



HAL
open science

Conception de Ligands Protéiques par Bioinformatique et Modélisation Moléculaire

Cedrik Magis

► **To cite this version:**

Cedrik Magis. Conception de Ligands Protéiques par Bioinformatique et Modélisation Moléculaire. Sciences du Vivant [q-bio]. Museum national d'histoire naturelle - MNHN PARIS, 2007. Français. NNT: . tel-00553476

HAL Id: tel-00553476

<https://theses.hal.science/tel-00553476>

Submitted on 7 Jan 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ÉCOLE DOCTORALE DU MUSÉUM NATIONAL D'HISTOIRE NATURELLE

Thèse présentée pour obtenir le grade de :

Docteur du Muséum National d'Histoire Naturelle
Spécialité : Ingénierie des protéines

Sujet du travail de thèse :

Conception de Ligands Protéiques par Bioinformatique et Modélisation Moléculaire

par C. MAGIS

Présentée le 23 février 2007, devant la Commission d'Examen :

MM. Hugues Bédouelle
Philippe Cuniasse
Max Goyffon
Thomas Haertlé
Solange Lavielle
André Ménez

Rapporteur
Examineur
Président
Examineur
Rapporteur
Directeur de Thèse

Renseignements personnels :

MAGIS Cedrik, doctorant du Muséum National d'Histoire Naturelle

8, rue Théodore de Banville, 75017, PARIS

06 76 89 22 72 / 01 47 63 08 97

cedrik.1978@gmail.com

TABLE DES MATIÈRES

ABRÉVIATIONS/GLOSSAIRE	p.1
INTRODUCTION	p.3
CONCEPTION DE LIGANDS PROTÉIQUES	p.7
Chapitre I : Conception de ligands protéiques	p.9
I.1 De l'intérêt de la conception de ligands	p.13
I.1.1 De la découverte à la conception de composés actifs	p.13
I.1.2 La conception « rationnelle »	p.16
I.1.3 Identification/Conception de « molécule guide »	p.19
I.1.4 Conclusion	p.26
I.2 De l'utilisation des protéines comme ligands	p.27
I.2.1 Implications pour la conception de ligands	p.27
I.2.2 Les protéines comme agents thérapeutiques et/ou diagnostiques	p.33
I.3 Conception de ligands protéiques	p.37
I.3.1 Minimisation de protéines	p.37
I.3.2 Synthèse de plateformes <i>de novo</i>	p.41
I.3.3 Utilisation de plateformes naturelles	p.43
I.3.4 Conclusion	p.48
Chapitre II Analyse des interactions protéine-protéine	p.49
II.1 Description générale des interfaces des interfaces	p.53
II.1.1 Un ou plusieurs types de complexes ?	p.53
II.1.2 Caractéristiques des interfaces protéine-protéine	p.54
II.1.3 Composition des interfaces	p.60
II.1.4 Conclusion	p.65
II.2 Considérations énergétiques	p.67
II.2.1 Aspects thermodynamiques de l'interaction	p.67
II.2.2 Interaction ponctuelles entre atomes non-liés	p.69
II.2.3 Définition de « point chaud de l'interaction » ou « hotspot »	p.74
II.2.4 Conclusion	p.76
II.3 Le potentiel électrostatique dans les interactions protéiques	p.77
II.3.1 Aspects électrostatiques de l'interaction	p.77
II.3.2 Notion de potentiel électrostatique	p.79
II.3.3 Conclusion	p.81

Chapitre III : Conception de ligands protéiques	p.83
III.1 Principe de la méthode de conception de ligands protéiques	p.87
III.1.1 Du motif fonctionnel à la conception de ligands protéiques	p.87
III.1.2 Recherche de structures possédant un motif topologique équivalent	p.89
III.1.3 Sélection des plateformes	p.92
III.1.4 Conclusion	p.99
III.2 Le système de référence toxine-canal potassique	p.101
III.2.1 Les canaux de la famille Kv	p.101
III.2.2 Les toxines spécifiques des canaux Kv1	p.105
III.2.3 L'interaction entre la BgK et les canaux de la famille Kv	p.108
III.2.4 Conclusion	p.110
Chapitre IV : Recherche de similarité d'interaction parmi les ligands d'une même cible. Analyse de complexes protéine-protéine	p.111
IV.1 Systèmes d'étude	p.115
IV.1.1 Critères de sélection des systèmes	p.115
IV.1.2 Homogénéité des cibles	p.117
IV.1.3 Le site de liaison	p.120
IV.1.4 Diversité des ligands	p.121
IV.1.5 Conclusion	p.126
IV.2 Analyse des interfaces ligand-cible	p.127
IV.2.1 Propriétés globales des interfaces	p.127
IV.2.2 Aspects électrostatiques	p.132
IV.2.3 Conclusion	p.141
IV.3 Analyse des interactions intermoléculaires ligands / cible	p.143
IV.3.1 Interactions « pi-pi » et « cation-pi »	p.143
IV.3.2 Interactions coulombiennes	p.144
IV.3.3 Liaisons hydrogène	p.146
IV.3.4 Conclusion	p.149
Chapitre V : À la recherche de plateformes protéiques	p.151
V.1 Influence des paramètres de recherche	p.155
V.1.1 Influence de la taille des structures	p.156
V.1.2 Influence de la taille du motif	p.157
V.1.3 Influence du RMSD_MAX	p.159
V.1.4 Influence du Δd_MAX	p.161
V.1.5 Influence du filtre stérique	p.163
V.1.6 Conclusion : choix des paramètres de recherche	p.164
V.2 Identification des plateformes protéiques	p.165
V.2.1 Recherche de motifs topologiques	p.165
V.2.2 Sélection <i>in-silico</i> des plateformes	p.167
V.2.3 Critères additionnels	p.169

V.2.4 Conclusion	p.170
V.3 Analyse des plateformes	p.171
V.3.1 Les plateformes identifiées	p.171
V.3.2 Diversité structurale des plateformes	p.173
V.3.3 Analyse du potentiel électrostatique	p.177
V.3.4 Conclusion	p.179
Chapitre VI : Validation expérimentale	p.181
VI.1 Production par biologie moléculaire de la plateforme 1J75	p.185
VI.1.1 Plateforme 1J75	p.185
VI.1.2 Rendement de production	p.187
VI.1.3 Conclusion	p.189
VI.2 Production des plateformes par voie chimique	p.191
VI.2.1 Synthèse chimique sur support solide	p.191
VI.2.2 Obtention des plateformes	p.195
VI.2.3 Conclusion	p.198
VI.3 Caractérisation structurale des plateformes	p.199
VI.3.1 Caractérisation par dichroïsme circulaire	p.199
VI.3.2 Détermination de la structure tridimensionnelle de 1J75mt K30A	p.204
VI.3.3 Conclusion	p.210
VI.4 Évaluation expérimentale des ligands protéiques	p.211
VI.4.1 Caractérisation de l'affinité	p.211
VI.4.2 Modèles d'interaction ligands/Kv1.2	p.215
CONCLUSIONS ET PERSPECTIVES	p.221
ANNEXES	p.233
A.1 Bioinformatique et Modélisation Moléculaire	p.235
A.2 La « Protein Data Bank » ou PDB	p.236
A.3 L'équation de Poisson-Boltzmann	p.238
A.4 Les programmes de sélection de plateformes	p.241
B.1 Affinement de la structure de la BgK	p.243
B.2 Systèmes de complexes cristallographiques	p.244
B.3 Analyse d'interfaces protéine-protéine	p.249
C.1 Séquence des protéines synthétisées	p.251
C.2 Synthèse par biologie moléculaire	p.252
C.3 Synthèse par voie chimique	p.255
C.4 Obtention des protéines synthétisées par voie chimique	p.257
D.1 Caractérisation par dichroïsme circulaire	p.261
D.2 Détermination de structure par radiocristallographie	p.263
D.3 Tests de liaison par compétition	p.265
D.4 Réalisation de modèles moléculaires	p.267
RÉFÉRENCES BIBLIOGRAPHIQUES	p.269

ABRÉVIATIONS / GLOSSAIRE

Aspects bioinformatiques / Modélisation moléculaire

STAMP	Search for Three Dimensional Atomic Motif in Proteins
SPASM	Spatial Arrangement of Side chains and Main chains
APBS	Adaptative Poisson Boltzmann Solver
UHBD	University of Houston Brownian Dynamics
PIPSA	Protein Interaction and Property Similarity Analysis
CHARMM	Chemistry at HARvard Macromolecular Mechanics
PyMOL	Python-enhanced MOLEcular graphic system
BLAST	Basic Local Alignment Search Tool
PDB	Protein Data Bank
RMSD	Root Mean Square Deviation
3D/2D	tridimensionnelle/bidimensionnelle
ASA	Accessible solvent Surface Area

Méthodes expérimentales

SAR	Structure Activity Relationship
QSAR	Quantitative Structure Activity Relationship
CD	Circular Dichroism (Dichroïsme circulaire)
UV	Ultra-Violet
RMN	Résonance Magnétique Nucléaire
SDS-PAGE	Sodium Dodecyl Sulfate-PolyAcrylamide Gel Electrophoresis
RP-HPLC	Reverse Phase-High Performance Liquid Chromatography
ES-MS	Electro Spray-Mass Spectrometry
PCR	Polymerase Chain Reaction

Composés chimiques

IPTG	isopropyl- β -D-thiogalactopyranoside
PMSF	phenylmethylsulfonyl fluoride.
TCEP	tris(2-carboxyethyl) phosphine
Fmoc	fluorenylmethyloxycarbonyl
DCC	N,N'-dicyclohexylcarbodiimide
HOAt	1-hydroxy-7-azabenzotriazole
DIEA	N,N-diisopropylethylamine
NMP	N-methyl pyrrolidone
TFA	trifluoroacetic acid
TIS	tri-isopropyl-silane
TBME	tert-butyl methyl ether
TFE	trifluoroethanol
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
MES	2-(N-morpholino) ethanesulphonic acid
DTT	dithiothreitol
PEG	polyethylene glycol
ADN	acide désoxyribonucléique

Protéines et mini-protéines

BgK	bunodosoma granulifera K ⁺ channel -voltage gated
ChTx	charybdotoxine
AgTx2	agitoxine 2
DTX	dendrotoxine
BPTI	bovine pancreatic trypsin inhibitor
BBI	Bowman Birk inhibitor
SFI	sun flower inhibitor
MGI	mung bean inhibitor
SSI	sunflower seeds inhibitor
APPI	Alzheimer's amyloid beta-protein precursor inhibitor
PPTI	porcine pancreatic trypsin inhibitor
CPTI-II	cucurbita pepo trypsin inhibitor II
STI	soybean trypsin inhibitor
EETI-II	ecballium elaterium trypsin inhibitor II
LDTI	leech derived tryptase inhibitor
MCTI-A	momordica charantia trypsin inhibitor A
TFPI	tissue factor pathway inhibitor (second domain)
OMTKY-2	ovomucoid turkey egg white inhibitor 2
PSTI-3	pancreatic secretory trypsin inhibitor 3

Acides aminés / Bases azotées

Tableau des acides aminés et acides nucléiques (désoxyribose) selon leur dénomination courante, le code à 3 lettres ou à 1 lettre, utilisés dans le code international.

Acides aminés naturels											
Alanine	ALA	A	Cystéine	CYS	C	Leucine	LEU	L	Sérine	SER	S
Arginine	ARG	R	Glutamine	GLN	Q	Lysine	LYS	K	Thréonine	THR	T
Asparagine	ASN	N	Glycine	GLY	G	Méthionine	MET	M	Tryptophane	TRP	W
Ac. aspartique	ASP	D	Histidine	HIS	H	Proline	PRO	P	Tyrosine	TYR	Y
Ac. glutamique	GLN	E	Isoleucine	ILE	I	Phénylalanine	PHE	F	Valine	VAL	V
Acides Nucléiques (formant des désoxyriboses)											
Adénine	ADE	A	Guanine	GUA	G	Thymine	THY	T	Cytosine	CYT	C

INTRODUCTION

Le développement de molécules, à usage thérapeutique ou diagnostique, reste un enjeu majeur en biologie. Bien que ces recherches bénéficient des connaissances croissantes des phénomènes physiologiques et pathologiques, à l'échelle moléculaire, la demande ne cesse de s'accroître car les enjeux de santé publique sont considérables (cancer(s), SIDA, maladies neuro-dégénératives, résistance aux antibiotiques...). Cependant, malgré la diversité des approches mises en œuvre, le nombre de nouvelles molécules actives, identifiées, reste relativement faible. Dès lors, le développement de nouvelles méthodes, permettant d'augmenter l'efficacité et le rendement du processus de conception, apparaît nécessaire.

