriques est alors de P^i, où P est le nombre d'états possibles par torsion pour chaque angle dièdre, et i est le nombre d'angles dièdres contenus dans la molécule

Par exemple, une recherche systématique sur une molécule possédant 4 angles dièdres avec un pas de n = 30°, générerait 20 736 conformères (P = 360/30 = 12 → 12^4 = 20 736). Une telle approche s'avère inutilisable, même sur des molécules relativement simples, car trop coûteuse en temps de calcul du fait de sa complexité exponentielle par rapport au nombre de dièdres. Les méthodes stochastiques ou déterministes sont alors privilégiées.

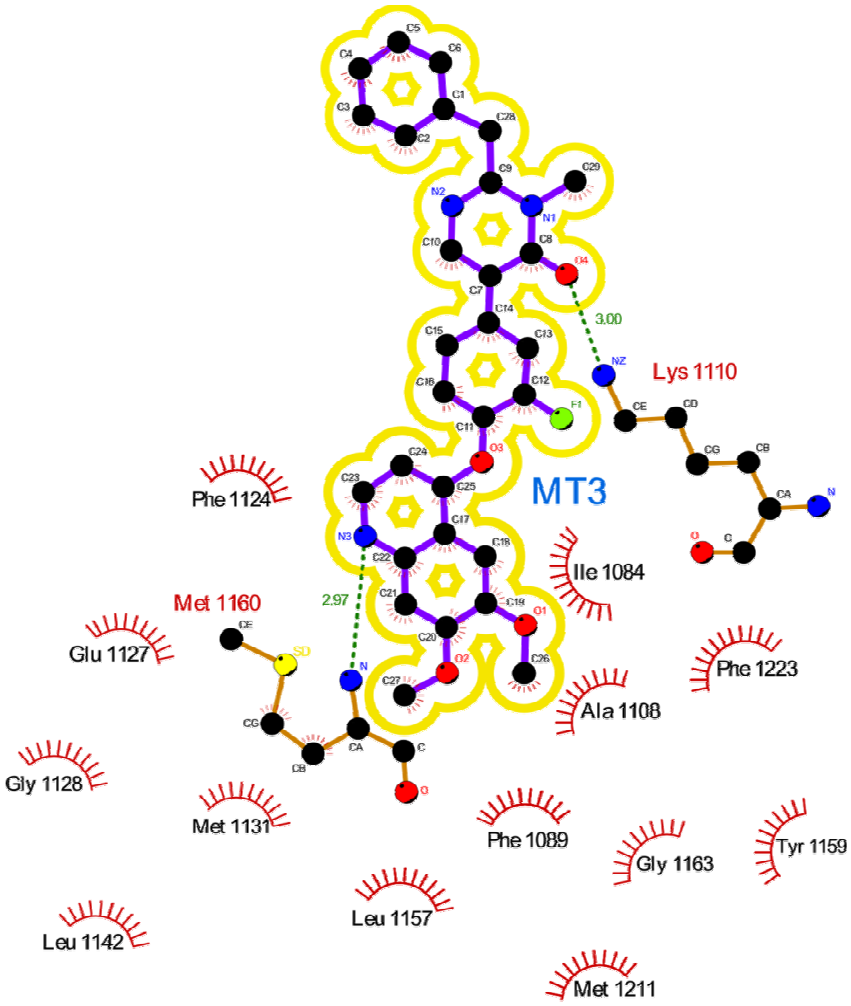
C) Liste des figures représentants les interactions entre les structures RX de c-Met issues de la PDB database et leur ligands respectifs. Les complexes sont rassemblés selon la classification établie sur notre premier article (Asses and al.).

