

N° ORDRE : 4444

**THÈSE DE DOCTORAT**

PRÉSENTÉE A

**L'UNIVERSITÉ DE BORDEAUX 1**

ÉCOLE DOCTORALE DES SCIENCES CHIMIQUES

PAR **JUDITH ELKAÏM**

POUR OBTENIR LE GRADE DE

**DOCTEUR**

SPÉCIALITÉ : CHIMIE-PHYSIQUE

---

*DRUG DESIGN IN SILICO*  
*CRIBLAGE VIRTUEL DE PROTÉINES À VISÉE*  
*THÉRAPEUTIQUE*

---

Soutenue le 20 décembre 2011

**Après avis de :**

G. LABESSE            Directeur de Recherche, CNRS  
B. VILLOUTREIX      Directeur de Recherche, INSERM

**Devant la commission d'examen formée de :**

|                 |                                              |                    |
|-----------------|----------------------------------------------|--------------------|
| G. LABESSE      | Directeur de Recherche, CNRS                 | Rapporteur         |
| B. VILLOUTREIX  | Directeur de Recherche, INSERM               | Rapporteur         |
| N. GENESTE      | Maître de Conférences, Université Bordeaux I | Examineur          |
| L. MORIN-ALLORY | Professeur, Université d'Orléans             | Président de Jury  |
| J. DESSOLIN     | Chargé de Recherche, CNRS                    | Directeur de Thèse |



## - REMERCIEMENTS -

Cette thèse a été réalisée au sein de l'Institut Européen de Chimie et Biologie (IECB) et de l'UMR de Chimie et Biologie des Membranes et Nano-objets (CBMN). Je tiens à remercier Messieurs Jean-Jacques Toulmé et Erick Dufourc pour m'avoir accueillie dans le laboratoire.

Je remercie les membres du jury pour l'intérêt porté à mon travail. Merci à Messieurs Gilles Labesse et Bruno Villoutreix d'avoir accepté de juger mon manuscrit. Merci également à Luc Morin-Allory d'avoir bien voulu présider le jury. Un remerciement tout particulier à Nathalie Geneste, pour votre participation au jury, mais surtout pour votre implication dans ma formation depuis ma première année. Vous m'avez poussée vers la chimie, initiée à la modélisation et soutenue tout au long de mes études afin que je poursuive en thèse. Sans vous, je n'en serais pas là.

Je remercie Michel Laguerre pour la place qu'il m'a offerte au sein de son équipe, ainsi que pour les conseils avisés qu'il a pris le temps de me donner à chaque fois que je l'ai sollicité. Merci à Jean Dessolin de m'avoir donné ce sujet de thèse qui correspondait à mes souhaits d'abord, et de m'avoir si bien encadré ensuite. En me donnant une liberté totale dans tous mes choix tout en vous montrant toujours disponible lorsque j'avais besoin d'aide, vous m'avez permis de me construire en tant que chercheur. J'ai beaucoup apprécié ces trois années, et j'espère trouver autant de plaisir à faire ce métier pendant longtemps, que j'en ai trouvé pendant ma thèse.

Je tiens également à remercier toutes les personnes ayant collaboré à mes recherches, ce travail ne serait pas ce qu'il est sans votre participation.

Merci à tous les membres de l'équipe de modélisation, passés et présents, ainsi qu'à toutes les personnes que j'ai eu l'occasion de croiser au laboratoire pour l'ambiance générale. Nada, merci pour ta bonne humeur et ta vision positive qui m'ont réconfortés dans certains moments difficiles. JM et Mélanicita, Clément et Mélanie, ces trois années n'auraient pas été les mêmes sans vous. Merci pour tous les à-cotés. Une petite pensée pour Chayan, Hélène et d'autres qui nous ont accompagnés par périodes.

Merci à mes amis de longue date que ces trois années ont un peu éloignés de moi mais que j'espère retrouver rapidement, et à mes anciens compagnons du master de chimie de l'environnement.

Merci à mes parents et à ma sœur, à qui je dois beaucoup.

Et David, merci pour tout.



# - SOMMAIRE -

|                                                                      |           |
|----------------------------------------------------------------------|-----------|
| - REMERCIEMENTS -.....                                               | 3         |
| - SOMMAIRE -.....                                                    | 6         |
| - PRÉAMBULE -.....                                                   | 12        |
| - LISTE DES ABRÉVIATIONS -.....                                      | 13        |
| - LEXIQUE -.....                                                     | 14        |
| <b>CHAPITRE 1 : INTRODUCTION.....</b>                                | <b>16</b> |
| <b>A.DÉCOUVERTE DE NOUVEAUX MÉDICAMENTS.....</b>                     | <b>18</b> |
| <b>B.MÉTHODES DE CRIBLAGE IN SILICO.....</b>                         | <b>20</b> |
| <b>C.CRIBLAGE VIRTUEL « STRUCTURE-BASED ».....</b>                   | <b>22</b> |
| <b>I.DOCKING.....</b>                                                | <b>22</b> |
| 1.RECHERCHE SYSTÉMATIQUE.....                                        | 23        |
| 2.RECHERCHE ALÉATOIRE.....                                           | 24        |
| 3.MÉTHODES DE SIMULATION.....                                        | 25        |
| <b>II.SCORING.....</b>                                               | <b>25</b> |
| 1.FONCTIONS DE SCORE BASÉES SUR LES CHAMPS DE FORCE.....             | 26        |
| 2.FONCTIONS DE SCORE EMPIRIQUES.....                                 | 26        |
| 3.FONCTIONS DE SCORE « KNOWLEDGE-BASED ».....                        | 27        |
| 4.CONSENSUS SCORING.....                                             | 28        |
| <b>III.CHOIX DES PROGRAMMES À UTILISER.....</b>                      | <b>30</b> |
| 1.REPRODUCTION DE POSES EXPÉRIMENTALES.....                          | 31        |
| 2.PRÉDICTION D’AFFINITÉ.....                                         | 32        |
| 3.CRIBLAGE DE BASES DE LIGANDS.....                                  | 33        |
| <b>D.PROBLÉMATIQUES DU CRIBLAGE VIRTUEL « STRUCTURE-BASED ».....</b> | <b>35</b> |
| <b>I.PROBLÉMATIQUES LIÉES À LA PROTÉINE.....</b>                     | <b>35</b> |
| 1.DISPONIBILITÉ DES STRUCTURE DE PROTÉINES.....                      | 35        |
| 2.PRÉPARATION DES PROTÉINES.....                                     | 37        |

|                                                                                |           |
|--------------------------------------------------------------------------------|-----------|
| 3.IMPORTANCE DU SOLVANT.....                                                   | 38        |
| 4.GESTION DE LA FLEXIBILITÉ DE LA PROTÉINE.....                                | 39        |
| <b>II.PROBLÉMATIQUES LIÉES AUX LIGANDS.....</b>                                | <b>42</b> |
| 1.BASES DE LIGANDS.....                                                        | 42        |
| 2.STRUCTURE DES LIGANDS.....                                                   | 43        |
| <b>III.AUTOMATISATION DES CALCULS ET GESTION DES FICHIERS.....</b>             | <b>44</b> |
| <b>E.OBJECTIFS.....</b>                                                        | <b>45</b> |
| <br>                                                                           |           |
| <b>CHAPITRE 2 : ÉTUDE MÉTHODOLOGIQUE DE LA FLEXIBILITÉ DE LA PROTÉINE.....</b> | <b>48</b> |
| <br>                                                                           |           |
| <b>A.GÉNÉRALITÉS.....</b>                                                      | <b>50</b> |
| <br>                                                                           |           |
| <b>B.DÉTAILS DU SYSTÈME ÉTUDIÉ.....</b>                                        | <b>51</b> |
| <b>I.STRUCTURES DU RÉCEPTEUR.....</b>                                          | <b>51</b> |
| <b>II.STRUCTURES DES LIGANDS.....</b>                                          | <b>54</b> |
| <b>III.OUTILS UTILISÉS.....</b>                                                | <b>56</b> |
| 1.DOCKING.....                                                                 | 56        |
| 2.SCORING.....                                                                 | 58        |
| <b>IV.STRATÉGIE DE CRIBLAGE.....</b>                                           | <b>60</b> |
| <br>                                                                           |           |
| <b>C.MATÉRIEL ET MÉTHODES.....</b>                                             | <b>61</b> |
| <b>I.PRÉPARATION DES RÉCEPTEURS.....</b>                                       | <b>61</b> |
| <b>II.RELAXED COMPLEX SCHEME.....</b>                                          | <b>61</b> |
| <b>III.PRÉPARATION DES LIGANDS.....</b>                                        | <b>62</b> |
| <b>IV.PROTOCOLES.....</b>                                                      | <b>62</b> |
| 1.DOCKING AVEC DOCK.....                                                       | 62        |
| 2.DOCKING AVEC VINA.....                                                       | 63        |
| 3.RESCORING.....                                                               | 64        |
| <br>                                                                           |           |
| <b>D.RÉSULTATS ET DISCUSSION.....</b>                                          | <b>64</b> |
| <b>I.CONFORMATIONS DES LIGANDS.....</b>                                        | <b>64</b> |
| <b>II.PRÉDICTIONS D’AFFINITÉ.....</b>                                          | <b>66</b> |
| 1.COMPARAISON ENTRE LES TROIS STRUCTURES CRISTALLOGRAPHIQUES.....              | 66        |
| 2.COMPARAISON ENTRE LES STRUCTURES RELAXÉES.....                               | 67        |
| 3.DOCKING FLEXIBLE.....                                                        | 69        |

|                                                                                       |            |
|---------------------------------------------------------------------------------------|------------|
| E.CONCLUSION.....                                                                     | 70         |
| <b>CHAPITRE 3 : APPLICATION À LA PHOSPHOLIPASE HUMAINE sPLA2 DE GROUPE X</b><br>..... | <b>73</b>  |
| <b>A.GÉNÉRALITÉS.....</b>                                                             | <b>75</b>  |
| <b>B.RÔLES PHYSIOLOGIQUES ET PATHOLOGIQUES.....</b>                                   | <b>77</b>  |
| I.DÉFENSE CONTRE LES PATHOGÈNES.....                                                  | 79         |
| II.HYDROLYSE DES PHOSPHOLIPIDES ALIMENTAIRES.....                                     | 79         |
| IV.ATHÉROSCLÉROSE.....                                                                | 80         |
| V.TUMOROGÈNE.....                                                                     | 81         |
| VI.INTÉRÊT THÉRAPEUTIQUE.....                                                         | 83         |
| <b>C.DÉTAILS DU SYSTÈME ÉTUDIÉ.....</b>                                               | <b>83</b>  |
| I.STRUCTURE DE LA sPLA2-X HUMAINE.....                                                | 83         |
| II.INHIBITEURS CONNUS.....                                                            | 85         |
| III.BASE DE LIGANDS INNOVANTS.....                                                    | 87         |
| <b>D.MATÉRIEL ET MÉTHODES.....</b>                                                    | <b>88</b>  |
| I.PRÉPARATION DU RÉCEPTEUR.....                                                       | 88         |
| II.RECHERCHE CONFORMATIONNELLE SUR LES LIGANDS.....                                   | 89         |
| III.PRÉPARATION DES LIGANDS.....                                                      | 90         |
| IV.DOCKING ET RESCORING.....                                                          | 90         |
| <b>E.RÉSULTATS ET DISCUSSION.....</b>                                                 | <b>91</b>  |
| I.CONFORMATION OBTENUES.....                                                          | 91         |
| II.POSES DES INHIBITEURS CONNUS.....                                                  | 95         |
| III.POSES DES LIGANDS TRIAZÉPANEDIONE.....                                            | 98         |
| IV.COMPARAISON DES POSES DES DEUX BASES DE LIGANDS.....                               | 101        |
| V.CLASSEMENTS.....                                                                    | 102        |
| <b>F.BILAN.....</b>                                                                   | <b>106</b> |
| <b>CHAPITRE 4 : APPLICATION À LA PONTINE.....</b>                                     | <b>109</b> |

|                                                                 |            |
|-----------------------------------------------------------------|------------|
| <b>A.GÉNÉRALITÉS.....</b>                                       | <b>111</b> |
| <b>B.RÔLES PHYSIOLOGIQUES ET PATHOLOGIQUES.....</b>             | <b>114</b> |
| <b>I.INTERACTIONS AVEC DES FACTEURS DE TRANSCRIPTION.....</b>   | <b>114</b> |
| <b>II.COMPLEXES DE REMODELAGE DE LA CHROMATINE.....</b>         | <b>116</b> |
| <b>III.INTERACTIONS AVEC DES RIBONUCLÉOPROTÉINES.....</b>       | <b>117</b> |
| <b>IV.ACTIVITÉ HÉLICASE.....</b>                                | <b>118</b> |
| <b>V.RÔLE SPÉCIFIQUE DANS LE CANCER.....</b>                    | <b>119</b> |
| <b>VI.INTÉRÊT THÉRAPEUTIQUE.....</b>                            | <b>120</b> |
| <b>C.DÉTAILS DU SYSTÈME ÉTUDIÉ.....</b>                         | <b>121</b> |
| <b>I.ALIGNEMENT DE SÉQUENCE.....</b>                            | <b>121</b> |
| <b>II.STRUCTURE DE LA PONTINE.....</b>                          | <b>122</b> |
| <b>III.COMPLEXES FORMÉS PAR PONTINE ET REPTINE.....</b>         | <b>125</b> |
| <b>IV.COOPÉRATIVITÉ ENTRE DEUX MONOMÈRES ADJACENTS.....</b>     | <b>126</b> |
| <b>V.BASES DE LIGANDS.....</b>                                  | <b>127</b> |
| <b>VI.ELABORATION D'UN TEST ENZYMATIQUE.....</b>                | <b>129</b> |
| <b>VII.CRIBLAGE IN CELLULO.....</b>                             | <b>130</b> |
| <b>VIII.STRATÉGIE DE CRIBLAGE VIRTUEL MISE EN PLACE.....</b>    | <b>130</b> |
| <b>D.MATÉRIEL ET MÉTHODES.....</b>                              | <b>131</b> |
| <b>I.PRÉPARATION DU RÉCEPTEUR.....</b>                          | <b>131</b> |
| <b>II.PRÉPARATION DES LIGANDS.....</b>                          | <b>132</b> |
| <b>III.DOCKING ET RESCORING.....</b>                            | <b>134</b> |
| <b>IV.TESTS ENZYMATIQUES.....</b>                               | <b>135</b> |
| <b>E.RÉSULTATS ET DISCUSSION.....</b>                           | <b>136</b> |
| <b>I.QUALITÉ DU MODÈLE.....</b>                                 | <b>136</b> |
| <b>II.CLASSEMENTS PAR FONCTIONS DE SCORE INDIVIDUELLES.....</b> | <b>138</b> |
| <b>III.CONSENSUS SCORING.....</b>                               | <b>139</b> |
| <b>IV.TRAITEMENT POST-EXPÉRIMENTAL.....</b>                     | <b>142</b> |
| <b>V.ÉLABORATION DE LISTES DE COMPOSÉS À TESTER.....</b>        | <b>143</b> |
| <b>VI.RÉSULTATS DES TESTS EXPÉRIMENTAUX.....</b>                | <b>144</b> |

|                                                                       |            |
|-----------------------------------------------------------------------|------------|
| <b>VII.ANALYSE DES MODES DE LIAISON.....</b>                          | <b>145</b> |
| <b>VIII.DOCKING FLEXIBLE.....</b>                                     | <b>147</b> |
| <b>IX.NATURE DES INHIBITEURS.....</b>                                 | <b>148</b> |
| <b>X.BLIND DOCKING.....</b>                                           | <b>149</b> |
| <b>F.BILAN.....</b>                                                   | <b>153</b> |
| <b>CONCLUSION.....</b>                                                | <b>156</b> |
| <b>RÉFÉRENCES.....</b>                                                | <b>161</b> |
| <b>ANNEXES.....</b>                                                   | <b>192</b> |
| - ANNEXE 1. FORMATS DE FICHIERS DE MOLÉCULES -.....                   | 194        |
| - ANNEXE 2. GESTION DES FICHIERS ET AUTOMATISATION DES CALCULS -..... | 200        |
| - ANNEXE 3. STRUCTURES 2D DES INDIRUBINES -.....                      | 204        |
| - ANNEXE 4. STRUCTURES 2D DES INHIBITEURS DE HGX -.....               | 208        |
| - ANNEXE 5. PUBLICATION -.....                                        | 216        |
| - RÉSUMÉ -.....                                                       | 249        |
| - ABSTRACT -.....                                                     | 251        |



## - PRÉAMBULE -

Ce manuscrit représente la somme de trois années de travail scientifique en modélisation moléculaire, mais aussi à l'interface d'autres disciplines scientifiques comme la biologie, la biochimie ou la chimie organique. Le lecteur pourra donc rencontrer fréquemment des expressions déroutantes qui seront, dans la mesure du possible, explicitées afin de permettre une compréhension immédiate. Outre des expressions scientifiques, mais néanmoins françaises, certaines expressions anglaises ont été utilisées en l'état lors de la rédaction de ce mémoire. Ainsi traduire « drug design » par « conception de médicaments » est possible mais demeure un pis-aller où « conception de principe actif » serait plus proche de la réalité, mais devient quasiment une périphrase et alourdit le texte. Les anglicismes seront donc présents dans la mesure où ce sont des expressions usuelles, dont il n'existe pas toujours d'équivalent français. De plus, leur utilisation devenue courante rend souvent plus aisée l'extraction de renseignements, par exemple lors d'une interrogation internet via un moteur de recherche courant, à partir de ces termes que de leur traductions. Les expressions françaises correspondantes seront utilisées de temps en temps, dans le but de faire le lien entre les termes anglais et français, mais sans traduire systématiquement ces expressions qui sont comprises par l'ensemble de la communauté scientifique. A toutes fins utiles, un lexique a été intégré pour ce vocabulaire particulier.

## - LISTE DES ABRÉVIATIONS -

AAA+ : ATPase Associated with diverse cellular Activities

ADP : Adénosine DiPhosphate

ADN : Acide DésoxyriboNucléique

ARDS : Acute Respiratory Distress Syndrome

ARN : Acide RiboNucléique

ATP : Adénosine TriPhosphate

hGX : Phospholipase sPLA2 Humaine de Groupe X

hGV : Phospholipase sPLA2 Humaine de Groupe V

HTS : High Throughput Screening

IC50 : Concentration Inhibitrice à 50%

ICM : Internal Coordinate Mechanics

LDL : Low Density Lipoproteins

LPC : LysoPhosphatidylCholine

NTP : Nucléoside TriPhosphate

OD : Optical Density

PC : PhosphatidylCholine

Pi : Phosphate Inorganique

PMF : Potential of Mean Force

POPC : 1-Palmitoyl-2-OleoylPhosphatidylCholine

POPG : 1-Palmitoyl-2-OleoylPhosphatidylGlycérol

RNP : RibuNucléoProtéine

RMS(D) : Root Mean Square (Deviation)

snoARN : Small Nucleolar ARN

snoRNP : Small Nucleolar RNP

sPLA2 : Secreted Phospholipase A2

## - LEXIQUE -

**Blind docking** : Docking réalisé à l'aveugle, c'est à dire sans limiter la recherche conformationnelle à un site actif prédéfini.

**Clash stérique** : Interpénétration des rayons de van der Waals de deux atomes.

**Clusterisation** : Processus de rassemblement de molécules selon une propriété commune (en général, une similarité de structure ou de pose).

**Consensus scoring** : Méthodes de scoring utilisant la combinaison des scores obtenus à partir de plusieurs fonctions.

**Drug design** : Processus consistant à modifier une molécule dans le but d'améliorer ses capacités en tant que médicament.

**Drug discovery** : Ensemble des processus menant à l'identification de molécules actives pouvant être commercialisées en tant que médicament.

**Drug-like (propriétés)** : Caractéristiques des molécules actives ayant un impact sur leur propriétés pharmacocinétiques (ADME) et/ou sur leur éventuelle toxicité.

**Drug-like (molécules)** : Molécules dont les propriétés drug-like sont satisfaisantes pour une application comme candidat médicament.

**Decoy** : Molécule inactive, utilisée pour évaluer l'efficacité des méthodes employées.

**Faux positif** : Molécule identifiée comme active mais en réalité inactive.

**Faux négatif** : Molécule identifiée comme inactive mais en réalité active.

**Induced fit** : Phénomène d'adaptation d'une protéine à son ligand au cours de la complexation.

**Pose** : Agencement 3D d'un composé dans un complexe, conformation et position d'un ligand au sein du site actif d'une protéine.

**Repositionnement** : Redirection d'une molécule commercialisée pour une application vers une autre application.

**Seed** : Nombre aléatoire ou fixé qui permet d'initialiser la génération de variables aléatoires dans un calcul.

**Test set** : Ensemble de ligands soumis à un criblage.

**Top** : Tête de liste obtenue via le scoring des molécules d'une base.

**Training set** : Ensemble de ligands utilisé pour le calibrage d'un outil ou d'une stratégie.



# **CHAPITRE 1 : INTRODUCTION**



## A. DÉCOUVERTE DE NOUVEAUX MÉDICAMENTS

Par le passé, un grand nombre de médicaments ont été découverts tout simplement grâce à l'identification de principes actifs extraits de substances naturelles historiquement utilisées dans la médecine non-conventionnelle, ou même par hasard, ce qu'on nomme sérendipité. Mais plus le nombre de médicaments connus augmente et plus les probabilités de faire une telle découverte sont faibles. Par la suite, les avancées dans le domaine de la synthèse chimique et de la pharmacologie ont conduit à une démarche de recherche systématique permettant l'élaboration de nouveaux médicaments de plus grande efficacité [Lombardino & Lowe, 2004].

La découverte de nouveaux médicaments (« drug discovery ») est un processus extrêmement long et fastidieux. Les différentes étapes sont illustrées de façon schématique dans la Figure 1. Au cours des phases successives, des milliers de molécules doivent être triées et sélectionnées, pour aboutir à un nombre très restreint de candidats. Pour chaque nouveau médicament, le temps passé depuis la première étape, au cours de laquelle une cible biologique pertinente dans un processus pathologique donné est identifiée, jusqu'à la mise sur le marché d'un composé tourne autour d'une moyenne d'environ 12 à 14 ans. Le coût global estimé est d'au moins 800 millions de dollars (estimation réalisée au taux en vigueur en 2000) par composé mis sur le marché [DiMasi *et al.*, 2003; Hubbard, 2005; Lombardino & Lowe, 2004]. Les nouvelles méthodes permettant la découverte de nouveaux médicaments se doivent donc d'innover afin de mettre en évidence des molécules encore inconnues ayant un certain potentiel d'activité sur des cibles biologiques connues [Davis *et al.*, 2003]. Les outils mis en place doivent être capables de guider les chimistes médicaux dans le choix des molécules à cribler et à synthétiser [Lombardino & Lowe, 2004; Walters *et al.*, 1998].



Figure 1. Étapes du processus de découverte des médicaments

Le choix d'une cible thérapeutique revient à sélectionner, à l'aide de données issues de la littérature, une biomolécule (en général une protéine ou un complexe protéique) impliquée

dans un certain processus pathologique. On estime que l'ensemble des médicaments sur le marché cible environ 500 molécules biologiques. Certaines familles de protéines sont plus représentées dans le génome humain, ou plus souvent impliquées dans des processus pathologiques, mais en contrepartie elles peuvent alors posséder un grand nombre de protéines homologues responsables d'effets indésirables. Certaines sont tout simplement plus aisées à cibler que d'autres, en particulier, il est plus facile pour une petite molécule de perturber la liaison d'un ligand endogène que des interactions protéine-protéine [Bleicher *et al.*, 2003]. Ces paramètres régulent ce que l'on appelle la « druggabilité » de la cible [Cheng *et al.*, 2007; Owens, 2007; Vistoli *et al.*, 2008].

La seconde étape consiste à identifier des hits, c'est à dire des molécules ayant la capacité d'interagir avec la cible choisie et susceptibles de moduler ses effets sur le processus biologique en question. Ces molécules sont sélectionnées à partir de bibliothèques de ligands naturels, ou de bibliothèques propriétaires, ciblées ou non, le plus souvent à l'aide des techniques de criblage à haut débit (HTS pour High Throughput Screening). Cependant, les tests expérimentaux, qui nécessitent le plus souvent une connaissance approfondie des systèmes étudiés, ne sont pas toujours aisés à mettre en œuvre, surtout à grande échelle, en automatisant les procédures, et sont très coûteux [Tuccinardi, 2009]. Ces difficultés ont un impact particulièrement important lorsque l'on se situe dans la phase d'identification des hits qui nécessite le criblage de milliers de molécules au minimum. Les stratégies de criblage virtuel, ou *in silico*, sont donc depuis quelques années employées en tant qu'alternative ou de façon complémentaire. Ces techniques sont en général assez faciles à mettre en place, pour un coût bien moindre que les criblages expérimentaux [Moitessier *et al.*, 2008]. De plus, l'évolution technologique constante de ces dernières décennies a permis d'accélérer considérablement les temps de calculs nécessaires à la simulation de systèmes complexes ou de bases de données de plusieurs milliers de molécules. Le criblage virtuel est donc aujourd'hui employé dans de nombreux projets, afin de sélectionner, au sein de vastes bibliothèques de molécules, un nombre restreint de composés à cribler expérimentalement [Walters *et al.*, 1998].

A ce niveau, un certain nombre de hits vont être transformés en leads, qui vont ensuite permettre de créer des séries de composés ayant une activité démontrée sur la cible choisie. Le choix porte sur des composés qui possèdent un motif structural de base sur lequel il est possible de greffer un certain nombre de substituants permettant de faire varier les propriétés des ligands formés. Cette étape est souvent considérée comme critique pour la sélection de composés de qualité [Oprea & Matter, 2004]. Les séries de leads sont ensuite optimisées d'une part afin de présenter une activité pour la cible la plus forte possible tout en conservant une spécificité suffisante vis à vis d'autres protéines [Keiser *et al.*, 2009], mais

également afin de répondre à un certain nombre de critères régissant les propriétés ADMETox et « drug-like » de la molécule, qui déterminent la capacité de celle-ci à être un bon médicament [Leeson & Springthorpe, 2007; Lipinski *et al.*, 1997; Vistoli *et al.*, 2008]. Cette phase correspond à l'expression « drug design » (conception de médicaments) souvent utilisée à tort comme un équivalent de « drug discovery » (découverte de médicaments).

Enfin, les phases pré-cliniques et cliniques, qui sont les plus longues et les plus coûteuses du processus, ont pour but d'évaluer l'efficacité et la sûreté du composé directement à partir d'essais réalisés chez des patients [Hubbard, 2005].

Le rôle du chimiste médicinal au cours de ce processus est donc dans un premier temps l'identification de hits, puis leur transformation en leads et l'optimisation de ceux-ci, afin d'obtenir des candidats médicaments susceptibles de devenir un jour un médicament commercialisé [Vistoli *et al.*, 2008].

## **B. MÉTHODES DE CRIBLAGE *IN SILICO***

Le terme criblage virtuel regroupe un ensemble de techniques computationnelles ayant pour objectif l'exploration de bases de composés à la recherche de molécules d'intérêt [Walters *et al.*, 1998]. Une analogie souvent utilisée compare ces techniques à des filtres qui permettraient de constituer des ensembles de molécules partageant certaines propriétés et de sélectionner les plus susceptibles d'interagir avec une cible donnée.

On regroupe ces méthodes en deux familles, le criblage virtuel « structure-based » et le criblage virtuel « ligand-based ». Comme son nom l'indique, le criblage virtuel « structure-based » part de la structure de la cible afin d'identifier les interactions possibles au sein du site de liaison et ainsi sélectionner des molécules capables de s'y arrimer. Par opposition, les techniques de criblage virtuel « ligand-based » utilisent les ligands connus comme guides pour définir les propriétés responsables de leur affinité pour la cible et identifier des molécules similaires [Clark, 2008; Tuccinardi, 2009]. Ces deux approches, qui peuvent s'avérer complémentaires, font appel à des procédés expérimentaux différents, et leur application nécessite l'accès soit à la structure du récepteur, soit à un ensemble d'inhibiteurs connus. Le choix de l'approche utilisée dépend donc des données disponibles au démarrage du projet [Clark, 2008].

Le criblage virtuel est utilisé dans les premières phases du développement de nouveaux médicaments, et en particulier lors de l'identification de hits. Il a alors pour but de sélectionner au sein de chimiothèques variées des ensembles réduits de molécules dont le

potentiel d'activité envers la cible thérapeutique visée est supérieur à celui des autres molécules [Enyedy & Egan, 2008]. Il peut également être utilisé lors de l'optimisation des leads. Dans ce cas, le criblage a pour objectif l'identification de motifs structuraux essentiels dans la liaison ligand-récepteur, et la discrimination des meilleurs composés au sein de chimiothèques orientées comprenant des molécules appartenant à une même série.

Aujourd'hui, les techniques de criblage virtuel sont devenues des outils indispensables de la chimie médicinale [Brooijmans & Kuntz, 2003; Leach *et al.*, 2006] et sont utilisées de manière quotidienne aussi bien dans les laboratoires de recherche publics que dans les grands laboratoires pharmaceutiques. Le nombre de publications dans le domaine est en augmentation constante depuis plus d'une dizaine d'années (Figure 2). En 2008, Clark montrait huit cas concrets dans lesquels des criblages virtuels « structure-based » ou « ligand-based » ont permis l'identification de molécules arrivées en phase clinique [Clark, 2008]. Un livre édité par Wiley-VCH en 2011 est entièrement consacré au criblage virtuel, et tout un chapitre est dédié aux applications et aux succès du criblage virtuel [Matter & Sotriffer, 2011]. Il regroupe environ 150 publications qui font état de divers projets dans lesquelles ces méthodes ont permis la découverte de nouveaux inhibiteurs pour de très nombreuses protéines issues de familles variées.

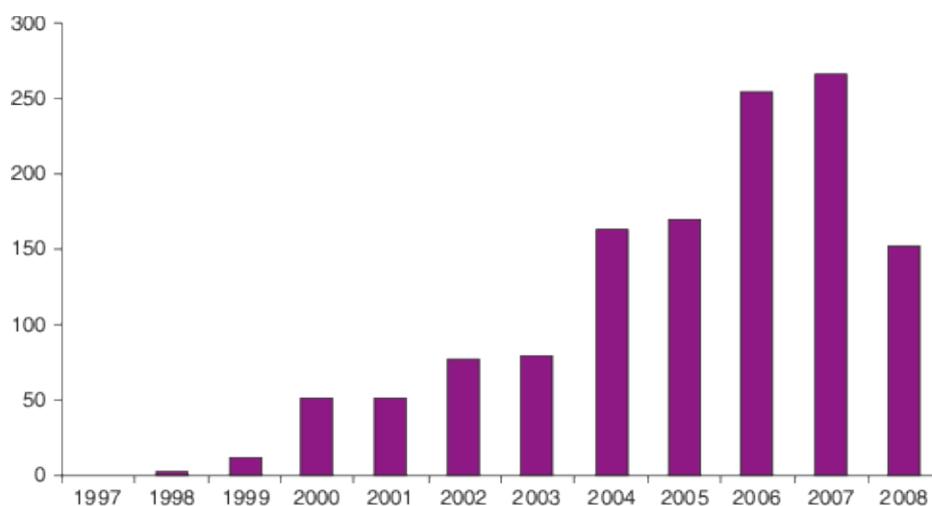


Figure 2. Evolution du nombre d'articles publiés par an dans le domaine du criblage virtuel entre 1997 et 2008 [Clark, 2008]

## C. CRIBLAGE VIRTUEL « STRUCTURE-BASED »

Au cours de ce travail, nous nous sommes intéressés uniquement aux techniques de criblage virtuel « structure-based ». Le processus de criblage « structure-based » correspond à un équivalent *in silico* d'un test de compétition ou de liaison expérimental puisque le but est d'observer la complexation entre un ligand et une biomolécule [Cummings *et al.*, 2005]. Il repose sur la combinaison de deux composantes. Dans un premier temps, le docking a pour objectif de prédire les positions que peut adopter le ligand au sein du site actif du récepteur, ainsi que la conformation du complexe. Les solutions sont appelées les poses. Ceci implique donc l'exploration des six degrés de liberté de translation et de rotation du ligand, ainsi que lorsque c'est possible des degrés de liberté conformationnels existants sur le ligand ou la protéine [Kitchen *et al.*, 2004; Leach *et al.*, 2006]. Dans un second temps, des fonctions de scoring vont être utilisées pour calculer les interactions formées entre le ligand et le récepteur dans les poses identifiées par le docking et en déduire un score représentatif de leur affinité. L'objectif de ce processus est, d'une part, de différencier les molécules susceptibles de se lier au récepteur des « decoys » qui en sont incapables, mais également les poses correctes parmi toutes les poses obtenues [Kitchen *et al.*, 2004; Moitessier *et al.*, 2008; Sousa *et al.*, 2006].

### I. DOCKING

De façon générale, le but du docking est de prédire la structure du complexe formé par un ligand et son récepteur. Plus particulièrement, dans les travaux présentés ici, les ligands sont des petites molécules destinées à inhiber l'activité d'une protéine, qui constitue le récepteur.

Historiquement, les premiers outils de docking obéissaient au principe dit « lock-and-key » (principe « clef-serrure ») selon lequel le ligand qui représente la clef, est complémentaire au niveau géométrique du site actif du récepteur, qui représente la serrure [Yuriev *et al.*, 2011]. Les algorithmes de l'époque permettaient donc de simuler un processus complètement statique au cours duquel un ligand dans une conformation donnée était docké au sein d'une protéine rigide, comme c'était le cas dans les années 1980 avec la première version de Dock par exemple [Kuntz *et al.*, 1982], on parlait alors de docking « rigid-body » [Dias & de Azevedo, 2008]. Depuis, il a été assez clairement établi que l'arrimage du ligand est un processus dynamique au cours duquel le ligand mais également la protéine sont susceptibles de subir des changements conformationnels importants [Yuriev *et al.*, 2011]. Cette représentation simpliste du problème a donc été améliorée par l'introduction de la

flexibilité sur le ligand, permettant l'exploration des conformations adoptées par celui-ci lors de l'arrimage dans le site actif de la protéine. Les liaisons dites flexibles sont toutes celles dont la rotation autour de l'axe est chimiquement valide et provoque le déplacement d'une « branche », c'est à dire d'un groupe d'atomes. Cependant, ce docking ligand flexible ne permet toujours pas de rendre compte de l'adaptation de la protéine.

Aujourd'hui, de nombreux programmes de docking permettent d'ajouter un niveau de complexité au problème en incluant également une certaine flexibilité sur le récepteur [Sousa *et al.*, 2006; Yuriev *et al.*, 2011], mais l'augmentation du nombre de degrés de liberté que cela représente est synonyme d'un décuplement du temps de calcul. Le docking ligand flexible-protéine rigide reste donc la méthode la plus souvent employée, car elle permet une évaluation rapide de bases de milliers, voire de millions de molécules [Mohan *et al.*, 2005]. La gestion de la flexibilité a été traitée comme un problème à part et fait l'objet d'un paragraphe dédié dans ce chapitre (voir « Problématiques liées à la protéine »).

En principe, un programme de docking doit être capable de générer les modes de liaison attendus pour des ligands dont la position adoptée au sein du site actif est connue (en général, via la structure cristallographique de la protéine co-cristallisée avec le ligand) dans un temps raisonnable. Pour cela, il est nécessaire que l'algorithme de recherche conformationnelle puisse explorer l'espace conformationnel le plus exhaustivement possible et de façon efficace. Classiquement, on juge de la qualité du docking en mesurant le RMSD (Root Mean Square Deviation) sur les atomes lourds entre la pose obtenue en docking, et la pose observée expérimentalement si elle existe [Sousa *et al.*, 2006].

Il existe trois types d'algorithmes pour le traitement de la flexibilité du ligand : les méthodes systématiques, les méthodes stochastiques ou aléatoires, et les méthodes de simulation [Kitchen *et al.*, 2004; Sousa *et al.*, 2006].

## 1. RECHERCHE SYSTÉMATIQUE

Ces méthodes ont pour but d'explorer la totalité des degrés de liberté du ligand. Pour cela, toutes les torsions du ligand subissent des rotations de 0 à 360°, en utilisant un pas incrémental fixé, afin de générer la totalité des conformations possibles. Rapidement, ceci provoque une augmentation très importante du nombre de conformations créées que l'on nomme explosion combinatoire, c'est pourquoi ces techniques ne sont pas employées dans leur forme originelle, mais en association avec un certain nombre de contraintes destinées à réduire le nombre de solutions [Sousa *et al.*, 2006].

Les méthodes de reconstruction incrémentale sont les algorithmes de recherche systématique les plus couramment appliqués. Celles-ci séparent la molécule en fragments rigides, et dockent ces fragments de façon indépendante avant de faire grandir étape par

étape le ligand dans le site actif en rajoutant les différents fragments sur le premier fragment docké qui sert de base à la reconstruction [Kitchen *et al.*, 2004; Sousa *et al.*, 2006]. Ce type d'algorithme est notamment employé dans les programmes FlexX [Kramer *et al.*, 1999; Rarey *et al.*, 1996], LUDI [Böhm, 1992a; Böhm, 1992b] et les versions flexibles du programme Dock [Ewing *et al.*, 2001; Moustakas *et al.*, 2006] (Figure 3).

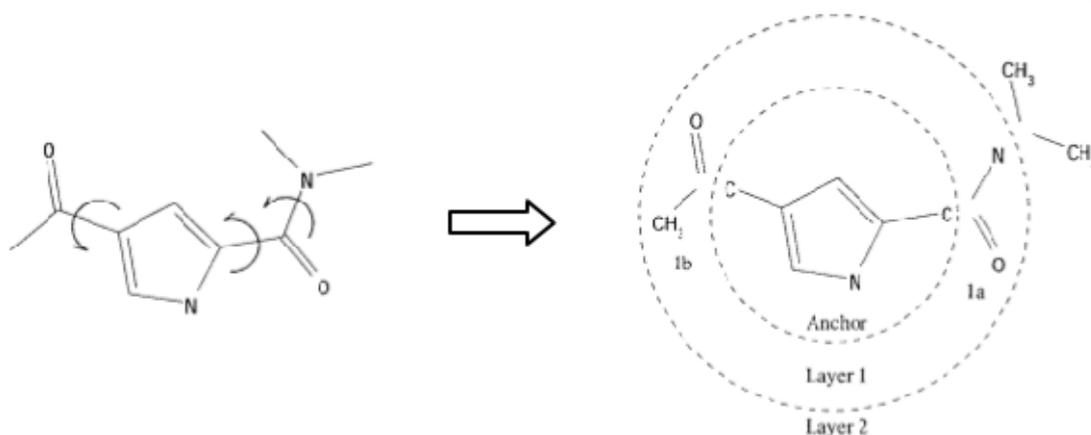


Figure 3. Reconstruction incrémentale de type Anchor and Grow (programme Dock) : la molécule est divisée en fragments puis dockée et reconstruite couche par couche dans le site actif [Ewing *et al.*, 2001]

Certains programmes utilisent des ensembles de conformations pré-établies afin de contrer le phénomène d'explosion combinatoire, comme FLOG [Kearsley *et al.*, 1994; Miller *et al.*, 1994].

## 2. RECHERCHE ALÉATOIRE

Ces méthodes fonctionnent en opérant des changements aléatoires sur un ligand ou une population de ligands, lesquels sont acceptés ou rejetés sur la base d'une fonction de probabilité (« fitness »), comme par exemple une fonction dérivée de Boltzmann dans la recherche de type Monte Carlo. Le programme ICM (Internal Coordinate Mechanics) est un outil de docking courant basé sur un algorithme de type Monte Carlo [Abagyan *et al.*, 1997; Abagyan *et al.*, 1994].

Les algorithmes génétiques utilisent des procédés inspirés de la théorie de l'évolution. En partant d'une population de conformations définies par un ensemble de variables assimilées à des gènes, ils appliquent des opérateurs génétiques (mutations, crossovers) permettant de faire varier les conformations jusqu'à l'obtention d'une population finale qui satisfait la fonction de « fitness » choisie [Sousa *et al.*, 2006]. Ce type l'algorithme est utilisé notamment dans les programmes Gold [Jones *et al.*, 1995; Jones *et al.*, 1997] et Autodock

[Morris *et al.*, 1998].

L'algorithme de recherche tabou explore l'espace conformationnel en interdisant les espaces déjà explorés et accepte les nouvelles solutions sur la base de leur RMSD par rapport aux précédentes conformations [Sousa *et al.*, 2006]. Il est utilisé par exemple dans le programme PRO\_LEADS [Baxter *et al.*, 1998].

### 3. MÉTHODES DE SIMULATION

Ces méthodes sont basées sur la résolution des équations du mouvement de Newton. Elles comprennent notamment les techniques de dynamique moléculaire et des algorithmes de minimisation. Les premières ne sont jamais utilisées pour générer de la flexibilité sur le ligand car elles demandent un temps de calcul non compatible avec la gestion de bases de molécules. Les secondes en revanche sont parfois utilisées dans les programmes de docking, en complément d'un autre algorithme de recherche, afin d'atteindre une conformation de basse énergie [Sousa *et al.*, 2006]. C'est le cas par exemple dans les versions récentes du programme Dock [Ewing *et al.*, 2001; Moustakas *et al.*, 2006], et dans le programme ICM [Abagyan *et al.*, 1997; Abagyan *et al.*, 1994].

## II. SCORING

Les fonctions de scoring sont des méthodes mathématiques permettant d'estimer l'affinité de liaison entre un récepteur et un ligand dans une pose donnée [Dias & de Azevedo, 2008]. Il convient ici de distinguer la fonction de scoring primaire, de la fonction de scoring secondaire. Les outils de docking sont tous pourvus d'une fonction de scoring qui permet de fournir les meilleures poses à la fin de la recherche conformationnelle, c'est ce que l'on appelle la fonction de scoring secondaire. La fonction de scoring primaire, elle, est la fonction utilisée pendant le docking afin d'évaluer les meilleures poses qui seront conservées au cours du processus [Mohan *et al.*, 2005]. Dans les programmes de docking les plus couramment utilisés, une seule fonction de scoring primaire existe, et elle est souvent équivalente à la (ou une lorsqu'il en existe plusieurs) fonction secondaire [Stahl & Rarey, 2001]. La distinction est donc souvent inexistante, mais leur rôle étant sensiblement différent, il est important de différencier les deux processus de scoring. En effet, s'il est possible d'effectuer un rescoring afin de compenser d'éventuelles défaillances dans la fonction de scoring secondaire utilisée par le programme de docking, la qualité des poses générées dépend de façon critique de la capacité de la fonction primaire à distinguer les bonnes poses des mauvaises [Mohan *et al.*, 2005; Perola *et al.*, 2004].

Les fonctions de score peuvent être divisées en trois classes : les fonctions basées sur des champs de force, les fonctions empiriques, et les fonctions « knowledge-based » [Kitchen *et al.*, 2004; Sousa *et al.*, 2006].

## 1. FONCTIONS DE SCORE BASÉES SUR LES CHAMPS DE FORCE

Dans ce type de fonctions, le score est la somme de l'énergie interne du ligand et de l'énergie d'interaction entre le ligand et la protéine, l'énergie interne de la protéine étant négligeable lorsque la conformation de celle-ci ne varie pas. Les énergies sont calculées à partir d'un terme de type van der Waals, en général un potentiel de Lennard-Jones, et d'un terme électrostatique coulombien assorti d'une fonction diélectrique dépendante de la distance qui permet d'atténuer les contributions des interactions charge-charge.

Ces fonctions sont largement représentées dans les programmes de docking les plus utilisés. Les fonctions de scoring primaires de Dock [Ewing *et al.*, 2001; Moustakas *et al.*, 2006] et d'Autodock [Morris *et al.*, 1998] ainsi que la fonction Goldscore [Verdonk *et al.*, 2003] de Gold [Jones *et al.*, 1995; Jones *et al.*, 1997] sont basées sur le champ de force AMBER [Weiner & Kollman, 1981; Weiner *et al.*, 1984; Weiner *et al.*, 1986]. Les fonctions G-score et D-score du module CScore de SYBYL s'appuient sur le champ de force Tripos (<http://tripos.com/>).

Ces fonctions sont limitées par l'absence de termes de solvatation et d'entropie, ainsi que la nécessité d'introduire des distances de cutoff pour le traitement des interactions non liantes à longue distance [Kitchen *et al.*, 2004; Sousa *et al.*, 2006]. Des modèles permettant de simuler le solvant de façon implicite par l'utilisation d'un continuum diélectrique comme les modèles PB/SA (Poisson-Boltzman/Surface Area) [Rocchia *et al.*, 2002], et GB/SA (Generalized Born/Surface Area) [Liu *et al.*, 2004] implémenté dans Dock [Ewing *et al.*, 2001; Moustakas *et al.*, 2006] existent, mais ne peuvent être utilisés qu'en scoring secondaire vu le temps de calcul qu'ils demandent [Huang *et al.*, 2010]. De plus, elles ne peuvent être adaptées à une application dans le cadre d'un docking protéine flexible sans un coût en temps de calcul très important, puisque l'utilisation de conformations multiples pour la protéine obligerait alors l'ajout du terme concernant l'énergie interne de celle-ci [Kitchen *et al.*, 2004].

## 2. FONCTIONS DE SCORE EMPIRIQUES

Les fonctions de score empiriques ont été conçues pour reproduire des données d'affinité obtenues expérimentalement en se basant sur l'idée que l'énergie de liaison peut être évaluée approximativement comme la somme pondérée de termes non corrélés. Ces termes peuvent représenter les contributions électrostatiques, celles des liaisons hydrogène

ou des contacts hydrophobes, certaines pénalités entropiques liées par exemple au nombre de torsions par ligand, ou à la désolvatation, etc. [Kitchen *et al.*, 2004; Sousa *et al.*, 2006].

Les fonctions de score empiriques ont été inaugurées par les programmes de docking LUDI [Böhm, 1994; Böhm, 1998], puis FlexX [Rarey *et al.*, 1996] qui a basé sa propre fonction de score sur une version dérivée de celle présente dans LUDI [Böhm, 1992a; Böhm, 1992b]. D'autres fonctions empiriques ont ensuite vu le jour, comme ChemScore [Eldridge *et al.*, 1997] qui a été implémentée dans Gold [Jones *et al.*, 1995; Jones *et al.*, 1997] et PRO\_LEADS [Baxter *et al.*, 1998] par exemple, ou le programme standalone Xscore [Wang *et al.*, 2002].

Ces fonctions sont bien plus simples à calculer que celles basées sur les champs de force tout en incluant des termes parfois équivalents. Leur principal désavantage est leur forte dépendance envers les ensembles de données utilisés lors de leur calibration [Kitchen *et al.*, 2004; Sousa *et al.*, 2006; Tuccinardi, 2009], la plupart des fonctions de scoring empiriques existantes ayant été entraînées sur des training set d'une centaine de complexes au mieux. Cependant, l'augmentation constante du nombre de complexes pour lesquels il existe des données structurales et d'affinité permet d'envisager l'utilisation d'un grand nombre de complexes, ce qui devrait théoriquement permettre la conception d'une fonction empirique assez généraliste [Huang *et al.*, 2010].

### 3. FONCTIONS DE SCORE « KNOWLEDGE-BASED »

Tout comme les fonctions de score empiriques, les fonctions « knowledge-based » ont pour objectif de reproduire des données expérimentales, mais il ne s'agit plus d'énergie de liaison mais de structure. En effet, elles dérivent de l'analyse statistique de complexes de structure connue qui permet de calculer la fréquence d'occurrence des contacts par paires d'atomes afin d'en extraire des potentiels de force moyenne (PMF ou Potential of Mean Force).

La fonction PMF, implémentée dans CScore (SYBYL, <http://tripos.com/>) par exemple, est la première approche de ce type à avoir été imaginée [Muegge, 2006; Muegge & Martin, 1999]. Les fonctions du programme standalone Drugscore [Gohlke *et al.*, 2000; Velec *et al.*, 2005] sont également « knowledge-based ».

Les fonctions de ce type sont extrêmement simples et rapides à calculer [Kitchen *et al.*, 2004; Sousa *et al.*, 2006; Tuccinardi, 2009]. De plus, bien qu'elles soient dépendantes des structures ayant servi à leur calibration, l'augmentation constante du nombre de complexes connus permet la création de fonctions de plus en plus précises. Toutefois, l'absence de données « négatives », représentant des contacts défavorables, dans les structures cristallographiques traditionnellement utilisées lors de la conception de ces fonctions peut

être un obstacle pour la discrimination de poses incorrectes [Baxter *et al.*, 2000]. Leur intégration bien qu'encore marginale pourrait permettre de mieux prendre en compte ces contacts dans le calcul des scores [Leach *et al.*, 2006].

#### 4. CONSENSUS SCORING

Comme indiqué plus haut, le calcul de l'énergie de liaison est réalisé sur la base d'un certain nombre de simplifications et d'approximations qui expliquent que toutes les fonctions de scoring existantes possèdent des avantages, mais aussi des inconvénients [Sousa *et al.*, 2006]. Pour contourner les défauts des fonctions de scoring individuelles, l'approche de consensus scoring propose de les combiner afin d'augmenter la probabilité d'identifier des molécules ayant effectivement une affinité pour la cible [Kitchen *et al.*, 2004]. Elle se base sur l'hypothèse que les interactions variées qui ont lieu entre la protéine et le ligand sont mises en évidence par différentes fonctions de scoring. Elles seraient donc complémentaires et de fait une combinaison appropriée de celles-ci pourrait permettre de décrire tous les aspects de la liaison de façon plus correcte que chaque fonction individuelle [Bar-Haim *et al.*, 2009]. De façon plus abrupte, elle suppose tout simplement que les composés classés en tête de liste dans plusieurs fonctions ont une probabilité plus élevée d'être effectivement capables de se lier à la cible que les composés trouvés en tête dans une seule fonction [Wang & Wang, 2001].

Plusieurs publications ont montré que le consensus scoring était capable de fournir de meilleurs résultats que les fonctions de scoring individuelles pour une utilisation dans le cadre de la sélection de hits [Bissantz *et al.*, 2000; Charifson *et al.*, 1999; Clark *et al.*, 2002; Stahl & Rarey, 2001]. Dans une expérience théorique, Wang et Wang proposent un modèle mathématique permettant de mettre en évidence l'augmentation du taux de hits grâce à l'utilisation du consensus scoring [Wang & Wang, 2001]. Cette amélioration est notamment due à une réduction du nombre de faux positifs en tête de liste [Charifson *et al.*, 1999; Wang & Wang, 2001]. Il existe toutefois des limitations, par exemple dans le cas où une fonction de scoring est très efficace sur un système et où l'association avec d'autres fonctions moins performantes provoquerait une baisse de la qualité des résultats [Yang *et al.*, 2005]. Il est également risqué d'associer des fonctions corrélées car il est alors possible que les effets des erreurs de chacune soient amplifiés au lieu d'être gommés [Kitchen *et al.*, 2004]. Les fonctions utilisées doivent donc être indépendantes, ce qui peut vouloir dire qu'elles sont de nature différente (empirique, basée sur les champs de force ou « knowledge-based »), ou qu'elles sont issues d'un programme différent, les fonctions issues d'un même programme faisant souvent intervenir des composantes identiques [Wang & Wang, 2001].

Il existe plusieurs méthodologies de consensus scoring. La méthode « rank-by-

number » (classement par le nombre) utilise directement les scores fournis par chaque fonction de scoring pour créer une moyenne des scores. Cette stratégie quantitative est quasiment inapplicable à cause du manque de cohérence entre les différentes unités pratiquées par chaque fonction. La méthodologie « rank-by-rank » (classement par le rang) propose une approche presque aussi réaliste en réalisant une moyenne des rangs obtenus pour chaque molécule. Elle s'affranchit ainsi des problèmes d'unités, mais gomme les éventuelles variations d'énergie qui pourraient constituer une information intéressante, par exemple si on observe un saut dans les énergies mesurées correspondant à la distinction entre des ligands très affins et d'autres moins affins pour une protéine. Enfin, la méthode « rank-by-vote » (classement par vote) réalise une intersection mathématique des têtes de listes obtenues pour les différentes fonctions, en attribuant des points aux ligands qui apparaissent en tête de liste (Figure 4). En comparaison des deux autres, cette méthode est non quantitative, et ne fournit pas une liste classée de ligands, mais un ensemble de molécules trouvées en tête dans deux fonctions ou plus [Bar-Haim *et al.*, 2009; Wang & Wang, 2001].

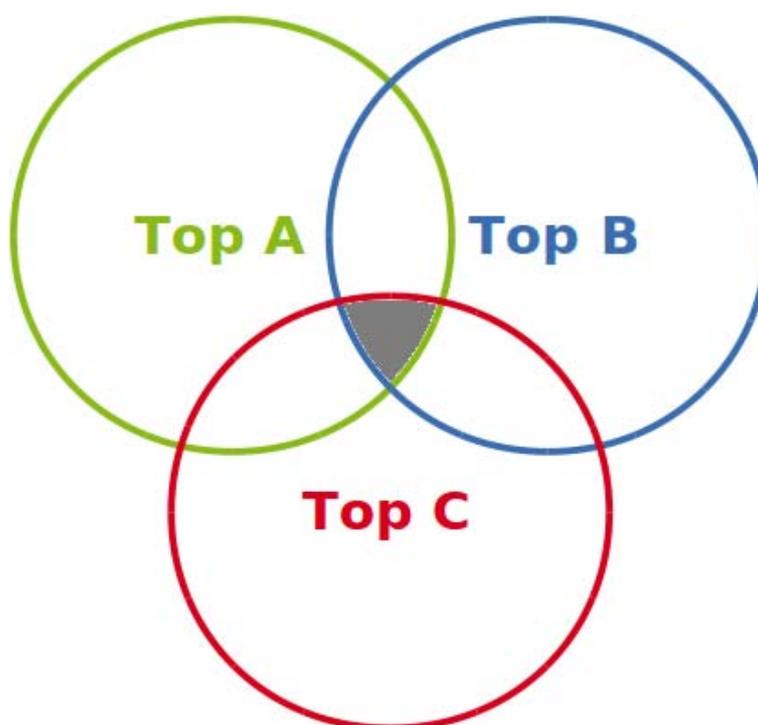


Figure 4. Consensus scoring « rank-by-vote » : la partie grisée correspond aux molécules conservées

### III. CHOIX DES PROGRAMMES À UTILISER

Les applications de ces techniques de criblage « structure-based » sont multiples. L'utilisation la plus répandue concerne le criblage à proprement parler, c'est à dire la recherche de composés actifs au sein de bases de ligands, qui correspond à l'étape de sélection des hits dans le processus de découverte de nouveaux médicaments. Cependant, ces méthodes peuvent également être utilisées lors de l'étape suivante d'optimisation des leads, ce qui signifie qu'elles doivent être capables de prédire de façon quantitative l'activité des ligands. Enfin, ces techniques sont parfois employées dans des études structurales afin d'explorer les modes de liaisons et de comprendre les interactions qui ont lieu entre le ligand et la protéine [Jain & Nicholls, 2008].

Deux paramètres régissent donc la qualité des outils de criblage virtuel : leur habilité à reproduire une pose observée expérimentalement et leur capacité à différencier des molécules actives dans une base de composés, voire de classer les ligands en fonction de leur affinité pour une cible. L'utilisation de ces techniques progressant de façon rapide, la nécessité de comparer les performances des différents programmes existants s'est fait sentir. De nombreuses publications ont donc proposé des expériences parallèles menés avec plusieurs programmes afin d'évaluer leur capacité à gérer des systèmes impliquant des types de protéines et de ligands variés [Bar-Haim *et al.*, 2009; Bissantz *et al.*, 2000; Charifson *et al.*, 1999; Cheng *et al.*, 2009; Clark *et al.*, 2002; Cross *et al.*, 2009; Cummings *et al.*, 2005; Deng & Verlinde, 2008; Englebienne & Moitessier, 2009; Good *et al.*, 2003; Jenkins *et al.*, 2003; Kellenberger *et al.*, 2004; Perola *et al.*, 2004; Stahl & Rarey, 2001; Teramoto & Fukunishi, 2007; Wang *et al.*, 2004; Wang *et al.*, 2003; Warren *et al.*, 2006; Yang *et al.*, 2005]. De plus, un numéro complet du Journal of Computer Aided Molecular Design a été consacré en 2008 à l'évaluation de ces méthodes [Clark & Webster-Clark, 2008; Cleves & Jain, 2008; Enyedy & Egan, 2008; Good & Oprea, 2008; Hawkins *et al.*, 2008; Irwin, 2008; Jain, 2008; Jain & Nicholls, 2008; Kirchmair *et al.*, 2008; Liebeschuetz, 2008; Nicholls, 2008; Sheridan *et al.*, 2008]. Si ces publications convergent sur plusieurs points, elles poussent également à la prudence concernant une comparaison trop précise, notamment en raison du manque de détail concernant les protocoles expérimentaux et les versions des programmes utilisés [Jain & Nicholls, 2008; Moitessier *et al.*, 2008]. Le point essentiel sur lequel la littérature présentée ci-dessus s'accorde toutefois, c'est qu'un outil de criblage virtuel universel doté d'un algorithme de docking et d'une fonction de scoring capables de donner de bons résultats sur la totalité des systèmes dans un temps raisonnable est actuellement hors de portée [Leach *et al.*, 2006; Perola *et al.*, 2004; Yuriev *et al.*, 2011]. Il est donc essentiel lors de la mise en place d'un criblage, d'établir un ensemble de procédures de

contrôle ayant pour objectif de vérifier la validité des résultats obtenus.

## 1. REPRODUCTION DE POSES EXPÉRIMENTALES

Dans l'absolu, une fonction de scoring primaire parfaite devrait être capable d'identifier, parmi les différentes solutions proposées par n'importe quel programme de docking, le meilleur mode de liaison pour un ligand, c'est à dire celui qui est observé expérimentalement [Charifson *et al.*, 1999; Huang *et al.*, 2010]. Avec le nombre de torsions sur le ligand la difficulté de trouver des poses correctes augmente, mais globalement, cet objectif est atteint de façon satisfaisante puisque les programmes de docking sont généralement calibrés pour reproduire des poses cristallographiques [Charifson *et al.*, 1999; Deng & Verlinde, 2008; Kitchen *et al.*, 2004]. Toutefois, la comparaison des performances de plusieurs programmes de docking ou de fonctions de scoring sur des protéines possédant des caractéristiques différentes montre des variations importantes dans la capacité du programme à réaliser un docking correct [Wang *et al.*, 2004]. La nature du site actif, plutôt hydrophile ou hydrophobe, les interactions qui régissent de façon critique la liaison, l'enfouissement et les contraintes stériques, ainsi que l'accessibilité au solvant ont une influence non négligeable sur les résultats de docking [Deng & Verlinde, 2008; Perola *et al.*, 2004].

Théoriquement, le choix du programme utilisé dans un projet pourrait donc se faire de façon éclairée, en sélectionnant le programme adéquat pour le système étudié. Dans la pratique, prédire quel outil est le approprié dans un cas précis est quasiment impossible [Bissantz *et al.*, 2000; Charifson *et al.*, 1999; Clark *et al.*, 2002; Schulz-Gasch & Stahl, 2003; Stahl & Rarey, 2001], surtout en l'absence de données approfondies concernant la protéine [Warren *et al.*, 2006]. Certains programmes peuvent être également plus efficaces mais beaucoup plus lents que d'autres, ce qui peut constituer un critère de choix dans le cas de criblages sur des bases de ligands de grande taille [Clark *et al.*, 2002; Kellenberger *et al.*, 2004]. De plus, les moyens permettant d'évaluer la qualité des résultats ne sont pas nécessairement adaptés. Le calcul du RMSD, qui est le plus souvent utilisé comme indicateur pour la reproduction de poses, accentue les défauts du docking, puisqu'un ligand de faible poids moléculaire, même placé de façon aléatoire, peut présenter un RMSD plus faible qu'une molécule de plus haut poids moléculaire. Ce paramètre peut être inefficace dans le cas de composés symétriques ou presque symétriques [Cole *et al.*, 2005]. En outre, un composé présentant un substituant de grande taille exposé au solvant et mal positionné peut posséder un RMSD assez élevé alors que les motifs situés dans le site actif sont correctement positionnés [Huang *et al.*, 2010] (Figure 5). Enfin, les poses reproduites sont le plus souvent issues de structures cristallographiques, dans lesquelles les modes de liaisons des ligands peuvent être influencés par les conditions de cristallisation [Sperandio *et al.*,

2006]. Sur 1300 structures, Sondergaard et al. ont remarqué que 36% d'entre elles présentaient des contacts cristallins susceptibles de conduire à un positionnement biaisé du ligand [Søndergaard *et al.*, 2009].

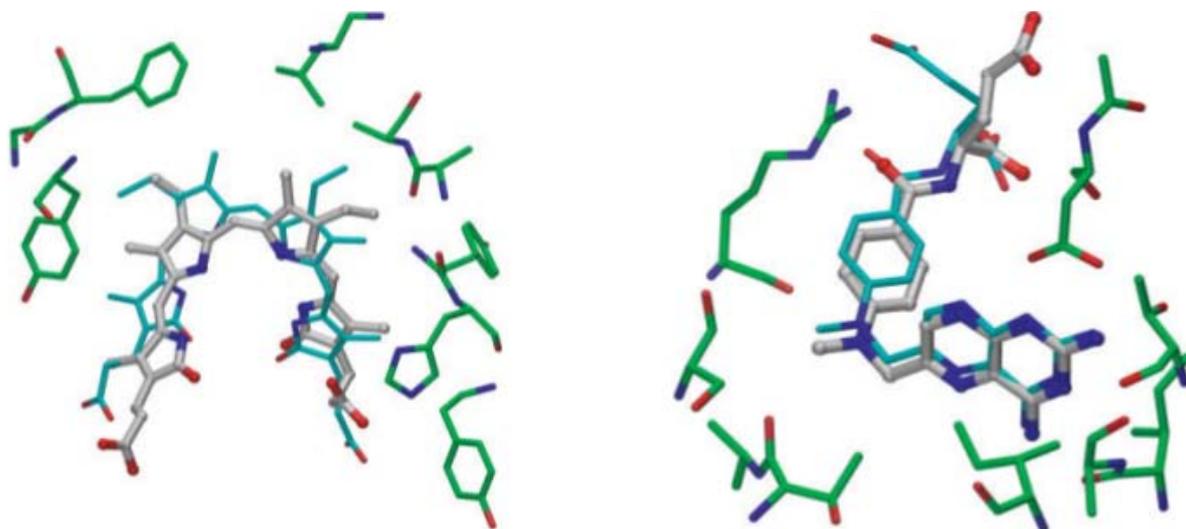


Figure 5. Poses de docking pour lesquels le RMSD calculé ne reflète pas la qualité de la pose. A gauche cas d'un composé symétrique, à droite cas d'un composé présentant un substituant dirigé vers le solvant, avec en vert les résidus de la protéine, en bleu la pose cristallographique et en gris la pose dockée [Cole *et al.*, 2005].

## 2. PRÉDICTION D’AFFINITÉ

Une fonction de scoring idéale devrait être capable de classer des ligands connus en fonction de leur affinité pour la cible [Clark, 2008]. Malgré l'existence d'un grand nombre de fonctions de scoring, aucune ne peut être appliquée de façon générale à tous les systèmes car chacune présente des avantages mais également des limitations [Huang *et al.*, 2010]. Un biais bien connu, est par exemple le meilleur classement des ligands de haute masse moléculaire, qui en raison d'un nombre d'atomes plus important, sont tout simplement capables de créer un plus grand nombre d'interactions [Cole *et al.*, 2005; Deng & Verlinde, 2008; Kitchen *et al.*, 2004]. De plus, tout comme la reproduction de poses expérimentales, le scoring est fortement dépendant de la nature de la protéine étudiée, selon les poids respectifs accordés aux interactions variées qui ont lieu entre les deux partenaires [Wang *et al.*, 2004].

Globalement, les fonctions de scoring disponibles à l'heure actuelle permettent au mieux de distinguer les molécules très actives des candidats peu susceptibles d'interagir avec la cible mais sont le plus souvent incapables de classer les ligands identifiés en

fonction de leur l'affinité [Clark, 2008; Leach *et al.*, 2006; Wang *et al.*, 2004] et parfois même de repérer les ligands de faible affinité [Charifson *et al.*, 1999]. En raison de cette limitation, les méthodologies « structure-based » sont plus adaptées à la phase d'identification de hits qu'à l'optimisation de leads qui semble relativement hors de portée des fonctions de scores actuelles.

### 3. CRIBLAGE DE BASES DE LIGANDS

L'objectif du criblage est la création d'un ensemble enrichi en molécules actives et de taille réduite, à partir de bases de ligands importantes [Enyedy & Egan, 2008; Knox *et al.*, 2005]. De fait, la réussite du criblage est dans plupart des études évalué non pas à l'aide d'un classement des ligands, mais d'un taux d'enrichissement, qui correspond au rapport entre le taux d'actifs présents dans la tête du classement (souvent exprimé en pourcentage), et le taux d'actifs global dans la base [Bender & Glen, 2005] (voir la formule ci-dessous).

$$E = \text{Tactifs}_{\text{TOP}} / \text{Tactifs}_{\text{GLOBAL}}$$

Ce taux d'enrichissement ne rend pas compte du classement des ligands et correspond donc mieux à ce que sont capables de fournir les fonctions de scoring actuelles. L'enjeu majeur dans ce processus est la réduction du nombre de faux positifs dans la tête de liste (ou top). En effet, si on imagine une base de 1000 ligands, dans laquelle on veut sélectionner un sous-ensemble de l'ordre de la dizaine de molécules pour les tester expérimentalement, un taux de faux positifs de 1% à peine peut résulter en l'identification de dix molécules supposées actives mais inactives en réalité ce qui réduit à néant les chances de trouver un composé actif [Bar-Haim *et al.*, 2009]. La stratégie de consensus scoring « rank-by-vote », qui permet d'obtenir un ensemble de ligands non classés comprenant un nombre réduit de faux positifs, va elle aussi dans le sens d'une distinction sous la forme de groupes de ligands d'affinité plus ou moins élevée, et s'est imposée comme un standard en matière de consensus [Bar-Haim *et al.*, 2009]. Pour que cette stratégie soit efficace, il convient de choisir avec précaution le cutoff choisi pour qualifier la tête de liste (le top) [Liebeschuetz, 2008].

Le taux d'enrichissement s'affranchit également de l'évaluation de la qualité de la pose [Cole *et al.*, 2005]. De fait, plusieurs études ont montré que le scoring prévalait sur le docking pour la sélection de molécules actives, avec une corrélation quasiment inexistante entre la qualité des poses obtenues et le taux de hits observé [Bissantz *et al.*, 2000; Leach *et al.*, 2006; Warren *et al.*, 2006]. Cependant, un criblage qui sélectionne les bonnes molécules mais ne trouve pas les bonnes poses reste suspect bien que plusieurs expériences

couronnées de succès aient montré ce type de profil [Cummings *et al.*, 2005]. Dans le cadre du criblage de bases importantes (dès la centaine de molécules), l'observation des poses est rarement réalisée [Clark, 2008]. Cependant, il est plus prudent de considérer que l'examen des poses doit rester systématique, au moins sur un nombre limité de composés choisis au hasard [Liebeschuetz, 2008]. Classiquement, une inspection visuelle post-criblage des ligands prioritaires est réalisée afin de vérifier la crédibilité des interactions observées entre le ligand et la protéine. Cette étape peut également servir à repérer d'éventuelles régions susceptibles d'accueillir un substituant pour l'optimisation de leads et le « drug design » [Tuccinardi, 2009].

Outre l'impact de la nature de la protéine sur les résultats de docking, les bases de ligands utilisées ont également une influence sur la capacité à sélectionner des inhibiteurs connus parmi des decoys, bien que cette problématique ait plus rarement été abordée dans la littérature. Verdonk *et al.* ainsi que Huang *et al.* ont pointé l'importance de l'utilisation de bases orientées lors de la validation des outils. En effet, il est fréquent d'observer des propriétés communes chez les molécules actives connues, qui dérivent souvent d'un même lead, ou appartiennent à la même famille chimique. Dans ce cas, la création d'un training set doit absolument prendre en compte ces similitudes, et les decoys choisis doivent posséder les mêmes caractéristiques que les actifs afin de ne pas biaiser le calcul vers une simple reconnaissance de ces propriétés [Huang *et al.*, 2006; Verdonk *et al.*, 2004].

De manière générale, certains programmes sont plus performants en ce qui concerne la reproduction de poses, que pour l'identification de molécules actives, ou vice versa [Cross *et al.*, 2009]. La qualité des résultats obtenus dépend donc également du type d'application mise en œuvre. De plus, dans un laboratoire un programme est souvent utilisé non pas pour ces performances sur un système particulier, mais par habitude, ce qui est également un avantage par exemple en ce qui concerne la gestion des fichiers et des paramètres, qui sera plus aisée pour l'expérimentateur s'il est familier avec les outils qu'il utilise, ou tout simplement l'analyse des données [Cole *et al.*, 2005; Kitchen *et al.*, 2004; Schulz-Gasch & Stahl, 2003]. Enfin, le bon déroulement à la fois du docking et du scoring va influencer de manière critique le succès global de l'expérience, mais il existe également un certain nombre d'autres problématiques liées directement à la protéine ou aux ligands utilisés et à la préparation des calculs qui vont entrer en ligne de compte pour la validité des résultats obtenus.

## **D. PROBLÉMATIQUES DU CRIBLAGE VIRTUEL « STRUCTURE-BASED »**

Le choix et la préparation des deux partenaires, protéine d'un côté, base de ligands de l'autre, doit être fait de manière éclairée, et représente une partie non négligeable du temps global de l'expérience [Sperandio *et al.*, 2006]. De manière générale, la pratique de la chimie médicinale demande une connaissance et une compréhension des mécanismes qui régissent les interactions entre la protéine et le ligand qui prennent une grande importance lors des étapes de préparation [Bissantz *et al.*, 2010; Davis & Teague, 1999] qu'il s'agisse de la sélection de la protéine, ou du choix, voire de la création de bases de ligands.

### **I. PROBLÉMATIQUES LIÉES À LA PROTÉINE**

La base de toute stratégie de criblage virtuel « structure-based » est la structure tridimensionnelle de la protéine étudiée. Il est donc primordial, afin de pouvoir mettre en œuvre ces méthodes, d'avoir accès à ces structures et d'être en mesure de préparer la protéine pour le docking, ce qui implique parfois une étude approfondie de son mécanisme d'action [Søndergaard *et al.*, 2009].

#### **1. DISPONIBILITÉ DES STRUCTURE DE PROTÉINES**

La plus grande base de données de structures protéiques est la RCSB Protein Data Bank (RCSB PDB, <http://www.pdb.org>). Née en 1971 au Brookhaven National Laboratory (Upton, NY, USA) sur l'impulsion d'un petit groupe d'experts désireux de stocker les données cristallographiques de macromolécules biologiques, elle réunissait à l'origine sept structures [Bernstein *et al.*, 1977]. Dans les années 1980, les progrès en matière de cristallographie et l'introduction de structures obtenues par RMN fait exploser le nombre d'entrées et conduit à une ouverture du projet au plus grand nombre. En octobre 1998, la PDB passe sous la gestion du RSCB (Research Collaboratory for Structural Bioinformatics) dont le but est la création d'une ressource basée sur les technologies modernes qui faciliterait l'usage et l'analyse de données structurales et serait propice à la recherche biologique ainsi qu'à l'éducation [Berman, 2008; Berman *et al.*, 2000]. A ce jour, la PDB réunit plus de 70000 structures de protéines, issues de classes et d'espèces très diverses. En 2010, plus de 7000 nouvelles structures y ont été incluses et le nombre de structures ajoutées augmente presque chaque année depuis les années 1980. Elle contient aujourd'hui également des structures d'acides nucléiques et de complexes divers [Bourne *et al.*, 2011; Rose *et al.*, 2011].

La grande majorité des structures présentes dans la PDB sont issues de la

cristallographie aux rayons X mais on trouve également un certain nombre de structures obtenues par RMN. Les premières permettent de rendre compte d'un état cristallin de la protéine, soit sous une forme holo, c'est à dire en co-cristallisation avec un ligand, soit sous une forme apo, sans ligand. Les secondes, qui sont obtenues à partir d'échantillons en solution aqueuse, fournissent plutôt une représentation dynamique de la protéine et se présentent la plupart du temps sous la forme d'un ensemble de structures. Les structures cristallines, qui sont en quelque sorte une photo prise de la protéine dans un état dépendant de l'environnement et conditionné par la présence éventuelle et la nature du ligand, sont les plus commodes à utiliser dans le cadre de la modélisation et donnent les meilleurs résultats [Fan *et al.*, 2009; McGovern & Shoichet, 2003]. Elles sont donc la base de toute procédure de criblage virtuel.

Dans les structures holo, les changements conformationnels qui peuvent avoir lieu au sein de la protéine lorsqu'elle se lie à son ligand pour s'adapter à celui-ci, sont un handicap potentiel dans le docking, lorsque les molécules testées sont structurellement éloignées du ligand co-cristallisé et présentent un mode de liaison différent [Fan *et al.*, 2009]. S'il existe plusieurs structures de la protéine cible, il est donc préférable que celle choisie soit représentative du système étudié. Il est recommandé de superposer les modèles, afin de les comparer et de vérifier que celui choisi ne diffère pas sensiblement des autres sans motif apparent. Lorsque la protéine existe sous plusieurs formes en fonction du ligand co-cristallisé, par exemple lorsque la liaison d'un antagoniste provoque un changement conformationnel par rapport à celle d'un agoniste, la structure utilisée devra si possible rendre compte de la nature des ligands recherchés [Liebeschuetz, 2008]. De la même façon, les structures apo, qui sont cristallisées en l'absence de ligand, peuvent être inadaptées à la liaison d'un ligand.