Par ailleurs, ces dernières décennies, la découverte de nouvelles molécules actives ne se limite plus aux molécules organiques, de faible poids moléculaire : de nouvelles voies sont envisagées, pour le développement de telles molécules. Parmi celles-ci, le développement de composés peptido mimétiques ou, de protéines issues de l'ingénierie, a permis des avancées importantes. Les protéines apparaissent attractives de part leur capacité à interagir à haute affinité et surtout avec une grande spécificité avec d'autres partenaires. Bien que les molécules organiques conservent un attrait considérable, pour les groupes pharmaceutiques, lié en particulier à leur faible coût de production, l'exploration de l'utilisation des protéines dans des contextes thérapeutiques et diagnostiques a conduit à identifier des protéines, issues de l'ingénierie, dont l'utilité a pu être démontrée, lors d'études cliniques. Ces résultats, ainsi que les progrès réalisés en matière de production, de purification et de caractérisation des protéines (en particulier sur le plan structural) ont contribué à l'intérêt porté à cette classe de molécules. D'autre part, l'accroissement continu du nombre de structures 3D de protéines, connues et reportées dans la « PDB » (« Protein Data Bank »), constitue un atout considérable tant sur le plan de la compréhension des phénomènes biologiques que de l'ingénierie des protéines.

Le travail réalisé dans le cadre de cette thèse s'inscrit dans le contexte de l'utilisation des protéines comme ligand et a pour objectifs (1) de contribuer à l'amélioration de la compréhension des interactions entre protéines et (2) de mettre au point une méthode de conception ligands protéiques. Cette méthode, repose sur le transfert d'un motif de résidus fonctionnels (énergétiquement important pour une interaction donnée), identifiés sur un ligand, dit de référence, de la cible considérée, par « alanine scanning ». Le transfert a lieu sur une plateforme protéique, capable de reproduire la topologie fonctionnel de ce même motif. Des travaux antérieurs, utilisant le transfert d'un motif fonctionnel sur plateforme protéique, pour la conception de ligands, ont permis de démontrer l'intérêt d'une telle approche. Cependant,

l'augmentation croissante d'informations structurales, concernant les protéines, offre de nouvelles perspectives, dans cette optique. Ainsi, l'identification des plateformes susceptibles d'accueillir le motif fonctionnel, peut-elle être réalisée grâce à un examen systématique de la topologie des structures protéiques contenues dans la PDB. Les petites protéines (de taille comprise de 20 à 100 résidus), dénommées mini-protéines, possèdent plusieurs caractéristiques qui les rendent particulièrement attractive pour la conception de ligands. L'examen systématique des structures, déposées dans la PDB, permettrait d'identifier toutes les mini-protéines, possédant des structures 3D compatibles avec le transfert du motif considéré, en reproduisant sa topologie fonctionnelle. Il serait alors possible d'augmenter la diversité des possibilités, accroissant, de ce fait, le taux de succès d'une telle approche.

Une méthode de conception de ligands, par transfert d'un motif fonctionnel sur une mini-protéine, doit prendre en compte l'ensemble des propriétés de la plateforme. En effet, tous les résidus de celle-ci sont susceptibles de contribuer, de manière favorable ou défavorable, à l'interaction avec la cible considérée. Il est donc essentiel que la sélection des plateformes implique des critères concernant divers aspects des interactions protéine-protéine. A notre connaissance, dans les études de conception de ligands protéiques par transfert de site fonctionnel reportées antérieurement, ces aspects ne sont jamais explicitement pris en compte. Dans le cadre de ce travail, nous avons réalisé une analyse de ces interactions, dans des systèmes relevant, par rapport aux questions posées par le développement d'une mini-protéine mimétique. Nous avons pu identifier, dans la PDB, diverses protéines, pour lesquelles plusieurs complexes, avec des ligands protéiques, interagissant à un même site de liaison, étaient disponibles. Ce travail nous a permis de mettre en évidence les critères importants pour la reproduction de l'interaction et de caractériser leurs fluctuations. La complémentarité de formes de partenaires protéiques et la similarité des potentiels électrostatiques des différents ligands, peuvent ainsi être utilisés, comme critères de sélections des protéines hôtes. Par ailleurs, ce travail a également montré que les interactions ponctuelles n'étaient pas systématiquement reproduites, par l'ensemble des ligands, en interaction avec la même cible protéique. Les informations issues de cette analyse ont été utilisées pour la caractérisation des modes d'interactions des ligands protéiques conçus. Cette approche générale de conception de ligand protéique, basée sur la structure, a été appliquée à la conception de mini-protéines capables de lier le canal Kv1.2, cible de la toxine BgK. Les résultats obtenus ouvrent la voie à la possibilité d'optimiser les ligands protéiques, du canal Kv1.2, conçus dans le cadre de ce travail et à l'application de cette approche à d'autres cibles d'intérêt, en santé humaine.

Conception de ligands protéiques

« La Science ne cherche pas à énoncer des vérités éternelles ou de dogmes immuables ; loin de prétendre que chaque étape est définitive et qu'elle a dit son dernier mot, elle cherche à cerner la vérité par approximations successives » (B. Russell).

Chapitre I : Conception de ligands protéiques

« Scientific discovery consists in the interpretation for our own convenience of a system of existence, which has been made with no eyes for our convenience at all » (Norbert Wiener).

Les macromolécules biologiques assurent de multiples rôles au sein d'un organisme (structure, défense, métabolisme, énergie, information...), mais chaque fonction ne peut s'exercer qu'après une étape de reconnaissance et d'interaction avec un ou plusieurs partenaires. La quasi totalité des interactions correspond à la formation de complexes non-covalents, de hautes affinité et spécificité, caractéristiques fondamentales des macromolécules biologiques.

Parmi ces macromolécules biologiques, les protéines sont les principaux médiateurs des fonctions d'un organisme. De nombreux travaux ont été réalisés pour comprendre comment celles-ci interagissent avec leurs partenaires, dont la nature peut être extrêmement diverse (acides nucléiques, ions, métaux, sucres, lipides, molécules organiques, ou autres protéines). Les interactions entre protéines en représentent la majeure partie.

La compréhension des interactions entre protéines est fondamentale : d'une part, afin de mieux appréhender les bases moléculaires des processus biologiques et d'autre part, pour développer des méthodes de conception de ligands en vue d'obtenir des agents thérapeutiques et/ou diagnostiques.

Le chapitre suivant portera sur l'état actuel des approches de conception de médicaments, et l'importance de la conception de ligands par rapport à cette dernière. Puis le choix de ligands de nature protéique sera expliqué et justifié par l'énumération d'exemples de protéines actuellement utilisées en tant que médicaments. Enfin les diverses méthodes de conception de ligands déjà décrites, ainsi que leurs limitations, seront présentées.

I.1 De l'intérêt de la conception de ligands

Les stratégies mises en place afin d'obtenir de nouveaux médicaments (*) ont radicalement changé au cours du siècle dernier, de la découverte hasardeuse de nouveaux composés à la planification de la conception, comme schématisé figure I-1 [1]. Les différentes évolutions sont indiquées par ordre chronologique dans le paragraphe suivant.

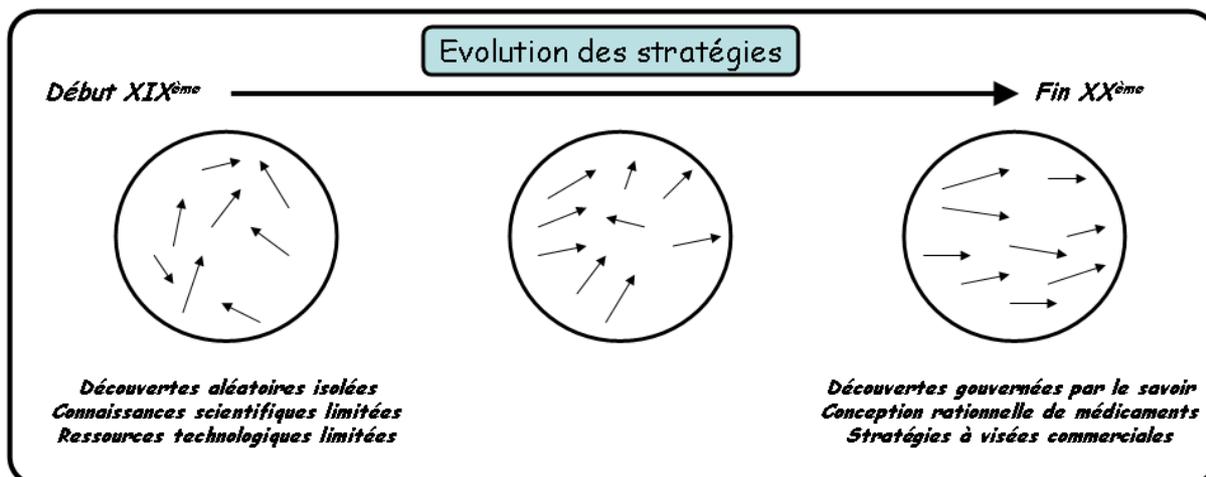


Figure I-1 : Représentation schématique des stratégies d'obtention de médicaments du début du XIX^{ème} siècle jusqu'à la fin du XX^{ème} siècle [1]. La découverte aléatoire de médicament, basée sur l'empirisme, est remplacée par une approche planifiée, dont le coût et la durée sont estimés au préalable, correspondant à une approche dénommée pour ces raisons « rationnelle ».

I.1.1 De la découverte à la conception de composés actifs

Jusqu'au XIX^{ème} siècle, la découverte de nouveaux médicaments était liée à l'observation empirique des effets produits par diverses substances naturelles vis-à-vis des maladies. Il est à noter que, déjà au XVI^{ème} siècle, Paracelse (**), prônait déjà l'idée de la nécessité de disposer de médicaments spécifiques pour chaque maladie. Au cours du XIX^{ème} siècle, grâce au progrès de la chimie et de la physiologie, les premiers principes actifs sont isolés (morphine, émétine, quinine, acide acétylsalicylique, digitaline). La première moitié du XX^{ème} siècle est marquée par le développement de nombreuses classes de médicaments. Apparaissent successivement les anti-syphilitiques (1906), les anti-paludéens (1927), les antibiotiques (1928), les anti-infectieux (1935), les anti-histaminiques (1942), les anti-coagulants (1947), les anti-tuberculeux (1951), les anti-dépresseurs (1957)...rejoignant ainsi les idées de Paracelse, émises 400 ans auparavant.

*** Médicament**

Est considéré comme médicament, toute substance ou composition présentée comme possédant des propriétés curatives ou préventives à l'égard des maladies humaines ou animales, ainsi que tout produit pouvant être administré à l'homme ou à l'animal en vue d'établir un diagnostic médical ou de restaurer, corriger ou modifier leurs fonctions organiques. (Article L.511 du Code de la Santé Publique).

**** Paracelse, de son véritable nom Philippus Aureolus Theophrastus Bombast von Hohenhiem (1493-1541)**

Chimiste et médecin, considéré comme l'un des pères de la médecine expérimentale, à qui l'on doit notamment : « *Tout est poison, rien n'est poison. Seule la dose fait le poison* »

La seconde moitié du XX^{ème} siècle, montre un changement radical : les premières méthodes d'évaluation préclinique et de criblage sont introduites, faisant suite au drame de la thalidomide. La création de banques de composés naturels et/ou synthétiques, leur criblage puis l'identification des molécules actives constitue l'approche dite combinatoire. La 26 janvier 1965, la mise sur le marché de médicaments est soumise à une réglementation plus contraignante par la directive 65/65/CEE. Elle stipule que chaque composé actif identifié doit obtenir au préalable une Autorisation de Mise sur le Marché (AMM), délivrée par l'Agence Française de Sécurité Sanitaire des Produits de Santé (AFSSAPS) ou l'Agence Européenne pour l'Evaluation des Médicaments (AMEA). L'AMM n'est décernée qu'à l'issue de différentes phases, détaillées sur la figure I-2, A.