ATP

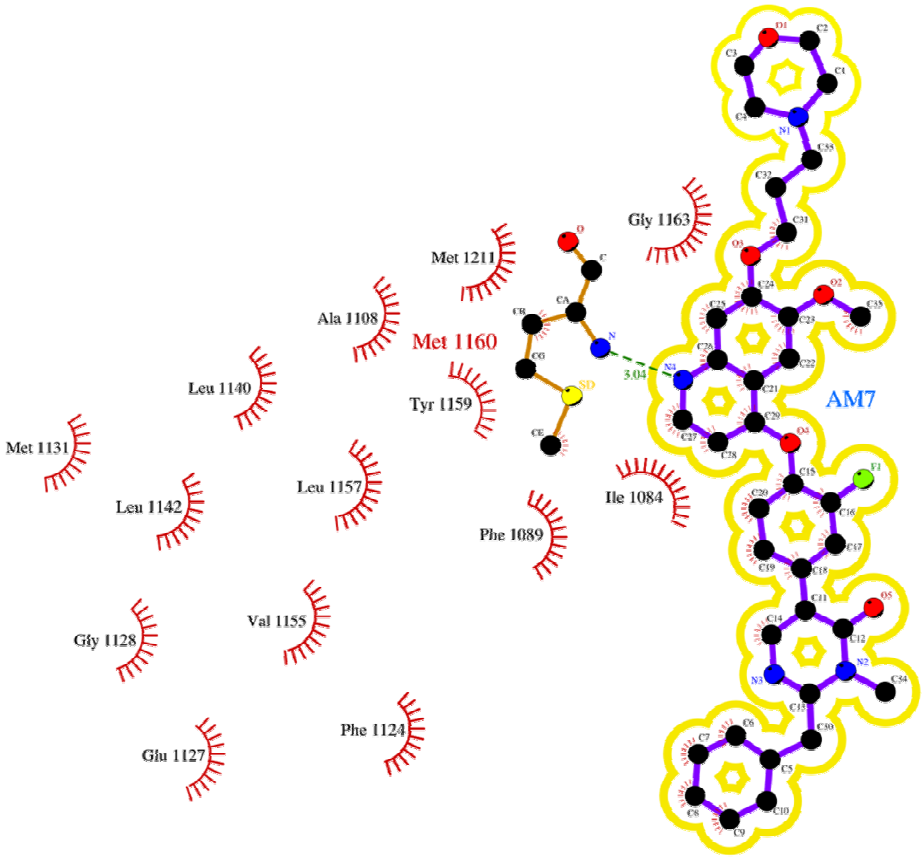


3DKC/Atp

C helix out

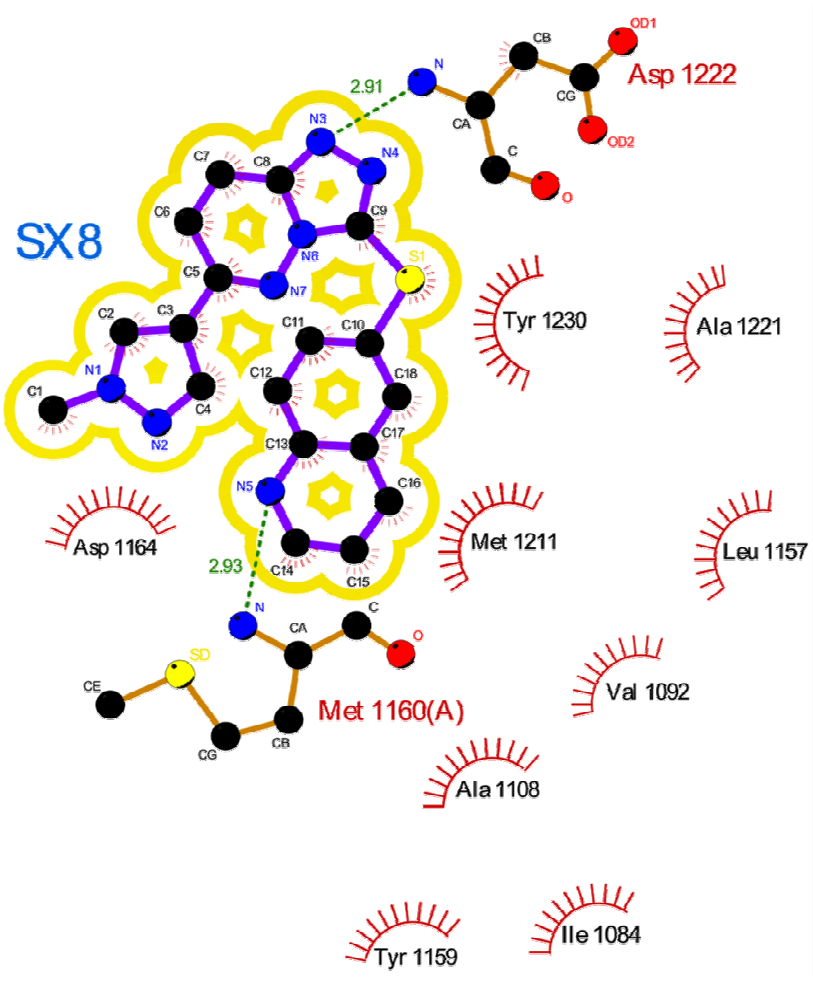


3EFJ/mt3

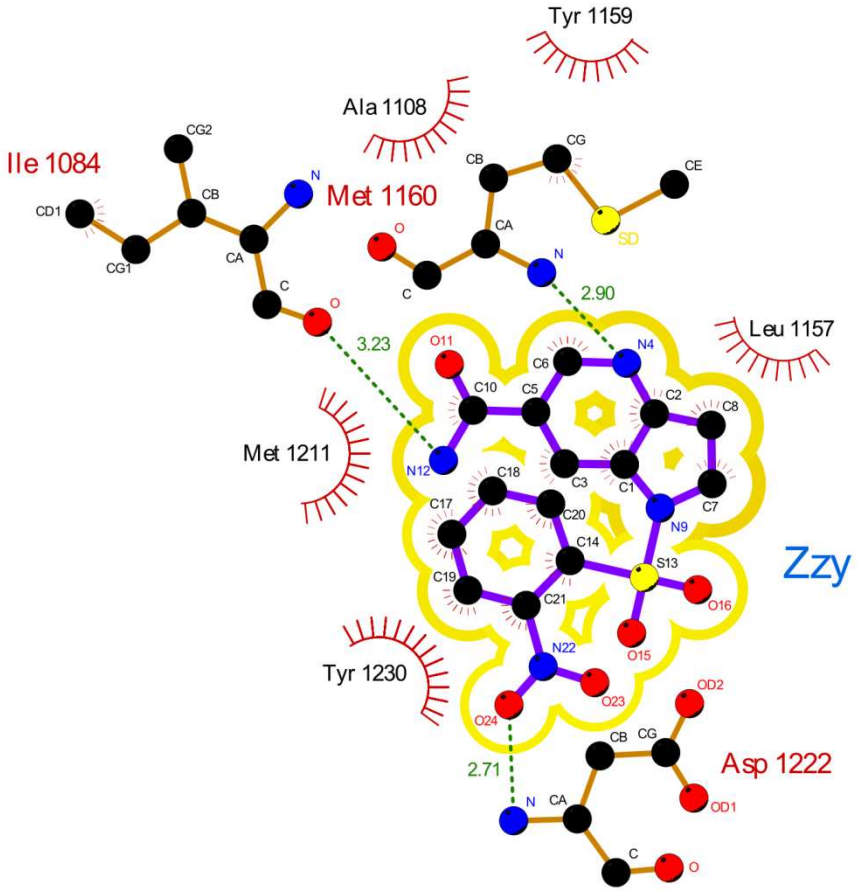

Gly 1163
Met 1211
Ala 1108
Met 1160
Tyr 1159
Leu 1140
Met 1131
Leu 1142
Leu 1157
Phe 1089
Ile 1084
AM7
Gly 1128
Val 1155
Phe 1124
Glu 1127
3.04