Dans certains cas, il peut arriver que la structure de la protéine d'intérêt n'ait pas été résolue et que l'on ne connaisse pas sa structure tridimensionnelle. A l'aide de la reconstruction par homologie de séquence, il est possible de reconstruire la protéine à partir de structures de protéines homologues, dont la séquence peptidique est suffisamment proche pour pouvoir servir de modèle. La position des chaînes latérales est généralement calculée à partir de banques de rotamères, une étape de minimisation permettant ensuite de relaxer la structure et d'éviter les clashes stériques liés au nouvel environnement. Ces techniques sont à utiliser avec précaution, mais elles peuvent avoir un intérêt dans le cas où une protéine ne possédant pas de structure tridimensionnelle présente une homologie de séquence importante avec une autre protéine dont la structure a été résolue [Kryshtafovych & Fidelis, 2009; Rockey & Elcock, 2006]. Dans certains cas, cette technique possède l'avantage de gommer les effets de l'adaptation de la protéine à son ligand qui peuvent

influencer le docking [Kairys *et al.*, 2006]. Un examen de la littérature concernant le criblage virtuel « structure-based » publiée entre 2000 et 2009 montre que presque 25% des expériences ont été menées sur des modèles reconstruits par homologie de séquence [Tuccinardi, 2009]. Au delà de 50% d'homologie, les structures proposées peuvent être considérées comme fiables, avec un enrichissement cinq fois supérieur aux structures créées à partir de modèle de plus faible homologie [Oshiro *et al.*, 2004]. Sur ce type de structures, l'enrichissement est comparable, voire occasionnellement meilleur que celui obtenu à partir de la structure cristallographique, bien que celles-ci restent très souvent la solution la plus fiable [Kairys *et al.*, 2006]

## 2. PRÉPARATION DES PROTÉINES

Les structures extraites de la Protein Data Bank (PDB) ne sont généralement pas utilisées dans des expériences de docking directement. Tout d'abord, il existe des contraintes liées au format (voir les annexes 1 « Formats des fichiers de molécules » et 2 « Gestion des fichiers et automatisation des calculs ») et aux paramètres qui doivent être intégrés au fichier d'entrée qui vont dépendre du programme utilisé. De plus, on trouve parfois des erreurs dans les structures qui impliquent donc un traitement préalable [Davis *et al.*, 2003].

Tout d'abord, on retrouve dans les fichiers un certain nombre de molécules qui ne font pas partie de la structure de la protéine, notamment des molécules d'eau et éventuellement des sels ou d'autres corps étrangers ayant servi à la cristallisation. En règle générale, tous ces corps étrangers doivent être supprimés pour le docking (voir le paragraphe « Importance du solvant » dans ce même chapitre) pour ne conserver que la structure de la protéine [Søndergaard *et al.*, 2009].

Au sein même de la structure de la protéine, il peut également exister un certain nombre d'erreurs sur les positions des atomes, qui sont déduites d'une position moyenne au cours du temps. L'interprétation des données de densité électronique, qui nécessite l'emploi de programmes d'affinage permettant de faire correspondre ces données avec un modèle reconstruit par l'expérimentateur, peut mener à quelques approximations qui dépendent considérablement de la résolution en Å [Davis *et al.*, 2003; Tuccinardi, 2009]. A une résolution de 2.0 Å, par exemple, ce qui est un chiffre assez classique, l'erreur moyenne sur la position des atomes peut atteindre 0.3 Å [Sousa *et al.*, 2006]. Les structures résultantes présentent donc à l'occasion des angles et longueurs de liaison atypiques qui peuvent être ramenés à une valeur connue par le biais de la minimisation afin d'obtenir une structure relaxée de plus basse énergie.

Certaines structures présentent parfois des parties manquantes. Lorsque ces boucles sont suffisamment éloignées du site actif, elles peuvent être ignorées, mais des outils

permettent également de les reconstruire (par exemple à l'aide du programme MODELLER [Eswar *et al.*, 2006; Fiser *et al.*, 2000; Martí-Renom *et al.*, 2000; Sali & Blundell, 1993]). Toutefois, lorsque de telles parties manquantes sont situées à proximité ou dans le site actif, il est souvent plus prudent de considérer la structure comme impropre à l'utilisation dans un criblage virtuel.

### 3. IMPORTANCE DU SOLVANT

La grande majorité des procédures de docking de routine font intervenir une protéine et un ligand sans prendre en compte le solvant, que ce soit de façon implicite ou explicite. Cependant, cette approche est assez éloignée de la réalité si on veut simuler le milieu biologique dans lequel la protéine et le ligand vont entrer en contact. Ce problème est parfois traité *a posteriori* dans le scoring des poses obtenues, avec des fonctions de score comme GB/SA dans Dock [Liu *et al.*, 2004] par exemple qui simulent la solvation de façon implicite. Ces méthodes demandent un temps de calcul assez long et sont rarement employées en tant que fonction de scoring primaire, c'est à dire pour évaluer les poses générées à chaque étape pendant le processus de docking.

Globalement, un ligand est capable d'interagir avec une protéine si la variation de l'énergie du système est négative, l'énergie du complexe étant plus basse que l'énergie des deux parties isolées et solvatées. Ceci implique donc que l'ensemble des interactions nouvellement formées soit suffisamment favorable pour contrebalancer les pertes énergétiques occasionnées par la désolvation des deux parties. Lors de la liaison du ligand à la protéine, un grand nombre de liaisons hydrogène entre la protéine et le solvant ainsi qu'entre le ligand et le solvant vont être rompues. Dans le même temps, des liaisons hydrogène vont se former entre le ligand et la protéine ainsi qu'entre les molécules d'eau éjectées du site actif et le solvant environnant. Cet échange, selon s'il est favorable ou non, peut avoir une grande influence sur la capacité du ligand à se lier à la protéine [Davis & Teague, 1999; Søndergaard *et al.*, 2009].

L'hydrophobicité du ligand et du site actif vont également jouer un rôle important dans l'assemblage du complexe. En effet, plus le site actif et le ligand sont hydrophobes, plus leur désolvation et leur liaison seront favorables, puisque les molécules d'eau éjectées seront alors dans un état de plus basse énergie. Ceci est particulièrement parlant lorsque l'on sait qu'en moyenne le nombre d'atomes de nature hydrophobe dans les molécules actives est de seize, alors que ces mêmes molécules ne comprennent qu'un ou deux donneurs et trois ou quatre accepteurs de liaisons hydrogène [Davis & Teague, 1999].

Dans certains cas, le solvant peut jouer un rôle important dans les interactions entre la protéine et les petites molécules, en agissant comme un médiateur de liaison hydrogène

entre un donneur/accepteur de l'hôte et un accepteur/donneur du ligand lorsqu'ils sont trop éloignés pour interagir directement. Selon une étude réalisée par Sondergaard et al., sur 1300 structures plus de 88% présentaient au moins une molécule d'eau à proximité du ligand (dans un rayon de 3.5 Å) susceptible de créer un relai pour la formation d'une liaison hydrogène avec la protéine [Søndergaard *et al.*, 2009]. Ces molécules d'eau peuvent même s'avérer nécessaires pour l'arrimage du ligand, et dans ce cas il paraît donc essentiel d'en tenir compte pendant le docking. Bien que leur rôle essentiel dans la liaison ligand-protéine ait souvent été souligné, peu d'applications concrètes les prennent en compte [Sousa *et al.*, 2006]. Une solution assez simple pour remédier à ce problème est de garder une molécule d'eau au sein du site actif pendant le calcul, celle-ci étant considérée comme faisant partie du récepteur [Liebeschuetz, 2008]. Cependant, du point de vue biologique, une molécule d'eau présente dans le site actif sera très probablement mobile et l'incorporation d'une molécule d'eau fixe ne sera certainement pas capable de rendre compte des interactions possibles. De façon approximative, la molécule d'eau peut éventuellement être déplacée manuellement pour créer plusieurs modèles du site actif [Clark, 2008], ou autorisée à apparaître et à disparaître pendant le calcul comme cela a été proposé avec GOLD [Verdonk *et al.*, 2005]. Pour une approche vraiment réaliste, il faudrait en quelque sorte effectuer un double docking simultané du ligand et de la molécule d'eau. Or ceci n'est pas possible dans les différents programmes de docking à notre disposition. Lorsque les informations disponibles sont trop peu nombreuses pour évaluer l'influence du solvant sur un système particulier, il existe différents programmes qui permettent de prédire si une molécule de solvant est très fortement liée à une partie de la cavité tels que GRID [Goodford, 1985; Wade *et al.*, 1993; Wade & Goodford, 1993], HINT [Amadasi *et al.*, 2006; Amadasi *et al.*, 2008] ou SuperStar [Verdonk *et al.*, 1999; Verdonk *et al.*, 2001]. Il existe aussi des méthodes empiriques comme Waterscore [García-Sosa *et al.*, 2003] qui sont capable d'évaluer si une molécule de solvant est déplaçable [Cole *et al.*, 2011].

#### 4. GESTION DE LA FLEXIBILITÉ DE LA PROTÉINE

Classiquement, le docking fait interagir des ligands flexibles sur une protéine rigide. Pourtant, tout comme les ligands vont adopter une conformation particulière lors de la liaison, la structure du site actif de la protéine peut elle aussi être influencée par son ligand. Ceci est facilement observable lorsque l'on compare deux structures de la même protéine liée à des ligands différents, où de subtiles mais parfois essentielles différences vont apparaître [Kitchen *et al.*, 2004]. Ces modifications concernent en majorité la position des chaînes latérales des acides aminés, mais on peut également, selon les protéines, observer un déplacement du squelette, et jusqu'à des mouvements de domaines entiers de la protéine

qui vont permettre par exemple l'activation ou l'inactivation du système [Carlson, 2002; Durrant & McCammon, 2010]. Deux explications cohabitent concernant ce phénomène. Le premier modèle implique l'existence de plusieurs états conformationnels distincts pour la protéine, et la sélection par les ligands d'un état préférentiel. Dans le second modèle, les changements conformationnels sont induits par la présence du ligand, c'est ce que l'on appelle l'« induced fit » [Perola & Charifson, 2004]. Qu'elles soient possibles uniquement en présence du ligand ou qu'elles aient lieu spontanément en son absence, ces variations dans la structure de la protéine sont un obstacle lors de la mise en place de procédures de criblage virtuel basé sur la structure. La flexibilité de la protéine est actuellement considérée comme le défi majeur dans ce domaine [Sousa *et al.*, 2006]. Dans la pratique, une mobilité complète de la protéine n'est pas envisageable lors du docking dans un délai adaptable au criblage virtuel, mais des méthodes ont été proposées qui permettent de tenir compte d'une certaine adaptation de la structure, soit au niveau local, soit de l'ensemble de la protéine. On peut les diviser en trois catégories, selon la phase au cours de laquelle la flexibilité de la protéine est explorée.

- *Exploration pré-docking*

Dans une optique assez proche du modèle de sélection conformationnelle, il est possible de considérer plusieurs modèles pour la protéine et d'effectuer le docking sur un ensemble de structures. Ces différents modèles peuvent être créés à partir de plusieurs structures cristallographiques issues de la PDB contenant des ligands co-cristallisés différents, ou de structures obtenues par RMN [Carlson, 2002]. On peut également générer plusieurs conformations à partir de différentes minimisations, ou encore à l'aide de la dynamique moléculaire [Durrant & McCammon, 2010]. Ensuite, selon les cas, il est possible de comparer les modèles afin de déterminer lequel est le plus adapté au système étudié ou même de réaliser une sorte de « consensus docking » en considérant cet ensemble comme un tout et en ne tenant compte que des meilleurs résultats pour tous les modèles confondus. Un exemple est la procédure dite RCS (Relaxed Complex Scheme) mise en place par Amaro *et al.*, qui utilise un ensemble de structures générées par dynamique moléculaire avant de réaliser un docking avec Autodock [Amaro *et al.*, 2008]. Une autre solution nommée IFREDA (pour ICM Flexible REceptor Docking Algorithm) a été proposée par Cavasotto et Abagyan. Elle consiste à réaliser un premier docking avec ICM pendant lequel plusieurs ligands flexibles sont dockés dans le récepteur en autorisant les mouvements sur les chaînes latérales et le squelette afin de créer un ensemble de structures de la protéine, qui sont ensuite utilisées au cours d'un docking classique ligand flexible [Cavasotto & Abagyan, 2004]. Ces méthodologies sont les plus adaptées à l'étude de changements

conformationnels de grande échelle [Sousa *et al.*, 2006].

- *Exploration au cours du docking*

Certains programmes de docking sont capables de prendre en compte la flexibilité de la protéine au cours du calcul, à travers l'utilisation de banques de rotamères afin de faire varier les torsions sur les chaînes latérales des résidus à proximité du site actif [Leach, 1994], ou en appliquant à l'échantillonnage des chaînes latérales des résidus l'algorithme de recherche conformationnelle utilisé pour les ligands. C'est le cas notamment dans Autodock [Morris *et al.*, 1998] et Autodock Vina [Trott & Olson, 2010]. Bien entendu, l'augmentation du nombre de liaisons flexibles dans le système (ligand plus protéine) s'accompagne d'une augmentation exponentielle du temps de calcul et cette possibilité doit donc être gérée de manière à rester dans des limites raisonnables. L'appréciation de ces limites dépend d'une part du type d'expérience menée : discrimination d'actifs potentiels ou exploration des mécanismes de liaison par exemple et d'autre part du nombre de ligands étudiés qui peut aller de plusieurs dizaines lors de criblages méthodologiques utilisant des bases orientées à quelques centaines ou milliers de molécules au moins dans le cadre du criblage d'une base de molécules commerciales.

Une étape critique dans ce processus est la sélection des acides aminés sur lesquels intégrer de la flexibilité pendant le calcul. Pour cela, Anderson *et al.* ont développé l'outil SOFTSPOTS, qui permet d'identifier les résidus à proximité du ligand qui sont susceptibles de subir un changement conformationnel pendant l'arrimage du ligand. Selon eux, les résidus hydrophobes sont les plus disposés à changer d'orientation, tandis que les résidus hydrophiles sont généralement peu mobiles [Anderson *et al.*, 2001]. Le programme FlexE, dérivé de FlexX, se base sur la comparaison de plusieurs structures, généralement obtenues expérimentalement mais qui peuvent être issues d'une dynamique moléculaire ou de minimisations, pour définir des régions mobiles et des régions fixes dans la protéine. De nouvelles conformations sont ensuite générées en faisant varier les domaines mobiles sur une base fixe déduite des positions moyennes des régions non mobiles pour créer un ensemble de structures pertinentes pharmacologiquement [Claussen *et al.*, 2001].

- *Exploration post-docking*

Le « soft-docking » a été l'une des premières méthodes simulant une flexibilité locale de la protéine, dès 1991 [Jiang & Kim, 1991]. Grâce au remplacement du potentiel de Lennard-Jones par une fonction plus tolérante qui accepte une interpénétration mineure des rayons de van der Waals des atomes, cette méthode permet d'introduire de manière implicite la flexibilité sur la protéine en autorisant dans une certaine mesure l'existence de clashes

stériques entre le ligand et la protéine [B-Rao *et al.*, 2009; Durrant & McCammon, 2010].

Beaucoup de programmes de docking permettent en outre une optimisation du complexe ligand-protéine après le docking, soit à travers des algorithmes de minimisation, soit par le biais de la dynamique moléculaire.

## II. PROBLÉMATIQUES LIÉES AUX LIGANDS

La nature et les propriétés des ligands criblés ont une influence considérable sur l'obtention de hits. En forçant le trait, il paraît évident qu'un criblage impliquant des molécules aux propriétés totalement incompatibles avec une utilisation biologique est voué à l'échec. Or l'espace chimique couvert par les molécules biologiquement pertinentes représente une portion extrêmement faible de l'espace chimique global, même réduit aux molécules carbonées de taille moléculaire inférieure à 500 Da typiques des molécules biologiques [Dobson, 2004]. Comme dans le choix et la préparation de la protéine, les décisions concernant les ligands doivent donc être faites de manière raisonnée afin de garantir le succès du criblage virtuel.

### 1. BASES DE LIGANDS

Deux types de bases de ligands peuvent être distinguées selon leur origine.

Les bases de ligands les plus couramment utilisées sont issues de banques de données disponibles sur internet, dans la littérature ou par le biais des vendeurs [Tuccinardi, 2009]. Elles présentent de multiples intérêts, dont le plus évident est leur disponibilité immédiate. S'agissant de molécules commerciales en particulier, elles peuvent être achetées facilement afin de procéder à un test expérimental, dès lors qu'une liste de composés prioritaires a été obtenue. De plus, ces bases sont le plus souvent annotées, ce qui permet de se renseigner sur les éventuelles activités biologiques déjà connues pour ces molécules. Elles peuvent ainsi être choisies par exemple pour leurs propriétés en accord avec le système étudié, ou au contraire pour leur diversité.

La seconde solution consiste à créer des bases de ligands à partir du matériel disponible dans le laboratoire, par exemple en demandant aux chimistes organiciens des modèles de molécules qu'ils pensent pouvoir synthétiser, ou tout simplement à partir de fragments [Tuccinardi, 2009]. Le plus souvent, cette stratégie est utilisée sur des séries de molécules possédant un noyau commun et plusieurs points de diversité, afin de créer des bases orientées destinées à répondre à une cible bien précise ou à un savoir-faire local.

Quelle que soit l'origine des composés criblés, la construction d'une base de molécules implique généralement l'application d'un certain nombre de filtres employés pour s'assurer

que les molécules possèdent des propriétés compatibles avec une application biologique. On parle de propriétés « drug-like ». Les bases de ces caractéristiques ont été établies par Lipinski et al. à travers la « rule-of-five » ou « règle des cinq ». Elles décrivent de façon simple quelques indicateurs de la biodisponibilité que sont la masse moléculaire, la lipophilicité et le nombre de donneurs et d'accepteurs de liaisons hydrogène [Lipinski *et al.*, 1997]. Ces règles ont évolué depuis pour répondre aux nouvelles connaissances dans les domaines de la pharmacodynamique et de la pharmacocinétique, mais le principe est resté le même, à savoir établir un certain nombre de critères permettant de réduire la base de ligands à l'essentiel, ce qui revient idéalement à sélectionner uniquement des molécules pertinentes biologiquement, tout en n'en excluant aucune [Leeson & Springthorpe, 2007; Vistoli *et al.*, 2008].

## 2. STRUCTURE DES LIGANDS

Les bases de ligands disponibles sur internet sont accessibles le plus souvent au format SMILES ou SD, mais ne rendent pas compte de la structure 3D des molécules (voir l'annexe 1 « Formats de fichiers de molécules »). Avant de pouvoir utiliser ces molécules dans un docking, il est donc nécessaire de générer des conformations 3D chimiquement valides et énergétiquement favorables. Lorsque des ligands sont créés au sein du laboratoire, ils peuvent être dessinés directement en 3D à partir d'une interface graphique, mais ceci est relativement long. De plus, quand il s'agit de créer de la diversité à partir d'un motif structural de base, il est facile d'automatiser la procédure de génération des composés en utilisant le code SMILES par exemple.

Il est donc essentiel de pouvoir générer de façon automatique les structures 3D des molécules, à partir du code SMILES ou de structures en 2D [Knox *et al.*, 2005]. Il existe pour cela plusieurs programmes comme CAESAR (Accelrys, San Diego, USA) [Li *et al.*, 2007], CORINA (Molecular Networks GmbH, Erlangen, Germany) [Sadowski *et al.*, 1994], OMEGA (OpenEye Scientific Software Inc., Santa Fe, USA) [Boström *et al.*, 2003; Hawkins *et al.*, 2010], Catalyst (Accelrys, San Diego, USA) [Kurogi & Güner, 2001; Smellie *et al.*, 1995]. Au cours de cette étape, il peut être intéressant de vérifier l'existence de tautomères, d'états de protonation dans une gamme de pH raisonnable, ainsi que d'éventuels stéréoisomères lorsque la chiralité de la molécule n'est pas définie [Mohan *et al.*, 2005]. Sachant qu'il est difficile de prévoir quel état est le plus probable, notamment si on considère que cet état peut être influencé par la liaison à la protéine, il est préférable de réaliser une énumération complète des différentes formes lors de la génération de la base [Cummings *et al.*, 2007; Knox *et al.*, 2005]. Cette étape augmente la taille de la base, ce qui est potentiellement pénalisant en terme de temps de calcul, mais peut s'avérer nécessaire pour l'identification de

molécules d'intérêt, comme l'ont récemment montré plusieurs études [Kalliokoski *et al.*, 2009; Martin, 2009; ten Brink & Exner, 2009].

Si elle est nécessaire pour le criblage virtuel de bases de ligands de taille importante, la gestion automatique des structures peut être une source d'erreurs [Liebeschuetz, 2008]. Dans ce cas, la vérification des conformations dockées se fait généralement *a posteriori*, seulement sur les composés identifiés par le criblage comme présentant un intérêt [Cummings *et al.*, 2007].

### III. AUTOMATISATION DES CALCULS ET GESTION DES FICHIERS

Les moyens de calcul à disposition actuellement permettent de cribler virtuellement des milliers voire des millions de petites molécules dans un temps raisonnable [Cummings *et al.*, 2007]. Lorsque l'on travaille avec des bases de données de cette taille, et même à partir de quelques centaines de ligands, il devient très difficile et extrêmement chronophage de préparer chaque ligand indépendamment de façon manuelle. En plus de cela, il existe de nombreux formats de fichiers de structures, qui possèdent chacun leurs spécificités, et les différents programmes de docking et de scoring font appel à des fichiers d'entrée dans des formats bien précis. Ceci pose un problème de compatibilité lorsque l'on passe de la structure de la protéine et du ligand au docking, puis du docking au rescoring. Outre les corrections et les transformations qui sont nécessaires sur les structures brutes des deux partenaires, les fichiers doivent donc être convertis vers les formats requis en entrée des programmes de docking. De même, les fichiers de sortie du docking nécessitent parfois une conversion pour pouvoir servir d'entrée pour les programmes utilisés dans le rescoring ou pour l'analyse des données.

Les programmes de visualisation tels que Discovery Studio (Accelrys, San Diego, USA) ou PyMOL (PyMOL Molecular Graphics System, Schrödinger, LLC) permettent de gérer plusieurs formats de fichiers, et donc de convertir un fichier d'un format vers un autre, mais uniquement à partir d'une interface graphique, ce qui ne convient pas à des bases de données de grande taille, ou grâce à l'emploi de scripts intégrés par le biais d'une API (Application Programming Interface), ce qui nécessite une expertise en programmation. La suite OpenBabel ([www.openbabel.org](http://www.openbabel.org)) [Guha *et al.*, 2006; O'Boyle *et al.*, 2011] a été spécifiquement créée pour permettre la conversion des fichiers de structure à partir de la ligne de commande, mais ne permet pas de gérer la totalité des formats existants. De plus, si certains formats de fichiers comme le PDB sont devenus des standards, beaucoup de programmes développent leur propre format de structure selon les besoins, ce qui complique encore la tâche.

Il est donc important de mettre en place des outils permettant d'automatiser la préparation des ligands, le lancement des calculs (y compris si besoin la création de fichiers de configuration) et la gestion des fichiers d'entrée et de sortie, voire la transformation des fichiers de sortie de docking vers des fichiers d'entrée pour le rescoring lorsque les formats sont incompatibles. Chaque étape de conversion ou d'automatisation peut être à l'origine d'erreurs, et doit donc être contrôlée, avant le lancement des calculs, ou de manière post-expérimentale. L'annexe 2 « Gestion des fichiers et automatisation des calculs » fournit le détail des actions mises en place dans le laboratoire pour automatiser les calculs et gérer les différents formats de fichiers.

## E. OBJECTIFS

Au cours de ce travail, nous nous sommes intéressés au criblage virtuel « structure-based » pour ses applications dans le cadre de la découverte de nouveaux médicaments et notamment de la recherche de hits. Notre but a été, à partir de bases de ligands commerciaux ou innovants, d'aboutir à une sélection de molécules enrichie en composés potentiellement actifs. Ce processus peut être schématisé en quatre étapes, dont les enjeux sont décrits ci-dessus.

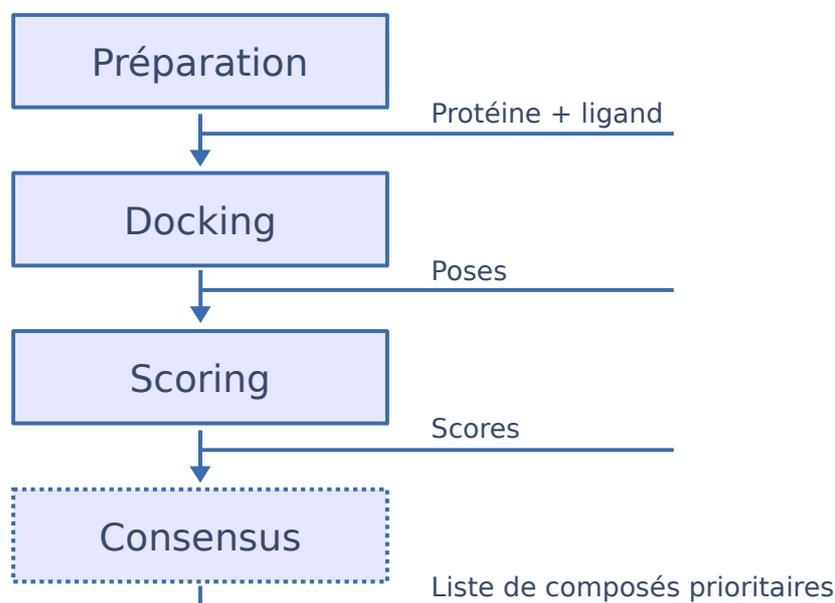


Figure 6. Étapes du protocole de criblage virtuel « structure-based »

Dans un premier temps, nous avons exploré les techniques du point de vue méthodologique, à travers l'exploration de la flexibilité du récepteur qui est aujourd'hui un enjeu majeur dans ce domaine. Suite à cela, nous avons développé des stratégies de criblage et les avons appliquées à des cas concrets de protéines citées dans la littérature pour leur potentiel thérapeutique, en particulier dans les cancers. Grâce à plusieurs collaborations avec des chimistes organiciens pour la partie synthèse, et des biochimistes pour la partie expérimentale, nous avons étudié deux cibles thérapeutiques, la sPLA2 de groupe X, et la Pontine. Les résultats obtenus nous ont permis d'identifier des composés commerciaux qui pourraient faire l'objet d'un repositionnement et qui sont actuellement encore à l'étude, ainsi que d'orienter la synthèse de molécules actives sur les cibles choisies.



## **CHAPITRE 2 : ÉTUDE MÉTHODOLOGIQUE DE LA FLEXIBILITÉ DE LA PROTÉINE**



Avec la participation de Marie-France Bakaï à l'occasion de son stage de licence.

### A. GÉNÉRALITÉS

Dans un premier temps, nous nous sommes intéressés aux aspects méthodologiques du criblage virtuel et plus particulièrement à la flexibilité de la protéine, qui est considérée par beaucoup comme l'enjeu majeur de ces techniques. Pour cela, nous avons considéré deux stratégies de criblage permettant de simuler la flexibilité de la protéine parmi celles décrites dans le chapitre précédent (voir le paragraphe « Gestion de la flexibilité de la protéine »). Dans un premier temps, nous avons mis en place une expérience mettant en œuvre le modèle de sélection de conformation. Pour cela, nous avons réalisé plusieurs criblages parallèles en utilisant trois structures cristallographiques de la protéine, ainsi qu'un ensemble de structures obtenues par minimisation. Ensuite, nous avons exploré une méthode permettant de considérer les mouvements des chaînes latérales au cours du docking, afin d'observer l'« induced fit » sur la protéine.

Pour le docking, les programmes utilisés sont Dock version 5.4 [Ewing *et al.*, 2001; Moustakas *et al.*, 2006] avec ligand flexible et Autodock Vina version 1.0.2 [Trott & Olson, 2010] avec ligand et/ou protéine flexible. Nous avons également utilisé les programmes standalone Drugscore version PDB et CSD [Gohlke *et al.*, 2000; Velec *et al.*, 2005] et Xscore [Wang *et al.*, 2002] pour effectuer un rescoring des résultats.

Pour cette étude, nous avons choisi de nous baser sur des protéines qui ont été intensivement étudiées ces dernières années, les kinases. Ce choix a été motivé par l'abondante littérature concernant cette famille de protéine, y compris dans le domaine de la modélisation moléculaire [Cavasotto & Abagyan, 2004; Dessalew & Bharatam, 2007; Katritzky *et al.*, 2006; Kim *et al.*, 2008; Martin *et al.*, 2006; Naerum *et al.*, 2002; Richardson *et al.*, 2007; Rockey & Elcock, 2005; Taha *et al.*, 2008], et en particulier l'existence de nombreuses structures cristallographiques et de ligands d'affinité connue.

Le kinome humain décrit par Manning *et al.* dans une publication qui fait aujourd'hui office de référence réunit plus de 500 protéines kinases (518 selon Manning *et al.*, 538 d'après les plus récentes estimations [Schwartz & Murray, 2011]), et représente presque 2% du génome humain [Manning *et al.*, 2002], ce qui en fait une des familles de gènes les plus importantes en nombre [Margutti & Laufer, 2007].

Les kinases catalysent la réaction de phosphorylation qui consiste à transférer le groupement  $\gamma$ -phosphate d'un nucléoside triphosphate (généralement l'ATP) vers un acide aminé porteur d'un groupe hydroxyle libre c'est à dire la sérine, la thréonine ou la tyrosine

[Schwartz & Murray, 2011]. On estime à environ 500 000 le nombre de sites de phosphorylation potentiels dans le protéome humain, et actuellement 25 000 événements de phosphorylation ont été décrits dans 7 000 protéines humaines [Lemeer & Heck, 2009]. Cette transformation, extrêmement courante, est connue pour son implication dans un très grand nombre de processus biologiques [Schwartz & Murray, 2011]. Elle permet l'induction de changements conformationnels permettant d'altérer les interactions d'une protéine avec ses partenaires, ou son activité, voire simplement d'inactiver la protéine [Ventura & Nebreda, 2006]. Par ce biais, les kinases sont capables d'affecter la très grande majorité des voies de signalisation cellulaire, et elles ont également un rôle dans de nombreux autres processus biologiques tels que le métabolisme, la transcription, l'apoptose et la différenciation cellulaire par exemple [Fedorov *et al.*, 2007; Manning *et al.*, 2002]. En règle générale, la perturbation des voies de signalisation est un élément majeur dans le développement de pathologies [Fedorov *et al.*, 2007], ce qui fait des kinases des cibles particulièrement attractives du point de vue thérapeutique [Vieth *et al.*, 2004]. Un dysfonctionnement dans les mécanismes de phosphorylation ou une dérégulation des protéines kinases sont souvent associés à des mécanismes pathologiques [Ventura & Nebreda, 2006]. En particulier, des liens très forts ont été identifiés entre la perturbation de la signalisation en raison de mutations ou d'une sur-expression des protéines kinases et le développement de cancers [Fedorov *et al.*, 2010; Margutti & Laufer, 2007]. Les kinases sont d'ailleurs les protéines les plus fréquemment mutées dans les tumeurs, c'est pourquoi elles sont devenues ces dernières années les cibles les plus étudiées dans la recherche de médicaments en oncologie [Fedorov *et al.*, 2010]. A ce jour, quatorze inhibiteurs de kinases ont été approuvés, dont douze pour le traitement de divers types de cancers [Schwartz & Murray, 2011]. En raison du grand nombre de kinases existantes, et de leur forte homologie, un problème récurrent reste la sélectivité des inhibiteurs pour une kinase particulière [Margutti & Laufer, 2007]. De fait, la plupart des inhibiteurs connus possèdent une sélectivité limitée et interagissent avec de multiples kinases [Fedorov *et al.*, 2010], ce qui peut être la source de nombreux effets secondaires indésirables.

## **B. DÉTAILS DU SYSTÈME ÉTUDIÉ**

### **I. STRUCTURES DU RÉCEPTEUR**

Parmi les 538 kinases répertoriées à ce jour, nous avons choisi de nous intéresser à GSK3, qui est non seulement impliquée dans les mécanismes de cancérogenèse, mais

également dans d'autres maladies telles que le diabète de type II, et la maladie d'Alzheimer [Bertrand *et al.*, 2003]. GSK3 est une kinase de type sérine/thréonine dont il existe deux isoformes,  $\alpha$  et  $\beta$ , chez l'homme [Mazanetz *et al.*, 2008]. L'isoforme  $\beta$  notamment a été cristallisée à plusieurs reprises, sous sa forme apo ou en complexe avec divers ligands, les différentes structures disponibles dans la PDB sont présentées dans la Table 1.

| Code pdb | Ligand co-cristallisé                                                                                                 | Référence                            |
|----------|-----------------------------------------------------------------------------------------------------------------------|--------------------------------------|
| 1l09     | Structure apo                                                                                                         | [ter Haar <i>et al.</i> , 2001]      |
| 1h8f     | 4-(2-hydroxyethyl)-1-piperazine                                                                                       | [Dajani <i>et al.</i> , 2001]        |
| 1gng     | Structure apo                                                                                                         | [Bax <i>et al.</i> , 2001]           |
| 1j1b     | Phosphoaminophosphonic acid-adenylate ester                                                                           | [Aoki <i>et al.</i> , 2004]          |
| 1j1c     | ADP                                                                                                                   | [Aoki <i>et al.</i> , 2004]          |
| 1o9u     | 9-methyl-9h-purin-6-amine                                                                                             | [Dajani <i>et al.</i> , 2003]        |
| 1pyx     | Phosphoaminophosphonic acid-adenylate ester                                                                           | [Bertrand <i>et al.</i> , 2003]      |
| 1q3d     | Staurosporine                                                                                                         | [Bertrand <i>et al.</i> , 2003]      |
| 1q3w     | Alsterpaullone                                                                                                        | [Bertrand <i>et al.</i> , 2003]      |
| 1q4l     | 2-chloro-5-[4-(3-chloro-phenyl)-2,5-dioxo-2,5-dihydro-1h-pyrrol-3-ylamino]-benzoic acid                               | [Bertrand <i>et al.</i> , 2003]      |
| 1q41     | Indirubin-3'-monoxime                                                                                                 | [Bertrand <i>et al.</i> , 2003]      |
| 1q5k     | N-(4-methoxybenzyl)-n'-(5-nitro-1,3-thiazol-2-yl)urea                                                                 | [Bhat <i>et al.</i> , 2003]          |
| 1uv5     | 6-bromoindirubin-3'-oxime                                                                                             | [Meijer <i>et al.</i> , 2003]        |
| 2o5k     | 2-(2,4-dichloro-phenyl)-7-hydroxy-1h-benzoimidazole-4-carboxylic acid [2-(4-methanesulfonylamino-phenyl)-ethyl]-amide | [Shin <i>et al.</i> , 2007]          |
| 2ow3     | Bis-(indole)maleimide pyridinophane                                                                                   | [Zhang <i>et al.</i> , 2007]         |
| 3du8     | (7s)-2-(2-aminopyrimidin-4-yl)-7-(2-fluoroethyl)-1,5,6,7-tetrahydro-4h-pyrrolo[3,2-c]pyridin-4-one                    | [Menichincheri <i>et al.</i> , 2009] |

|      |                                                                                                      |                                        |
|------|------------------------------------------------------------------------------------------------------|----------------------------------------|
| 2jld | Ruthenium pyridocarbazole                                                                            | [Atilla-Gokcumen <i>et al.</i> , 2008] |
| 3f7z | 2-(1,3-benzodioxol-5-yl)-5-[(3-fluoro-4-methoxybenzyl)sulfanyl]-1,3,4-oxadiazole                     | [Saitoh <i>et al.</i> , 2009]          |
| 3f88 | 5-[1-(4-methoxyphenyl)-1h-benzimidazol-6-yl]-1,3,4-oxadiazole-2(3h)-thione                           | [Saitoh <i>et al.</i> , 2009]          |
| 3gb2 | 2-methyl-5-(3-{4-[(s)-methylsulfinyl]phenyl}-1-benzofuran-5-yl)-1,3,4-oxadiazole                     | [Saitoh <i>et al.</i> , 2009]          |
| 3i4b | N-[(1s)-2-hydroxy-1-phenylethyl]-4-[5-methyl-2-(phenylamino)pyrimidin-4-yl]-1h-pyrrole-2-carboxamide | [Aronov <i>et al.</i> , 2009]          |
| 3pup | Ruthenium octasporine                                                                                | [Feng <i>et al.</i> , 2011]            |
| 3q3b | 4-(4-hydroxy-3-methylphenyl)-6-phenylpyrimidin-2(5h)-one                                             | [Coffman <i>et al.</i> , 2011]         |

Table 1. Liste des structures cristallines publiées de GSK3 $\beta$ , avec le nom du ligand co-cristallisé lorsqu'il existe et la référence bibliographique

La comparaison de ces structures montre l'existence de plusieurs conformations en fonction de l'inhibiteur co-cristallisé, mettant en évidence de façon expérimentale le phénomène d' « induced fit » sur cette protéine, ou de sélection de conformation.

En 2003, Bertrand *et al.* ont cristallisé cinq complexes impliquant GSK3 $\beta$  et des inhibiteurs de différente nature. Parmi ces cinq modèles, nous en avons étudié deux, la structure 1q3d, qui contient l'inhibiteur staurosporine, et 1q3w co-cristallisé avec l'alsterpallone [Bertrand *et al.*, 2003]. A ceux-ci, nous avons rajouté la structure 1uv5, résolue par Meijer *et al.* avec l'inhibiteur 6-bromoindirubin-3'-monoxime [Meijer *et al.*, 2003]. La différence majeure entre les trois structures se situe au niveau de la boucle riche en glycine (glycine-rich loop) situées entre les résidus G63 et G68<sup>1</sup> [Aoki *et al.*, 2004]. Celle-ci est en effet déplacée de façon notable entre les structures 1q3d et 1q3w qui contiennent des ligands volumineux et non plans, et 1uv5 dont le ligand est plus petit et plan (Figure 7 page suivante).

1 Les acides aminés cités dans ce manuscrit seront indiqués dans le texte en utilisant leur code à une lettre accompagné de la position du résidu au sein de la chaîne.

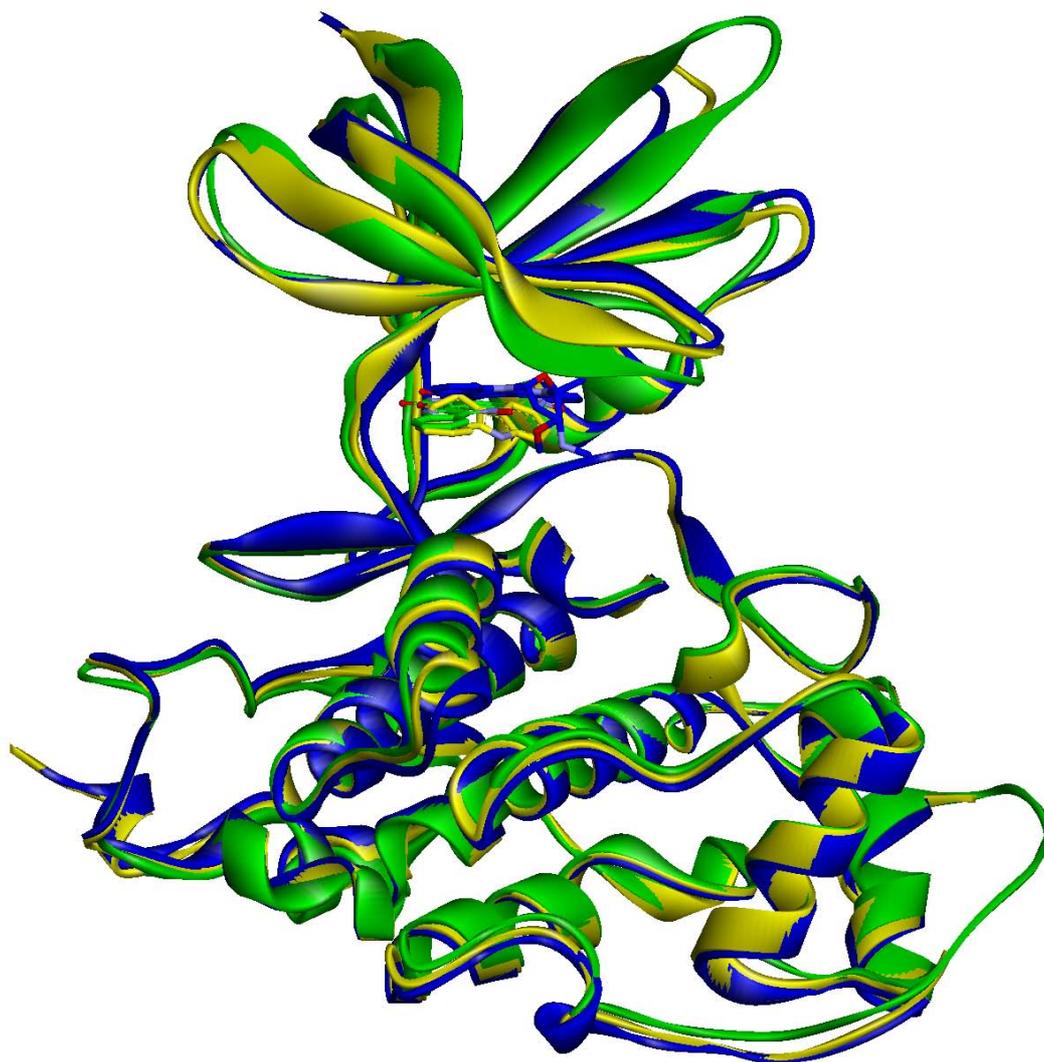


Figure 7. Superposition des structures cristallographiques de 1q3d (bleu), 1q3w (jaune) et 1uv5 (vert) avec les ligands co-cristallisés dans le site actif. La boucle riche en glycine est la boucle située au-dessus à droite des ligands

## II. STRUCTURES DES LIGANDS

Plusieurs familles de petites molécules présentant une grande diversité structurale sont connues pour leur activité inhibitrice sur les kinases, et en particulier sur GSK3 $\beta$ , telles que les indirubines [Beauchard *et al.*, 2006; Leclerc *et al.*, 2001; Meijer *et al.*, 2003; Polychronopoulos *et al.*, 2004], différentes maléimides [Engler *et al.*, 2005; Kuo *et al.*, 2003; Smith *et al.*, 2001; Zhang *et al.*, 2005; Zhang *et al.*, 2003], les paullones [Kunick *et al.*, 2004; Leost *et al.*, 2000] et les pyrazolopyrimidines, pyrazolopyridines et pyrazolopyridazines [Peat

*et al.*, 2004a; Peat *et al.*, 2004b; Tavares *et al.*, 2004; Witherington *et al.*, 2003a; Witherington *et al.*, 2003b; Witherington *et al.*, 2003c; Witherington *et al.*, 2003d].

Pour cette étude méthodologique, nous avons choisi de nous intéresser à une famille particulière qui sont les indirubines et leurs dérivés, pour laquelle des valeurs d'IC<sub>50</sub> (concentration inhibitrice à 50%) sur la protéine GSK3 $\beta$  humaine sont connues. Les indirubines sont des composés constitués de deux indoles reliés par une double liaison, et sont donc entièrement plans si l'on excepte les points de diversités présents sur les deux bicycles (Figure 8). Leurs propriétés anti-tumorales ont été découvertes grâce à leur identification dans un remède chinois traditionnel, le Danggui Longhui Wan, utilisé dans le traitement des leucémies chroniques [Leclerc *et al.*, 2001].



*Figure 8. Structure de l'indirubine non substituée indiquant la numérotation des points de substitution possibles*

Plusieurs publications font état de mesures d'affinité de ces molécules envers GSK3 $\beta$ . Afin de garantir que le même protocole expérimental avait été utilisé pour toutes les mesures d'IC<sub>50</sub> et donc d'assurer la cohérence des données utilisées les valeurs d'IC<sub>50</sub> ont toutes été extraites de la publication de Leclerc *et al.* [Leclerc *et al.*, 2001]. Le test set ainsi créé est constitué de trente-quatre molécules dont les structures sont visibles dans l'annexe 3 « Structures 2D des indirubines ». Dans le tableau regroupant les IC<sub>50</sub>, on peut constater que les différentes molécules ont une activité extrêmement variable, entre 9 nM pour la plus active, la 5-iodoindirubin-3'-monoxime, et 550  $\mu$ M pour la moins active, l'indigo [Leclerc *et al.*, 2001] (Table 2 page suivante).

| Ligand  | IC50 (nM) |
|---------|-----------|
| molla15 | 9         |
| molla14 | 22        |
| molla22 | 33        |
| molla20 | 40        |
| molla23 | 40        |
| molla9  | 42        |
| molla6  | 50        |
| molla5  | 55        |
| molla8  | 62        |
| molla4  | 68        |
| molla7  | 78        |
| molla19 | 80        |
| molla24 | 110       |
| molla16 | 130       |
| molla21 | 180       |
| molla12 | 250       |
| molla10 | 280       |
| molla11 | 350       |
| molla1  | 600       |
| molla13 | 4000      |
| molla38 | 29000     |
| molla43 | 70000     |
| molla39 | 75000     |
| molla37 | 85000     |
| molla42 | 180000    |
| molla18 | 200000    |
| molla26 | 280000    |
| molla34 | 310000    |
| molla40 | 320000    |
| molla25 | 350000    |
| molla27 | 360000    |
| molla41 | 380000    |
| molla44 | 400000    |
| molla2  | 550000    |

Table 2. IC50 des indirubines sur la protéine GSK3 $\beta$

### III. OUTILS UTILISÉS

#### 1. DOCKING

- *Dock 5.4*

Le premier programme Dock [Kuntz *et al.*, 1982] est le précurseur des programmes de docking. A l'époque, l'algorithme était conçu pour réaliser le docking d'un ligand rigide dans le site actif d'une protéine rigide. Aujourd'hui, la version la plus récente de Dock, Dock 6

[Lang *et al.*, 2009], permet le traitement d'un ligand flexible, ainsi que l'utilisation de la dynamique moléculaire pour simuler la flexibilité de la protéine. La version 5.4, que nous avons utilisée, autorise le docking flexible du ligand dans une protéine rigide [Ewing *et al.*, 2001; Moustakas *et al.*, 2006].

L'algorithme de recherche conformationnelle utilisé est un algorithme de reconstruction incrémentale de type « anchor-and-grow » basé sur le docking d'un fragment rigide de la molécule (nommé « ancre »), puis la croissance du ligand complet à partir de ce fragment. Le docking se fait à l'aide de sphères qui constituent une empreinte négative du site actif et qui représentent les positions que peuvent adopter les atomes des fragments. La première étape est donc le découpage du ligand en plusieurs fragments, séparés par les liaisons flexibles, puis la sélection du plus gros fragment qui sera utilisé comme ancre. Ensuite, ce fragment rigide est docké dans le site actif en faisant correspondre les atomes avec les sphères du site actif, puis sa position est optimisée à l'aide d'une minimisation. Toutes les poses obtenues sont ensuite scorées (fonction de score primaire), et seules les meilleures sont conservées pour l'étape suivante. Ce processus est ensuite reconduit avec la couche suivante, c'est à dire en rajoutant un fragment sur l'ancre, et ainsi couche par couche jusqu'à la reconstruction complète du ligand. La conformation finale obtenue est enfin évaluée à l'aide de la fonction de scoring secondaire.

La fonction de scoring par défaut de Dock est une fonction basée sur le champ de force AMBER [Ewing & Kuntz, 1997; Meng *et al.*, 1992] qui ne prend en compte que les interactions entre le ligand et la protéine [Moustakas *et al.*, 2006]. Le récepteur n'étant pas flexible, il peut être modélisé à l'aide d'une grille, c'est à dire que l'espace de recherche est divisé en points dans lesquels sont stockés les contributions des atomes de la protéine. Le calcul de score est ainsi simplifié puisqu'il est calculé à partir des interactions entre les atomes du ligand et les points de la grille. L'accélération des calculs résultant de cette approche est très importante, en raison du grand nombre de poses générées à chaque étape qui doivent être évaluées par la fonction de scoring primaire.

- *Autodock Vina 1.0.2*

Autodock Vina [Trott & Olson, 2010] est un programme de docking dérivé d'Autodock [Morris *et al.*, 1998]. Cependant, malgré son nom, l'algorithme de recherche ainsi que la fonction de scoring utilisés dans Autodock Vina sont différents de ceux d'Autodock, ce qui en fait un nouveau logiciel à part. Ces programmes sont tous les deux conçus pour être capables de gérer la flexibilité de la protéine au cours du docking, via des torsions sur les chaînes latérales de résidus choisis qui sont traitées de la même façon que les torsions des ligands. Une comparaison de ces deux outils a été publiée très peu de temps après

Autodock Vina et montre une plus grande efficacité de ce dernier, notamment dans le traitement des ligands avec un nombre de torsions important. De plus, il est plus rapide et permet de tirer avantage de multiples processeurs [Chang *et al.*, 2010].

Autodock Vina utilise un algorithme de recherche d'optimisation globale stochastique de type Iterated Local Search déjà éprouvé par le programme de docking ICM [Abagyan *et al.*, 1994]. Au cours du docking, plusieurs calculs parallèles vont être effectués à partir de conformations aléatoires. Chacun de ces calculs est réalisé sur un certain nombre de pas, qui consistent en une perturbation aléatoire de la conformation suivie d'une étape d'optimisation locale réalisée selon la méthode Broyden-Fletcher-Goldfarb-Shanno (BFGS) [Nocedal & Wright, 2006]. Après chaque pas, la conformation obtenue est acceptée ou rejetée selon le critère de Metropolis, le nombre de pas étant établi en fonction de la complexité du problème étudié, c'est à dire de la taille de l'espace de recherche et du nombre de torsions autorisées sur le ligand, voire sur la protéine lorsque c'est le cas. Les résultats de tous les calculs sont ensuite rassemblés et clusterisés afin de fournir une liste de solutions définitives classées en fonction de leur score. Le paramètre « exhaustiveness » (exhaustivité) régit le nombre de conformations aléatoires et donc de calculs menés sur le système. Ces calculs pouvant être réalisés en parallèle sur des processeurs différents, ce paramètre est directement lié au nombre de processeurs employés. C'est le seul élément qui permette d'influencer le résultat de docking, Autodock Vina étant conçu pour pouvoir être utilisé de façon très simple.

La fonction de scoring utilisée par Autodock Vina permet de faire la somme des contributions intramoléculaires et intermoléculaires du système d'une manière inspirée par le programme Xscore [Wang *et al.*, 2002] qui utilise une fonction de score empirique. Elle s'appuie sur un training set issu de la PDBbind [Wang *et al.*, 2005] cependant, contrairement à son modèle, la fonction de score d'Autodock Vina n'exploite pas seulement les valeurs d'affinité des ligands, mais également les préférences conformationnelles des complexes du training set. Elle combine ainsi les avantages des fonctions de score empiriques et « knowledge-based » [Trott & Olson, 2010].

Pour des raisons de commodité, le programme Autodock Vina, ou la fonction de scoring issue de ce programme seront appelés Vina dans la suite de ce travail.

## 2. SCORING

- *Xscore*

Les fonctions de score implémentées dans le programme standalone Xscore [Wang *et al.*, 2002] sont de type empirique, c'est à dire qu'elles sont conçues pour être capable de

reproduire les valeurs d'affinité observées expérimentalement en réalisant la somme de termes simples et indépendants. Cette méthodologie possède un avantage certain en terme de temps de calcul en comparaison d'approches basées sur la mesure des interactions que sont les fonctions dérivées des champs de force. Cependant, ces fonctions sont directement dérivées à partir d'un training set de complexes d'affinité connue, et possèdent donc par nature une forte dépendance envers l'ensemble utilisé lors du calibrage. L'utilisation d'un training set varié de complexes possédant des caractéristiques diverses peut permettre d'envisager une fonction non-biaisée, ne répondant pas à un type de protéine particulier [Wang *et al.*, 2002].

Les fonctions de Xscore décrivent la variation de l'énergie au cours de la liaison ligand-protéine sous la forme :

$$\Delta G_{\text{bind}} = \Delta G_{\text{vdw}} + \Delta G_{\text{Hbond}} + \Delta G_{\text{déformation}} + \Delta G_{\text{hydrophobe}} + \Delta G_0$$

où  $\Delta G_{\text{vdw}}$  représente les interactions de van der Waals entre le ligand et la protéine,  $\Delta G_{\text{Hbond}}$  les liaisons hydrogène,  $\Delta G_{\text{déformation}}$  les pénalités de déformation et  $\Delta G_{\text{hydrophobe}}$  les interactions hydrophobes.  $\Delta G_0$  est une constante qui inclut de manière implicite la perte entropique liée à la formation du complexe. Les trois fonctions indépendantes, HPscore, HMscore et HSscore implémentées dans Xscore diffèrent par leur façon de traiter les interactions hydrophobes. La dernière fonction, AVE est un consensus des trois autres matérialisé sous la forme d'une simple moyenne arithmétique.

Pour faciliter la compréhension, les fonctions de Xscore seront notées de la façon suivante tout au long de ce manuscrit: HPscore, HMscore, HSscore et AVEscore.

- *Drugscore*

Les fonctions du programme Drugscore [Gohlke *et al.*, 2000; Velec *et al.*, 2005] sont des fonctions de type « knowledge-based ». Comme les fonctions empiriques, elles exploitent les données expérimentales de complexes connus, mais au lieu de se baser sur les affinités mesurées, elles utilisent directement des données fournies par les structures. Elles se basent sur le principe selon lequel les interactions favorables entre paires d'atomes sont reflétées dans les statistiques de distribution des distances interatomiques [Gohlke *et al.*, 2000]. Cette méthode, bien que reposant sur des résultats empiriques, utilise toutes les valeurs obtenues par paires atomiques dans le training set, ce qui représente statistiquement un nombre bien plus important de données que les fonctions empiriques classiques, d'où une plus faible dépendance envers l'ensemble utilisé pour le calibrage.

Deux versions de Drugscore existent, qui varient selon le training set utilisé pour la

dérivation des données statistiques. Dans la première version, Drugscore PDB, [Gohlke *et al.*, 2000] les potentiels ont été extraits de structures de complexes ligand-protéine issus de la PDB. Au contraire, dans la seconde version, Drugscore CSD, [Veleg *et al.*, 2005], les données ont été obtenues à partir des structures cristallographiques de la CSD (Cambridge Structural Database) qui ne contient que des petites molécules. En s'appuyant sur l'idée que les phénomènes physiques qui régissent les interactions non liantes sont les mêmes dans les petites molécules qu'au sein de complexes, l'utilisation de structures de petites molécules permet en effet de s'affranchir du bruit créé par le très grand nombre d'atomes et donc de paires atomiques dans les complexes macromoléculaires de type ligand-protéine [Veleg *et al.*, 2005].

Les potentiels obtenus sont de deux sortes, soit des potentiels de paire dépendants de la distance, soit des potentiels atomiques de surface accessible au solvant. Il existe donc deux fonctions de score pour chaque version de Drugscore, qui sont les fonctions PAIR et SURF, sachant que les secondes sont identiques pour les deux versions. Comme pour Xscore, une fonction combinant les deux autres est également présente dans chaque version, qui est calculée comme la somme des fonctions PAIR et SURF et porte le nom de PAIRSURF.

Les différentes fonctions issues de Drugscore seront notées SURF, PAIR\_pdb, PAIRSURF\_pdb, PAIR\_csd et PAIRSURF\_csd dans la suite de ce travail.

#### **IV. STRATÉGIE DE CRIBLAGE**

Le but de cette étude était d'observer l'influence de la flexibilité du récepteur lors du criblage virtuel. Plus précisément, nous nous sommes penchés sur le classement de ligands appartenant à une même famille, un objectif réputé difficile à atteindre (voir le chapitre « Introduction »). Ce travail était également une manière de comparer les outils à notre disposition, afin de mettre en place des procédures viables pour des applications concrètes impliquant le criblage de bases de ligands et/ou de cibles inconnues.

Trois expériences ont été réalisées afin de simuler la flexibilité de la protéine. Dans un premier temps un docking a été réalisé sur les trois structures rigides de GSK3 $\beta$  extraites de la PDB qui montrent des conformations différentes. Ensuite, nous avons créé un ensemble de modèles relaxés à partir de ces trois structures en utilisant plusieurs étapes de minimisation, dans une procédure inspirée du Relaxed Complex Scheme proposé par Amaro *et al.* [Amaro *et al.*, 2008]. Enfin, un docking flexible a été réalisé avec Vina.

## C. MATÉRIEL ET MÉTHODES

### I. PRÉPARATION DES RÉCEPTEURS

Les structures de GSK3 $\beta$  co-cristallisée avec la staurosporine, code PDB 1q3d, avec l'alsterpaullone, code PDB 1q3w [Bertrand *et al.*, 2003] et avec la 6-bromoindirubin-3'-monoxime, code PDB 1uv5 [Meijer *et al.*, 2003] ont été extraites de la PDB. Dans les trois fichiers, les molécules d'eau ont été supprimées, puis tous les hydrogènes manquants ont été rajoutés à l'aide du champ de force Charmm dans Discovery Studio 2.1 (Accelrys, San Diego, USA). Dans les fichiers 1q3d et 1q3w, deux chaînes identiques sont présentes, la chaîne B a donc été supprimée afin de conserver uniquement la chaîne A. Le fichier 1uv5 présente seulement la chaîne A.

Afin de corriger d'éventuelles erreurs de structure, les complexes nettoyés ont subi une double minimisation, incluant 1000 pas avec un algorithme de type Steepest Descent, gradient 0,01, puis 1000 pas de Conjugate Gradient, gradient 0,01. Au cours de cette minimisation les atomes du squelette et le ligand ont été immobilisés à l'aide d'une contrainte de force (tethering). Le maintien du ligand dans le site actif permet de conserver la spécificité de la structure envers le ligand co-cristallisé.

### II. RELAXED COMPLEX SCHEME

Au delà de la première minimisation qui a pour but de corriger la structure cristallographique, nous avons réalisé une succession de minimisations en utilisant des contraintes sur le squelette de la protéine de moins en moins élevées, afin de créer un ensemble de structure de plus en plus relaxées (Figure 9). A chaque étape, la structure obtenue à l'étape précédente est réutilisée comme état initial du nouveau cycle. Comme précédemment, la minimisation comprend 1000 pas de Steepest Descent, gradient 0,01, et 1000 pas de Conjugate Gradient, gradient 0,01.

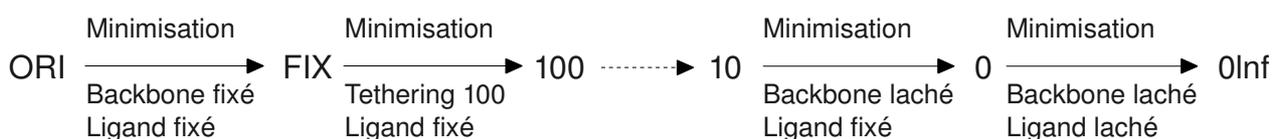


Figure 9. Protocole de minimisation du Relaxed Complex Scheme

Au final, en partant de la structure FIX, dans laquelle squelette et ligand sont fixés, jusqu'aux structures T0 dans laquelle plus aucune contrainte ne pèse sur le squelette et

T0LNF où le ligand est également libre, on obtient sept modèles différents. Les valeurs de RMSD entre la structure de départ et les modèles obtenus indiquent que le squelette (RMSD C $\alpha$ ) est relativement peu déformé jusqu'à l'étape T10. Au contraire, les chaînes latérales (RMSD SC) sont fortement déplacées dès la première étape de minimisation, puis subissent un second changement brutal à l'étape T0 (Table 3).

| Etape | RMSD (C $\alpha$ ) | RMSD (SC) |
|-------|--------------------|-----------|
| FIX   | 0                  | 0.92      |
| T100  | 0.019              | 0.928     |
| T50   | 0.042              | 0.954     |
| T25   | 0.071              | 0.965     |
| T10   | 0.113              | 0.985     |
| T0    | 0.951              | 1.541     |
| T0LNF | 0.964              | 1.559     |

Table 3. RMSD entre la structure de départ et les structures obtenues par minimisation

### III. PRÉPARATION DES LIGANDS

Les trente-quatre ligands dont l'IC50 est fournie dans la publication de Leclerc et al. [Leclerc *et al.*, 2001] ont été reconstruits manuellement en 3D dans l'interface graphique de Discovery Studio 2.1, puis minimisés à l'aide du protocole *Dreiding Minimize* afin d'obtenir une conformation de basse énergie. Les douze ligands cités dans la publication de Leclerc et al. pour lesquels une IC50 n'a pas pu être mesurée en raison d'une affinité trop faible pour GSK3 $\beta$  ont été écartés de l'étude, dans laquelle le but ultime était d'établir une corrélation entre la mesure expérimentale et le score obtenu dans le criblage. Les ligands ont été enregistrés en MOL2, puis convertis au format PDBQT pour le docking avec Vina à l'aide du script *prepare\_ligand4.py* fourni par MGLTools [Sanner, 1999].

### IV. PROTOCOLES

#### 1. DOCKING AVEC DOCK

Les sphères ont été générées automatiquement à l'aide des outils dms et sphgen fournis avec Dock. Seules celles positionnées dans un rayon de 10 Å autour du ligand staurosporine, qui est le ligand le plus volumineux des trois ligand co-cristallisés, ont été sélectionnées pour le docking. Les grilles ont été créées par le programme grid, en utilisant les paramètres par défaut, sauf le cutoff pour le calcul des interactions à longue distance

(energy cutoff) fixé à 9999, soit le nombre maximal, afin de prendre en compte toutes les interactions, y compris à longue et très longue distance. Plusieurs jeux de paramètres ont été utilisés pour le docking. En particulier, plusieurs valeurs ont été testées pour certains paramètres déterminants de la recherche conformationnelle comme le nombre maximal d'orientations de l'ancre générées (max orientations), ou le nombre maximal d'orientations et de conformations conservées à chaque étape (num anchor orient for growth, number confs for next growth). Les paramètres de la minimisation ont également été explorés.

### 2. DOCKING AVEC VINA

- Docking rigide

Pendant le docking, tous les résidus de la protéine sont gardés fixes tandis que les ligands sont considérés comme flexibles. Les torsions sur les ligands sont attribuées automatiquement lors de la conversion du MOL2 au PDBQT. Le site actif a été défini manuellement de manière à englober la totalité de l'espace couvert par les trois ligands co-cristallisés dans les structures d'origine, plus une marge d'au moins 3 Å dans toutes les directions. Les dimensions de la boîte résultante sont de 18×14×20 Å.

- Docking flexible

Au contraire du docking rigide, certaines chaînes latérales des résidus à proximité du ligand sont flexibles pendant le docking. Ici aussi, les torsions des ligands sont attribuées automatiquement. En ce qui concerne les torsions autorisées sur la protéine, elles ont été choisies en fonction de leur proximité avec le ligand en utilisant une sphère de 6 Å autour du centre de masse du ligand co-cristallisé. Dans ce cas, le site actif doit être suffisamment étendu pour englober le ligand, mais aussi les chaînes latérales des résidus mobiles, quelle que soit leur nouvelle position. Les boîtes de recherche ont été construites automatiquement pour permettre cette flexibilité. Les structures des récepteurs étant différentes, on obtient trois boîtes de calculs de taille très différente, en fonction du ligand co-cristallisé. Pour 1q3d, les résidus flexibles sont I62, V70, L188, C199 et les dimensions de la boîte sont de 11,6×15,6×11,1 Å. Pour 1q3w, des torsions sont incluses sur les résidus I62, V70, Y134, V135, T138, Q185, L188 et C199 et la boîte correspondante fait 15,3×16,1×14,1 Å. Enfin, les acides aminés flexibles pour 1uv5 sont I62, V70, Y134, V135, T138, L188 et C199 et les dimensions de l'espace de recherche sont de 12,5×15,1×14,7 Å. Ces boîtes sont plus petites que celle utilisée pour le docking rigide que nous avons volontairement élargie afin de vérifier la préférence des ligands pour le site actif.

*NB : A posteriori, nous émettons un doute sur cette stratégie qui est biaisée par la différence entre les résidus flexibles pour les trois structures, cependant nous n'avons pas eu l'occasion de reconduire l'expérience de façon plus juste.*

Les paramètres par défaut de Vina ont été utilisés pendant les calculs en rigide et en flexible, soit:

- exhaustiveness = 8
- nombre de poses générées = 9

Les fichiers de sortie de Vina en PDBQT ont été convertis en PDB à l'aide de scripts écrits dans le laboratoire, et les hydrogènes non polaires, non considérés par Vina, ont été ajoutés automatiquement à l'aide d'OpenBabel 2.2.3 [Guha *et al.*, 2006].

### 3. RESCORING

Pour les deux programmes, tous les docking a été menés cinq fois pour chaque ligand, en utilisant cinq seed fixés à l'avance. Le rescoring complet de toutes les poses obtenues a été réalisé avec les fonctions de scoring fournies par Drugscore [Gohlke *et al.*, 2000; Velec *et al.*, 2005] et Xscore [Wang *et al.*, 2002], ce qui donne au total dix fonctions de score distinctes : PAIR\_csd, PAIRSURF\_csd, PAIR\_pdb, PAIRSURF\_pdb et SURF, HPscore, HMscore, HSscore et AVEscore, et enfin Vina.

Les scores calculés par fonction de score pour une pose dans les cinq expériences ont été moyennés pour donner le score final par pose et par fonction. Ensuite, seul le meilleur score obtenu par fonction de score pour un ligand (toutes poses confondues) a été conservé pour créer le classement définitif des ligands.

## D. RÉSULTATS ET DISCUSSION

### I. CONFORMATIONS DES LIGANDS

Les résultats obtenus avec Dock montrent un problème dans la gestion des liaisons flexibles. En effet, les indirubines qui constituent notre test set sont des molécules presque entièrement rigides, où seuls les substituants des deux cycles indoles peuvent présenter des torsions (voir Figure 8). Pourtant, Dock considère qu'il existe une torsion entre les deux cycles et les conformations obtenues montrent un angle de déformation par rapport au plan souvent très important (Figure 10 page suivante). Au contraire, avec Vina, la structure centrale des indirubines est traitée comme rigide, ce qui correspond à la nature chimique de ces molécules.

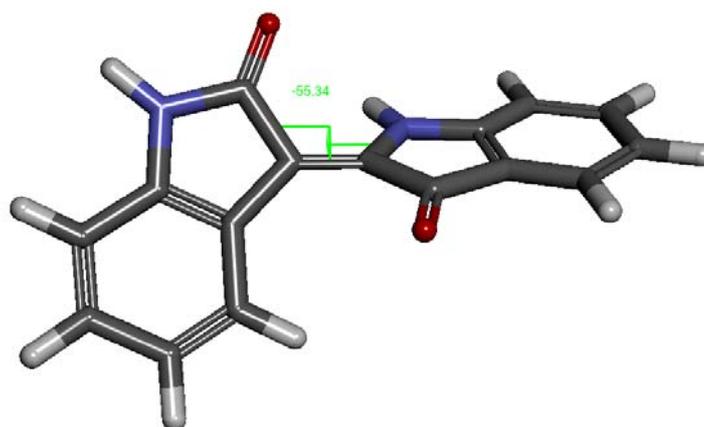


Figure 10. Exemple de pose montrant une indirubine déformée après le docking avec Dock

Une grande différence entre Dock et Vina se situe dans la façon d'attribuer les torsions autorisées. Dans les deux cas, les torsions sont attribuées automatiquement, soit lors de la conversion vers le PDBQT pour Vina, soit directement pendant le docking pour Dock, à l'aide de fichiers de paramètres internes. Pour Vina, ces torsions sont écrites clairement dans le fichier PDBQT (voir l'annexe 1 « Formats des fichiers de molécules »), et peuvent être modifiées directement dans le fichier, ou en passant par l'interface graphique de ADT (Autodock Tools [Sanner, 1999]). Au contraire, dans Dock 5.4, il n'existe pas de moyen direct de contrôler les torsions autorisées (Note : cette option existait dans les versions précédentes de Dock mais a été retirée). Au vu de ces résultats, nous avons donc essayé de comprendre ce qui dans le fichier de paramètres utilisé par Dock pouvait poser problème pour identifier cette liaison comme rigide, mais nous ne sommes pas parvenus à corriger ce défaut. L'exploration des paramètres dictant la recherche conformationnelle et la minimisation n'a pas non plus permis d'éliminer les conformations déformées et a d'ailleurs montré peu d'impact sur les résultats obtenus, excepté au niveau du temps de calcul.

Pour contourner le problème, nous avons mis en place des scripts permettant de filtrer les conformations *a posteriori*, en autorisant un écart au plan de 20°, toutes les conformations présentant un écart supérieur étant rejetées. Cependant, pour des molécules porteuses de substituants encombrants, les poses présentant un fort écart à la planarité sont souvent plus favorables que celles se rapprochant du plan, et malgré cette tolérance certains composés ne présentaient aucune conformation acceptable. De plus, même lorsqu'il en existe, ce type de filtrage post-expérimental met en évidence une quantité de déchet non négligeable, et représente une perte en terme de temps de calcul, puisque de nombreuses

conformations sont calculées pour être ensuite rejetées. Enfin, cette approche est uniquement possible dans les cas où les structures des molécules sont connues et où les angles de torsions à vérifier sont identifiés, ce qui constitue un handicap majeur dans une stratégie de criblage virtuel.

Une approche permettant de s'affranchir de ce défaut de Dock consisterait à réaliser un docking rigide à partir de plusieurs conformations représentatives comme par exemple proposé dans la procédure MS-DOCK [Sauton *et al.*, 2008], cependant, cette méthode reste beaucoup plus lente que le docking avec Vina. Pour toutes ces raisons, Dock ne nous a pas semblé suffisamment fiable pour la suite des expériences et son utilisation a été abandonnée au profit de Vina. Les résultats présentés ci-dessous concernent donc uniquement les solutions proposées par ce programme.

## II. PRÉDICTIONS D’AFFINITÉ

A partir des meilleurs scores obtenus pour chaque ligand, un classement par fonction de score a été réalisé. L'utilisation d'un classement et non des scores permet de s'affranchir des unités qui varient selon les différentes fonctions de score utilisées et les valeurs d'IC50 mesurées. Afin d'évaluer la capacité à reproduire les affinités expérimentales, deux indices ont été calculés qui sont le coefficient de corrélation entre les classements par fonction et le classement par IC50, ainsi que la moyenne des différences de rang entre ces classements. Le premier, qui évalue la relation entre les deux séries de valeurs, doit être le plus élevé possible pour un criblage efficace. Au contraire, le second qui mesure l'écart entre le classement attendu et le classement observé doit être le plus bas possible.

En référence, les mêmes indices ont été calculés à partir de la masse molaire des composés. Pour ce système, la masse molaire des composés et les IC50 sont anti-corrélés, avec des IC50 supérieures en moyenne pour les composés de plus faible masse molaire. Le coefficient de corrélation est faible, de l'ordre de 24% seulement, tandis que la différence moyenne de rang est de 9,21, ce qui semble élevé pour un test set de seulement 34 molécules.

### 1. COMPARAISON ENTRE LES TROIS STRUCTURES CRISTALLOGRAPHIQUES

Dans l'expérience réalisée sur les structures cristallographiques co-cristallisées avec des ligands différents, la structure 1q3d se détache des deux autres, avec quatre fonctions de score pour lesquelles le coefficient de corrélation est supérieur à 60% et seulement une pour laquelle il est inférieur à 40%. Au contraire, aucune fonction ne parvient à fournir une corrélation meilleure que 54% avec la structure 1uv5, et seules trois fonctions présentent un

coefficient de corrélation supérieur à 40%. Les résultats obtenus avec 1q3w sont intermédiaires, avec deux fonctions supérieures à 60%, mais toutes les autres inférieures à 40% (Table 4). La répartition obtenue par le calcul des différences moyennes de rang est identique (Table 4).

Ces résultats peuvent paraître surprenant, puisque la structure 1uv5 est celle qui contient le ligand co-cristallisé de type indirubine, tandis que les autres structures correspondent à des ligands de nature différente. La base de ligands étant constituée exclusivement d'indirubines, cette structure devrait, en principe, être le mieux adaptée à l'étude de ce système, ce qui n'est pas le cas. La superposition des trois modèles (voir Figure 1) montre un déplacement du squelette de la protéine assez important au niveau de la boucle riche en glycine (glycin-rich loop) chez 1q3d et 1q3w, qui élargit le site actif et permet à la totalité des ligands de se placer dans la cavité. Le site actif de 1uv5 est en quelque sorte trop adapté à son ligand pour être capable d'accommoder d'autres molécules, y compris structurellement proches.

| Fonction | 1q3d | 1q3w | 1uv5 |
|----------|------|------|------|
| Vina     | 0.41 | 0.38 | 0.38 |
| PS csd   | 0.68 | 0.63 | 0.54 |
| PS pdb   | 0.67 | 0.36 | 0.27 |
| P csd    | 0.71 | 0.64 | 0.5  |
| P pdb    | 0.68 | 0.37 | 0.27 |
| Surf     | 0.31 | 0.3  | 0.46 |
| HP       | 0.52 | 0.2  | 0.36 |
| HM       | 0.49 | 0.24 | 0.25 |
| HS       | 0.46 | 0.32 | 0.36 |

| Fonction | 1q3d | 1q3w | 1uv5 |
|----------|------|------|------|
| Vina     | 8.29 | 8.41 | 8.76 |
| PS csd   | 6.24 | 6.29 | 7.24 |
| PS pdb   | 6.35 | 9.24 | 9.18 |
| P csd    | 5.82 | 6.12 | 7.65 |
| P pdb    | 6.29 | 9.41 | 9.24 |
| Surf     | 9.53 | 9.06 | 7.82 |
| HP       | 7    | 10   | 8.94 |
| HM       | 7.65 | 9.65 | 9.76 |
| HS       | 7.59 | 8.76 | 8.88 |

Table 4. Corrélation (droite) et différences moyennes de rang (gauche) sur les trois modèles rigides pour toutes les fonctions de score utilisées. Les meilleurs résultats sont figurés en jaune et les moins bons en bleu.

On remarque également que les fonctions de Drugscore, en dehors de SURF semblent les plus efficaces sur ce système, puisqu'elles donnent systématiquement les meilleures corrélations. En particulier, avec les fonctions PAIR\_csd et PAIRSURF\_csd la corrélation est toujours supérieure à 50%, pour les trois structures.

## 2. COMPARAISON ENTRE LES STRUCTURES RELAXÉES

- *Modèle 1q3d*

La comparaison des sept modèles relaxés de 1q3d montre une fracture entre les modèles FIX à T25, et les modèles T10, T0 et T0LNF. Les premiers modèles donnent des

résultats comparables, avec des coefficients de corrélation situés entre 65 et 75% environ, tandis que les seconds sont beaucoup moins efficaces, avec un grand nombre de fonctions de scores dont le coefficient de corrélation diminue brutalement. La même distribution est observée sur les différences moyennes de rang. Il semble donc qu'il existe un seuil, à partir duquel la minimisation déforme la protéine de façon trop importante, empêchant les ligands de se lier à la protéine et compromettant la qualité des résultats obtenus (Table 5).