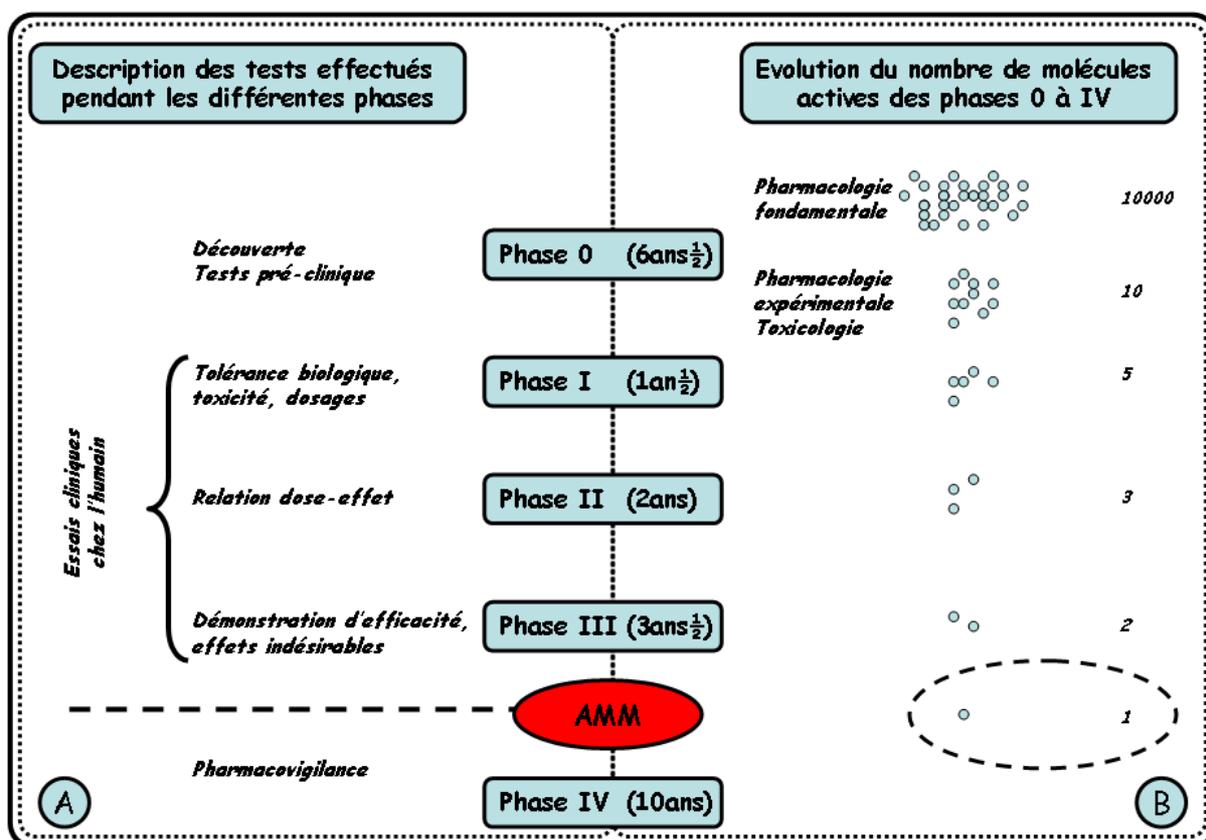


Figure I-2 : Phases nécessaires pour le développement d'un médicament jusqu'à l'obtention de son autorisation de mise sur le marché (AMM). La partie A définit les étapes nécessaires pour l'obtention de l'AMM. La partie B montre l'évolution du nombre de médicaments obtenus après les différentes phases nécessaires pour l'obtention de l'AMM, et le rendement de chacune [<http://www.chups.jussieu.fr/polys/pharmaco/poly>].

A titre d'indication, il est estimé qu'à la fin de celles-ci, 1 molécule active sur 10000 obtient finalement l'AMM, comme indiqué sur la figure I-2, B. L'ensemble de la procédure couvre une période de 10 à 15 ans, de la découverte d'une molécule active à sa mise sur le marché. En tenant compte de ces limitations de temps et de rendement, le coût de développement d'un médicament est devenu prohibitif. La dernière estimation (bien que surévaluée d'après certains [2]) est évaluée à 802 millions de dollars [3], expliquant l'engouement des industries pharmaceutiques

pour la mise au point de méthodes de conception de médicaments plus rapides et plus efficaces. Depuis une vingtaine d'années, l'industrie pharmaceutique s'est tournée vers la « conception rationnelle » (« **Rational Drug Design** ») pour l'obtention de nouveaux médicaments. Le but de l'approche rationnelle est de réduire le nombre d'échecs et limiter ainsi les coûts de développement. A cette fin, la conception est orientée vers une cible moléculaire, connue pour son implication dans un processus pathologique, et non plus par l'observation d'effets physiologiques. La conception rationnelle est une approche interdisciplinaire, représentée sur la figure I-3, qui s'appuie, entre autre, sur les concepts de l'approche combinatoire (chimie combinatoire [4], « phage display » [5], évolution dirigée [6]), ainsi que sur diverses techniques de bioinformatique et de modélisation moléculaire [7] (annexe A.1), de biologie moléculaire, de biophysique et de chimie.

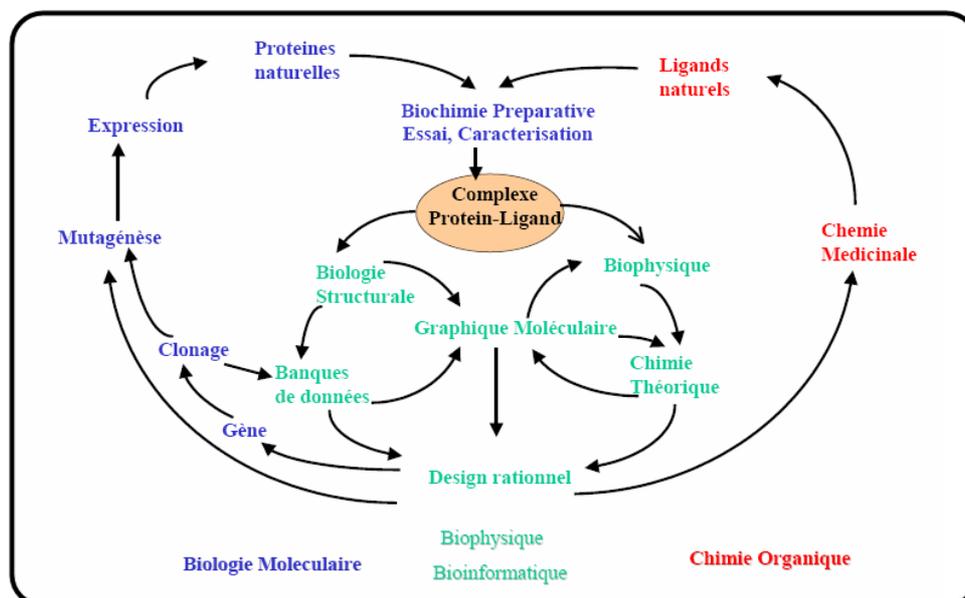


Figure I-3 : Représentation schématisée des relations entre les différentes disciplines impliquées dans la conception rationnelle. [didier.rognan@pharma.u-strasbg.fr]

I.1.2 La conception « rationnelle »

Le principe de la conception rationnelle repose sur la connaissance préalable d'un système biologique ciblé (cause moléculaire d'une pathologie). La conception rationnelle de médicaments comprend plusieurs étapes successives, indiquées sur la figure I-4 :

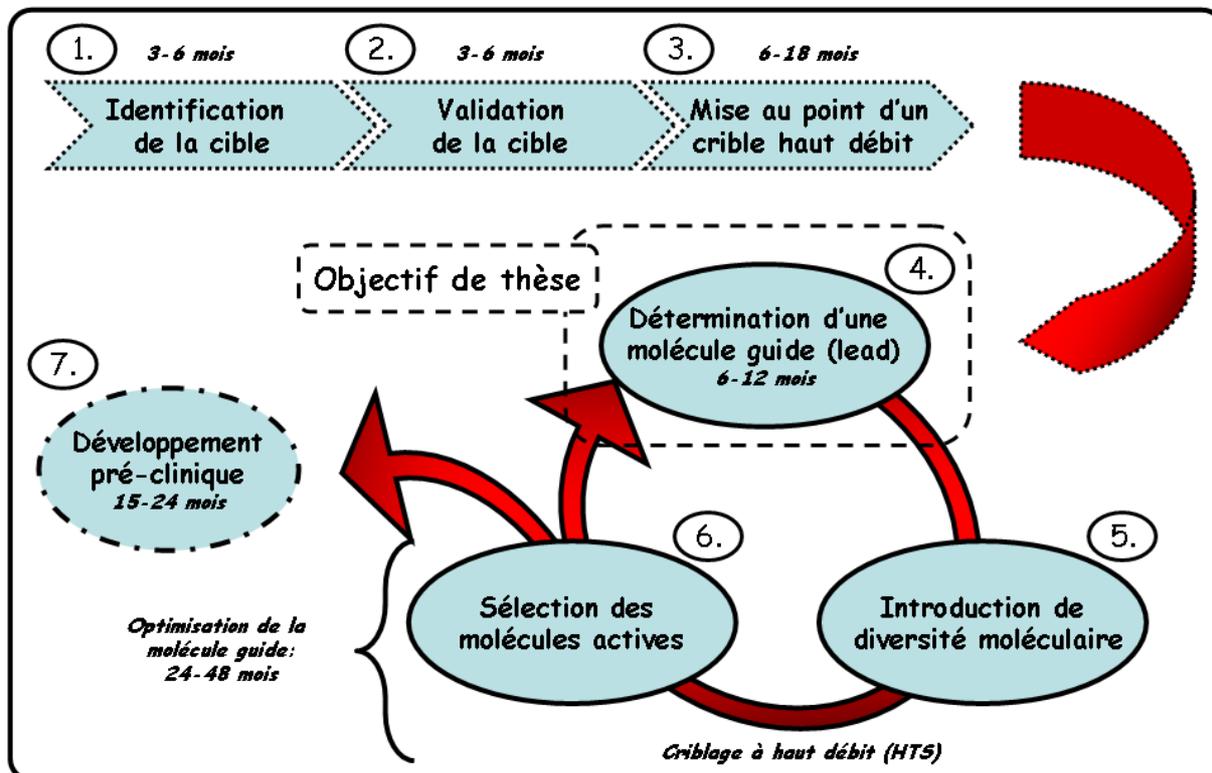


Figure I-4 : Schéma général de la conception rationnelle de médicaments. Les premières étapes concernent l'identification et la validation d'une cible moléculaire d'intérêt. Les étapes suivantes concernent le développement à proprement dit de médicament vis à vis de cette cible. L'étape de détermination d'une molécule guide correspond à ce travail de thèse.

Étapes 1 et 2 : identification et validation de cibles moléculaires d'intérêt. Les deux premières étapes concernent la cible moléculaire d'intérêt [8]. Les développements récents de l'informatique et de la bioinformatique (annexe A.1) combinés aux techniques de biologie moléculaire et de biochimie, ont permis l'émergence de domaines nouveaux tels que la génomique fonctionnelle, la génomique structurale ou la protéomique [7]. Ceux-ci permettent d'améliorer notre compréhension de processus biologiques et pathologiques, de la séquence des gènes aux interactions moléculaires impliquées. Ainsi, les cibles moléculaires associées aux pathologies peuvent-elles être identifiées et validées [7, 8].

Étape 3 : mise au point d'un test d'évaluation des composés actifs. Avant toute conception, il convient de disposer d'une méthode de sélection la plus spécifique possible afin d'identifier les composés les plus actifs. L'approche rationnelle nécessite, le plus souvent, une étape

combinatoire, laquelle génère une grande diversité de composés [9, 10, 11]. Selon la taille de la banque de composés générés lors de l'étape combinatoire, des méthodes de criblage à haut débit ou HTS (« High Throughput Screening ») [12], seront nécessaires. Plus récemment, les développements de la modélisation moléculaire laissent envisager, à l'avenir, des méthodes de criblage *in-silico* ou vHTS (« virtual HTS ») [13].

Étape 4 : identification de molécules guides. Après la mise au point d'une méthode d'évaluation, le processus dans son ensemble est itératif (figure I-4, étapes 4 à 7). La première de ces étapes correspond à l'identification ou la conception d'une ou plusieurs molécules guides (« lead »), capables d'interagir avec la cible choisie. Le seul critère d'affinité est utilisé à cette fin. Une approche plus récente, ou « criblage *in-vivo* », consiste à sélectionner, dès la première étape, les molécules possédant des critères biologiques compatibles avec le développement de médicaments (biodisponibilité, pharmacocinétique...). De cette façon, l'étape d'optimisation s'avère plus efficace.

Étape 5 : optimisation des propriétés biologiques de la molécule guide. Ultérieurement, les propriétés biologiques (affinité, spécificité, biodisponibilité, toxicité...) des molécules guides sont optimisées par diverses modifications de leur composition. Ces modifications peuvent être déterminées à partir, soit de l'analyse structurale et/ou fonctionnelle du système, soit de façon aléatoire. Toutefois compte tenu de l'impossibilité actuelle à prédire les propriétés biologiques d'un composé donné à partir de sa seule structure, il est nécessaire d'introduire la diversité moléculaire la plus complète possible. Dans ce but, des banques de composés variés de nature diverse, aussi bien organiques que protéiques, sont construites. La chimie combinatoire, qui trouve son origine dans la chimie des peptides [4, 14] permet de construire d'importantes banques de composés organiques ou peptidiques. La grande diversité ainsi générée [10, 11] explique l'importance de cette approche dans la découverte de nouveaux composés actifs [9]. Plus récemment le développement de la synthèse à diversité orientée (ou DOS) [15, 16] a permis d'augmenter la complexité des composés organiques synthétisés. En outre, dans le cas de composés peptidiques, la diversité ou la modification de fonctions peut être obtenue par des techniques de biologie moléculaire regroupées sous le terme d'évolution dirigée [6]. La méthode de « phage display » [5, 17], couplée à celles-ci, rend possible l'identification des peptides actifs contenus dans ces banques [18].

Étape 6 : sélection des composés actifs. Enfin, dans la dernière étape de chaque cycle, les molécules issues des banques créées à partir des molécules guides sont testées grâce, notamment, aux méthodes de criblage à haut débit (HTS [12] ou vHTS [13]) et, seules les molécules les plus efficaces, sont conservées.