2RFN/am7

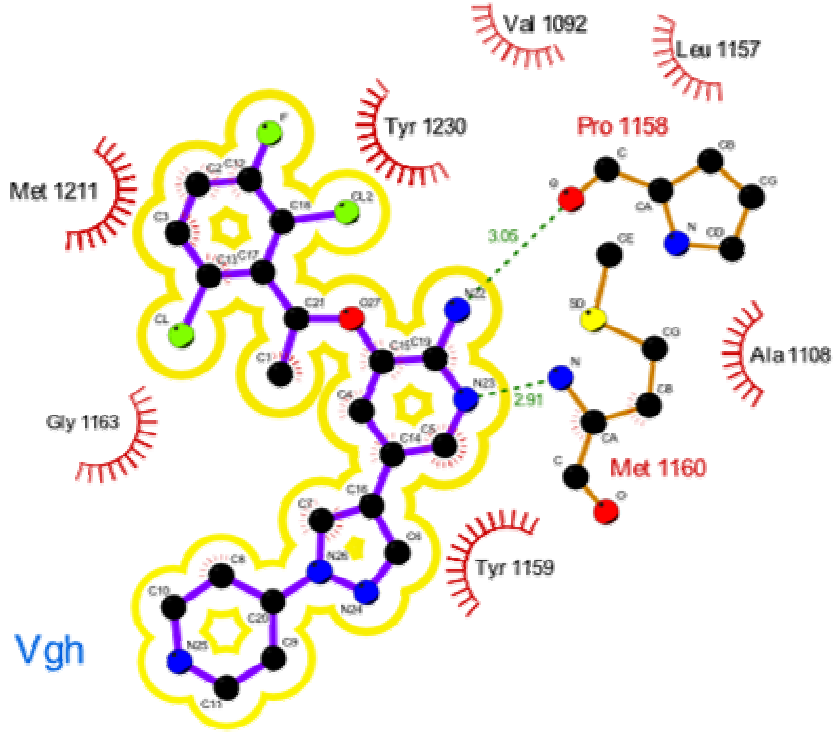
Type I



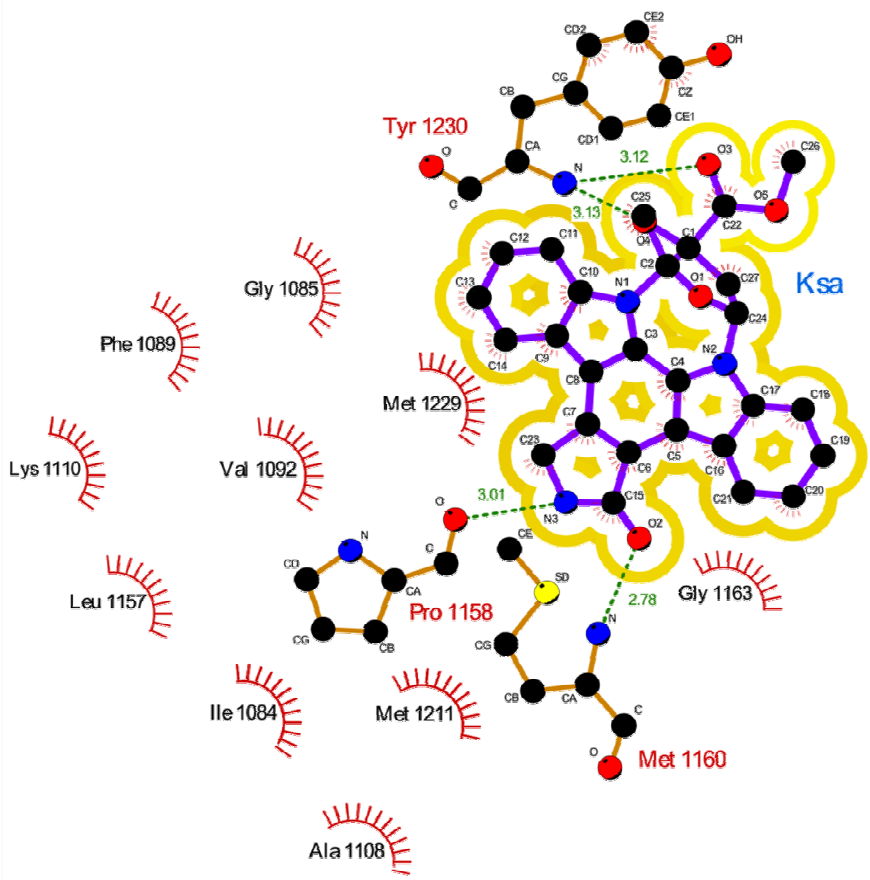
3DKF/sx8



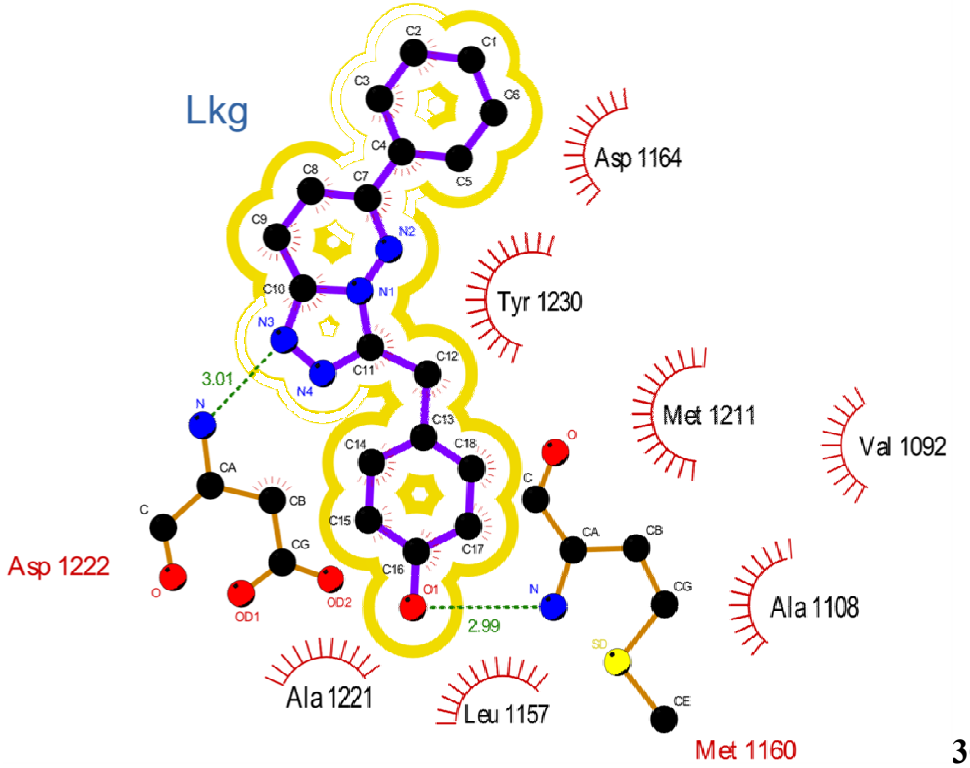
2WD1/zzy



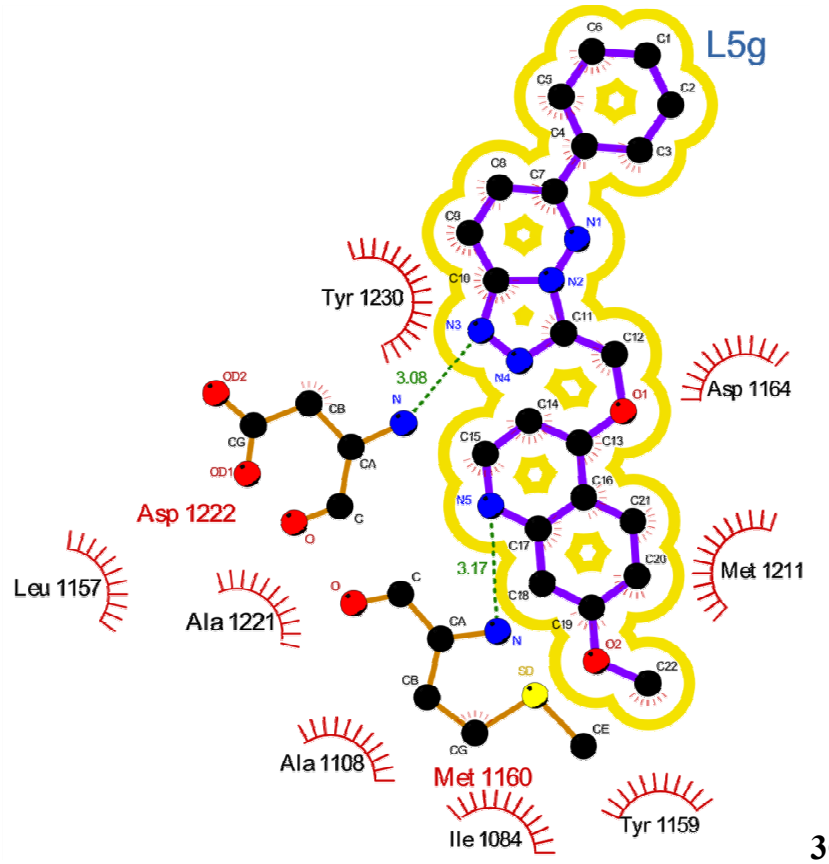
2WGJ/vgh



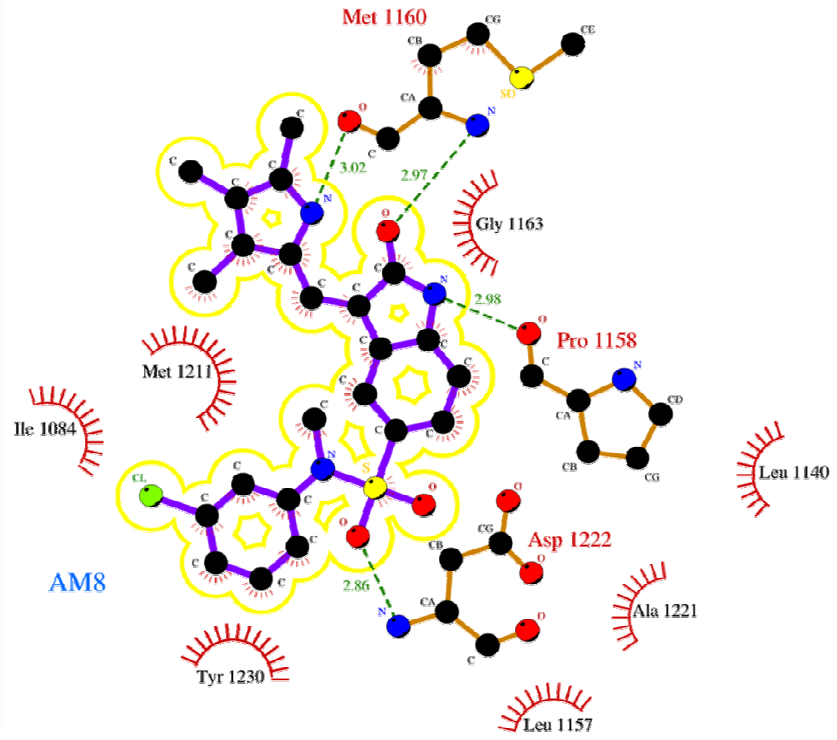
1R0P/ksa