Comme pour l'expérience précédente, les fonctions PAIR\_csd et PAIRSURF\_csd donnent les meilleurs résultats, confirmant leur efficacité sur ce système. La fonction SURF quant à elle donne des résultats assez mauvais, avec une corrélation du même ordre que celle obtenue en utilisant la masse molaire des composés, voire plus faible.

| Fonction | fix  | t100 | t50  | t25  | t10  | t0   | t0Inf |
|----------|------|------|------|------|------|------|-------|
| Vina     | 0.41 | 0.43 | 0.44 | 0.43 | 0.35 | 0.33 | 0.22  |
| PS csd   | 0.68 | 0.69 | 0.71 | 0.72 | 0.48 | 0.58 | 0.62  |
| PS pdb   | 0.67 | 0.65 | 0.65 | 0.59 | 0.48 | 0.19 | 0.16  |
| P csd    | 0.71 | 0.71 | 0.73 | 0.77 | 0.5  | 0.6  | 0.61  |
| P pdb    | 0.68 | 0.65 | 0.67 | 0.64 | 0.46 | 0.22 | 0.14  |
| Surf     | 0.31 | 0.16 | 0.12 | 0.16 | 0.26 | 0.16 | 0.31  |
| HP       | 0.52 | 0.57 | 0.54 | 0.53 | 0.57 | 0.16 | 0.19  |
| HM       | 0.49 | 0.47 | 0.5  | 0.52 | 0.45 | 0.21 | 0.21  |
| HS       | 0.46 | 0.41 | 0.4  | 0.45 | 0.46 | 0.21 | 0.23  |

| Fonction | fix  | t100  | t50   | t25   | t10  | t0    | t0Inf |
|----------|------|-------|-------|-------|------|-------|-------|
| Vina     | 8.29 | 8.12  | 8.06  | 8.12  | 8.76 | 9     | 10    |
| PS csd   | 6.24 | 5.94  | 5.94  | 6.06  | 8    | 7     | 6.82  |
| PS pdb   | 6.35 | 6.82  | 6.94  | 7.53  | 8.06 | 10.29 | 10.65 |
| P csd    | 5.82 | 5.71  | 5.71  | 5.18  | 7.88 | 7.18  | 7     |
| P pdb    | 6.29 | 6.65  | 6.65  | 7.18  | 8.29 | 10.18 | 10.71 |
| Surf     | 9.53 | 10.65 | 10.65 | 10.18 | 9.59 | 10    | 9.06  |
| HP       | 7    | 6.88  | 7.29  | 7.35  | 6.82 | 9.71  | 9.35  |
| HM       | 7.65 | 7.88  | 7.94  | 7.35  | 8    | 9.35  | 9.76  |
| HS       | 7.59 | 8.18  | 8.06  | 7.59  | 7.65 | 9.65  | 9.41  |

Table 5. Corrélation (haut) et différences moyennes de rang (bas) sur les modèles relaxés de 1q3d pour toutes les fonctions de score utilisées. Les meilleurs résultats sont figurés en jaune et les moins bons en bleu.

- *Modèle 1uv5*

Dans la première expérience, le modèle 1uv5 était beaucoup moins performant que les deux autres, et notamment que 1q3d. On remarque ici que l'utilisation de plusieurs structures relaxées ne permet pas d'améliorer les résultats obtenus à partir de ce modèle. Aucune des fonctions de scoring utilisées ne donne de coefficient de corrélation supérieur à 60% et

seules les fonctions PAIR\_csd et PAIRSURF\_csd identifiées comme les plus efficaces sur ce système parviennent à atteindre les 50% (Table 6).

| Fonction | fix  | t100 | t50  | t25  | t10  | t0   | t0Inf |
|----------|------|------|------|------|------|------|-------|
| Vina     | 0.38 | 0.39 | 0.41 | 0.41 | 0.4  | 0.39 | 0.26  |
| PS csd   | 0.54 | 0.53 | 0.52 | 0.53 | 0.54 | 0.39 | 0.48  |
| PS pdb   | 0.27 | 0.28 | 0.28 | 0.27 | 0.23 | 0.28 | 0.19  |
| P csd    | 0.5  | 0.5  | 0.52 | 0.53 | 0.52 | 0.43 | 0.48  |
| P pdb    | 0.27 | 0.29 | 0.3  | 0.31 | 0.27 | 0.27 | 0.16  |
| Surf     | 0.46 | 0.19 | 0.17 | 0.18 | 0.02 | 0.15 | 0.31  |
| HP       | 0.36 | 0.35 | 0.39 | 0.41 | 0.39 | 0.36 | 0.27  |
| HM       | 0.25 | 0.28 | 0.33 | 0.32 | 0.31 | 0.39 | 0.3   |
| HS       | 0.36 | 0.39 | 0.41 | 0.41 | 0.36 | 0.37 | 0.29  |

| Fonction | fix  | t100  | t50  | t25  | t10   | t0    | t0Inf |
|----------|------|-------|------|------|-------|-------|-------|
| Vina     | 8.76 | 8.53  | 8.53 | 8.35 | 8.41  | 8.29  | 9.94  |
| PS csd   | 7.24 | 7.35  | 7.53 | 7.53 | 7.06  | 8.71  | 8.29  |
| PS pdb   | 9.18 | 9     | 8.94 | 9.12 | 9.35  | 9.59  | 10.18 |
| P csd    | 7.65 | 7.47  | 7.47 | 7.29 | 7.24  | 8.41  | 8.18  |
| P pdb    | 9.24 | 9     | 8.88 | 8.82 | 9.12  | 9.53  | 10.18 |
| Surf     | 7.82 | 10.12 | 9.94 | 9.88 | 11.12 | 10.06 | 8.65  |
| HP       | 8.94 | 9.06  | 8.88 | 8.59 | 8.76  | 8.82  | 9.65  |
| HM       | 9.76 | 9.82  | 9.29 | 9.35 | 9.53  | 8.41  | 9.47  |
| HS       | 8.88 | 8.88  | 8.53 | 8.53 | 8.76  | 8.47  | 9.41  |

Table 6. Corrélation (haut) et différences moyennes de rang (bas) sur les modèles relaxés de 1uv5 pour toutes les fonctions de score utilisées. Les moins bons résultats sont figurés en bleu.

### 3. DOCKING FLEXIBLE

Pour les modèles 1q3d et 1q3w, les résultats du docking flexible montrent une perte d'efficacité en comparaison du docking rigide. Les meilleurs coefficients de corrélation sont toujours obtenus avec PAIR\_csd et PAIRSURF\_csd mais ils passent de 71% et 68% à 62% et 61% respectivement pour 1q3d et de 64% et 63% à 52% et 57% pour 1q3w. Au contraire, pour 1uv5, le docking flexible induit une amélioration des corrélations, jusqu'à atteindre des coefficients de 61% et 60% pour ces deux fonctions. La flexibilité des chaînes latérales autorisée par Vina permet apparemment une plus grande liberté dans les poses des ligands au sein du site actif de 1uv5, et compense le volume plus réduit de la cavité en comparaison des deux autres structures. La différence entre les résultats obtenus pour les trois structures est gommée par l'introduction de degrés de libertés sur les résidus du site actif. Toutefois, ces résultats restent moins bons que ceux obtenus lors du docking rigide des modèles, relaxés ou non, construits à partir de la structure de 1q3d (Table 7).

| Fonction | 1q3d | 1q3w | 1uv5 |
|----------|------|------|------|
| Vina     | 0.44 | 0.59 | 0.43 |
| PS csd   | 0.61 | 0.57 | 0.6  |
| PS pdb   | 0.49 | 0.4  | 0.32 |
| P csd    | 0.62 | 0.52 | 0.61 |
| P pdb    | 0.5  | 0.41 | 0.32 |
| Surf     | 0.06 | 0.25 | 0.22 |
| HP       | 0.48 | 0.43 | 0.33 |
| HM       | 0.31 | 0.31 | 0.34 |
| HS       | 0.29 | 0.37 | 0.42 |

| Fonction | 1q3d  | 1q3w | 1uv5 |
|----------|-------|------|------|
| Vina     | 8     | 6.88 | 8.24 |
| PS csd   | 6.82  | 7.24 | 6.71 |
| PS pdb   | 8.18  | 8.71 | 8.53 |
| P csd    | 6.82  | 8    | 6.71 |
| P pdb    | 8.18  | 8.65 | 8.71 |
| Surf     | 11.18 | 9.35 | 9.71 |
| HP       | 7.76  | 7.71 | 8.88 |
| HM       | 9.06  | 9.18 | 9.12 |
| HS       | 9     | 8.76 | 8.47 |

Table 7. Corrélation (droite) et différences moyennes de rang (gauche) sur les trois modèles flexibles pour toutes les fonctions de score utilisées. Les meilleurs résultats sont figurés en jaune et les moins bons en bleu.

## E. CONCLUSION

Les résultats obtenus lors de cette étude montrent une grande dépendance envers la structure de la protéine utilisée pendant le criblage. Dans le cas présent, une tendance naturelle aurait sans doute été de considérer que le meilleur modèle, pour le criblage d'une base de molécules construites à partir d'un motif structural commun, était celui qui justement avait été obtenu en co-cristallisation avec un membre de cette famille. Notre étude montre que cette intuition peut se révéler fautive, avec des conséquences très importantes sur les résultats de criblage. Au contraire, une structure obtenue avec un ligand beaucoup plus volumineux et un site actif plus large permet d'accueillir un plus grand nombre de ligands et d'obtenir des corrélations bien meilleures. L'introduction de la flexibilité, soit à travers un ensemble de structures soit directement au cours du docking, semble moins critique, au moins pour ce système, que le choix d'une structure protéique adéquate. Ces résultats peuvent être rapprochés de l'étude publiée par Rockey et al. dans laquelle un criblage virtuel sur des kinases avait montré un taux de succès plus fort pour les structures reconstruites par homologie que pour les structures cristallographiques [Rockey & Elcock, 2006]. Dans leur cas, les modèles obtenus par homologie avaient été basés sur des structures de complexes comprenant un ligand de la même famille que ceux testés. Le succès de cette technique avait été relié à cette similarité entre les ligands, mais il est possible que la dé-spécialisation du site actif induite par la reconstruction soit également un facteur non négligeable dans la réussite de l'expérience.

De plus, quelle que soit l'approche utilisée pour simuler la flexibilité de la protéine, cela

représente un temps de calcul bien plus important que l'utilisation d'un récepteur rigide. Ici, ce temps de calcul supplémentaire équivaut à une perte de temps, puisque les résultats obtenus dans les approches flexibles ne montrent pas, ou très peu, d'amélioration en comparaison du modèle rigide.

En complément, cette étude nous a également permis de comparer les outils à notre disposition. En ce qui concerne le docking, les résultats obtenus avec Dock nous ont paru incompatibles avec le criblage de ligands inconnus, à cause de défauts dans la gestion des liaisons flexibles. Au contraire, le programme Vina a fourni des résultats très satisfaisants. La convergence des meilleures poses sur cinq expériences est totale, et les résultats obtenus sont indépendants de la conformation d'origine, ce qui n'est pas forcément vrai pour tous les programmes de docking [Yuriev *et al.*, 2011]. Les bons résultats obtenus avec ce programme sont confirmés par le succès de Vina, malgré sa publication très récente. En effet, en moins de deux ans Vina a déjà fait l'objet d'une vingtaine de publications [Bieganski & Yarmush, 2011; Błażewska *et al.*, 2011; Danielson *et al.*, 2011; Kukol, 2011; Maldonado-Rojas & Olivero-Verbel, 2011; Musa *et al.*, 2011; Palermo & Natarajan, 2011; Peterson *et al.*, 2011; Seclaman *et al.*, 2011; Segura-Cabrera *et al.*, ; Shen *et al.*, 2011; Turabekova *et al.*, 2010; Wiggers *et al.*, 2011] qui comprennent une comparaison avec le programme Autodock dont il dérive qui plaide en faveur de l'utilisation de Vina [Chang *et al.*, 2010], et trois publications proposant l'intégration de la méthode de docking de Vina dans une interface graphique [Lill & Danielson, 2011; Sandeep *et al.*, 2011; Seeliger & de Groot, 2010]. En conséquence, Vina semble adapté à l'étude de bases de ligands inconnus, et a été utilisé pour les applications présentées dans la suite de ce travail.



**CHAPITRE 3 : APPLICATION À LA PHOSPHOLIPASE  
HUMAINE SPLA2 DE GROUPE X**



En collaboration avec Gilles Guichard et Claire Venin, de l'équipe Peptidomimetics chemistry, IECB/CNRS UMR 5248.

## A. GÉNÉRALITÉS

Les phospholipases sont des protéines spécialisées dans la catalyse de l'hydrolyse des phospholipides libérant des acides gras ainsi que d'autres substances lipophiles dépendantes de la classe de phospholipase concernée. Elles sont divisées en quatre groupes selon le type de clivage caractéristique de chacune (Figure 11). Les phospholipases A1 et A2 (PLA1 et PLA2) sont capables d'hydrolyser spécifiquement l'ester en position sn-1 et sn-2 respectivement, tandis que les phospholipases B (PLB) hydrolysent indifféremment sur ces deux mêmes positions (sn-1 et sn-2). Les phospholipases C (PLC) catalysent la coupure juste avant le phosphate, et les phospholipases D (PLD) réalisent le clivage juste après le phosphate<sup>2</sup>.

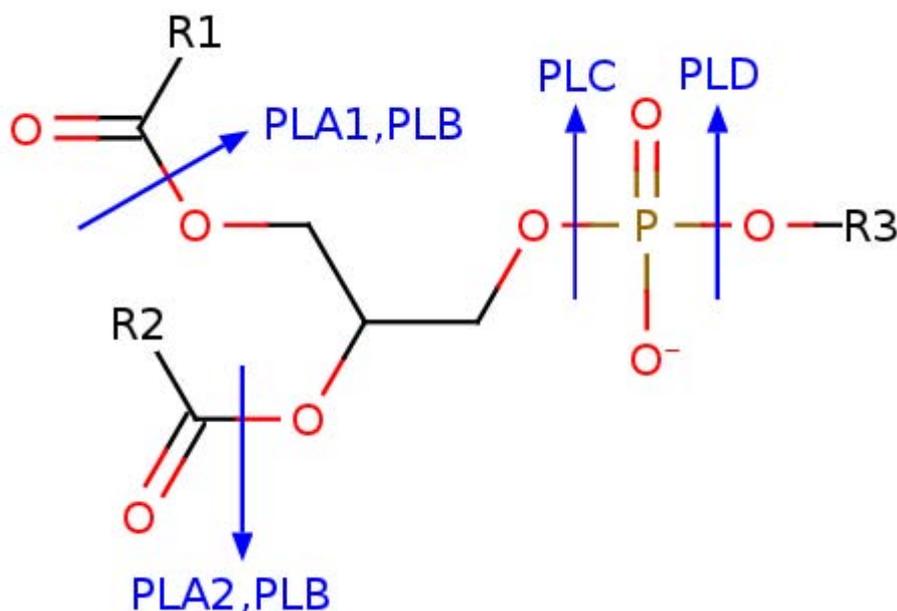


Figure 11. Sites de clivage des différentes phospholipases

Bien que catalysant la même réaction, les PLA2 forment une classe de protéines

<sup>2</sup> De façon classique, on utilise la notation « sn » (stereochemical numbering) pour les phospholipides. Cette nomenclature prévoit que la chaîne numérotée sn-2 qui correspond au groupe hydroxyle du second carbone du glycérol soit située à gauche sur une projection de Fischer, tandis que les groupes sn-1 et sn-3 se placent respectivement en haut et en bas.

diverses, que ce soit en terme de fonction, de localisation, de structure et de mécanisme, en particulier en ce qui concerne la dépendance fonctionnelle de certaines phospholipases à des ions métalliques. Ceci rend donc leur classification relativement complexe, mais la littérature les classe généralement en trois grandes familles: les PLA2 intracellulaires ou cytosoliques (cPLA2) et les PLA2 extracellulaires ou sécrétées (sPLA2) dépendantes du calcium d'un côté, et les PLA2 indépendantes du calcium (iPLA2) de l'autre [Dennis, 1994; Laye & Gill, 2003].

Les sPLA2 (secreted PLA2), qui regroupent donc les formes extracellulaires de PLA2, sont des enzymes dites interfaciales, c'est à dire qu'elle interagissent avec les phospholipides lorsque ceux-ci sont organisés en micelles ou en bicouche (membranes). En effet, elles sont solubles dans l'eau, alors que les phospholipides qui leur servent de substrat sont très fortement hydrophobes, ce qui leur impose une fixation sur l'interface membranaire préalable afin d'accéder au substrat phospholipidique [Lambeau & Gelb, 2008]. En conséquence, leur mécanisme d'action comporte deux étapes, la première étant l'adsorption de l'enzyme sur l'interface lipidique, et la seconde la catalyse du phospholipide qui ne peut avoir lieu qu'après la première étape (Figure 12) [Berg *et al.*, 2001; Winget *et al.*, 2006].

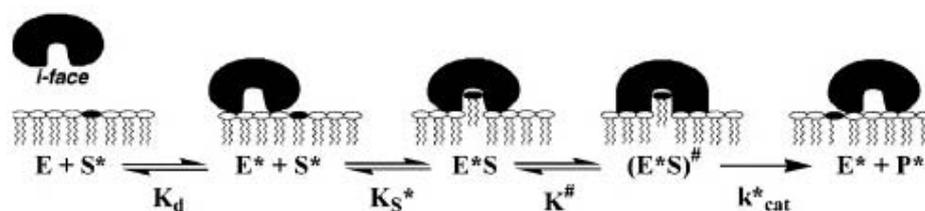


Figure 12. Mécanisme d'action interfaciale

Le site de reconnaissance interfaciale est différent du site catalytique et il existe donc un double phénomène de spécificité en rapport avec la nature de la membrane (spécificité interfaciale) et en rapport avec le phospholipide hydrolysé (spécificité du site catalytique) qui définit la spécificité globale de l'enzyme envers son substrat [Singer *et al.*, 2002].

Dix sPLA2 différentes ont été identifiées chez la souris : les sPLA2 des groupes IB, IIA, IIC, IID, IIE, IIF, III, V, X et XII. Neuf sont également présentes chez l'homme, la dixième (sPLA2-IIC) étant retrouvée uniquement sous forme de pseudogène [Lambeau & Gelb, 2008; Singer *et al.*, 2002]. Parmi les dix formes, les groupes I, II, V et X sont plus particulièrement proches. Il s'agit de protéines de faible poids moléculaire (entre 14 et 19 kDa) dont la structure est caractérisée par la présence de nombreux ponts disulfures dont au moins six absolument conservés. Elles possèdent également un site actif fortement conservé

au sein duquel on retrouve une histidine qui forme une dyade catalytique avec un aspartate situé à proximité, ainsi qu'une boucle de liaison chargée de retenir un atome de calcium [Murakami & Kudo, 2004]. En effet, comme indiqué précédemment, ces protéines requièrent la présence de  $\text{Ca}^{2+}$  dans le milieu pour catalyser la réaction d'hydrolyse [Burke & Dennis, 2009; Dennis, 1994; Schaloske & Dennis, 2006].

Toutefois l'homologie de séquence entre les formes reste assez faible, entre 15 et 50% d'homologie comparées deux par deux. Ajouté au fait que les distributions tissulaires des isoformes suivent des schémas bien distincts, ceci semble indiquer des fonctions physiologiques différentes [Mitsubishi *et al.*, 2006; Singer *et al.*, 2002]. Cette hypothèse est renforcée par les variations au niveau des spécificités interfaciales et liées au site actif des différentes formes de sPLA2.

Ainsi, alors que l'ensemble des sPLA2 possède une affinité très forte pour les vésicules de phospholipides anioniques, les sPLA2 des groupes V et X sont les seules sPLA2 à posséder également une affinité élevée pour les vésicules de type zwitterionique [Lambeau & Gelb, 2008] comme les phosphatidylcholines (PC). Ces phospholipides sont particulièrement présents dans la partie extracellulaire des membranes des cellules chez les mammifères [Singer *et al.*, 2002]. Dans le cas de hGX (sPLA2 humaine de Groupe X), on trouve même une affinité pour les vésicules de POPC (1-palmitoyl-2-oleoylphosphatidylcholine) supérieure à celle pour les vésicules de POPG (1-palmitoyl-2-oleoylphosphatidylglycérol), ainsi qu'une préférence catalytique pour les phosphatidylcholines. On observe que c'est aussi le groupe pour lequel la préférence pour les chaînes d'acides gras sn-2 poly-insaturées de type arachidonyle est la plus marquée (affinité à peu près doublée), bien que cette préférence soit tout de même assez faible, comme chez les autres sPLA2 qui la présentent. A l'inverse, hGV (sPLA2 humaine de Groupe V) possède une préférence pour les chaînes saturées (affinité triplée) [Singer *et al.*, 2002].

## **B. RÔLES PHYSIOLOGIQUES ET PATHOLOGIQUES**

Les différentes sPLA2 ont été identifiées dans de nombreux processus physiologiques et pathologiques chez les mammifères, mais leur mode d'action n'est pas toujours très clair et leurs interventions dans un certain nombre de ces processus manquent encore de confirmation *in vivo* [Lambeau & Gelb, 2008]. Il est même probable que toutes leurs fonctions n'aient pas encore été élucidées. Les paragraphes suivants ne sont donc pas une liste exhaustive des tous les processus pathologiques liés aux sPLA2 mais des extraits de la

littérature concernant les fonctions les plus souvent documentées de ces protéines permettant de mettre en évidence plusieurs grands rôles pathologiques des différentes formes de sPLA2 et de faire apparaître leurs divergences.

Les produits de l'hydrolyse de l'ester en position sn-2 réalisée par les sPLA2 sont d'une part des lysophospholipides, et d'autre part des acides gras, deux familles que l'on peut qualifier de médiateurs lipidiques. En effet, ce sont d'importants messagers secondaires dans la communication cellulaire et la transduction du signal [Curfs *et al.*, 2008; Karabina *et al.*, 2006]. En particulier, la libération d'acide arachidonique, un acide gras essentiel impliqué dans de très nombreux processus physiologiques et pathologiques, est principalement contrôlée par les sPLA2 [Balsinde *et al.*, 1999]. Celui-ci est ensuite de nouveau métabolisé par le biais des cyclo-oxygénases (COX-1 ou COX-2), des lipoxygénases (LOX) ou des cytochromes P450 (CYP450) et se trouve être le principal précurseur des eicosanoïdes [Laye & Gill, 2003], des molécules impliquées dans l'inflammation et les réponses immunitaires par exemple ainsi que la signalisation au niveau du système nerveux central (Figure 13) [Schaloske & Dennis, 2006].

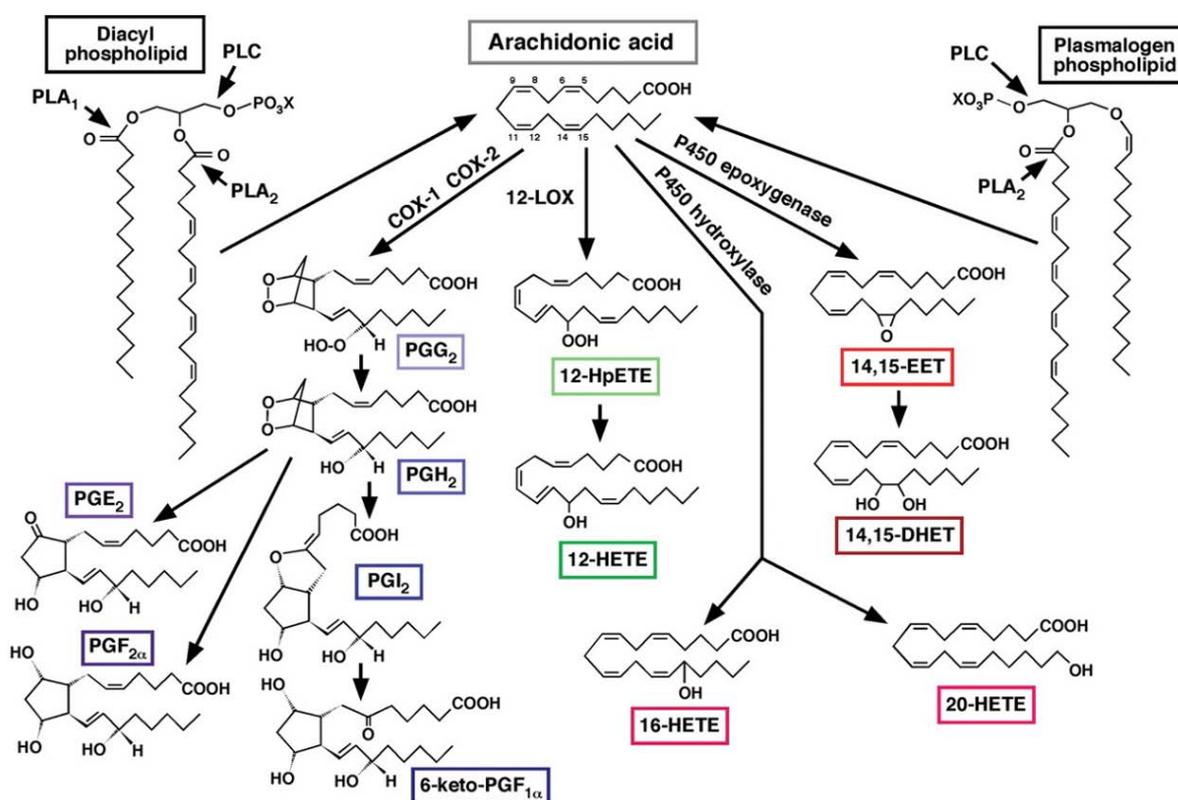


Figure 13. Voies de signalisations découlant de l'acide arachidonique produit par les sPLA2 [Schaloske & Dennis, 2006]

Il est à noter que les sPLA2 sont également présentes dans les venins de serpents ou

d'abeilles, cependant les fonctions décrites ci-dessous ne concernent que les observations faites chez les mammifères, homme ou souris le plus souvent.

## I. DÉFENSE CONTRE LES PATHOGÈNES

Le rôle des sPLA2 dans la réponse immunitaire innée est l'une de leurs fonctions les plus documentées, notamment à travers les études concernant sPLA2-IIA réalisées dès les années 1980 [Lambeau & Gelb, 2008; Murakami & Kudo, 2004]. Mais sPLA2-IIA, surtout connue pour ses propriétés antibactériennes, et que l'on retrouve par exemple en quantité importante dans les larmes, n'est pas la seule à jouer un rôle dans les défenses contre divers agents pathogènes et plusieurs sPLA2 sont capables de dégrader les membranes bactériennes. *In vitro*, les différentes sPLA2 peuvent être classées dans l'ordre suivant selon leur capacité à tuer les bactéries à Gram positif : IIA > X > V > XII > IIE > IB, IIF [Koduri *et al.*, 2002].

Les groupes X et V ont également montré une capacité à empêcher l'infection par des adénovirus, responsables en particulier de maladies respiratoires, en agissant sur la membrane plasmique des cellules hôtes. Ces deux formes étant exprimées au niveau du système respiratoire, des études montrent qu'elles jouent certainement un rôle dans la défense de celui-ci contre les adénovirus [Mitsuishi *et al.*, 2006]. La sPLA2 de groupe X a également montré une activité antivirale contre plusieurs virus enveloppés de type lentivirus tels que HIV-1, par dégradation de la membrane virale, particulièrement riche en phospholipides zwitterioniques [Kim *et al.*, 2007].

## II. HYDROLYSE DES PHOSPHOLIPIDES ALIMENTAIRES

La sPLA2-IB, aussi appelée sPLA2 pancréatique, est présente en grande quantité dans les sucs pancréatiques et libérée en réponse à l'ingestion de nourriture dans le tractus gastro-intestinal où elle a pour rôle l'hydrolyse des phospholipides issus de l'alimentation. Des études sur des souris déficientes en sPLA2-IB ont révélé que dans des conditions de régime pauvre en lipides, celles-ci montraient un développement normal et un métabolisme lipidique ainsi que des fonctions reproductives similaires à des souris de référence. En revanche, les mêmes expériences en conditions de régime riche en lipides ont permis de constater que les souris déficientes en sPLA2 étaient moins sujettes à l'obésité ainsi qu'à la résistance à l'insuline liée à l'obésité [Huggins *et al.*, 2002]. Les sPLA2 semblent donc constituer un facteur de risque dans les maladies liées au surpoids.

D'autres études ont montré que les formes IIA, IID, IIE, IIF, V et X sont également exprimées à hauteur variable dans le tractus digestif, mais leur implication dans la digestion

des phospholipides alimentaires n'a pas été explorée plus avant [Eerola *et al.*, 2006].

### III. INFLAMMATION

Grâce à leur capacité à se lier à des membranes riches en phosphatidylcholine (PC), les groupes X et V sont les sPLA2 qui présentent la plus forte production d'acide arachidonique. En règle générale, les divers médiateurs lipidiques produits notamment par le biais de la libération d'acide arachidonique sont des activateurs de voies de signalisation de l'inflammation [Natarajan & Nadler, 2004]. De plus, une augmentation de la quantité extracellulaire de sPLA2 dans les fluides inflammatoires et le plasma a été observé dans plusieurs maladies inflammatoires ainsi qu'au niveau de tissus enflammés [Karabina *et al.*, 2006; Murakami & Kudo, 2004] suggérant un rôle de ces protéines dans toutes sortes de mécanismes liés à l'inflammation.

En particulier, certains eicosanoïdes tels que les leucotriènes et les prostaglandines sont impliqués dans la pathogenèse de l'asthme et des modèles de souris déficientes en sPLA2-X laissent supposer que son inhibition pourrait réduire considérablement les inflammations des voies respiratoires [Henderson *et al.*, 2007]. Les niveaux de LPC (lysophosphatidylcholine) résultant de l'hydrolyse du surfactant pulmonaire et l'activité des sPLA2 dans les liquides de lavage broncho-alvéolaires de patients atteints d'ARDS (Acute respiratory distress syndrome ou syndrome de détresse respiratoire aiguë) ou de formes sévères d'asthme sont en corrélation avec la gravité de la maladie [Chilton *et al.*, 1996; Touqui & Arbibe, 1999]. L'augmentation des taux de LPC est globalement un facteur aggravant de ces maladies [Murakami & Kudo, 2004].

Bien que plusieurs modèles animaux expérimentaux d'ARDS aient jusque là permis de mettre en avant l'activité de sPLA2-IIA dans la dégradation du surfactant pulmonaire [Wu *et al.*, 2003], une étude *in vitro* a démontré que les sPLA2 V et X en sont également capables [Chabot *et al.*, 2003]. Une sur-expression des sPLA2-X dans les cellules épithéliales des voies aériennes et dans les macrophages bronchiques des patients atteints d'asthme a par ailleurs été mise en évidence [Hallstrand *et al.*, 2007].

### IV. ATHÉROSCLÉROSE

Trois membres au moins de la famille des sPLA2, sPLA2-IIA, V et X, ont été détectés dans des plaques d'athérome et semblent avoir un lien avec le phénomène d'athérosclérose [Lambeau & Gelb, 2008; Murakami, 2004], qui correspond à une accumulation de lipides et d'autres produits dans les artères pouvant aboutir à l'obstruction de celles-ci. Dans ce processus, les LDL (low density lipoproteins ou lipoprotéines de basse densité) chargées du

transport des lipides agissent comme promoteur de l'athérosclérose en étant retenues sur la paroi artérielle et phagocytées par les macrophages. Or l'hydrolyse des LDL par les sPLA2 citées conduit à la libération de LDL plus petites et d'affinité supérieure pour les protéoglycanes contenus dans la paroi artérielle [Murakami & Kudo, 2004; Webb, 2005], ce qui augmente leur potentiel pro-athérogénique [Karabina *et al.*, 2006]. Ces LDL provoquent également une augmentation de la formation de cellules spumeuses (macrophages chargés en lipides) [Webb, 2005], et selon Murakami et al. [Murakami & Kudo, 2004], les LDL modifiées par sPLA2-X sont efficacement intégrées aux macrophages, accélérant la formation de plaques d'athérome. L'activité d'hydrolyse des PC des lipoprotéines chez sPLA2-X et sPLA2-V étant plus importante que chez sPLA2-IIA [Murakami & Kudo, 2004], ces deux formes sont plus susceptibles de libérer une grande quantité de LDL fortement pro-athérogéniques.

S'agissant d'une maladie inflammatoire, les divers médiateurs lipidiques pro-inflammatoires produits par les sPLA2-IIA, V et X sont également des facteurs d'aggravations de l'athérosclérose [Natarajan & Nadler, 2004]. En adéquation avec ces résultats, des études ont montré que des taux plasmatiques élevés d'activité sPLA2 constituaient un facteur de risque pour les cardiopathies coronariennes [Packard *et al.*, 2000; Rader, 2000].

## V. TUMOROGENÈSE

Différentes formes de PLA2 intracellulaires et extracellulaires semblent impliquées dans la tumorigenèse, principalement par le biais des médiateurs lipidiques qu'elles produisent. En particulier, la métabolisation des lysophospholipides en acide lysophosphatidique contribue à la croissance cellulaire et la plupart des métabolites de l'acide arachidonique sont capables d'induire la croissance tumorale et la prolifération cellulaire *in vitro* [Cummings, 2007]. Plusieurs études font d'ailleurs état de doses élevées d'eicosanoïdes dans des tumeurs [Laye & Gill, 2003; Marks *et al.*, 2000; Tapiero *et al.*, 2002]. La Figure 14 [Cummings, 2007] montre le rôle des PLA2 dans la production de médiateurs lipidiques impliqués dans la tumorigenèse.

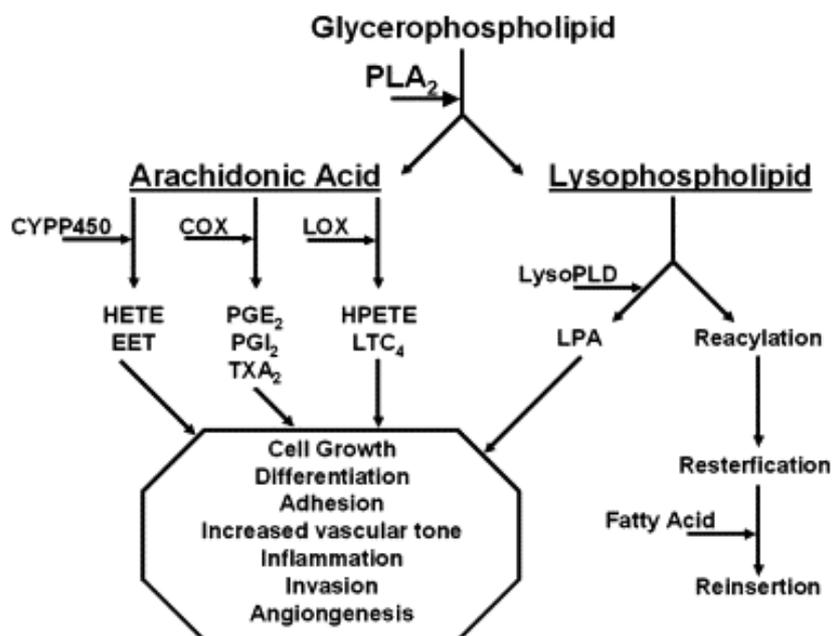


Figure 14. Rôle des espèces produites par les sPLA2 dans la tumorigénèse [Cummings, 2007]

La littérature rapporte que les taux de plusieurs PLA2 sont plus élevés que la normale dans certains cancers, cependant, le groupe ou même la famille de PLA2 concernés ne sont pas toujours précisés [Cummings, 2007; Denizot *et al.*, 2005; Dong *et al.*, 2006; Graff *et al.*, 2001; Jiang *et al.*, 2002; Laye & Gill, 2003; Yamashita *et al.*, 1994]. En ce qui concerne les sPLA2, une augmentation de l'expression et de l'activité, notamment de sPLA2-IIA, a été observée dans les cancers du sein [Yamashita *et al.*, 1994], du pancréas [Kiyohara *et al.*, 1993], de la prostate [Dong *et al.*, 2006; Jiang *et al.*, 2002; Kallajoki *et al.*, 1998; Menschikowski *et al.*, 2008], du foie [Ying *et al.*, 1994], de la peau [Yamashita *et al.*, 1994] et colorectal [Denizot *et al.*, 2003; Morioka *et al.*, 2000]. Dans le cancer de la prostate, une corrélation négative a même été établie entre l'expression de sPLA2-IIA et la survie des patients [Graff *et al.*, 2001]. Dans les cas de cancer du colon, la sPLA2 de groupe X ainsi que la voie de métabolisation de l'acide arachidonique COX-2 sont très fortement exprimées et permettent la production de la prostaglandine E2 connue pour son activité oncogénique [Morioka *et al.*, 2000].

Toutefois, si la présence de sPLA2 dans divers cancers est fréquemment observée, les fonctions exercées par ces protéines dans les mécanismes de tumorigénèse restent peu documentées ou controversées [Lambeau & Gelb, 2008].

## **VI. INTÉRÊT THÉRAPEUTIQUE**

Comme indiqué plus haut, la famille des sPLA2 est impliquée dans un grand nombre de processus physiologiques et surtout pathologiques chez les mammifères, fonctionnant de façon pro ou anti-maladie selon les cas. Outre les processus décrits en détail ci-dessus, d'autres rôles encore peu documentés ou controversés dans la littérature sont également à noter. Par exemple, les sPLA2 sont retrouvées dans toutes les régions du cerveau, et la forme du groupe X, présente dans les neurones du système nerveux périphérique, est capable de provoquer la neurogenèse à travers la production de LPC [Farooqui *et al.*, 2006; Masuda *et al.*, 2005], indiquant un rôle possible dans la maladie d'Alzheimer [Schaeffer *et al.*, 2010]. On observe également les sPLA2 dans les organes génitaux mâles de plusieurs espèces animales [Masuda *et al.*, 2004], et une équipe a démontré l'influence de la sPLA2-X sur la motilité spermatique [Escoffier *et al.*, 2010; Escoffier *et al.*, 2011]. Enfin, l'observation de ces protéines dans les venins d'abeilles ou de serpents, et d'une activité anticoagulante *in vitro* pour les sPLA2 issues de venins ou de mammifères, laisse supposer qu'elles jouent également un rôle dans la coagulation *in vivo* [Lambeau & Gelb, 2008; Mounier *et al.*, 2001].

Le rôle majeur de ces protéines, commun à toutes les pathologies qui leur sont associées, est la production d'acide arachidonique pour le métabolisme des eicosanoïdes. Or la capacité de chaque sPLA2 à produire celui-ci est très variable. La comparaison des activités exercées par les différentes formes de sPLA2 montre une grande dépendance au type de vésicule rencontré. On observe également une corrélation très forte entre leur capacité à se lier à des vésicules riches en phosphatidylcholine et leur aptitude à induire la libération d'acide arachidonique [Singer *et al.*, 2002]. Ceci implique que les sPLA2 V et X, qui se lient plus fortement à des vésicules de ce type soient des cibles thérapeutiques de choix parmi les différentes sPLA2 [Smart *et al.*, 2004], à condition quelle soient exprimées dans les processus pathologiques, ce qui est effectivement le cas, en particulier pour sPLA2-X. Celle-ci, qui présente en outre une préférence pour les chaînes sn-2 de type arachidonyle [Singer *et al.*, 2002], est donc une protéine particulièrement pertinente dans le cadre de la recherche de nouveaux médicaments, et a fait l'objet d'une étude approfondie.

## **C. DÉTAILS DU SYSTÈME ÉTUDIÉ**

### **I. STRUCTURE DE LA SPLA2-X HUMAINE**

Sur la base des classifications établies notamment par Dennis, en 1997 [Dennis,

1997], Cupillard et al. [Cupillard *et al.*, 1997] identifiaient la même année une sPLA2 humaine faisant partie d'un nouveau groupe unique de sPLA2. En effet, en se basant sur ses caractéristiques structurales, celle-ci ne pouvait se classer au sein d'aucun des groupes déjà connus. D'abord l'homologie de séquence entre hGX et les autres sPLA2 est située entre 27 et 35%, sans profil permettant *a priori* de la rapprocher d'un des groupes connus plutôt que d'un autre, bien qu'une comparaison plus complète de la séquence protéique d'hGX révèle un niveau d'identité légèrement plus élevé avec le groupe II. Ensuite, comme toutes les sPLA2 elle présente de nombreuses cystéines dont on peut supposer qu'elles sont capables de s'assembler pour former un certain nombre de ponts disulfures. En fait, parmi les seize cystéines incluses dans la structure d'hGX, deux, en position 11 et 77, sont caractéristiques des sPLA2 de groupe I, et deux, en position 50 et 137, sont caractéristiques des sPLA2 du groupe II, les autres étant situées dans des positions que l'on retrouve indifféremment dans les deux groupes. Elle présente également une extension du côté C-terminal similaire aux sPLA2 du groupe II, mais ne montre pas la boucle des élapidés (elapid loop) caractéristique des sPLA2 IA, ni la boucle pancréatique (pancreatic loop) caractéristique des sPLA2 IB, pas plus que l'insertion caractéristique des sPLA2 IIC. En revanche, elle présente bien tous les résidus conservés chez les sPLA2 actives, permettant de considérer qu'elle est potentiellement active catalytiquement, ce qui sera confirmée par la suite.

La structure cristallographique de la phospholipase A2 de groupe X humaine sans ligand co-cristallisé a été résolue par Pan et al. en 2002 [Pan *et al.*, 2002], avec une résolution de 1.97 Å (code PDB 1le6). Comme les autres sPLA2, hGX montre une distinction très nette entre la surface de liaison interfaciale ou i-face, qui détermine l'arrimage de la protéine sur la membrane et le site actif à proprement parler. L'i-face constitue une région quasiment plane située autour de l'ouverture qui constitue le site actif, qui est lui formé d'une poche de forme conique d'environ 15 Å de profondeur capable d'accueillir un phospholipide unique, au fond de laquelle se trouvent les résidus catalytiques nécessaires à son hydrolyse (Figure 15).

Les résidus du site actif étant fortement conservés, la structure de celui-ci possède une forte similarité avec les structures d'autres sPLA2 cristallisées (groupe IB, code PDB 1g4i [Steiner *et al.*, 2001], groupe IIA, code PDB 1pod [Scott *et al.*, 1991] et 1db4 [Schevitz *et al.*, 1995]), mais d'importantes différences structurales apparaissent au niveau de l'i-face, en particulier sur une boucle formée par les résidus 14 à 20 qui s'avère critique pour la liaison sur la membrane. Ces différences expliquent aussi la présence d'une cavité bien plus ouverte, en accord avec la capacité de hGX à hydrolyser des phospholipides contenant des chaînes sn-2 insaturées de type arachidonyle. De plus, l'i-face d'hGX est globalement

apolaire et donc électrostatiquement neutre alors que hGIIA présente une i-face fortement chargée positivement (Figure 15) [Pan *et al.*, 2002]. Des études de mutagenèse ont montré que l'i-face de hGX contient un tryptophane dont l'importance est capitale pour la liaison de la protéine à des vésicules riches en PC. Ce résidu, absent chez hGIIA, a plus d'influence sur l'affinité de liaison aux membranes riches en PC que la présence ou l'absence de résidus cationiques de type lysine et arginine [Singer *et al.*, 2002].

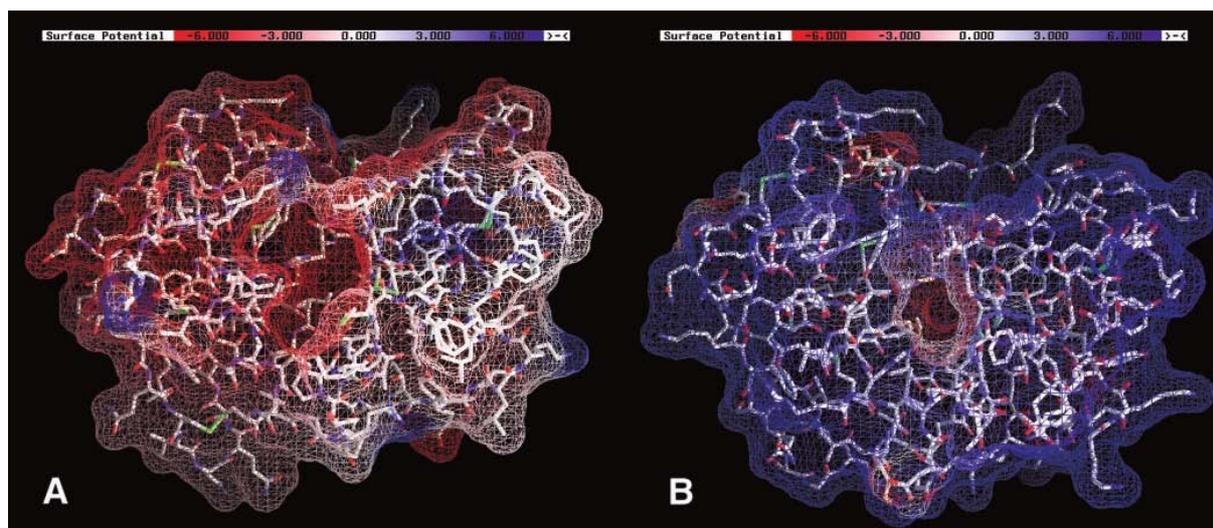


Figure 15. Comparaison de l'i-face et du site actif de hGX et hGIIA [Pan *et al.*, 2002]

Comme indiqué précédemment, l'ion calcium présent dans le site actif de la protéine est indispensable à l'activité de catalyse, la vitesse de la réaction en l'absence de calcium étant nulle pour toutes les formes de sPLA2 humaines et murines alors qu'elle augmente de façon hyperbolique avec la concentration en calcium [Singer *et al.*, 2002]. Dans la structure cristallographique, l'ion  $\text{Ca}^{2+}$  se trouve stabilisé dans une géométrie octaédrique à l'aide de sept atomes d'oxygène situés à des distances entre 2.3 et 2.6 Å. Trois d'entre eux sont des oxygènes de type carbonyle fournis par F26, G28 et G30 de la boucle de liaison au calcium, deux proviennent du carboxylate de D47, et les autres sont issus de deux molécules d'eau contenues dans le site actif [Pan *et al.*, 2002].

## II. INHIBITEURS CONNUS

Comme indiqué précédemment, les sPLA2, et hGX en particulier sont impliquées dans un grand nombre de processus physiologiques et pathologiques, d'où l'intérêt de développer des inhibiteurs pour ces protéines. A ce jour, il existe plusieurs inhibiteurs connus de PLA2, qui peuvent se classer en trois catégories. Les premiers, non spécifiques, sont capables de se lier à toutes les formes de PLA2 intracellulaires ou extracellulaires. Les seconds sont

capables de différencier les PLA2 selon leur localisation extra ou intracellulaire en détectant le type de résidu catalytique (histidine pour les sPLA2 et sérine pour les cPLA2). Enfin les derniers peuvent distinguer les isoformes de la protéine au sein d'une même classe [Cummings, 2007], s'avérant ainsi les plus utiles dans le cadre du développement de candidats médicaments.

Parmi les inhibiteurs de sPLA2 connus, on trouve des analogues de phospholipides tels que les phosphonates, mais aussi de nombreuses séries de molécules non-analogues de phospholipides [Smart *et al.*, 2004]. C'est dans la séries des indoles et des indolizines, identifiées dès le milieu des années 1990, que se présentent les inhibiteurs les plus fortement actifs et les plus susceptibles de constituer de bons candidats médicaments en terme de propriétés pharmacocinétiques. Sur l'ensemble des sPLA2, c'est le Me-Indoxam (Figure 34) qui paraît le plus puissant parmi les divers inhibiteurs testés provenant de plusieurs familles chimiques [Singer *et al.*, 2002]. Smart *et al.* [Smart *et al.*, 2004] rapportent avoir résolu une structure cristallographique de hGX co-cristallisée avec le composé Me-Indoxam mais celle-ci n'a pas été publiée et est introuvable sur la PDB. Selon cette publication, cette structure montre une coordination du calcium similaire à celle observée dans la structure apo [Pan *et al.*, 2002], avec sept atomes d'oxygènes, la seule différence se situant au niveau des oxygènes fournis par les deux molécules d'eau qui sont remplacés par les atomes d'oxygène de l'amide et du carboxyle de Me-Indoxam. A l'opposé, la chaîne N1 qui porte un méthyl-biphényle dépasse vers l'extérieur de la cavité, tout comme le substituant en position 6 sur l'indole. La position 2 correspond à un point de diversité particulièrement important car il fait face à un renflement dans le site actif caractéristique de hGX qui devrait permettre d'accueillir un groupement plus large que le méthyle de Me-Indoxam [Smart *et al.*, 2006].

Les composés LY315920 (Varespladib), Indoxam et Me-Indoxam (Figure 16) issus des séries indoles et indolizines ont servi de base à Oslund *et al.* [Oslund *et al.*, 2008] pour la synthèse d'une série de composés qui ont ensuite été testés sur les phospholipases de groupe IB, IIA, IID, IIE, IIF, V et X humaines et murines.

Ces composés nous ont permis de construire un training set de vingt-huit molécules dont les IC50 sur hGX varient entre 10 nM environ et plus de 1500 nM. Les structures des 25 molécules dérivées des modèles présentés en Figure 16 figurent dans l'annexe 4, « Structures 2D des inhibiteurs connus de hGX ».

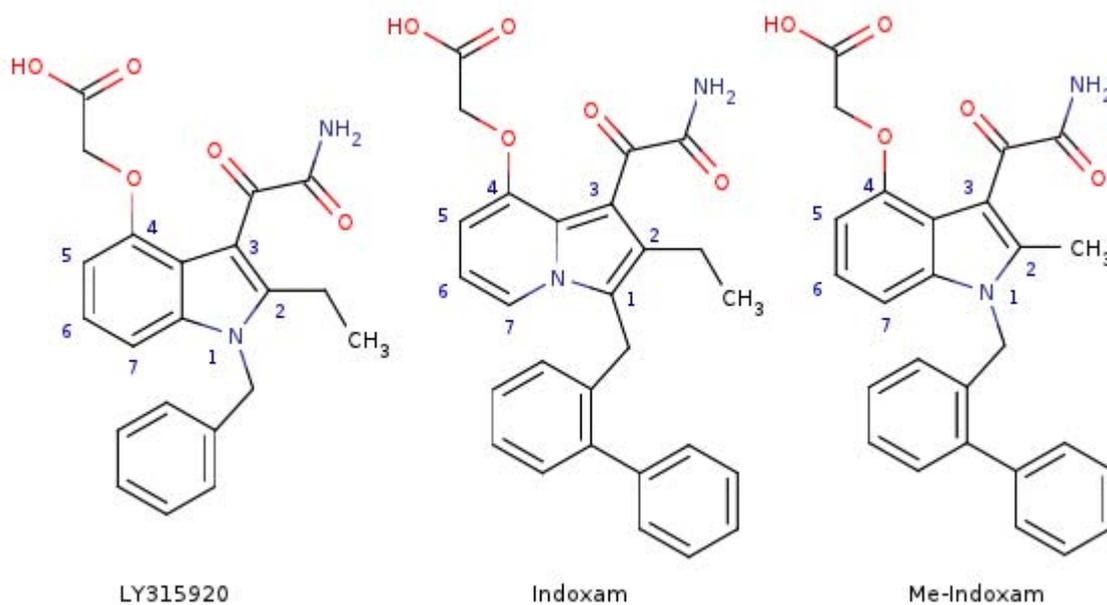


Figure 16. Structure 2D des inhibiteurs modèles de sPLA2

### III. BASE DE LIGANDS INNOVANTS

Au sein du laboratoire, l'équipe de Gilles Guichard avec laquelle nous avons collaboré réalise actuellement la synthèse de composés pour lesquels on soupçonne une activité inhibitrice sur les sPLA2. Les activités biologiques observées pour des molécules contenant un motif diazépine ou triazépine, et plus particulièrement de celles présentant un cycle à sept atomes comportant une urée ont fourni un point de départ à ce projet de synthèse [Lena *et al.*, 2006]. Les ligands étudiés sont donc des composés heptacycliques de type triazépanedione comprenant une urée et un amide. Ces hétérocycles dérivés de peptides, inspirés de la pipérazine-2,5-dione qui fait figure de modèle dans ce domaine, possèdent des motifs similaires à de nombreux produits naturels et peuvent être fonctionnalisés de manière extensive, grâce à cinq points de diversité.

A l'aide d'un criblage virtuel inverse au cours duquel cinq composés représentatifs ont été dockés sur 2148 cibles protéiques extraites de la PDB à l'aide du programme Gold, et de tests enzymatiques, les sPLA2, et plus spécialement hGV et hGX ont été identifiées comme des cibles potentielles pour cette série de molécules [Muller *et al.*, 2006].

La structure des molécules étudiées indiquant les points de diversité possibles est montrée en Figure 28. En plus des diverses substitutions envisageables, quelques molécules contenant des cycles à cinq ou six maillons, substitués ou non, entre les positions

R1 et R4 ou R et R3, ont également été synthétisées.

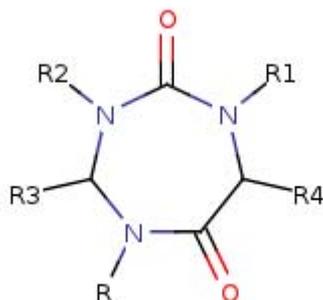


Figure 17. Structure de base des ligands innovants de sPLA2 synthétisés dans l'équipe de Gilles Guichard

## D. MATÉRIEL ET MÉTHODES

### I. PRÉPARATION DU RÉCEPTEUR

La structure 3D de la phospholipase humaine de groupe X issue de la PDB (code PDB 1le6) [Pan *et al.*, 2002] nous a fourni un modèle pour le criblage virtuel. Le fichier contient trois chaînes, A, B et C, mais seule la chaîne A a été conservée. A l'aide de Discovery Studio 2.1 (Accelrys, San Diego, USA), les molécules d'eau et les ions phosphates contenus dans la structure ont été retirés et les hydrogènes rajoutés à l'aide du champ de force Charmm. La protéine est complète et ne comporte aucun résidu manquant. Afin de corriger d'éventuels problèmes de longueurs de liaisons ou d'angles incorrects, une minimisation de la protéine a été réalisée. La structure utilisée est une structure apo ce qui signifie qu'elle ne contient pas de ligand co-cristallisé. Pour éviter que le site actif ne soit bloqué par des chaînes latérales qui auraient bougé pendant la minimisation, le ligand MJ33 extrait de la structure de la sPLA2 porcine de groupe X [Yu *et al.*, 2005] a été inséré dans le site actif. Les atomes du squelette de la protéine et le ligand ont été immobilisés à l'aide d'une contrainte, et la structure a ensuite subi une minimisation de type Steepest Descent, 2000 pas, avec un gradient fixé à 0,01.

## II. RECHERCHE CONFORMATIONNELLE SUR LES LIGANDS

La structure de l'heptacycle qui constitue la base des molécules synthétisées dans l'équipe de Gilles Guichard peut adopter différentes conformations. Des études en diffraction des rayons X et  $^1\text{H}$  RMN ont permis d'isoler des configurations préférentielles pour ces cycles particuliers. Le cycle adopte le plus souvent une forme en « V » avec les substituants R2 et R4 en position pseudo-équatoriale. L'angle entre les plans formés par l'urée et l'amide indiquent deux conformations possibles, l'une avec un angle d'environ  $120^\circ$  et l'autre avec un angle de  $160^\circ$  [Lena *et al.*, 2006]. Le programme MacroModel 6.5 (Schrödinger, LLC, New York, USA) [Chang *et al.*, 1989; Saunders *et al.*, 1990], a été utilisé pour réaliser une recherche conformationnelle de type Monte Carlo sur l'heptacycle étudié. Deux molécules représentatives de nos composés ont été sélectionnées : MR1 (nommée G3 dans notre base et mol2 dans la publication de Muller *et al.* [Muller *et al.*, 2006]) et MR2 (G27 ou mol30). Elles incluent plusieurs substituants différents en terme de nature et de position, et l'une des deux montre une cyclisation entre les positions R et R3 ce qui permet d'étudier l'influence de ce cycle annexe sur la conformation du macrocycle (Figure 18).

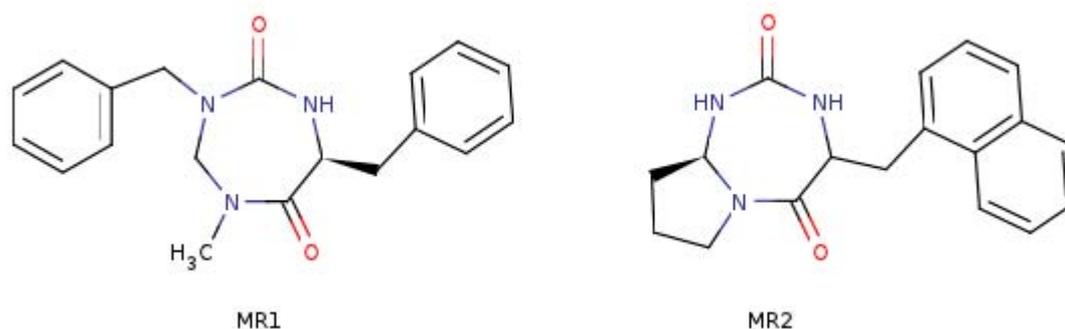


Figure 18. Structures 2D des molécules étudiées lors de la recherche conformationnelle

A la suite de cette première étude, une seconde, portant sur les mêmes molécules débarrassées de tous leurs substituants (structures simplifiées MRS1 et MRS2) a également été réalisée afin de vérifier la validité des conformations trouvées dans des cas généraux (Figure 19).

Les paramètres de configuration automatique ont été conservés et la recherche effectuée pour chaque structure a été faite sur 2000 pas. Chaque conformère obtenu a ensuite été minimisé à l'aide d'un algorithme de type TNCG (truncated Newton Conjuguate Gradient), 5000 pas, fenêtre énergétique autorisée  $10 \text{ kcal.mol}^{-1}$ ,  $\text{RMS} \leq 0,01 \text{ kJ.Å}^{-1}.\text{mol}^{-1}$ , et du champ de force MMFF94 [Halgren, 1996]. Seules les configuration trouvées plus de

trois fois ont été considérées comme convergentes. Les conformères obtenus ont été clusterisés à l'aide de XCluster 1.5 [Shenkin & McDonald, 1994], en utilisant le RMS sur les atomes lourds uniquement comme critère de sélection. Dans chaque cluster, un conformère représentatif a été sauvé pour servir de base à la construction des molécules.

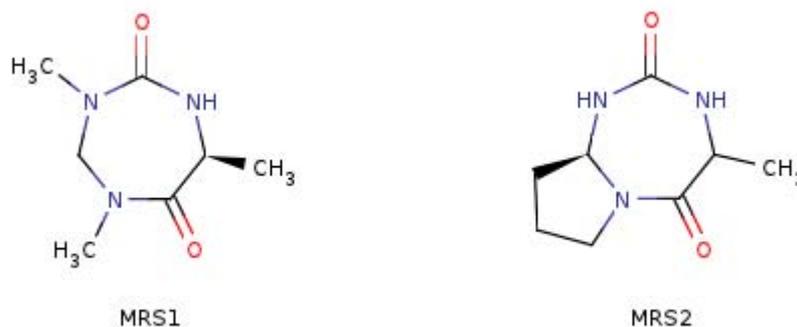


Figure 19. Structures simplifiées des molécules étudiées

### III. PRÉPARATION DES LIGANDS

Les ligands du training set tiré de Oslund et al. [Oslund *et al.*, 2008] ont été construits manuellement dans Discovery Studio 2.1. Le protocole *Dreiding Minimize* a été utilisé pour atteindre une conformation de basse énergie. Les structures obtenues ont été enregistrées en MOL2 et converties en PDBQT, le format d'entrée requis pour le docking, à l'aide du script *prepare\_ligand4.py* fourni par MGLTools [Sanner, 1999].

Concernant la banque de molécules fournies par l'équipe Guichard, les molécules ont été reconstruites sur la base des conformations obtenues dans l'analyse conformationnelle, auxquelles les substituants ont été rajoutés dans Discovery Studio 2.1. Pour obtenir des conformations de basse énergie au niveau des substituants sans altérer les structures du macrocycle identifiées par MacroModel, les atomes de celui-ci ont été immobilisés à l'aide d'une contrainte avant d'appliquer le protocole *Dreiding Minimize*. Comme pour le training set, les molécules sauvées en MOL2 ont été converties en PDBQT grâce au script *prepare\_ligand4.py* [Sanner, 1999].

### IV. DOCKING ET RESCORING

Les docking ont été réalisés avec le programme Autodock Vina version 1.0.2 [Trott & Olson, 2010]. La protéine a été gardée rigide pendant le docking tandis que les ligands ont été considérés comme flexibles. Les cycles sont considérés comme fixes même si plusieurs

conformations peuvent exister. L'ensemble des torsions autorisées pour chaque ligand a été attribué automatiquement lors du passage en PDBQT.

Le site actif dans lequel le docking a été effectué correspond au site de liaison des phospholipides dans hGX, dont la localisation est clairement explicitée dans la littérature. Pour les besoins du criblage, nous avons donc défini un espace de recherche comprenant la totalité de la cavité plus une marge de quelques ångströms dans toutes les directions libres. Les dimensions de la boîte de recherche conformationnelle résultante ont été fixées à 16×18×16 Å.

Les paramètres par défaut de Vina ont été utilisés pour les calculs, à savoir :

- exhaustiveness = 8
- nombre de poses générées = 9

Pour chaque ligand, le docking a été répété cinq fois, en utilisant cinq seed différents pour chaque expérience, mais égaux pour tous les ligands dans une expérience.

Les fichiers de sortie de Vina en PDBQT ont été convertis en PDB à l'aide de scripts écrits dans le laboratoire (voir l'annexe 2 « Gestion des fichiers et automatisation des calculs »), et les hydrogènes non polaires, non considérés par Vina, ont été rajoutés automatiquement à l'aide d'OpenBabel 2.2.3 [Guha *et al.*, 2006].

Un rescoring complet des poses générées par Vina a été effectué avec toutes les fonctions de score incluses dans les programmes Drugscore [Gohlke *et al.*, 2000; Velec *et al.*, 2005] et Xscore [Wang *et al.*, 2002]. Les cinq fonctions de score de Drugscore sont PAIR\_csd, PAIRSURF\_csd, PAIR\_pdb, PAIRSURF\_pdb et SURF. Les quatre fonctions de score de Xscore sont HPscore, HMscore, HScore et AVEScore. Avec Vina, dix fonctions de scoring ont donc été utilisées.

Pour chacune des poses d'un ligand, les scores calculés par toutes les fonctions de score dans les cinq expériences ont été moyennés pour donner le score final par pose et par fonction. Parmi ces scores, le meilleur score obtenu pour chaque fonction de score pour un ligand a été sauvé pour créer un classement des molécules.

## **E. RÉSULTATS ET DISCUSSION**

### **I. CONFORMATION OBTENUES**

Pour les deux molécules représentatives choisies, il existe une bonne convergence dans les conformations obtenues comme indiqué par le grand nombre de solutions trouvées pour chaque conformation unique. La première molécule étudiée (MR1) (Figure 20), qui

possède des substituants en R2, R4 et R, mais pas de cyclisation annexe au cycle à sept maillons, présente vingt et une conformations uniques dans une fenêtre énergétique de 10 kcal.mol<sup>-1</sup>.

```

Final report:
  21 unique conformations found
  21 minimized with good convergence
Found    2 confs within  1.00 kcal/mol ( 4.18 kJ/mol) of glob. min.
Found    5 confs within  2.00 kcal/mol ( 8.37 kJ/mol) of glob. min.
Found   11 confs within  3.00 kcal/mol (12.55 kJ/mol) of glob. min.
Found   16 confs within  5.00 kcal/mol (20.92 kJ/mol) of glob. min.
Found   21 confs within 10.00 kcal/mol (41.84 kJ/mol) of glob. min.
Global minimum E = -225.04 found  330 times.
Total number of structures processed = 5000
Conformations with poor convergence marked with a *
Conformation    1 (  -225.0416    kJ/mol) was found  330 times
Conformation    2 (  -221.0315    kJ/mol) was found  239 times
Conformation    3 (  -219.5412    kJ/mol) was found  269 times
Conformation    4 (  -219.0174    kJ/mol) was found  247 times
Conformation    5 (  -216.7891    kJ/mol) was found  228 times
Conformation    6 (  -216.3648    kJ/mol) was found  243 times
Conformation    7 (  -216.1818    kJ/mol) was found  420 times
Conformation    8 (  -214.9409    kJ/mol) was found  151 times
Conformation    9 (  -214.1784    kJ/mol) was found  189 times
Conformation   10 (  -214.0757    kJ/mol) was found  142 times
Conformation   11 (  -212.6844    kJ/mol) was found  205 times
Conformation   12 (  -210.9162    kJ/mol) was found  200 times
Conformation   13 (  -207.4201    kJ/mol) was found  245 times
Conformation   14 (  -206.8462    kJ/mol) was found  200 times
Conformation   15 (  -206.8250    kJ/mol) was found  200 times
Conformation   16 (  -204.4034    kJ/mol) was found  258 times
Conformation   17 (  -203.9420    kJ/mol) was found  110 times
Conformation   18 (  -201.6047    kJ/mol) was found  253 times
Conformation   19 (  -193.5280    kJ/mol) was found   77 times
Conformation   20 (  -193.0786    kJ/mol) was found  123 times
Conformation   21 (  -187.1156    kJ/mol) was found  116 times

```

Figure 20. Résultats de la recherche conformationnelle pour MR1

La seconde (MR2) (Figure 21), qui possède un cycle à cinq maillons entre R et R3 et un substituant en R4, présente vingt conformations uniques qui sont toutes dans une fenêtre énergétique de 5 kcal.mol<sup>-1</sup> (bien que la fenêtre autorisée ait été de 10 kcal.mol<sup>-1</sup> comme pour MR1).

```

Final report:
  20 unique conformations found
  20 minimized with good convergence
Found    6 confs within  1.00 kcal/mol ( 4.18 kJ/mol) of glob. min.
Found   10 confs within  2.00 kcal/mol ( 8.37 kJ/mol) of glob. min.
Found   17 confs within  3.00 kcal/mol (12.55 kJ/mol) of glob. min.
Found   20 confs within  5.00 kcal/mol (20.92 kJ/mol) of glob. min.
Global minimum E = -272.00 found 188 times.
Total number of structures processed = 5000
Conformations with poor convergence marked with a *
Conformation    1 (  -271.9994    kJ/mol) was found 188 times
Conformation    2 (  -271.9619    kJ/mol) was found 203 times
Conformation    3 (  -271.7383    kJ/mol) was found 227 times
Conformation    4 (  -271.7190    kJ/mol) was found 216 times
Conformation    5 (  -268.2822    kJ/mol) was found 314 times
Conformation    6 (  -267.9987    kJ/mol) was found 329 times
Conformation    7 (  -267.2603    kJ/mol) was found 112 times
Conformation    8 (  -267.0277    kJ/mol) was found 146 times
Conformation    9 (  -266.5312    kJ/mol) was found 189 times
Conformation   10 (  -266.5161    kJ/mol) was found 175 times
Conformation   11 (  -262.9549    kJ/mol) was found 101 times
Conformation   12 (  -262.3165    kJ/mol) was found 121 times
Conformation   13 (  -261.8476    kJ/mol) was found 108 times
Conformation   14 (  -261.8151    kJ/mol) was found 135 times
Conformation   15 (  -260.5089    kJ/mol) was found 214 times
Conformation   16 (  -259.5156    kJ/mol) was found 247 times
Conformation   17 (  -259.5123    kJ/mol) was found 230 times
Conformation   18 (  -259.2506    kJ/mol) was found 251 times
Conformation   19 (  -255.1301    kJ/mol) was found 148 times
Conformation   20 (  -254.8903    kJ/mol) was found 134 times

```

Figure 21. Résultats de la recherche conformationnelle pour MR2

La clusterisation avec XCluster, qui a normalement pour but de rassembler les conformations proches après superposition automatique, s'avère non concluante, à cause des substituants qui peuvent adopter une multitude de positions différentes grâce à plusieurs torsions. Afin d'identifier les structures représentatives du cycle étudié pour nos deux composés, il a donc été nécessaire de procéder à une superposition manuelles des conformation uniques obtenues en supprimant ces substituants. Les superpositions obtenues mettent en évidence la présence de quatre conformations représentatives pour les deux molécules étudiées, les conformations C1, C13, C3 et C4 pour MR1 et C1, C5, C2 et C7 pour MR2. Pour les conformations C1, et C13 de MR1 et C1 et C5 de MR2, les substituants se placent en position axiales, tandis qu'ils sont en position équatoriale pour les conformations C3 et C4 de MR1 et C2 et C7 de MR2.

Les structures obtenues sont également superposables malgré la présence du cycle sur le composé MR2. En effet, le cycle supplémentaire conduit à une réduction du nombre de conformations possibles, puisque les conformations C1 et C5 avec les substituants en position axiale sont équivalentes entre elles et superposables à la conformation C1 de MR1

si on exclut de prendre en compte la conformation du cycle à cinq atomes. De même, les conformations C2 et C7 de MR2 avec les substituants en position équatoriale sont équivalentes et superposables à la conformation C4 de MR1. Les structures cristallines rapportées précédemment [Lena *et al.*, 2006] sont superposables aux conformères C4 de MR1 et C2 de MR2.

La recherche conformationnelle sur les molécules représentatives simplifiées confirme la validité de ces conformations pour les deux composés. La recherche conformationnelle sur MR1 simplifiée (MRS1) donne quatre conformations uniques identiques à celles obtenues pour MR1 (Figure 22), tandis que la recherche concernant MR2 simplifiée (MRS2) en donne deux, également superposables à celles obtenues avec MR2 (Figure 23).

```
Final report:
  4 unique conformations found
  4 minimized with good convergence
Found    3 confs within  1.00 kcal/mol ( 4.18 kJ/mol) of glob. min.
Found    4 confs within  5.00 kcal/mol (20.92 kJ/mol) of glob. min.
Global minimum E = -378.88 found  495 times.

Total number of structures processed = 5000
Conformations with poor convergence marked with a *
Conformation    1 (  -378.8823    kJ/mol) was found  495 times
Conformation    2 (  -376.5991    kJ/mol) was found  396 times
Conformation    3 (  -375.0109    kJ/mol) was found  620 times
Conformation    4 (  -359.0926    kJ/mol) was found  534 times
```

Figure 22. Résultats de la recherche conformationnelle pour MRS1

```
Final report:
  2 unique conformations found
  2 minimized with good convergence
Found    1 confs within  1.00 kcal/mol ( 4.18 kJ/mol) of glob. min.
Found    2 confs within  3.00 kcal/mol (12.55 kJ/mol) of glob. min.
Global minimum E = -437.27 found 1264 times.

Total number of structures processed = 5000
Conformations with poor convergence marked with a *
Conformation    1 (  -437.2674    kJ/mol) was found 1264 times
Conformation    2 (  -427.0163    kJ/mol) was found  771 times
```

Figure 23. Résultats de la recherche conformationnelle pour MRS1

Les quatre conformations obtenues sont visibles en Figure 24, avec en haut à gauche, la conformation C4 de MR1 superposée avec C2 et C7 de MR2, en haut à droite, la conformation C1 de MR1 superposée avec C1 et C5 de MR2, et en bas C3 et C13 de MR1 qui ne sont pas retrouvées pour MR2.

Comme indiqué précédemment, les cycles sont considérés comme des entités non déformables pendant le docking avec Vina. Suite à cette étude, et afin de ne pas biaiser les résultats, il paraît donc plus juste de considérer, pour chaque ligand de notre banque de molécules, quatre conformères (ou deux selon les cas) correspondant à une conformation unique trouvée par le biais de la recherche conformationnelle effectuée. Nous en avons tenu compte et la base de ligands créée contient donc plusieurs exemplaires de chaque molécule. La base finale comprend 625 conformères représentatifs pour 155 ligands différents.

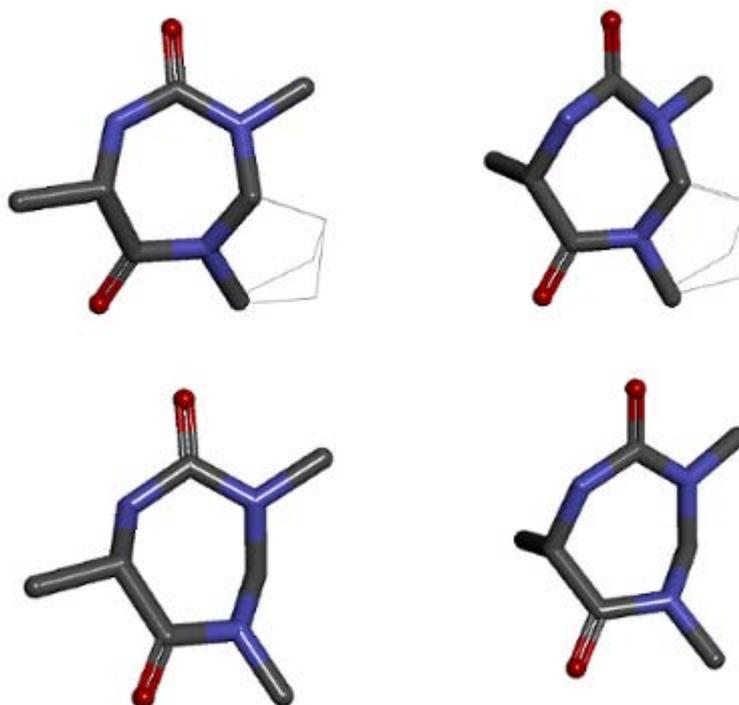


Figure 24. Conformations obtenues après traitement par MacroModel

## II. POSES DES INHIBITEURS CONNUS

A l'aide des informations contenues dans la publication de Smart et al. [Smart *et al.*, 2004] qui mentionne une structure de hGX co-cristallisée avec le Me-Indoxam et des différentes structures cristallographiques de sPLA2 disponibles dans la PDB, nous avons reconstruit un modèle d'hGX avec le Me-Indoxam dans le site actif. Comme décrit dans la littérature, nous avons positionné le Me-Indoxam dans la cavité de manière à faire ressortir vers l'extérieur la chaîne N1, tout en plaçant les oxygènes de l'amide et du carboxylate en position pour complexer le calcium. La pose obtenue est visible sur la Figure 25. Il existe

également dans la PDB, une structure de la sPLA2 de groupe IIA co-cristallisée avec le composé LY311727 (voir la structure en annexe 4 « Structures 2D des inhibiteurs connus de hGX »), qui est apparenté à la molécule LY315920 (code PDB 1db4 [Schevitz *et al.*, 1995]). Ces deux structures ont servi de référence pour l'analyse des résultats obtenus lors du docking. En effet, vu la similarité entre les molécules de la série et la nécessité de stabiliser le calcium au sein du site actif, on peut supposer que les autres inhibiteurs synthétisés adoptent une pose très similaire à celles de ces deux modèles.