Étape 7 : mise au point de médicament. Le processus étant itératif, les molécules identifiées comme les plus actives peuvent être de nouveau utilisées comme guide pour un cycle supplémentaire d'optimisation/sélection. Lorsque l'activité d'un composé est estimée acceptable, le processus de développement de médicament se poursuit par la réalisation d'essais précliniques et de formulation adéquate

Remarques

Malgré la mise au point et l'utilisation généralisée de la conception rationnelle, la conception de nouveaux médicaments est en déclin [19]. Plusieurs travaux incriminent cette approche, du fait d'une prise en compte incomplète des systèmes et processus biologiques. Une même pathologie peut, en effet, être associée à plusieurs protéines, de ce fait ; l'inhibition d'une seule cible ne peut suffire. Les limitations concernent donc essentiellement le choix de la cible d'intérêt et non pas l'activité des composés vis-à-vis de celle-ci. Certains envisagent une approche, combinant à la fois conception rationnelle et conception basée sur la physiologie (« **Physiology Based Drug Design** »), via l'utilisation de criblage *in vivo* notamment.

I.1.3 Identification/Conception de « molécule guide »

Le travail effectué au cours de cette thèse concerne l'étape de détermination de « molécules guides » (conception de ligands). Deux objectifs à celle-ci : améliorer la compréhension des interactions protéine-protéine et développer une méthode de conception de ligands, pouvant être utilisée dans un processus plus large de conception de médicaments. La conception rationnelle s'est largement répandue notamment grâce à l'apport croissant d'informations structurales, ou « **Structure Based Drug Design** » [20, 21, 22]. Cette dernière repose sur l'utilisation de la bioinformatique et de la modélisation moléculaire (annexe A.1), désigné sous le nom de « **Computer Aided Drug Design** » [23]. L'accroissement en parallèle, représenté sur la figure I-5, des informations structurales de protéines contenues dans la PDB (figure I-5, A ; annexe A.2) ainsi que de la puissance des calculateurs (figure I-5, B) ont largement contribué au développement de ces approches.

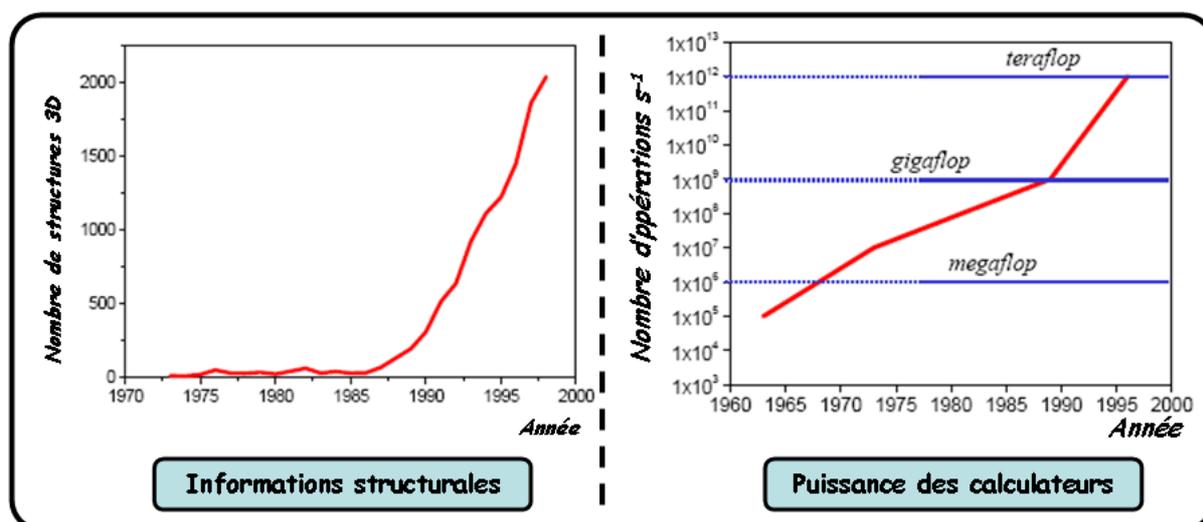


Figure I-5 : Évolution parallèle des informations structurales contenues dans la PDB et de la puissance des calculateurs. Ces évolutions facilitent les approches basées sur la structure. [didier.rognan@pharma.u-strasbg.fr].

La connaissance d'informations structurales autorise l'analyse, à l'échelle atomique, des déterminants d'une interaction donnée, autrement dénommés « pharmacophores » [24, 25]. Ce concept a conduit à proposer l'idée que les déterminants fonctionnels sont distincts de déterminants structuraux, ou épitopes structuraux [26]. L'identification de pharmacophores permet de construire un modèle (2D ou 3D) qui représente l'agencement des fonctions chimiques ou des résidus nécessaires à l'interaction avec la cible choisie [27, 28]. Les figures I-6 et I-6bis illustrent l'utilisation de modèle pharmacophorique pour la conception de ligand de la protéase HIV-1 [29].

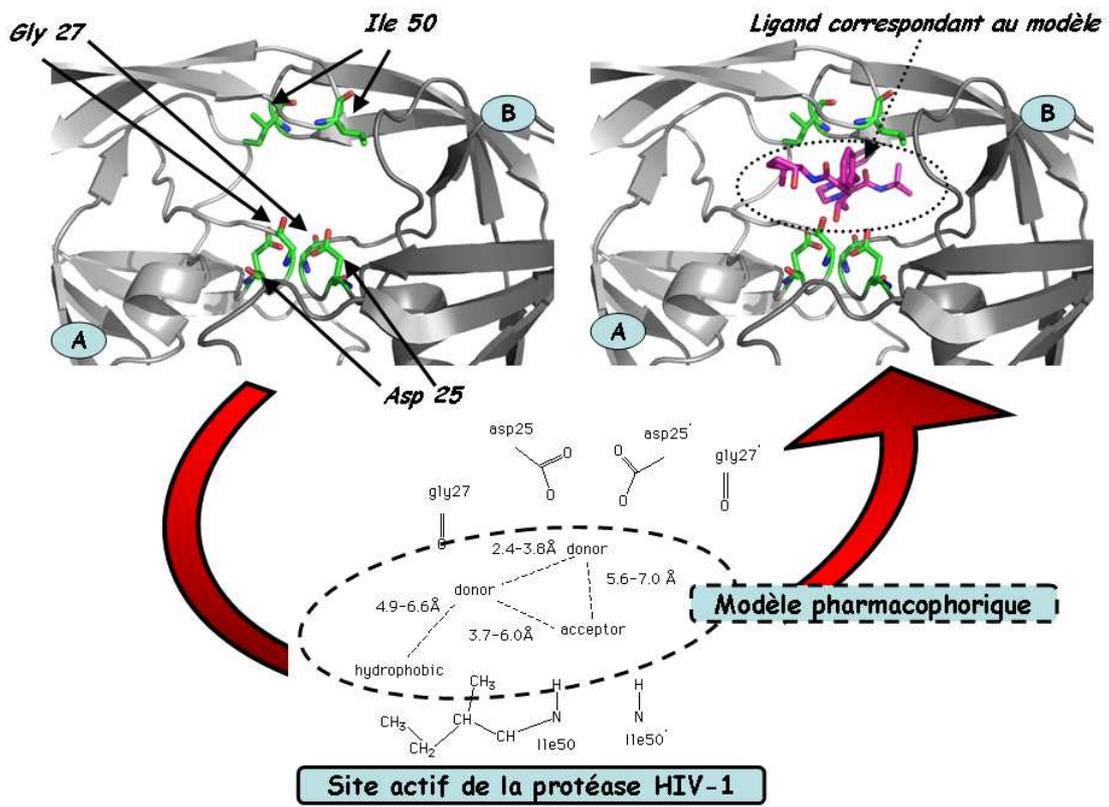


Figure I-6 : Fichier PDB 2BPV [29]. La protéase HIV-1 (en gris, monomères A et B) est représentée en mode « cartoon ». Le ligand L735-524 (en magenta) ainsi que les résidus du site de liaison de la protéase (en vert) sont représentés en mode « stick ».

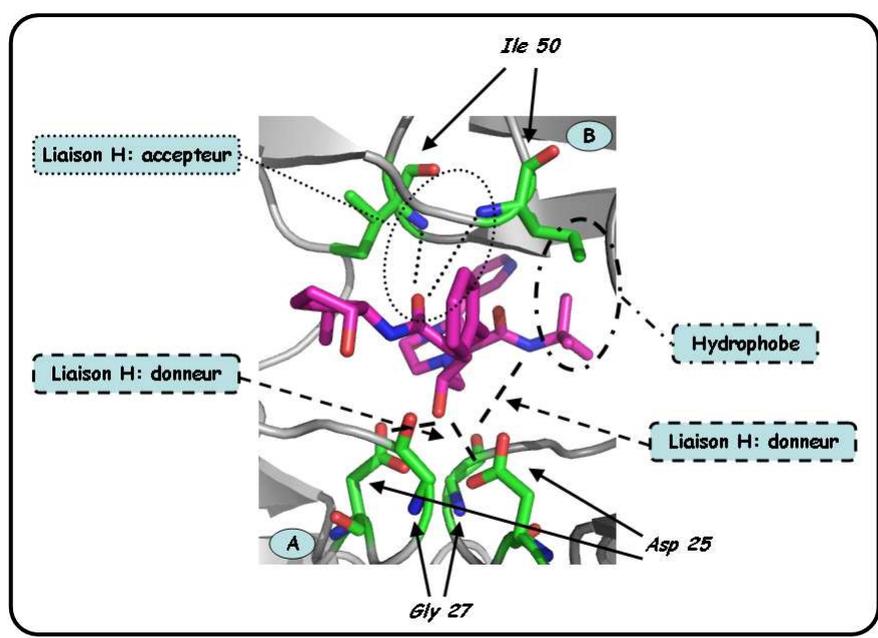


Figure I-6bis : Représentation agrandie du ligand L735-524 dans le site de liaison de la protéase HIV-1. Le ligand correspond au modèle pharmacophorique.

La connaissance structurale du site actif a permis de mettre en évidence la composition chimique de sites potentiels pour le développement d'inhibiteur. L'inhibiteur L735-524 (*Indinavir*, AMM en 1996), développé par MERCK correspond à ce modèle pharmacophorique.

Pour cibler un processus biologique donné, différentes voies, schématisées figure I-7, sont empruntées pour construire le modèle d'interaction, selon la nature des informations structurales disponibles, contenues dans la PDB notamment (annexe A.2). Ces dernières sont les suivantes :

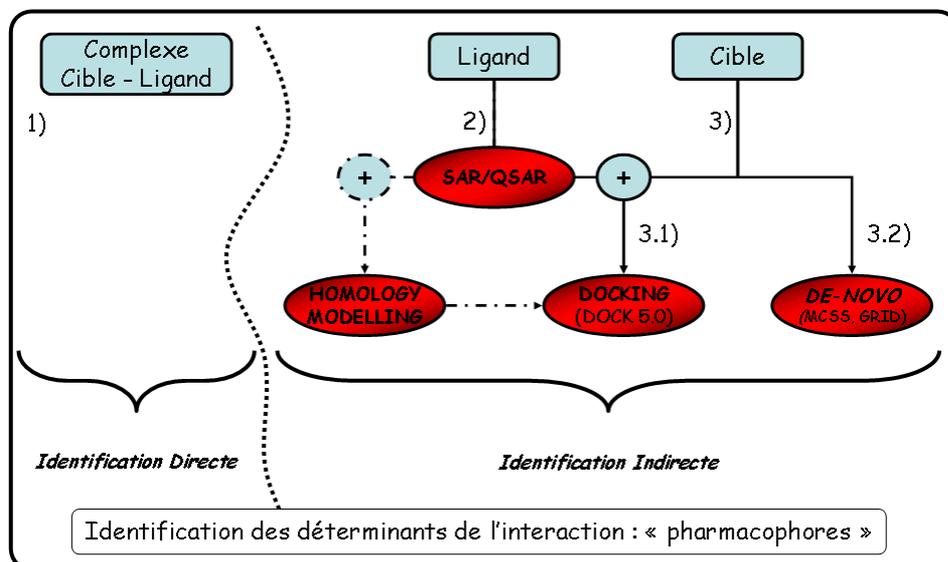


Figure I-7 : Identification des pharmacophores à partir d'informations structurales disponibles.

1) La structure d'un complexe cible/ligand est disponible. Cela rend possible l'identification du site de liaison et des interactions entre la cible et son ligand. Les pharmacophores sont alors directement déduits de l'analyse de la structure du complexe selon des critères de distance et de géométrie. Ces observations peuvent, en outre, être corrélées avec divers résultats expérimentaux : microcalorimétrie [30], « alanine scanning » [31], méthodes d'évolution dirigée [6], BIAcore (1990). Ces différentes techniques permettent de quantifier l'implication énergétique de chaque fonction chimique dans la formation du complexe cible-ligand. Ce cas est optimal pour la conception de nouveaux ligands étant donné que des observations structurales fiables peuvent être directement reliées à diverses informations expérimentales.