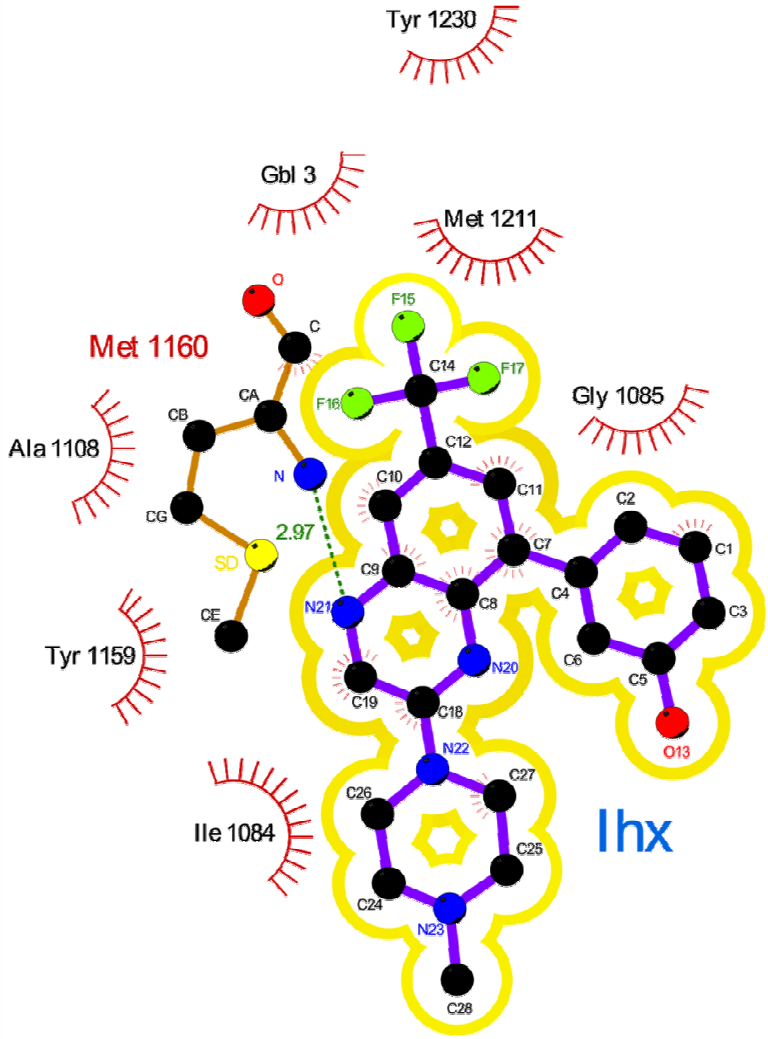
3CCN/lkg



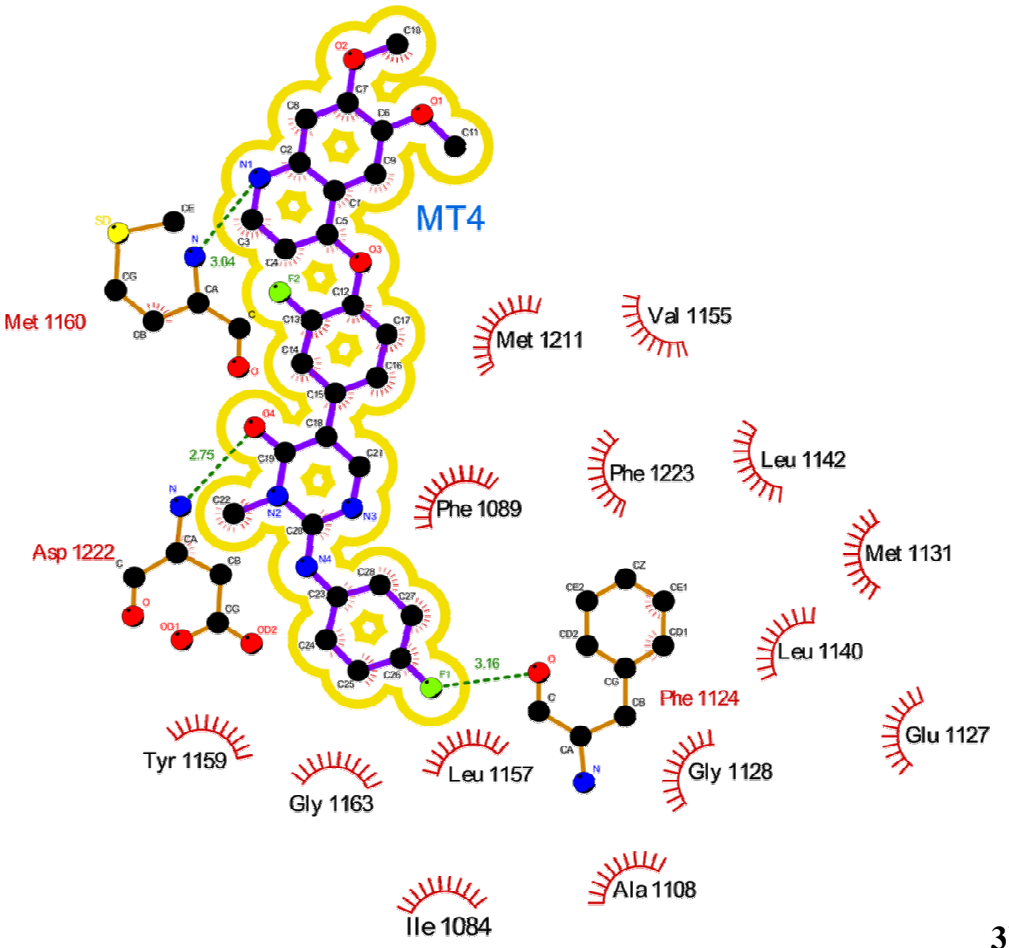
3CD8/l5g



2RFS/am8

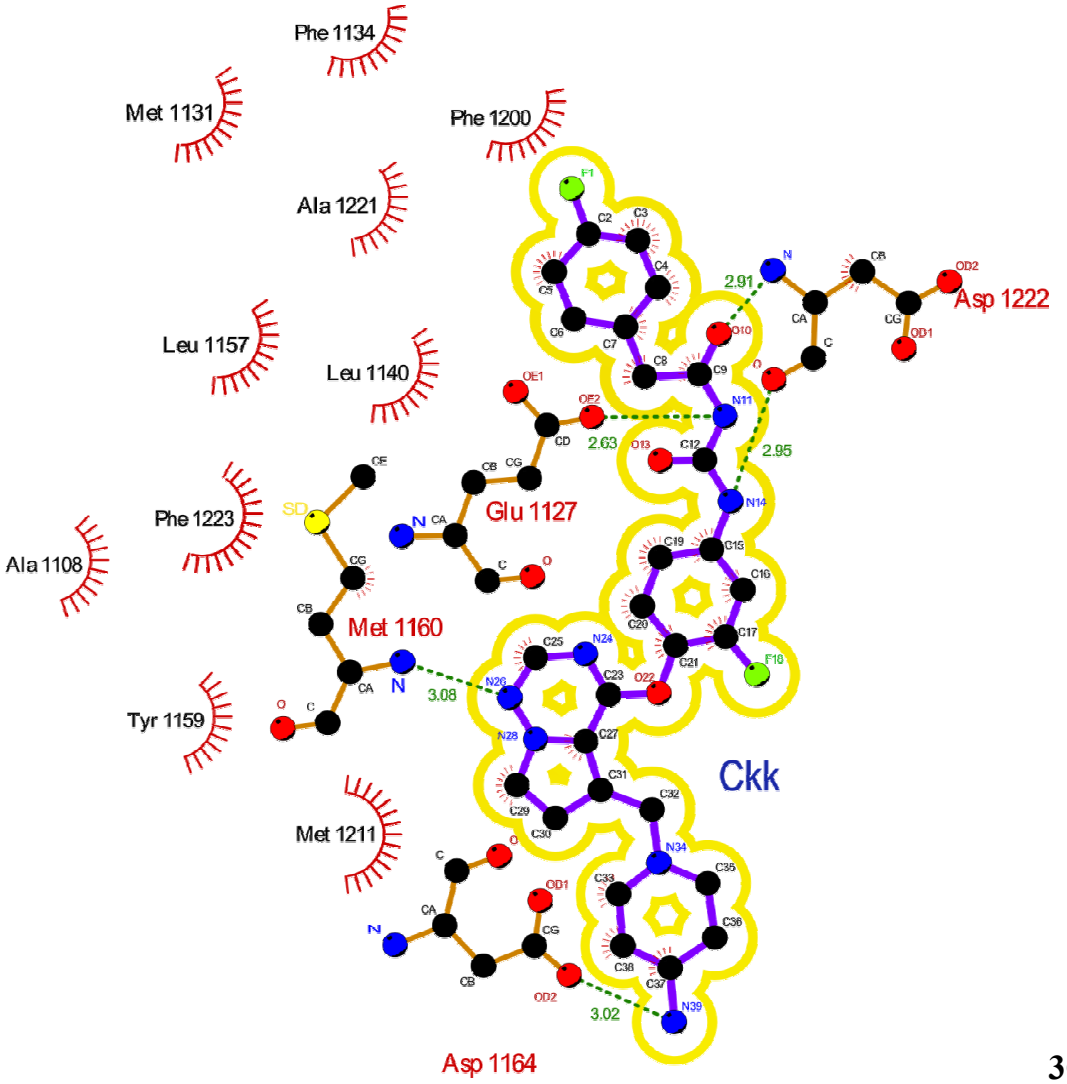

Tyr 1230
Gbl 3
Met 1211
Met 1160
Gly 1085
Ala 1108
Tyr 1159
Ile 1084
lhx
O
C
CA
CB
CG
SD
CE
N
2.97
F15
C14
F17
F16
C12
C10
C11
C2
C1
C7
C9
C8