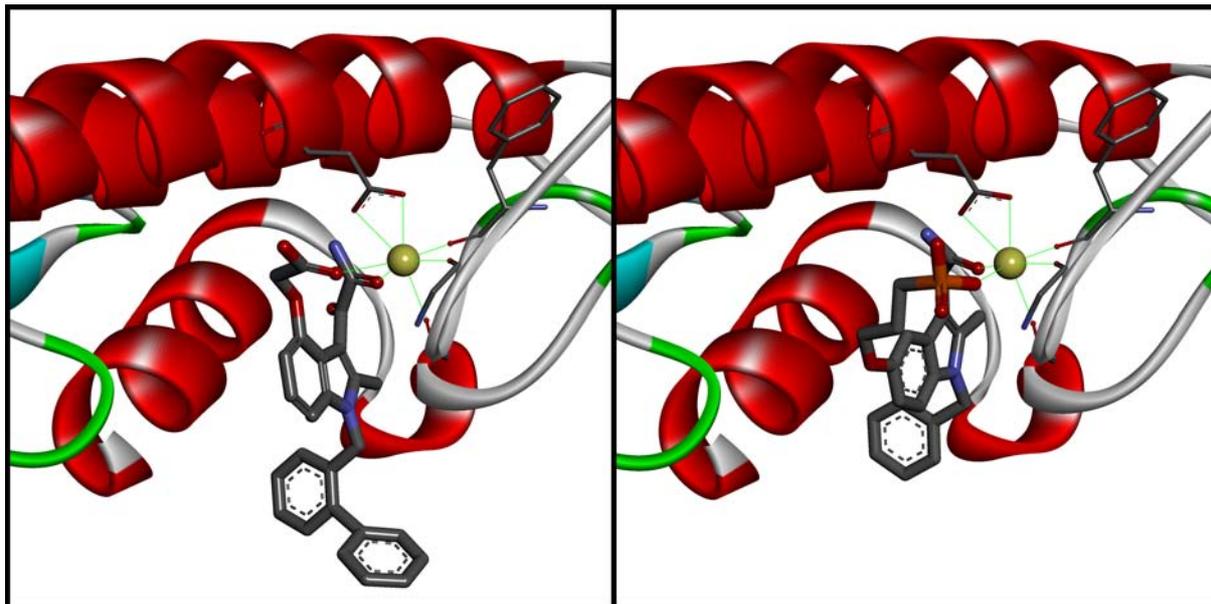


Figure 25. Modèles de poses dans le site actif d'hGX. A gauche, pose reconstruite du Me-Indoxam dans hGX, à droite pose de LY311727 d'après sa pose dans hGIIA. Le calcium est représenté par une sphère jaune et ses interactions avec la cible et les ligands sont en vert

Le docking de Me-Indoxam montre une première pose assez éloignée du modèle reconstruit en ce qui concerne la position de l'indole et de la chaîne N1, mais qui assure tout de même la coordination du calcium. En fait, on trouve toujours un oxygène du carboxylate dirigé vers l'ion métallique, mais au lieu de l'oxygène de l'amide, c'est le carbonyle situé juste avant qui complète la coordination du calcium. La seconde pose présente la même coordination, et un profil similaire à la pose reconstruite n'est retrouvé qu'en cinquième position dans le classement fourni par les scores de Vina, plus haut en énergie que deux poses ne permettant même pas la coordination du calcium. Les poses retrouvées sont en fait plus proches de la pose de l'inhibiteur LY311727 dans hGIIA [Schevitz *et al.*, 1995].

En ce qui concerne l'Indoxam, la meilleure pose obtenue présente à proximité du calcium les groupements NH<sub>2</sub> de l'amide et l'éthyle en position 2. Une coordination équivalente au modèle reconstruit est retrouvée dès la deuxième meilleure pose, sans que le

reste de la molécule soit positionné de façon similaire. En revanche, pour LY315920, les deux premières poses permettent la coordination du calcium. La première montre un profil de coordination similaire au modèle tandis que la seconde se rapproche de la première pose du Me-Indoxam avec l'oxygène du carbonyle dirigé vers le calcium au lieu de celui de l'amide et est parfaitement superposable à la pose de LY311727 dans hGIIA [Schevitz *et al.*, 1995].

De la même façon, dans la série de ligands synthétisés par Oslund *et al.* [Oslund *et al.*, 2008], les meilleures poses obtenues pour les composés 12a et 12b présentent un oxygène du carboxylate et l'oxygène du carbonyle dirigés vers le calcium, mais le groupement en N1 orienté à l'opposé du modèle et dans le même sens que LY311727. Les composés 12e et 12f, créés sur le modèle de 12a et 12b mais contenant un benzoindole à la place de l'indole, ainsi que 16a, qui est constitué d'une indolizine et porte un méthyl-biphényle similaire au Me-Indoxam en position 1, adoptent les mêmes poses. Le composé 16b, similaire au 16a sauf qu'il porte une méthyl-cétone à la place de l'acide carboxylique de la chaîne 4, permet également une coordination du calcium à partir de la deuxième meilleure pose. Le composé 15c qui présente à la fois un méthyle sur l'amide ainsi qu'une méthyl-cétone à la place de l'acide carboxylique de la chaîne 4 montre lui aussi une double coordination à la seconde pose.

A l'opposé, le composé 15b, construit également sur la base de 12e mais qui porte un méthyle sur l'amide, ne présente pas de position permettant une coordination correcte du calcium. Les composés 14a, 14b, et 16c dont la chaîne 4 a été remplacée par un simple groupement méthyl-ether sont également dans ce cas puisqu'il sont dans l'incapacité structurelle de former la « pince » constituée de deux oxygènes nécessaire à cette coordination. Les poses obtenues présentent le plus souvent un cycle et au mieux un seul oxygène à proximité du calcium.

En ce qui concerne les composés 11c, 11d, 11g, 11h et 13a, 13b, 13c, 13d, 13e, 13f, 13g, ainsi que la molécule 15a qui possèdent un groupement phénylsulfonamide très encombrant à la place de l'hydroxyle de l'acide carboxylique, aucune des poses obtenues ne peut être rapprochée du modèle fourni par le Me-Indoxam et aucune ne montre de coordination même partielle du calcium. En revanche, les composés 13h et 13i pour lesquels le phénylsulfonamide est remplacé par un méthylsulfonamide et un trifluorométhylsulfonamide présentent une meilleure pose exactement superposable à celle obtenue pour LY315920 avec l'oxygène de l'amide de la chaîne 3 dirigé vers le calcium et la chaîne 4 à proximité.

### III. POSES DES LIGANDS TRIAZÉPANEDIONE

Parmi les molécules de la base synthétisée dans l'équipe de Gilles Guichard, nous nous sommes particulièrement intéressés aux composés étudiés dans le criblage *in silico* ayant permis d'identifier les sPLA2 comme des cibles potentielles [Muller *et al.*, 2006] et notamment les six molécules qui ont montré une activité inhibitrice importante sur hGX (Structures chimiques en Figure 26).

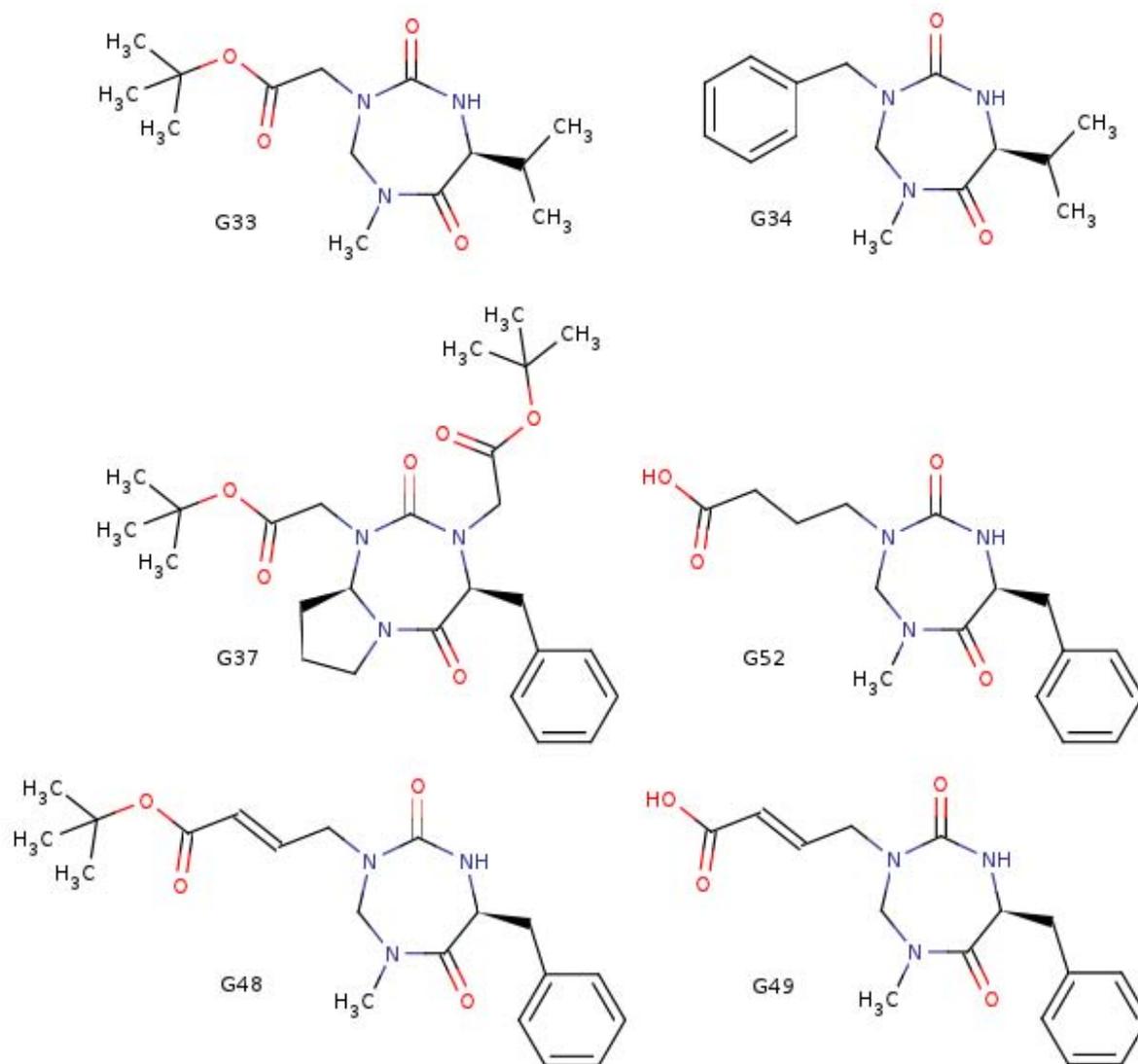


Figure 26. Structure 2D des triazépanediones actives sur hGX

A l'aide des descriptions et des images des modes de liaison fournies, les poses des molécules G49 (mol13 dans Muller *et al.* [Muller *et al.*, 2006]) et G33 (mol19) ont été reconstruites et comparées à celles obtenues lors de notre criblage (voir Figure 27 page

suivante). Les profils de liaison des deux molécules sont assez différents. La première, qui possède en R2 une longue chaîne terminée par un acide carboxylique, place le cycle triazépanedione au centre du site actif, et les substituants en R4 vers les hélices 1 et 2 constituées des résidus 1 à 20. Ceci permet de placer la chaîne R2 dans une position telle que l'oxygène de l'acide est capable de coordonner le calcium, tandis que le septième oxygène de la configuration typiquement adoptée est fourni par l'urée de l'heptacycle. La seconde en revanche, qui est faiblement substituée et qui possède une chaîne R2 plus courte adopte une position avec l'heptacycle déplacé vers les hélices 1 et 2 qui ne permet pas la coordination directe du calcium, laquelle devra donc certainement être assurée par un relai formé par des molécules d'eau comme dans la structure apo [Muller *et al.*, 2006]. Seule la pose de la molécule G49 est ressemblante à la pose adoptée par le composé LY311727 dans hGIIA [Schevitz *et al.*, 1995] et à la structure reconstruite de Me-Indoxam dans le site actif de hGX [Smart *et al.*, 2004].

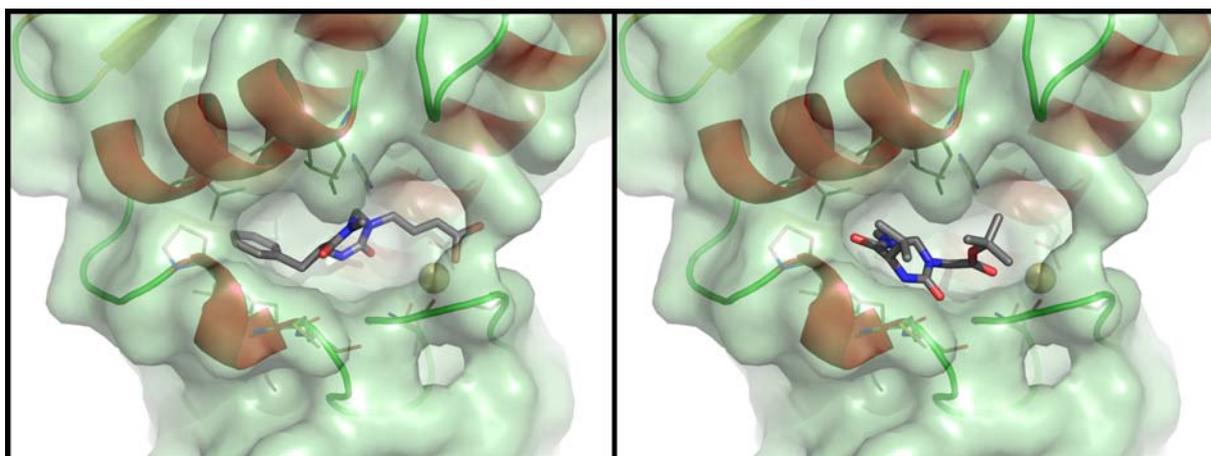


Figure 27. Poses reconstruites de G49 (gauche) et G33 (droite) dans le site actif de hGX. Le calcium est représenté par une sphère jaune

Parmi les molécules actives, on peut attendre un profil de liaison similaire à G33 pour G34 (mol20), alors que G48 (mol8) et G52 (mol10) seront probablement plus proches de la pose de G49. Il est plus difficile de classer G37 (mol32) qui possède une chaîne R2 assez courte comme G33, mais est très fortement encombrée stériquement ce qui devrait probablement empêcher une position proche des hélices 1 et 2.

Les meilleures poses obtenues pour la molécule G33 sont toutes différentes du modèle reconstruit et sont presque indifféremment positionnées loin ou près du calcium sans qu'aucune ne montre de coordination directe (Figure 28). C'est le cas également pour la molécule G34, qui présente plusieurs poses situées à proximité du calcium et des résidus Y20 et M21 de l'hélice 2 et C27, G28 et L29 de la boucle de liaison, mais aussi d'autres

placées plutôt vers l'hélice 1.

La molécule G37 ne présente aucune pose qui permette la coordination ni aucune pose située vers le fond de la cavité, ses divers substituants l'empêchant même d'adopter des poses qui ne soient pas à moitié en dehors du site actif (Figure 28).

En revanche, plusieurs poses parmi les meilleures obtenues pour le composé G48 sont assez proches de la pose reconstruite de G49, même si les deux oxygènes supposés intervenir dans la coordination du calcium ne sont pas toujours orientés vers celui-ci. C'est également le cas pour G49 qui se place de façon assez similaire au modèle (Figure 28), et pour la molécule G52.

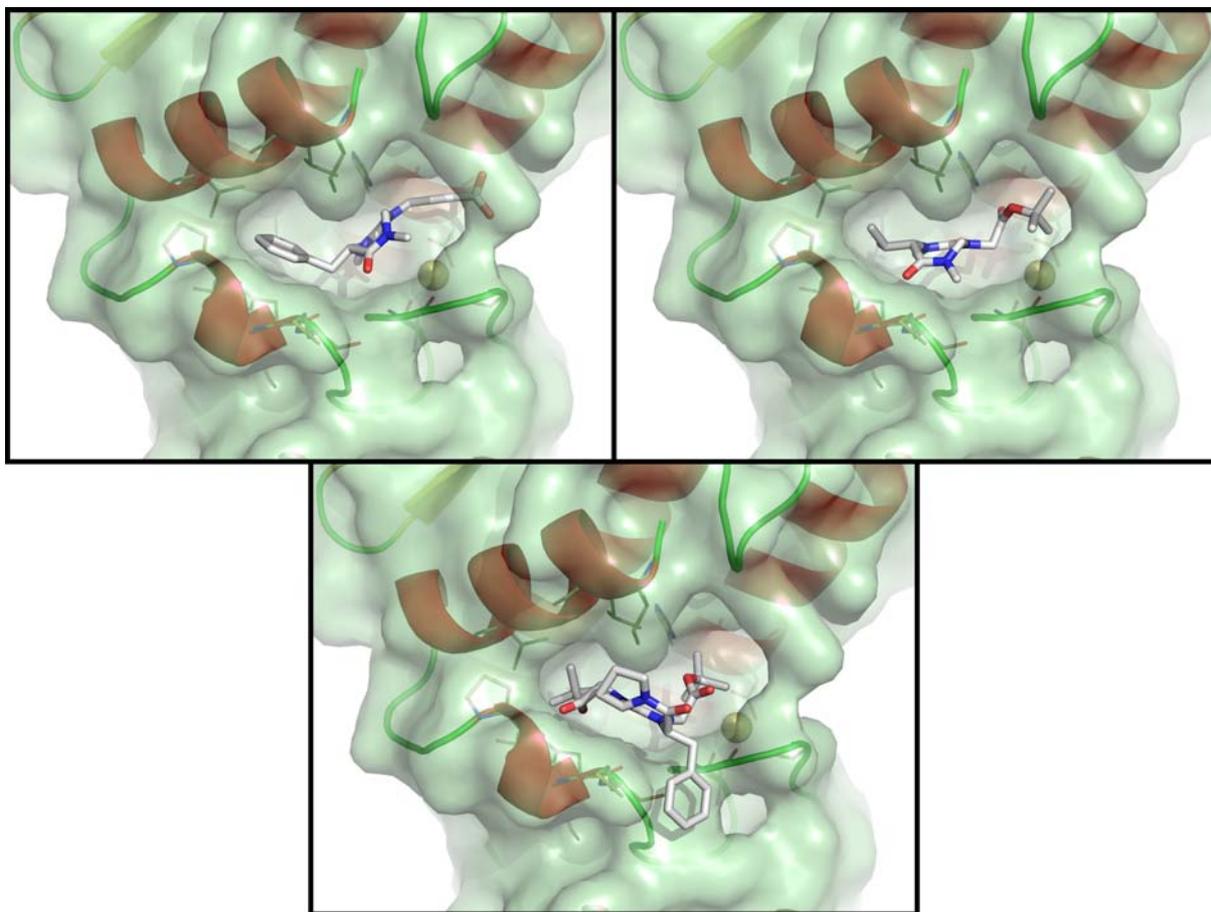


Figure 28. Meilleures poses obtenues lors du docking pour G49 (en haut à gauche), G33 (en haut à droite) et G37 (en bas). Le calcium est représenté par une sphère jaune

La comparaison des poses adoptés par les ligands en fonction de la conformation du cycle ne montre pas de différence très significative. Dans l'ensemble, le conformère 3 donne plus souvent que les autres des poses vraiment absurdes, tandis que les conformères 2 et 4 fournissent des poses légèrement plus cohérentes avec les modèles établis. Le modèle 4 correspond d'ailleurs à la conformation du cycle obtenue le plus souvent dans les études

cristallographiques effectuées précédemment [Lena *et al.*, 2006]. Toutefois ces variations sont minimales et ne concernent le plus souvent pas les poses les mieux classées, ce qui les rend négligeables.

#### IV. COMPARAISON DES POSES DES DEUX BASES DE LIGANDS

Parmi les trois molécules ayant servi de base à la synthèse des inhibiteurs connus, c'est LY315920 (Figure 16) qui semble le mieux positionné dans le site actif. Il mime parfaitement la position adoptée par l'inhibiteur LY311727 (voir la structure dans l'annexe 4 « Structures 2D des inhibiteurs connus de hGX ») dans le site actif de hGIIA observée dans la structure PDB 1db4 [Schevitz *et al.*, 1995]. Ceci est en accord avec les activités observées puisque c'est le seul des trois composés qui possède une affinité remarquable pour hGX.

La plupart des molécules capables de stabiliser le calcium de la façon attendue ont en commun une fonction amide non substituée en position 3 et une fonction acide carboxylique libre en position 4, ce qui semble indiquer que ces deux fonctions sont essentielles pour que la pose de docking obtenue présente un profil de coordination similaire au modèle de Me-Indoxam, ainsi qu'à l'Indoxam et à LY315920. Pourtant, on retrouve des poses proches de LY315920 avec les composés 13h et 13i qui présentent un seul oxygène susceptible de fournir une coordination au calcium, et qui sont de bons ligands de hGX. La coordination correcte du calcium pourrait donc être un facteur mineur ou, en tous les cas, non-déterminant dans le classement de ces poses en tête de liste, ce qui dans ce cas-là concorde avec l'activité des ligands.

En revanche, la corrélation entre la pose et l'affinité n'est pas toujours vraie puisque LY315920, 12a, 12b, 12e, 12f sont de bons inhibiteurs de hGX alors que le Me-Indoxam, l'Indoxam et la molécule 16a ont une activité beaucoup plus faible et sont pourtant dockés de façon équivalente. De plus, si les poses générées par Vina sont capables de fournir une coordination sur l'ion calcium, aucune ne montre une orientation de l'indole (ou indolizine) similaire à celle attendue, et surtout, aucune ne place la chaîne située en position 1 complètement vers l'extérieur du site comme c'est le cas dans la structure cristallographique décrite par Smart *et al.* [Smart *et al.*, 2004]. En effet, cette chaîne est le plus souvent dirigée vers le fond de la cavité et les hélices constituées des résidus 1 à 20, ou seulement partiellement orientée vers l'extérieur lorsque le phényle est remplacé par un biphenyle beaucoup plus encombrant.

Enfin, pour l'ensemble des composés comportant un phénylsulfonamide sur la chaîne 4, aucune pose correcte n'a pu être obtenue, en raison d'un encombrement stérique trop important. Ceci indique clairement que les classements tirés de ce docking doivent être

analysés en gardant un œil très critique sur ces molécules qui sont pourtant de bons inhibiteurs de hGX.

De même, l'étude des ligands synthétisés par l'équipe de Gilles Guichard n'a pas permis de reproduire les résultats provenant de la littérature. Les poses obtenues ne correspondent à celles attendues que lorsque les ligands ne sont ni trop encombrés, ni de trop petite taille, ce qui fausse complètement les résultats de scoring.

## V. CLASSEMENTS

Pour faciliter l'analyse des scores, les ligands synthétisés par Oslund et al. [Oslund *et al.*, 2008] ont été classés en quatre catégories en fonction de leur affinité pour hGX. Les molécules possédant une IC50 supérieure à 1500 nM ont été considérées comme de mauvais inhibiteurs, les composés avec une IC50 entre 1499 et 500 nM comme des inhibiteurs moyens, ceux avec une IC50 entre 499 et 50 nM comme de bons inhibiteurs et enfin ceux pour lesquels l'IC50 est inférieure à 49 nM sont considérés comme de très bons inhibiteurs. Ces catégories ont été différenciées par un code couleur allant du rouge pour les moins bons inhibiteurs au vert pour les meilleurs en passant par le orange puis le jaune.

Les IC50 ainsi que les résultats de scoring pour tous les inhibiteurs connus sont présentés dans la Table 8, page suivante.

Comme on peut le voir rapidement, aucune fonction de score n'est capable de classer correctement les ligands connus qui nous ont servi de training set. Au vu des poses obtenues pour certains inhibiteurs ayant des affinités particulièrement élevées pour hGX qui ne correspondent pas du tout aux poses attendues, ce résultat était prévisible. Les ligands ayant été mal positionnés dans le site actif, ils sont classés en fonctions de scores calculés sur la base d'interactions qui ne correspondent pas à celles observées dans les cristallographies accessibles.

D'après l'analyse des poses ci-dessus, nous avons essayé de trouver une corrélation entre les classement obtenus et les poses ou les motifs communs aux différents ligands. Les deux ligands qui sont le mieux classés en moyenne par toutes les fonctions de scoring sont les ligands 11g et 11h qui sont d'assez bons inhibiteurs de hGX mais pour lesquels les poses obtenues semble réfutables du fait de leur fort encombrement. On retrouve également dans le haut du classement la molécule 15c, qui est quasiment inactive sur hGX et ne fournit pas de coordination au calcium à cause d'un groupement méthyle sur l'amide et du remplacement du carboxyle par une méthyl-cétone. L'Indoxam et le Me-Indoxam sont également assez bien classés alors que le LY315920 qui est, parmi les trois molécules modèles, le meilleur inhibiteur de hGX, se retrouve à la toute dernière place avec un

classement moyen de 27.4 sur 28 pour l'ensemble des dix fonctions de scoring.

Lors de l'analyse des poses issues du docking, nous avons constaté un problème récurrent sur certaines des molécules les plus actives qui étaient positionnées à l'extérieur du site actif à cause d'un substituant de type phényl-sulfonamide très encombrant, tandis que les ligands moins encombrés se positionnaient de façon plus similaire aux poses cristallines. Afin de déterminer l'influence du mauvais positionnement de ces composés particuliers sur l'échec du classement global des inhibiteurs, de nouveaux tableaux de score ont été établis, dans lesquels ces composés n'étaient pas comptabilisés. Les tableaux obtenus sont très similaires à ceux montrés ici et mettent en évidence un problème plus profond dans le criblage de ce système.

Chapitre 3 : Application à la phospholipase humaine sPLA2 de groupe X

| Ligand    | IC50  |
|-----------|-------|
| pla13h    | 7     |
| pla12f    | 10    |
| pla13f    | 11    |
| pla13d    | 14    |
| pla11c    | 15    |
| pla12e    | 20    |
| pla11d    | 21    |
| pla12a    | 22    |
| pla11h    | 30    |
| pla13i    | 30    |
| pla12b    | 50    |
| pla13g    | 70    |
| LY315920  | 75    |
| pla13a    | 80    |
| pla11g    | 90    |
| pla13c    | 140   |
| pla13e    | 320   |
| pla13b    | 540   |
| pla15a    | 600   |
| pla15b    | 600   |
| Melndoxam | 1000  |
| pla14b    | 1000  |
| pla14a    | 1500  |
| Indoxam   | >1600 |
| pla15c    | >1600 |
| pla16a    | >1600 |
| pla16b    | >1600 |
| pla16c    | >1600 |

| Ligand    | Vina |
|-----------|------|
| pla12f    | 9.76 |
| Melndoxam | 9.76 |
| pla16c    | 9.7  |
| pla16b    | 9.58 |
| pla13i    | 9.56 |
| Indoxam   | 9.54 |
| pla16e    | 9.5  |
| pla11h    | 9.48 |
| pla13h    | 9.46 |
| pla14a    | 9.4  |
| pla11g    | 9.32 |
| pla14b    | 9.3  |
| pla13a    | 9.28 |
| pla12b    | 9.2  |
| pla12e    | 9.18 |
| pla13g    | 8.94 |
| pla15a    | 8.78 |
| pla13e    | 8.76 |
| pla13e    | 8.74 |
| pla13c    | 8.66 |
| pla11d    | 8.64 |
| pla16a    | 8.6  |
| pla13d    | 8.6  |
| pla12a    | 8.58 |
| pla11c    | 8.52 |
| LY315920  | 8.5  |
| pla15b    | 8.46 |
| pla13f    | 8.46 |
| pla13b    | 8.42 |

| Ligand    | PAIRcsd  |
|-----------|----------|
| pla11g    | 286717   |
| pla13a    | 271493.8 |
| pla14b    | 264121.2 |
| pla15c    | 261170   |
| pla11h    | 259109.6 |
| pla16b    | 251960.2 |
| pla15a    | 249998.8 |
| pla13i    | 249982   |
| pla15b    | 247386.8 |
| pla12f    | 246901.8 |
| Indoxam   | 246242.8 |
| pla16c    | 245047.4 |
| pla12e    | 237882.2 |
| pla14a    | 236534.6 |
| pla13f    | 235833.4 |
| pla13h    | 235445   |
| Melndoxam | 235407.2 |
| pla13g    | 235123.2 |
| pla13d    | 233941.6 |
| pla12b    | 233622   |
| pla16a    | 232685.6 |
| pla13b    | 227002.2 |
| pla13e    | 226496.4 |
| pla12a    | 221068.4 |
| pla11d    | 220980   |
| pla13c    | 218856.2 |
| pla11c    | 216338.8 |
| LY315920  | 210683.6 |

| Ligand    | Pscsd    |
|-----------|----------|
| pla11g    | 334872.6 |
| pla13a    | 315074.4 |
| pla11h    | 306799.4 |
| pla14b    | 304199.6 |
| pla15c    | 301478.6 |
| pla15a    | 296171.2 |
| pla16b    | 289669.2 |
| pla12f    | 286539   |
| pla15b    | 284687   |
| pla13i    | 284629.8 |
| Indoxam   | 281938.6 |
| pla16c    | 280395.2 |
| pla13f    | 278322.8 |
| pla12e    | 275074.2 |
| Melndoxam | 273022.2 |
| pla14a    | 272477.6 |
| pla13d    | 272354   |
| pla16a    | 271439.8 |
| pla13g    | 267550.6 |
| pla13h    | 267262.8 |
| pla12b    | 265487.2 |
| pla13e    | 264518.6 |
| pla13b    | 259433.2 |
| pla11d    | 255866.8 |
| pla13c    | 254684.2 |
| pla11c    | 252548   |
| pla12a    | 252461   |
| LY315920  | 243590.8 |

| Ligand    | PAIRpdb  |
|-----------|----------|
| pla11g    | 658648.4 |
| pla11h    | 640323.4 |
| pla13a    | 621694.8 |
| pla13i    | 599851.8 |
| pla16a    | 597951.4 |
| pla15c    | 589508.4 |
| Indoxam   | 574771.8 |
| pla13h    | 572644.6 |
| pla16b    | 572144   |
| pla12f    | 566126   |
| Melndoxam | 562626.6 |
| pla13g    | 558987.4 |
| pla15a    | 557246.8 |
| pla13f    | 551228   |
| pla14b    | 547337.2 |
| pla15b    | 545662   |
| pla13e    | 538701   |
| pla12e    | 537866.8 |
| pla13c    | 537714.8 |
| pla11d    | 535067.4 |
| pla12b    | 531991.6 |
| pla16c    | 531774.8 |
| pla12a    | 529445.6 |
| pla13b    | 527594   |
| pla11c    | 520357   |
| pla14a    | 519831   |
| pla13d    | 517350.2 |
| LY315920  | 498178.6 |

| Ligand    | Pspdb    |
|-----------|----------|
| pla11g    | 706804.2 |
| pla11h    | 688412.4 |
| pla13a    | 665275.4 |
| pla13i    | 634499.6 |
| pla16a    | 631666.2 |
| pla15c    | 630716.8 |
| Indoxam   | 613487.8 |
| pla16b    | 610917   |
| pla12f    | 605763.4 |
| pla13h    | 603775.4 |
| pla15a    | 603419.6 |
| Melndoxam | 600241.4 |
| pla13f    | 593717.8 |
| pla13g    | 591415   |
| pla14b    | 587415.8 |
| pla15b    | 583276.4 |
| pla13e    | 575634.4 |
| pla12e    | 575318.6 |
| pla13c    | 572048.6 |
| pla11d    | 572032.2 |
| pla16c    | 565798.4 |
| pla12b    | 563857   |
| pla13b    | 562379   |
| pla12a    | 560838.2 |
| pla13d    | 555762.2 |
| pla14a    | 555081.6 |
| pla11c    | 553645.6 |
| LY315920  | 531085.8 |

| Ligand    | SURF     | Ligand    | HPScore | Ligand    | HMScore | Ligand    | HSScore | Ligand    | AVEScore |
|-----------|----------|-----------|---------|-----------|---------|-----------|---------|-----------|----------|
| pla11g    | 49895.84 | pla11g    | 7.624   | pla16a    | 8.684   | pla15c    | 7.162   | pla16a    | 7.778    |
| pla11h    | 48088.86 | pla11h    | 7.598   | pla11h    | 8.624   | pla11g    | 7.154   | pla11h    | 7.752    |
| pla15a    | 46172.66 | pla16a    | 7.53    | pla12f    | 8.512   | pla16a    | 7.122   | Melndoxam | 7.65     |
| pla13a    | 43560.62 | pla15c    | 7.46    | Melndoxam | 8.51    | pla16b    | 7.118   | pla15c    | 7.646    |
| pla13f    | 42489.58 | Melndoxam | 7.404   | pla11g    | 8.482   | pla13a    | 7.116   | Indoxam   | 7.576    |
| pla15c    | 41208.52 | Indoxam   | 7.366   | Indoxam   | 8.394   | pla11h    | 7.1     | pla16b    | 7.574    |
| pla12f    | 41000.7  | pla16b    | 7.352   | pla13a    | 8.324   | Melndoxam | 7.036   | pla12f    | 7.566    |
| Melndoxam | 40594.92 | pla12f    | 7.254   | pla15c    | 8.316   | Indoxam   | 6.99    | pla11g    | 7.552    |
| pla14b    | 40078.62 | pla16c    | 7.23    | pla16c    | 8.3     | pla13i    | 6.928   | pla16c    | 7.42     |
| pla13i    | 40049.38 | pla13i    | 7.18    | pla13c    | 8.284   | pla12f    | 6.924   | pla13a    | 7.38     |
| pla16b    | 39503.82 | pla13a    | 7.154   | pla16b    | 8.272   | pla16c    | 6.866   | pla13i    | 7.378    |
| pla13h    | 38871.86 | pla13h    | 7.092   | pla13e    | 8.262   | pla13h    | 6.832   | pla14b    | 7.36     |
| pla16a    | 38754.3  | pla14b    | 7.062   | pla14b    | 8.26    | pla14b    | 6.77    | pla13h    | 7.292    |
| Indoxam   | 38716.24 | pla12e    | 6.99    | pla13d    | 8.228   | pla12e    | 6.758   | pla12e    | 7.29     |
| pla13d    | 38412.16 | pla13f    | 6.942   | pla13b    | 8.21    | pla12a    | 6.738   | pla13e    | 7.272    |
| pla13c    | 38071.4  | pla13e    | 6.914   | pla12e    | 8.12    | pla12b    | 6.716   | pla13c    | 7.254    |
| pla12e    | 38039.88 | pla13c    | 6.904   | pla15a    | 8.112   | pla15a    | 6.696   | pla13d    | 7.204    |
| pla13e    | 38022    | pla15b    | 6.896   | pla13f    | 8.092   | pla13f    | 6.668   | pla12b    | 7.154    |
| pla15b    | 37614.76 | pla11d    | 6.896   | pla13g    | 8.084   | pla13e    | 6.638   | pla13f    | 7.148    |
| pla11d    | 36964.84 | pla12b    | 6.878   | pla13i    | 8.03    | pla14a    | 6.58    | pla11d    | 7.132    |
| pla12a    | 36694.16 | pla15a    | 6.838   | pla11d    | 7.998   | pla13c    | 6.566   | pla13b    | 7.132    |
| pla14a    | 36821.6  | pla13d    | 6.828   | pla13h    | 7.954   | pla13d    | 6.56    | pla15a    | 7.102    |
| pla13b    | 36725.76 | pla12a    | 6.806   | pla12b    | 7.866   | pla11d    | 6.53    | pla12a    | 7.08     |
| pla11c    | 36209.08 | pla13g    | 6.798   | pla14a    | 7.794   | pla15b    | 6.512   | pla15b    | 7.046    |
| pla12b    | 35814.22 | pla14a    | 6.784   | pla11c    | 7.764   | pla13b    | 6.518   | pla14a    | 7.038    |
| pla16c    | 35347.74 | pla11c    | 6.7     | pla15b    | 7.726   | LY315920  | 6.484   | pla13g    | 7.016    |
| LY315920  | 34930.88 | pla13b    | 6.686   | pla12a    | 7.696   | pla11c    | 6.414   | pla11c    | 6.888    |
| pla13g    | 32427.5  | LY315920  | 6.604   | LY315920  | 7.424   | pla13g    | 6.274   | LY315920  | 6.836    |

Table 8. IC50 et scores des inhibiteurs connus pour toutes les fonctions de score

## F. BILAN

L'analyse des poses et des scores obtenus lors du criblage d'une base d'inhibiteurs connus montre de façon évidente que les outils employés ne sont pas adaptés au système étudié. Bien que plusieurs composés aient été correctement positionnés dans le site actif, d'autres, notamment parmi les composés les plus actifs, ne présentent aucune pose similaire aux modèles fournis par les structures cristallographiques holo d'autres sPLA2 ou la reconstruction d'une structure d'hGX complexée par le Me-Indoxam. Pour les composés les plus petits, les poses adoptées sont rarement en contact avec le calcium qui se trouve dans une partie plutôt externe du site actif, laquelle offre un nombre réduit d'interaction en comparaison avec des zones plus enfouies de la cavité. Pour les composés encombrés stériquement, les interactions créées par la coordination de l'ion métallique ont un poids relatif trop faible et les poses qui les proposent sont supplantées par des poses plus favorables en terme de complémentarité de forme. Seules les poses obtenues pour les ligands de taille moyenne montrent une coordination complète (deux oxygènes) ou partielle (un oxygène) du calcium, mais même dans ce cas, il ne s'agit pas nécessairement des poses les mieux classées. En effet, Vina ne semble pas considérer la coordination du calcium comme un critère déterminant lors de la génération des poses. Cet échec partiel du docking explique au moins en partie les classements obtenus qui ne rendent pas compte de l'affinité des molécules pour la protéine. Dans ce cas, les multiples fonctions de scoring utilisées ou un consensus scoring ne permettent pas de corriger le problème, car ce n'est pas le calcul des interactions qui est en cause, mais bien les modes de liaison prédits par le programme de docking.

De la même façon, il est difficile de savoir si les ligands innovants de la base fournie par l'équipe de Gilles Guichard sont correctement positionnés dans le site, mais la comparaison avec de précédents travaux et les résultats peu concluants concernant les positions adoptées par les inhibiteurs connus ainsi que leur classement mettent en doute la capacité des outils de docking utilisés à produire des résultats corrects sur ce système. Toutefois, les molécules nouvellement synthétisées étant très différentes de celles du training set de ligands connus, il peut sembler prématuré de conclure sur la non-validité des poses et des classements obtenus. En l'absence de données expérimentales concernant des composés de la même famille, il est donc difficile de sélectionner, à partir des résultats montrés ici, une liste de ligands plus fortement susceptibles de se lier à hGX que les autres, ce qui reste l'objectif ultime de ce projet.

Lors de ce projet, nous avons donc constaté les limites du programme Vina lors de l'évaluation d'un système faisant intervenir un ion métallique. Bien que celui-ci apparaisse dans le site actif, il n'est pas géré de façon correcte et les poses générées ne prennent pas suffisamment en compte la nécessité d'une coordination. Les récepteurs contenant du métal sont de façon notoire assez difficilement traités en général par les programmes de modélisation moléculaire [Moitessier *et al.*, 2008]. La continuation de ce projet doit donc maintenant passer par l'expérimentation d'autres outils de docking ou d'autres stratégies. En particulier, de précédents criblages menés à l'aide du programme Gold ont montré plus de succès dans le traitement de cette protéine [Muller *et al.*, 2006] et ont servi de base à ce projet. Gold, qui semble donc adapté à l'étude de ce système, sera prochainement utilisé pour la suite de ce projet qui sera mené par Marie-France Bakaï dans le cadre de son stage de fin de Master. Sachant qu'en l'absence de ligand, la complexation du calcium est assurée par deux molécules d'eau, en réintroduire une dans le site actif pourrait également fournir de précieux indices sur les possibilités de liaisons hydrogène entre le ligand et la protéine et constitue donc une piste pour une nouvelle étude.



## **CHAPITRE 4 : APPLICATION À LA PONTINE**



En collaboration avec Patrick Lestienne et Jean Rosenbaum, de l'équipe Physiopathologie du cancer du foie, INSERM U1053, pour la partie expérimentation biochimique.

Avec la participation d'un autre membre de l'équipe, Marc Lamblin, ainsi que de François-Xavier Felpin de l'équipe Wine, Imagery, Medicinal & Organometallic Chemistry de l'ISM/CNRS UMR 5255 pour les bases de ligands notées ML\* et FX\*. Avec également la participation de Gilles Guichard et Claire Venin de l'équipe Peptidomimetics chemistry, IECB/CNRS UMR 5248 pour la base de ligands G\*.

Les résultats obtenus dans cette partie ont fait l'objet d'une publication acceptée chez *Biochemical Journal* : « First identification of small molecule inhibitors of Pontin by combining virtual screening and enzymatic assay » (voir l'annexe 5).

## A. GÉNÉRALITÉS

Les ATPases, ou plus généralement les NTPases, sont des protéines qui catalysent l'hydrolyse des nucléotides depuis leur forme tri-phosphate pour relarguer un nucléotide diphosphate et un ion phosphate. Cette réaction, qui se trouve être productrice d'énergie, est nécessaire à l'accomplissement de très nombreux processus biologiques, en particulier lorsque ceux-ci impliquent des complexes macromoléculaires (aussi appelés machines moléculaires) dans lesquels plusieurs protéines interagissent, s'assemblent, se réarrangent, le plus souvent de manière énergie-dépendante [Neuwald *et al.*, 1999]. Les processus majeurs qui ont lieu dans la cellule impliquent le plus souvent des complexes de plus de dix protéines, qui comprennent des parties mobiles fortement coordonnées dont les réarrangements sont déclenchés par l'hydrolyse de nucléotides [Alberts, 1998]. Les NTPases sont donc des protéines ubiquitaires regroupant un grand nombre de familles de fonctions et de structures diverses. Parmi celles-ci, les plus communément rencontrées sont les ATPases de type P-loop ou Walker, caractérisées par leurs motifs Walker A (ou P-loop) et Walker B [Ogura & Wilkinson, 2001] qui interviennent respectivement dans la liaison et l'hydrolyse de l'ATP [Ammelburg *et al.*, 2006].

Les protéines AAA+ (pour ATPases Associated with diverse cellular Activities ou ATPases associées à diverses activités cellulaires), constituent une sous-famille de ces ATPases de type Walker. On les retrouve chez les eucaryotes, les procaryotes et les archéobactéries, ce qui est révélateur de leur ancienneté et de leur importance pour toutes les formes de vie [Vale, 2000]. En plus des motifs Walker A et Walker B, ces protéines présentent plusieurs résidus fortement conservés reliés à l'activité ATPase qui sont l'Arg

finger, et les sensors 1 et 2 [Lupas & Martin, 2002]. Elles forment un grand groupe de fonctions diverses parmi lesquelles on trouve des protéines chaperonnes, des protéases, et des enzymes spécialisées dans le traitement des acides nucléiques telles que les hélicases [Ammelburg *et al.*, 2006]. Elles ont en commun leur capacité à former des complexes en anneau, le plus souvent hexamériques [Ogura *et al.*, 2004] ainsi que leurs fonctions dans l'assemblage, la dissociation et le réarrangement ATP-dépendant de différentes machines moléculaires [Lupas & Martin, 2002], y compris de complexes impliquant l'ADN [Neuwald *et al.*, 1999].

Parmi ces ATPases, la Pontine et son homologue la Reptine, présentes chez les eucaryotes, montrent une homologie de séquence avec RuvB, une protéine bactérienne de la famille AAA+ qui a pour rôle de catalyser la migration des jonctions de Holliday au cours de la recombinaison homologue et intervient dans la réparation de l'ADN endommagé. RuvB est notamment connue pour ses propriétés ATPase dépendante de l'ADN et hélicase [Makino *et al.*, 1999; Matias *et al.*, 2006]. Pontine et Reptine partagent avec RuvB deux régions conservées où sont inclus les motifs Walker A et Walker B caractéristiques des ATPases. En outre, elles possèdent également les sensors 1 et 2 ainsi que l'Arg finger typiques des protéines de type AAA+. Le rôle des sensors est de différencier l'ADP et l'ATP grâce à des liaisons hydrogène avec les  $\beta$  et  $\gamma$ -phosphate de ces molécules [Matias *et al.*, 2006], tandis que celui de l'Arg finger est de favoriser l'hydrolyse en formant des interactions avec le  $\gamma$ -phosphate pendant l'état de transition de la réaction [Ogura *et al.*, 2004]. La Pontine humaine possède pour ces deux régions une identité et une similarité de séquence de 38% et 25% et de 55% et 46% respectivement avec la RuvB de *Thermus thermophilus*, l'organisme pour lequel cette homologie est la plus importante [Makino *et al.*, 1998] (Figure 29 page suivante).

Entre 1997 et 2000, la Pontine et la Reptine ont été identifiées de manière isolée par plusieurs équipes dans de multiples complexes et processus cellulaires. Leur découverte simultanée, au sein de complexes variés et par des équipes indépendantes, est révélatrice de la grande diversité des fonctions qu'elles peuvent exercer et est à l'origine des nombreuses appellations qui leur ont été attribuées. Au cours de ce travail, les noms Pontine et Reptine seront utilisés systématiquement, mais pour plus de clarté, le tableau ci-dessous répertorie les différentes appellations de ces deux protéines dans la littérature et les publications associées (Table 9 page suivante).

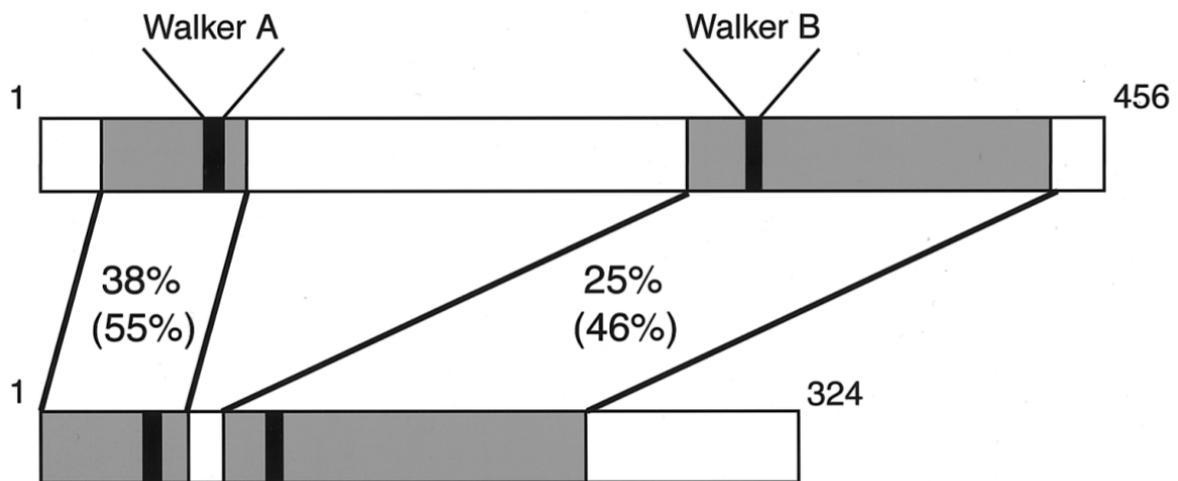


Figure 29. Alignement de séquence entre Pontine (en haut) et RuvB (en bas) mettant en évidence les deux régions conservées [Makino et al., 1998]

| Référence               | Pontine        | Reptine       | Signification                                                                    |
|-------------------------|----------------|---------------|----------------------------------------------------------------------------------|
| [Kanemaki et al., 1997] | TIP49          |               | TBP (TATA-binding protein) interacting protein                                   |
| [Qiu et al., 1998]      | RUVBL1         | RuvBL2        | RuvB-like                                                                        |
| [Bauer et al., 1998]    | Pontin52       |               | Référence à sa capacité à ponter l'interaction entre TBP et la $\beta$ -caténine |
| [Makino et al., 1999]   | TIP49a         |               | TBP (TATA-binding protein) interacting protein                                   |
| [Kanemaki et al., 1999] |                | TIP49b        | TBP (TATA-binding protein) interacting protein                                   |
| [Wood et al., 2000]     |                | TIP48         | TBP (TATA-binding protein) interacting protein                                   |
| [Shen et al., 2000]     | Rvb1           | Rvb2          | RuvB-like                                                                        |
| [Ikura et al., 2000]    | TAP54 $\alpha$ | TAP54 $\beta$ | TIP60-associated protein                                                         |
| [Bauer et al., 2000]    |                | Reptin52      | Repressing Pontin52                                                              |

Table 9. Noms donnés à Pontine et Reptine et références associées

## B. RÔLES PHYSIOLOGIQUES ET PATHOLOGIQUES

Comme de nombreuses protéines de la famille des ATPases de type AAA+, la Pontine et la Reptine sont des protéines ubiquitaires impliquées dans de nombreuses fonctions cellulaires. Leur forte conservation au cours de l'évolution, de la levure à l'homme, suggère un rôle dans certains processus majeurs [Bauer *et al.*, 1998]. *In vivo*, les drosophiles déficientes en *dpon* ou *drep* (les gènes orthologues de Pontine et Reptine) montrent une mortalité au premier stade larvaire, indiquant que ces deux gènes ont des fonctions essentielles et non redondantes dès les premiers stades du développement [Bauer *et al.*, 2000]. En plus de *Drosophila melanogaster*, la Pontine et la Reptine sont également indispensables pour la viabilité de la levure *Saccharomyces cerevisiae* [Lim *et al.*, 2000; Qiu *et al.*, 1998] et du nématode *Caenorhabditis elegans* [Matias *et al.*, 2006]. Chez la levure, la suppression de Pontine ou de Reptine provoque l'arrêt de la croissance à toutes les étapes du cycle cellulaire, tout comme l'introduction, dans les motifs Walker A et Walker B, de mutations empêchant la liaison ou l'hydrolyse de l'ATP [Jónsson *et al.*, 2001].

Malgré leur découverte il y a plus d'une dizaine d'années, et l'identification de plusieurs de leurs partenaires, les implications de ces deux protéines dans des processus pathologiques ainsi que leurs fonctions physiologiques restent encore à clarifier. Actuellement, les données disponibles dans la littérature indiquent que la Pontine et la Reptine sont des protéines nucléaires [Bauer *et al.*, 1998; Makino *et al.*, 1999] qui interviennent dans plusieurs complexes liés à la transcription [Gallant, 2007] ou au remodelage de la chromatine [Jónsson *et al.*, 2001; Shen *et al.*, 2000] et pourraient exercer une activité hélicase [Makino *et al.*, 1999]. En outre, elles ont récemment été identifiées comme des cibles dans la thérapie contre le cancer et sont sur-exprimées dans plusieurs formes de tumeurs [Grigoletto *et al.*, 2011]. Ces deux protéines sont capables de s'associer dans des complexes *in vitro* et *in vivo*, lesquels présentent une activité ATPase accrue en comparaison des protéines isolées [Ikura *et al.*, 2000; Kanemaki *et al.*, 1999]. On les retrouve le plus souvent sous la forme d'hexamères homo- ou hétérophiles, mais également sous la forme de dodécamères qui sont en fait des double hexamères [Jha & Dutta, 2009]. L'existence de ces différentes formes sera discutée dans le paragraphe « Complexes formés par Pontine et Reptine ».

### I. INTERACTIONS AVEC DES FACTEURS DE TRANSCRIPTION

En 1997, Kanemaki *et al.* [Kanemaki *et al.*, 1997] identifient la Pontine chez le rat comme partenaire du facteur de transcription TBP (TATA-binding protein), dont la fonction

consiste à activer la formation du complexe PIC (PreInitiation Complex) permettant d'initier la transcription. Ils la nomment TIP49 pour TBP-Interacting Protein 49.

Qiu et al. [Qiu *et al.*, 1998] découvrent de leur côté la Pontine humaine dans le complexe holoenzyme de l'ARN polymérase II, qui intervient également dans l'activation de la transcription, et la baptisent RuvBL1 pour RuvB-like 1, à cause de son homologie avec la protéine RuvB bactérienne.

Plus tard, Bauer et al. [Bauer *et al.*, 1998] identifient un nouveau partenaire inconnu de la  $\beta$ -caténine, impliquée notamment dans la voie de signalisation des protéines Wnt par l'intermédiaire d'un complexe formé avec les facteurs de transcription LEF-1/TCF (Lymphoid enhancer-binding factor 1/T-cell factor). Cette voie de signalisation joue un rôle dans plusieurs processus du développement chez les animaux comme la différenciation, l'adhésion et la prolifération cellulaire [Bienz & Clevers, 2000; Peifer & Polakis, 2000] et ses dysfonctionnements peuvent être reliés à l'oncogenèse ainsi qu'à différentes maladies dégénératives [Morin, 1999; Nusse, 2005; Polakis, 2000]. La nouvelle protéine identifiée possède une homologie de séquence très forte avec la protéine nommée TIP49 chez le rat et présentée quelques mois avant par Kanemaki et al. [Kanemaki *et al.*, 1997]. Ils la nomment Pontin52 (Pontine) en raison de sa capacité potentielle à se lier à la fois (« ponter ») à la  $\beta$ -caténine et au facteur de transcription TBP. Leurs résultats montrent effectivement une interaction directe entre la  $\beta$ -caténine et la Pontine, ainsi qu'entre la TBP et la Pontine, et l'existence d'un complexe triple formé de la  $\beta$ -caténine, de la Pontine et du facteur de transcription LEF-1 [Bauer *et al.*, 2000; Bauer *et al.*, 1998]. Plus tard, cette même équipe identifie une autre protéine qui possède les mêmes partenaires mais une action antagoniste sur les activités Wnt liées à la  $\beta$ -caténine et lui donnent le nom de Reptin52 (Reptine) pour repressing Pontin52 [Bauer *et al.*, 2000].

Pontine s'avère être un activateur de la transcription des gènes dépendants du TCF et de la transformation néoplasique par le biais de la  $\beta$ -caténine. L'activité oncogénique de la  $\beta$ -caténine semble fortement liée à son rôle dans la transcription [Kolligs *et al.*, 1999]. L'utilisation du mutant de Pontine D302N dépourvu d'activité ATPase a permis de mettre en évidence une inhibition de la transformation néoplasique induite par la  $\beta$ -caténine dans des lignées de cellules épithéliales immortalisées de rat. Ce même mutant a également induit une diminution de la croissance tumorale de lignées de cancer du côlon présentant une dérégulation de la  $\beta$ -caténine. En présence d'un tel mutant, l'expression de gènes dépendants du TCF régulés par la  $\beta$ -caténine est inhibée, et cette inhibition s'accompagne de la diminution de l'acétylation des histones à proximité des sites de liaison du TCF. Ces résultats semblent indiquer que la Pontine régule de l'expression des gènes TCF via la  $\beta$ -caténine à travers des modifications de la chromatine [Feng *et al.*, 2003].

La Pontine et la Reptine sont également des partenaires du facteur de transcription c-Myc [Etard *et al.*, 2005; Li *et al.*, 2010; Wood *et al.*, 2000], qui fait partie des oncogènes les plus largement représentés dans les cancers humains [Cole, 1986]. Elles sont en effet capables d'interagir avec la partie M<sub>h</sub>II (pour Myc homology box II) de c-Myc, un domaine fortement conservé de l'extrémité N-terminale du facteur de transcription qui s'avère indispensable pour l'ensemble de ses fonctions biologiques et en particulier la transformation oncogénique et l'induction de l'apoptose. En présence du mutant D302N de la Pontine dépourvu d'activité ATPase, la totalité de l'action oncogénique de c-Myc est inhibée de manière spécifique, tandis que la croissance cellulaire normale n'est pas perturbée. La Pontine semble donc un co-facteur essentiel pour la transformation oncogénique induite par c-Myc [Wood *et al.*, 2000].

La Pontine interagit aussi avec le facteur de transcription E2F1 dont les activités sont assez similaires à celles de c-Myc et qui fonctionne également comme oncogène et inducteur d'apoptose. La sur-expression d'un mutant de la Pontine incapable de réaliser l'hydrolyse de l'ATP permet d'augmenter considérablement le déclenchement de l'apoptose induite par E2F1 [Dugan *et al.*, 2002].

## II. COMPLEXES DE REMODELAGE DE LA CHROMATINE

De façon plus spécifique, la Pontine et la Reptine sont également impliquées dans la transcription via des interactions avec différents complexes de remodelage de la chromatine. La présence de Pontine ou de Reptine au sein de tels complexes, qu'ils impliquent des facteurs de remodelage de la chromatine ATP-dépendant ou des enzymes de modification des histones, ont été d'ailleurs été confirmées à plusieurs reprises [Jha & Dutta, 2009].

La Pontine et la Reptine sont présentes dans le complexe INO80, chez la levure comme chez l'homme, avec une stœchiométrie de six pour un en comparaison des autres partenaires impliqués (en accord avec leur capacité à former des hexamères). Ce complexe comporte au moins douze sous-unités dont l'Actine, ainsi que trois ARP (Actin Related Proteins) et intervient dans la réparation de l'ADN endommagé [Jónsson *et al.*, 2004; Shen *et al.*, 2000]. Il est notamment impliqué dans le déplacement des nucléosomes le long du brin d'ADN [Jin *et al.*, 2005]. Chez la levure, l'absence de Pontine et Reptine affecte la transcription de plus de 5% des gènes actifs appartenant à différentes classes et processus cellulaires, avec un effet notablement plus important lors de la suppression de la Reptine [Jónsson *et al.*, 2001]. Cette extinction abolit également l'activité de remodelage de la chromatine du complexe INO80 [Jónsson *et al.*, 2004]. Cette forme inactive est dépourvue de l'ARP5p, qui joue un rôle essentiel dans l'activité de remodelage du complexe [Shen *et*

*al.*, 2003], indiquant que la Pontine et la Reptine sont nécessaires à l'assemblage d'un complexe INO80 fonctionnel [Jónsson *et al.*, 2004]. Une étude a également montré que le knock-down de la Reptine ou de INO80 entraîne une augmentation de la sensibilité aux dommages à l'ADN, et que cette augmentation reste la même lors du knock-down simultané des deux protéines. Ceci suggère qu'elles sont toutes les deux impliquées dans le même mécanisme cellulaire de réponse aux dommages à l'ADN [Wu *et al.*, 2007].

La Pontine et la Reptine, ainsi que plusieurs autres composants du complexe INO80 sont également présents dans les complexes SWR1 (levure) et SRCAP (mammifères), mais aucune étude précise sur les fonctions des différents partenaires n'a été réalisée à notre connaissance [Choi *et al.*, 2009; Mizuguchi *et al.*, 2004; Wu *et al.*, 2005].

La Pontine et la Reptine sont également requises dans le complexe TIP60 chez l'homme. Ce complexe constitué de 14 sous-unités distinctes montre une activité acétylase sur l'histone, en particulier dans les processus de réparation de l'ADN et d'apoptose, dépendante de la présence de Pontine et Reptine [Feng *et al.*, 2003; Ikura *et al.*, 2000; Jha *et al.*, 2008]. De façon surprenante, toutes les deux sont absentes du complexe équivalent chez la levure (NuA4 pour Nucleosomal Acetyltransferase of H4) [Jha & Dutta, 2009].

### III. INTERACTIONS AVEC DES RIBONUCLÉOPROTÉINES

La Pontine et la Reptine interviennent dans la synthèse des snoARN (small nucleolar ARN) et dans la biogenèse des snoRNP (small nucleolar RNP), des complexes formés de protéines dites RNP (ribonucléoprotéine) et de snoARN [Watkins *et al.*, 2002]. Ces complexes ont pour fonction le clivage et les modifications de plusieurs types d'ARN et jouent donc un rôle important dans la réplication de l'ADN. En l'absence de Reptine, ou en présence de la protéine comprenant une mutation dans le domaine Walker A requis pour la liaison à l'ATP, l'accumulation de snoARN et de snoRNP est fortement diminuée chez la levure [King *et al.*, 2001]. Dans des lignées cellulaires humaine, la suppression de la Pontine et de la Reptine entraînent une réduction de la quantité de snoARN [Watkins *et al.*, 2004].

La Pontine et la Reptine ont aussi été trouvée dans le complexe télomérase, qui a pour fonction l'ajout de petites séquences répétitives d'ADN sur les télomères, pour stabiliser et protéger les chromosomes. Dans cette étude, la perte de la Pontine et de la Reptine provoque une diminution très importante de l'activité télomérase (jusqu'à 90%). De plus cette activité télomérase semble dépendante de l'activité ATPase de la Pontine, mais pas de la Reptine. Le complexe télomérase est sur-exprimé dans les cancers humains où il sert de promoteur à la prolifération cellulaire et à la survie des cellules malades grâce au maintien de télomères fonctionnels [Venteicher *et al.*, 2008].

#### IV. ACTIVITÉ HÉLICASE

Les hélicases, qui sont elles aussi caractérisées par la présence des motifs Walker A et B, utilisent l'énergie produite par l'hydrolyse de l'ATP pour désappairier l'ADN double brin. Ce sont donc des acteurs essentiels des processus cellulaires impliquant l'ADN sous sa forme simple brin et elles sont de fait impliquées dans tous les processus de réplication, réparation et de transcription de l'ADN, ainsi que dans la synthèse des ribosomes et le remodelage de la chromatine [Makino *et al.*, 1999; Tsaneva *et al.*, 1993]. Leur homologie avec la protéine RuvB, suggère que la Pontine et la Reptine présentent également une activité hélicase et pourraient donc avoir un rôle dans l'initiation de la transcription [Bauer *et al.*, 1998].

L'étude de Pontine à partir d'extraits cellulaires de foie de rat réalisée par Makino *et al.* montre une activité ATPase intrinsèque et une stimulation de cette activité jusqu'à dix fois en présence d'ADN simple brin. Elle met également en évidence une activité hélicase avec un déroulement du double brin d'ADN dans le sens 3' vers 5' (opposé à celui observé chez RuvB) [Kanemaki *et al.*, 1997; Makino *et al.*, 1999]. Des résultats similaires ont été obtenus dans la même équipe à partir de la Reptine humaine, mis à part le sens de désappariement du brin qui se fait cette fois de 5' vers 3' [Kanemaki *et al.*, 1999]. Pour Gribun *et al.*, la Pontine et la Reptine de la levure possèdent une activité ATPase modérée et non stimulée par la présence d'ADN lorsqu'elles sont seules, alors que sous la forme d'un hétéro-hexamère elles présentent une activité ATPase forte et une augmentation de celle-ci en présence d'ADN. Dans la même étude, ils montrent également une activité hélicase faible pour la Pontine et la Reptine et forte pour le complexe Pontine/Reptine, préférentiellement dans le sens 5' vers 3' [Gribun *et al.*, 2008].

A l'inverse, aucune activité hélicase ni ATPase n'a pu être détectée par Qiu *et al.* lors de l'étude de la Pontine humaine exprimée à partir de baculovirus d'insecte et ce même en présence de différents types d'acides nucléiques [Qiu *et al.*, 1998]. Selon Ikura *et al.*, la Pontine et la Reptine humaine exprimées à partir d'*E. Coli* ne présentent pas non plus d'activité hélicase. La purification de la Pontine seule par voie chromatographique, montre que celle-ci présente un profil d'élution sous la forme d'un monomère et une activité ATPase nulle, tandis que Reptine se présente sous une forme dimérique ou trimérique d'activité ATPase faible. Ensemble, elles forment un complexe d'activité ATPase plus élevée mais sans activité hélicase [Ikura *et al.*, 2000]. Chez Matias *et al.*, l'hexamère de Pontine humaine seule montre une activité ATPase extrêmement faible et non-stimulée par la présence d'acides nucléiques, et pas d'activité hélicase [Matias *et al.*, 2006]. Puri *et al.* rapportent l'existence d'un dodécamère constitué d'un mélange équimolaire de Pontine et Reptine humaine pourvu d'une activité ATPase non dépendante de l'ADN et dépourvu d'activité

hélicase [Puri *et al.*, 2007].

Dans les études réalisées par Jonsson *et al.*, aucune activité hélicase n'a pu être détectée dans le complexe de remodelage de la chromatine formé par le dodécamère de Pontine et Reptine avec INO80, bien qu'il présente une activité ATPase stimulée en présence d'ADN [Jónsson *et al.*, 2001; Jónsson *et al.*, 2004]. Au contraire, ce même complexe présente une activité hélicase (dans le sens 3' vers 5') selon Shen *et al.* [Shen *et al.*, 2000].

Enfin, plusieurs études ont montré une augmentation de l'activité ATPase en présence de différents oligonucléotides [Jónsson *et al.*, 2001; Kanemaki *et al.*, 1999; Makino *et al.*, 1999], un résultat qui appuie l'hypothèse d'une interaction entre ces protéines et l'ADN, et donc d'une possible activité hélicase. A l'inverse, certaines équipes n'ont pas pu détecter de stimulation de ce type lors d'expériences similaires [Matias *et al.*, 2006; Puri *et al.*, 2007; Qiu *et al.*, 1998; Torreira *et al.*, 2008]. Les résultats disponibles mettent donc en évidence une controverse à propos de l'activité hélicase de ces deux protéines, qu'elles soient testées seules, ensemble, sous forme d'oligomères ou au sein de complexes de plus grande échelle.

## V. RÔLE SPÉCIFIQUE DANS LE CANCER

Les études menées au sein de l'équipe de Jean Rosembaum sur la Pontine et la Reptine ont conclu à un rôle majeur de ces deux protéines dans plusieurs types de cancers (voir les revues [Grigoletto *et al.*, 2011] et [Huber *et al.*, 2008]). Ils ont notamment remarqué que les patients atteints de carcinomes hépatocellulaires présentaient des taux de Pontine et de Reptine dans la tumeur plus élevés que dans les tissus sains environnants [Blanc *et al.*, 2005]. Des résultats équivalents ont également été observés au sein de plusieurs autres équipes [Acevedo *et al.*, 2008; Iizuka *et al.*, 2006; Li *et al.*, 2005]. Suite à cela, des études ont été menées *in vitro* qui montrent une diminution de la croissance tumorale lors de l'extinction de la Pontine ou de la Reptine grâce à la transfection de siARN (small interfering ARN) [Haurie *et al.*, 2009; Rousseau *et al.*, 2007]. *In vivo*, l'extinction de la Reptine dans des xenogreffes de carcinomes hépatocellulaires réalisées sur la souris ont également montré un recul de la progression tumorale [Ménard *et al.*, 2010].

Des taux d'expression élevés ont également été rapportés pour la Pontine dans des cas de cancers colorectaux [Carlson *et al.*, 2003; Lauscher *et al.*, 2007], et pour la Reptine dans des cas de cancers du sein [Maslon *et al.*, 2010].

En plus de ces données issues de la littérature, l'exploration de la base Oncomine<sup>TM</sup> (Compendia Bioscience, Ann Harbor, USA) a permis de mettre en évidence un dérèglement de ces protéines dans un certain nombre d'autres cancers humains. On trouve des rapports

indiquant une sur-expression la Pontine et/ou de la Reptine dans les tumeurs solides colorectales [Graudens *et al.*, 2006; Ki *et al.*, 2007], gastriques [Li *et al.*, 2010], de la vessie [Dyrskjøt *et al.*, 2004; Sanchez-Carbayo *et al.*, 2006], les mésothéliomes [Zhan *et al.*, 2007] et les cancers du poumon non à petites cellules [Dehan *et al.*, 2007], ainsi que dans plusieurs types de leucémies aiguës [Andersson *et al.*, 2007] ou chroniques [Haslinger *et al.*, 2004], les myélomes multiples [Zhan *et al.*, 2007], les lymphomes de haut grade [Nishiu *et al.*, 2002] et les lymphomes de Burkitt [Basso *et al.*, 2005].

Ces deux protéines sont donc fréquemment observées à des niveaux d'expression élevés dans les tissus cancéreux, toutefois, les mécanismes qui régissent cette sur-expression ainsi que leur rôle précis dans ces cancers reste encore à élucider. La Pontine et la Reptine sont associées à plusieurs processus cellulaires en relation avec la tumorigenèse, à travers leur implication dans la croissance cellulaire et la transcription, ou dans des complexes liés à la réparation de l'ADN comme TIP60. De nombreux partenaires de la Pontine et de la Reptine cités précédemment ont également été associés aux mécanismes de tumorigenèse. En particulier, elles interagissent avec un certain nombre de facteurs de transcription à fort potentiel carcinogène. La  $\beta$ -caténine, par exemple, est capable d'activer la transcription de plusieurs gènes impliqués dans la progression tumorale, et ses voies de signalisation sont activées dans les cancers colorectaux et les carcinomes hépatocellulaires [de La Coste *et al.*, 1998; Fodde *et al.*, 2001]. c-Myc est également un oncogène bien connu, très fréquemment rencontré dans les cancers humains [Cole, 1986], et dont les effets sont inhibés en présence du mutant D302N de la Pontine sans activité ATPase.

## VI. INTÉRÊT THÉRAPEUTIQUE

Les protéines Pontine et Reptine sont donc capables d'interagir avec divers facteurs de transcription, et leur présence semble indispensable à l'activité biologique normale de plusieurs d'entre eux. De même, ces deux protéines sont présentes dans plusieurs complexes de remodelage de la chromatine qui sont incapables de fonctionner en leur absence. De plus, on trouve parmi leurs partenaires au moins trois oncogènes connus qui sont la  $\beta$ -caténine, c-Myc et E2F1. Ces données indiquent un rôle certain de la Pontine et de la Reptine dans des processus cellulaires majeurs comme la transcription, ainsi qu'un intérêt potentiel dans la thérapie contre le cancer. De récentes publications ont effectivement confirmé que ces deux protéines peuvent constituer des cibles anti-cancer prometteuses [Grigoletto *et al.*, 2011]. Les données concernant plusieurs complexes dans lesquelles ces protéines ont été impliquées, dans diverses fonctions biologiques, montrent un

comportement assez semblable, ce qui laisse supposer que leur mode d'action et leur rôle dans tous ces processus sera probablement très similaire [Huen *et al.*, 2010].

Cependant, le rôle essentiel de ces protéines dans la viabilité cellulaire ainsi que la gamme très large de processus qu'elle touchent sont également un désavantage dans la recherche de molécules inhibitrices, car il est fort probable que l'inhibition de ces protéines soit associée *in vitro* ou *in vivo* à une forte toxicité non compatible avec la conception d'un médicament [Jha & Dutta, 2009].

Ces protéines possèdent des motifs ATPase typiques et probablement une activité ATPase essentielle pour la plupart de leurs fonctions. Ceci est appuyé par les nombreuses études qui ont montré que l'utilisation d'un mutant de l'une ou l'autre de ces protéines dans la région du Walker B permettait d'inhiber les activités liées à ces protéines. Un tel mutant (D302N), dans lequel un acide aspartique fortement conservé du motif Walker B « DExH » est remplacé par une asparagine pour former le motif « NExH », est basé sur des études concernant RuvB dans lesquelles un mutant similaire est incapable d'hydrolyser l'ATP [Mézard *et al.*, 1997]. Plus occasionnellement, des mutations dans le motif Walker A, qui perturbent la liaison à l'ATP, ont également été utilisées, avec les mêmes conséquences [Jónsson *et al.*, 2001; King *et al.*, 2001]. Ces résultats indiquent que le site de liaison à l'ATP de la Pontine et de la Reptine constitue une cible thérapeutique de choix permettant d'inhiber leur activité ATPase et de ce fait la plupart de leurs fonctions biologiques.

## **C. DÉTAILS DU SYSTÈME ÉTUDIÉ**

### **I. ALIGNEMENT DE SÉQUENCE**

La Pontine et la Reptine sont des protéines homologues impliquées dans de nombreux complexes où on les retrouve souvent associées l'une à l'autre. Il existe des différences entre elles, et leur activité peut parfois être distinguée comme indiqué dans la partie précédente, mais de nombreuses études montrent une activité similaire, voire une interdépendance entre ces deux protéines. Elles agissent le plus souvent de façon concertée, et la régulation de l'expression de l'une entraîne des conséquences sur l'expression de l'autre [Radovic *et al.*, 2007; Venteicher *et al.*, 2008].



que le potentiel électrostatique du tunnel semblent indiquer une préférence pour les simples brins d'ADN. La structure contient également six molécules d'ADP co-cristallisées au sein de l'hexamère, soit une par monomère, qui sont situées à l'interface entre deux sous-unités adjacentes dans une cavité suffisamment large pour accueillir l'ATP ou des molécules de plus grande taille [Matias *et al.*, 2006] (Figure 31).