2) La structure d'un ligand est disponible. La structure 3D de la cible étant inconnue, l'approche de conception ne repose que sur la connaissance de la structure d'un ou plusieurs ligands identifiés, correspondant au concept de « **Ligand Based Drug Design** » [32]. L'étude de la relation entre la structure du ligand et son activité vis à vis de la cible (ou SAR, « Structure Activity Relationship ») est alors nécessaire. Dans ce contexte, l'utilisation de la modélisation moléculaire facilite cette étude, via des méthodes quantitatives. Dans les études QSAR (« Quantitative Structure Activity Relationship »), les modifications de structure sont corrélées avec une propriété donnée (interaction avec une cible par exemple). Plusieurs exemples décrits dans la littérature décrivent la mise en application d'une telle approche basée sur la structure

d'un ligand pour l'identification et l'optimisation de molécules guides. Par exemple, la cyclophiline A (CypA) a été identifiée comme cible moléculaire d'intérêt pour son rôle dans l'infection du HIV [33]. A partir de la structure de ligands de la CypA une molécule guide a été identifiée, d'affinité 0,3 μM , qui, après optimisation, a fourni deux ligands, d'affinité 14 et 20 nM [34].

Par ailleurs, une alternative peut être envisagée en l'absence d'information structurale sur la cible. Il a été démontré que deux séquences fortement homologues conservaient le même repliement [35]. La modélisation par homologie ou « Homology Modelling » [36] permet de remplacer l'information structurale manquante, à condition de disposer de la structure d'une protéine présentant une forte homologie de séquence avec la cible. Il est estimé que, deux structures peuvent être considérées identiques, lorsque leur RMSD (obtenue *via* la superposition des atomes de leurs chaînes principales respectives) est inférieur à 2 Å. D'après le tableau I-8, à partir de 60% d'homologie, 70% des prédictions peuvent donc être considérées comme correctes. Du fait de la diversité des structures présentes dans la PDB, cette approche peut être couramment employée. En revanche, d'après ces résultats, elle ne peut s'appliquer dans tous les cas.

Percent sequence identity ^a	Total number of models ^b	Percent ^c models with rmsd lower than 1 Å	Percent models with rmsd lower than 2 Å	Percent models with rmsd lower than 3 Å	Percent models with rmsd lower than 4 Å	Percent models with rmsd lower than 5 Å	Percent models with rmsd higher than 5 Å
25-29	125	0	10	30	46	67	33
30-39	222	0	18	45	66	77	23
40-49	156	9	44	63	78	91	9
50-59	155	18	55	79	86	91	9
60-69	145	38	72	85	91	92	8
70-79	137	42	71	82	85	88	12
80-89	173	45	79	86	94	95	5
90-95	88	59	78	83	86	91	9

Tableau I-8: Conservation du repliement en fonction du degré d'homologie [35]. Pour des valeurs de RMSD supérieures à 2 Å, deux structures peuvent être considérées comme similaires. A partir de 60% d'homologie, la modélisation par homologie permet une prédiction correcte dans 70% des cas.

3) La structure de la cible est disponible. L'approche correspond alors au concept de « Target Based Drug Design ». Deux cas sont à distinguer :

3.1) La structure d'au moins un ligand est disponible. Ne disposant pas de la structure du complexe, il est envisageable de positionner le ligand à la surface de la cible au site considéré

comme le plus favorable, énergétiquement et/ou stériquement. Cette approche, dénommée ancrage (ou « docking ») [37, 38], est appliquée par de nombreux logiciels [39] dont le programme DOCK [40, *], couramment utilisé pour la conception de ligands. Actuellement, la procédure d'ancrage est relativement systématique, comme décrit dans la figure I-9. Elle peut différer en revanche par la fonction d'énergie utilisée pour le calcul de l'énergie d'association. Un grand nombre d'approches ont été proposées afin d'estimer cette énergie ; toutefois ces dernières peuvent être regroupées selon trois catégories distinctes [41] : celles basées sur un modèle physique (ou PEEF, pour « Physical Effective Energy Functions »), sur l'analyse statistique de structures 3D de protéines (ou SEEF, pour « Statistical Effective Energy Functions ») ou reposant sur l'analyse d'informations thermodynamiques, obtenues expérimentalement (ou EEEF, pour « Empirical Effective Energy Functions »).

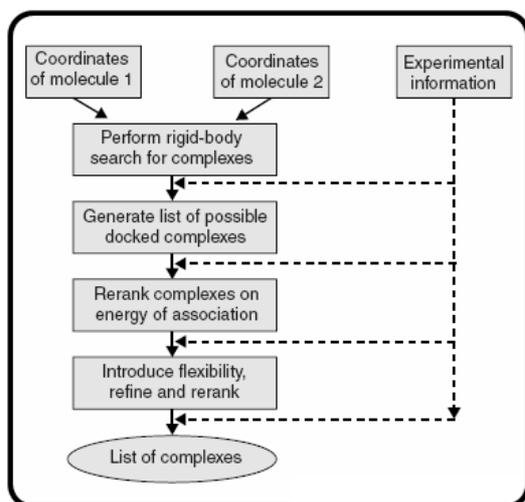


Figure I-9 : Schéma général de la procédure d'ancrage [39]. Les diverses méthodes ou logiciels existants diffèrent selon la fonction d'énergie utilisée pour estimer la validité des solutions proposées.

3.2) Aucune structure de ligands n'est disponible. La détermination de pharmacophores ne peut être obtenue que par l'utilisation de méthodes de modélisation moléculaire. A cet égard, plusieurs programmes ont été écrits, dont le programme GRID [42, **]. Dans ce programme, différentes sondes moléculaires (groupement méthyle, molécule d'eau, sphère chargée...) sont utilisées pour déterminer, à la surface d'une protéine donnée, les sites de liaison énergétiquement favorables à un type d'interaction particulier. Ces stratégies sont représentées par la figure I-10. Les pharmacophores, ainsi déterminés, autorisent la recherche de molécule correspondante dans diverses bases de données de structures, schématisé sur la figure I-10, partie A. Une autre option repose sur l'utilisation couplée des programmes MCSS pour « Multiple Copy Simultaneous Search » [43] et HOOK [44].

* <http://dock.compbio.ucsf.edu/>

** http://www.moldiscovery.com/soft_grid.php

Le premier programme (MCSS) permet de définir les groupements fonctionnels et leur positionnement les plus adaptés pour un site de fixation déterminé, d'une manière analogue au programme GRID. Le second programme (HOOK) détermine les structures chimiques capables de relier les groupements fonctionnels entre eux dans la topologie décrite par MCSS, comme schématisé sur la figure I-10, partie B.

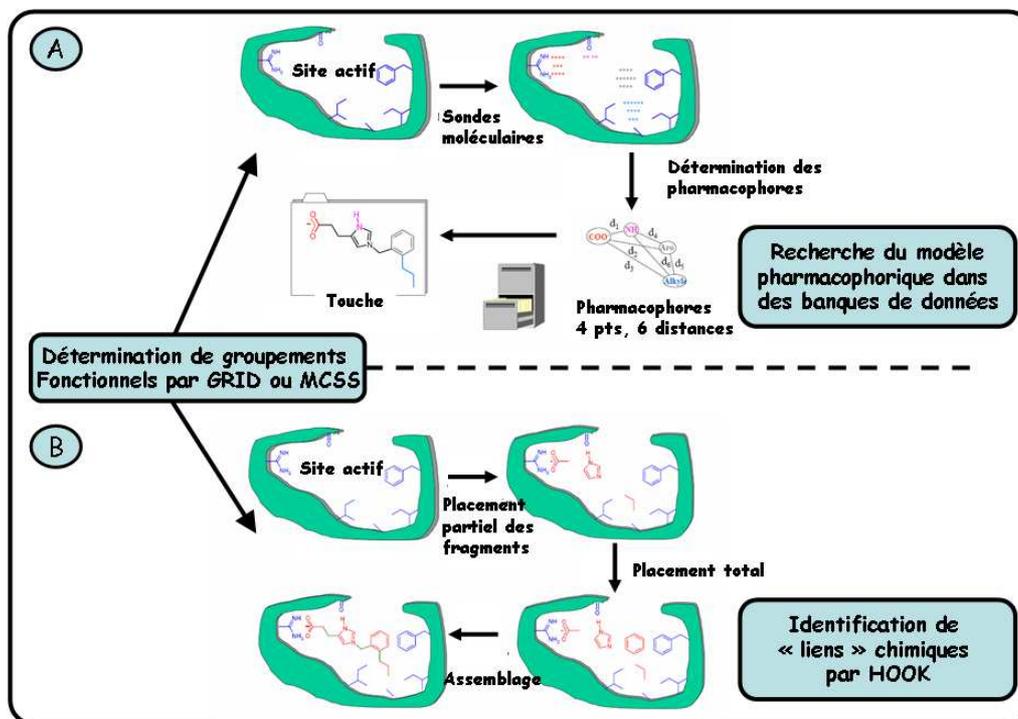


Figure I-10 : Conception de novo par modélisation moléculaire. Partie A : l'utilisation des programmes GRID ou MCSS suivie de recherche dans des bases de données. Partie B : l'utilisation couplée de MCSS et HOOK pour la conception de nouveaux ligands (partie B). [didier.rognan@pharma.u-strasbg.fr]

Ainsi, ne disposant que de la structure du récepteur, un ensemble de pharmacophores sont identifiés. Ces approches correspondent au concept de « *De Novo Drug Design* ». Plusieurs exemples montrent leur mise en application dans la conception de ligands. A partir de la seule structure cristallographique de la sialidase du virus *influenza*, deux inhibiteurs ont été conçus par l'utilisation du programme GRID [45]. Ces derniers présentent des constantes d'inhibition pour la sialidase de $5 \cdot 10^{-8}$ M et $2 \cdot 10^{-10}$ M, et constituent les inhibiteurs les plus puissants décrits dans la littérature. Ce travail a abouti à la mise sur le marché du *Relenza*®, développé par GlaxoSmithKline pour le traitement des virus *influenza* de type A et B. Dans un autre exemple, le programme MCSS a été utilisé pour le développement d'inhibiteur de la thrombine. Plusieurs molécules ont-elles ainsi été générées, dont l'une possédant une constante d'inhibition de $1,7 \mu\text{M}$ [46].

Remarque

Cependant, l'approche de conception de médicament *de novo* peut être envisagée, en absence de pharmacophores identifiés, en utilisant la méthode d'ancrage précédemment décrite. A partir d'un site de liaison localisé, les bases de données de structures sont criblées, chaque structure étant positionnée, au mieux, au niveau de ce site. La capacité de chaque molécule à se lier étant évaluée par un calcul d'énergie de liaison. Cette stratégie, schématisée sur la figure I-11, reste actuellement peu fiable compte tenu de l'absence d'informations concernant les ligands et les sites de liaisons potentiels.

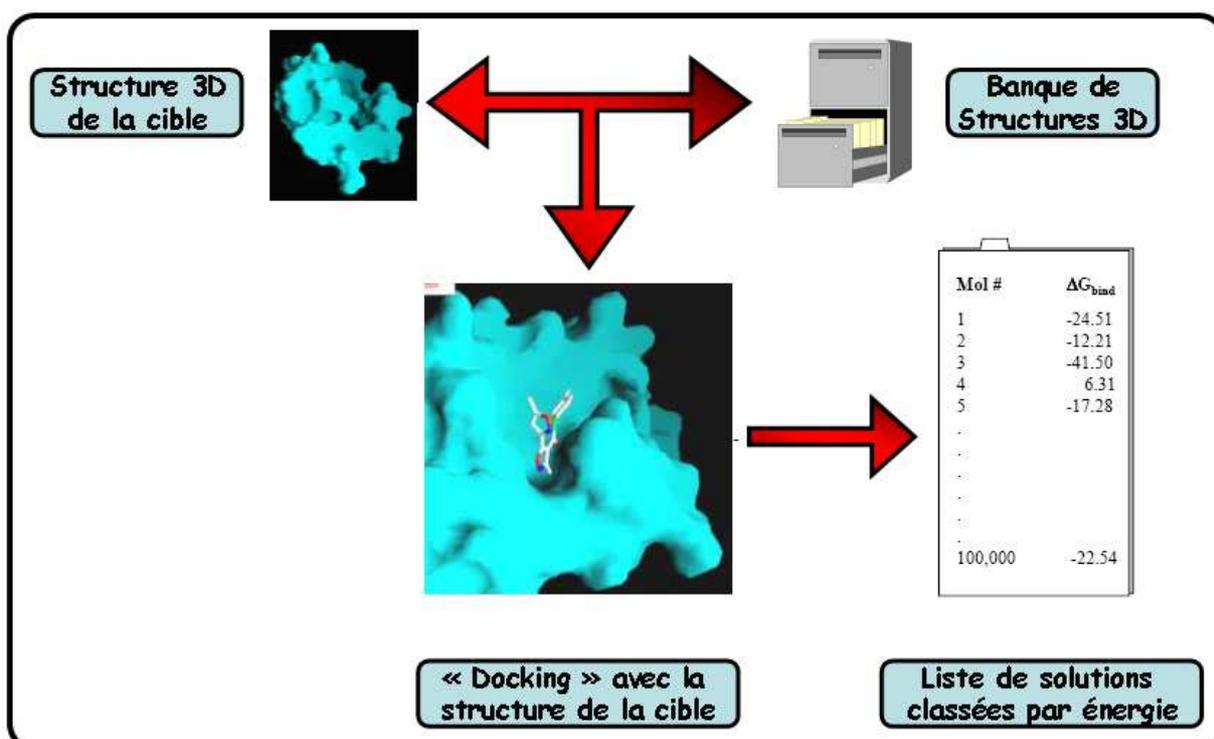


Figure I-11 : Utilisation de méthode d'ancrage pour l'identification de molécules guides à partir de la seule structure tridimensionnelle de la cible. [didier.rognan@pharma.u-strasbg.fr]

Utilisation du modèle pharmacophorique

Une fois le modèle pharmacophorique de l'interaction avec la cible choisie défini par l'une de ces méthodes, celui-ci est utilisé pour rechercher, dans des bases de données, toute structure moléculaire (plateforme structurale) lui correspondant. Cette recherche est largement facilitée par nombres d'outils bioinformatiques d'analyse de banques de structures, stratégie couramment utilisée dans le cas de composés organiques [27, 47], mais beaucoup moins dans le cas de ligands de nature protéique.