C4
C3
N21
N20
C6
C5
C19
C18
O13
N22
C27
C26
C25
C24
N23
C28

3F66/ihx

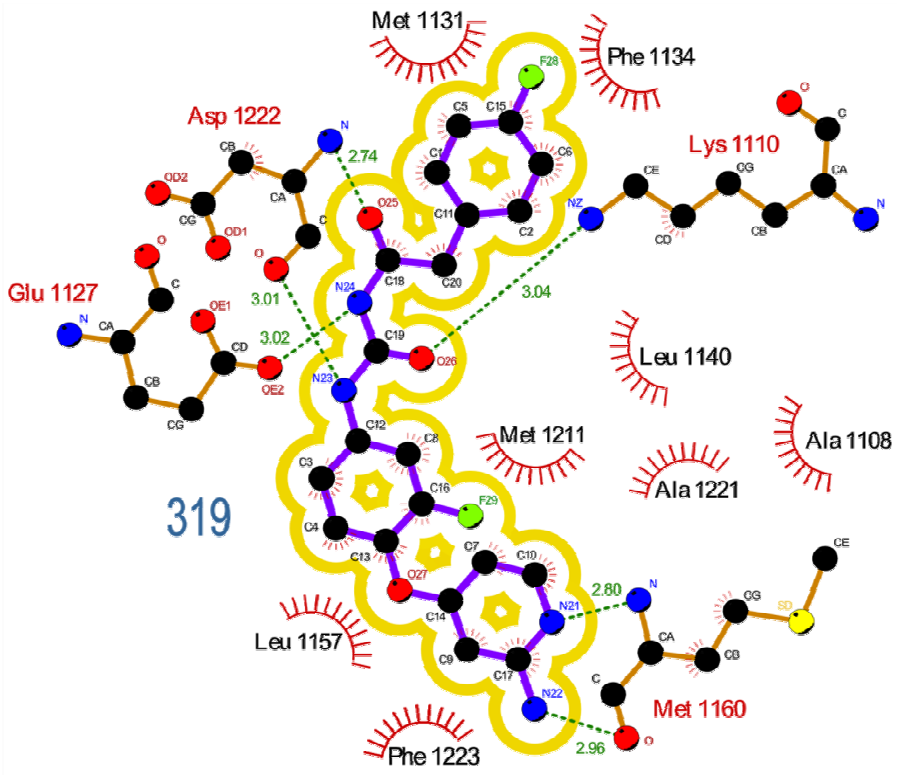

MT4
Met 1160
Asp 1222
Tyr 1159
Gly 1163
Leu 1157
Ile 1084
Ala 1108
Gly 1128
Phe 1124
Glu 1127
Leu 1140
Met 1131
Leu 1142
Phe 1223
Phe 1089
Met 1211
Val 1155
3.04
2.75
3.16