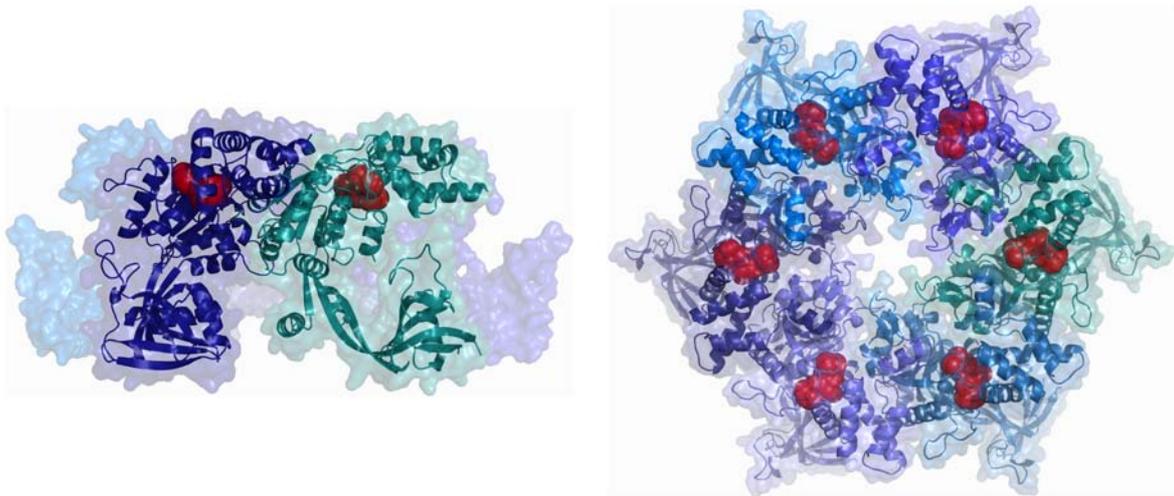
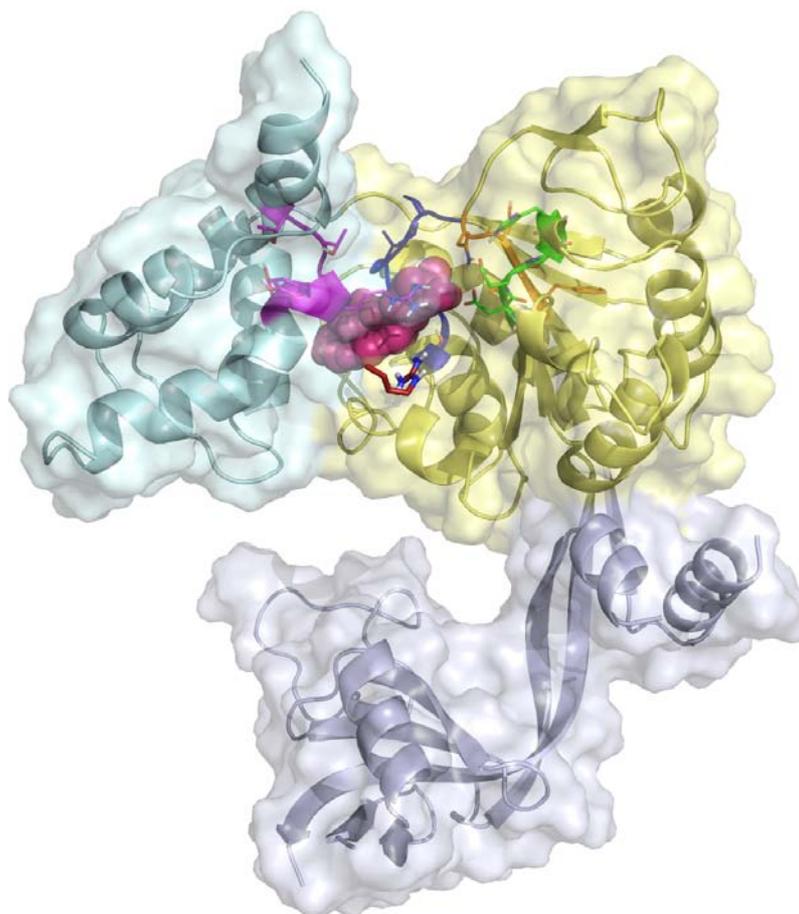


Figure 31. Vue de côté et de dessus de l'hexamère de Pontine cristallisé avec les monomères de Pontine dans des tons de bleus, et l'ADP co-cristallisé en rouge

Chaque sous-unité présente trois domaines distincts. Le domaine I, qui réunit les acides aminés 1 à 120 et 296 à 365 s'apparente au module ATPase commun aux AAA+ ainsi qu'à d'autres protéines proches. Il constitue le cœur de la protéine et contient le site de liaison à l'ATP. Le domaine III, qui court de l'acide aminé 366 jusqu'au C-terminal, est similaire au domaine II de la RuvB et sert à couvrir l'entrée du site de liaison du nucléotide. Le domaine II, un tout nouveau domaine qui n'a pour l'instant été identifié que dans cette protéine, est relié au domaine I par l'intermédiaire de deux brins bêta potentiellement mobiles. Il est formé des acides aminés 121 à 295 et coupe donc le domaine I en deux. Il présente sept brins bêta dont l'arrangement spatial rappelle la structure du domaine de liaison aux oligonucléotides de plusieurs protéines impliquées dans le métabolisme de l'ADN et semble jouer un rôle dans l'interaction de la protéine avec les acides nucléiques. Des expériences menées sur ce domaine seul ont confirmé qu'il était capable d'interagir avec de l'ADN double et simple brin, ainsi qu'avec de l'ARN, sans discrimination en fonction de la séquence nucléotidique des acides nucléiques testés [Matias *et al.*, 2006].

Le motif Walker A (GxxxxGKT) et le Walker B (DExH/N) [Gorbalenya *et al.*, 1989; Koonin, 1993; Walker *et al.*, 1982] se situent dans le domaine I, entre les résidus 70 à 77, et

302 à 305 respectivement, tout comme le sensor 1 qui est constitué des acides aminés 329 à 332. Le sensor 2 se trouve dans le domaine III, et est formé des résidus 401 à 406. Enfin, l'Arg finger intervient dans le site actif à partir de la sous-unité adjacente, et se situe en position 357 [Matias *et al.*, 2006] (Figure 32).



*Figure 32. Structure tridimensionnelle du monomère de Pontine, représentée en cartoon et surface. Le domaine I est en jaune, le domaine II en mauve, le domaine III en cyan. L'ADP co-cristallisé est représenté en sphères de couleur rose. Les motifs Walker A, Walker B, sensor1, sensor2 et Arg finger sont en stick et colorés comme dans l'alignement de séquence*

L'interface entre deux sous-unités est constituée exclusivement de résidus des domaines I et III qui présentent une complémentarité de forme assez importante. Située à l'interface, l'entrée de la cavité est inaccessible en cas d'hexamérisation [Matias *et al.*, 2006]. Cette localisation laisse penser qu'une coopération entre les sous-unités pourrait jouer un rôle dans le mécanisme d'hydrolyse de l'ATP. Cette hypothèse est renforcée par le fait que

l'arginine R357 issue de la seconde sous-unité se place de façon idéale pour jouer le rôle d'Arg finger. Ceci est en accord avec la littérature concernant RuvB, selon laquelle cette arginine, extrêmement conservée parmi les protéines AAA+, crée des interactions avec le  $\gamma$ -phosphate de l'ATP de la sous-unité adjacente qui sont essentielles pour l'hydrolyse de celui-ci [Hishida *et al.*, 2004]. L'arginine R404, du sensor 2, joue elle aussi un rôle majeur dans la liaison de l'ATP, son hydrolyse et dans les interactions entre les sous-unités [Ogura *et al.*, 2004]. Toutefois, l'importance de ces résidus est déduit d'études concernant plusieurs membres de la famille AAA+, et notamment RuvB, mais n'incluant pas la Pontine et la Reptine. A notre connaissance, aucune étude n'a prouvé directement par le biais de mutations le rôle de ces acides aminés particuliers pour ces deux protéines, et seule la mutation du domaine Walker B (D302N) est largement documentée.

### III. COMPLEXES FORMÉS PAR PONTINE ET REPTINE

Comme attendu pour des protéines de cette famille, de nombreuses sources montrent que la Pontine et la Reptine sont capables de former des oligomères et en particulier des hexamères ou des dodécamères, soit en association l'une avec l'autre, soit isolément [Cheung *et al.*, 2010; Gribun *et al.*, 2008; Matias *et al.*, 2006; Puri *et al.*, 2007; Torreira *et al.*, 2008]. Elles présentent également un profil hexamérique ou dodécamérique dans des complexes de plus grande échelle comme le complexe INO80 [Jónsson *et al.*, 2004; Shen *et al.*, 2000].

Testées séparément et à faible concentration (5  $\mu$ M), la Pontine et la Reptine montrent toutes les deux un profil d'éluion sous la forme de monomères, tandis qu'elle hexamérisent spontanément à plus forte concentration (40  $\mu$ M). Testées ensemble, les deux protéines éluent sous la forme d'un complexe hétéro-hexamérique, y compris à faible concentration [Gribun *et al.*, 2008]. Ces résultats sont assez cohérents avec un précédent rapport qui montraient un profil d'éluion monomérique pour la Pontine et dimérique ou trimérique pour la Reptine, tandis que le mélange éluait sous la forme d'un complexe comprenant autant de sous-unités de Pontine que de Reptine [Ikura *et al.*, 2000].

Des études en microscopie électronique des complexes formés par la Reptine et la Pontine chez l'homme et la levure montrent qu'elles forment un dodécamère constitué de deux hexamères circulaires distincts, mais dont la nature homo- ou hétéromérique n'a pas été résolue [Puri *et al.*, 2007; Torreira *et al.*, 2008]. Des hétéro-hexamères simples ont également été rapportés par Gribun *et al.* [Gribun *et al.*, 2008]. La structure cristalline de l'hexamère de Pontine seule a quant à elle été résolue [Matias *et al.*, 2006] et a été décrite ci-dessus. Celle de l'hétéro-dodécamère est sur le point d'être publiée. Elle présente un

complexe symétrique formé de deux hétéro-hexamères circulaires, collés l'un sur l'autre et constitués de sous-unités de Pontine et de Reptine en alternance. La surface de contact entre les deux anneaux est assurée par le domaine II des protéines, et chaque monomère de Pontine ou de Reptine est en contact avec une Pontine et une Reptine de l'autre hexamère. La structure de chaque monomère ne montre pas de différence significative par rapport au monomère de l'hexamère de Pontine seule [Gorynia *et al.*, 2011]. Globalement, le complexe correspond à la structure proposée par Gribun *et al.*, bien que ceux-ci aient rapporté avoir observé uniquement des hexamères simples [Gribun *et al.*, 2008] et est en désaccord avec les complexes asymétriques observés par Puri *et al.*, et Torreira *et al.* [Puri *et al.*, 2007; Torreira *et al.*, 2008], bien que la nature homomérique des hexamères constitutifs de ceux-ci n'ait pas été établie. Il faut également noter que les complexes observés proviennent de deux espèces différentes, la levure [Gribun *et al.*, 2008; Torreira *et al.*, 2008] et l'homme [Gorynia *et al.*, 2011; Matias *et al.*, 2006; Puri *et al.*, 2007] et que différents systèmes ont été utilisés pour l'expression des protéines, ce qui pourrait expliquer ces divergences. Enfin on peut aussi envisager la possibilité que ces assemblages représentent plusieurs états conformationnels qui seraient requis pour des fonctions différentes [Cheung *et al.*, 2010].

#### IV. COOPÉRATIVITÉ ENTRE DEUX MONOMÈRES ADJACENTS

Plusieurs études rapportent une activité ATPase plus importante dans le cas d'un complexe constitué d'un ratio stœchiométrique de Pontine et Reptine (structure non précisée) [Ikura *et al.*, 2000], et dans les hétéro-hexamères et dodécamères [Gribun *et al.*, 2008; Puri *et al.*, 2007]. Puri *et al.* ont montré que l'activité ATPase du complexe dodécamérique est dix fois plus importante que celle du monomère de Pontine, et cinq fois plus importante que celle du monomère de Reptine. A l'aide de complexes dans lesquelles la Pontine ou la Reptine de type naturel était remplacée par un mutant dépourvu d'activité ATPase, ils ont également montré que les deux protéines catalysent l'hydrolyse de l'ATP au sein du complexe [Puri *et al.*, 2007].

L'activité ATPase de chaque protéine isolément ayant parfois été difficilement observable [Ikura *et al.*, 2000; Qiu *et al.*, 1998], il se pourrait qu'elle soit au moins en partie dépendante d'une éventuelle association entre deux monomères de Pontine et Reptine ou de l'une ou l'autre. Plusieurs équipes indiquent d'ailleurs la formation d'hexamères ou de dodécamères présentant une importante coordination entre les sous-unités et dans lesquels la présence dans le site actif de l'Arg finger issu de la seconde sous-unité semble jouer un rôle crucial dans l'hydrolyse de l'ATP [Matias *et al.*, 2006; Moffitt *et al.*, 2009; Ogura *et al.*,

2004; Zhang & Wigley, 2008]. Dans l'homo-hexamère cristallisé par Matias et al., six molécules d'ADP sont présentes dans le site actif situé à l'interface entre deux sous-unités [Matias *et al.*, 2006]. La structure circulaire semble donc permettre l'hydrolyse d'un ATP par monomère, grâce à la coopération de deux sous-unités adjacentes.

## V. BASES DE LIGANDS

Il n'existe actuellement aucun inhibiteur connu pour la Pontine ou la Reptine, il est donc impossible de se baser sur des composés actifs existants afin d'identifier de nouvelles molécules susceptibles d'interagir avec ces deux protéines. Dans ce contexte, on ne dispose d'aucun critère permettant de sélectionner des composés à tester expérimentalement, et l'utilisation de bases de ligands diversifiées est la solution la plus logique afin de couvrir le plus largement possible l'espace chimique. Nous avons donc décidé de tester plusieurs bases de molécules différentes à notre disposition, sans *a priori* sur leur contenu. Les premières bases utilisées ont été obtenues à partir de catalogues commerciaux, qui regroupent des composés existants et d'activité connue, puis des bases de ligands innovants créés au sein même du laboratoire ont également été testées.

Dans un premier temps, l'utilisation de bases commerciales permet d'obtenir des composés facilement accessibles pour des tests enzymatiques et qui répondent à des besoins précis. Plus particulièrement, les bases utilisées réunissent des molécules ayant une activité biologique connue. Celles-ci possèdent donc des propriétés chimiques qui les rendent biologiquement pertinentes et « drug-like », donc biodisponibles, et ont également déjà passé un certain nombre de tests ADME-Tox permettant de juger de leur utilisabilité en tant que médicament. Les bases sélectionnées sont les bases Calbiochem Database of Inhibitors® (qui sera notée Calbiochem par la suite) de Merck, et la Prestwick Chemical Library® (qui sera notée Prestwick), qui présentent toutes les deux une forte diversité chimique et pharmacologique. Le coefficient de Tanimoto moyen [Flower, 1998], qui permet d'évaluer la diversité structurale au sein d'une base de molécules, est de 0.1839 pour Calbiochem, et de 0.1265 pour Prestwick (calculé avec OpenBabel 2.2.99 et FP2 [Guha *et al.*, 2006; O'Boyle *et al.*, 2011]). Ce chiffre signifie qu'il existe seulement environ 18 et 12% de similarité entre les composés des bases Calbiochem et Prestwick respectivement.

Afin d'aller un peu plus loin, nous avons également effectué le criblage de deux bases de ligands innovants, synthétisés au sein du laboratoire. A l'origine, ces molécules ont été conçues à partir de données concernant des cibles protéiques bien précises, mais pas la Pontine. Cependant, n'ayant aucun *a priori* sur la structure des ligands susceptibles d'interagir avec elle, nous avons jugé intéressant de tester ces molécules sur la Pontine,

dans l'éventualité où l'une d'entre elle se révélerait être un inhibiteur de l'activité ATPase de cette protéine. La cible originelle de la première base de ligands, qui regroupe les composés ML\* et FX\*, est la phosphatase Cdc25. Les ligands ont été conçus à partir de la structure du TMC-69, un inhibiteur de Cdc25 aux propriétés antitumorales [Fürstner *et al.*, 2003; Fürstner *et al.*, 2004; Hirano *et al.*, 2001; Kohno *et al.*, 2001] (Figure 33).

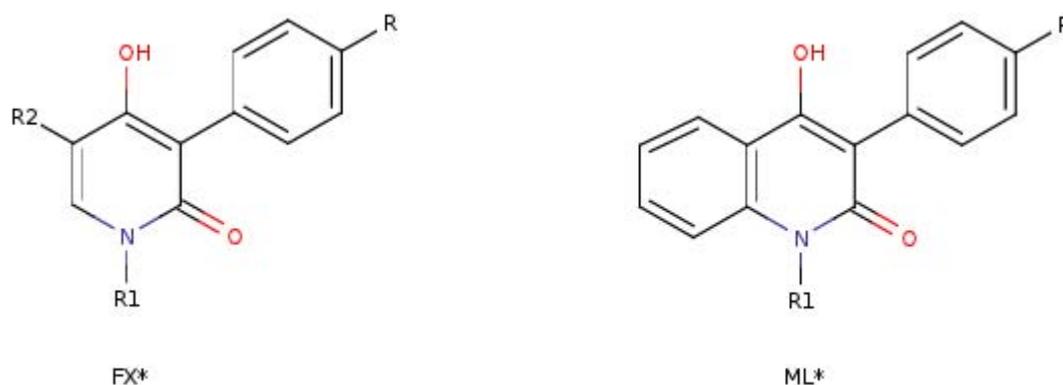


Figure 33. Motifs structurels de base des ligands FX\* et ML\*

La structure des ligands de la seconde base, appelée G\*, repose sur un heptacycle central de type triazépanedione conçu au départ pour inhiber les protéines phospholipases A2 [Lena *et al.*, 2006; Muller *et al.*, 2006] (Figure 34).

NB : Le criblage de ces ligands sur leur cible originelle fait l'objet du chapitre 3 de cette thèse, « Application à la phospholipase humaine sPLA2 de groupe X », les structures de ces molécules y sont décrites plus en détail.

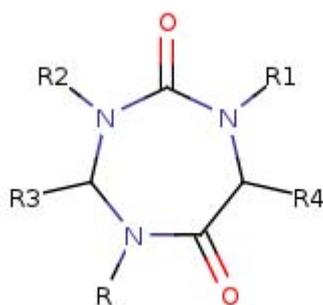


Figure 34. Motif structurel de base des ligands G\*

## VI. ELABORATION D'UN TEST ENZYMATIQUE

*NB : Partie réalisée par Patrick Lestienne (INSERM U1053).*

Afin de bloquer les fonctions de la Pontine, nous avons donc noté que le moyen le plus évident était d'inhiber son activité ATPase. Cependant, nous avons aussi montré plusieurs études dans lesquelles aucune activité ATPase n'a pu être détectée, rendant les mesures d'inhibition impossibles. Toutefois, un test enzymatique fonctionnel a pu être élaboré au sein de l'équipe de Jean Rosenbaum, permettant la mesure de l'activité ATPase de la Pontine. Celui-ci fait appel à une méthode dérivée des tests utilisant le vert de malachite pour l'analyse du phosphate inorganique (Pi). Il utilise donc le dosage colorimétrique du complexe formé par le vert de malachite et le molybdate avec l'orthophosphate libéré lors de l'hydrolyse de l'ATP. Comme on peut le constater sur les droites de corrélation obtenues, la densité optique à 620nm (OD 620 nm) mesurée est proportionnelle à la concentration en phosphate inorganique, et la vitesse de la réaction dépend de la concentration en enzyme dans la solution (Figure 35).

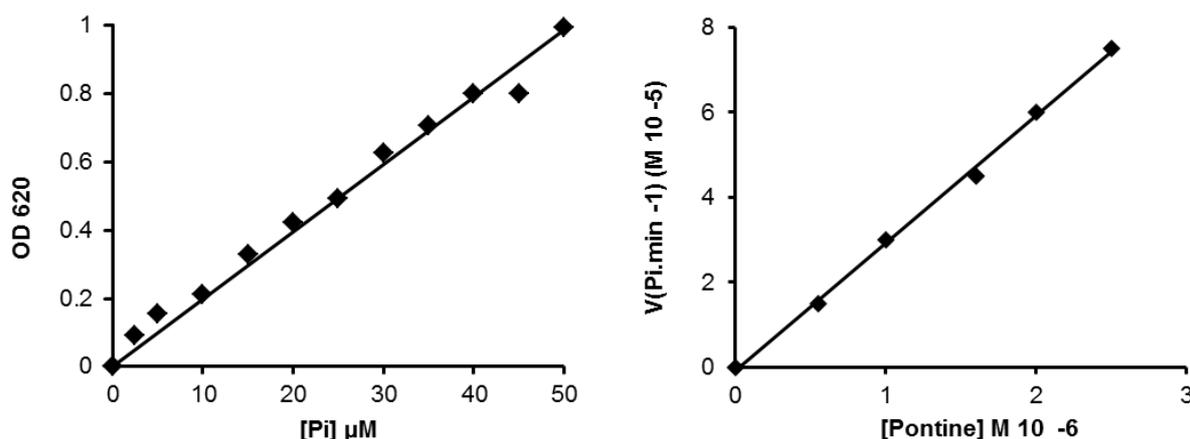


Figure 35. Corrélation entre OD 620 nm et concentration en Pi (à gauche) et entre vitesse de la réaction et concentration en protéine (à droite). Figure, P. Lestienne.

Comme indiqué précédemment, une activation de la vitesse de réaction en présence d'oligonucléotides avait été rapportée par plusieurs équipes [Jónsson *et al.*, 2001; Kanemaki *et al.*, 1999; Makino *et al.*, 1999], mais contestée par d'autres [Matias *et al.*, 2006; Puri *et al.*, 2007; Qiu *et al.*, 1998; Torreira *et al.*, 2008]. Dans cet essai, aucune variation de la vitesse de réaction d'hydrolyse de l'ATP n'a pu être détectée en l'absence d'oligonucléotides, alors que l'introduction en excès d'ADN double ou simple brin ( $[ADN]/[enzyme] = 1-10$ ) a permis d'augmenter jusqu'à quatre fois la vitesse de réaction. Cette stimulation est essentielle pour la détection d'une activité ATPase suffisante à la mesure de l'IC50 des composés (Figure

36). Techniquement, elle est rendue possible par la présence dans la structure cristallographique de la Pontine d'un domaine capable d'interagir avec des oligonucléotides [Matias *et al.*, 2006].

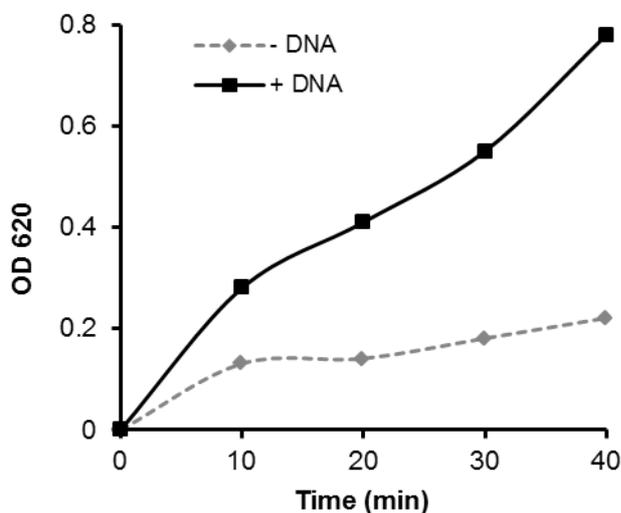


Figure 36. OD 620 nm au cours du temps, en présence (en noir) ou en absence (en gris) d'ADN.  
Figure, P. Lestienne.

Vingt-neuf molécules issues de la Chimiothèque Nationale ont été sélectionnées sur la base d'une évaluation visuelle de leur analogie structurale avec l'ATP pour servir de training set lors de l'élaboration du test enzymatique décrit ci-dessus. Pour aucune de ces molécules une perturbation notable de l'activité ATPase de la Pontine n'a pu être mise en évidence, ces molécules ont donc été considérées comme inactives.

## VII. CRIBLAGE *IN CELLULO*

Parallèlement à l'élaboration d'un test enzymatique, nous avons également fait appel à la plateforme de criblage ImaGIF (Institut de Chimie des Substances Naturelles, CNRS, UPR 2301) afin de réaliser des tests de cytotoxicité *in cellulo*. Ces tests ont été effectués en triplicats sur plusieurs lignées tumorales humaines, à savoir les lignées KB (carcinome buccal), MCF7 (carcinome mammaire), HCT116 (carcinome du colon) et HL60 (lymphoblastome).

## VIII. STRATÉGIE DE CRIBLAGE VIRTUEL MISE EN PLACE

L'objectif de ce travail est l'élaboration d'une liste réduite de composés prioritaires à

tester expérimentalement. Grâce à l'identification expérimentale préalable de vingt-neuf molécules inactives (ou decoys), nous disposons, malgré l'absence d'inhibiteurs connus, d'un moyen d'évaluer les résultats obtenus. En effet, si les fonctions de score utilisées sont efficaces, ces composés devraient être placés plutôt en queue de classement, tandis que se positionneraient en tête les composés potentiellement actifs sur notre cible. La pertinence des fonctions de score a donc été jugée en fonction de leur capacité à discriminer les decoys des inhibiteurs potentiels, de façon individuelle dans un premier temps, puis de manière conjointe, à l'aide d'un consensus scoring. Pour cela, nous avons utilisé la base Calbiochem complétée par ces vingt-neuf molécules inactives comme un training set, afin de mettre en place une stratégie de criblage permettant la sélection de ligands au potentiel inhibiteur élevé. Les autres bases de ligands ont ensuite été criblées selon les conclusions obtenues à partir de Calbiochem.

## D. MATÉRIEL ET MÉTHODES

### I. PRÉPARATION DU RÉCEPTEUR

La structure tridimensionnelle de la Pontine co-cristallisée avec l'ADP est disponible dans la Protein Data Bank sous le code PDB 2c9o [Matias *et al.*, 2006] et nous a fourni un modèle pour le docking. La structure présente un assemblage en forme d'anneau incluant six sous-unités de Pontine numérotées de A à F, et six molécules d'ADP situées dans une cavité à l'interface entre deux sous-unités. A l'exception de deux boucles entre les résidus 142 et 155 ainsi que 248 et 276, la structure est complète. Ces boucles étant assez éloignées du site de liaison à l'ADP, leur influence sur les calculs a été considérée comme négligeable et elles n'ont pas été reconstruites. La structure de la protéine a été nettoyée à l'aide de Discovery Studio 2.1 (Accelrys, San Diego, USA). Les molécules d'eau présentes dans la structure ont été supprimées, tandis que les hydrogènes manquants ont été rajoutés à l'aide du champ de force Charmm.

Comme indiqué précédemment, le site de liaison à l'ATP semble être une cible idéale pour le design de composés ayant pour objectif l'inhibition de l'activité de la Pontine ou de la Reptine. Nous avons également décrit plus haut que l'ADP co-cristallisé se trouve dans une cavité située à l'interface entre deux sous-unités de Pontine. De plus, la coopération entre deux sous-unités semble importante pour l'hydrolyse de l'ATP, en particulier, l'Arg finger, qui interagit avec le  $\gamma$ -phosphate de l'ATP et permet son hydrolyse, est un résidu du monomère adjacent qui pointe dans le site catalytique [Hishida *et al.*, 2004]. En accord avec ces

données, le docking a été effectué sur un dimère de Pontine, afin de prendre en compte les interactions avec les deux sous-unités. Les chaînes A et F de l'hexamère ont donc été conservées, ainsi que l'ADP situé dans la cavité entre les deux monomères. La structure ainsi formée a ensuite subi une minimisation à l'aide d'un algorithme de type Steepest Descent, 2000 pas, gradient 0,01, au cours de laquelle une contrainte a été utilisée pour fixer les atomes du squelette de la protéine et le ligand. L'ADP, conservé pendant l'étape de minimisation afin de garder le site actif libre et éviter que les chaînes latérales des résidus de la cavité ne se positionnent en travers du site de liaison, a ensuite été retiré pour obtenir la structure finale utilisée pendant le docking.

## II. PRÉPARATION DES LIGANDS

Les structures des ligands issus de la base commerciale Calbiochem database of Inhibitors® ont été obtenues en 2D au format SD à partir du site web de Merck Chemicals (références catalogue 539743, 539744, 539745, 539746). Les structures tridimensionnelles ont été générées de façon automatique à l'aide de Catalyst (Accelrys, San Diego, USA) [Kurogi & Güner, 2001; Smellie *et al.*, 1995] via le protocole *Prepare Ligands* de Discovery Studio 2.1 et sauvées au format MOL2. Elles incluent pour chaque molécule tous les états de protonation possibles entre un pH de 6.5 et de 8.5. Les ligands ont ensuite été filtrés à l'aide de DBFilter 2.2.8 (A drug-like analyzer for chemical library, distribué par l'auteur S.-H. Wang, 2005, non publié), afin d'éliminer les composés qui ne répondaient pas à des critères de Lipinski étendus. Les critères de sélection ont été fixés comme suit : masse moléculaire comprise entre 200 et 800, nombre de liaisons flexibles inférieur ou égal à 10, atomes autorisés H, C, N, O, F, Cl, Br, I, P, S. Les fichiers MOL2 ont été convertis au format PDBQT, le format d'entrée du docking, à l'aide du script *prepare\_ligand4.py* fourni par MGLTools [Sanner, 1999]. La base commerciale Calbiochem ainsi préparée contient exactement 900 ligands.

Les structures des ligands de la base Prestwick Chemical Library® ont été obtenues auprès de Prestwick sur demande (<http://www.prestwickchemical.com/index.php?pa=26>) en 2D et au format SD. En suivant la même procédure que pour les ligands de la base Calbiochem, les structures tridimensionnelles des composés ont été générées et sauvées au format MOL2 à l'aide du protocole *Prepare Ligands* de Discovery Studio 2.1 et de DBFilter, puis les fichiers PDBQT ont été obtenus grâce au script *prepare\_ligand4.py*. La base Prestwick finale comprend 1299 structures.

Les structures des vingt-neuf composés sélectionnés dans la Chimiothèque Nationale lors de l'élaboration du test enzymatique ont été manuellement reconstruites dans Discovery

Studio 2.1 et minimisées à l'aide du protocole *Dreiding Minimize* afin d'obtenir une conformation de basse énergie. Comme pour les deux bases précédentes, les structures ont été sauvées au format MOL2 puis converties en PDBQT. En l'absence d'inhibiteurs connus, cet ensemble de composés inactifs nous a servi de repère lors de l'évaluation des résultats de criblage obtenus. Ils ont donc été intégrés à la base Calbiochem pour servir de training set. La base Calbiochem réunit donc au total 929 composés dont les analogues de l'ATP incapables d'inhiber l'activité ATPase de la Pontine (decoys).

Pour la création des bases de ligands ML\* et FX\*, qui comprennent respectivement 988 et 160 molécules, un script en langage Python a été élaboré afin de générer automatiquement en SMILES un ensemble de ligands à partir d'un ou plusieurs motifs de base et de substituants. Grâce à la simplicité du format SMILES il est en effet facile, en partant d'une structure commune, de greffer à la position désirée des branches variées permettant de générer une base de ligands divers. Les molécules au format SMILES ont ensuite été transformées en MOL2 et les structures 3D ont été générées à l'aide d'OpenBabel et de l'option gen3D qui permet de créer les structures tridimensionnelles à partir de la 1D (SMILES) ou de la 2D (SD).

Pour la base de molécules G\*, le macrocycle triazépanedione qui sert de motif commun à toutes les molécules peut adopter plusieurs conformations. Les outils de docking utilisés ici ne permettant pas de flexibilité au sein des cycles pendant le docking, nous avons jugé préférable de réaliser une recherche conformationnelle sur ce cycle et de créer autant d'exemplaires de chaque ligand que de conformations obtenues. La base résultante est composée de 652 conformations pour 155 ligands. *NB : Le détail de cette étude conformationnelle et de la constructions de ces ligands est détaillé dans le chapitre consacré à ces molécules et ne sera pas décrit de nouveau ici.*

Les deux bases de ligands innovants ont elles aussi été converties en PDBQT via le script *prepare\_ligand4.py* fourni par MGLTools [Sanner, 1999].

On peut noter ici que les ligands n'ont pas tous été générés de la même façon. En terme de structure, la conformation de basse énergie obtenue avec l'un ou l'autre des programmes utilisés est dépendante de sa capacité à franchir les barrières énergétique et à trouver les minima énergétiques. Cependant, ce biais de calcul est compensé par la flexibilité du ligand au cours du docking. En effet, à partir du moment où les longueurs et angles de liaison ont été fixées à des valeurs correctes, le programme Vina utilisé pour le docking est indépendant des angles de torsions de départ. La seule exception est le cas d'un cycle pouvant adopter plusieurs conformations, puisque les conformations des cycles ne sont échantillonnées pendant le docking avec Vina. Dans ce cas, et comme nous l'avons fait, une recherche conformationnelle s'impose pour déterminer la ou les conformations à docker.

Les différents protocoles ont également pour conséquence l'attribution de charges partielles différentes d'une base à une autre. Là encore, cette divergence n'a pas d'effet sur les résultats obtenus puisque le programme de docking utilisé ainsi que les programmes de rescoring réattribuent leurs propres charges au cours du calcul.

### III. DOCKING ET RESCORING

Tous les docking ont été réalisés avec Autodock Vina 1.0.2 [Trott & Olson, 2010], et trois types d'expériences différentes ont été menées. Un premier docking protéine rigide/ligand flexible a été d'abord été mené sur toutes les bases de molécules avec le dimère de Pontine. Ensuite, un docking protéine flexible/ligand flexible a été réalisé sur la base Calbiochem (decoys inclus) uniquement, avec le dimère également. Enfin quelques composés sélectionnés ont été testés au cours d'un blind docking (docking sans indication de site de liaison) protéine rigide/ligand flexible sur un monomère de Pontine. Les torsions autorisées sur les ligands ont été automatiquement attribuées lors de la conversion du MOL2 au PDBQT par le script `prepare_ligand4.py`.

- Dimère rigide

Pendant le docking, tous les résidus de la protéine sont gardés fixes tandis que les ligands sont considérés comme flexibles. Le site actif a été défini comme la cavité située à l'interface entre les sous-unités A et F dans laquelle se trouve l'ADP dans la structure cristallographique. L'espace de recherche inclut la totalité de cette cavité, plus une marge d'au moins 3 Å dans toutes les directions. Les dimensions de la boîte de recherche conformationnelle résultante sont de 20×22×20 Å. Au total, soixante-quatorze acides aminés de la première sous-unité et six acides aminés de la seconde sous-unité, y compris l'Arg finger (R357), sont totalement ou partiellement situés à l'intérieur de la boîte.

- Dimère flexible

Pour cette expérience, les ligands sont toujours flexibles, tout comme les chaînes latérales de certains résidus du site actif de la protéine. Un nombre élevé de torsions augmente le temps de calcul de façon considérable, c'est pourquoi nous avons limité de le nombre de torsions sur le récepteur. Les résidus mobiles ont été sélectionnés de façon classique, à l'aide d'une sphère de sélection de 5 Å autour du centre de masse du ligand co-cristallisé, l'ADP. Cette sphère a été choisie car elle permet d'inclure un total de onze torsions sur le récepteur, ce qui correspond à peu près au nombre de torsions maximal par ligand et représente une augmentation du temps de calcul raisonnable. Ces onze torsions sont réparties sur quatre résidus différents, tout issus de la chaîne A, qui sont S17, H18, K76, T77. La boîte de calcul utilisée est identique à celle de l'expérience sur le dimère rigide.

- Monomère rigide, blind docking

Cette dernière expérience, menée sur un nombre très restreint de composés, a été réalisée à partir d'un monomère de Pontine seul dans un premier temps, puis complexé avec l'ADP ou l'ATP ensuite. Dans ces deux derniers cas, les nucléotides sont considérés comme partie intégrante de la protéine et sont donc complètement fixes. Afin de s'affranchir de la notion de site actif et de chercher des poses alternatives à un ancrage dans le site de liaison à l'ATP, la totalité de la protéine a été incluse dans la nouvelle boîte de recherche conformationnelle. Les dimensions de la boîte ainsi créée sont 58×76×70 Å.

Dans tous les cas, le docking a été reconduit cinq fois pour chaque ligand, en utilisant cinq seed fixés à l'avance. Les paramètres par défaut de Vina ont été utilisés pendant les calculs, à savoir :

- exhaustiveness = 8
- nombre de poses générées = 9

Les fichiers de sortie de Vina en PDBQT ont été convertis en pdb à l'aide de scripts écrits dans le laboratoire (voir l'annexe 2 « Gestion des fichiers et automatisation des calculs »), et les hydrogènes non polaires, non considérés par Vina, ont été rajoutés automatiquement à l'aide d'OpenBabel 2.2.3 [Guha *et al.*, 2006]. Le rescoring complet de toutes les poses obtenues a été réalisé avec les fonctions de scoring fournies par Drugscore [Gohlke *et al.*, 2000; Velec *et al.*, 2005] et Xscore [Wang *et al.*, 2002], ce qui donne au total dix fonctions de score distinctes qui sont PAIR\_csd, PAIRSURF\_csd, PAIR\_pdb, PAIRSURF\_pdb et SURF pour Drugscore ; HPscore, HMscore, HSscore et AVEscore pour Xscore, et enfin Vina.

Les scores calculés par fonction de score pour une pose dans les cinq expériences ont été moyennés pour donner le score final par pose et par fonction. Ensuite, seul le meilleur score obtenu par fonction de score pour un ligand (toutes poses confondues) a été conservé pour créer le classement définitif des ligands.

Les consensus scoring ont été réalisés à partir de ces classements, à l'aide du protocole *Consensus Scoring* de Discovery Studio 2.1.

#### IV. TESTS ENZYMATIQUES

*NB : Partie réalisée par Patrick Lestienne (INSERM U1053).*

La Pontine recombinante humaine exprimée chez *Escherichia coli* a été purifiée par chromatographie. Les fractions d'élution résultantes montrent des proportions variables de Pontine sous forme dodécamérique, hexamérique et monomérique. Seule la fraction monomérique a montré une activité ATPase dans le test enzymatique décrit plus haut, et les

essais ont donc tous été réalisés sur cette fraction.

La réaction démarre au moment de l'addition d'ATP et est suivie au cours du temps par la mesure de l'OD 620 nm dans le test colorimétrique mis en place qui détecte la quantité de phosphate inorganique produite. Pour chaque concentration en inhibiteur testée, la vitesse de la réaction est mesurée avec une concentration initiale en ATP de 50  $\mu$ M, ce qui correspond au  $K_m$  de la réaction. Chaque mesure a été réalisée deux fois, à partir de fractions issues de deux purifications différentes, et les valeurs présentées sont les moyennes de ces deux expériences. Les  $IC_{50}$  obtenues sont reproductibles avec un intervalle de confiance de 20%.

Les molécules testées sont dissoutes dans le DMSO, c'est pourquoi l'absence d'interférence entre le DMSO et la vitesse de réaction ou le test colorimétrique a été préalablement vérifiée. De même, l'OD 620 nm de tous les composés testés a également été mesurée à des concentrations allant jusqu'à 50  $\mu$ M. Aucun n'a montré d'interférence avec le test colorimétrique.

Le test étant réalisé en présence d'ADN, trois oligonucléotides différents ont été utilisés pour étudier leur impact respectif sur la stimulation de la réaction d'hydrolyse de l'ATP. Aucune différence significative n'a été observée entre les trois. Les séquences sont disponibles ci-dessous, les parties soulignées sont complémentaires.

- Simple brin (longueur 43 bases)

5'GCT CGC TAC CCG GGG ATC CTC TAG AGT CAT CAG TGC AAG ACC G 3'

- Épingle (longueur 44 bases)

5'TCG CTC TTC TAC TAT GAA CCC CCC TCC CC ATT TTT GGG GAG GGG 3'

- Double brin (longueur 64 bases)

5' TCG CTC TTC TAC TAT GAA CCC CCC TCC CC ATT TTT GGG GAG GGG GGT  
TCA TAG TAG AAG AGC GA 3'

Le protocole complet est décrit dans la publication « First identification of small molecule inhibitors of Pontin by combining virtual screening and enzymatic assay » (publication acceptée chez Biochemical Journal, voir l'annexe 5).

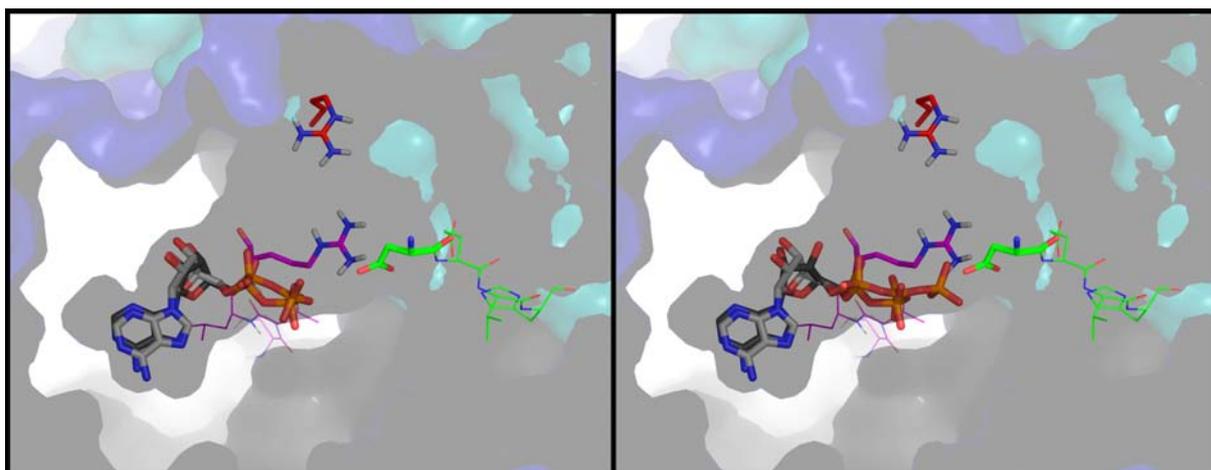
## E. RÉSULTATS ET DISCUSSION

### I. QUALITÉ DU MODÈLE

Dans la plupart des stratégies de criblage virtuel, la capacité à reproduire la pose cristallographique d'un inhibiteur est un critère très important pour l'évaluation de la qualité

des résultats obtenus. Ici, nous ne disposons ni d'un inhibiteur, ni à plus forte raison d'une structure tridimensionnelle présentant le mode de liaison d'un tel ligand. Par contre, la structure PDB disponible présente une molécule d'ADP co-cristallisée avec la protéine. Dans un premier temps, avant le criblage de nos bases de ligands, nous avons donc réalisé le docking de l'ADP et de l'ATP dans le site actif visé, afin de vérifier si l'outil de docking utilisé permettait de reproduire la pose cristallographique, et donc vérifier la qualité de notre modèle.

Les poses obtenues sont très proches de la position adoptée par l'ADP dans le site actif de la protéine (Figure 37). Le groupement adénine des deux molécules dockées est parfaitement superposé à celui de l'ADP co-cristallisé. De même, le ribose est très légèrement déplacé par rapport à la position attendue. La seule différence notable se situe au niveau de la chaîne phosphate, avec celle de l'ADP docké superposée à celle de l'ADP cristallisé, tandis que celle de l'ATP, plus longue et très flexible, est déplacée de façon remarquable sur plusieurs poses.



*Figure 37. Superposition de la pose cristallographique de l'ADP et des poses de docking de l'ADP (à gauche) et de l'ATP (à droite). L'Arg finger, le motif Walker B et le sensor 2 sont figurés en rouge, vert et rose respectivement*

Au moins une quinzaine de résidus sont en contact à la fois avec l'ADP et l'ATP, en particulier plusieurs résidus du Walker A et l'arginine R404 du sensor 2. Ces acides aminés sont P72, G73, T74, G75, K76, A78 (Walker A), R404 (sensor 2), S17, H18, H20, G38, L39, G41, Y366, I374 et R378. L'ATP, mais pas l'ADP, est en outre en contact avec l'acide aspartique D302 du Walker B dont la mutation est couramment utilisée pour empêcher l'hydrolyse de l'ATP. La chaîne phosphate de l'ADP est stabilisée par un réseau d'au moins huit liaisons hydrogène impliquant six résidus différents, alors que celle de l'ATP compte un

réseau de treize liaisons hydrogène venant de huit résidus, dont cinq des six engagés dans la liaison de l'ADP. Ces poses confirment la capacité de Vina à reproduire la pose cristallographique.

En outre, cinq expériences ont été conduites sur chaque ligand, en utilisant pour chacune un seed différent. Quelques molécules au hasard ont été sélectionnées dans la base Calbiochem et les meilleurs poses obtenues ont été superposées. Les poses montrent une parfaite superposition, ce qui indique que les poses calculées par Vina sont extrêmement reproductibles. Cette notion est essentielle car elle met en évidence la grande convergence de résultats obtenue avec Vina. Lors du rescoring, les fonctions de score utilisées ont elles aussi montré une très forte convergence.

## II. CLASSEMENTS PAR FONCTIONS DE SCORE INDIVIDUELLES

A l'aide du training set constitué de la base Calbiochem et des decoys, toutes les fonctions de scoring ont d'abord été évaluées individuellement en fonction de leur capacité à éliminer les decoys du haut du classement. Pour cela, nous avons comparé le nombre de decoys présents dans le Top 10% pour chaque fonction de scoring, c'est à dire les composés qui se classent dans les dix premiers pourcent de la base, soit les 93 molécules les mieux classées. Nous avons ensuite élargi le champ, en comparant également le nombre de decoys dans le Top 15%, qui représente les 139 molécules les mieux classées (Table 10).

| Fonctions    | Decoys                       |
|--------------|------------------------------|
| Vina         | 1 + 1 (51 + 109)             |
| PAIR_csd     | 5 (21, 32, 57, 58, 81)       |
| PAIRSURF_csd | 4 + 1 (30, 36, 43, 65 + 104) |
| PAIR_pdb     | 1 (51)                       |
| PAIRSURF_pdb | 1 (53)                       |
| SURF         | 1 (37)                       |
| HPscore      | 1 (95)                       |
| HMscore      | -                            |
| HSscore      | -                            |
| AVEScore     | 1 (83)                       |

Table 10. Nombre et classement (entre parenthèse) des decoys dans les top 10% + 15% du training set (base Calbiochem, decoys inclus)

En premier lieu, nous nous sommes intéressés aux résultats de la fonction de score fournie par Vina. La qualité de cette fonction est un point crucial pour l'exactitude de l'expérience dans son ensemble, puisqu'elle a non seulement été utilisée pour classer les poses des ligands obtenues, mais aussi et surtout pendant la génération de ces poses. Les résultats obtenus sont corrects, avec seulement un decoy dans le top 10%, mais montrent tout de même une petite faiblesse avec deux decoys dans le top 15%.

De toutes les fonctions de score utilisées, les deux fonctions de Drugscore CSD, PAIR\_csd et PAIRSURF\_csd donnent les plus mauvais résultats et semblent ne pas être capables de distinguer les decoys au sein de la base. Elles présentent en effet respectivement cinq et quatre decoys parmi les composés du top 10%, et un de plus dans le top 15% pour PAIRSURF\_csd. Les mêmes fonctions mais issues de la version PDB de Drugscore, PAIR\_pdb et PAIRSURF\_pdb, donnent de bien meilleurs résultats, avec seulement un decoy dans les top 10% et 15%. C'est également le cas de la fonction SURF, pour laquelle un premier decoy est trouvé dans le top 10% en position 37, mais aucun autre n'est observé avant la 218ème place.

Enfin, les fonctions fournies par Xscore sont les plus satisfaisantes, avec seulement un decoy dans les top 10% et 15% pour AVEscore et HPscore et aucun pour HScore et HMscore, les premiers apparaissant aux positions 187 et 198 respectivement.

Les fonctions de Xscore ont donc été les plus efficaces en ce qui concerne l'élimination des decoys de la tête du classement, suivies de près par SURF, et PAIR\_pdb ou PAIRSURF\_pdb, puis Vina. Au contraire, les deux fonctions PAIR\_csd et PAIRSURF\_csd échouent complètement et n'ont pas été utilisées plus avant.

### III. CONSENSUS SCORING

En raison du nombre très limité de résultats expérimentaux et de l'absence de molécules actives parmi les composés testés, il est difficile de juger de façon fiable de la qualité d'une fonction de scoring. Dans ce cas, l'utilisation du consensus scoring pour la création d'un ensemble réduit de composés à tester expérimentalement peut être un bon moyen d'éliminer de la tête du classement un grand nombre de faux positifs [Charifson *et al.*, 1999].

Pour des raisons de commodité, tous les consensus décrits par la suite vont être notés de la façon suivante : « Fonction1/Fonction2 X », avec X le pourcentage considéré pour chacune des fonctions utilisées dans le consensus. Tous les consensus possibles avec les huit fonctions restantes ont été examinés, excepté les combinaisons de fonctions issues du même programme, puisque ces fonctions ne sont pas indépendantes, ce qui est un

prérequis pour réaliser un consensus valide [Wang & Wang, 2001] (voir le chapitre « Introduction »). Les consensus réalisés sont du type « rank-by-vote », ce qui correspond à réaliser l'intersection mathématique des top pour plusieurs fonctions de score.

Afin d'obtenir une vue d'ensemble des consensus permettant le mieux d'éliminer les decoys, un consensus assez large impliquant le top 30% (soit 279 composés) des fonctions a été réalisé. Les intersections impliquant les fonctions les plus efficaces individuellement, soit toutes les fonctions de Xscore et les fonctions PAIR\_pdb, PAIRSURF\_pdb et SURF de Drugscore, ont été réalisées en priorité, puis la fonction de score de Vina a également été croisée avec toutes les autres fonctions (Table 11).

| <b>Fonctions</b>  | <b>Mol communes</b> | <b>Decoys communs</b> |
|-------------------|---------------------|-----------------------|
| HP/PAIR_pdb       | 228                 | 1                     |
| HM/PAIR_pdb       | 197                 | 4                     |
| HS/PAIR_pdb       | 207                 | 2                     |
| AVE/PAIR_pdb      | 215                 | 3                     |
| HP/PAIRSURF_pdb   | 230                 | 1                     |
| HM/PAIRSURF_pdb   | 197                 | 4                     |
| HS/PAIRSURF_pdb   | 207                 | 2                     |
| AVE/PAIRSURF_pdb  | 216                 | 3                     |
| HP/SURF           | 242                 | 1                     |
| HM/SURF           | 204                 | 2                     |
| HS/SURF           | 217                 | 1                     |
| AVE/SURF          | 220                 | 2                     |
| Vina/HP           | 125                 | -                     |
| Vina/HM           | 129                 | 2                     |
| Vina/HS           | 119                 | -                     |
| Vina/AVE          | 121                 | 1                     |
| Vina/PAIR_pdb     | 118                 | 3                     |
| Vina/PAIRSURF_pdb | 117                 | 3                     |
| Vina/SURF         | 127                 | 2                     |
| Vina/HS/SURF      | 100                 | -                     |
| Vina/HP/SURF      | 108                 | -                     |

Table 11. Nombre de molécules communes et de decoys dans les consensus 30%

Globalement, les résultats de Xscore et des fonctions conservées de Drugscore sont fortement corrélés, avec entre 242 et 197 ligands en commun sur 279 molécules dans le top 30%. Dans les consensus « Xscore/PAIR\_pdb 30 » et « Xscore/PAIRSURF\_pdb 30 » (avec Xscore qui équivaut à n'importe quelle fonction issue de Xscore), on obtient systématiquement le même résultat, avec quatre decoys pour HMscore, trois pour AVEscore, deux pour HSscore et un seul pour HPscore, ce qui permet à première vue d'établir un classement entre les fonctions de Xscore. Ce classement ne correspond pas exactement aux résultats des fonctions individuelles qui distinguaient HMscore et HSscore d'un côté, avec aucun decoy dans les top 10% et 15%, et HPscore et AVEscore de l'autre, avec un decoy dans le top 10%. Cependant, il est en accord avec les résultats obtenus dans le consensus « Xscore/SURF 30 », qui présente deux decoys pour HMscore et AVEscore et seulement un pour HPscore et HSscore. Les résultats obtenus montrent que les consensus impliquant HPscore et HSscore sont les meilleurs, de plus HPscore est la fonction qui présente le plus grand nombre de composés commun avec les trois fonctions de Drugscore, avec jusqu'à 242 molécules pour « HPscore/SURF 30 ».

Si l'on compare les deux consensus « Xscore/PAIR\_pdb 30 » et « Xscore/PAIRSURF\_pdb 30 » avec « Xscore/SURF 30 », on constate également que ce dernier est apparemment plus performant, avec au maximum deux decoys dans l'intersection, alors que l'on monte jusqu'à quatre avec les deux autres fonctions. Les deux meilleurs résultats sont donc obtenus avec les consensus « HPscore/SURF 30 » et « HSscore/SURF 30 », cependant, aucun des consensus n'a permis l'élimination complète des decoys.

Lorsque l'on croise les fonctions de Drugscore ou de Xscore avec Vina, on obtient un nombre beaucoup plus réduit de molécules communes, compris entre 117 et 129, ce qui signifie une corrélation moins importante qu'entre Drugscore et Xscore. Cependant les consensus « Vina/HPscore 30 » et « Vina/HSscore 30 » sont particulièrement intéressants puisqu'ils montrent un nombre de decoys égal à zéro. Ces consensus sont donc capable de discriminer totalement les decoys. Comme précédemment, ces consensus sont plus efficaces que « Vina/HMscore30 » et « Vina/AVEscore 30 » qui comprennent respectivement deux et un decoy(s). Dans une moindre mesure, le consensus « Vina/SURF 30 » est lui aussi meilleur que les consensus « Vina/PAIR\_pdb 30 » et « Vina/PAIRSURF 30 », avec deux decoys contre trois.

Les résultats précédents indiquent que les fonctions SURF pour Drugscore et HPscore ou HSscore pour Xscore permettent d'obtenir les meilleurs consensus. De plus, les consensus créés avec Vina fournissent un nombre beaucoup plus réduit de composés communs, mais les conjonctions HPscore ou HSscore avec Vina sont les seuls procédures

capables d'éliminer totalement les decoys de la liste des tops. En outre, comme mentionné plus haut, la totalité des poses ont été générées avec Vina, ce qui rend cette fonction plus critique que les autres pour la qualité des résultats obtenus et nous a incité à l'intégrer dans le consensus final. Pour ces raisons, nous avons mené deux consensus supplémentaires, impliquant non plus deux, mais trois fonctions de score, qui sont « Vina/HPscore/SURF 30 » et « Vina/HSscore/SURF 30 ». Ces deux consensus donnent des listes de 108 et 100 molécules respectivement, dont aucun decoy. Le nombre un peu plus élevé de composés communs, et l'efficacité légèrement supérieure de HPscore dans les autres consensus nous ont poussé à sélectionner le consensus « Vina/HPscore/SURF 30 » pour la suite de notre travail.

#### IV. TRAITEMENT POST-EXPÉRIMENTAL

Un examen visuel de la centaine de ligands sélectionnés grâce à ce consensus dans la base Calbiochem a été effectué, afin de vérifier que les structures ont été correctement générées et que les poses obtenues sont correctes (pas de clashes intra ou intermoléculaires). Cette vérification post-expérimentale montre que pour environ 10% des molécules la structure automatiquement générée par Catalyst présente des erreurs, probablement à cause d'une mauvaise annotation dans les bases obtenues au format SD. En effet, certains motifs n'ont pas été compris par le programme et la structure obtenue ne correspond pas à la molécule présente dans la base à l'origine et vendue par Merck. En particulier, les hétérocycles aromatiques sont très souvent désaromatisés, ce qui a pour conséquence directe la perte de la planarité du cycle et engendre une grande différence en terme de géométrie. Le même constat a été fait lors du criblage des molécules de la base Prestwick. En conséquence, une vérification de la structure de tous les composés issus du consensus « Vina/HPscore/SURF 30 » a été effectuée manuellement sur les deux bases commerciales. Toutes les molécules présentant des erreurs ont donc été corrigées, puis criblées à nouveau, et réintégrées aux classements avant la sélection des composés à tester expérimentalement.

Cette méthodologie est particulièrement pertinente lorsque l'on traite des bases de molécules de grande taille (de manière arbitraire, dès que l'on dépasse la centaine de ligands), pour lesquelles une construction manuelle de tous les composés requiert un temps de travail trop important. En réalisant ces vérifications uniquement sur les molécules d'intérêt, on accélère le processus, même si cette stratégie a également pour conséquence de ne pas traiter les faux négatifs éventuels classés de façon erronée à cause d'une mauvaise gestion de la structure.

Les composés des bases FX\* et ML\*, convertis du SMILES au MOL2 par l'intermédiaire de l'option gen3D d'OpenBabel n'ont pas montré de telles erreurs de structure et n'ont donc pas nécessité de correction, tout comme ceux de la base G\*, générés manuellement.

## V. ÉLABORATION DE LISTES DE COMPOSÉS À TESTER

Les fonctions Vina, HPscore et SURF ont donc été choisies afin de créer un consensus efficace sur notre système, c'est à dire capable d'éliminer les decoys de la liste de composés prioritaires. Afin de réduire le nombre de ligands à tester à un nombre acceptable pour nos collaborateurs, le premier consensus 30% a été réduit à un consensus 15%, conduisant à la sélection de vingt-deux molécules dans la base Calbiochem. Parmi ces molécules, dix ont été choisies au hasard pour être testées dans l'essai enzymatique décrit précédemment.

De la même façon, la base commerciale Prestwick a été criblée puis triée sur la base du consensus « Vina/HPscore/SURF 15 » qui a conduit à un ensemble de vingt-trois molécules, dont quinze ont de nouveau été sélectionnées aléatoirement.

Pour les molécules de la base G\* synthétisées dans l'équipe de Gilles Guichard, chaque ligand est présent en deux ou quatre exemplaires selon la conformation du cycle central (voir Préparation des ligands, et le chapitre Phospholipase). Les composés ont donc d'abord été triés à partir des consensus à 10%, 15% et 20%. Dans les listes ainsi créées, plusieurs ligands ont été identifiés à travers un, deux, trois ou quatre conformères. A partir de ces données, nous avons réalisé une liste de priorité à échelons, en pondérant l'ordre de priorité des ligands par le nombre de conformations trouvées dans les consensus. Au final, deux composés ont été classés en priorité 1 (la plus forte), deux en priorité 2, et quatre en priorité 3, ce qui donne un total de huit molécules à tester expérimentalement.

*NB : Ces composés sont actuellement en cours de criblage via le test enzymatique décrit précédemment et les résultats ne sont pas encore connus, c'est pourquoi il ne sont pas mentionnés dans la partie « Résultats des tests expérimentaux ».*

En ce qui concerne les bases FX\* et ML\*, seule la synthèse d'un nombre réduit de ces composés avait déjà été réalisée avant la création des bases virtuelles. Ces molécules ont donc été abordées dans une approche orientée « drug design » dans laquelle le but était de guider la synthèse. Dans un premier temps, deux bases d'environ 100 molécules pour FX\* et 600 molécules pour ML\* ont été dessinées et criblées. Les scores obtenus sur les meilleures molécules de ces bases se sont révélés supérieurs à ceux obtenus dans le criblage de toutes les autres bases, indiquant un fort potentiel inhibiteur sur l'activité ATPase de Pontine. Cependant, à partir des premières discussions avec nos collaborateurs responsables de la

synthèse, nous avons constaté que plusieurs molécules bien classées dans le criblage ne pourraient pas être préparées. En effet, lors de la construction des bases, nous avons inclus plusieurs séries dont les voies de synthèse n'avaient pas encore été explorées et qui se sont révélées être difficiles à obtenir dans un délai raisonnable (nombre d'étapes important). Nous avons donc établi plusieurs listes en fonction de ce qui paraissait possible ou pas, au niveau de la synthèse, et créé de nouvelles molécules à partir des motifs envisageables. Ce processus a été répété plusieurs fois, jusqu'à l'aboutissement de la synthèse de six molécules identifiées à l'aide du consensus « Vina/HPscore/SURF » qui ont été criblées expérimentalement via le test enzymatique et les tests cellulaires décrits dans la partie précédente.

## VI. RÉSULTATS DES TESTS EXPÉRIMENTAUX

*NB : Partie réalisée par Patrick Lestienne (INSERM U1053).*

Sur les dix molécules sélectionnées dans la base Calbiochem, trois ont montré un effet inhibiteur sur l'activité ATPase de la Pontine. Le ligand Calb124, identifié dans la base Calbiochem comme un inhibiteur des kinases Akt1, Akt2 et Akt3, possède une IC<sub>50</sub> de 22 µM (Figure 38). Le composé Calb539, un inhibiteur non compétitif allostérique de la phosphatase PTP1B, présente une IC<sub>50</sub> de 15 µM (Figure 38). Enfin, la molécule Calb557, qui possède une activité inhibitrice non sélective sur la famille des PKC, a une IC<sub>50</sub> de 16 µM (Figure 38). Les K<sub>i</sub> résultants sont de 11, 7 et 8 µM respectivement.

Sur les quinze molécules de la base Prestwick testées, seule la molécule Prest148, un inhibiteur du récepteur de leukotriène 1, possède une activité inhibitrice sur la Pontine, avec une IC<sub>50</sub> de 13 µM, soit un K<sub>i</sub> de 6 µM (Figure 38).

Ces quatre molécules étant déjà dotées d'une activité biologique connue, il est intéressant de noter la spécificité de ces ligands envers leurs différentes cibles. Par exemple, la molécule Prest148 possède une IC<sub>50</sub> pour sa cible originelle entre 4 et 7 nM, ce qui est beaucoup plus élevé que pour la Pontine. De même, le composé Calb124 possède une affinité très élevée pour les kinases de la famille Akt, avec des IC<sub>50</sub> de 58 nM, 210 nM et 2 µM pour Akt1, Akt2 et Akt3. En revanche, les molécules Calb539 et Calb557, sont assez peu spécifiques, et présentent une activité sur leurs cibles originelles du même ordre de grandeur que celle observée ici, avec une IC<sub>50</sub> d'environ 4 µM sur la PTP1B pour Calb539, et d'environ 3 µM sur la PKC $\delta$  et 30 à 100 µM sur les PKC $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\epsilon$ ,  $\zeta$ ,  $\eta$  pour Calb557. Selon les données fournies par Merck Chemicals et Prestwick, ces molécules sont biodisponibles et non toxiques.

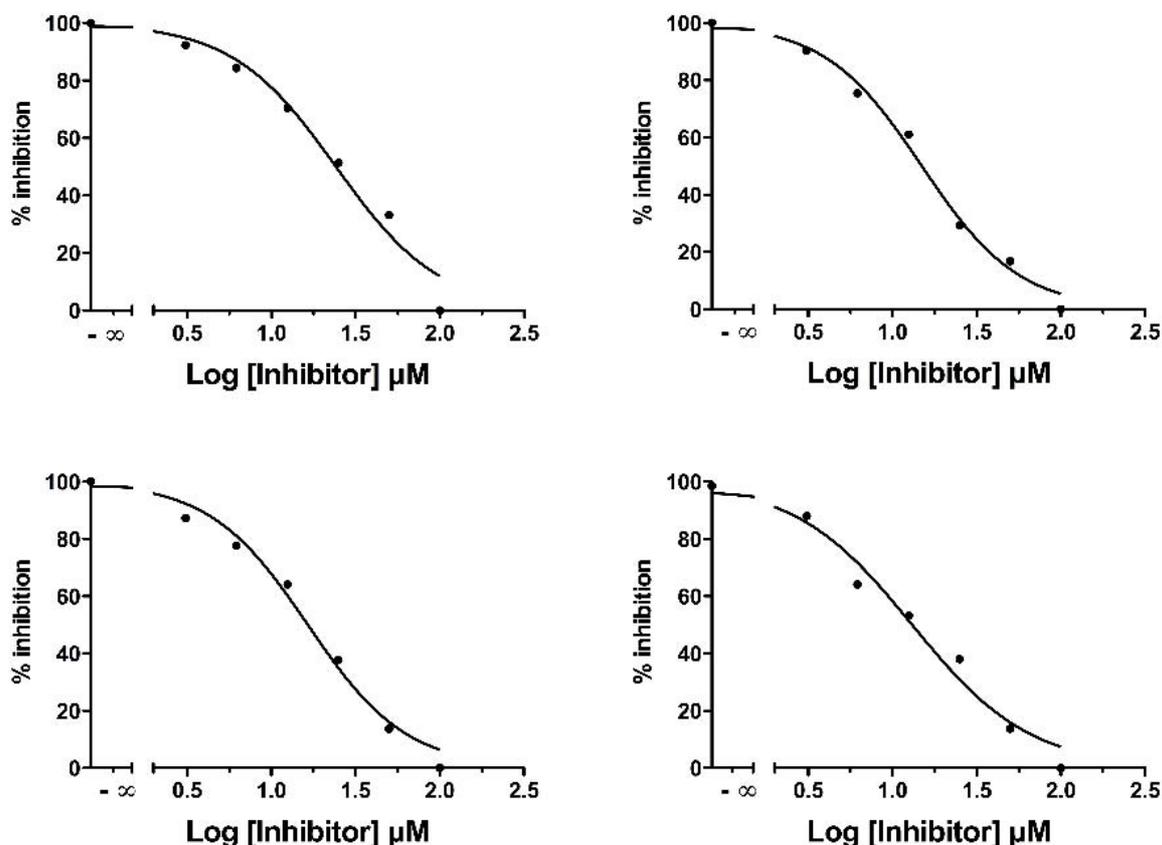


Figure 38. Courbes d'inhibition pour les quatre molécules actives identifiées. Figure, P. Lestienne.

Parmi les six composés de la base ML\*, une seule molécule a montré un effet inhibiteur sur l'activité d'hydrolyse de l'ATP de la Pontine. L'IC50 obtenue pour cet inhibiteur innovant est de 17  $\mu\text{M}$  et son  $K_i$  de 8  $\mu\text{M}$ . Lors du criblage *in cellulo*, cette molécule est également la seule à présenter une activité cytotoxique. Cette activité a été retrouvée sur les quatre lignées cellulaires testées.

## VII. ANALYSE DES MODES DE LIAISON

Dès lors que le criblage virtuel concerne des bases de ligands assez importantes et structurellement diverses, l'évaluation des scores prend le plus souvent le pas sur l'observation des poses. A plus forte raison, lorsque l'on ne connaît pas d'inhibiteurs pour une protéine comme c'est le cas ici, il n'est pas possible de comparer les poses de nouveaux ligands avec les poses cristallographiques de molécules connues. Pourtant l'étude du site actif et des interactions qui s'y font entre le ligand et la protéine peut être un guide très utile pour le design de nouvelles molécules actives. Avec l'identification de quatre inhibiteurs, une

étude approfondie des meilleures poses adoptées par ceux-ci a permis d'identifier des résidus souvent engagés dans la liaison des inhibiteurs qui sont également impliqués dans la liaison et l'hydrolyse de l'ATP. Le positionnement des molécules dans le site actif présente par ailleurs quelques points de similitude avec celui de l'ADP (Figure 39).

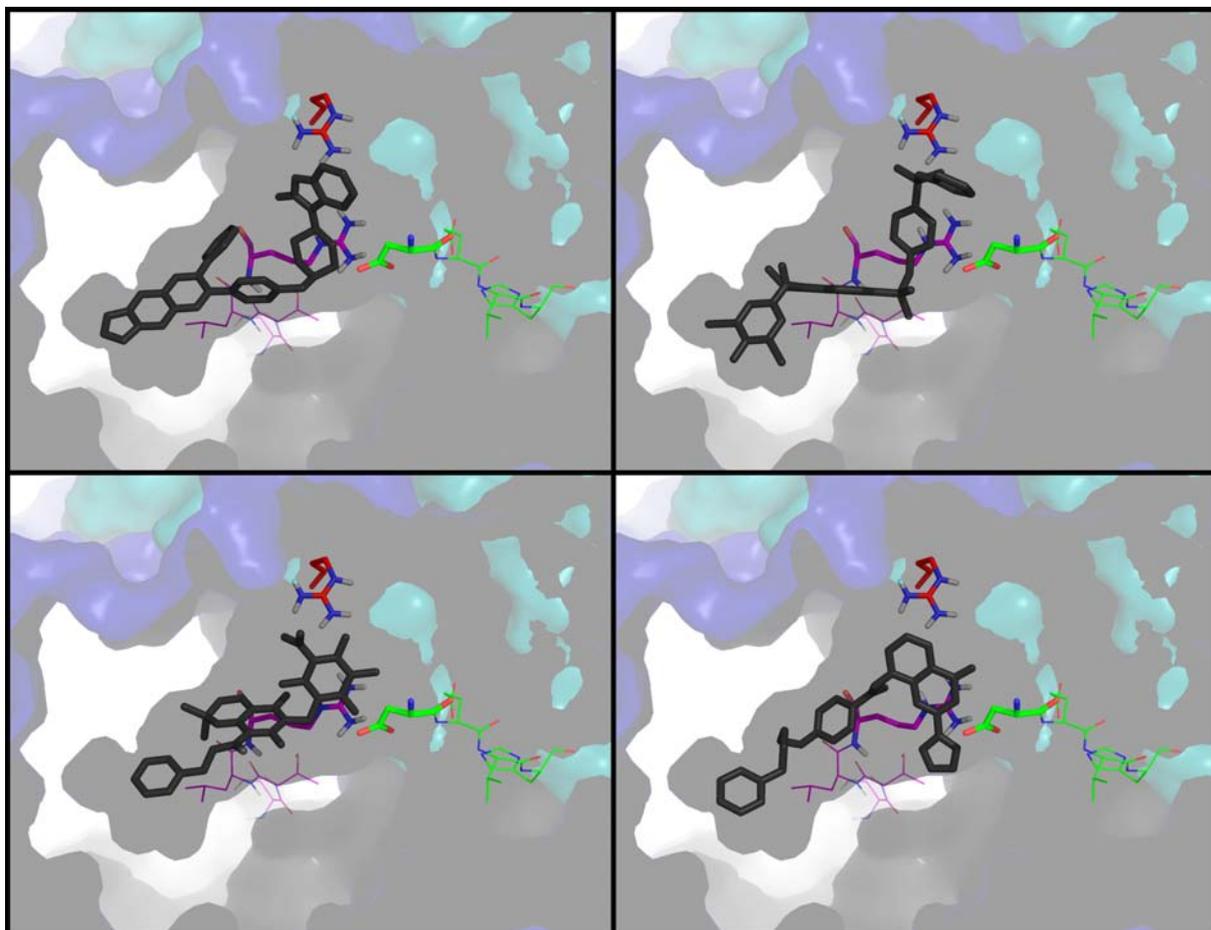


Figure 39. Meilleures poses de docking des quatre inhibiteurs. L'Arg finger, le motif Walker B et le sensor 2 sont figurés en rouge, vert et rose respectivement

Parmi la quinzaine de résidus en contact avec l'ADP, dix sont en contact avec tous les ligands. Le nombre de résidus communs aux modes de liaison de l'ADP et des ligands va de onze pour la molécule Calb557, à treize pour Prest148, et quatorze pour Calb124 et Calb539. La position où se place le groupement adénine est occupée dans le cas des ligands par des cycles aromatiques variables mais qui permettent de conserver la planarité de l'adénine. Les molécules Calb539 et Prest148 possèdent toutes les deux un groupement polaire accepteur de liaison hydrogène superposé au  $\gamma$ -phosphate de l'ATP.

Plusieurs résidus du Walker A, qui conditionne la liaison de l'ATP, sont présents dans les modes de liaison de tous les ligands comme de l'ADP et de l'ATP, et chaque pose est en

contact avec au moins trois acides aminés de ce motif. Pour Calb539 et Prest148, on compte également une possibilité de liaison hydrogène avec le squelette de G73 similaire à celle observée à la fois chez l'ATP et l'ADP.

Le résidu D302 du Walker B, dont l'importance capitale dans la catalyse de l'ATP a été démontrée plus d'une fois à l'aide du mutant D302N, est en contact avec tous les ligands ainsi que l'ATP, mais pas l'ADP. Aucun autre acide aminé du Walker B n'est situé à proximité des ligands ou de l'ATP.

La molécule Calb557 est la seule en dehors de l'ATP à présenter une liaison hydrogène avec l'arginine R404 du sensor 2 qui est essentielle pour la liaison et l'hydrolyse de l'ATP ainsi que la coordination entre les sous-unités.

A première vue, l'espace occupé par l'ADP ou même l'ATP est largement plus restreint que la cavité et une zone du site est totalement vide. En fait, toute la partie du site actif où affleure la seconde sous-unité est laissée libre par ces deux molécules, qui ne sont en contact avec aucun des résidus de celle-ci. Au contraire, tous les inhibiteurs identifiés s'étendent dans cette partie de la cavité et sont en contact avec la seconde sous-unité par le biais du résidu D356, voisin de l'Arg finger R357. Le ligand Calb557 est également en contact avec D353, tandis que Calb539 et Prest148 sont situés à proximité de l'Arg finger, avec dans le premier cas, deux accepteurs de liaison hydrogène qui pointent vers le groupement guanidinium.

Ces modes de liaisons, qui correspondent uniquement aux meilleures poses pour chaque ligand, ne sont que des indications, mais ils permettent de repérer quelques motifs communs à toutes les molécules et qui ont donc certainement eu un poids important lors du docking. Ceux-ci correspondent d'ailleurs aux résidus mis en avant dans la littérature pour leur rôle dans la liaison ou l'hydrolyse de l'ATP ou les interactions entre les sous-unités.

## VIII. DOCKING FLEXIBLE

Dans un système biologique, rien n'est jamais complètement rigide. Lors de l'arrimage d'un ligand à une protéine, le ligand va adopter une conformation spécifique pour s'adapter à son hôte, tout comme la cavité va se déformer pour accommoder le ligand. On appelle ce phénomène l'« induced fit ». Le docking rigide, s'il permet de traiter un grand nombre de ligands dans un temps réduit, fait donc une approximation assez importante en considérant la protéine comme une entité figée. Aujourd'hui cependant, de nombreux programmes de docking, Vina inclus, permettent d'introduire une certaine flexibilité au sein de la protéine, en autorisant des torsions sur les chaînes latérales des résidus. Si cette possibilité ne permet pas de rendre compte d'un changement conformationnel majeur ni d'aucun mouvement du

squelette de la protéine, elle permet tout de même d'observer quelques réorientations de résidus ayant pour but de créer de nouvelles interactions avec les ligands. Un criblage flexible a donc été mené sur le training set constitué de la base Calbiochem et des decoys, en incluant des torsions sur plusieurs résidus du site actif.

Le consensus « Vina/HPscore/SURF 15 » sur le docking flexible donne un total de trente-et-une molécules, dont un decoy. Parmi celles-ci, dix-sept font partie des vingt-deux molécules présentes dans le même consensus réalisé sur le docking rigide, soit plus de 77%. On note également la présence des trois molécules actives identifiées expérimentalement. Les deux docking sont donc très cohérents et les résultats obtenus indiquent que la stratégie de docking rigide est suffisamment performante. La comparaison des poses de liaison obtenues confirme cette forte corrélation, avec assez peu de changements significatifs. Cette constatation n'est pas très surprenante, car la configuration du site actif, qui est complètement enfoui et donc fortement contraint, ne permet qu'un nombre limité de mouvements sur les chaînes latérales des résidus.

## IX. NATURE DES INHIBITEURS

*NB : Partie réalisée par Patrick Lestienne (INSERM U1053).*

Lors de la mise en place du criblage virtuel ayant permis d'identifier les composés, le site de liaison de l'ATP a été désigné comme le meilleur site pour l'inhibition de l'activité de la Pontine et ciblé spécifiquement au cours des calculs. Cette stratégie correspond donc à la recherche d'inhibiteurs compétitifs, qui ont le même site de liaison que le substrat naturel, l'ATP.