I.1.4 Conclusion

Les différentes approches de conception de ligands, basées sur la structure ont prouvé, dans un bon nombre de cas, leur capacité à identifier ou créer de nouvelles molécules, capables de lier la cible désignée. Toutefois, la conception rationnelle présente des limitations, notamment concernant sa productivité, en termes de nouveaux médicaments produits [19]. Cette observation est, d'après l'auteur, la conséquence d'une prise en compte partielle et simplifiée des systèmes biologiques [19]. Dans le cadre de cette thèse, la conception de ligands a été envisagée par une approche basée sur la structure du ligand seul, à partir de la définition d'un modèle pharmacophorique. En revanche, la structure de la cible de ce ligand, s'est révélée nécessaire pour la sélection des protéines les plus aptes à reproduire l'interaction ciblée. La méthode proposée, dans le cadre de cette thèse, sera entièrement détaillée dans le chapitre III.

I.2 De l'utilisation des protéines comme ligands

Différents types de molécules peuvent être utilisés en tant que ligands : composés organiques, peptides, peptido-mimétiques ou protéines (*). Ce travail de thèse n'a concerné que le développement de molécules de nature protéique.

I.2.1 Implications pour la conception de ligands

La conception d'un ligand dans l'approche rationnelle, à partir d'une cible identifiée, repose sur deux étapes : la détermination d'une molécule guide et son optimisation. Tandis que la première a pour but d'identifier des molécules affines, quelle que soit leur nature, la seconde dépend grandement des caractéristiques des molécules utilisées.

I.2.1-A Optimisation des ligands

Lors de l'optimisation, l'introduction de la diversité moléculaire la plus complète possible ne peut se faire qu'à condition que le nombre de positions à modifier soit restreint. Dans le cas contraire, le nombre de molécules générées serait trop important pour que chacune soit analysée et testée. Par ailleurs, la synthèse chimique de molécules de nature peptidique est limitée approximativement à une centaine de résidus, pour une utilisation routinière (annexe C.3). Le recours à la synthèse chimique autorise, en outre, l'optimisation par introduction de résidus non-naturels, de conformation contrainte ou exposant des fonctions chimiques supplémentaires (en comparaison aux résidus naturels). Ces considérations expliquent le recours à des molécules de faible poids moléculaire, facilitant la mise en application d'une approche de conception de ligand. Ceci est d'autant plus justifié puisque l'analyse de complexes protéiques a montré que seul un faible nombre de résidus de l'interface, contribuait majoritairement à l'énergie de liaison [26, 48, 49]. Or, la conception de ligands de très faible poids moléculaire s'est révélée peu évidente pour plusieurs raisons [50, 51], notamment en raison de la taille de l'interface. Cette dernière, supérieure à 1000\AA^2 dans de nombreux complexes protéiques de structure connue [52], dépasse largement celle de molécules organiques et de peptides de poids moléculaire inférieur à 500Da. L'optimisation de ligands, dont la surface d'interaction est de faible taille, est donc limitée du fait d'un faible nombre de fonctions chimiques modifiables. Dans le cas des molécules protéiques, les peptides sont un bon exemple de ligands optimisables.

*** Protéines / Mini-protéines / Peptides**

Les protéines sont des polymères d'acides aminés (alors appelés résidus), liés entre eux par des liaisons amide, entre la fonction acide carboxylique du résidu N et la fonction amide du résidu N+1. Selon le nombre de résidus N présents dans le polymère on parlera de peptide ($N < 20-50$), de polypeptide ($20-50 < N < 100$), autrement appelé mini-protéines selon certains auteurs, ou de protéine ($N > 100$).

Malgré leur faible taille (inférieure à 20 résidus), les peptides sont dotés de nombreuses activités biologiques (inhibition enzymatique, antihypertenseurs, antioxydants, antimicrobiens, neurotransmetteurs, opioïdes, propriétés antigéniques...). Leur faible taille facilite les approches combinatoires générant ainsi une grande diversité, constituant une source importante d'information pour la compréhension des interactions avec diverses cibles moléculaires [9]. D'autre part, leur capacité à mimer des ligands protéiques de taille plus importante a été démontrée. Par exemple, les fragments constants (Fc) des immunoglobulines G (IgG) humaines interagissent avec diverses protéines, au même site de liaison, comme illustré sur la figure I-12.

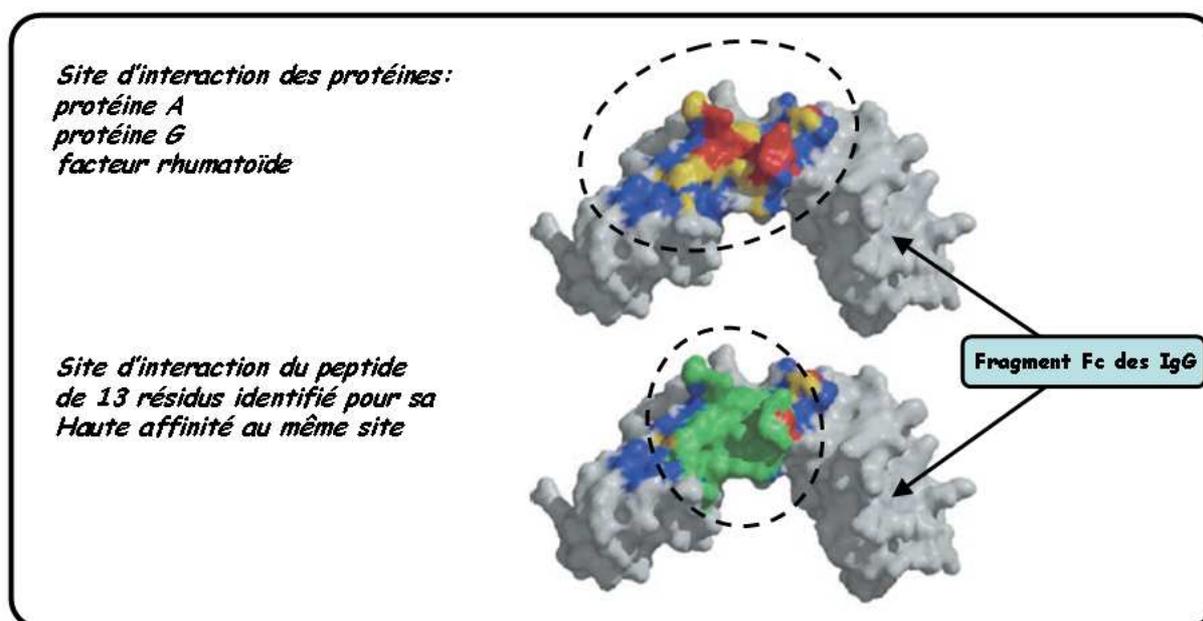


Figure I-12 : Représentation du fragment Fc des IgG en interaction avec la protéine A, la protéine G et le facteur rhumatoïde. Les résidus sont colorés en bleu, jaune et rouge selon qu'ils soient impliqués dans une, deux ou trois interactions (partie haute). Le site de liaison du peptide identifié (en vert) est représenté dans la partie basse. [53]

Ces protéines, domaine B1 de la protéine A, domaine C2 de la protéine G, facteur rhumatoïde et récepteur néo-natal Fc, possèdent des tailles variables, toutes supérieures à 60 résidus [53]. La conception de banques de peptides (de 13 résidus) par « phage display » a permis d'identifier plusieurs ligands des fragments Fc des IgG. Un peptide notamment possède une affinité comparable (de l'ordre du nanomolaire) avec les ligands connus, et ceci au même site de liaison, illustré sur la figure I-12. Toutefois, l'utilisation de ligands peptidiques est fortement limitée notamment en raison d'une sensibilité importante à la dégradation enzymatique dans le cas des peptides linéaires [54, 55, 56] et surtout une trop grande flexibilité [57, 58].

I.2.1-B Importance de la structure pour la conception de ligands

Du fait de cette flexibilité, les peptides sont présents sous différentes conformations, énergétiquement proches. Dès lors, il est relativement difficile de connaître la conformation

bioactive d'un peptide donné [58]. La connaissance de la structure 3D d'un peptide ne peut alors être utilisée pour une approche basée sur la structure. Par ailleurs, l'existence de plusieurs conformères se traduit par un coût entropique important, réduisant l'activité de ligands peptidiques en comparaison avec des structures plus rigides. Pour limiter ces inconvénients, l'approche courante repose sur le recours à des composés « peptido-mimétiques » [57, 59, 54], illustrée par la figure I-13.

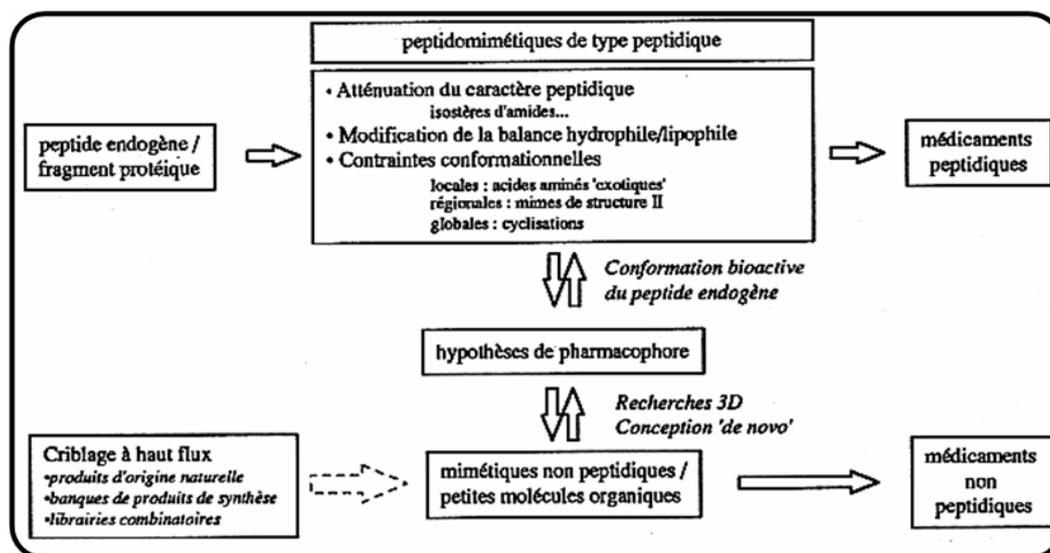


Figure I-13 : L'approche « peptido-mimétique ». [Chantal Rubin-Carrez, 1998, Ateliers de Formation INSERM]

Leur structure est généralement dérivée de la conformation bioactive de peptides endogènes (identifiée par des études SAR) par diverses modifications chimiques permettant d'atténuer le caractère peptidique [57]. L'intérêt principal consiste à restreindre l'espace conformationnel accessible des peptides [60] par formation de cycles intra-chaîne et introduction de résidus contraints. A titre d'exemple, ceci a permis d'améliorer la capacité de liaison des peptides C, correspondant à la partie C terminale de la protéine gp41 de HIV-1 [61]. Les peptides C inhibent l'entrée du virus HIV-1, en liant la protéine gp41 ; toutefois leur affinité est trop faible pour envisager leur utilisation à des fins thérapeutiques. L'introduction de restrictions conformationnelles a amélioré l'affinité pour la protéine gp41 d'un facteur 7, et diminué l'entrée du virus d'un facteur 15 [61]. Il est, par ailleurs, possible d'améliorer leurs propriétés pharmacocinétiques [57] par introduction de résidus non-naturels ou de liens non-peptidiques.