3EFK/mt4

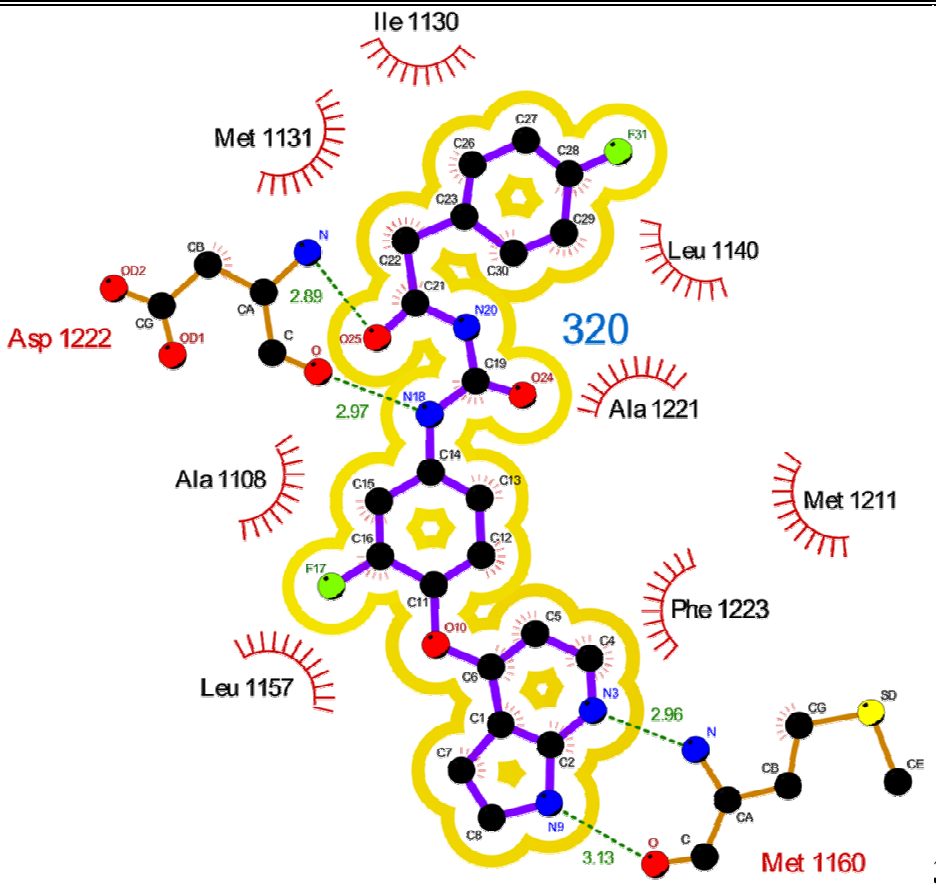
DFG out



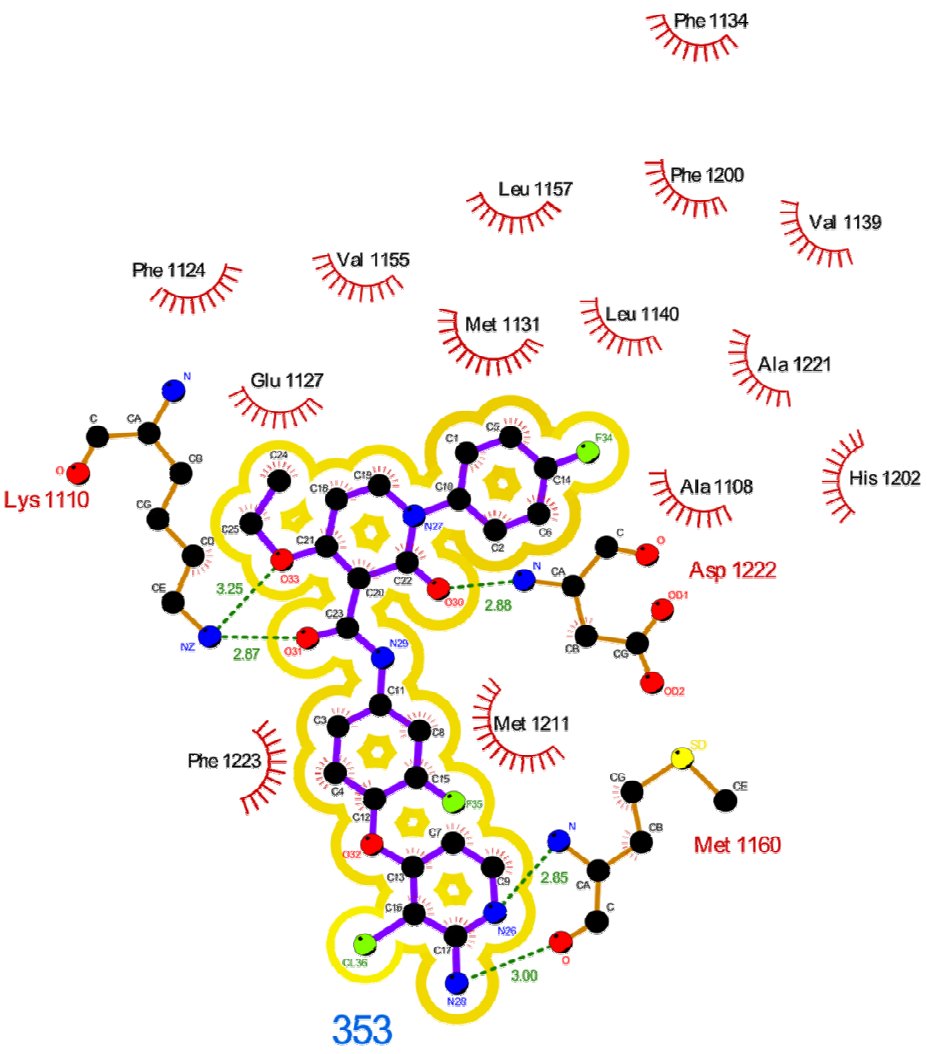
3C1X/ckk



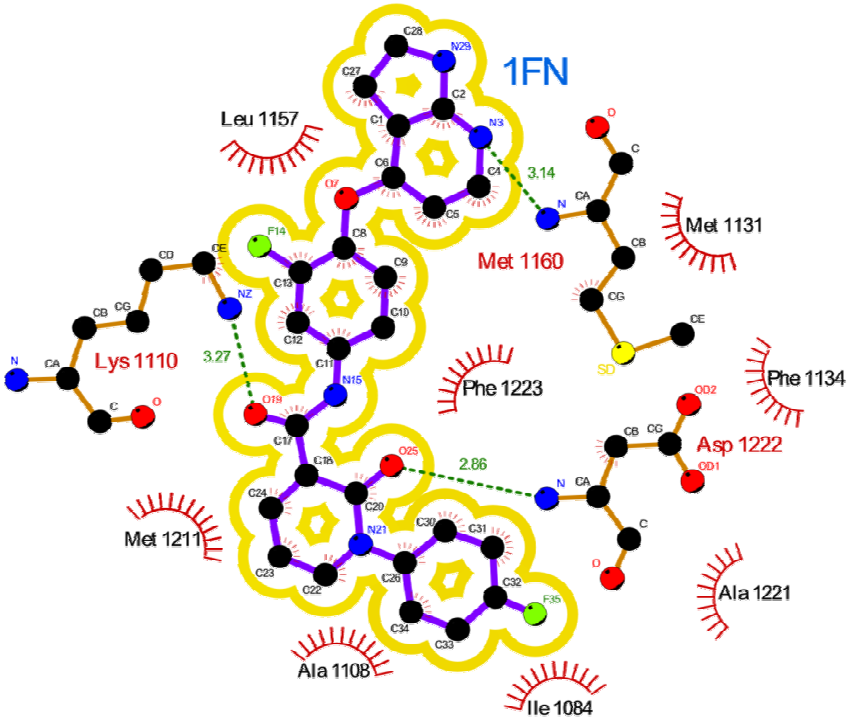
3CTH/319



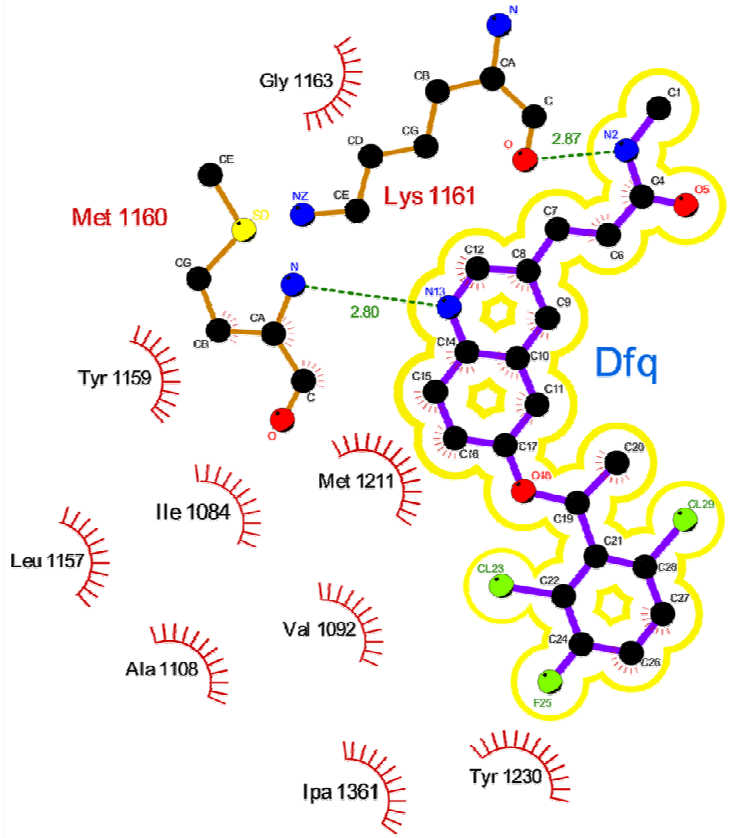
3CTH/320



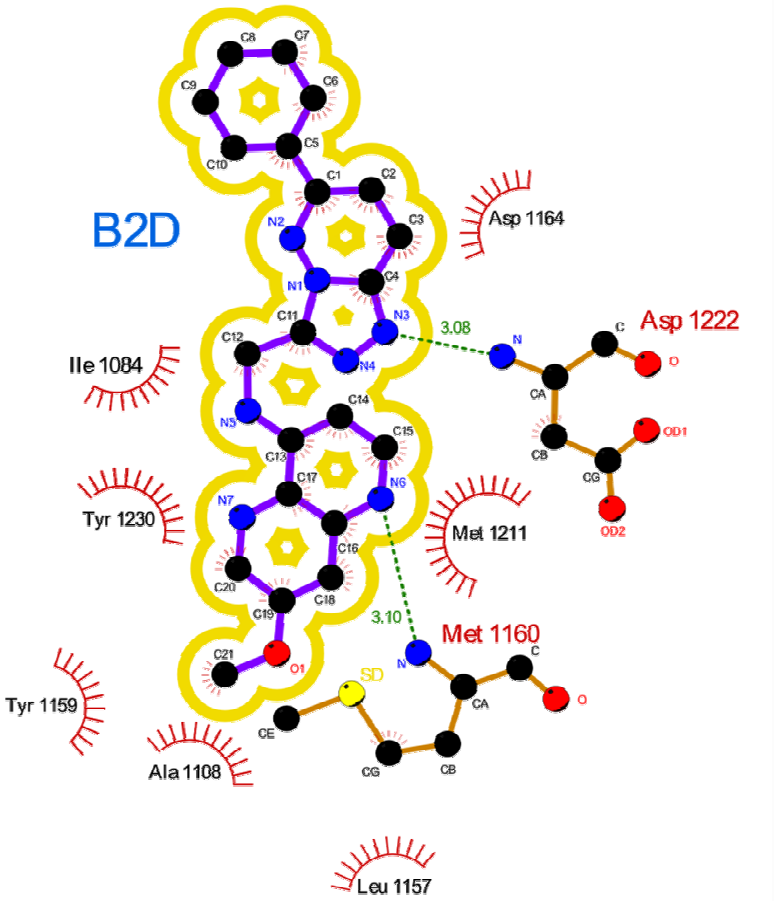
3F82/353



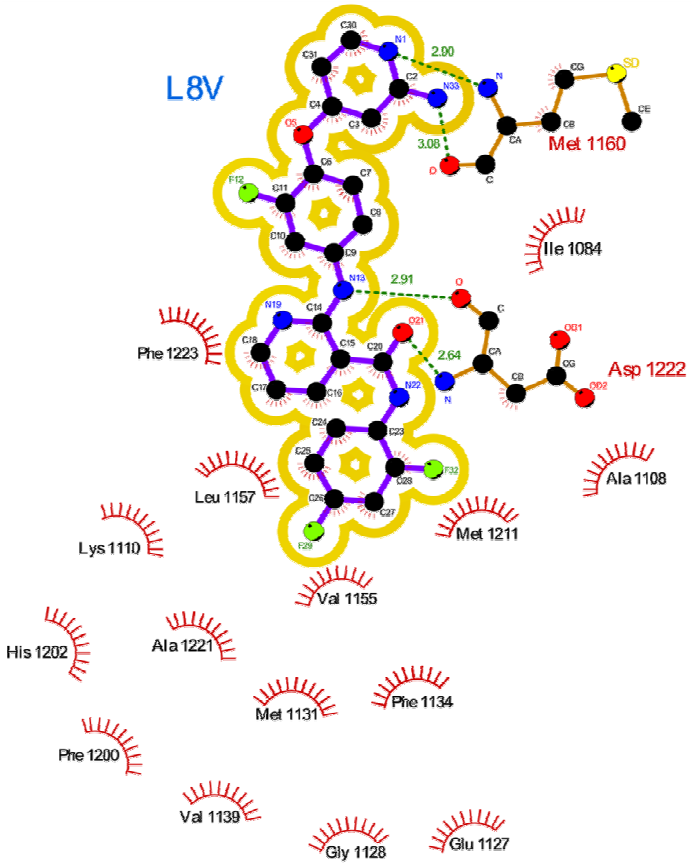
3CE3/1FN



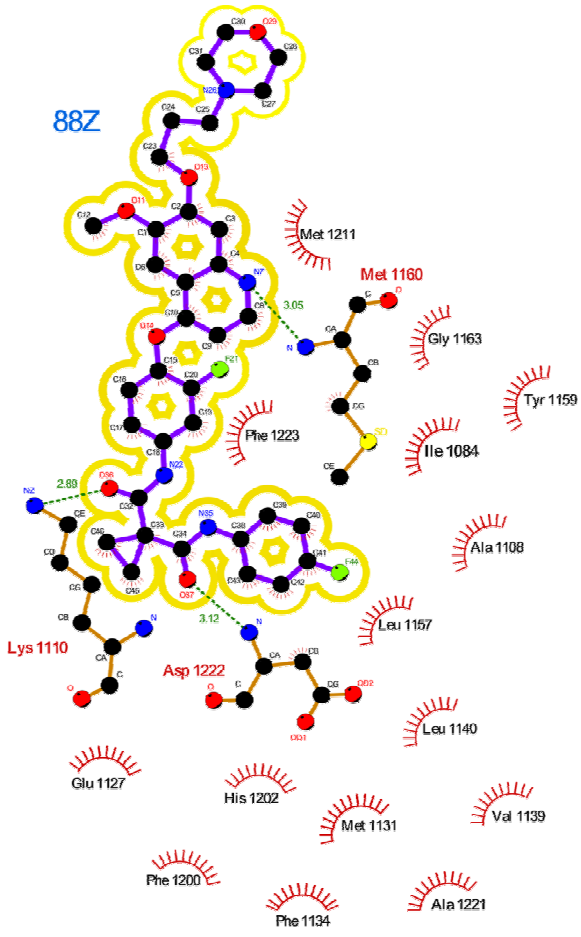
3A4P/dfq



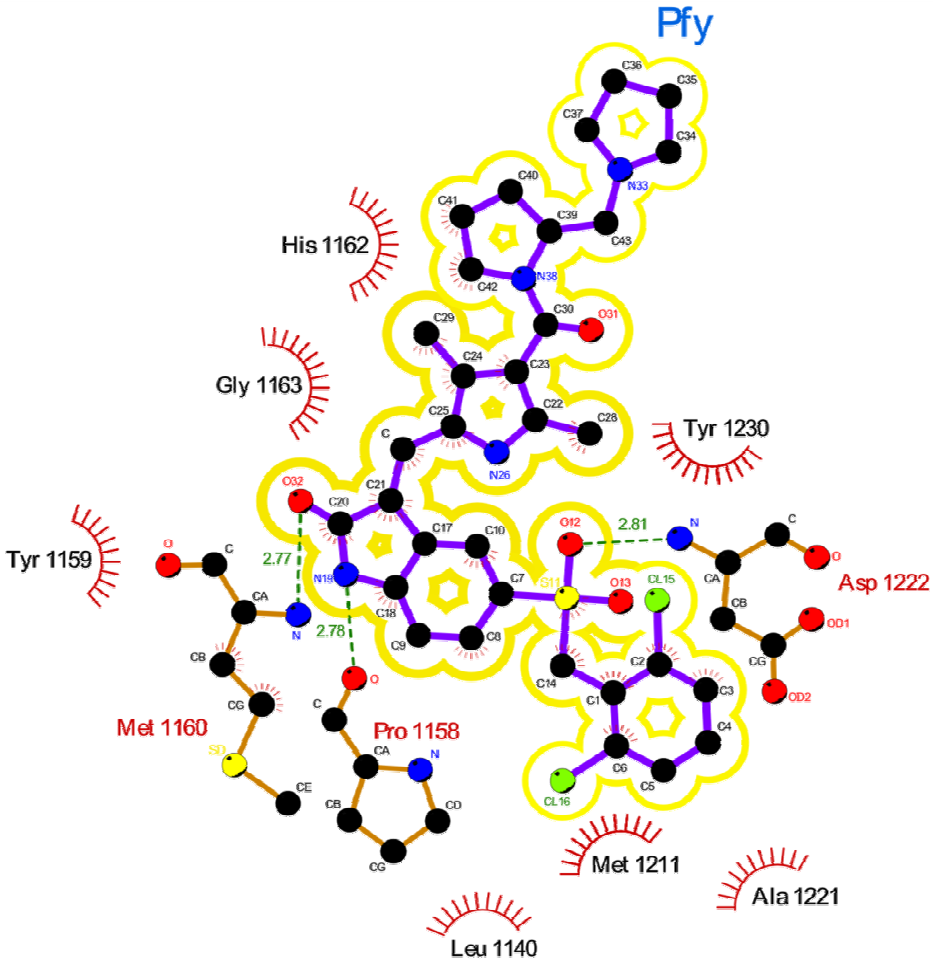
3I5N/b2d



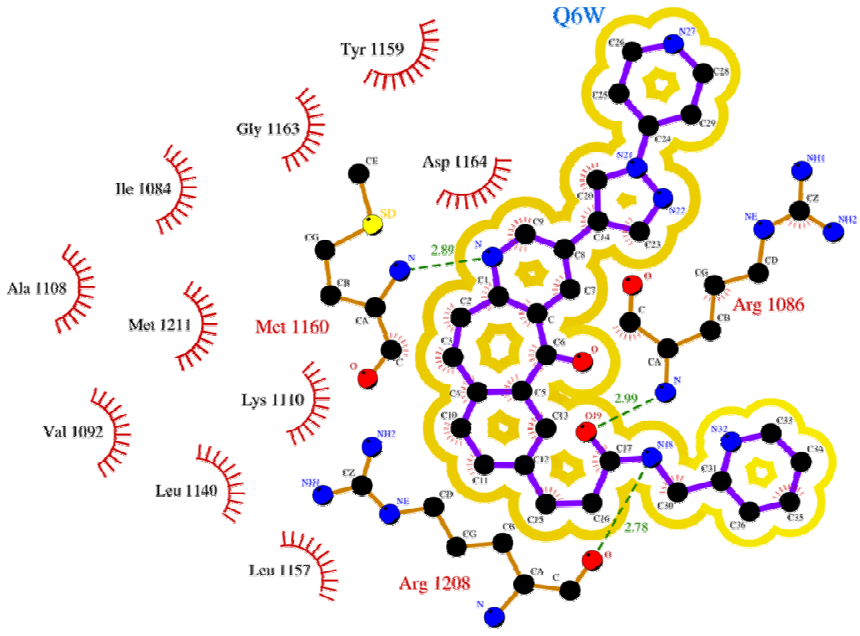
3L8V/l8v



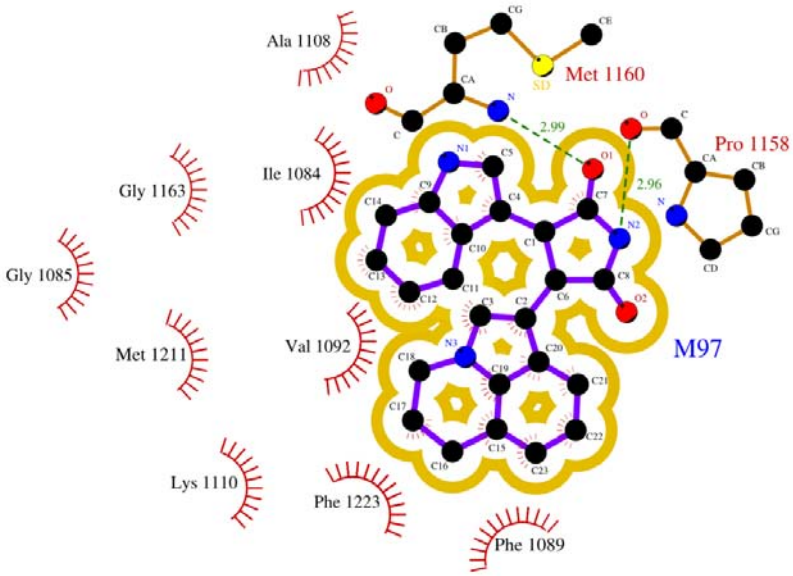
3LQ8/88z



2WKM/pfy



3Q6W/q6w


Ala 1108
Met 1160
Pro 1158
Ile 1084
Gly 1163
Gly 1085
Met 1211
Val 1092
M97
Lys 1110
Phe 1223
Phe 1089
2.99
2.96

3RHK/m97

Résumé

L'enjeu des travaux effectués au cours de cette thèse est l'extraction *in silico* de molécules potentiellement intéressantes dans le processus d'inhibition du récepteur tyrosine kinase c-Met. La faculté de cette protéine à interagir dans les phénomènes d'embryogenèse et de réparation tissulaires rendent son inhibition cruciale dans les traitements contre les développements tumoraux où c-Met se trouve impliquée.

Pour cela, la stratégie employée implique l'utilisation de méthodes *in silico* de conception rationnelle de médicaments. Nous avons utilisé comme support les multiples structures cristallographiques publiées sur la ProteinData Base. Un travail de modélisation par homologie fut tout d'abord nécessaire pour combler les lacunes des structures cristallographiques collectées. Afin d'échantillonner au mieux l'espace conformationnel