La nature des inhibiteurs identifiés a été déterminée expérimentalement à l'aide du test enzymatique décrit plus haut en mesurant la réaction d'hydrolyse de l'ATP en fonction de la concentration en inhibiteur à partir de cinq concentrations initiales en ATP. La représentation des résultats obtenus en double-réciproque nous a ainsi permis d'observer les variations de  $K_m$  et de  $V_{max}$  représentatives du type d'inhibition (Figure 40).

La représentation en double réciproque obtenue pour la molécule Calb 557 montre une augmentation du  $K_m$ , mais pas de changement dans la vitesse maximale  $V_{max}$ , ce qui est typique d'une inhibition compétitive. Au contraire, celles tracées pour les composés Calb539 et Prest148 présentent des variations à la fois du  $K_m$  et de  $V_{max}$ , ce qui correspond à une inhibition mixte ou incompétitive. Enfin, la molécule Calb124 possède un  $K_m$  invariable, au contraire de  $V_{max}$ , indiquant une inhibition de type non compétitive.

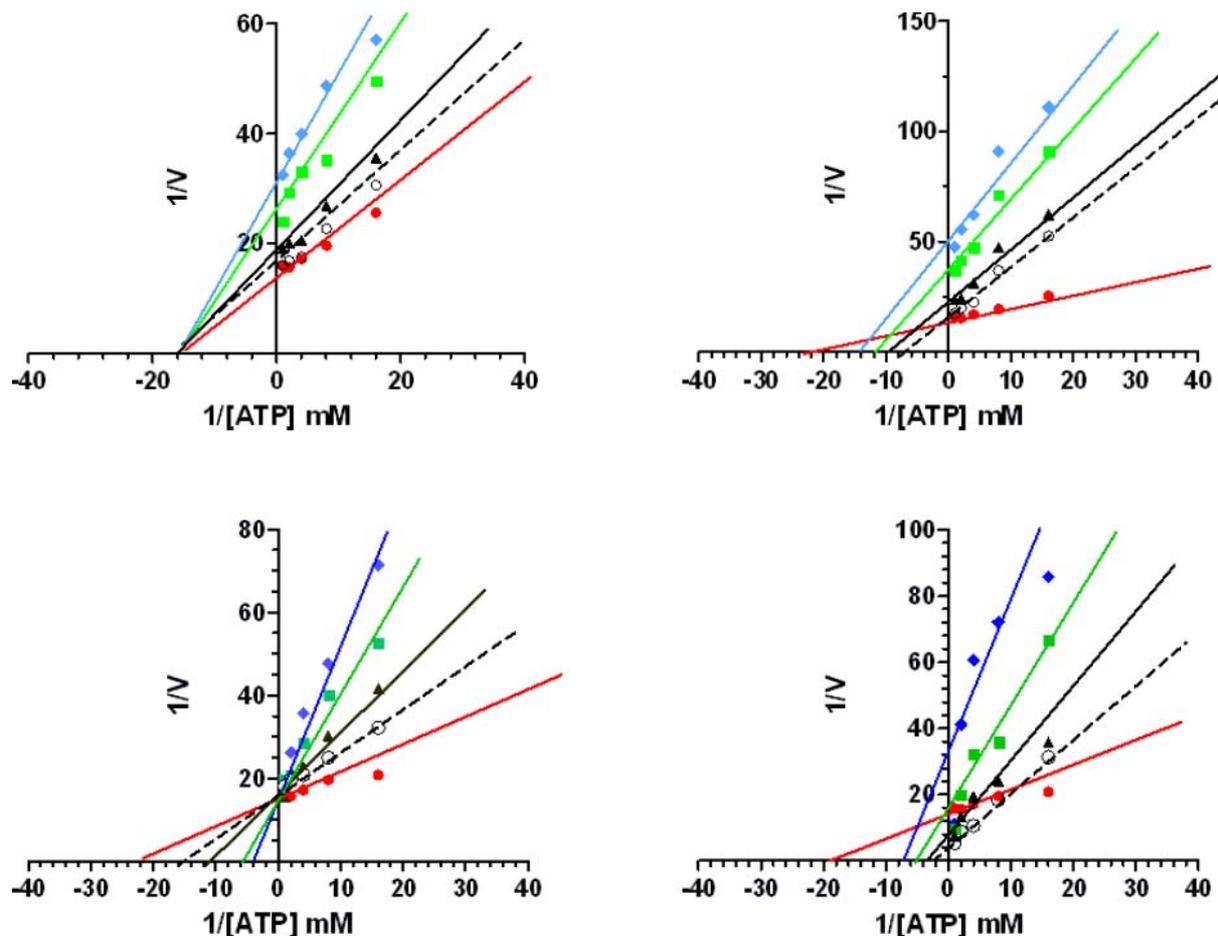


Figure 40. Double réciproque pour les quatre inhibiteurs. La droite correspondant au contrôle est en rouge, les droites bleues, vertes et noires représentant les différentes concentrations testées. Figure, P. Lestienne.

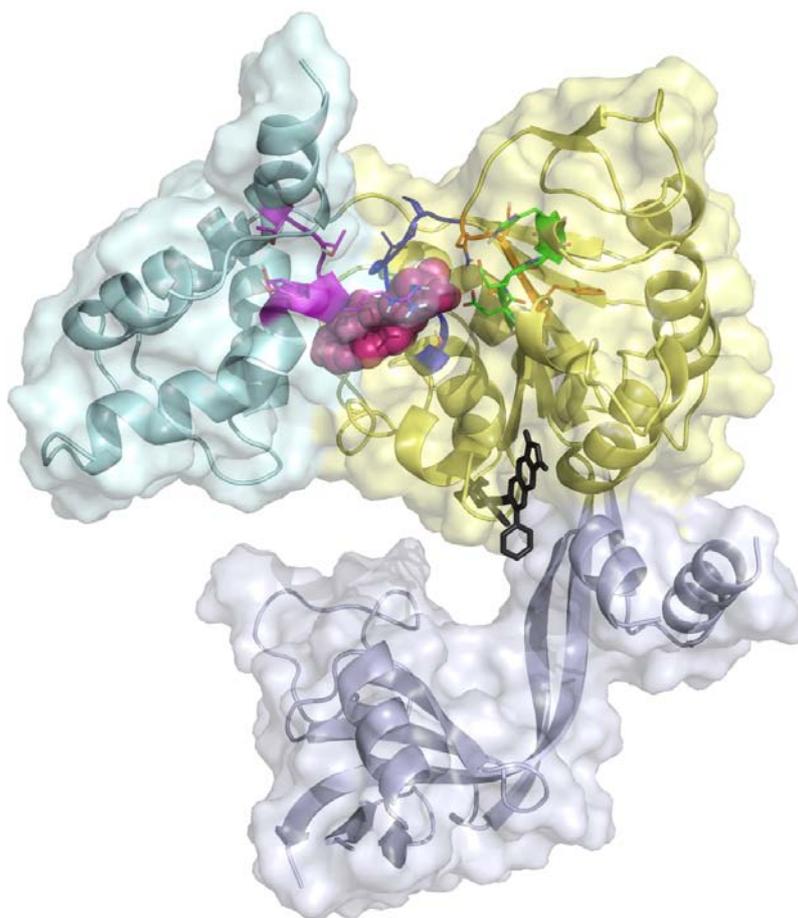
Contre toute attente, vu le mode de sélection des ligands à tester, parmi les quatre molécules actives seule Calb557 inhibe donc de manière compétitive l'activité ATPase de la Pontine tandis que les composés Calb 539 et Prest148 montrent une inhibition mixte ou incompétitive et Calb124 un profil non compétitif.

## X. BLIND DOCKING

Afin de comprendre ces résultats, de nouveaux docking ont été menés dans le but de trouver d'autres modes de liaison pour les trois ligands ne présentant pas une inhibition compétitive. Ces expériences impliquent uniquement le monomère de Pontine seul ou en présence d'ADP ou d'ATP et ont été conduites « à l'aveugle » (blind docking), c'est à dire que la protéine entière a été considérée dans la boîte de recherche conformationnelle. Dans

les expériences précédentes menées sur le dimère, le site actif était entièrement fermé grâce à l'intervention de la seconde sous-unité qui recouvrait la cavité. Celle-ci, même dans les cas où elle n'interagissait pas directement avec les ligands, contribuait à contraindre la pose dans un espace restreint et était en partie responsable de la forme globale de celui-ci, jouant ainsi un rôle majeur sur le mode de liaison des molécules. Le blind docking réalisé ici permet aux ligands, non seulement d'explorer la surface autour de la cavité recouverte par la seconde sous-unité dans le dimère, mais également celle de totalité de la protéine, les autorisant à explorer de nouveaux sites de liaison.

Pour le composé Calb124, les résultats obtenus en l'absence ou en la présence d'ADP ou d'ATP sont les mêmes. En dehors de quelques poses situées soit en partie dans le site actif quand il est libre, soit juste à l'entrée lorsque l'ATP se trouve dans le site, presque la totalité des poses sont situées dans un sillon à l'interface entre les domaines I et II de la protéine (Figure 41).

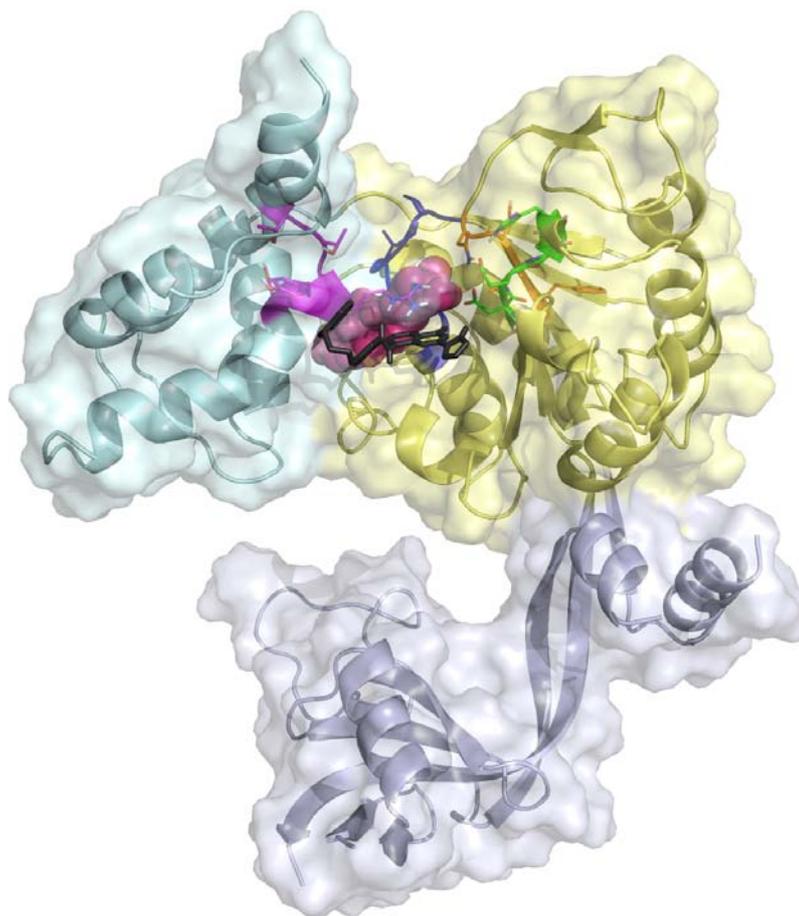


*Figure 41. Pose préférentielle de la molécule Calb124 dans le blind docking. Les motifs Walker A et B et les sensor 1 et 2 sont figurés en bleu, vert, orange et rose respectivement*

Ces poses ne sont pas incompatibles avec l'existence d'un multimère similaire à celui observé en cristallographie (Matias 2006, JMolBiol), et sont apparemment trop éloignées du site de liaison de l'ATP pour interférer directement avec la liaison de celui-ci, mais elles sont situées dans le domaine II, qui comprend la région supposée du site de liaison des oligonucléotides. La molécule Calb124 pourrait donc interagir avec le site de liaison à l'ADN, or nous avons montré lors de l'élaboration du test enzymatique que la présence d'ADN provoquait une stimulation très importante de la réaction. Le potentiel inhibiteur de Calb124 pourrait donc être lié à sa capacité à interférer avec la liaison de l'ADN, ce qui est en accord avec la nature non compétitive de cette inhibition par rapport à l'ATP.

Pour la molécule Prest148, au contraire, les résultats du docking sur la protéine seule ou complexée avec l'ADP et ceux obtenus en présence d'ATP sont très différents. Lorsque le site actif est occupé par l'ADP, le docking fournit un grand nombre de solutions différentes, ne permettant pas vraiment de déterminer une pose préférentielle. Lorsqu'il est vide, comme dans le cas de Calb124, les poses les plus nombreuses sont localisées à moitié dans le site actif. Ceci est d'une part en contradiction avec les résultats expérimentaux, et peut s'expliquer par un biais du calcul d'autre part, puisque le nombre d'interactions possibles à l'entrée de la cavité est supérieur à celui en surface de la protéine, tout en ne présentant pas les contraintes imposées par une pose située complètement dans le site. En présence d'ATP dans le site, le docking montre de façon claire une pose préférentielle à l'entrée du site actif (Figure 42). Cette pose, qui est en contact avec le  $\gamma$ -phosphate de l'ATP, pourrait empêcher l'hydrolyse de l'ATP, et est placée de telle sorte qu'elle ne permet pas la sortie de l'ADP. Elle est également incompatible avec l'existence d'un multimère tel qu'il est décrit dans la littérature [Matias *et al.*, 2006], puisque la molécule Prest148 se place dans une position qui devrait alors être occupée par différents résidus de la seconde sous-unité. Ce résultat est compatible avec l'hypothèse d'une inhibition incompétitive, dans laquelle le ligand ne peut interagir avec la protéine que lorsque celle-ci est déjà en complexe avec son substrat. Ce serait donc la liaison de l'ATP qui permettrait à Prest148 de se lier à la protéine, et du même coup d'inhiber son activité ATPase.

En ce qui concerne le composé Calb539, le blind docking n'a pas permis de mettre en évidence une pose préférentielle, que ce soit en présence de la protéine seule ou du complexe avec l'ADP ou l'ATP. Dans ce dernier cas, on observe un nombre légèrement plus élevé de poses situées à l'entrée du site actif, dans une position similaire à celle de la molécule Prest148 décrite ci-dessus, mais également une grande quantité de poses diverses, de telle sorte qu'il nous a été difficile d'identifier de façon déterminante une pose concluante.



*Figure 42. Pose préférentielle de la molécule Prest148 dans le blind docking. Les motifs Walker A et B et les sensor 1 et 2 sont figurés en bleu, vert, orange et rose respectivement*

Il est important de noter que dans les expériences de blind docking, il est beaucoup plus difficile d'observer une convergence dans les poses obtenues, en raison de l'augmentation de la taille de l'espace de recherche conformationnelle qui entraîne une explosion du nombre de solutions possibles. Ceci est à mettre en relation avec les résultats décrits ci-dessus. D'un côté, le profil de Calb539, qui ne montre pas de pose préférentielle, est le résultat direct de ce manque de convergence et n'est pas surprenant dans une expérience de ce type. De l'autre, les résultats convergents obtenus avec Calb124 et Prest148 ont encore plus de poids, puisque la pose préférentielle est retrouvée un très grand nombre de fois malgré le défaut de convergence typique du blind docking.

## F. BILAN

En combinant un criblage virtuel avec un test enzymatique, nous avons donc identifié pour la première fois quatre inhibiteurs de l'activité ATPase de la Pontine, une protéine impliquée dans de nombreux processus biologiques et une cible intéressante dans la thérapie contre le cancer. En outre, nous avons également guidé la synthèse d'un ligand innovant pour cette protéine, qui montre une activité cytotoxique *in cellulo*, et d'autres sont actuellement en cours de criblage.

L'utilisation du consensus scoring, pour contre-balancer le manque de données concernant d'éventuelles molécules actives, nous a permis de sélectionner des composés à tester en priorité avec succès puisque parmi la vingtaine de composés testés, quatre se sont révélés capables d'inhiber l'activité ATPase de la Pontine. L'analyse *a posteriori* des différents consensus scoring établis pendant l'évaluation de toutes les possibilités montre qu'aucun autre n'aurait sélectionné les quatre molécules qui se sont révélées actives, ce qui confirme que nous avons fait le bon choix en choisissant ce consensus. L'examen individuel des fonctions de scoring utilisées révèle également qu'aucune d'entre elles n'avait classé toutes les quatre molécules en tête.

Cependant, la détermination expérimentale de la nature des inhibiteurs a produit des résultats surprenants, et plutôt en désaccord avec l'approche de criblage virtuel ayant permis de sélectionner ces molécules. Le composé Calb557 montre une inhibition compétitive par rapport à l'ATP qui confirme l'approche utilisée pour la création de listes de priorité. En revanche, la nature incompétitive ou mixte de l'inhibition de Calb539 et Prest148 ainsi que la non compétitivité de Calb124 suggèrent des interactions indirectes avec l'activité ATPase de la Pontine qui n'expliquent pas leur présence parmi les composés prioritaires. A l'aide d'expériences de blind docking, nous avons essayé de comprendre leurs modes d'action. Dans le cas de Calb124, l'inhibition semble impliquer une perturbation de la liaison des oligonucléotides au domaine II de la protéine ce qui sous-entend que l'interaction avec l'ADN, ainsi que la stimulation de l'activité ATPase qu'elle implique sont des processus essentiels. Pourtant, la dépendance de l'activité ATPase de la Pontine envers l'ADN reste controversée dans la littérature. Dans le cas de Prest148 le mécanisme d'action semble impliquer une perturbation de l'hydrolyse de l'ATP et de la sortie de l'ADP après hydrolyse ou de la formation de multimères. Cette seconde possibilité implique que l'interaction entre deux sous-unités est nécessaire pour l'hydrolyse de l'ATP. Bien que souvent évoquée dans la littérature, cette coopération n'est pas mise en évidence dans le test enzymatique mis en place ici, puisque seule la fraction monomérique de la protéine s'est révélée active.

Des études complémentaires concernant les mécanismes qui régissent l'inhibition de

l'activité ATPase de la Pontine par ces différentes molécules sont donc nécessaires afin de clarifier les points qui restent en suspens. L'identification de ces composés ouvre la voie à l'obtention des structures cristallographiques des complexes formés par la Pontine et ces inhibiteurs qui fournirait des informations structurales essentielles sur leurs modes de liaison.

Une étude similaire menée sur la Reptine serait très intéressante, sachant que ces protéines possèdent une forte homologie et sont impliquées dans les mêmes processus. Enfin, une étude structurale approfondie du multimère mixte cristallisé récemment pourrait également permettre de vérifier la validité de nos résultats dans le cas d'un dimère constitué non plus de deux sous-unités de Pontine, mais d'une sous-unité de Pontine et d'une de Reptine.

Globalement, la méthodologie mise en place a permis un gain de temps et d'argent considérable en comparaison d'un criblage expérimental complet des bases de ligands. Les composés identifiés pour le moment sont modérément actifs, avec une IC50 de l'ordre du micromolaire, mais les résultats obtenus sont encourageants pour la découverte de nouvelles molécules d'intérêt thérapeutique sur la Pontine. Le mode de liaison de la molécule Calb557 pourra être utilisé pour guider le design de nouvelles molécules sur le modèle de l'inhibiteur innovant déjà identifié, afin d'augmenter leur activité jusqu'au nanomolaire.



## **CONCLUSION**



Au cours du travail présenté dans ce manuscrit, nous avons exploré plusieurs outils et stratégies de criblage virtuel « structure-based » pour leur application dans le cadre de la découverte de nouveaux médicaments.

La modélisation moléculaire et le criblage virtuel sont des domaines de la chimie qui exigent une interaction très forte avec différents métiers afin de relier les résultats obtenus *in silico* avec des données expérimentales. De plus, les données structurales fournies par la cristallographie en particulier constituent la base de ces méthodologies. Le modélisateur est donc souvent au centre d'un réseau de collaborations qui fait intervenir des structuralistes, des biologistes ou des biochimistes et des chimistes de synthèse. Au cours de ces trois années de thèse, j'ai eu l'occasion de participer à deux projets impliquant des partenaires venant de la biochimie, de la biologie cellulaire et de la chimie de synthèse. Deux cibles présentant un intérêt thérapeutique ont été étudiées, qui sont la phospholipase sPLA2 de groupe X et l'ATPase Pontine. La première est notamment impliquée dans diverses maladies inflammatoires tandis que la seconde est sur-exprimée dans plusieurs types de cancers.

Concernant la phospholipase sPLA2 de groupe X, un ensemble de ligands connus a été utilisé pour établir un protocole de criblage. Cependant, les complexes ainsi que les classements obtenus sont en désaccord avec les données expérimentales fournies par la littérature, ne permettant pas de mettre en place une procédure validée et donc de sélectionner des molécules potentiellement actives. Ces difficultés sont probablement liées à la présence au sein du site actif d'un ion calcium essentiel au fonctionnement de la protéine mais qui s'avère être un obstacle au cours des calculs. La continuation de ce projet passe donc par l'exploration de nouveaux outils de criblage et notamment de docking.

Pour la Pontine, aucun inhibiteur n'avait été rapporté dans la littérature, et nous avons donc utilisé des decoys identifiés par nos collaborateurs lors de l'élaboration d'un test enzymatique pour cette protéine. Lors du criblage de deux bases de molécules commerciales, l'utilisation d'un consensus scoring nous a permis d'établir des listes de composés prioritaires parmi lesquels quatre se sont révélés actifs. En parallèle, une base virtuelle de molécules a été conçue à partir d'un motif structurel correspondant à l'expertise de synthèse d'un membre de l'équipe et a mené à la synthèse de cinq molécules innovantes dont une active.

Ce projet a donc mené à l'identification des premiers inhibiteurs de la Pontine, par le biais d'un criblage virtuel. Toutefois, les calculs menés sur le site de liaison de l'ATP sont en désaccord avec la nature de ces inhibiteurs dont seul un est compétitif, et les poses

obtenues ne reflètent donc pas un complexe probable. Ceci peut être expliqué en considérant que le protocole mis en place a été capable d'éliminer les molécules non susceptibles de se lier à la protéine plutôt que de sélectionner les composés potentiellement actifs. De plus, les activités anti-ATPase des inhibiteurs identifiés ont pu être reliées à des poses alternatives n'entrant pas en compétition avec l'ATP. Des données structurales complémentaires pourraient être obtenues grâce à la résolution de la structure cristalline de la Pontine en présence de ces inhibiteurs et permettraient de confirmer les résultats obtenus.

Des tests cellulaires récemment obtenus (voir la publication en annexe 5) ont permis d'observer une diminution de la prolifération cellulaire chez quatre molécules sur cinq, y compris une activité cytotoxique pour la molécule innovante conçue dans le laboratoire. L'étude des relations entre ces effets et l'activité anti-ATPase des inhibiteurs concernant la Pontine est actuellement en cours.



## **RÉFÉRENCES**



Abagyan R., Batalov S., Cardozo T., Totrov M., Webber J., Zhou Y. (1997). Homology modeling with internal coordinate mechanics: deformation zone mapping and improvements of models via conformational search. *Proteins Suppl* **1**, 29-37.

Abagyan R., Totrov M., Kuznetsov D. (1994). ICM—A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* **15**, 488-506.

Acevedo L.G., Bieda M., Green R., Farnham P.J. (2008). Analysis of the mechanisms mediating tumor-specific changes in gene expression in human liver tumors. *Cancer Res.* **68**, 2641-2651.

Alberts B. (1998). The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* **92**, 291-294.

Amadasi A., Spyrakis F., Cozzini P., Abraham D.J., Kellogg G.E., Mozzarelli A. (2006). Mapping the Energetics of Water–Protein and Water–Ligand Interactions with the “Natural” HINT Forcefield: Predictive Tools for Characterizing the Roles of Water in Biomolecules. *J. Mol. Biol.* **358**, 289-309.

Amadasi A., Surface J.A., Spyrakis F., Cozzini P., Mozzarelli A., Kellogg G.E. (2008). Robust Classification of “Relevant” Water Molecules in Putative Protein Binding Sites. *J. Med. Chem* **51**, 1063-1067.

Amaro R.E., Baron R., McCammon J.A. (2008). An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *J. Comput. Aided Mol. Des.* **22**, 693-705.

Ammelburg M., Frickey T., Lupas A.N. (2006). Classification of AAA+ proteins. *J. Struct. Biol.* **156**, 2-11.

Anderson A.C., O’Neil R.H., Surti T.S., Stroud R.M. (2001). Approaches to solving the rigid receptor problem by identifying a minimal set of flexible residues during ligand docking. *Chem. Biol.* **8**, 445-457.

Andersson A., Ritz C., Lindgren D., Edén P., Lassen C., Heldrup J., Olofsson T., Råde J., Fontes M., Porwit-Macdonald A. et al. (2007). Microarray-based classification of a consecutive series of 121 childhood acute leukemias: prediction of leukemic and genetic subtype as well as of minimal residual disease status. *Leukemia* **21**, 1198-1203.

Aoki M., Yokota T., Sugiura I., Sasaki C., Hasegawa T., Okumura C., Ishiguro K., Kohno T., Sugio S., Matsuzaki T. (2004). Structural insight into nucleotide recognition in tau-protein kinase I/glycogen synthase kinase 3 beta. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 439-446.

Aronov A.M., Tang Q., Martinez-Botella G., Bemis G.W., Cao J., Chen G., Ewing N.P., Ford P.J., Germann U.A., Green J. et al. (2009). Structure-guided design of potent and selective pyrimidylpyrrole inhibitors of extracellular signal-regulated kinase (ERK) using

conformational control. *J. Med. Chem.* **52**, 6362-6368.

Atilla-Gokcumen G.E., Pagano N., Streu C., Maksimoska J., Filippakopoulos P., Knapp S., Meggers E. (2008). Extremely tight binding of a ruthenium complex to glycogen synthase kinase 3. *Chembiochem* **9**, 2933-2936.

B-Rao C., Subramanian J., Sharma S.D. (2009). Managing protein flexibility in docking and its applications. *Drug Discov. Today* **14**, 394-400.

Balsinde J., Balboa M.A., Insel P.A., Dennis E.A. (1999). Regulation and inhibition of phospholipase A2. *Annu. Rev. Pharmacol. Toxicol.* **39**, 175-189.

Bar-Haim S., Aharon A., Ben-Moshe T., Marantz Y., Senderowitz H. (2009). SeleX-CS: a new consensus scoring algorithm for hit discovery and lead optimization. *J. Chem. Inf. Model.* **49**, 623-633.

Basso K., Margolin A.A., Stolovitzky G., Klein U., Dalla-Favera R., Califano A. (2005). Reverse engineering of regulatory networks in human B cells. *Nat. Genet.* **37**, 382-390.

Bauer A., Chauvet S., Huber O., Usseglio F., Rothbacher U., Aragnol D., Kemler R., Pradel J. (2000). Pontin52 and reptin52 function as antagonistic regulators of beta-catenin signalling activity. *EMBO J.* **19**, 6121-6130.

Bauer A., Huber O., Kemler R. (1998). Pontin52, an interaction partner of beta-catenin, binds to the TATA box binding protein. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14787-14792.

Bax B., Carter P.S., Lewis C., Guy A.R., Bridges A., Tanner R., Pettman G., Mannix C., Culbert A.A., Brown M.J. et al. (2001). The structure of phosphorylated GSK-3beta complexed with a peptide, FRATtide, that inhibits beta-catenin phosphorylation. *Structure* **9**, 1143-1152.

Baxter C., Murray C., Waszkowycz B., Li J., Sykes R., Bone R., Perkins T., Wylie W. (2000). New approach to molecular docking and its application to virtual screening of chemical databases. *J. Chem. Inf. Comput. Sci.* **40**, 254-262.

Baxter C.A., Murray C.W., Clark D.E., Westhead D.R., Eldridge M.D. (1998). Flexible docking using Tabu search and an empirical estimate of binding affinity. *Proteins* **33**, 367-382.

Beauchard A., Ferandin Y., Frère S., Lozach O., Blairvacq M., Meijer L., Thiéry V., Besson T. (2006). Synthesis of novel 5-substituted indirubins as protein kinases inhibitors. *Bioorg. Med. Chem.* **14**, 6434-6443.

Bender A., Glen R.C. (2005). A discussion of measures of enrichment in virtual screening: comparing the information content of descriptors with increasing levels of sophistication. *J. Chem. Inf. Model.* **45**, 1369-1375.

Berg O.G., Gelb M.H., Tsai M.D., Jain M.K. (2001). Interfacial enzymology: the secreted phospholipase A(2)-paradigm. *Chem. Rev.* **101**, 2613-2654.

Berman H.M. (2008). The Protein Data Bank: a historical perspective. *Acta Crystallogr.*,

A, Found. Crystallogr. **64**, 88-95.

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.

Bernstein F.C., Koetzle T.F., Williams G.J., Meyer Jr E.F., Brice M.D., Rodgers J.R., Kennard O., Shimanouchi T., Tasumi M. (1977). The protein data bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535-542.

Bertrand J.A., Thieffine S., Vulpetti A., Cristiani C., Valsasina B., Knapp S., Kalisz H.M., Flocco M. (2003). Structural characterization of the GSK-3beta active site using selective and non-selective ATP-mimetic inhibitors. *J. Mol. Biol.* **333**, 393-407.

Bhat R., Xue Y., Berg S., Hellberg S., Ormö M., Nilsson Y., Radesäter A., Jerning E., Markgren P., Borgegård T. et al. (2003). Structural insights and biological effects of glycogen synthase kinase 3-specific inhibitor AR-A014418. *J. Biol. Chem.* **278**, 45937-45945.

Bieganski R.M., Yarmush M.L. (2011). Novel ligands that target the mitochondrial membrane protein mitoNEET. *J. Mol. Graph. Model.* **29**, 965-973.

Bienz M., Clevers H. (2000). Linking colorectal cancer to Wnt signaling. *Cell* **103**, 311-320.

Bissantz C., Folkers G., Rognan D. (2000). Protein-Based Virtual Screening of Chemical Databases. 1. Evaluation of Different Docking/Scoring Combinations. *J. Med. Chem.* **43**, 4759-4767.

Bissantz C., Kuhn B., Stahl M. (2010). A medicinal chemist's guide to molecular interactions. *J. Med. Chem.* **53**, 5061-5084.

Blanc J., Lalanne C., Plomion C., Schmitter J., Bathany K., Gion J., Bioulac-Sage P., Balabaud C., Bonneu M., Rosenbaum J. (2005). Proteomic analysis of differentially expressed proteins in hepatocellular carcinoma developed in patients with chronic viral hepatitis C. *Proteomics* **5**, 3778-3789.

Błażewska K.M., Ni F., Haiges R., Kashemirov B.A., Coxon F.P., Stewart C.A., Baron R., Rogers M.J., Seabra M.C., Ebetino F.H. et al. (2011). Synthesis, stereochemistry and SAR of a series of minodronate analogues as RGGT inhibitors. *Eur J Med Chem* **46**, 4820-4826.

Bleicher K.H., Böhm H., Müller K., Alanine A.I. (2003). Hit and lead generation: beyond high-throughput screening. *Nat Rev Drug Discov* **2**, 369-378.

Böhm H.J. (1992a). The computer program LUDI: a new method for the de novo design of enzyme inhibitors. *J. Comput. Aided Mol. Des.* **6**, 61-78.

Böhm H.J. (1992b). LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. *J. Comput. Aided Mol. Des.* **6**, 593-606.

Böhm H.J. (1994). The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J.*

Comput. Aided Mol. Des. **8**, 243-256.

Böhm H.J. (1998). Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs. *J. Comput. Aided Mol. Des.* **12**, 309-323.

Boström J., Greenwood J.R., Gottfries J. (2003). Assessing the performance of OMEGA with respect to retrieving bioactive conformations. *Journal of Molecular Graphics and Modelling* **21**, 449-462.

Bourne P.E., Beran B., Bi C., Bluhm W.F., Dimitropoulos D., Feng Z., Goodsell D.S., Prlić A., Quinn G.B., Rose P.W. et al. (2011). The evolution of the RCSB Protein Data Bank website. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **1**, 782-789.

Brooijmans N., Kuntz I.D. (2003). Molecular recognition and docking algorithms. *Annu Rev Biophys Biomol Struct* **32**, 335-373.

Burke J.E., Dennis E.A. (2009). Phospholipase A2 biochemistry. *Cardiovasc. Drugs Ther.* **23**, 49-59.

Carlson H.A. (2002). Protein flexibility and drug design: how to hit a moving target. *Curr. Opin. Chem. Biol.* **6**, 447-452.

Carlson M.L., Wilson E.T., Prescott S.M. (2003). Regulation of COX-2 transcription in a colon cancer cell line by Pontin52/TIP49a. *Mol. Cancer* **2**, 42.

Cavasotto C.N., Abagyan R.A. (2004). Protein flexibility in ligand docking and virtual screening to protein kinases. *J. Mol. Biol.* **337**, 209-225.

Chabot S., Koumanov K., Lambeau G., Gelb M.H., Balloy V., Chignard M., Whitsett J.A., Touqui L. (2003). Inhibitory effects of surfactant protein A on surfactant phospholipid hydrolysis by secreted phospholipases A2. *J. Immunol.* **171**, 995-1000.

Chang G., Guida W.C., Still W.C. (1989). An internal coordinate Monte Carlo method for searching conformational space. *J. Am. Chem. Soc* **111**, 4379-4386.

Chang M.W., Ayeni C., Breuer S., Torbett B.E. (2010). Virtual screening for HIV protease inhibitors: a comparison of AutoDock 4 and Vina. *PLoS ONE* **5**, e11955.

Charifson P.S., Corkery J.J., Murcko M.A., Walters W.P. (1999). Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **42**, 5100-5109.

Cheng A.C., Coleman R.G., Smyth K.T., Cao Q., Soulard P., Caffrey D.R., Salzberg A.C., Huang E.S. (2007). Structure-based maximal affinity model predicts small-molecule druggability. *Nat. Biotechnol.* **25**, 71-75.

Cheng T., Li X., Li Y., Liu Z., Wang R. (2009). Comparative assessment of scoring functions on a diverse test set. *J. Chem. Inf. Model.* **49**, 1079-1093.

Cheung K.L.Y., Huen J., Houry W.A., Ortega J. (2010). Comparison of the multiple

oligomeric structures observed for the Rvb1 and Rvb2 proteins. *Biochem. Cell Biol.* **88**, 77-88.

Chilton F.H., Averill F.J., Hubbard W.C., Fonteh A.N., Triggiani M., Liu M.C. (1996). Antigen-induced generation of lyso-phospholipids in human airways. *J. Exp. Med.* **183**, 2235-2245.

Choi J., Heo K., An W. (2009). Cooperative action of TIP48 and TIP49 in H2A.Z exchange catalyzed by acetylation of nucleosomal H2A. *Nucleic Acids Res.* **37**, 5993-6007.

Clark D.E. (2008). What has virtual screening ever done for drug discovery?. *Expert Opin. Drug Discov* **3**, 841-851.

Clark R.D., Strizhev A., Leonard J.M., Blake J.F., Matthew J.B. (2002). Consensus scoring for ligand/protein interactions. *J. Mol. Graph. Model.* **20**, 281-295.

Clark R.D., Webster-Clark D.J. (2008). Managing bias in ROC curves. *J. Comput. Aided Mol. Des.* **22**, 141-146.

Claussen H., Buning C., Rarey M., Lengauer T. (2001). FlexE: efficient molecular docking considering protein structure variations. *J. Mol. Biol.* **308**, 377-395.

Cleves A.E., Jain A.N. (2008). Effects of inductive bias on computational evaluations of ligand-based modeling and on drug discovery. *J. Comput. Aided Mol. Des.* **22**, 147-159.

Coffman K., Brodney M., Cook J., Lanyon L., Pandit J., Sakya S., Schachter J., Tseng-Lovering E., Wessel M. (2011). 6-amino-4-(pyrimidin-4-yl)pyridones: novel glycogen synthase kinase-3 $\beta$  inhibitors. *Bioorg. Med. Chem. Lett.* **21**, 1429-1433.

Cole J.C., Korb O., Olsson T.S.G., Liebeschuetz J.. The Basis for Target-Based Virtual Screening: Protein Structures. In *Virtual Screening: Principles, Challenges, and Practical Guidelines*. C. Sotriffer, R. Mannhold, H. Kubinyi, G. Folkers (Ed.). 2011. 87-114.

Cole J.C., Murray C.W., Nissink J.W.M., Taylor R.D., Taylor R. (2005). Comparing protein-ligand docking programs is difficult. *Proteins* **60**, 325-332.

Cole M.D. (1986). The myc oncogene: its role in transformation and differentiation. *Annu. Rev. Genet.* **20**, 361-384.

Cross J.B., Thompson D.C., Rai B.K., Baber J.C., Fan K.Y., Hu Y., Humblet C. (2009). Comparison of several molecular docking programs: pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* **49**, 1455-1474.

Cummings B.S. (2007). Phospholipase A2 as targets for anti-cancer drugs. *Biochem. Pharmacol.* **74**, 949-959.

Cummings M.D., DesJarlais R.L., Gibbs A.C., Mohan V., Jaeger E.P. (2005). Comparison of automated docking programs as virtual screening tools. *J. Med. Chem.* **48**, 962-976.

Cummings M.D., Gibbs A.C., DesJarlais R.L. (2007). Processing of Small Molecule

Databases for Automated Docking. *Medicinal Chemistry* **3**, 107-113.

Cupillard L., Koumanov K., Mattéi M.G., Lazdunski M., Lambeau G. (1997). Cloning, chromosomal mapping, and expression of a novel human secretory phospholipase A2. *J. Biol. Chem.* **272**, 15745-15752.

Curfs D.M.J., Ghesquiere S.A.I., Vergouwe M.N., van der Made I., Gijbels M.J.J., Greaves D.R., Verbeek J.S., Hofker M.H., de Winther M.P.J. (2008). Macrophage secretory phospholipase A2 group X enhances anti-inflammatory responses, promotes lipid accumulation, and contributes to aberrant lung pathology. *J. Biol. Chem.* **283**, 21640-21648.

Dajani R., Fraser E., Roe S.M., Yeo M., Good V.M., Thompson V., Dale T.C., Pearl L.H. (2003). Structural basis for recruitment of glycogen synthase kinase 3beta to the axin-APC scaffold complex. *EMBO J.* **22**, 494-501.

Dajani R., Fraser E., Roe S.M., Young N., Good V., Dale T.C., Pearl L.H. (2001). Crystal structure of glycogen synthase kinase 3 beta: structural basis for phosphate-primed substrate specificity and autoinhibition. *Cell* **105**, 721-732.

Dalby A., Nourse J.G., Hounshell W.D., Gushurst A.K.I., Grier D.L., Leland B.A., Laufer J. (1992). Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited. *J. Chem. Inf. Comput. Sci.* **32**, 244-255.

Danielson M.L., Desai P.V., Mohutsky M.A., Wrighton S.A., Lill M.A. (2011). Potentially increasing the metabolic stability of drug candidates via computational site of metabolism prediction by CYP2C9: The utility of incorporating protein flexibility via an ensemble of structures. *Eur J Med Chem* **46**, 3953-3963.

Davis A.M., Teague S.J. (1999). Hydrogen Bonding, Hydrophobic Interactions, and Failure of the Rigid Receptor Hypothesis. *Angewandte Chemie International Edition* **38**, 736-749.

Davis A.M., Teague S.J., Kleywegt G.J. (2003). Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. *Angew. Chem. Int. Ed. Engl.* **42**, 2718-2736.

de La Coste A., Romagnolo B., Billuart P., Renard C.A., Buendia M.A., Soubrane O., Fabre M., Chelly J., Beldjord C., Kahn A. et al. (1998). Somatic mutations of the beta-catenin gene are frequent in mouse and human hepatocellular carcinomas. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8847-8851.

Dehan E., Ben-Dor A., Liao W., Lipson D., Frimer H., Rienstein S., Simansky D., Krupsky M., Yaron P., Friedman E. et al. (2007). Chromosomal aberrations and gene expression profiles in non-small cell lung cancer. *Lung Cancer* **56**, 175-184.

Deng W., Verlinde C.L.M.J. (2008). Evaluation of different virtual screening programs for docking in a charged binding pocket. *J. Chem. Inf. Model.* **48**, 2010-2020.

- Denizot Y., Chianéa T., Labrousse F., Truffinet V., Delage M., Mathonnet M. (2005). Platelet-activating factor and human thyroid cancer. *Eur. J. Endocrinol.* **153**, 31-40.
- Denizot Y., Gainant A., Guglielmi L., Bouvier S., Cubertafond P., Mathonnet M. (2003). Tissue concentrations of platelet-activating factor in colorectal carcinoma: inverse relationships with Dukes' stage of patients. *Oncogene* **22**, 7222-7224.
- Dennis E.A. (1994). Diversity of group types, regulation, and function of phospholipase A2. *J. Biol. Chem.* **269**, 13057-13060.
- Dennis E.A. (1997). The growing phospholipase A2 superfamily of signal transduction enzymes. *Trends Biochem. Sci.* **22**, 1-2.
- Dessalew N., Bharatam P.V. (2007). Identification of potential glycogen kinase-3 inhibitors by structure based virtual screening. *Biophys. Chem.* **128**, 165-175.
- Dias R., de Azevedo W.F.J. (2008). Molecular docking algorithms. *Curr Drug Targets* **9**, 1040-1047.
- DiMasi J.A., Hansen R.W., Grabowski H.G. (2003). The price of innovation: new estimates of drug development costs. *J Health Econ* **22**, 151-185.
- Dobson C.M. (2004). Chemical space and biology. *Nature* **432**, 824-828.
- Dong Q., Patel M., Scott K.F., Graham G.G., Russell P.J., Sved P. (2006). Oncogenic action of phospholipase A2 in prostate cancer. *Cancer Lett.* **240**, 9-16.
- Dugan K.A., Wood M.A., Cole M.D. (2002). TIP49, but not TRRAP, modulates c-Myc and E2F1 dependent apoptosis. *Oncogene* **21**, 5835-5843.
- Durrant J.D., McCammon J.A. (2010). Computer-aided drug-discovery techniques that account for receptor flexibility. *Curr Opin Pharmacol* **10**, 770-774.
- Dyrskjøt L., Kruhøffer M., Thykjaer T., Marcussen N., Jensen J.L., Møller K., Ørntoft T.F. (2004). Gene expression in the urinary bladder: a common carcinoma in situ gene expression signature exists disregarding histopathological classification. *Cancer Res.* **64**, 4040-4048.
- Eerola L.I., Surrel F., Nevalainen T.J., Gelb M.H., Lambeau G., Laine V.J.O. (2006). Analysis of expression of secreted phospholipases A2 in mouse tissues at protein and mRNA levels. *Biochim. Biophys. Acta* **1761**, 745-756.
- Eldridge M.D., Murray C.W., Auton T.R., Paolini G.V., Mee R.P. (1997). Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput. Aided Mol. Des.* **11**, 425-445.
- Englebienne P., Moitessier N. (2009). Docking ligands into flexible and solvated macromolecules. 4. Are popular scoring functions accurate for this class of proteins?. *J. Chem. Inf. Model.* **49**, 1568-1580.
- Engler T.A., Malhotra S., Burkholder T.P., Henry J.R., Mendel D., Porter W.J., Furness

K., Diefenbacher C., Marquart A., Reel J.K. et al. (2005). The development of potent and selective bisarylmaleimide GSK3 inhibitors. *Bioorg. Med. Chem. Lett.* **15**, 899-903.

Enyedy I.J., Egan W.J. (2008). Can we use docking and scoring for hit-to-lead optimization?. *J. Comput. Aided Mol. Des.* **22**, 161-168.

Escoffier J., Jemel I., Tanemoto A., Taketomi Y., Payre C., Coatrieux C., Sato H., Yamamoto K., Masuda S., Pernet-Gallay K. et al. (2010). Group X phospholipase A2 is released during sperm acrosome reaction and controls fertility outcome in mice. *J. Clin. Invest.* **120**, 1415-1428.

Escoffier J., Pierre V.J., Jemel I., Munch L., Boudhraa Z., Ray P.F., De Waard M., Lambeau G., Arnoult C. (2011). Group X secreted phospholipase A<sub>2</sub> specifically decreases sperm motility in mice. *J. Cell. Physiol.* **226**, 2601-2609.

Eswar N., Webb B., Marti-Renom M.A., Madhusudhan M.S., Eramian D., Shen M., Pieper U., Sali A. (2006). Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics* **Chapter 5**, Unit 5.6.

Etard C., Gradl D., Kunz M., Eilers M., Wedlich D. (2005). Pontin and Reptin regulate cell proliferation in early *Xenopus* embryos in collaboration with c-Myc and Miz-1. *Mech. Dev.* **122**, 545-556.

Ewing T.J., Makino S., Skillman A.G., Kuntz I.D. (2001). DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J. Comput. Aided Mol. Des.* **15**, 411-428.

Ewing T.J.A., Kuntz I.D. (1997). Critical evaluation of search algorithms for automated molecular docking and database screening. *J. Comput. Chem.* **18**, 1175-1189.

Fan H., Irwin J.J., Webb B.M., Klebe G., Shoichet B.K., Sali A. (2009). Molecular docking screens using comparative models of proteins. *J. Chem. Inf. Model.* **49**, 2512-2527.

Farooqui A.A., Ong W., Horrocks L.A. (2006). Inhibitors of brain phospholipase A<sub>2</sub> activity: their neuropharmacological effects and therapeutic importance for the treatment of neurologic disorders. *Pharmacol. Rev.* **58**, 591-620.

Fedorov O., Müller S., Knapp S. (2010). The (un)targeted cancer kinome. *Nat. Chem. Biol.* **6**, 166-169.

Fedorov O., Sundström M., Marsden B., Knapp S. (2007). Insights for the development of specific kinase inhibitors by targeted structural genomics. *Drug Discov. Today* **12**, 365-372.

Feng L., Geisselbrecht Y., Blanck S., Wilbuer A., Atilla-Gokcumen G.E., Filippakopoulos P., Kräling K., Celik M.A., Harms K., Maksimoska J. et al. (2011). Structurally sophisticated octahedral metal complexes as highly selective protein kinase inhibitors. *J. Am. Chem. Soc.* **133**, 5976-5986.

Feng Y., Lee N., Fearon E.R. (2003). TIP49 regulates beta-catenin-mediated neoplastic

transformation and T-cell factor target gene induction via effects on chromatin remodeling. *Cancer Res.* **63**, 8726-8734.

Fiser A., Do R.K., Sali A. (2000). Modeling of loops in protein structures. *Protein Sci.* **9**, 1753-1773.

Flower D.R. (1998). On the Properties of Bit String-Based Measures of Chemical Similarity. *J. Chem. Inf. Comput. Sci.* **38**, 379-386.

Fodde R., Smits R., Clevers H. (2001). APC, signal transduction and genetic instability in colorectal cancer. *Nat. Rev. Cancer* **1**, 55-67.

Fürstner A., Feyen F., Prinz H., Waldmann H. (2003). Total synthesis and reassessment of the phosphatase-inhibitory activity of the antitumor agent TMC-69-6H. *Angew. Chem. Int. Ed. Engl.* **42**, 5361-5364.

Fürstner A., Feyen F., Prinz H., Waldmann H. (2004). Synthesis and evaluation of the antitumor agent TMC-69-6H and a focused library of analogs. *Tetrahedron* **60**, 9543-9558.

Gallant P. (2007). Control of transcription by Pontin and Reptin. *Trends Cell Biol.* **17**, 187-192.

García-Sosa A.T., Mancera R.L., Dean P.M. (2003). WaterScore: a novel method for distinguishing between bound and displaceable water molecules in the crystal structure of the binding site of protein-ligand complexes. *Journal of Molecular Modeling* **9**, 172-182.

Gohlke H., Hendlich M., Klebe G. (2000). Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* **295**, 337-356.

Good A.C., Cheney D.L., Sitkoff D.F., Tokarski J.S., Stouch T.R., Bassolino D.A., Krystek S.R., Li Y., Mason J.S., Perkins T.D.J. (2003). Analysis and optimization of structure-based virtual screening protocols. 2. Examination of docked ligand orientation sampling methodology: mapping a pharmacophore for success. *J. Mol. Graph. Model.* **22**, 31-40.

Good A.C., Oprea T.I. (2008). Optimization of CAMD techniques 3. Virtual screening enrichment studies: a help or hindrance in tool selection?. *J. Comput. Aided Mol. Des.* **22**, 169-178.

Goodford P.J. (1985). A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **28**, 849-857.

Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. (1989). Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res.* **17**, 4713-4730.

Gorynia S., Bandejas T.M., Pinho F.G., McVey C.E., Vorrhein C., Round A., Svergun D.I., Donner P., Matias P.M., Carrondo M.A. (2011). Structural and functional insights into a dodecameric molecular machine - The RuvBL1/RuvBL2 complex. *J. Struct. Biol.* [Epub

ahead of print] DOI:10.1016/j.jsb.2011.09.001.

Graff J.R., Konicek B.W., Deddens J.A., Chedid M., Hurst B.M., Colligan B., Neubauer B.L., Carter H.W., Carter J.H. (2001). Expression of group IIa secretory phospholipase A2 increases with prostate tumor grade. *Clin. Cancer Res.* **7**, 3857-3861.

Graudens E., Boulanger V., Mollard C., Mariage-Samson R., Barlet X., Grémy G., Couillault C., Lajémi M., Piatier-Tonneau D., Zaborski P. et al. (2006). Deciphering cellular states of innate tumor drug responses. *Genome Biol.* **7**, R19.

Gribun A., Cheung K.L.Y., Huen J., Ortega J., Houry W.A. (2008). Yeast Rvb1 and Rvb2 are ATP-dependent DNA helicases that form a heterohexameric complex. *J. Mol. Biol.* **376**, 1320-1333.

Grigoletto A., Lestienne P., Rosenbaum J. (2011). The multifaceted proteins Reptin and Pontin as major players in cancer. *Biochim. Biophys. Acta* **1815**, 147-157.

Guha R., Howard M.T., Hutchison G.R., Murray-Rust P., Rzepa H., Steinbeck C., Wegner J., Willighagen E.L. (2006). The Blue Obelisk-interoperability in chemical informatics. *J. Chem. Inf. Model.* **46**, 991-998.

Halgren T.A. (1996). Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *J. Comput. Chem.* **17**, 490-519.

Hallstrand T.S., Chi E.Y., Singer A.G., Gelb M.H., Henderson W.R.J. (2007). Secreted phospholipase A2 group X overexpression in asthma and bronchial hyperresponsiveness. *Am. J. Respir. Crit. Care Med.* **176**, 1072-1078.

Haslinger C., Schweifer N., Stilgenbauer S., Döhner H., Lichter P., Kraut N., Stratowa C., Abseher R. (2004). Microarray gene expression profiling of B-cell chronic lymphocytic leukemia subgroups defined by genomic aberrations and VH mutation status. *J. Clin. Oncol.* **22**, 3937-3949.

Haurie V., Ménard L., Nicou A., Touriol C., Metzler P., Fernandez J., Taras D., Lestienne P., Balabaud C., Bioulac-Sage P. et al. (2009). Adenosine triphosphatase pontin is overexpressed in hepatocellular carcinoma and coregulated with reptin through a new posttranslational mechanism. *Hepatology* **50**, 1871-1883.

Hawkins P.C.D., Skillman A.G., Warren G.L., Ellingson B.A., Stahl M.T. (2010). Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **50**, 572-584.

Hawkins P.C.D., Warren G.L., Skillman A.G., Nicholls A. (2008). How to do an evaluation: pitfalls and traps. *J. Comput. Aided Mol. Des.* **22**, 179-190.

Henderson W.R.J., Chi E.Y., Bollinger J.G., Tien Y., Ye X., Castelli L., Rubtsov Y.P., Singer A.G., Chiang G.K.S., Nevalainen T. et al. (2007). Importance of group X-secreted

phospholipase A2 in allergen-induced airway inflammation and remodeling in a mouse asthma model. *J. Exp. Med.* **204**, 865-877.

Hirano N., Kohno J., Tsunoda S., Nishio M., Kishi N., Okuda T., Kawano K., Komatsubara S., Nakanishi N. (2001). TMC-69, a new antitumor antibiotic with Cdc25A inhibitory activity, produced by *Chrysosporium* sp. TC1068. Taxonomy, fermentation and biological activities. *J. Antibiot.* **54**, 421-427.

Hishida T., Han Y., Fujimoto S., Iwasaki H., Shinagawa H. (2004). Direct evidence that a conserved arginine in RuvB AAA+ ATPase acts as an allosteric effector for the ATPase activity of the adjacent subunit in a hexamer. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 9573-9577.

Huang N., Shoichet B.K., Irwin J.J. (2006). Benchmarking sets for molecular docking. *J. Med. Chem.* **49**, 6789-6801.

Huang S., Grinter S.Z., Zou X. (2010). Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. *Phys Chem Chem Phys* **12**, 12899-12908.

Hubbard R.E. (2005). 3D structure and the drug-discovery process. *Mol. BioSyst* **1**, 391-406.

Huber O., Ménard L., Haurie V., Nicou A., Taras D., Rosenbaum J. (2008). Pontin and reptin, two related ATPases with multiple roles in cancer. *Cancer Res.* **68**, 6873-6876.

Huen J., Kakihara Y., Uguw F., Cheung K.L.Y., Ortega J., Houry W.A. (2010). Rvb1-Rvb2: essential ATP-dependent helicases for critical complexes. *Biochem. Cell Biol.* **88**, 29-40.

Huggins K.W., Boileau A.C., Hui D.Y. (2002). Protection against diet-induced obesity and obesity-related insulin resistance in Group 1B PLA2-deficient mice. *Am. J. Physiol. Endocrinol. Metab.* **283**, E994-E1001.

Iizuka N., Tsunedomi R., Tamesa T., Okada T., Sakamoto K., Hamaguchi T., Yamada-Okabe H., Miyamoto T., Uchimura S., Hamamoto Y. et al. (2006). Involvement of c-myc-regulated genes in hepatocellular carcinoma related to genotype-C hepatitis B virus. *J. Cancer Res. Clin. Oncol.* **132**, 473-481.

Ikura T., Ogryzko V.V., Grigoriev M., Groisman R., Wang J., Horikoshi M., Scully R., Qin J., Nakatani Y. (2000). Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis. *Cell* **102**, 463-473.

Irwin J.J. (2008). Community benchmarks for virtual screening. *J. Comput. Aided Mol. Des.* **22**, 193-199.

Jain A.N. (2008). Bias, reporting, and sharing: computational evaluations of docking methods. *J. Comput. Aided Mol. Des.* **22**, 201-212.

Jain A.N., Nicholls A. (2008). Recommendations for evaluation of computational

methods. *J. Comput. Aided Mol. Des.* **22**, 133-139.

Jenkins J.L., Kao R.Y.T., Shapiro R. (2003). Virtual screening to enrich hit lists from high-throughput screening: a case study on small-molecule inhibitors of angiogenin. *Proteins* **50**, 81-93.

Jha S., Dutta A. (2009). RVB1/RVB2: running rings around molecular biology. *Mol. Cell* **34**, 521-533.

Jha S., Shibata E., Dutta A. (2008). Human Rvb1/Tip49 is required for the histone acetyltransferase activity of Tip60/NuA4 and for the downregulation of phosphorylation on H2AX after DNA damage. *Mol. Cell. Biol.* **28**, 2690-2700.

Jiang F., Kim S.H. (1991). "Soft docking": matching of molecular surface cubes. *J. Mol. Biol.* **219**, 79-102.

Jiang J., Neubauer B.L., Graff J.R., Chedid M., Thomas J.E., Roehm N.W., Zhang S., Eckert G.J., Koch M.O., Eble J.N. et al. (2002). Expression of group IIA secretory phospholipase A2 is elevated in prostatic intraepithelial neoplasia and adenocarcinoma. *Am. J. Pathol.* **160**, 667-671.

Jin J., Cai Y., Yao T., Gottschalk A.J., Florens L., Swanson S.K., Gutiérrez J.L., Coleman M.K., Workman J.L., Mushegian A. et al. (2005). A mammalian chromatin remodeling complex with similarities to the yeast INO80 complex. *J. Biol. Chem.* **280**, 41207-41212.

Jones G., Willett P., Glen R.C. (1995). Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* **245**, 43-53.

Jones G., Willett P., Glen R.C., Leach A.R., Taylor R. (1997). Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **267**, 727-748.

Jónsson Z.O., Dhar S.K., Narlikar G.J., Auty R., Wagle N., Pellman D., Pratt R.E., Kingston R., Dutta A. (2001). Rvb1p and Rvb2p are essential components of a chromatin remodeling complex that regulates transcription of over 5% of yeast genes. *J. Biol. Chem.* **276**, 16279-16288.

Jónsson Z.O., Jha S., Wohlschlegel J.A., Dutta A. (2004). Rvb1p/Rvb2p recruit Arp5p and assemble a functional Ino80 chromatin remodeling complex. *Mol. Cell* **16**, 465-477.

Kairys V., Fernandes M.X., Gilson M.K. (2006). Screening drug-like compounds by docking to homology models: a systematic study. *J. Chem. Inf. Model.* **46**, 365-379.

Kallajoki M., Alanen K.A., Nevalainen M., Nevalainen T.J. (1998). Group II phospholipase A2 in human male reproductive organs and genital tumors. *Prostate* **35**, 263-272.

Kalliokoski T., Salo H.S., Lahtela-Kakkonen M., Poso A. (2009). The effect of ligand-based tautomer and protomer prediction on structure-based virtual screening. *J. Chem. Inf.*

Model. **49**, 2742-2748.

Kanemaki M., Kurokawa Y., Matsu-ura T., Makino Y., Masani A., Okazaki K., Morishita T., Tamura T.A. (1999). TIP49b, a new RuvB-like DNA helicase, is included in a complex together with another RuvB-like DNA helicase, TIP49a. *J. Biol. Chem.* **274**, 22437-22444.

Kanemaki M., Makino Y., Yoshida T., Kishimoto T., Koga A., Yamamoto K., Yamamoto M., Moncollin V., Egly J.M., Muramatsu M. et al. (1997). Molecular cloning of a rat 49-kDa TBP-interacting protein (TIP49) that is highly homologous to the bacterial RuvB. *Biochem. Biophys. Res. Commun.* **235**, 64-68.

Karabina S., Brochériou I., Le Naour G., Agrapart M., Durand H., Gelb M., Lambeau G., Ninio E. (2006). Atherogenic properties of LDL particles modified by human group X secreted phospholipase A2 on human endothelial cell function. *FASEB J.* **20**, 2547-2549.

Katritzky A.R., Pacureanu L.M., Dobchev D.A., Fara D.C., Duchowicz P.R., Karelson M. (2006). QSAR modeling of the inhibition of glycogen synthase kinase-3. *Bioorg. Med. Chem.* **14**, 4987-5002.

Kearsley S.K., Underwood D.J., Sheridan R.P., Miller M.D. (1994). Flexibases: a way to enhance the use of molecular docking methods. *J. Comput. Aided Mol. Des.* **8**, 565-582.

Keiser M.J., Setola V., Irwin J.J., Laggner C., Abbas A.I., Hufeisen S.J., Jensen N.H., Kuiler M.B., Matos R.C., Tran T.B. et al. (2009). Predicting new molecular targets for known drugs. *Nature* **462**, 175-182.

Kellenberger E., Rodrigo J., Muller P., Rognan D. (2004). Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins* **57**, 225-242.

Ki D.H., Jeung H., Park C.H., Kang S.H., Lee G.Y., Lee W.S., Kim N.K., Chung H.C., Rha S.Y. (2007). Whole genome analysis for liver metastasis gene signatures in colorectal cancer. *Int. J. Cancer* **121**, 2005-2012.

Kim H., Choo H., Cho Y.S., No K.T., Pae A.N. (2008). Novel GSK-3beta inhibitors from sequential virtual screening. *Bioorg. Med. Chem.* **16**, 636-643.

Kim J., Chakrabarti B.K., Guha-Niyogi A., Louder M.K., Mascola J.R., Ganesh L., Nabel G.J. (2007). Lysis of human immunodeficiency virus type 1 by a specific secreted human phospholipase A2. *J. Virol.* **81**, 1444-1450.

King T.H., Decatur W.A., Bertrand E., Maxwell E.S., Fournier M.J. (2001). A well-connected and conserved nucleoplasmic helicase is required for production of box C/D and H/ACA snoRNAs and localization of snoRNP proteins. *Mol. Cell. Biol.* **21**, 7731-7746.

Kirchmair J., Markt P., Distinto S., Wolber G., Langer T. (2008). Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection--what can we learn from earlier mistakes?. *J. Comput. Aided Mol. Des.* **22**, 213-228.

- Kitchen D.B., Decornez H., Furr J.R., Bajorath J. (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* **3**, 935-949.
- Kiyohara H., Egami H., Kako H., Shibata Y., Murata K., Ohshima S., Sei K., Suko S., Kurano R., Ogawa M. (1993). Immunohistochemical localization of group II phospholipase A2 in human pancreatic carcinomas. *Int. J. Pancreatol.* **13**, 49-57.
- Knox A.J.S., Meegan M.J., Carta G., Lloyd D.G. (2005). Considerations in Compound Database Preparations“Hidden” Impact on Virtual Screening Results. *J. Chem. Inf. Model* **45**, 1908-1919.
- Koduri R.S., Grönroos J.O., Laine V.J.O., Le Calvez C., Lambeau G., Nevalainen T.J., Gelb M.H. (2002). Bactericidal properties of human and murine groups I, II, V, X, and XII secreted phospholipases A(2). *J. Biol. Chem.* **277**, 5849-5857.
- Kohno J., Hirano N., Sugawara K., Nishio M., Hashiyama T., Nakanishi N., Komatsubara S. (2001). Structure of TMC-69, a new antitumor antibiotic from *Chrysosporium* sp. TC 1068. *Tetrahedron* **57**, 1731-1735.
- Kolligs F.T., Hu G., Dang C.V., Fearon E.R. (1999). Neoplastic transformation of RK3E by mutant beta-catenin requires deregulation of Tcf/Lef transcription but not activation of c-myc expression. *Mol. Cell. Biol.* **19**, 5696-5706.
- Koonin E.V. (1993). A common set of conserved motifs in a vast variety of putative nucleic acid-dependent ATPases including MCM proteins involved in the initiation of eukaryotic DNA replication. *Nucleic Acids Res.* **21**, 2541-2547.
- Kramer B., Rarey M., Lengauer T. (1999). Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. *Proteins* **37**, 228-241.
- Kryshtafovych A., Fidelis K. (2009). Protein structure prediction and model quality assessment. *Drug Discov. Today* **14**, 386-393.
- Kukol A. (2011). Consensus virtual screening approaches to predict protein ligands. *Eur J Med Chem* **46**, 4661-4664.
- Kunick C., Lauenroth K., Wieking K., Xie X., Schultz C., Gussio R., Zaharevitz D., Leost M., Meijer L., Weber A. et al. (2004). Evaluation and comparison of 3D-QSAR CoMSIA models for CDK1, CDK5, and GSK-3 inhibition by paullones. *J. Med. Chem.* **47**, 22-36.
- Kuntz I.D., Blaney J.M., Oatley S.J., Langridge R., Ferrin T.E. (1982). A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **161**, 269-288.
- Kuo G., Prouty C., DeAngelis A., Shen L., O'Neill D.J., Shah C., Connolly P.J., Murray W.V., Conway B.R., Cheung P. et al. (2003). Synthesis and discovery of macrocyclic polyoxygenated bis-7-azaindolyImaleimides as a novel series of potent and highly selective glycogen synthase kinase-3beta inhibitors. *J. Med. Chem.* **46**, 4021-4031.
- Kurogi Y., Güner O.F. (2001). Pharmacophore modeling and three-dimensional

database searching for drug design using catalyst. *Curr Med Chem* **8**, 1035-1055.

Lambeau G., Gelb M.H. (2008). Biochemistry and physiology of mammalian secreted phospholipases A2. *Annu. Rev. Biochem.* **77**, 495-520.

Lang P.T., Brozell S.R., Mukherjee S., Pettersen E.F., Meng E.C., Thomas V., Rizzo R.C., Case D.A., James T.L., Kuntz I.D. (2009). DOCK 6: combining techniques to model RNA-small molecule complexes.. *Rna New York Ny* **15**, 1219-1230.

Lauscher J.C., Loddenkemper C., Kosel L., Gröne J., Buhr H.J., Huber O. (2007). Increased pontin expression in human colorectal cancer tissue. *Hum. Pathol.* **38**, 978-985.

Laye J.P., Gill J.H. (2003). Phospholipase A2 expression in tumours: a target for therapeutic intervention?. *Drug Discov. Today* **8**, 710-716.

Leach A.R. (1994). Ligand docking to proteins with discrete side-chain flexibility. *J. Mol. Biol.* **235**, 345-356.

Leach A.R., Shoichet B.K., Peishoff C.E. (2006). Prediction of protein-ligand interactions. Docking and scoring: successes and gaps. *J. Med. Chem.* **49**, 5851-5855.

Leclerc S., Garnier M., Hoessel R., Marko D., Bibb J.A., Snyder G.L., Greengard P., Biernat J., Wu Y.Z., Mandelkow E.M.et al. (2001). Indirubins inhibit glycogen synthase kinase-3 beta and CDK5/p25, two protein kinases involved in abnormal tau phosphorylation in Alzheimer's disease. A property common to most cyclin-dependent kinase inhibitors?. *J. Biol. Chem.* **276**, 251-260.

Leeson P.D., Springthorpe B. (2007). The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat Rev Drug Discov* **6**, 881-890.

Lemeer S., Heck A.J.R. (2009). The phosphoproteomics data explosion. *Curr. Opin. Chem. Biol.* **13**, 414-420.

Lena G., Lallemand E., Gruner A.C., Boeglin J., Roussel S., Schaffner A., Aubry A., Franetich J., Mazier D., Landau I.et al. (2006). 1,3,5-Triazepane-2,6-diones as structurally diverse and conformationally constrained dipeptide mimetics: identification of malaria liver stage inhibitors from a small pilot library. *Chemistry* **12**, 8498-8512.

Leost M., Schultz C., Link A., Wu Y.Z., Biernat J., Mandelkow E.M., Bibb J.A., Snyder G.L., Greengard P., Zaharevitz D.W.et al. (2000). Paullones are potent inhibitors of glycogen synthase kinase-3beta and cyclin-dependent kinase 5/p25. *Eur. J. Biochem.* **267**, 5983-5994.

Li C., Tan Y., Zhou H., Ding S., Li S., Ma D., Man X., Hong Y., Zhang L., Li L.et al. (2005). Proteomic analysis of hepatitis B virus-associated hepatocellular carcinoma: Identification of potential tumor markers. *Proteomics* **5**, 1125-1139.

Li J., Ehlers T., Sutter J., Varma-O'brien S., Kirchmair J. (2007). CAESAR: a new conformer generation algorithm based on recursive buildup and local rotational symmetry consideration. *J. Chem. Inf. Model.* **47**, 1923-1932.

- Li W., Zeng J., Li Q., Zhao L., Liu T., Björkholm M., Jia J., Xu D. (2010). Reptin is required for the transcription of telomerase reverse transcriptase and over-expressed in gastric cancer. *Mol. Cancer* **9**, 132.
- Liebescu J.W. (2008). Evaluating docking programs: keeping the playing field level. *J. Comput. Aided Mol. Des.* **22**, 229-238.
- Lill M.A., Danielson M.L. (2011). Computer-aided drug design platform using PyMOL. *J. Comput. Aided Mol. Des.* **25**, 13-19.
- Lim C.R., Kimata Y., Ohdate H., Kokubo T., Kikuchi N., Horigome T., Kohno K. (2000). The *Saccharomyces cerevisiae* RuvB-like protein, Tih2p, is required for cell cycle progression and RNA polymerase II-directed transcription. *J. Biol. Chem.* **275**, 22409-22417.
- Lipinski C.A., Lombardo F., Dominy B.W., Feeney P.J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **23**, 3-25.
- Liu H., Kuntz I.D., Zou X. (2004). Pairwise GB/SA Scoring Function for Structure-based Drug Design. *The Journal of Physical Chemistry B* **108**, 5453-5462.
- Lombardino J.G., Lowe J.A.3. (2004). The role of the medicinal chemist in drug discovery--then and now. *Nat Rev Drug Discov* **3**, 853-862.
- Lupas A.N., Martin J. (2002). AAA proteins. *Curr. Opin. Struct. Biol.* **12**, 746-753.
- Makino Y., Kanemaki M., Kurokawa Y., Koji T., Tamura T.A. (1999). A rat RuvB-like protein, TIP49a, is a germ cell-enriched novel DNA helicase. *J. Biol. Chem.* **274**, 15329-15335.
- Makino Y., Mimori T., Koike C., Kanemaki M., Kurokawa Y., Inoue S., Kishimoto T., Tamura T. (1998). TIP49, homologous to the bacterial DNA helicase RuvB, acts as an autoantigen in human. *Biochem. Biophys. Res. Commun.* **245**, 819-823.
- Maldonado-Rojas W., Olivero-Verbel J. (2011). Potential interaction of natural dietary bioactive compounds with COX-2. *J. Mol. Graph. Model.* **30**, 157-166.
- Manning G., Whyte D.B., Martinez R., Hunter T., Sudarsanam S. (2002). The protein kinase complement of the human genome. *Science* **298**, 1912-1934.
- Margutti S., Laufer S.A. (2007). Are MAP kinases drug targets? Yes, but difficult ones. *ChemMedChem* **2**, 1116-1140.
- Marks F., Müller-Decker K., Fürstenberger G. (2000). A causal relationship between unscheduled eicosanoid signaling and tumor development: cancer chemoprevention by inhibitors of arachidonic acid metabolism. *Toxicology* **153**, 11-26.
- Martí-Renom M.A., Stuart A.C., Fiser A., Sánchez R., Melo F., Sali A. (2000). Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* **29**, 291-325.

Martin L., Catherinot V., Labesse G. (2006). kinDOCK: a tool for comparative docking of protein kinase ligands. *Nucleic Acids Res.* **34**, W325-9.

Martin Y.C. (2009). Let's not forget tautomers. *J. Comput. Aided Mol. Des.* **23**, 693-704.

Maslon M.M., Hrstka R., Vojtesek B., Hupp T.R. (2010). A divergent substrate-binding loop within the pro-oncogenic protein anterior gradient-2 forms a docking site for Reptin. *J. Mol. Biol.* **404**, 418-438.

Masuda S., Murakami M., Matsumoto S., Eguchi N., Urade Y., Lambeau G., Gelb M.H., Ishikawa Y., Ishii T., Kudo I. (2004). Localization of various secretory phospholipase A2 enzymes in male reproductive organs. *Biochim. Biophys. Acta* **1686**, 61-76.

Masuda S., Murakami M., Takanezawa Y., Aoki J., Arai H., Ishikawa Y., Ishii T., Arioka M., Kudo I. (2005). Neuronal expression and neuritogenic action of group X secreted phospholipase A2. *J. Biol. Chem.* **280**, 23203-23214.

Matias P.M., Gorynia S., Donner P., Carrondo M.A. (2006). Crystal structure of the human AAA+ protein RuvBL1. *J. Biol. Chem.* **281**, 38918-38929.

Matter H., Sotriffer C.. Applications and Success Stories in Virtual Screening. In *Virtual Screening: Principles, Challenges, and Practical Guidelines*. C. Sotriffer, R. Mannhold, H. Kubinyi, G. Folkers (Ed.). 2011. 319-358.

Mazanetz M.P., Withers I.M., Laughton C.A., Fischer P.M. (2008). Exploiting glycogen synthase kinase 3beta flexibility in molecular recognition. *Biochem. Soc. Trans.* **36**, 55-58.

McGovern S.L., Shoichet B.K. (2003). Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. *J. Med. Chem.* **46**, 2895-2907.

Meijer L., Skaltsounis A., Magiatis P., Polychronopoulos P., Knockaert M., Leost M., Ryan X.P., Vonica C.A., Brivanlou A., Dajani R. et al. (2003). GSK-3-selective inhibitors derived from Tyrian purple indirubins. *Chem. Biol.* **10**, 1255-1266.

Ménard L., Taras D., Grigoletto A., Haurie V., Nicou A., Dugot-Senant N., Costet P., Rousseau B., Rosenbaum J. (2010). In vivo silencing of Reptin blocks the progression of human hepatocellular carcinoma in xenografts and is associated with replicative senescence. *J. Hepatol.* **52**, 681-689.

Meng E.C., Shoichet B.K., Kuntz I.D. (1992). Automated docking with grid-based energy evaluation. *J. Comput. Chem.* **13**, 505-524.

Menichincheri M., Bargiotti A., Berthelsen J., Bertrand J.A., Bossi R., Ciavoletta A., Cirila A., Cristiani C., Croci V., D'Alessio R. et al. (2009). First Cdc7 kinase inhibitors: pyrrolopyridinones as potent and orally active antitumor agents. 2. Lead discovery. *J. Med. Chem.* **52**, 293-307.

Menschikowski M., Hagelgans A., Gussakovsky E., Kostka H., Paley E.L., Siegert G. (2008). Differential expression of secretory phospholipases A2 in normal and malignant

prostate cell lines: regulation by cytokines, cell signaling pathways, and epigenetic mechanisms. *Neoplasia* **10**, 279-286.

Mézard C., Davies A.A., Stasiak A., West S.C. (1997). Biochemical properties of RuvBD113N: a mutation in helicase motif II of the RuvB hexamer affects DNA binding and ATPase activities. *J. Mol. Biol.* **271**, 704-717.

Miller M.D., Kearsley S.K., Underwood D.J., Sheridan R.P. (1994). FLOG: a system to select 'quasi-flexible' ligands complementary to a receptor of known three-dimensional structure. *J. Comput. Aided Mol. Des.* **8**, 153-174.

Mitsuishi M., Masuda S., Kudo I., Murakami M. (2006). Group V and X secretory phospholipase A2 prevents adenoviral infection in mammalian cells. *Biochem. J.* **393**, 97-106.

Mizuguchi G., Shen X., Landry J., Wu W., Sen S., Wu C. (2004). ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science* **303**, 343-348.

Moffitt J.R., Chemla Y.R., Aathavan K., Grimes S., Jardine P.J., Anderson D.L., Bustamante C. (2009). Intersubunit coordination in a homomeric ring ATPase. *Nature* **457**, 446-450.