Une alternative, à l'utilisation de peptides et peptido-mimétiques, repose sur l'utilisation de mini-protéines. Leur taille, comprise entre 20 et 100 résidus, rend possible les approches combinatoires et la synthèse chimique. De plus, malgré une taille faible, ces mini-protéines possèdent, de façon générale, un repliement relativement rigide [62]. Leur faible taille exclut la

présence d'un cœur hydrophobe contrairement aux protéines de haut poids moléculaire ; néanmoins la stabilité structurale est assurée par une forte densité de ponts disulfures ou la présence de sites de liaisons de métaux [62]. L'existence d'une structure bien définie facilite l'utilisation de méthodes de résolution de structures (RMN, radiocristallographie). Il est ainsi possible de mettre à profit les informations de la PDB (annexe A.2), contenant de nombreuses structures de mini-protéines naturelles : toxines animales [63], inhibiteurs de protéases [64], « knottins » (*) [65], protéines à doigt de zinc [66], protéines à « Leucin zippers » [67]. Les différents repliements sont représentés sur la figure I-14 [68, 69, 70, 71, 72].

I.2.1-C Conclusion

Les mini-protéines apparaissent comme de bons candidats pour la conception de ligands, de part leur taille compatible avec différentes approches de synthèse et d'optimisation, et leur repliement bien défini facilitant les approches basées sur la structure.

*** Knottin**

Type de repliement particulier caractérisé par la présence d'un pont disulfure pénétrant un macrocycle formé par deux autres ponts qui connectent les segments de la chaîne polypeptidique. Ces protéines sont de petite taille et très riches en cystéine.

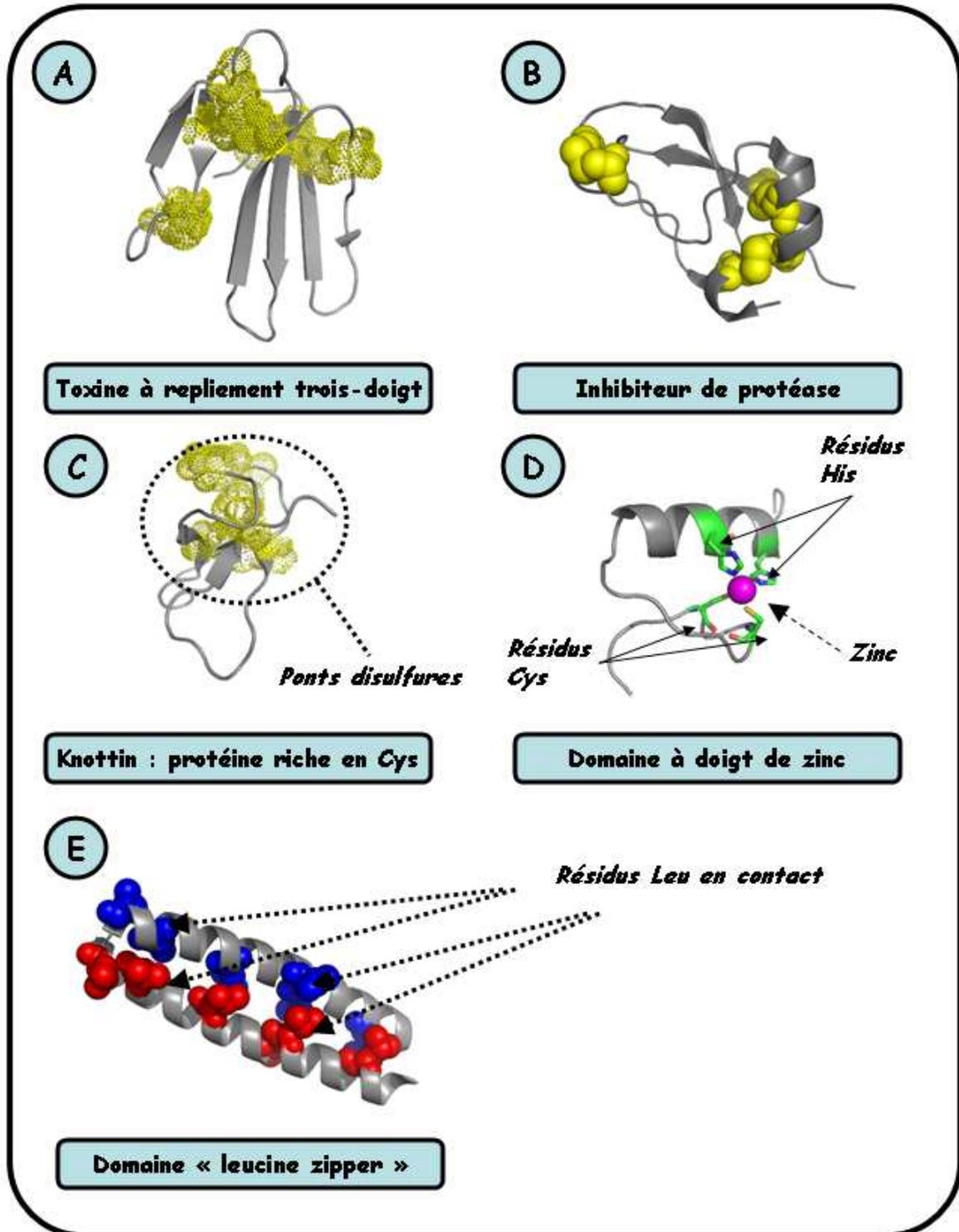


Figure I-14 : Différents repliements existants parmi les mini-protéines. Les mini-protéines (en gris) sont représentées en « cartoon », les ponts disulfures (en jaune) en mode « sphere » ou « dot ». A : repliement trois doigts de la toxine bucardine [1IJC, 68]; B : repliement du BPTI [1BPI, 69]; C : repliement du knottin Alo-3 [1Q3J, 70]; D : structure du facteur de transcription Sp1, l'atome de zinc (magenta) est représenté en mode « sphere », les résidus chélatants (C5, C8, H21 et H25) sont représentés en mode « ribbon » [1SP1, 71]; E : structure du facteur de transcription GCN4, les résidus Leu (bleu et rouge pour les sous unités A et B respectivement) sont représentés en mode « sphere » [2ZTA, 72].

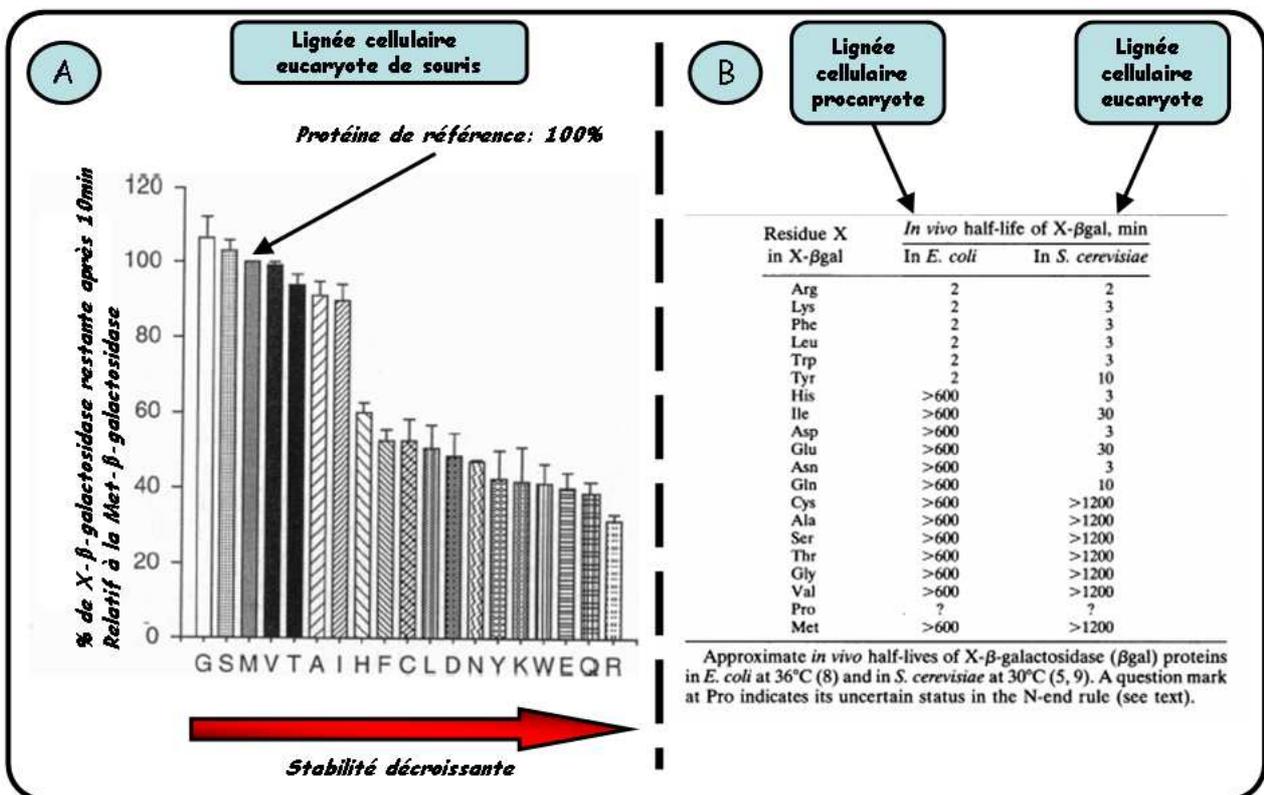


Figure 1-15 : Estimation de la dégradation de protéines *in-vivo* en fonction de la nature du résidu N terminal (« N end rule »). La partie A illustre cette dégradation dans une lignée cellulaire eucaryote de souris par rapport à la quantité de protéine restante après 10min [82]. Le tableau, partie B, liste les valeurs de demi-vie attribuée à chaque protéine, dans une lignée cellulaire procaryote ou eucaryote [83].

I.2.2 Les protéines comme agents thérapeutiques et/ou diagnostiques

Un des intérêts de la conception de ligands, concerne le processus de conception de médicaments. L'accroissement du nombre de protéines utilisées, en tant qu'agents thérapeutiques et/ou diagnostiques, met en évidence l'intérêt de ce type de ligands.

I.2.2-A Protéines à usage thérapeutique

Depuis une dizaine d'années, le nombre de protéines à usage thérapeutique ne cesse d'augmenter [73] et les prévisions, estimées par plusieurs travaux, pour les années à venir sont optimistes [73, 74]. Initialement, leur utilisation se limitait à compenser une fonction déficiente ou insuffisante : injection d'insuline en cas de diabète ou hormone de croissance en cas de nanisme par exemple. Désormais, le recours à l'ingénierie des protéines permet d'élargir les possibilités d'utilisation de celles-ci [75, 76, 77].

Les méthodes d'optimisation [75] peuvent être utilisées pour améliorer les critères d'affinité et spécificité, ou d'autres propriétés liées à l'utilisation de molécules thérapeutiques, dont la stabilité *in-vivo* et la toxicité ou l'immunogénicité [78, 79]. Les protéines, même d'origine humaine, sont capables d'induire des réactions du système immunitaire de l'organisme. Différentes stratégies d'optimisation ont été mises au point de façon à réduire cet inconvénient [78, 80, 81] : modifier la séquence primaire de façon à supprimer les épitopes antigéniques [81] et/ou améliorer la solubilité évitant ainsi la formation d'agrégats immunogènes [80], masquer les sites d'interactions avec les anticorps par ajout de PEG à la surface des protéines [80].

Les protéines ont une durée de vie *in-vivo* limitée, du fait de la dégradation par les protéases et le protéasome. Dans ce dernier cas par exemple, la nature du résidu en position N terminale est déterminante pour la durée de vie de la protéine [82, 83]. La figure I-15, partie A, montre l'évolution de la quantité d'une protéine de référence (la β -galactosidase) restante après dix minutes selon la nature du résidu en position N terminale, dans une lignée cellulaire eucaryote. Pour 12, parmi les 20 acides aminés naturels, la dégradation des protéines est supérieure à 40%. Le temps de demi-vie de cette protéine de référence a été estimé dans deux lignées, bactérienne (*E. Coli*) et eucaryote (*S. Cerevisiae*), et est présenté dans le tableau I-15, partie B. D'autre part, l'action des protéases est, quant à elle, ralentie lorsque la structure des protéines est repliée. Il en résulte une stabilité variable d'une protéine à l'autre. Le cas des anticorps, dont la durée de vie dépasse un jour, rend concevable l'utilisation de protéines pour des applications *in-vivo*.