de la kinase c-Met et de caractériser sa flexibilité, une longue campagne de simulation de Dynamique Moléculaire fut menée. Pour compléter ces simulations, nous avons également utilisé la méthode des modes normaux de vibration. De ces 2 approches, nous avons extrait un ensemble de 10 conformères considérés comme les plus représentatifs de l'espace conformationnel simulé pour la kinase c-Met et avons proposé un mode de fonctionnement de ce récepteur. Utilisant les conformations représentatives, nous avons ensuite mené une importante campagne de criblage virtuel sur plusieurs chimiothèques constituant environ 70.000 composés. L'analyse des résultats de l'arrimage moléculaire nous a conduits à la sélection de composés intéressants possédant théoriquement une bonne affinité pour la kinase c-Met. Ces molécules ont été soumises aux tests expérimentaux.

Mots-clés : modélisation moléculaire ; protéines tyrosine kinase ; c-Met ; cancer ; modélisation par homologie ; échantillonnage conformationnel ; dynamique moléculaire ; modes normaux de vibration ; docking moléculaire ; clustering ; criblage virtuel ; inhibiteurs.

Abstract

The challenge of this PhD work is the *in silico* identification of potentially interesting molecules concerning the inhibitory process of tyrosine kinase receptor c-Met. The faculty of this protein to interact in embryogenesis and tissue repair phenomena makes its inhibition crucial for treatments against tumor development in which c-Met is involved.

For that purpose, the employed strategy involves the use of several *in silico* methods for rational drug design. As the basement of this work, we used the multiple crystal structures published in the ProteinData Base (PDB). A preliminary homology modeling work was needed to fill gaps in the crystal structures. To sample at best the c-Met kinase conformational space and to characterize its flexibility, a long Molecular Dynamics (MD) simulation campaign was carried out both on *apo* and *holo* forms of available crystal structures. To complete these simulations, part of this work consisted to use normal modes of vibration (NM) method. From these two approaches (DM and NM), we extracted a set of 10 conformers considered as the most representative of the kinase simulated conformational space and we suggested a mode of operation of this kinase. Using extracted conformations from the conformational sampling has enabled us to conduct an extensive campaign on several virtual screening libraries constituting a total of approximately 70,000 compounds. Analysis of the molecular docking results has led us to the selection of several theoretically interesting molecules with good potential affinity for c-Met kinase. These molecules were submitted to experimental tests performed by the biologist team associated to our work.

Keywords : molecular modelling ; tyrosine protein kinases ; c-Met; cancer ; homology modelling; conformational sampling ; molecular dynamics ; normal modes of vibration ; molecular docking ; clustering ; virtual screening ; inhibitors.