Mohan V., Gibbs A.C., Cummings M.D., Jaeger E.P., DesJarlais R.L. (2005). Docking: successes and challenges. *Curr. Pharm. Des.* **11**, 323-333.

Moitessier N., Englebienne P., Lee D., Lawandi J., Corbeil C.R. (2008). Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *Br. J. Pharmacol.* **153 Suppl 1**, S7-26.

Morin P.J. (1999). beta-catenin signaling and cancer. *Bioessays* **21**, 1021-1030.

Morioka Y., Ikeda M., Saiga A., Fujii N., Ishimoto Y., Arita H., Hanasaki K. (2000). Potential role of group X secretory phospholipase A(2) in cyclooxygenase-2-dependent PGE(2) formation during colon tumorigenesis. *FEBS Lett.* **487**, 262-266.

Morris G.M., Goodsell D.S., Halliday R.S., Huey R., Hart W.E., Belew R.K., Olson A.J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **19**, 1639-1662.

Mounier C.M., Bon C., Kini R.M. (2001). Anticoagulant venom and mammalian secreted phospholipases A(2): protein- versus phospholipid-dependent mechanism of action. *Haemostasis* **31**, 279-287.

Moustakas D.T., Lang P.T., Pegg S., Pettersen E., Kuntz I.D., Brooijmans N., Rizzo R.C. (2006). Development and validation of a modular, extensible docking program: DOCK 5. *J. Comput. Aided Mol. Des.* **20**, 601-619.

Muegge I. (2006). PMF scoring revisited. *J. Med. Chem.* **49**, 5895-5902.

- Muegge I., Martin Y.C. (1999). A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J. Med. Chem.* **42**, 791-804.
- Muller P., Lena G., Boilard E., Bezzine S., Lambeau G., Guichard G., Rognan D. (2006). In silico-guided target identification of a scaffold-focused library: 1,3,5-triazepan-2,6-diones as novel phospholipase A2 inhibitors. *J. Med. Chem.* **49**, 6768-6778.
- Murakami M. (2004). Hot topics in phospholipase A2 field. *Biol. Pharm. Bull.* **27**, 1179-1182.
- Murakami M., Kudo I. (2004). Secretory phospholipase A2. *Biol. Pharm. Bull.* **27**, 1158-1164.
- Musa M.A., Cooperwood J.S., Khan M.O.F., Rahman T. (2011). In-vitro Antiproliferative Activity of Benzopyranone Derivatives in Comparison with Standard Chemotherapeutic Drugs. *Arch. Pharm. Chem. Life Sci.* **2**, 102-110.
- Naerum L., Nørskov-Lauritsen L., Olesen P.H. (2002). Scaffold hopping and optimization towards libraries of glycogen synthase kinase-3 inhibitors. *Bioorg. Med. Chem. Lett.* **12**, 1525-1528.
- Natarajan R., Nadler J.L. (2004). Lipid inflammatory mediators in diabetic vascular disease. *Arterioscler. Thromb. Vasc. Biol.* **24**, 1542-1548.
- Neuwald A.F., Aravind L., Spouge J.L., Koonin E.V. (1999). AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res.* **9**, 27-43.
- Nicholls A. (2008). What do we know and when do we know it?. *J. Comput. Aided Mol. Des.* **22**, 239-255.
- Nishiu M., Yanagawa R., Nakatsuka S., Yao M., Tsunoda T., Nakamura Y., Aozasa K. (2002). Microarray analysis of gene-expression profiles in diffuse large B-cell lymphoma: identification of genes related to disease progression. *Jpn. J. Cancer Res.* **93**, 894-901.
- Nocedal J., Wright S.J.. Quasi-Newton Methods. In *Numerical Optimization*. T. V. Mikosch, S. I. Resnick, S. M. Robinson (Ed.). 2006. 135-163.
- Nusse R. (2005). Wnt signaling in disease and in development. *Cell Res.* **15**, 28-32.
- O'Boyle N.M., Banck M., James C.A., Morley C., Vandermeersch T., Hutchison G.R. (2011). Open Babel: An open chemical toolbox. *J. Cheminform* **3**, 33.
- Ogura T., Whiteheart S.W., Wilkinson A.J. (2004). Conserved arginine residues implicated in ATP hydrolysis, nucleotide-sensing, and inter-subunit interactions in AAA and AAA+ ATPases. *J. Struct. Biol.* **146**, 106-112.
- Ogura T., Wilkinson A.J. (2001). AAA+ superfamily ATPases: common structure-diverse function. *Genes Cells* **6**, 575-597.
- Oprea T.I., Matter H. (2004). Integrating virtual screening in lead discovery. *Curr. Opin.*

Chem. Biol. **8**, 349-358.

Oshiro C., Bradley E.K., Eksterowicz J., Evensen E., Lamb M.L., Lanctot J.K., Putta S., Stanton R., Grootenhuis P.D.J. (2004). Performance of 3D-database molecular docking studies into homology models. *J. Med. Chem.* **47**, 764-767.

Oslund R.C., Cermak N., Gelb M.H. (2008). Highly specific and broadly potent inhibitors of mammalian secreted phospholipases A2. *J. Med. Chem.* **51**, 4708-4714.

Owens J. (2007). Determining druggability. *Nat Rev Drug Discov* **6**, 187-187.

Packard C.J., O'Reilly D.S., Caslake M.J., McMahon A.D., Ford I., Cooney J., Macphee C.H., Suckling K.E., Krishna M., Wilkinson F.E. et al. (2000). Lipoprotein-associated phospholipase A2 as an independent predictor of coronary heart disease. West of Scotland Coronary Prevention Study Group. *N. Engl. J. Med.* **343**, 1148-1155.

Palermo N.Y., Natarajan A. (2011). Beyond the frog: the evolution of homology models of human IKK $\beta$ . *Bioorg. Med. Chem. Lett.* **21**, 6081-6084.

Pan Y.H., Yu B., Singer A.G., Ghomashchi F., Lambeau G., Gelb M.H., Jain M.K., Bahnon B.J. (2002). Crystal structure of human group X secreted phospholipase A2. Electrostatically neutral interfacial surface targets zwitterionic membranes. *J. Biol. Chem.* **277**, 29086-29093.

Peat A.J., Boucheron J.A., Dickerson S.H., Garrido D., Mills W., Peckham J., Preugschat F., Smalley T., Schweiker S.L., Wilson J.R. et al. (2004a). Novel pyrazolopyrimidine derivatives as GSK-3 inhibitors. *Bioorg. Med. Chem. Lett.* **14**, 2121-2125.

Peat A.J., Garrido D., Boucheron J.A., Schweiker S.L., Dickerson S.H., Wilson J.R., Wang T.Y., Thomson S.A. (2004b). Novel GSK-3 inhibitors with improved cellular activity. *Bioorg. Med. Chem. Lett.* **14**, 2127-2130.

Peifer M., Polakis P. (2000). Wnt signaling in oncogenesis and embryogenesis--a look outside the nucleus. *Science* **287**, 1606-1609.

Perola E., Charifson P.S. (2004). Conformational analysis of drug-like molecules bound to proteins: an extensive study of ligand reorganization upon binding. *J. Med. Chem.* **47**, 2499-2510.

Perola E., Walters W.P., Charifson P.S. (2004). A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins* **56**, 235-249.

Peterson M.W., Fairchild S.Z., Otto T.C., Mohtashemi M., Cerasoli D.M., Chang W.E. (2011). VX hydrolysis by human serum paraoxonase 1: a comparison of experimental and computational results. *PLoS ONE* **6**, e20335.

Polakis P. (2000). Wnt signaling and cancer. *Genes Dev.* **14**, 1837-1851.

Polychronopoulos P., Magiatis P., Skaltsounis A., Myrianthopoulos V., Mikros E.,

Tarricone A., Musacchio A., Roe S.M., Pearl L., Leost M. et al. (2004). Structural basis for the synthesis of indirubins as potent and selective inhibitors of glycogen synthase kinase-3 and cyclin-dependent kinases. *J. Med. Chem.* **47**, 935-946.

Puri T., Wendler P., Sigala B., Saibil H., Tsaneva I.R. (2007). Dodecameric structure and ATPase activity of the human TIP48/TIP49 complex. *J. Mol. Biol.* **366**, 179-192.

Qiu X.B., Lin Y.L., Thome K.C., Pian P., Schlegel B.P., Weremowicz S., Parvin J.D., Dutta A. (1998). An eukaryotic RuvB-like protein (RUVBL1) essential for growth. *J. Biol. Chem.* **273**, 27786-27793.

Rader D.J. (2000). Inflammatory markers of coronary risk. *N. Engl. J. Med.* **343**, 1179-1182.

Radovic S., Rapisarda V.A., Tosato V., Bruschi C.V. (2007). Functional and comparative characterization of *Saccharomyces cerevisiae* RVB1 and RVB2 genes with bacterial Ruv homologues. *FEMS Yeast Res.* **7**, 527-539.

Rarey M., Kramer B., Lengauer T., Klebe G. (1996). A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **261**, 470-489.

Richardson C.M., Nunns C.L., Williamson D.S., Parratt M.J., Dokurno P., Howes R., Borgognoni J., Drysdale M.J., Finch H., Hubbard R.E. et al. (2007). Discovery of a potent CDK2 inhibitor with a novel binding mode, using virtual screening and initial, structure-guided lead scoping. *Bioorg. Med. Chem. Lett.* **17**, 3880-3885.

Rocchia W., Sridharan S., Nicholls A., Alexov E., Chiabrera A., Honig B. (2002). Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects. *J. Comput. Chem.* **23**, 128-137.

Rockey W.M., Elcock A.H. (2005). Rapid computational identification of the targets of protein kinase inhibitors. *J. Med. Chem.* **48**, 4138-4152.

Rockey W.M., Elcock A.H. (2006). Structure selection for protein kinase docking and virtual screening: homology models or crystal structures?. *Curr. Protein Pept. Sci.* **7**, 437-457.

Rose P.W., Beran B., Bi C., Bluhm W.F., Dimitropoulos D., Goodsell D.S., Prlic A., Quesada M., Quinn G.B., Westbrook J.D. et al. (2011). The RCSB Protein Data Bank: redesigned web site and web services. *Nucleic Acids Res.* **39**, D392-401.

Rousseau B., Ménard L., Haurie V., Taras D., Blanc J., Moreau-Gaudry F., Metzler P., Hugues M., Boyault S., Lemièrre S. et al. (2007). Overexpression and role of the ATPase and putative DNA helicase RuvB-like 2 in human hepatocellular carcinoma. *Hepatology* **46**, 1108-1118.

Sadowski J., Gasteiger J., Klebe G. (1994). Comparison of Automatic Three-

Dimensional Model Builders Using 639 X-ray Structures. *J. Chem. Inf. Comput. Sci.* **34**, 1000-1008.

Saitoh M., Kunitomo J., Kimura E., Iwashita H., Uno Y., Onishi T., Uchiyama N., Kawamoto T., Tanaka T., Mol C.D. et al. (2009). 2-{3-[4-(Alkylsulfinyl)phenyl]-1-benzofuran-5-yl}-5-methyl-1,3,4-oxadiazole derivatives as novel inhibitors of glycogen synthase kinase-3 $\beta$  with good brain permeability. *J. Med. Chem.* **52**, 6270-6286.

Sali A., Blundell T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779-815.

Sanchez-Carbayo M., Socci N.D., Lozano J., Saint F., Cordon-Cardo C. (2006). Defining molecular profiles of poor outcome in patients with invasive bladder cancer using oligonucleotide microarrays. *J. Clin. Oncol.* **24**, 778-789.

Sandeep G., Nagasree K.P., Hanisha M., Kumar M.M.K. (2011). AUDocker LE: A GUI for virtual screening with AUTODOCK Vina. *BMC Res Notes* **4**, 445.

Sanner M.F. (1999). Python: a programming language for software integration and development. *J. Mol. Graph. Model.* **17**, 57-61.

Saunders M., Houk K.N., Wu Y., Still W.C., Lipton M., Chang G., Guida W.C. (1990). Conformations of Cycloheptadecane. A Comparison of Methods for Conformational Searching. *J. Am. Chem. Soc.* **112**, 1419-1427.

Sauton N., Lagorce D., Villoutreix B.O., Miteva M.A. (2008). MS-DOCK: accurate multiple conformation generator and rigid docking protocol for multi-step virtual ligand screening. *BMC Bioinformatics* **9**, 184.

Schaeffer E.L., da Silva E.R., Novaes B.D.A., Skaf H.D., Gattaz W.F. (2010). Differential roles of phospholipases A2 in neuronal death and neurogenesis: implications for Alzheimer disease. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **34**, 1381-1389.

Schaloske R.H., Dennis E.A. (2006). The phospholipase A2 superfamily and its group numbering system. *Biochim. Biophys. Acta* **1761**, 1246-1259.

Schevitz R.W., Bach N.J., Carlson D.G., Chirgadze N.Y., Clawson D.K., Dillard R.D., Draheim S.E., Hartley L.W., Jones N.D., Mihelich E.D. et al. (1995). Structure-based design of the first potent and selective inhibitor of human non-pancreatic secretory phospholipase A2. *Nat. Struct. Biol.* **2**, 458-465.

Schulz-Gasch T., Stahl M. (2003). Binding site characteristics in structure-based virtual screening: evaluation of current docking tools. *J Mol Model* **9**, 47-57.

Schwartz P.A., Murray B.W. (2011). Protein kinase biochemistry and drug discovery. *Bioorg. Chem.* **39**, 192-210.

Scott D.L., White S.P., Browning J.L., Rosa J.J., Gelb M.H., Sigler P.B. (1991). Structures of free and inhibited human secretory phospholipase A2 from inflammatory

exudate. *Science* **254**, 1007-1010.

Seclaman E., Bora A., Avram S., Simon Z., Kurunczi L. (2011). MTD-PLS and docking study for a series of substituted 2-phenylindole derivatives with oestrogenic activity. *Chemical Papers* **65**, 566-576.

Seeliger D., de Groot B.L. (2010). Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *J. Comput. Aided Mol. Des.* **24**, 417-422.

Segura-Cabrera A., Guo X., Rojo-Domínguez A., Rodríguez-Pérez M. (2011). Integrative computational protocol for the discovery of inhibitors of the *Helicobacter pylori* nickel response regulator (NikR). *Journal of Molecular Modeling*. [Epub ahead of print] DOI: 10.1007/s00894-011-0962-2

Shen Q., Xiong B., Zheng M., Luo X., Luo C., Liu X., Du Y., Li J., Zhu W., Shen J. et al. (2011). Knowledge-based scoring functions in drug design: 2. Can the knowledge base be enriched?. *J. Chem. Inf. Model.* **51**, 386-397.

Shen X., Mizuguchi G., Hamiche A., Wu C. (2000). A chromatin remodelling complex involved in transcription and DNA processing. *Nature* **406**, 541-544.

Shen X., Ranallo R., Choi E., Wu C. (2003). Involvement of actin-related proteins in ATP-dependent chromatin remodeling. *Mol. Cell* **12**, 147-155.

Shenkin P.S., McDonald D.Q. (1994). Cluster Analysis of Molecular Conformations. *J. Comput. Chem.* **15**, 899-916.

Sheridan R.P., McGaughey G.B., Cornell W.D. (2008). Multiple protein structures and multiple ligands: effects on the apparent goodness of virtual screening results. *J. Comput. Aided Mol. Des.* **22**, 257-265.

Shin D., Lee S., Heo Y., Lee W., Cho Y., Kim Y.E., Hyun Y., Cho J.M., Lee Y.S., Ro S. (2007). Design and synthesis of 7-hydroxy-1H-benzimidazole derivatives as novel inhibitors of glycogen synthase kinase-3 $\beta$ . *Bioorg. Med. Chem. Lett.* **17**, 5686-5689.

Singer A.G., Ghomashchi F., Le Calvez C., Bollinger J., Bezzine S., Rouault M., Sadilek M., Nguyen E., Lazdunski M., Lambeau G. et al. (2002). Interfacial kinetic and binding properties of the complete set of human and mouse groups I, II, V, X, and XII secreted phospholipases A2. *J. Biol. Chem.* **277**, 48535-48549.

Smart B.P., Oslund R.C., Walsh L.A., Gelb M.H. (2006). The first potent inhibitor of mammalian group X secreted phospholipase A2: elucidation of sites for enhanced binding. *J. Med. Chem.* **49**, 2858-2860.

Smart B.P., Pan Y.H., Weeks A.K., Bollinger J.G., Bahnson B.J., Gelb M.H. (2004). Inhibition of the complete set of mammalian secreted phospholipases A(2) by indole analogues: a structure-guided study. *Bioorg. Med. Chem.* **12**, 1737-1749.

Smellie A., Tieg S.L., Towbin P. (1995). Poling: Promoting Conformational Variation. *J.*

Comput. Chem. **16**, 171-187.

Smith D.G., Buffet M., Fenwick A.E., Haigh D., Ife R.J., Saunders M., Slingsby B.P., Stacey R., Ward R.W. (2001). 3-Anilino-4-arylmaleimides: potent and selective inhibitors of glycogen synthase kinase-3 (GSK-3). *Bioorg. Med. Chem. Lett.* **11**, 635-639.

Søndergaard C.R., Garrett A.E., Carstensen T., Pollastri G., Nielsen J.E. (2009). Structural artifacts in protein-ligand X-ray structures: implications for the development of docking scoring functions. *J. Med. Chem.* **52**, 5673-5684.

Sousa S.F., Fernandes P.A., Ramos M.J. (2006). Protein-ligand docking: current status and future challenges. *Proteins* **65**, 15-26.

Sperandio O., Miteva M.A., Delfaud F., Villoutreix B.O. (2006). Receptor-based computational screening of compound databases: the main docking-scoring engines. *Curr. Protein Pept. Sci.* **7**, 369-393.

Stahl M., Rarey M. (2001). Detailed analysis of scoring functions for virtual screening. *J. Med. Chem.* **44**, 1035-1042.

Steiner R.A., Rozeboom H.J., de Vries A., Kalk K.H., Murshudov G.N., Wilson K.S., Dijkstra B.W. (2001). X-ray structure of bovine pancreatic phospholipase A2 at atomic resolution. *Acta Crystallogr. D Biol. Crystallogr.* **57**, 516-526.

Taha M.O., Bustanji Y., Al-Ghoussein M.A.S., Mohammad M., Zalloum H., Al-Masri I.M., Atallah N. (2008). Pharmacophore modeling, quantitative structure-activity relationship analysis, and in silico screening reveal potent glycogen synthase kinase-3beta inhibitory activities for cimetidine, hydroxychloroquine, and gemifloxacin. *J. Med. Chem.* **51**, 2062-2077.

Tapiero H., Ba G.N., Couvreur P., Tew K.D. (2002). Polyunsaturated fatty acids (PUFA) and eicosanoids in human health and pathologies. *Biomed. Pharmacother.* **56**, 215-222.

Tavares F.X., Boucheron J.A., Dickerson S.H., Griffin R.J., Preugschat F., Thomson S.A., Wang T.Y., Zhou H. (2004). N-Phenyl-4-pyrazolo[1,5-b]pyridazin-3-ylpyrimidin-2-amines as potent and selective inhibitors of glycogen synthase kinase 3 with good cellular efficacy. *J. Med. Chem.* **47**, 4716-4730.

ten Brink T., Exner T.E. (2009). Influence of protonation, tautomeric, and stereoisomeric states on protein-ligand docking results. *J. Chem. Inf. Model.* **49**, 1535-1546.

ter Haar E., Coll J.T., Austen D.A., Hsiao H.M., Swenson L., Jain J. (2001). Structure of GSK3beta reveals a primed phosphorylation mechanism. *Nat. Struct. Biol.* **8**, 593-596.

Teramoto R., Fukunishi H. (2007). Supervised consensus scoring for docking and virtual screening. *J. Chem. Inf. Model.* **47**, 526-534.

Torreira E., Jha S., López-Blanco J.R., Arias-Palomo E., Chacón P., Cañas C., Ayora S., Dutta A., Llorca O. (2008). Architecture of the pontin/reptin complex, essential in the

assembly of several macromolecular complexes. *Structure* **16**, 1511-1520.

Touqui L., Arbibe L. (1999). A role for phospholipase A2 in ARDS pathogenesis. *Mol Med Today* **5**, 244-249.

Trott O., Olson A.J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455-461.

Tsaneva I.R., Müller B., West S.C. (1993). RuvA and RuvB proteins of *Escherichia coli* exhibit DNA helicase activity in vitro. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 1315-1319.

Tuccinardi T. (2009). Docking-based virtual screening: recent developments. *Comb. Chem. High Throughput Screen.* **12**, 303-314.

Turabekova M.A., Rasulev B.F., Dzhakhangirov F.N., Leszczynska D., Leszczynski J. (2010). Aconitum and Delphinium alkaloids of curare-like activity. QSAR analysis and molecular docking of alkaloids into AChBP. *Eur J Med Chem* **45**, 3885-3894.

Vale R.D. (2000). AAA proteins. Lords of the ring. *J. Cell Biol.* **150**, F13-9.

Velec H.F.G., Gohlke H., Klebe G. (2005). DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *J. Med. Chem.* **48**, 6296-6303.

Venteicher A.S., Meng Z., Mason P.J., Veenstra T.D., Artandi S.E. (2008). Identification of ATPases pontin and reptin as telomerase components essential for holoenzyme assembly. *Cell* **132**, 945-957.

Ventura J., Nebreda A.R. (2006). Protein kinases and phosphatases as therapeutic targets in cancer. *Clin Transl Oncol* **8**, 153-160.

Verdonk M.L., Berdini V., Hartshorn M.J., Mooij W.T.M., Murray C.W., Taylor R.D., Watson P. (2004). Virtual screening using protein-ligand docking: avoiding artificial enrichment. *J. Chem. Inf. Comput. Sci.* **44**, 793-806.

Verdonk M.L., Chessari G., Cole J.C., Hartshorn M.J., Murray C.W., Nissink J.W.M., Taylor R.D., Taylor R. (2005). Modeling water molecules in protein-ligand docking using GOLD. *J. Med. Chem.* **48**, 6504-6515.

Verdonk M.L., Cole J.C., Hartshorn M.J., Murray C.W., Taylor R.D. (2003). Improved protein-ligand docking using GOLD. *Proteins* **52**, 609-623.

Verdonk M.L., Cole J.C., Taylor R. (1999). SuperStar: A Knowledge-based Approach for Identifying Interaction Sites in Proteins. *J. Mol. Biol.* **289**, 1093-1108.

Verdonk M.L., Cole J.C., Watson P., Gillet V., Willett P. (2001). Superstar: improved knowledge-based interaction fields for protein binding sites. *J. Mol. Biol.* **307**, 841-859.

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.

Vistoli G., Pedretti A., Testa B. (2008). Assessing drug-likeness – what are we missing?. *Drug Discov. Today* **13**, 285-294.

Wade R.C., Clark K.J., Goodford P.J. (1993). Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 1. Ligand probe groups with the ability to form two hydrogen bonds. *J. Med. Chem* **36**, 140-147.

Wade R.C., Goodford P.J. (1993). Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 2. Ligand probe groups with the ability to form more than two hydrogen bonds. *J. Med. Chem* **36**, 148-156.

Walker J.E., Saraste M., Runswick M.J., Gay N.J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1**, 945-951.

Walters W.P., Stahl M.T., Murcko M.A. (1998). Virtual screening – an overview. *Drug Discov. Today* **3**, 160-178.

Wang R., Fang X., Lu Y., Yang C., Wang S. (2005). The PDBbind Database: Methodologies and Updates. *J. Med. Chem.* **48**, 4111-4119.

Wang R., Lai L., Wang S. (2002). Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput. Aided Mol. Des.* **16**, 11-26.

Wang R., Lu Y., Fang X., Wang S. (2004). An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein-ligand complexes. *J. Chem. Inf. Comput. Sci.* **44**, 2114-2125.

Wang R., Lu Y., Wang S. (2003). Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **46**, 2287-2303.

Wang R., Wang S. (2001). How does consensus scoring work for virtual library screening? An idealized computer experiment. *J. Chem. Inf. Comput. Sci.* **41**, 1422-1426.

Warren G.L., Andrews C.W., Capelli A., Clarke B., LaLonde J., Lambert M.H., Lindvall M., Nevins N., Semus S.F., Senger S. et al. (2006). A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **49**, 5912-5931.

Watkins N.J., Dickmanns A., Lührmann R. (2002). Conserved stem II of the box C/D motif is essential for nucleolar localization and is required, along with the 15.5K protein, for the hierarchical assembly of the box C/D snoRNP. *Mol. Cell. Biol.* **22**, 8342-8352.

Watkins N.J., Lemm I., Ingelfinger D., Schneider C., Hossbach M., Urlaub H., Lührmann R. (2004). Assembly and maturation of the U3 snoRNP in the nucleoplasm in a

large dynamic multiprotein complex. *Mol. Cell* **16**, 789-798.

Webb N.R. (2005). Secretory phospholipase A2 enzymes in atherogenesis. *Curr. Opin. Lipidol.* **16**, 341-344.

Weiner P.K., Kollman P.A. (1981). AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions. *J. Comput. Chem.* **2**, 287-303.

Weiner S.J., Kollman P.A., Case D.A., Singh U.C., Ghio C., Alagona G., Profeta S., Weiner P. (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **106**, 765-784.

Weiner S.J., Kollman P.A., Nguyen D.T., Case D.A. (1986). An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem.* **7**, 230-252.

Weininger D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **28**, 31-36.

Wiggers H.J., Rocha J.R., Cheleski J., Montanari C.A. (2011). Integration of Ligand- and Target-Based Virtual Screening for the Discovery of Cruzain Inhibitors. *Molecular Informatics* **30**, 565-578.

Winget J.M., Pan Y.H., Bahnson B.J. (2006). The interfacial binding surface of phospholipase A2s. *Biochim. Biophys. Acta* **1761**, 1260-1269.

Witherington J., Bordas V., Gaiba A., Garton N.S., Naylor A., Rawlings A.D., Slingsby B.P., Smith D.G., Takle A.K., Ward R.W. (2003a). 6-aryl-pyrazolo[3,4-b]pyridines: potent inhibitors of glycogen synthase kinase-3 (GSK-3). *Bioorg. Med. Chem. Lett.* **13**, 3055-3057.

Witherington J., Bordas V., Gaiba A., Naylor A., Rawlings A.D., Slingsby B.P., Smith D.G., Takle A.K., Ward R.W. (2003b). 6-heteroaryl-pyrazolo[3,4-b]pyridines: potent and selective inhibitors of glycogen synthase kinase-3 (GSK-3). *Bioorg. Med. Chem. Lett.* **13**, 3059-3062.

Witherington J., Bordas V., Garland S.L., Hickey D.M.B., Ife R.J., Liddle J., Saunders M., Smith D.G., Ward R.W. (2003c). 5-aryl-pyrazolo[3,4-b]pyridines: potent inhibitors of glycogen synthase kinase-3 (GSK-3). *Bioorg. Med. Chem. Lett.* **13**, 1577-1580.

Witherington J., Bordas V., Haigh D., Hickey D.M.B., Ife R.J., Rawlings A.D., Slingsby B.P., Smith D.G., Ward R.W. (2003d). 5-aryl-pyrazolo[3,4-b]pyridazines: potent inhibitors of glycogen synthase kinase-3 (GSK-3). *Bioorg. Med. Chem. Lett.* **13**, 1581-1584.

Wood M.A., McMahon S.B., Cole M.D. (2000). An ATPase/helicase complex is an essential cofactor for oncogenic transformation by c-Myc. *Mol. Cell* **5**, 321-330.

Wu S., Shi Y., Mulligan P., Gay F., Landry J., Liu H., Lu J., Qi H.H., Wang W., Nickoloff J.A. et al. (2007). A YY1-INO80 complex regulates genomic stability through homologous recombination-based repair. *Nat. Struct. Mol. Biol.* **14**, 1165-1172.

Wu W., Alami S., Luk E., Wu C., Sen S., Mizuguchi G., Wei D., Wu C. (2005). Swc2 is a widely conserved H2AZ-binding module essential for ATP-dependent histone exchange. *Nat. Struct. Mol. Biol.* **12**, 1064-1071.

Wu Y., Medjane S., Chabot S., Kubrusly F.S., Raw I., Chignard M., Touqui L. (2003). Surfactant protein-A and phosphatidylglycerol suppress type IIA phospholipase A2 synthesis via nuclear factor-kappaB. *Am. J. Respir. Crit. Care Med.* **168**, 692-699.

Yamashita S., Yamashita J., Ogawa M. (1994). Overexpression of group II phospholipase A2 in human breast cancer tissues is closely associated with their malignant potency. *Br. J. Cancer* **69**, 1166-1170.

Yang J., Chen Y., Shen T., Kristal B.S., Hsu D.F. (2005). Consensus scoring criteria for improving enrichment in virtual screening. *J. Chem. Inf. Model.* **45**, 1134-1146.

Ying Z., Tojo H., Komatsubara T., Nakagawa M., Inada M., Kawata S., Matsuzawa Y., Okamoto M. (1994). Enhanced expression of group II phospholipase A2 in human hepatocellular carcinoma. *Biochim. Biophys. Acta* **1226**, 201-205.

Yu B., Pan Y.H., Janssen M.J.W., Bahnson B.J., Jain M.K. (2005). Kinetic and structural properties of disulfide engineered phospholipase A2: insight into the role of disulfide bonding patterns. *Biochemistry* **44**, 3369-3379.

Yuriev E., Agostino M., Ramsland P.A. (2011). Challenges and advances in computational docking: 2009 in review. *J. Mol. Recognit.* **24**, 149-164.

Zhan F., Barlogie B., Arzoumanian V., Huang Y., Williams D.R., Hollmig K., Pineda-Roman M., Tricot G., van Rhee F., Zangari M. et al. (2007). Gene-expression signature of benign monoclonal gammopathy evident in multiple myeloma is linked to good prognosis. *Blood* **109**, 1692-1700.

Zhang H., Boñaga L.V.R., Ye H., Derian C.K., Damiano B.P., Maryanoff B.E. (2007). Novel bis(indolyl)maleimide pyridinophanes that are potent, selective inhibitors of glycogen synthase kinase-3. *Bioorg. Med. Chem. Lett.* **17**, 2863-2868.

Zhang H., Derian C.K., McComsey D.F., White K.B., Ye H., Hecker L.R., Li J., Addo M.F., Croll D., Eckardt A.J. et al. (2005). Novel indolylindazolylmaleimides as inhibitors of protein kinase C-beta: synthesis, biological activity, and cardiovascular safety. *J. Med. Chem.* **48**, 1725-1728.

Zhang H., White K.B., Ye H., McComsey D.F., Derian C.K., Addo M.F., Andrade-Gordon P., Eckardt A.J., Conway B.R., Westover L. et al. (2003). Macrocyclic bisindolylmaleimides as inhibitors of protein kinase C and glycogen synthase kinase-3. *Bioorg. Med. Chem. Lett.* **13**, 3049-3053.

Zhang X., Wigley D.B. (2008). The 'glutamate switch' provides a link between ATPase activity and ligand binding in AAA+ proteins. *Nat. Struct. Mol. Biol.* **15**, 1223-1227.



# **ANNEXES**



## - ANNEXE 1. FORMATS DE FICHIERS DE MOLÉCULES -

Il existe plusieurs types de formats permettant de stocker des données chimiques et plus particulièrement les structures tridimensionnelles de molécules. Parmi les plus souvent rencontrés et ceux qui ont été utilisés au cours de ce travail, les formats SMILES, SDF, MOL2, PDB et PDBQT sont décrits ici en détail. Outre les lignes correspondant aux coordonnées des atomes, ces formats prévoient aussi des espaces dédiés aux informations basiques (nom de la molécule, nombre d'atomes, de résidus...) et spécifiques (détails techniques de la cristallographie pour le format PDB par exemple) dont seuls ceux qui ont une importance dans notre cas seront explicités.

### ***Format SMILES***

Le format SMILES (Simplified Molecular Input Line Entry Specification) [Weininger, 1988] est un langage permettant de décrire les structures chimiques sous la forme de simples chaînes de caractères, grâce à la représentation des molécules sous la forme de graphes. L'avantage du code SMILES est sa simplicité, qui permet de stocker un très grand nombre de structures dans des fichiers de petite taille. Cependant il ne contient aucune information concernant le type atomique ou les arrangements spatiaux des atomes en 2D ou en 3D.

Chaque atome est représenté par son symbole chimique entre crochets, sauf pour les atomes usuels (C, N, O, P, S, B, F, Cl, Br et I). Dans ce dernier cas, les valences libres peuvent être considérées comme remplies par des atomes d'hydrogènes non explicités. Les liaisons simples sont implicites et les liaisons doubles et triples sont représentées par les signes "=" et "#". Des parenthèses sont utilisées pour séparer les branches de substituants. Dans les cycles aromatiques, les doubles liaisons ne sont généralement pas représentées, mais des lettres minuscules sont utilisées pour les symboles des atomes. Plusieurs signes permettent également de préciser la stéréochimie des molécules.

### ***Format SDF***

Le format SDF ou SD (Structure Data Format) [Dalby *et al.*, 1992] permet la représentation de molécules en 2D ou en 3D, il est très souvent utilisé pour stocker des bases de ligands sur internet. Il est constitué d'une table de connectivité qui définit les propriétés des atomes, leurs coordonnées et leurs relations. Il est divisé en plusieurs blocs, requis ou non, et dans lesquels les différents champs sont séparés les uns des autres par un

simple espace, ou tout caractère d'espacement (tabulation). Tous les éléments ne sont pas obligatoires, ceux qui ne sont renseignés peuvent être remplacés par une valeur standard (0 par exemple pour les champs chiffrés).

- Bloc 1 : Obligatoire. Il s'agit d'une simple ligne comprenant au moins :
  - Le nombre d'atomes dans la molécule.
  - Le nombre de liaisons dans la molécule.
- Bloc 2 : Obligatoire. Il représente la structure à proprement parler, chaque ligne figurant un atome de la molécule.
  - Champs 1, 2 et 3 : Coordonnées de l'atome selon les axes X, Y et Z en Angströms. Dans le cas de structures en 2D, la coordonnée Z est laissée vide.
  - Champ 4 : Type atomique représenté le plus souvent par le symbole chimique de l'élément.
  - Champs 5 à 16 : Charge (nombre entier uniquement), détails sur la stéréochimie ou la valence de l'atome, etc.
- Bloc 3 : Non obligatoire mais complémentaire. Il permet de préciser les relations entre les atomes, chaque ligne représentant une liaison.
  - Champs 1 et 2 : Numéro des deux atomes concernés (c'est à dire leur position dans la liste des atomes du bloc précédent).
  - Champ 3 : Type de liaison, sous la forme d'un entier. "1" représente une liaison simple, "2" une liaison double, "3" une liaison triple, "4" une liaison aromatique, "5" une liaison simple ou double, "6" une liaison simple ou aromatique, "7" une liaison double ou aromatique, "8" une liaison quelconque, "9" une liaison de coordination, et "10" une liaison hydrogène.
  - Champs 4 à 7 : Autres précisions.
- Bloc 4 : Non obligatoire. Cette partie du fichier permet de rajouter des données optionnelles au besoin. Une ligne doit commencer par le signe "supérieur" (">") et le nom de la donnée doit être indiqué entre un signe "inférieur" et un signe "supérieur" (exemple : "<donnée1>").

Les fichiers SDF peuvent contenir de multiples structures. Dans ce cas, elles doivent être séparées par une ligne comprenant quatre signes dollar ("\$\$\$\$").

### ***Format MOL2***

Le format MOL2 possède des propriétés identiques au format SDF. Comme celui-ci, il permet la gestion de fichiers en 3D, il est séparé en plusieurs blocs distincts qui sont requis ou non pour que le fichier soit correct, et chaque ligne contient plusieurs éléments séparés

par des espaces ou des tabulations. La séparation entre les blocs est matérialisée par une ligne qui débute le bloc et en indique le contenu. Un champ non obligatoire et non renseigné peut être remplacé par une valeur standard, un champ obligatoire mais non renseigné doit être remplacé par la chaîne de caractères "\*\*\*\*".

- Bloc 1 : Obligatoire. Il commence par la ligne "@<TRIPOS>MOLECULE" et fournit les renseignements de base sur la molécule.
  - Ligne 1 : Nom de la molécule.
  - Ligne 2 : Au minimum, nombre d'atomes, nombre de liaisons, nombre de sous-structures.
  - Ligne 3 : Type de molécule. Peut prendre les valeurs SMALL, BIOPOLYMER, PROTEIN, NUCLEIC\_ACID, SACCHARIDE.
  - Ligne 4 : Type de charges appliquées. Par exemple, NO\_CHARGES, GASTEIGER, USER\_CHARGES.
  - Ligne 5 : Commentaire (non obligatoire).
- Bloc 2 : Obligatoire. Il représente la structure à proprement parler, chaque ligne figurant un atome de la molécule. Il commence par la ligne "@<TRIPOS>ATOM".
  - Champ 1 : Numéro de l'atome.
  - Champ 2 : Nom de l'atome.
  - Champs 3, 4, 5 : Coordonnées selon les X, les Y et les Z en Angströms.
  - Champ 6 : Type atomique
  - Champ 7 et 8 : Numéro et nom de la sous-structure à laquelle l'atome appartient.
  - Champ 9 : Charge partielle de l'atome.
- Bloc 3 : Obligatoire. Il fournit le détail des liaisons entre les atomes. Il commence par la ligne "@<TRIPOS>BOND".
  - Champ 1 : Numéro de la liaison.
  - Champ 2 et 3 : Numéro des deux atomes concernés.
  - Champ 4 : Type de liaison. "1" représente une liaison simple, "2" une liaison double, "3" une liaison triple, "am" une liaison amide, "ar" une liaison aromatique, "du" une liaison entre dummy atomes, "un" une liaison indéterminée et "nc" indique que les deux atomes ne sont pas connectés.
- Bloc 4 : Obligatoire. Il commence par la ligne "@<TRIPOS>SUBSTRUCTURE" et réunit les informations concernant les différentes sous-structures présentes dans le fichier.
  - Champ 1 : Numéro de la sous-structure.

- Champ 2 : Nom de la sous-structure.
- Champ 3 : Numéro de l'atome principal (root atom) de la sous-structure.
- Champ 4 : Type de sous-structure. Peut prendre les valeurs temp, perm, residue, group ou domain.
- Champ 5 : Type de dictionnaire associé à la sous-structure. Ce paramètre est obligatoire mais très rarement renseigné.
- Champ 6 et 7 : Chaîne dont dépend la sous-structure et type de chaîne.
- Autres blocs : De nombreux autres blocs peuvent être ajoutés en fonction des besoins mais ne sont pas nécessaires à la lecture du fichier.

Les fichiers MOL2 peuvent contenir plusieurs structures, simplement en les plaçant les unes à la suite des autres, les blocs étant toujours présentés dans le même ordre.

### **Format PDB**

Le format PDB est le format le plus répandu pour les protéines. Il s'agit du format développé par la Protein Data Bank pour le stockage des structures protéiques [Berman *et al.*, 2000].

Il peut être divisé en plusieurs sous-parties, qui sont différenciées par un marqueur constitué des six premiers caractères de chaque ligne, mais seule la partie structure à proprement parler est requise dans le fichier. Dans les structures issues de la PDB, un certain nombre de remarques situées en début de fichier permettent de décrire avec précision l'origine de la protéine (références bibliographiques, date de publication, information cristallographiques, séquence protéique, etc.). La partie structure est constituée de lignes de 80 caractères, chaque position sur la ligne correspondant à un champ particulier. Tous les champs ne sont pas obligatoires, ceux qui le sont sont marqués d'une étoile dans la liste ci-dessous (\*).

- Colonnes 1 à 6\* : Marqueur indiquant que l'on se trouve dans la partie coordonnées. Le format PDB étant à l'origine fait pour stocker les structures des protéines, ce marqueur peut prendre les valeurs "ATOM " pour les protéines ou "HETATM" pour les ligands, molécules d'eau ou tout autre espèce présente dans la structure.
- Colonnes 7 à 11\* : Identifiant de l'atome, compris entre 1 et 99999.
- Colonnes 13 à 16\* : Nom de l'atome, en alphanumérique. Un même nom peut être attribué plusieurs fois dans la même structure sans conséquence.
- Colonne 17 : Marqueur de position alternative. Lorsqu'il existe plusieurs conformations, tous les atomes pour une conformation donnée sont pointés par un marqueur de position alternative identique.

- Colonnes 18 à 20\* : Nom du résidu (protéine) ou de la molécule (autre), les acides aminés étant identifiées selon leur code à trois lettres.
- Colonne 22 : Nom de la chaîne (protéine), en caractère alphabétique.
- Colonnes 23 à 26\* : Identifiant du résidu ou de la molécule, compris entre 1 et 9999.
- Colonne 27 : Marqueur d'insertion, utilisé si deux résidus ont le même identifiant.
- Colonnes 31 à 38\* : Coordonnées de l'atome selon les X, exprimées en Angströms, jusqu'à trois chiffres après la virgule.
- Colonnes 39 à 46 et 47 à 54\* : Coordonnées en Y et Z.
- Colonnes 55 à 60 : Facteur d'occupation, jusqu'à deux chiffres après la virgule.
- Colonnes 61 à 66 : Facteur d'agitation thermique, jusqu'à deux chiffres après la virgule.
- Colonnes 77 à 78\* : Symbole chimique de l'élément, justifié à droite.
- Colonnes 79 à 80 : Charge de l'atome, constitué d'un seul chiffre suivi d'un signe "positif" ("+") ou "négatif" ("-").

On trouve très souvent une ligne qui débute par "TER" en fin de fichier, pour indiquer la terminaison de la structure. Plusieurs structures peuvent être présentes dans un même fichier. Chacune commence alors par la ligne "MODEL" et finit par la ligne "ENDMDL".

#### Format PDBQT

Ce format est un dérivé du précédent et utilise donc de façon générale le même formalisme. Cependant, ces fichiers sont utilisés en entrée de docking et deux différences majeures sont à noter.

- Les atomes doivent être pourvus d'une charge partielle (jusqu'à trois chiffres après la virgule) qui occupe les colonnes 71 à 76. Le symbole chimique de l'élément est repoussé dans les colonnes 79 à 80.
- Les liaisons flexibles sont matérialisées dans le fichier. La molécule est divisée en fragments le long des torsions, on distingue la branche principale (root), et les branches annexes. Une ligne contenant le mot "ROOT" précède le premier atome du fragment principal et une ligne contenant le mot "ENDROOT" suit son dernier atome. De la même façon, une ligne comprenant le mot "BRANCH" et les numéros des deux atomes impliqués dans la torsion précède chaque branche, et une avec le mot "ENDBRANCH" et ces deux numéros suit le dernier atome de celle-ci. A la fin de la molécule, une ligne qui commence par "TORSIDORF" indique le nombre de torsions total dans la structure.



## - ANNEXE 2. GESTION DES FICHIERS ET AUTOMATISATION DES CALCULS -

Lors de ce travail, tous les criblages ont été effectués avec les programmes de docking Dock 5.4 [Kuntz *et al.*, 1982; Moustakas *et al.*, 2006] et Vina 1.0.2 [Trott & Olson, 2010], et les programmes de scoring Drugscore PDB [Gohlke *et al.*, 2000], Drugscore CSD [Veleg *et al.*, 2005] et Xscore [Wang *et al.*, 2002].

Les programmes Dock, Drugscore et Xscore sont utilisables sur des fichiers au format PDB et MOL2 (voir Annexe 1), des formats assez classiques qui sont gérés par OpenBabel ([www.openbabel.org](http://www.openbabel.org)) [Guha *et al.*, 2006; O'Boyle *et al.*, 2011]. De plus, ils permettent l'emploi de fichiers multi-MOL2 contenant autant de ligands qu'il y en a dans la base, au lieu de fichiers individuels pour chaque ligands. Ceci ramène donc le lancement des calculs à une seule opération impliquant seulement les fichiers contenant la protéine (structure plus des fichiers spécifiques à chaque programme), un fichier contenant les ligands et éventuellement un fichier de configuration.

En revanche, le programme de docking Vina, qui est basé sur Autodock [Morris *et al.*, 1998] utilise son propre format de fichier dérivé du format PDB classique, le PDBQT (voir Annexe 1). Bien que peu différent du PDB, le PDBQT n'était pas pris en charge par OpenBabel au moment où les expériences décrites ici ont été menées. De plus, bien que Vina permette un criblage rapide depuis la ligne de commande (au lieu de l'interface graphique d'Autodock), il nécessite des fichiers mono-molécule pour chaque ligand, ce qui représente donc autant de fichiers de structure, de fichiers de configuration, et de commandes à lancer que de ligands présents dans la base, pour chaque criblage. La suite d'outils MGLTools 1.5.4 [Sanner, 1999] disponible avec Autodock fournit un certain nombre de scripts permettant de préparer les fichiers d'entrée des ligands pour le docking à partir de fichiers MOL2 ou PDB, en incluant de manière automatique des rotations sur les liaisons flexibles. Dans un premier temps, nous avons donc automatisé la génération des fichiers de configuration à l'aide d'un script python. Celui-ci crée pour chaque ligand dans un dossier un fichier configuration (.conf) contenant les informations nécessaires aux calculs, ceux-ci étant ensuite lancés les uns à la suite des autres à l'aide d'un fichier texte rendu exécutable et contenant les lignes de commande.

La ligne "*flex*" n'est requise que dans le cas d'un docking flexible, les paramètres "*exhaustiveness*" (complétude des calculs) et "*num\_modes*" (nombre de poses générées au maximum par ligand) ne sont pas obligatoires, et prennent en cas d'absence les valeurs par défaut indiquées ici, enfin, une seed aléatoire est choisie si la ligne "*seed*" n'est pas remplie.

```
receptor = recH_rig.pdbqt
flex = recH_flex.pdbqt
ligand = lig.pdbqt
out = 1_lig.pdbqt
center_x = 0
center_y = 0
center_z = 0

size_x = 20
size_y = 20
size_z = 20

cpu = 2
seed = 1
exhaustiveness = 8
num_modes = 9
```

En ce qui concerne les fichiers de sortie, les solutions proposées par MGLTools ne sont pas compatibles avec la procédure de rescoring mise en place, c'est pourquoi nous avons créé un script en langage python pour convertir les fichiers de sortie de Vina en fichier PDB ou MOL2 utilisables avec Drugscore et Xscore. Rapidement, nous avons également intégré le lancement des calculs de rescoring dans le même script de manière à obtenir directement les scores pour chacune des poses générées par Vina. Cet outil de gestion des résultats obtenus avec Vina a été nommé Vinapac.

Pour chaque ligand, les fichiers de sortie de Vina se présentent sous la forme de fichiers PDBQT contenant entre une et n poses, avec n le nombre de poses demandé au maximum (voir Annexe G2). Dans le cas d'un docking flexible, la partie mobile de la protéine est intégrée à chaque pose. La première étape consiste à séparer les différentes poses adoptées par le ligand et, s'il y a lieu, les parties flexibles du récepteur sont isolées du ligand correspondant. Ensuite, chaque pose (ligand uniquement) est écrite en PDB dans un fichier individuel. Les colonnes et les lignes supplémentaires sont supprimées et les séparations entre chaque colonne sont vérifiées pour corriger d'éventuelles erreurs qui peuvent empêcher le bon fonctionnement d'OpenBabel ou des programmes de rescoring. En particulier, pour les atomes dont le symbole est constitué de deux lettres, il faut s'assurer que la taille complète de l'identifiant (nom de l'atome + position) ne dépasse pas quatre caractères. En parallèle, le récepteur est également écrit en PDB en réinsérant les parties flexibles dans le fichier (dans le cas d'un docking rigide, le récepteur n'est pas réécrit

puisque'il n'a pas été modifié). Enfin les fichiers PDB sont convertis en MOL2 par l'intermédiaire d'OpenBabel, lequel est également utilisé pour rajouter sur les structures les hydrogènes non polaires, non considérés par Vina.

Les fichiers écrits sont utilisés pour lancer automatiquement les calculs de rescoring avec Drugscore et Xscore. Outre les fichiers MOL2 et PDB, le script écrit donc également un fichier texte de type CSV (Comma Separated Values) qui contient les valeurs associées à chaque pose unique pour toutes les fonctions de score. Une pose unique est définie par le numéro de pose, le ligand testé, et si le criblage a été réalisé en plusieurs exemplaires pour des raisons de reproductibilité, le numéro d'expérience (égal à la seed fixée dans notre cas). En plus de Vinapac, un autre script permet l'extraction de ces résultats. Pour chaque pose, les scores obtenus sur plusieurs expériences sont moyennés, puis pour chaque ligand, seule le score de la meilleure pose est conservé pour créer les classements définitifs pour toutes les fonctions de score utilisées.



### - ANNEXE 3. STRUCTURES 2D DES INDIRUBINES -

Toutes les structures présentées ici et dans le corps du manuscrit ont été dessinées à l'aide de MarvinSketch 5.3.2 (ChemAxon, Budapest, Hungary).

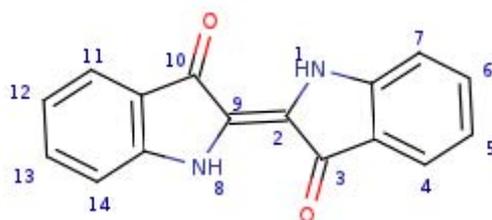
#### Molécules dérivées de l'indirubine



| Nom | R1 | R5                | R6 | R10  | R12 |
|-----|----|-------------------|----|------|-----|
| 1   |    |                   |    |      |     |
| 4   |    | I                 |    |      |     |
| 5   |    | Br                |    |      |     |
| 6   |    | Cl                |    |      |     |
| 7   |    | F                 |    |      |     |
| 8   |    | CH <sub>3</sub>   |    |      |     |
| 9   |    | NO <sub>2</sub>   |    |      |     |
| 10  |    | SO <sub>3</sub> H |    |      |     |
| 11  |    |                   |    |      | Br  |
| 12  |    | Br                |    |      | Br  |
| 13  |    | SO <sub>3</sub> H |    |      | Br  |
| 14  |    |                   |    | =NOH |     |
| 15  |    | I                 |    | =NOH |     |
| 16  |    |                   | I  |      |     |

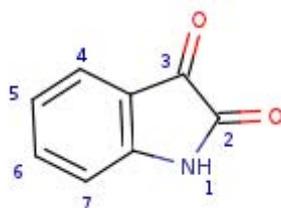
|    |    |                                                        |  |               |  |
|----|----|--------------------------------------------------------|--|---------------|--|
| 18 | Ph |                                                        |  |               |  |
| 19 |    | $\text{SO}_3^-$                                        |  | $=\text{NOH}$ |  |
| 20 |    | $\text{SO}_2\text{NH}_2$                               |  |               |  |
| 21 |    | $\text{SO}_2\text{N}(\text{CH}_3)_2$                   |  |               |  |
| 22 |    | $\text{SO}_2\text{NC}_2\text{H}_5\text{OH}$            |  |               |  |
| 23 |    | $\text{SO}_2\text{N}(\text{C}_2\text{H}_5\text{OH})_2$ |  |               |  |
| 24 |    | $\text{SO}_2\text{NHCH}_3$                             |  |               |  |

***Molécules dérivées de l'indigo***



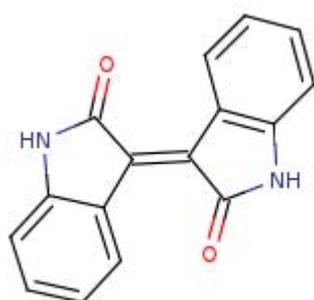
| Nom | R5              | R7              | R12             | R14             |
|-----|-----------------|-----------------|-----------------|-----------------|
| 2   |                 |                 |                 |                 |
| 25  | $\text{SO}_3^-$ | $\text{SO}_3^-$ | $\text{SO}_3^-$ | $\text{SO}_3^-$ |
| 26  | $\text{SO}_3^-$ |                 | $\text{SO}_3^-$ | $\text{SO}_3^-$ |
| 27  | $\text{SO}_3^-$ |                 | $\text{SO}_3^-$ |                 |

***Molécules dérivées de l'isatin***

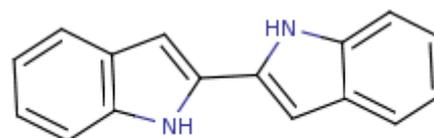


| Nom | R5                                                               | R6 |
|-----|------------------------------------------------------------------|----|
| 34  | NO <sub>2</sub>                                                  |    |
| 37  | SO <sub>2</sub> N(CH <sub>3</sub> ) <sub>2</sub>                 |    |
| 38  | SO <sub>2</sub> N(C <sub>2</sub> H <sub>5</sub> OH) <sub>2</sub> |    |
| 39  |                                                                  | I  |

**Autres molécules**



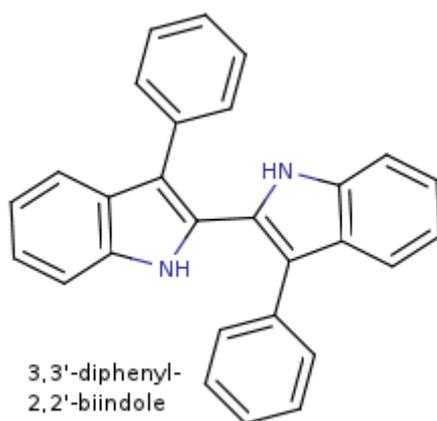
Isoindigo



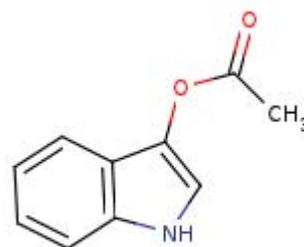
2,2'-biindole



5-bromo-3-indoxyl acetate



3,3'-diphenyl-  
2,2'-biindole



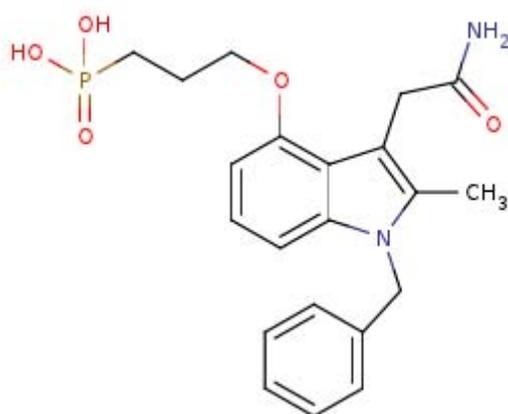
3-indoxyl acetate



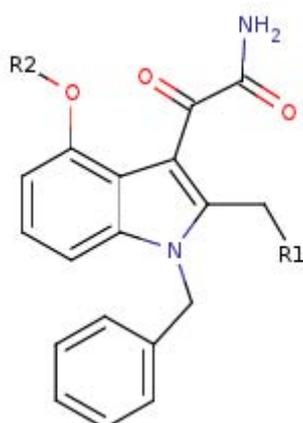
## - ANNEXE 4. STRUCTURES 2D DES INHIBITEURS DE HGX -

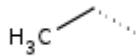
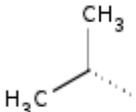
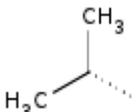
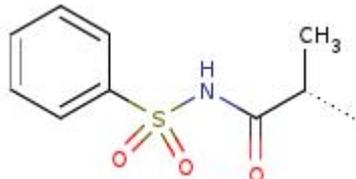
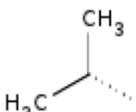
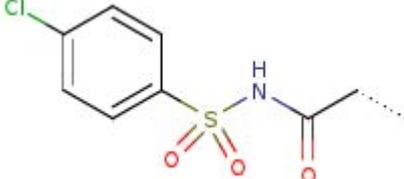
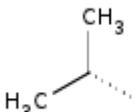
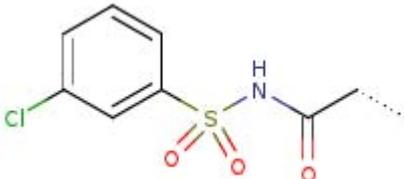
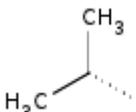
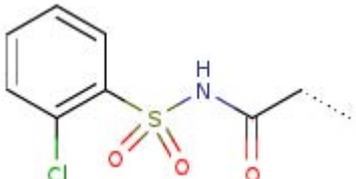
Toutes les structures présentées ici et dans le corps du manuscrit ont été dessinées à l'aide de MarvinSketch 5.3.2 (ChemAxon, Budapest, Hungary).

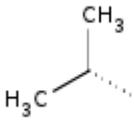
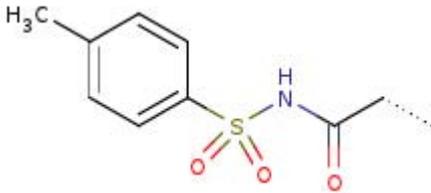
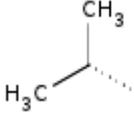
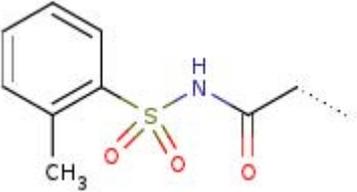
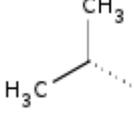
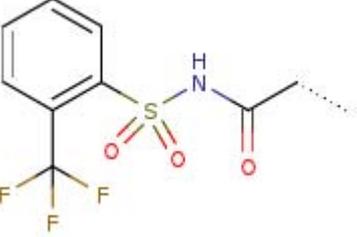
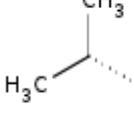
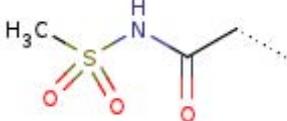
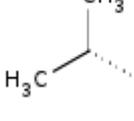
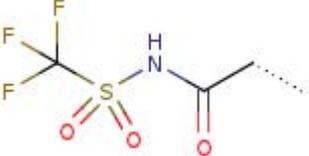
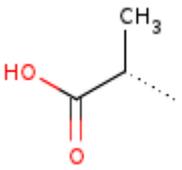
### LY311727 (extraite de la structure PDB 1db4 [Schevitz et al., 1995])

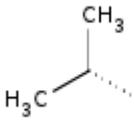
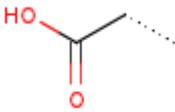


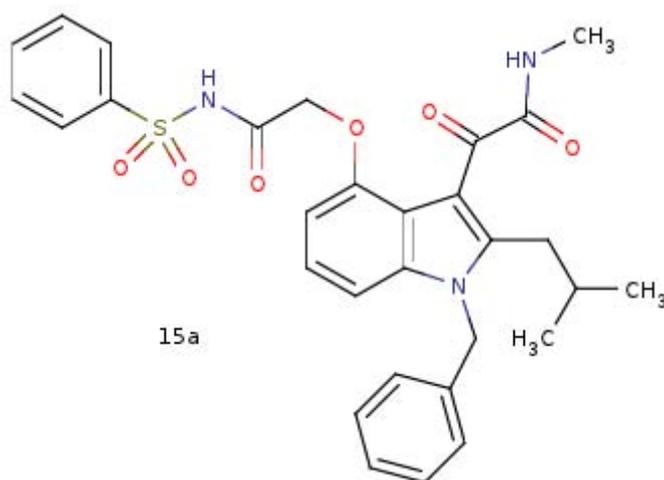
### Molécules dérivées de LY315920



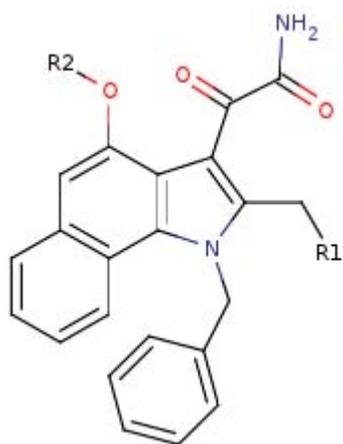
| Nom | R1                                                                                  | R2                                                                                   |
|-----|-------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|
| 11c |    |    |
| 11d |    |    |
| 13a |    |   |
| 13b |  |  |
| 13c |  |  |
| 13d |  |  |

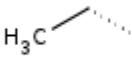
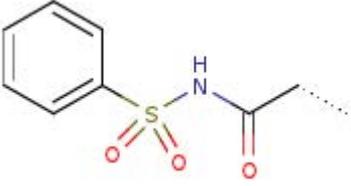
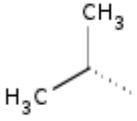
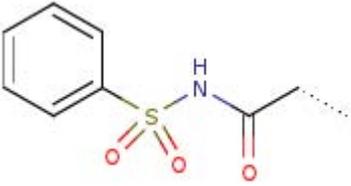
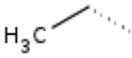
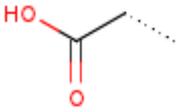
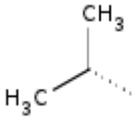
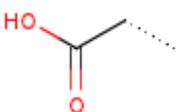
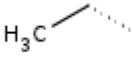
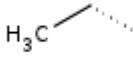
|     |                                                                                     |                                                                                       |
|-----|-------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
| 13e |    |     |
| 13f |    |     |
| 13g |    |     |
| 13h |  |   |
| 13i |  |   |
| 12a |  |  |

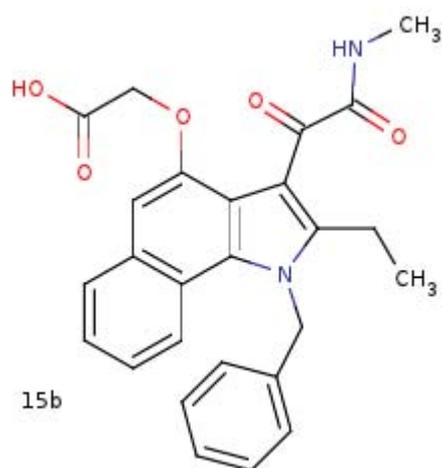
|     |                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                               |
|-----|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 12b |  <p>Chemical structure of (S)-2-butanol, showing a central carbon atom bonded to a methyl group (CH<sub>3</sub>), a hydroxyl group (OH), and two other groups (represented by a solid wedge and a dashed wedge).</p> |  <p>Chemical structure of (S)-2-butanoic acid, showing a central carbon atom bonded to a hydroxyl group (HO), a methyl group (CH<sub>3</sub>), and two other groups (represented by a solid wedge and a dashed wedge).</p> |
|-----|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|



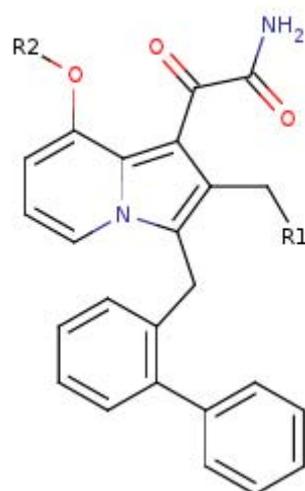
***Molécules dérivées de LY315920 de type benzo-indole***



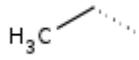
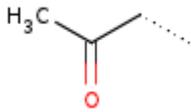
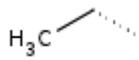
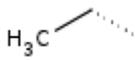
| Nom | R1                                                                                  | R2                                                                                    |
|-----|-------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
| 11g |    |     |
| 11h |    |     |
| 12e |    |    |
| 12f |  |  |
| 14a |  |  |

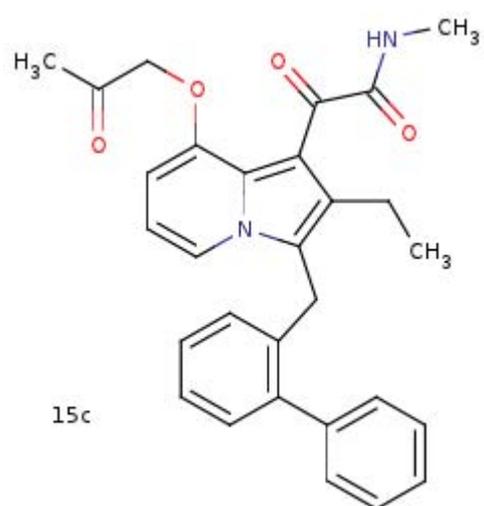


**Molécules dérivées de l'indoxam**



| Nom | R1 | R2 |
|-----|----|----|
| 16a |    |    |

|     |                                                                                   |                                                                                     |
|-----|-----------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| 16b |  |  |
| 16c |  |  |





**- ANNEXE 5. PUBLICATION -**

Publication acceptée dans Biochemical Journal.

**First identification of small molecule inhibitors of Pontin by combining  
virtual screening and enzymatic assay**

Judith ELKAIM\*, Michel CASTROVIEJO†, Driss BENNANI\*, Said TAOUJI‡, Nathalie ALLAIN‡,  
Michel LAGUERRE\*, Jean ROSENBAUM‡, Jean DESSOLIN\*<sup>1</sup> and Patrick LESTIENNE‡<sup>2</sup>

\* Molecular modeling group, IECB-CNRS-Université de Bordeaux, UMR 5248, 2 rue R.  
Escarpit, F-33607 Pessac, France

† Platform protein expression and purification, CNRS, UMR 5234, 146 rue L. Saignat, F-33076  
Bordeaux Cedex, France

‡ Physiopathologie du cancer du foie, INSERM U1053-Université de Bordeaux, 146 rue L.  
Saignat, F-33076 Bordeaux Cedex, France

\*Contacts:

<sup>1</sup> Correspondance in the field of Molecular Modeling

[j.dessolin@iecb.u-bordeaux.fr](mailto:j.dessolin@iecb.u-bordeaux.fr)

Tel: 33 (0)5 40 00 30 29

<sup>2</sup> Correspondance in the field of Biochemistry

[patrick.lestienne@inserm.fr](mailto:patrick.lestienne@inserm.fr)

Tel: 33 (0)5 57 57 47 27

**SHORT TITLE**

A combined approach to select Pontin inhibitors

**KEYWORDS**

Pontin, TIP49, enzymatic assay, virtual screening, Vina, ATPase activity

**SYNOPSIS**

The human protein Pontin, which belongs to the AAA+ family, is overexpressed in several cancers and its silencing *in vitro* leads to tumor cell growth arrest and apoptosis, making it a good target for cancer therapy. In particular, high levels of expression were found in hepatic tumors for which the therapeutic arsenal is rather limited. The 3D structure of Pontin had previously been resolved, revealing an hexameric assembly with one ADP molecule co-crystallized in each subunit. Using Vina, Drugscore and Xscore, structure-based virtual screening of 2,200 commercial molecules was conducted into the ATP binding site formed by a dimer of Pontin in order to prioritize the best candidates. Complementary to the *in silico* screening, a versatile and sensitive colorimetric assay was set up to measure the disruption of the ATPase activity of Pontin. This assay allowed the determination of inhibition curves for more than twenty top scoring compounds, resulting in the identification of four ligands presenting an inhibition constant in the micromolar concentration range. Three of them inhibited tumor cell proliferation. The association of virtual screening and experimental assay thus proved successful for the discovery of the first small molecule inhibitors of Pontin.

**ABBREVIATIONS**

AAA+ family, ATPases Associated with various cellular Activities family; TIP, TATA-binding protein-Interacting Protein; RUVBL, RuvB like; ATPase, Adenosine Triphosphatase; ADME-Tox, Absorption Distribution Metabolism Excretion and Toxicity; PDB, Protein Data Bank; IPTG, isopropyl- $\beta$ -D-thiogalactopyranoside; PMSF, *para*-methanesulfonyl fluoride; Pi, inorganic Phosphate; Akt, Protein Kinase B; PTP1B, Protein Tyrosine Phosphatase 1B.

## INTRODUCTION

Pontin, also known as TIP49 or RUVBL1, and its homolog Reptin (TIP48, RUVBL2), with which it shares 40% sequence identity and 65% homology [1], belong to the AAA+ family and display homologies with the bacterial RuvB helicase [2]. The structure of Pontin shows characteristic ATPase Walker A and B domains, the Walker A motif plays a role in nucleotide binding and in metal-ion coordination, while the Walker B domain contains residues involved in metal-ion coordination and ATP hydrolysis. Additional domains such as the sensors 1 and 2, and the Arg finger (Arg357) interact with the  $\gamma$ -phosphate and play an important role in inter-subunit communication and interactions [3]. The arrangement in hexameric complexes, shared with other AAA+[4], was deduced from X-ray diffraction analysis and from electron microscopy experiments either for Pontin alone or in complex with Reptin in yeast and human [5,6,7]. The three dimensional reconstruction of electron microscopy pictures revealed a dodecameric edifice with a central cavity forming a tunnel compatible with the diameter of potential polynucleotides. As confirmed by the crystallographic structure of Pontin alone published by Matias in 2006 [3], this complex was supposed to be composed of an homohexamer of Pontin superimposed on top of an homohexamer of Reptin, for which no 3D structure existed. However, another structure has been recently described by the same team, displaying important differences with the original one, mainly in the hexameric arrangements which are not composed of Pontin alone [8]. In fact, the proposed structure presents an heterohexamer composed of alternating subunits of Pontin and Reptin which will probably have a great significance in the mechanistic comprehension of this complex.

Most publications agree that human Pontin and Reptin are indeed endowed with an ATPase activity either alone, or in heteromeric complexes [9]. Pontin and Reptin belong to several multi-protein complexes in the nucleus [10], where they are thought to participate to chromatin remodeling [11,12], double-strand break DNA repair [13] and regulation of transcription [14]. They notably interact with the oncogenic transcription factors  $\beta$ -catenin and c-myc and modulate their activities [15,16]. They are also involved in the biogenesis and assembly of ribonucleoprotein complexes such as snoRNPs [17,18] and telomerase [19].

We have previously reported that both Pontin [20] and Reptin [21] were overexpressed in human hepatocellular carcinoma. Overexpression of these proteins was also found in a number of other human cancers [9]. In vitro silencing of either Pontin or Reptin led to tumor cell growth reduction [20,21]. Furthermore in vivo silencing of Reptin in xenografted tumors dramatically reduced tumor progression [22]. These features indicate that Pontin and Reptin could be good candidates for cancer therapy.

With the use of Walker A or B mutants of Pontin devoid of ATPase activity, several authors

have suggested that the ATPase activity of Pontin is required for growth and viability in yeast cells [1,23,24]. The same method allowed to show that the Asp302Asn Walker B mutant of Pontin did not support telomerase biogenesis in human cells [19] and inhibited cell transformation by several oncogenes like c-myc [16],  $\beta$ -catenin [25] or E1A [26]. Therefore, targeting the ATPase activity of Pontin appears to be a suitable strategy against cancer.

In this study, we present two complementary approaches for the discovery of inhibitors of Pontin. No report of small molecule inhibitors for Pontin or Reptin has been made to date, since a major drawback encountered by several teams in the design of an enzymatic assay for Pontin was the low activity of the protein, if any, when detected by the hydrolysis of radiolabeled ATP, yielding ADP and inorganic phosphate [3,6,27,28,29]. On top of that, this method is time consuming due to thin layer chromatography, presents difficulties due to the short half-life of  $^{32}\text{P}$ , and is slightly harmful and uneasy to handle in a standard laboratory. An easy to handle colorimetric assay that only requires a spectrophotometer was designed in order to overcome these limitations and allowed the identification of twenty-nine inactive compounds from the French Chimiothèque Nationale. These molecules presented no inhibitory effect on the ATP hydrolysis displayed by Pontin.

In order to select potential inhibitors, we took advantage of the three-dimensional structure of the hexameric complex of Pontin with bound ADP, that has been resolved by X-ray crystallography [3], to provide a model for structure-based virtual screening of commercial chemical compounds in the ATPase catalytic center of Pontin. The main criterion for the database selection was that the molecules were to be available for experimental testing, which is why two commercial databases of ligands were used in the virtual screening: Calbiochem database of Inhibitors® from Merck, and Prestwick Chemical Library®. Another advantage was that all the molecules from these databases display a known activity and are biologically relevant or “drug-like”, plus their ADME-Tox data are documented. The dockings were conducted with Vina [30], then the compounds were rescored with Drugscore [31,32] and Xscore [33] and the top scoring ligands were selected via consensus scoring.

Finally, priority compounds identified via virtual screening *in silico* were tested in the enzymatic assay *in vitro* for their activity against the ATPase activity of human recombinant Pontin, then in cultured tumor cells for their anti-proliferative effect.

## EXPERIMENTAL

### Material

HPLC purified oligonucleotides were from Eurogentech (Seraing, Belgium). Selected molecules were from Merck Bioscience, and Prestwick. Protease inhibitors tablets were from Roche. The pET21- N-ter His<sub>6</sub> TIP49 plasmid was a kind gift of I.R. Tsaneva (University College London). Superdex G 200 HR column was from Amersham Biosciences. Ni-NTA superflow cartridge was from Qiagen. Rottlerin, PTP1B inhibitor (3-(3,5-Dibromo-4-hydroxy-benzoyl)-2-ethyl-benzofuran-6-sulfonicacid-(4-(thiazol-2-ylsulfamyl)-phenyl)-amide) and Akt inhibitor 1/2 (1,3-Dihydro-1-(1-((4-(6-phenyl-1H-imidazo[4,5-g]quinoxalin-7-yl)phenyl)methyl)-4-piperidinyl)-2H-benzimidazol-2-one) were from Calbiochem (catalog #557370, 539741 and 124017, respectively). Pranlukast (N-[4-oxo-2-(1H-tetrazol-5-yl)-4H-chromen-7-yl]-4-(4-phenylbutoxy)benzamide) was from Prestwick. Imidazole buffer solution and other reagents were from Sigma.

The oligonucleotides used in these studies are listed below (underlined are complementary):

-Single strand (43 mer)

**5'GCT CGC TAC CCG GGG ATC CTC TAG AGT CAT CAG TGC AAG ACC G 3'**

-Hairpin (44 mer)

**5'TCG CTC TTC TAC TAT GAA CCC CCC TCC CC ATT TTT GGG GAG GGG 3'**

-Double stranded (64 mer)

**5' TCG CTC TTC TAC TAT GAA CCC CCC TCC CC ATT TTT GGG GAG GGG GGT TCA  
TAG TAG AAG AGC GA 3'**

### Receptor preparation for docking

The crystal structure of Pontin bound with ADP, available from the RCSB Protein Data Bank (PDB code 2C9O) was used for docking [3]. Except for two loops from residues 142 to 155 and 248 to 276, the structure was complete. Those loops were far enough from the region of interest and therefore were not reconstructed prior to calculations. The structure was visualized with Discovery Studio 2.1 (Accelrys, San Diego, USA). All water molecules were removed and missing hydrogen atoms were added using Charmm forcefield. A minimization of the structure was conducted in the presence of bound ADP, using a Steepest Descent algorithm, 2000 steps with a 0.01 gradient, keeping the backbone of the protein fixed. The ADP binding site is located at the interface between two subunits, therefore, the docking was conducted on a dimer of Pontin to take into account potential interactions with the second subunit.

### Ligand preparation for docking

The chosen commercial chemical databases are a collection of compounds selected for their high chemical and pharmacological diversity, as well as their documented bioavailability and safety in humans (according to the manufacturer). The mean Tanimoto coefficients are 0.1839 and 0.1265 respectively for Calbiochem and Prestwick, which mean a high structural diversity (calculated with OpenBabel 2.2.99 and FP2 from Daylight)[34].

Compounds from the Calbiochem Database of Kinase Inhibitors® were obtained in SD (structure file) format from the Merck Chemicals website (catalogue numbers are 539743, 539744, 539745 and 539746). The three dimensional structures in MOL2 format were automatically generated using Catalyst in Discovery Studio via the *Prepare Ligands* protocol. All parameters were turned to False, except for Change Ionization, the maximum pH was set to 8.5 and the minimum to 6.5 in order to generate different protonation states. The structures obtained were then filtered with DBfilter 2.2.8 (A drug-like analyzer for chemical library. Distributed by the author Wang S.-H. 2005, unpublished). Molecules with a molecular weight lower than 200 or higher than 800 were rejected, as well as molecules with more than 10 rotatable bonds. Only standard atoms such as H, C, N, O, F, Cl, Br, I, P, S were allowed. All compounds in MOL2 format were translated into PDBQT files suitable for docking with the script *prepare\_ligand4.py* from MGLTools 1.5.4 [35] .

The same treatment was applied to the compounds from Prestwick Chemical Library® (<http://www.prestwickchemical.com/index.php?pa=26>) available on request in SD format. The structures and PDBQT files were automatically generated using Discovery Studio and the *prepare\_ligand4.py* script.

Twenty-nine compounds from the French Chimiothèque Nationale were visually selected for experimental evaluation of their inhibition potential of Pontin ATPase. Because of a non-disclosure agreement, the authors cannot display these structures but the identifications are available upon request. These molecules were manually constructed in Discovery Studio, and *Dreiding Minimize* was used to reach a low energy conformation. As described for the compounds from Calbiochem, the PDBQT input files were automatically prepared with the *prepare\_ligand4.py* script. These compounds were incorporated into the Calbiochem database and used to evaluate the scoring functions on their ability to discriminate decoys from potential actives.

Charges automatically assigned to the ligands during the *Prepare Ligands* protocol are not equivalent to the ones assigned via *Dreiding Minimize*, but these charges are irrelevant, since Vina and all the scoring functions used here recalculate their own partial charges for the ligands during calculation.

## METHODS

### Docking and scoring

Autodock Vina 1.0.2 [30] was used for all the dockings in this study. Vina was derived from Autodock, but it achieves improvements in speed and accuracy as compared to the latest release of Autodock (Autodock4, [36]). Besides Vina calculations are run directly from the command line and automatically take advantage of multiple cores [37]. The ligands were docked using Vina, and then rescored thanks to the standalone programs Drugscore and Xscore. For each ligand in the database five docking experiments were conducted, each of them generating up to nine poses. The results from the five experiments were averaged and the compounds were then ranked based on their average scores for each individual scoring function or sorted via consensus scoring.

**Rigid dimer:** The ligands were considered flexible while the protein was held rigid. The docking grid was designed in order to include the whole cavity surrounding ADP in the crystallographic structure plus a margin of at least 3 Å in all directions. The resulting dimensions of the box were 20×22×20 Å<sup>3</sup>. Seventy four amino acids belonging to the first subunit and six amino acids belonging to the second subunit are fully or partially located inside the box. The parameters were kept at their default value. Each docking was conducted five times with five fixed seed numbers that were used for all ligands. The scores obtained for each function are the averages of these five experiments.

**Flexible dimer:** The system and parameters were similar to the rigid dimer, expect that eleven torsions from four different residues were considered rotatable during docking.

**Blind docking:** The blind docking was conducted on a monomer of Pontin. The parameters were kept at their default values once again, but the search space was considerably extended to encompass the whole protein. The box dimensions were 58×76×70 Å<sup>3</sup>.

The docking output PDBQT-formatted files were translated into PDB with in-house scripts and non-polar hydrogens were added in MOL2 format with OpenBabel 2.2.3 [38]. All poses generated by Vina were rescored using scoring functions from Drugscore and Xscore, resulting in ten different scores for each pose.

The *Consensus scoring* protocol from Discovery Studio was used to create a list of priority compounds.

### Post experimental check

Manual checking of the three-dimensional structures of the ligands was performed on the top scoring molecules only, after calculations. The automatic generation of the structures had eventually

produced some errors, mostly on heterocyclic aromatic features. Therefore, all molecules included in the best consensus for both databases were visually checked for errors, corrected, and redocked if needed prior to the identification of compounds selected for experimental screening.

### Enzyme purification

Human recombinant Pontin was purified from 2 liters of *E. coli* BL21 essentially as described [6] in 1-2 working days, with a two-step chromatographic procedure. Briefly, following 3 hours induction at 20°C with 1 mM IPTG, cells were lysed in 40 mL of buffer A containing 20 mM Tris-HCl pH 7.5, 300 mM NaCl, 10% Glycerol, 1 mM PMSF, 1 mM 2-β-mercapto-ethanol, protease inhibitor tablet, 0.5% Nonidet P 40 and 20 mg/ml lysozyme. Upon incubation for 15 min, the lysate was extensively sonicated and cleared by centrifugation for 15 min at 45000 g. The supernatant was loaded onto a 5 ml Nickel-NTA superflow cartridge column and washed with buffer B (20 mM Tris-HCl, pH 7.5, 20 mM imidazole, 300 mM NaCl, 10% glycerol, 1 mM 2-β-mercaptoethanol). The eluted fractions collected upon a 250 mM step with buffer B containing 500 mM imidazole, were then purified onto a Superdex G200 (1×30 cm) in buffer C (buffer B without imidazole). Fractions were analyzed by SDS PAGE, pooled, and dialyzed overnight in buffer C containing 50% glycerol, then stored at -80°C.

### Enzymatic assays

The hydrolysis of ATP into ADP and Pi is measured by the change in absorbance of the dye Malachite green in the presence of phosphomolybdate complexes using the PiColorLock Gold assay from Gentaur (Kampenhout, Belgium).

The enzymatic reactions were performed in a final volume of 100 μL at 37°C. The assay contained a final concentration of 20 mM Tris-Acetate buffer pH 7.5, 10 mM Mg Acetate, 0.5 mM 2-β-mercaptoethanol, 100 pmol DNA, 10% DMSO and catalytic amounts of Pontin. Various serial dilutions of inhibitors dissolved in DMSO were added to a final volume of 80 μL. The reactions were started by the addition of 20 μL of the required final concentration of ATP, and its hydrolysis into ADP and inorganic phosphate (Pi) was detected by the colorimetric assay. The reaction rates were measured by taking aliquots at several times and distributed in plates of 96 wells each containing 20 μL of Picolor Gold and 60 μL H<sub>2</sub>O for at least 10 min at room temperature. The stain was then stabilized with 10 μL of “Stabilizer” during 30 min at room temperature. The OD<sub>620</sub> was measured using a spectrophotometer (Labsystems Multiskan Bichromatic) and correlated with a standard curve of inorganic phosphate. The slopes were determined, thus providing the reaction rates. All experiments were done in triplicates.

The reproducibility of the assay was tested by measuring the Z'-value. This was done by

measuring the reaction rates during thirty minutes with  $10^{-6}$  M Pontin, and 1 mM ATP. The  $Z'$ -value was deduced from 48 reactions rates and found equal to 0.7. This value indicates the good reproducibility of the assay since accepted values of  $Z'$ -lie between 0.5 and 1 [39].

### **Inhibition studies**

Molecules were dissolved in DMSO. The final amount of DMSO was 10% in the assays and we checked that it did not interfere with the reaction rates. We also checked that at the highest tested concentration of 50  $\mu$ M, none of the molecules interfered with the assay in the presence of Pi (not shown). For each concentration of inhibitor, the reaction rates were measured upon addition of 50  $\mu$ M ATP. The  $IC_{50}$  (half maximal inhibitory concentration) was determined from two enzyme purifications. The reported  $IC_{50}$  are the mean values of three independent experiments

### **Cell proliferation experiments**

The human hepatic tumor cell lines HuH7 and Hep3B were used. Cells were seeded at an initial density of 2500 cells/well in 96-well plates in DMEM with 10% foetal calf serum. On the following day, they were treated with dilutions of test molecules in DMSO. The DMSO concentration was adjusted to 1% in every well. After 4 days, cell numbers were estimated colorimetrically at 492 nm with the CellTiter 96 Aqueous One Solution Cell Proliferation Assay (MTS assay) from Promega. Growth index was calculated as shown below (equation 1):

$$Growth\ Index = \frac{D4_T - D0}{D4_{DMSO} - D0}$$

where  $D4_T$  refers to the OD with the test molecule at Day 4,  $D0$  is the OD at Day 0,  $D4_{DMSO}$  is the OD at Day 4 with DMSO alone.

## RESULTS AND DISCUSSION

### Docking strategy

As mentioned before, no Pontin inhibitor had ever been reported in the literature, but according to the enzymatic assay described below, an initial test on analogues of ATP had allowed us to identify twenty-nine inactive compounds (data not shown). The ability to discriminate these decoys presented by our scoring functions either independently or in conjunction was used as a tool in order to evaluate the correctness of our results. At first, 900 compounds from Calbiochem database plus the twenty-nine decoys from the French Chimiothèque Nationale were used as a training set. Then, 1299 compounds from Prestwick were screened with regards to the results obtained with the Calbiochem training set.

Prior to the docking of the databases, ATP and ADP were docked into the active site to evaluate the quality of the model. The resulting poses were very similar to crystallized ADP (Figure 1). In particular the adenine groups and the ribose of both molecules superimposed with that of crystallized ADP, while only the highly flexible phosphate chain of ATP was notably displaced in some of the poses. At least fifteen residues were involved in hydrophobic contacts with both ADP and ATP. Besides, the phosphate chain of ADP was stabilized by a network of eight hydrogen bonds coming from six different residues, while that of ATP could create up to thirteen hydrogen bonds with eight residues including five out of the six found in the case of ADP.

### Constitution of priority lists of compounds by consensus scoring

Considering the very few experimental results, the approach based on individual scoring function was very likely to produce a large number of misranks. In this context, a consensus scoring strategy seemed a reasonable choice that could result in a reduction in the number of false negatives when selecting a limited number of molecules for experimental testing, [40,41] .

When considered individually, scoring functions from Drugscore CSD completely failed at eliminating the decoys from the top scorers while Vina, Drugscore PDB and Xscore provided far better results. In conjunction, a very good consensus was obtained with HPscore from Xscore and SURF from Drugscore, which provided a large number of common top compounds including only one decoy, and adding Vina in the consensus allowed to eliminate the last decoy in the top list. This improvement was followed by a significant decrease in the number of compounds in the intersection, but considering that the dockings were made with Vina and that its scoring function was used to generate the poses of the ligands, we decided that Vina scoring function should be included in the consensus (see Supplementary Information for detailed description of the virtual screening results).

The consensus carried out with these three independent scoring functions on the Calbiochem database led to the identification of twenty-two priority compounds ranking in the top 15% for the three functions. Ten of them, that is approximately 1% of the complete database, were randomly selected for experimental screening. The same procedure was applied to the 1299 compounds from the Prestwick library, leading to the prioritization of twenty-three new molecules and approximately the same percentage, namely fifteen compounds, were chosen for testing.

### **Design of a versatile ATPase assay to measure Pontin activity**

Human recombinant Pontin was purified via chromatography and the eluted fractions yielded variable amounts of dodecameric, hexameric and monomeric forms of Pontin as shown by elution volume and SDS PAGE. When tested in the ATPase assay described below, only the purified monomeric fraction (Figure 2A) proved active, therefore all enzymatic experiments were conducted on the monomeric fraction.

A Malachite green-derived assay that quantifies the green complex formed between Malachite Green, molybdate and free orthophosphate was used to measure the enzyme activity in 96 wells plates. As shown on Figure 2B, OD<sub>620</sub> was proportional to the Pi concentration. The reaction rates were proportional to the enzyme concentration, and did not display apparent cooperativity (Figure 2C).

Almost no variation of the reaction rate of the ATPase catalysis could be detected in assays performed in the absence of oligonucleotides. In contrast, upon addition of an excess of single or double-stranded DNA ([DNA]/[E]=1-10), a threefold stimulation of the reaction rate was observed as shown on Figure 2D (p<0.01). The effects of a single-stranded DNA of 43 mer, a hairpin of 44 mer and a perfect duplex DNA of 64 mer were studied. No significant differences were found whichever polynucleotide was used (data not shown). The ATP hydrolysis reaction rate stimulation in presence of DNA correlates with the presence of a polynucleotide binding site detected by gel shift assays with Pontin [3]. This activation by DNA is controversial in the literature, since it has been reported by many groups [1,42,43,44] while it could not be detected by others [3,6,7]. However, the presence of DNA was essential in our assay to detect a sufficient activity in order to measure the IC50 of the ligands.

In order to test for potential inhibitors, the ATP concentration corresponding to the initial velocity was determined. Catalysis followed an asymptotic curve, allowing the determination of the Km (Michaelis constant) with the Michaelis-Menten equation (equation 2):

$$V = \frac{V_{max} \times [S]}{K_m + [S]}$$

Using the double-reciprocal plot described by Lineweaver-Burk, the Km was found to be equal

to 50  $\mu\text{M}$  ( $\pm 15 \mu\text{M}$ ) as shown on Figure 2E. Thus, the inhibition curves of the compounds were obtained using a concentration of ATP of 50  $\mu\text{M}$ , and the  $\text{IC}_{50}$  as well as the resulting inhibition constant  $K_i$ , could be calculated thanks to the Cheng-Prusoff relationship (equation 3):

$$K_i = \frac{\text{IC}_{50}}{1 + \frac{[S]}{K_m}}$$

with  $[\text{ATP}]/K_m$  equals 1 [45] .

### Experimental inhibition

All compounds were initially tested at a 50  $\mu\text{M}$  concentration. When no significant variation of the reaction rates were measured (i.e. about 30%), we considered the compounds as inactive.

Akt inhibitor 1/2 inhibited Pontin ATPase activity with an  $\text{IC}_{50}$  of 24  $\mu\text{M}$  (Figure 3A). This compound was reported to inhibit preferentially the Akt1 kinase ( $\text{IC}_{50}$  58 nM) with also a good potency against Akt2 (210 nM) and less against Akt3 (2.1  $\mu\text{M}$ ).

A non-competitive specific inhibitor of the PTP1B phosphatase presented an  $\text{IC}_{50}$  of 15  $\mu\text{M}$  (Figure 3B), as compared to 4  $\mu\text{M}$  for its primary target.

Rottlerin inhibited Pontin activity with an  $\text{IC}_{50}$  of 10  $\mu\text{M}$  (Figure 3C). This molecule was also reported as a reversible inhibitor of the PKC family although this has been questioned and it may also inhibit several MAPK family members [46] Using the Cheng- Prusoff equation (see equation 3 [45]) the  $K_i$  of these 3 molecules was determined, and found to be equal to 12.1, 7.5 and 5.1  $\mu\text{M}$  respectively.

From the fifteen compounds prioritized in the Prestwick chemical collection, only Pranlukast, a potent and specific competitive antagonist of the cysteinyl leukotriene 1 receptor, inhibited the ATPase reaction with an  $\text{IC}_{50}$  of 13  $\mu\text{M}$  (Figure 3D), thus with a  $K_i$  of 6.5  $\mu\text{M}$ .

The chemical structures of these compounds are shown in Figure 4.

### Binding poses analysis

The observation of the active site topology led to the distinction of two spaces. The flattest part of the cavity was occupied by ADP and ATP while the other side, which includes all residues from the second subunit, was left empty. The best docking poses for all active compounds shared common features with ADP and ATP (Figure 5).

Out of the seventeen residues involved in the binding mode of ADP, ten were also implicated in the binding of all inhibitors. The number of residues common to the binding mode of ADP and to the ligands respectively ranged from eleven for Rottlerin, to thirteen for Pranlukast and fourteen for Akt and PTP1B inhibitors. Besides, various residues from the Walker A domain (residues 70 to 79)

were involved in the binding mode of all ligands as well as ADP and ATP, with at least four different residues out of ten in close proximity to the ligands. As for the Walker B domain (residues 302 to 305), all inhibitors and ATP interacted with Asp<sup>302</sup> only, while ADP did not make contact with it at all. This residue is crucial for the hydrolysis of ATP, as shown by the lack of ATPase activity for the Asp<sup>302</sup>Asn mutant [3] .

In all cases, an aromatic cycle lied in a position close to that of the adenine. Furthermore Pranlukast and the PTP1B inhibitor presented an hydrogen bond acceptor group superimposed to the  $\gamma$ -phosphate of ATP, and created hydrogen bonding with Gly<sup>73</sup> similar to both ATP and ADP. On the contrary, the pose of the Akt 1/2 inhibitor was preventing all hydrogen bonding with this side of the cavity. Rottlerin was the only ligand to form an hydrogen bond with Arg<sup>404</sup> from the sensor 2, like in the case of ATP.

As opposed to ADP or ATP, all compounds created interactions with the second area of the cavity. This included hydrophobic contacts with the second subunit, through Asp<sup>353</sup> for Rottlerin, as well as Asp<sup>356</sup> for all ligands. Besides, Pranlukast and the PTP1B inhibitor made electrostatic interactions with the Arg finger, that is the Arg<sup>357</sup> from the second subunit. The latter even accepted two hydrogen bonds between its sulfonamide and the guanidine group from the Arg finger.

A flexible docking was also conducted on the dimer, meaning that the rotations on the side chains of the residues located in a sphere of 5 Å around the center of mass of ADP were authorized during docking. This method aims at rendering the rearrangements induced by the binding of a ligand into the structure of the active site (induced fit), but the major drawback is the calculation time which increases dramatically with the number of allowed torsions. In the model, the active site was completely buried and highly constrained. The authorized movements on the side chains were thus very limited. As expected, we have observed no noticeable modifications in the results obtained or in the poses as compared to those obtained with the non-flexible dimer.

### **Type of enzyme inhibition**

The kind of inhibition presented by these compounds was determined by measuring the rates of ATP hydrolysis with the green malachite assay, using five concentrations of ATP, ranging from  $K_m$  (50  $\mu$ M) to 1 mM, together with increasing concentrations of inhibitors.

The double reciprocal plots (Figure 6) showed that Rottlerin was the only ligand that inhibited competitively the ATPase activity. In the experiments made with Pranlukast and the PTP1B inhibitor both  $V_m$  and  $K_m$  for ATP hydrolysis varied upon addition of the inhibitors revealing a mixed or uncompetitive inhibition. The analysis of the results obtained with the Akt 1/2 inhibitor revealed a more surprising inhibition profile since the double reciprocal plots indicated that the  $-1/K_m$  was

identical to the one measured for ATP proving that the inhibitor did not compete with the ATP binding site.

The compounds tested experimentally had been selected via structure-based virtual screening. According to the literature, the docking experiments were conducted by targeting the ATP binding site [1,19,23,24] and therefore we expected the inhibitors that were identified to bind competitively to the ATP binding site of Pontin. Yet, among four inhibitors, only Rottlerin appeared to be competitive. Pranlukast and the PTP1B inhibitor presented a mixed or uncompetitive profile and in the case of the Akt inhibitor, the double reciprocal plot indicated a non-competitive inhibitor.

In order to explore other potential binding modes, additional docking experiments were conducted on the monomer of Pontin, in the presence or in the absence of ATP and ADP, considering a search space that encompassed the whole protein (blind docking). In the case of a tandem dimer docking, the search box contained only the binding site of crystallized ADP, which was completely buried since the second monomer was closing the cavity, contributing to the shape of the active site. This subunit, even if it did not create interactions with the compounds, constrained them into a closed area, playing a key role in the ligands binding modes. On the contrary, in the case of the blind docking on the monomer, the ligands had the freedom not only to explore the surface surrounding the cavity that was covered up by the second subunit in the dimer, but also to bind everywhere on the surface of whole protein.

In the case of the Akt inhibitor, similar results were obtained either using Pontin alone or in complex with ADP and ATP. Except for a few poses halfway inside the active site when Pontin was used alone, or situated at the entrance of the active site when ATP was present, almost all the poses observed were located in a groove at the junction between the domain I and the domain II of the protein (Figure 7A). These poses were too far from the active site to interfere directly with the binding of ATP, but they could be interacting with the nucleotide binding site. Actually, the domain II is the seat of the DNA binding site, and since we have shown that the presence of DNA considerably influenced the reaction rate, it is possible that the Akt inhibitor interferes with this feature.

As for Pranlukast, the blind docking have shown that when ATP was complexed with Pontin, a binding mode located at the entrance of the cavity was clearly favored (Figure 7B). This pose could prevent ATP from being hydrolyzed and does not leave sufficient space for ADP to get out. It is also incompatible with the existence of a tandem dimer similar to that observed in crystallography [3], as Pranlukast would create clashes with several residues from the second subunit. On the contrary, no clear pose preference was observed when Pontin was complexed with ADP and the best poses obtained with Pontin alone were located halfway inside the active site which was in contradiction with the experimental data. These results are consistent with the hypothesis of an uncompetitive inhibition

which implies that the enzyme is complexed by its substrate prior to the binding of the inhibitor. The binding of ATP would thus enable Pranlukast to bind to the protein and to inhibit its enzymatic activity.

The blind docking of the PTP1B inhibitor did not reveal such a clear-cut profile. The complexation of ATP did slightly favor the pose at the entrance of the cavity just like in the case of Pranlukast, but numerous other poses were observed as well, and it was hard to define which pose was the most likely to be observed.

Taken together, these experiments would warrant further structural and biochemical studies due to the complexity of Pontin. The competitive inhibition observed with Rottlerin, indicated a pure competition with the ATP binding site which confirmed the *in silico* antagonists selection. Nonetheless, the mixed or uncompetitive inhibition measured with the PTP1B inhibitor and Pranlukast, and the non-competitive nature of the Akt inhibitor, suggest indirect interferences with the ATPase activity of Pontin, either by blocking the hydrolysis of ATP or the exit of ADP, or through interactions with the DNA binding site. Therefore, the modulation of the hydrolysis shown upon addition of DNA may be an important parameter in these processes. Actually, as shown by Mezard [2] and Rottbauer [44], DNA binding and ATPase activities are tightly coupled. Moreover, the catalytic site is formed by two subunits and we can assume that the Arg finger may indirectly modulate the catalytic activity upon inhibitor binding [47]. Similar inference may be suggested by the observations of Zhang [48] and Moffit [49], pointing to inter-subunits interactions.

### **Effects of the inhibitors on proliferation of cultured cell lines**

The hepatic tumor cell lines HuH7 and Hep3B were grown in the presence of these compounds for four days. As can be seen on Fig. 8, Rottlerin had a strong anti-proliferative effect on both cell lines. It was toxic at concentrations higher than 5  $\mu\text{M}$  for HuH7 and 1  $\mu\text{M}$  for Hep3B, as evidenced by cell numbers falling below the Day 0 values. The Akt inhibitor behaved similarly, although it was toxic only at higher concentrations. Pranlukast was not toxic in the range of tested concentrations, and dose-dependently decreased cell proliferation of both cell lines. Calculated  $\text{IC}_{50}$  in HuH7 and Hep3B cells were respectively 0.57 and 0.25  $\mu\text{M}$  for Rottlerin, 4.0 and 3.5  $\mu\text{M}$  for Akt 1/2 inhibitor, and 63.7 and 34.3  $\mu\text{M}$  for Pranlukast. Finally, the PTP1B inhibitor did not show any growth inhibition or cytotoxicity at concentrations up to 100  $\mu\text{M}$ .

Altogether, three out of four molecules reduced cell numbers as seen with Pontin silencing [20], suggesting that they may indeed target Pontin within cells. Because of their cognate targets, it was expectable that Rottlerin and the Akt inhibitor would have such an effect. Since these molecules are likely to act at the same time on those targets and on Pontin, it is difficult to strictly compare their

*in vitro* and *in vivo* IC<sub>50</sub> and we can thus only conclude that their effects are compatible with an effect on Pontin. PTP1B inhibition can lead either to increased or reduced cell proliferation, depending on the context [50]. In our hands, the PTP1B antagonist had no effect on the growth of the tested HCC cell lines. It may be that access of the inhibitor to Pontin is hampered *in vivo*, either because of internalization or solubility issues, or because the *in vivo* conformation of the target prevents the inhibitor to access the catalytic centre. Finally, Pranlukast, an antagonist of the Leukotriene receptor, reduced cell proliferation with a similar potency as for Pontin ATPase activity inhibition. This is of special interest since, contrary to the other three molecules which primary targets are nucleotides-using enzymes, Pranlukast is a competitive antagonist of a non-nucleotide molecule believed to act outside the plasma membrane.

## CONCLUSION

Using *in silico* and *in vitro* complementary approaches, our report discloses the first identification of inhibitors of the ATPase activity of Pontin, an activity required for several biological functions [9].

The first step involved structural modeling of the ATP binding centre, and virtual screening of about 2,200 commercial compounds with Vina. Thorough rescoring of the poses with Xscore and Drugscore, using twenty-nine decoys as a negative calibration set, allowed the selection of efficient scoring functions for this system. Various consensus using these functions were compared and the intersection of Vina, SURF score from Drugscore and HPscore from Xscore scores appeared to provide the best results. With regards to these results, this consensus was used to prioritize twenty-five compounds for experimental testing. This virtual screening strategy proved efficient, since post experimental analysis of the results indicated that the consensus scoring used to discriminate the compounds was the only one that allowed the selection of the four active molecules. Similarly, none of the single functions had detected all four ligands in their top scoring lists.

The enzymatic testing of chemicals selected by docking with a green malachite assay was made possible thanks to improvements in the procedure, and in particular the addition of polynucleotides that increased several-folds the reaction rates. Four ligands displayed an inhibition constant in the micromolar range. The colorimetric assay described here could also be used for high throughput screenings, because of its sensitivity and rapidity as compared to the current time-consuming use of radiolabeled ATP.

Finally, three out of the four tested compounds reduced cell growth at concentrations close to those inhibiting the ATPase activity of Pontin *in vitro*. Whether their anti-proliferative action is due to Pontin antagonism will require further studies.

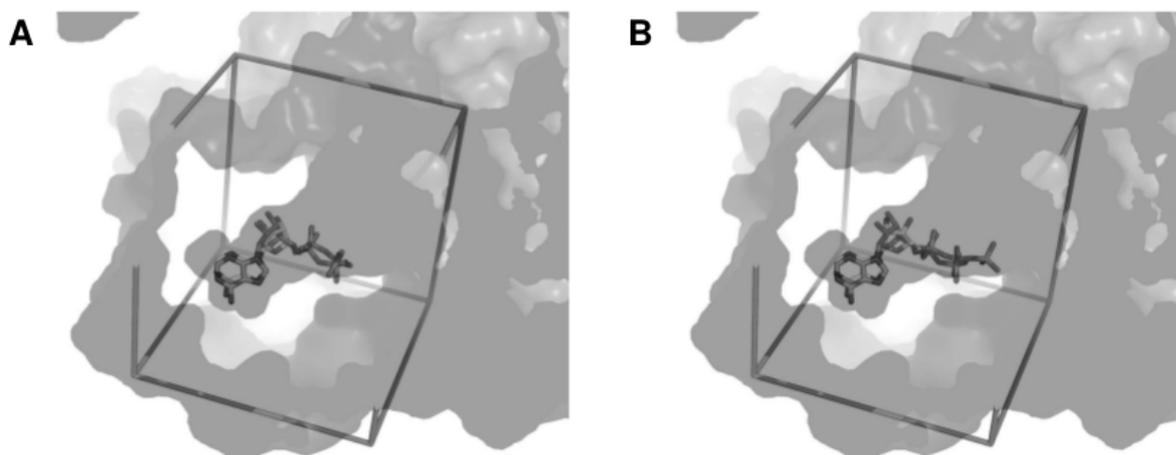
## ACKNOWLEDGEMENTS

We thank the French Chimiothèque Nationale for initial studies. We gratefully acknowledge Dr I.R. Tsaneva (University College London) for the plasmid kind gift and for helpful and friendly discussion. We express our warmest thanks to Dr S.H. Wang (Graduate Institute of Pharmaceutical Chemistry, China Medical University, Taiwan) for providing DBfilter, and Dr M.L. Jung from Prestwick.

## FUNDING

This project was supported by grants from Ligue Nationale Contre le Cancer, Association pour la Recherche sur le Cancer and Institut National du Cancer [PLBIO10-155]. The authors wish to thank the MENRT, CNRS and INSERM for their support.

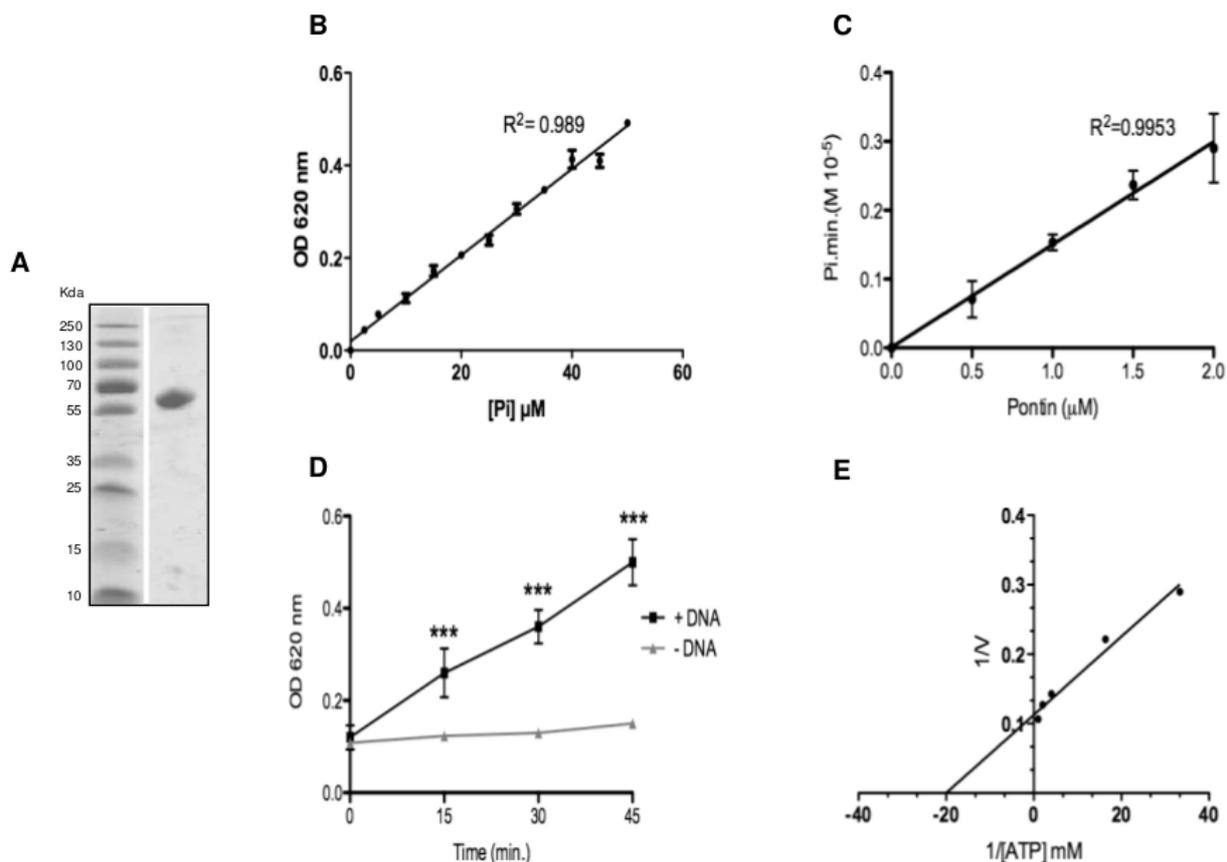
## FIGURES



### Figure 1. Docking of ADP and ATP

(A) Superimposition of crystallized ADP from 2C9O and docked ADP into the dimer of Pontin. The active site is represented in transparency, the search space box is represented in sticks.

(B) Similar representation with docked ATP.



**Figure 2. Enzymatic assays of Pontin ATPase activity**

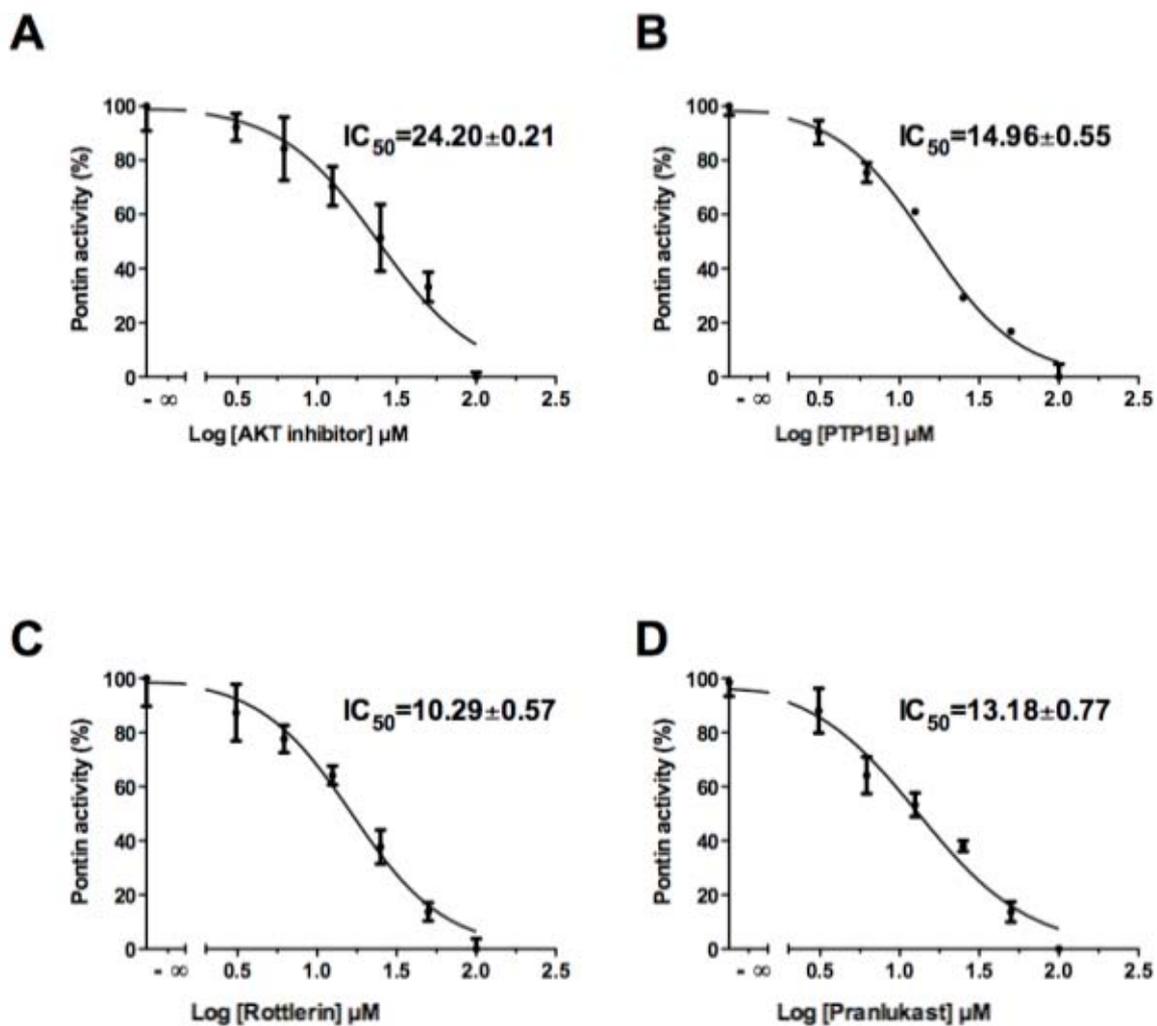
(A) SDS polyacrylamide gel electrophoresis of 5  $\mu$ g of the monomeric Pontin fraction purified by chromatography on Superdex G200, and stained with Coomassie brilliant blue.

(B) Relationship between Pi concentration and OD<sub>620</sub>. Increasing concentrations of Pi were used with the Malachite Green assay, as described in the Methods section.

(C) Reaction rates according to the concentration of Pontin. The reported rates were determined for each concentration of Pontin during a 40 min incubation at 37°C. Aliquots of 20  $\mu$ L were taken every 10 min and the release of Pi was measured as in Figure 2B. The graph shows the mean  $\pm$  SD of 3 experiments.

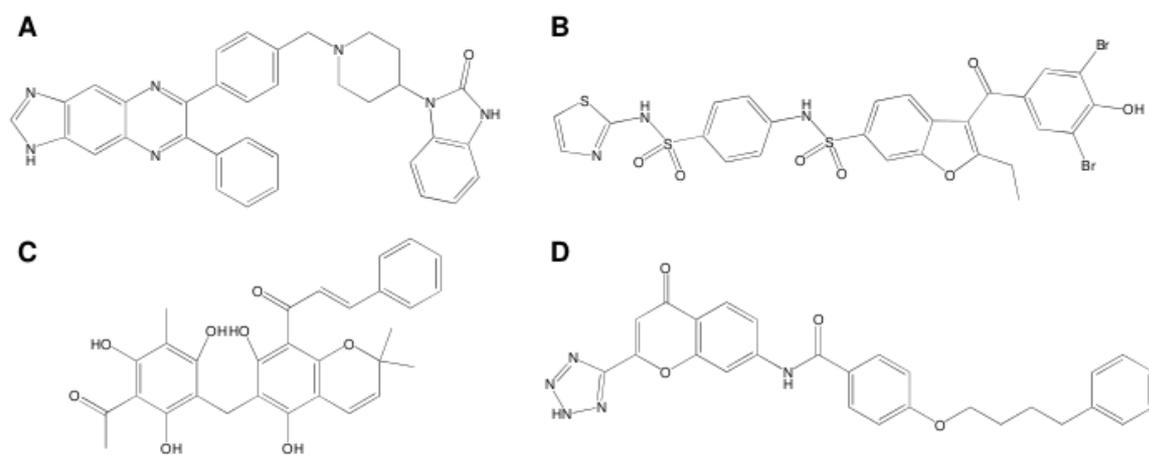
(D) Effect of 10  $\mu$ M DNA (43 mer) on the reaction rates in the presence of 2.5  $\mu$ M Pontin and 2 mM ATP. The graph shows the mean  $\pm$  SD of 3 experiments. The presence of DNA significantly increased the reaction rate (ANOVA,  $p < 0.001$ ). Similar results were found with the other oligonucleotides described in the Materials section.

(E) Lineweaver-Burk representation of the reaction rates according to the ATP concentration in the assays. The concentration of Pontin was 2  $\mu$ M. The graph shows the mean of 3 experiments.



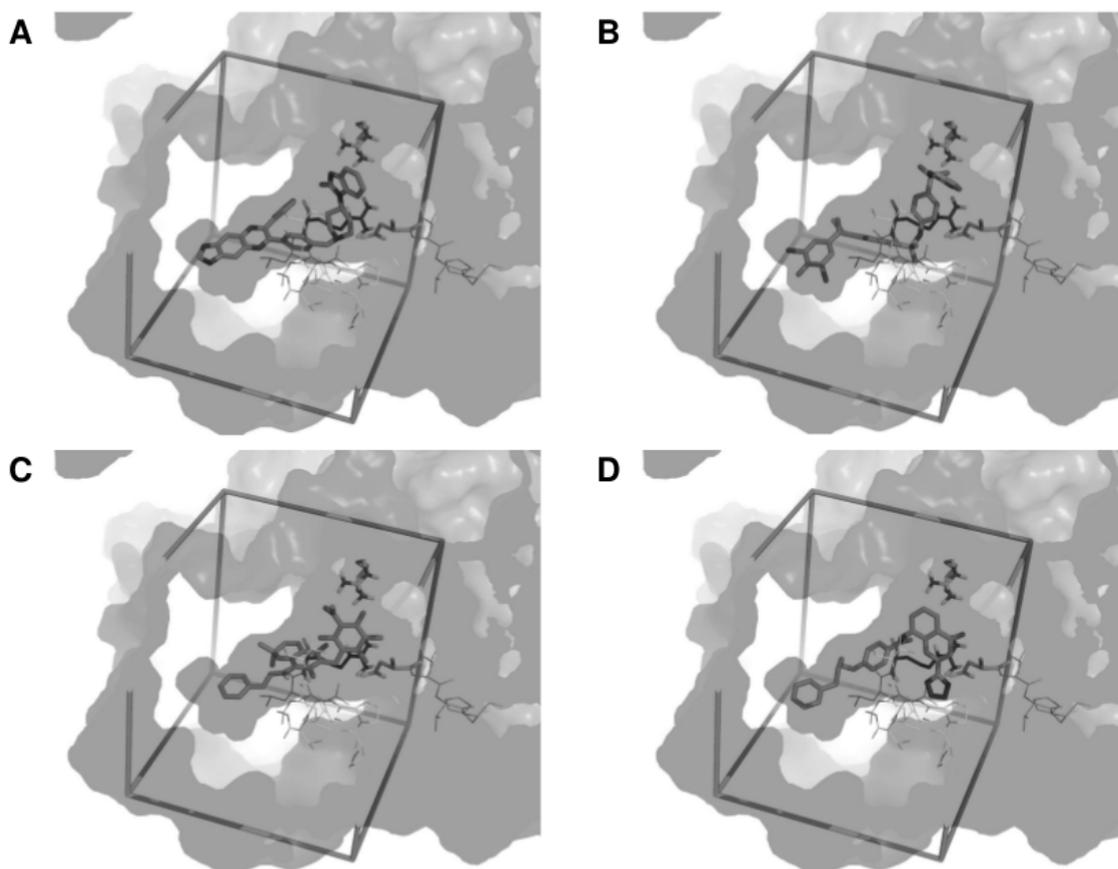
### Figure 3. Identification of four inhibitors

Determination of the  $\text{IC}_{50}$  of the Akt 1/2 inhibitor (A), the PTP1B inhibitor (B), Rottlerin (C) and Pranlukast (D) with 50  $\mu\text{M}$  ATP and 4  $\mu\text{M}$  Pontin. The graphs show the mean  $\pm$  SD of 4 experiments. Some standard deviations cannot be seen because they were too small.



**Figure 4. Chemical structures of the four active compounds**

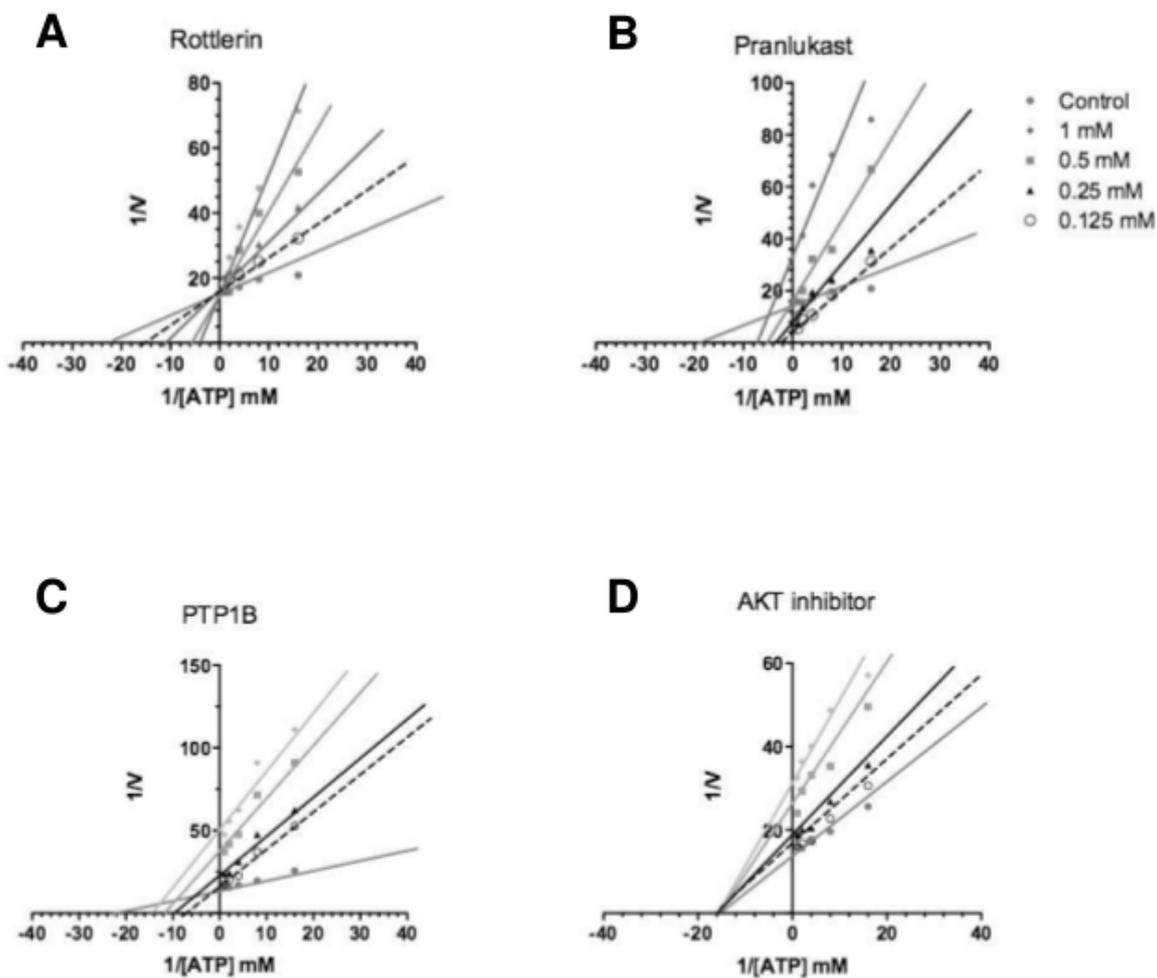
(A-D) Akt 1/2 inhibitor, PTP1B inhibitor, Rottlerin, Pranlukast.



### Figure 5. Docking poses in the dimer of Pontin

(A) The best docking pose for the Akt 1/2 inhibitor is shown in sticks. The active site is represented in transparency with surfaces of both subunits of Pontin shown in different shades of grey. Important domains from the ATP binding site are shown in lines with essential residues in sticks. From the first subunit, the Walker A domain, is in front of the ligands, the Walker B domain is at the right side of the image, with Asp<sup>302</sup> highlighted, and the sensor 2 is at the back of the cavity, with Arg<sup>404</sup> highlighted. From the second subunit, the Arg finger (Arg<sup>357</sup>) is shown in sticks. The search space box is represented in sticks.

(B-D) Similar representation of the PTP1B inhibitor (B), Rottlerin (C) and Pranlukast (D).

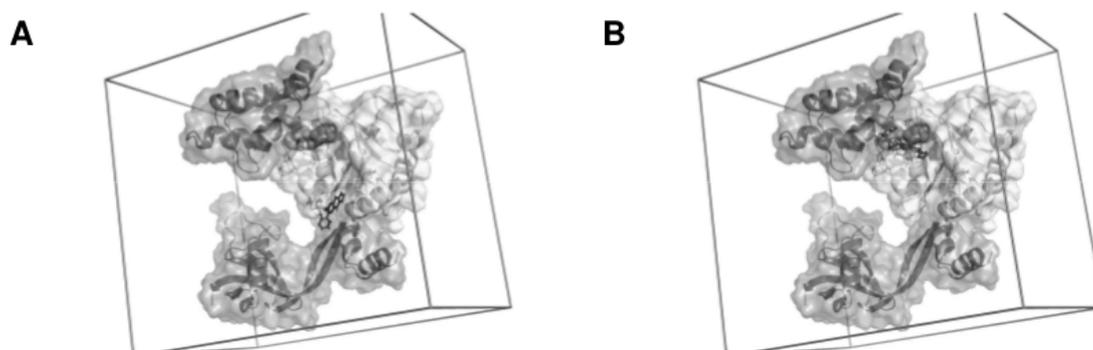


**Figure 6. Type of enzyme inhibition**

(A) Double reciprocal plots (Lineweaver-Burk plots) of the reaction rates with ATP alone, and increasing concentrations of Rottlerin. When no variation of  $1/[ATP]$  is observed, the inhibitor is non competitive and does not modify the binding of ATP. In contrast, when no variation of  $1/V_{max}$  is observed, the ligand is competitive and disrupts the ATP binding to the protein.

(B-D) Similar plots for Pranlukast (B), PTP1B inhibitor (C) and Akt 1/2 inhibitor (D).

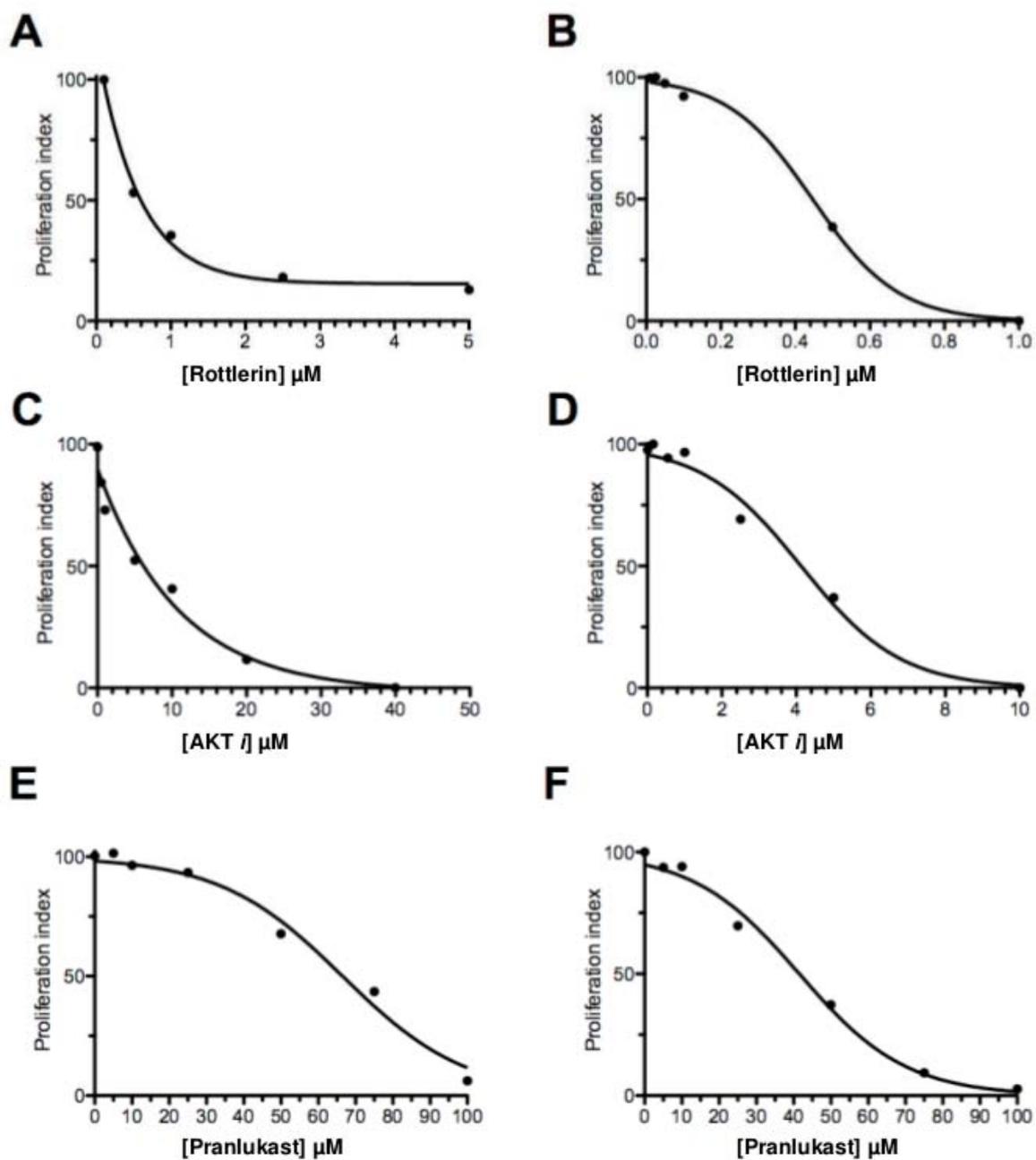
The graphs show the mean of two independent experiments performed in tetraplicates.



**Figure 7. Docking poses obtained via blind docking**

(A) The best docking pose for the Akt 1/2 inhibitor is shown in sticks. The monomer of Pontin is represented in transparent surfaces and cartoons, with domains I, II and III colored in different shades of grey. The inhibitor lays in a groove between domain I and domain II, in close proximity to the DNA binding site.

(B) Similar representation for Pranlukast. The inhibitor is located at the entrance of the catalytic site.



**Figure 8. Cell proliferation assays.**

Effects of Rottlerin (A and B), Akt inhibitor (C and D) and Pranlukast (E and F) on the growth of HUH7 (left column) and Hep3B (right column) hepatocarcinoma cell lines.

Cells were grown in the presence of various concentrations of the inhibitors indicated in  $\mu\text{M}$ . Cell numbers were estimated four days later with the MTS assay and the growth index was calculated as described in the Methods section. The graphs show the mean of two independent experiments performed with 5 replicates each.

## SUPPLEMENTARY ONLINE DATA

### Individual scoring functions

Drugscore provides three scoring functions: PAIR, SURF, and PAIRSURF and two alternative versions of Drugscore are available: Drugscore CSD [31] and Drugscore PDB [32]. For both versions, SURF score is identical and PAIRSURF score is the sum of PAIR and SURF score. Consequently, five distinct functions are available from Drugscore, that will be noted PAIR\_csd, PAIRSURF\_csd, PAIR\_pdb, PAIRSURF\_pdb and SURF. Xscore [33] offers four scoring functions that are HPscore, HMscore, HSscore, and AVEscore, which is the average of the former three.

At first, all scoring functions were evaluated individually by comparing the number of decoys ranked in the best 15% (139 compounds) of the training set (Table 1).

All dockings were performed with Vina [30], meaning that its scoring function was used during the process of poses generation. Therefore, the correctness of this function was a key factor for the accuracy of the whole experiment. Out of the 929 compounds present in the training set, two decoys appeared in the top 15% with the scoring function inherent to Vina.

The functions from Drugscore CSD clearly failed to discriminate the decoys. With PAIRSURF\_csd and PAIR\_csd, five decoys were found in the top 15%. The PDB version of Drugscore performed better with both PAIRSURF\_pdb and PAIR\_pdb ranking only one decoy in the best 15%. Similarly, SURF ranked only decoy in the top 15%.

The best results were obtained with the functions from Xscore. AVEscore and HPscore both ranked only one decoy in the best 15%, and as for HSscore and HMscore, they were very efficient at excluding the decoys from the top, since no decoy was observed in the best 15%.

### Consensus scoring

From now on, a consensus will be noted as follows: “Function1/Function2 X”, with X the percentage of base considered for each function. To discriminate the combinations that were able to exclude the decoys from the top scoring compounds, the consensus were first carried out by keeping up to 30% of the base for each function, that is 279 compounds, and then progressively reducing the percentage of base considered. Using the *Consensus Scoring* protocol from Discovery Studio 2.1 (Accelrys, San Diego, USA), the numerous combinations available with the ten scoring functions were explored. However, the combinations of different scoring functions coming from the same program were not considered since the functions are not independent [41].

The combinations of SURF with all functions from Xscore were explored first, since those functions were the best at discriminating the decoys individually. The consensus obtained showed that the results given by those functions were highly correlated, in particular in the consensus HPscore/SURF 30, where up to 242 compounds out of 279 in the best 30% were common to both top lists, including only one decoy. Similar results were observed with HSscore/SURF 30 with 217 common compounds, among which one was a decoy. HMscore/SURF 30 and AVEscore/SURF 30 gave poorer results, with respectively 204 and 220 molecules shared by both functions, including two decoys. Yet, none of these consensus was able to completely exclude the decoys.

The consensus made with Vina displayed a much lower number of common ligands. In the consensus Vina/SURF 30, 127 molecules were shared by both functions, but two of these were decoys. As for Vina/HMscore 30 and Vina/AVEscore 30, they provided respectively 128 and 121 shared compounds, including one decoy. Vina/HSscore 30 and Vina/HPscore 30 did best at eliminating the decoys from the top since respectively 119 and 125 compounds responded to both individual functions, including zero decoy.

Using three functions, the consensus Vina/HSscore/SURF 30 and Vina/HPscore/SURF 30 gave very similar results, with respectively 100 and 108 compounds responding to the three functions, including no decoy.

## TABLES

**Table 1. Decoy discrimination, single functions.** Number and ranks (into brackets) of the decoys found in the top 15% of the training set (Calbiochem database + decoys).

| <b>Functions</b> | <b>Decoys</b>           |
|------------------|-------------------------|
| Vina             | 2 (51, 109)             |
| PAIR_csd         | 5 (21, 32, 57, 58, 81)  |
| PAIRSURF_csd     | 5 (30, 36, 43, 65, 104) |
| PAIR_pdb         | 1 (51)                  |
| PAIRSURF_pdb     | 1 (53)                  |
| SURF             | 1 (37)                  |
| HPscore          | 1 (95)                  |
| HMscore          | -                       |
| HSscore          | -                       |
| AVEscore         | 1 (83)                  |

**Table 2. Decoy discrimination, consensus.** Number of common ligands in the consensuses 30% and number of decoys found among them (Calbiochem database + decoys).

| <b>Functions</b> | <b>Common</b> | <b>Nb of decoys</b> |
|------------------|---------------|---------------------|
| HP/SURF          | 242           | 1                   |
| HM/SURF          | 204           | 2                   |
| HS/SURF          | 217           | 1                   |
| AVE/SURF         | 220           | 2                   |
| Vina/SURF        | 127           | 2                   |
| Vina/HS          | 119           | -                   |
| Vina/HM          | 128           | 1                   |
| Vina/HP          | 125           | -                   |
| Vina/AVE         | 121           | 1                   |
| Vina/HS/SURF     | 100           | -                   |
| Vina/HP/SURF     | 108           | -                   |

## REFERENCES

- 1 Jonsson, Z.O., Dhar, S.K., Narlikar, G.J., Auty, R., Wagle, N., Pellman, D., Pratt, R.E., Kingston, R., and Dutta, A. (2001). Rvb1p and Rvb2p are essential components of a chromatin remodeling complex that regulates transcription of over 5% of yeast genes. *J. Biol. Chem.* 276, 16279-16288.
- 2 Mezard, C., Davies, A.A., Stasiak, A., and West, S.C. (1997). Biochemical properties of the RuvBD113N: a mutation in helicase motif II of the RuvB hexamer affects DNA binding and ATPase activities. *J. Mol. Biol.* 271, 704-717.
- 3 Matias, P.M., Gorynia, S., Donner, P., and Carrondo, M.A. (2006). Crystal structure of the human AAA+ protein RuvBL1. *J. Biol. Chem.* 281, 38918-38929.
- 4 Ammelburg, M., Frickey, T., and Lupas, A.N. (2006). Classification of AAA+ proteins. *J. Struct. Biol.* 156, 2-11.
- 5 Gribun, A., Cheung, K.L., Huen, J., Ortega, J., and Houry, W.A. (2008). Yeast Rvb1 and Rvb2 are ATP-dependent DNA helicases that form a heterohexameric complex. *J. Mol. Biol.* 376, 1320-1333.
- 6 Puri, T., Wendler, P., Sigala, B., Saibil, H., and Tsaneva, I.R. (2007). Dodecameric structure and ATPase activity of the human TIP48/TIP49 complex. *J. Mol. Biol.* 366, 179-192.
- 7 Torreira, E., Jha, S., Lopez-Blanco, J.R., Arias-Palomo, E., Chacon, P., Canas, C., Ayora, S., Dutta, A., and Llorca, O. (2008). Architecture of the pontin/reptin complex, essential in the assembly of several macromolecular complexes. *Structure* 16, 1511-1520.
- 8 Gorynia, S., Bandejas, T.M., Pinho, F.G., McVey, C.E., Vornrhein, C., Round, A., Svergun, D.I., Donner, P., Matias, P.M. and Carrondo, M.A. (2011). Structural and functional insights into a dodecameric molecular machine - The RuvBL1/RuvBL2 complex. *J. Struct. Biol.* [Epub ahead of print] DOI:10.1016/j.jsb.2011.09.001. *This article presenting a new structure for the hexameric assembly was released online while our work was submitted. Therefore, it was not taken into account in our simulations.*
- 9 Grigoletto, A., Lestienne, P., and Rosenbaum, J. (2011). The multifaceted proteins Reptin and Pontin as major players in cancer. *Biochim. Biophys. Acta* 1815, 147-157.
- 10 Jha, S., and Dutta, A. (2009) RVB1/RVB2: running rings around molecular biology. *Mol. Cell* 34, 521-533.
- 11 Jin, J., Cai, Y., Yao, T., Gottschalk, A.J., Florens, L., Swanson, S.K., Gutierrez, J.L., Coleman, M.K., Workman, J.L., Mushegian, A., Washburn M.P., Conaway, R.C., and Conaway, J.W. (2005). A mammalian chromatin remodeling complex with similarities to the yeast INO80 complex. *J.*

Biol. Chem. 280, 41207-41212.

12 Shen, X., Mizuguchi, G., Hamiche, A., and Wu, C. (2000). A chromatin remodelling complex involved in transcription and DNA processing. *Nature* 406, 541-544.

13 Jha, S., Shibata, E., and Dutta, A. (2008). Human Rvb1/Tip49 is required for the histone acetyltransferase activity of Tip60/NuA4 and for the downregulation of phosphorylation on H2AX after DNA damage. *Mol. Cell Biol.* 28, 2690-2700.

14 Gallant, P. (2007). Control of transcription by Pontin and Reptin. *Trends Cell. Biol.* 17, 187-192.

15 Bauer, A., Chauvet, S., Huber, O., Usseglio, F., Rothbacher, U., Aragnol, D., Kemler, R., and Pradel, J. (2000). Pontin52 and reptin52 function as antagonistic regulators of beta- catenin signalling activity. *Embo J.* 19, 6121-6130.

16 Wood, M.A., McMahon, S.B., and Cole, M.D. (2000). An ATPase/helicase complex is an essential cofactor for oncogenic transformation by c-Myc. *Mol. Cell* 5, 321-330.

17 Watkins, N.J., Dickmanns, A., and Luhrmann, R. (2002). Conserved stem II of the box C/D motif is essential for nucleolar localization and is required, along with the 15.5K protein, for the hierarchical assembly of the box C/D snoRNP. *Mol. Cell Biol.* 22, 8342-8352.

18 Boulon, S., Marmier-Gourrier, N., Pradet-Balade, B., Wurth, L., Verheggen, C., Jady, B.E., Rothe, B., Pescia, C., Robert, M.C., Kiss, T., Bardoni, B., Krol, A., Branlant, C., Allmang, C., Bertrand, E., and Charpentier, B. (2008). The Hsp90 chaperone controls the biogenesis of L7Ae RNPs through conserved machinery. *J. Cell Biol.* 180, 579-595.

19 Venteicher, A.S., Meng, Z., Mason, P.J., Veenstra, T.D., and Artandi, S.E. (2008). Identification of ATPases pontin and reptin as telomerase components essential for holoenzyme assembly. *Cell* 132, 945-957.

20 Haurie, V., Menard, L., Nicou, A., Touriol, C., Metzler, P., Fernandez, J., Taras, D., Lestienne, P., Balabaud, C., Bioulac-Sage, P., Prats, H., Zucman-Rossi, J., and Rosenbaum, J. (2009). Adenosine triphosphatase pontin is overexpressed in hepatocellular carcinoma and coregulated with reptin through a new posttranslational mechanism. *Hepatology* 50, 1871- 1883.

21 Rousseau, B., Menard, L., Haurie, V., Taras, D., Blanc, J., Moreau-Gaudry, F., Metzler, P., Hugues, M., Boyault, S., Lemiere, S., Canron, X., Costet, P., Cole, M., Balabaud, C., Bioulac- Sage, P., Zucman-Rossi, J., and Rosenbaum, J. (2007). Overexpression and role of the ATPase and putative DNA helicase RuvB-like 2 in human hepatocellular carcinoma. *Hepatology* 46, 1108-1118.

22 Menard, L., Taras, D., Grigoletto, A., Haurie, V., Nicou, A., Dugot-Senant, N., Costet, P., Rousseau, B., and Rosenbaum, J. (2010). In vivo silencing of Reptin blocks the progression of human hepatocellular carcinoma in xenografts and is associated with replicative senescence. *J. Hepatol.* 52,

681-689.

23 King, T.H., Decatur, W.A., Bertrand, E., Maxwell, E.S., and Fournier, M.J. (2001). A well-connected and conserved nucleoplasmic helicase is required for production of box C/D and H/ACA snoRNAs and localization of snoRNP proteins. *Mol. Cell Biol.* 21, 7731-7746.

24 Lim, C.R., Kimata, Y., Ohdate, H., Kokubo, T., Kikuchi, N., Horigome, T., and Kohno, K. (2000). The *Saccharomyces cerevisiae* RuvB-like protein, Tih2p, is required for cell cycle progression and RNA polymerase II-directed transcription. *J. Biol. Chem.* 275, 22409-22417.

25 Feng, Y., Lee, N., and Fearon, E.R. (2003). TIP49 regulates beta-catenin-mediated neoplastic transformation and T-cell factor target gene induction via effects on chromatin remodeling. *Cancer Res.* 63, 8726-8734.

26 Dugan, K.A., Wood, M.A., and Cole, M.D. (2002). TIP49, but not TRRAP, modulates c-Myc and E2F1 dependent apoptosis. *Oncogene* 21, 5835-5843.

27 Qiu, X.B., Lin, Y.L., Thome, K.C., Pian, P., Schlegel, B.P., Weremowicz, S., Parvin, J.D., and Dutta, A. (1998). An eukaryotic RuvB-like protein (RUVBL1) essential for growth. *J. Biol. Chem.* 273, 27786-27793.

28 Ikura, T., Ogryzko, V.V., Grigoriev, M., Groisman, R., Wang, J., Horikoshi, M., Scully, R., Qin, J., and Nakatani, Y. (2000). Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis. *Cell* 102, 463-473.

29 Choi, J., Heo, K., and An, W. (2009). Cooperative action of TIP48 and TIP49 in H2A.Z exchange catalyzed by acetylation of nucleosomal H2A. *Nucleic Acids Res.* 37, 5993-6007.

30 Trott, O., and Olson, A.J. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455-461.

31 Velec, H.F.G., Gohlke, H., and Klebe, G. (2005). DrugscoreCSD Knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near native ligand poses and better affinity prediction. *J. Med. Chem.* 48, 6296-6303.

32 Gohlke, H., Hendlich, M., and Klebe, G. (2000). Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* 295, 337-356.

33 Wang, R., Laib, L., and Wang, S. (2002). Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput-Aided Mol. Des.* 16, 11-26.

34 O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T., Hutchison, G.R. (2011). Open Babel: An open chemical toolbox. *J. Cheminform.* 3, 33.

35 Sanner, M.F. (1999). Python: a programming language for software integration and

development. *J. Mol. Graphics Mod.* 17, 57-61.

36 Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K., and Olson, A.J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* 19, 1639-1662.

37 Chang, M.W., Ayeni, C., Breuer, S., and Torbett, B.E. (2010). Virtual screening for HIV protease inhibitors: A comparison of AutoDock 4 and Vina. *Plos One* 5, e11955.

38 Guha, R., Howard, M.T., Hutchison, G.R., Murray-Rust, P., Rzepa, H., Steinbeck, C., Wegner, J.K., and Willighagen, E. (2006). The Blue Obelisk – Interoperability in chemical informatics. *J. Chem. Inf. Model.* 46, 991-998.

39 Zhang, J.H., Chung, T.D. and Oldenburg, K.R. (1999). A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J. Biomol. Screen.* 4, 67-73.

40 Charifson, P.S., Corkery, J.J., Murcko, M.A., and Walters, W.P. (1999). Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* 42, 5100-5109.

41 Wang, R., and Wang, S. (2001). How does consensus scoring work for virtual library screening? An idealized computer experiment. *J. Chem. Inf. Comput. Sci.* 41, 1422-1426.

42 Kanemaki, M., Kurokawa, Y., Matsu-ura, T., Makino, Y., Masani, A., Okazaki, K., Morishita, T., and Tamura, T.A. (1999). TIP49b, a new RuvB-like DNA helicase, is included in a complex together with another RuvB-like DNA helicase, TIP49a. *J. Biol. Chem.* 274, 22437-22444.

43 Makino, Y., Kanemaki, M., Kurokawa, Y., Koji, T., and Tamura, T-A. (1999). A rat Ruv-B like Protein, TIP49a, is a germ cell-enriched novel DNA helicase. *J. Biol. Chem.* 274, 15329- 15335.

44 Rottbauer, W., Saurin, A.J., Lickert, H., Shen, X., Burns, C.G., Wo, Z.G., Kemler, R., Kingston, R., Wu, C., and Fishman, M. (2002). Reptin and pontin antagonistically regulate heart growth in zebrafish embryos. *Cell* 111, 661-672.

45 Cheng, Y., and Prusoff, W.H. (1973). Relationship between inhibition constant ( $K_i$ ) and the concentration of inhibitor which causes 50 per cent inhibition ( $I_{50}$ ) of an enzymatic reaction. *Biochem. Pharmacol.* 22, 3099-3108.

46 Soltoff, SP. (2007). Rottlerin: an inappropriate and ineffective inhibitor of PKCdelta. *Trends Pharmacol. Sci.* 28, 453-458.

47 Hishida, T., Han, Y.W., Fujimoto, S., Iwasaki, H., and Shinagawa, H. (2004). Direct evidence that a conserved arginine in RuvB AAA+ ATPase acts as an allosteric effector for the ATPase activity of the adjacent subunit in hexamer. *Proc. Natl. Acad. Sci. USA* 101, 9573- 9577.

48 Zhang, X., and Wigley, D.B. (2008). The “Glutamate switch” provides a link, between ATPase activity and ligand binding in AAA+ proteins. *Nat. Struct. Mol. Biol.* 11, 1223-1227.

49 Moffitt, J.R., Chemila, Y.R., Aathavan, K., Grimes, S., Jardine, P.J., Anderson, D.L., and Bustamante, C. (2009). Intersubunit coordination in a homomeric ring ATPase. *Nature* 457, 446-450.

50 Lessard, L., Stuiblé, M. and Tremblay M. (2010) The two faces of PTP1B in cancer. *Biochim. Biophys. Acta* 1804, 613-619.



## - RÉSUMÉ -

Les processus qui mènent à la découverte de nouveaux médicaments sont longs et fastidieux, et les taux de succès sont relativement faibles. L'identification de candidats par le biais de tests expérimentaux s'avère coûteuse, et nécessite de connaître en profondeur les mécanismes d'action de la protéine visée afin de mettre en place des essais efficaces. Le criblage virtuel peut considérablement accélérer ces processus en permettant une évaluation rapide de chimiothèques de plusieurs milliers de molécules afin de déterminer lesquelles sont les plus susceptibles de se lier à une cible. Ces dernières années ont ainsi été témoins de quelques success stories dans ce domaine.

Le premier objectif de ce travail était de comparer différents outils et stratégies couramment utilisés dans le criblage virtuel "structure-based", puis de les appliquer à des cibles protéiques à visée thérapeutique, en particulier dans le cadre du cancer.

La protéine kinase GSK3 et un test set de ligands connus ont servi de modèle pour différentes études méthodologiques ayant pour but d'évaluer les programmes de docking et de scoring à notre disposition. En particulier, l'utilisation de plusieurs structures relaxées du récepteur ou l'insertion de torsions sur certains résidus du site actif pendant le docking ont permis d'évaluer l'influence de la flexibilité de la protéine. L'utilité et la pertinence d'outils permettant de générer automatiquement les structures 3D des ligands et de méthodes de consensus scoring ont également été étudiées.

Un criblage virtuel de la Pontine, une ATPase impliquée dans la croissance tumorale pour laquelle aucun inhibiteur n'était connu, a permis la sélection de candidats issus de banques de données commerciales. Ces molécules ont été testées dans un essai enzymatique par le biais d'une collaboration, et quatre d'entre elles se sont révélées capable d'inhiber l'activité ATPase de la Pontine. Le criblage de bases de ligands synthétisés et imaginés dans l'équipe a également fourni un inhibiteur original. Au contraire, l'étude de la sPLA2-X humaine, une phospholipase dont l'activité catalytique est dépendante d'un atome de  $\text{Ca}^{2+}$  localisé au sein du site actif, a montré les limites de nos outils de docking qui n'ont pas été capables de gérer cet ion métallique et mis en évidence la nécessité de mettre en place d'autres outils.



## - ABSTRACT -

The process of drug discovery is long and tedious. Besides, it is relatively inefficient in terms of hit rate. The identification of candidates through experimental testing is expensive and requires extensive data on the mechanisms of the target protein in order to develop efficient assays. Virtual screening can considerably accelerate the process by quickly evaluating large databases of compounds and determining the most likely to bind to a target. Some success stories have emerged in the field over the last few years.

The objectives of this work were first, to compare common tools and strategies for structure-based virtual screening, and second, to apply those tools to actual target proteins implied notably in carcinogenesis.

In order to evaluate the docking and scoring programs available, the protein kinase GSK3 and a test set of known ligands were used as a model to perform methodological studies. In particular the influence of the flexibility of the protein was explored via relaxed structures of the receptor or the insertion of torsions on the side chains of residues located in the binding site. Studies concerning the automatic generation of 3D structures for the ligands and the use of consensus scoring also provided insights on the usability of these tools while performing a virtual screening.

Virtual screening of the human protein Pontin, an ATPase implied in tumor cell growth for which no inhibitors were known, allowed the prioritization of compounds from commercial databases. These compounds were tested in an enzymatic assay via a collaboration, and led to the identification of four molecules capable of inhibiting the ATPase activity of Pontin. Additional screens of in-house oriented databases also provided at least one innovative inhibitor for this protein. On the contrary, a study of the human PLA2-X, a phospholipase that requires a  $\text{Ca}^{2+}$  atom to bind to its active site in order to catalyze the hydrolysis of its substrate, revealed the limits of our docking tools that could not handle the metal ion and the need for new tools.