Médicament	Protéine	Modifications	Indication	Impact des modifications
Betaseron [49] (Chiron)	IFN β -1b	Cys17→Ser, glycosylation, no Met at N terminus	Multiple sclerosis	Increased production yields
CEA-Scan* [50] (Immunomedics)	Anti-carcinoembryonic antigen Ab fragment	^{99m} Tc-labeled	Detection of colorectal cancer	Imaging
Cerezyme [51,52] (Genzyme)	Glucocerebrosidase	Arg495→His, glycosylation	Type I Gaucher's disease	Change introduced during cloning, cell targeting
Enbrel [53] (Immunex)	Soluble TNFR	Dimeric TNFR fused to IgG1 Fc	Rheumatoid arthritis	Enhanced affinity and half-life
Geref [54] (Serono)	Human GHRH (synthetic)	N terminus only	Idiopathic pediatric growth hormone deficiency	Minimal biologically active core
Herceptin [55] (Genentech/Roche)	Anti-HER2 Mab	Humanized Mab	Metastatic breast cancer	Reduced immunogenicity, enhanced affinity
Humalog [56] (Lilly)	Insulin	Reversal of Lys and Pro at C terminus of B chain	Diabetes	Faster onset of action
Infergen [57] (Amgen)	Type 1 IFN α	Consensus sequence of four α subtypes, four additional amino acid changes	Hepatitis C	Increased activity, facilitated molecular construction
Lantus [58] (Aventis)	Long-acting insulin	Asp21→Gly and addition of two Arg to C terminus of B chain	Diabetes	Increased duration of action
Leukine [59] (Immunex)	Human GM-CSF	Pro23→Leu, glycosylation	Bone marrow transplantation	Expression/production
Miacalcin [†] (Novartis)	Calcitonin-salmon (synthetic)	50% identical to human	Postmenopausal osteoporosis	50 Times more potent than human version
Mylotarg [60] (AHP/Celtech)	Anti-CD33 Mab	Humanized, fused calicheamicin	Acute myeloid leukemia	Reduced immunogenicity, target toxin
Neumega (Genetics Institute)	Human interleukin-11	Lacks N-terminal Pro	Thrombocytopenia (chemo-induced)	Expression
Neupogen [†] (Amgen)	Human G-CSF	Additional N-terminal Met, nonglycosylated	Neutropenia (chemo-induced)	Expression
OncoScint* [61] (Cytogen)	Anti-TAG-72 Mab	¹¹¹ In-labeled	Detection of metastases of colorectal or ovarian cancer	Imaging
Ontak [†] (Ligand Pharma.)	IL-2 fusion	Fused to diphtheria toxin	Persistent or recurrent cutaneous T-cell lymphoma	Cell targeting
Proleukin [62] (Chiron)	Human IL-2	Cys125→Ser, no N-terminal Ala, nonglycosylated	Metastatic melanoma and renal cell carcinoma	Increased specific activity
ProstaScint* [†] (Cytogen)	Anti-PMSA Mab	¹¹¹ In-labeled	Detection of prostate cancer and metastases	Imaging
Refacto [63] (Genetics Institute)	Factor VIII	B domain deleted	Hemophilia A	Increased production yields
Refludan [64] (Aventis)	Hirudin	Ile→Leu and no sulfate group on C-terminal Tyr	Heparin-induced thrombocytopenia type II	Changed binding affinity
Remicade [65] (Centocor/Schering Plough)	Anti-TNF α Mab	Chimeric human-murine Mab	Crohn's disease	Reduced immunogenicity
ReoPro [64] (Centocor/Lilly)	Anti-GPIIb/IIIa antibody	Chimeric human-murine Mab	Platelet aggregation; prevention of blood clots	Reduced immunogenicity
Retavase [†] (Centocor/Boehringer)	tPA variant	Kringle-2 and protease domains only	Treatment of acute myocardial nonglycosylated infarction	Longer half-life
Rituxan [60] (IDEC/Genentech/Roche)	Anti-CD20 Mab	Chimeric human-murine Mab	CD20-positive B-cell non-Hodgkin's lymphoma	Reduced immunogenicity
Roferon A [66] (Roche)	IFN α -2a	Rare IFN, contains Lys23 instead of common Arg23 (IFN α -2b)	Cancers and hepatitis C	Cloned rare IFN subtype

● **cytokines** ● **protéines de nature diverse**
● **anticorps**

Figure I-16 : Liste de plusieurs protéines mises sur le marché (USA), en tant qu'agents thérapeutiques ou diagnostiques (lignes entourées) [75].

I.2.2-B Exemples de protéines thérapeutiques

Les anticorps sont des protéines produites par le système immunitaire dont le but est de fixer les molécules étrangères à l'organisme. De part leur faculté à reconnaître des antigènes divers, les anticorps ont été les premières protéines utilisées comme médicament. Leur utilisation s'est

largement développée lors de la mise au point d'anticorps monoclonaux [84], capables de reconnaître une cible précise. Les anticorps présentent de nombreux avantages : ce sont des protéines circulantes, donc très solubles. Leur structure rigide rend la détermination de leur structure 3D aisée. Enfin, les régions impliquées dans la reconnaissance (fragments variables Fv) de structures diverses, sont peu sensibles aux mutations, facilitant les approches d'ingénierie et d'optimisation. Leurs applications sont multiples :

- **à des fins diagnostiques**, ils peuvent être associés à des marqueurs fluorescents ou radioactifs (test ELISA, test de grossesse, marquage de tumeurs...);
- **à des fins thérapeutiques**, leur capacité à déclencher certains mécanismes immunitaires (cytotoxicité, apoptose...) peut être mise à profit vis-à-vis de cellules cancéreuses. Leur capacité de liaison peut être utilisée afin de bloquer diverses protéines circulantes ou récepteurs membranaires, impliqués dans une pathologie.

La liste de plusieurs anticorps monoclonaux mis sur le marché est présentée dans le tableau I-16. Dans ce dernier, sont indiqués : le nom de la protéine utilisée, le nom du médicament, les modifications effectuées pour améliorer ses propriétés biologiques ou modifier son utilisation, et son indication. Les protéines peuvent donc être utilisées dans une grande variété de pathologies, et ne se limitent plus à compenser une fonction déficiente.

Par la suite, d'autres classes de protéines ont été mises à profit notamment différentes cytokines : les interférons, les facteurs de croissance hématopoïétiques, les facteurs de croissance, les interleukines, les CSF (« Colony Stimulating Factor »). Plusieurs d'entre elles sont utilisées en tant que médicaments et sont reportés dans le tableau I-16. Ces protéines ou glycoprotéines solubles (représentées figure I-17) de faible poids moléculaires (compris entre 8 et 30kDa, la plupart contenant moins de 150 résidus), sont caractérisées d'un point de vue structural par des repliements bien définis. Pour ces raisons, les cytokines peuvent donc facilement s'inscrire dans une approche de conception basée sur la structure. L'utilisation des cytokines, à des fins thérapeutiques, est essentiellement liée à leurs fonctions dans l'organisme. Elles sont, en effet, impliquées dans les voies de communication cellulaire, en particulier entre cellules du système immunitaire. De ce fait, leur utilisation présente un intérêt particulier pour le développement de médicaments dans diverses pathologies : maladies auto-immunes, cancers, transplantations, hépatites...[85]

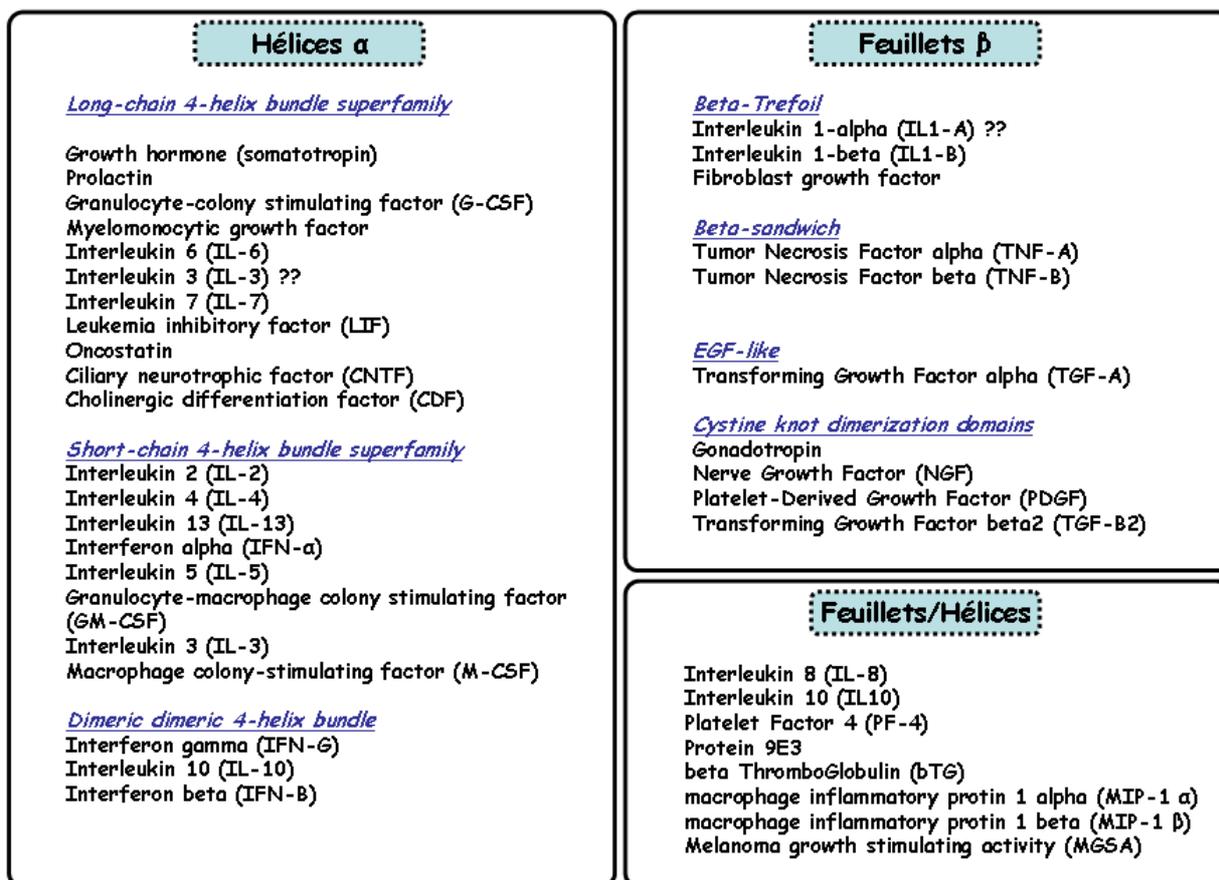


Figure I-17 : Classification des cytokines en fonction de leur repliement.

I.2.2-C Conclusion

Ces différents exemples démontrent la possibilité d'utiliser des protéines en vue d'applications thérapeutiques et/ou diagnostiques. Toutefois, les médicaments actuellement sur le marché (anticorps et cytokines majoritairement) sont utilisés en raison de leur fonction ou de leurs propriétés particulières intrinsèques. De nombreux cas existent, de composés organiques entièrement artificiels, dont le mode d'action est original et n'imité pas celui de composés naturels ou déjà connus. Aucun exemple similaire n'existe dans le cas de protéines. Le développement d'approches de conception de ligands protéiques, permettrait de faciliter l'obtention de nouvelles protéines, utilisées à des fins thérapeutiques et/ou diagnostiques.

I.3 Conception de ligands protéiques

Depuis une vingtaine d'années, le nombre d'exemples de conception de ligands de type mini-protéine n'a cessé d'augmenter. Plusieurs approches ont été proposées à cet égard, se différenciant, notamment, par la nature de la plateforme utilisée pour présenter les pharmacophores [86] ou par la méthode utilisée pour lui conférer sa fonction de liaison. Créer une nouvelle fonction peut être réalisé soit, par le transfert d'un site fonctionnel identifié, soit par des techniques d'évolution dirigée, réalisées de façon systématique [87] ou sur des positions déterminées. Les méthodes présentées dans cette partie sont classées en fonction de la plateforme structurale utilisée.

I.3.1 Minimisation de protéines

Le principe de la minimisation correspond à la réduction de la taille des protéines tout en conservant leur activité biologique initiale. Cette dernière étant supportée par la structure 3D des protéines, la minimisation permet de conserver le motif de liaison (pharmacophores) dans une topologie compatible avec l'interaction. Deux types d'approche sont possibles : les unes basées sur le maintien de la structure, les autres sur le maintien de la fonction [88]. D'après les auteurs, il est ainsi possible de réduire la taille de toute protéine naturelle jusqu'à l'obtention de mini-protéines, tout en conservant la structure portant le motif de liaison. Par exemple, le plus petit domaine protéique naturel possède au plus 50 à 60 résidus en absence d'éléments stabilisants [89]. L'approche par minimisation a démontré qu'il était possible de réduire la taille d'un domaine à doigt de zinc (repliement feuillet-feuillet-hélice) de moitié jusqu'à l'obtention d'une mini-protéine de 23 résidus [89], celle-ci conservant le même repliement que la protéine initiale et la capacité à fixer le zinc. La même stratégie a été utilisée pour la minimisation des domaines B et Z (59 résidus chacun) de la protéine A [90, 91]. Ces domaines lient, tous deux, le fragment constant Fc des immunoglobulines G. Le domaine initial comprend trois hélices α , dont seulement deux sont impliquées dans l'interaction. La minimisation du domaine Z a conduit à une mini-protéine de 38 résidus contenant seulement les deux hélices impliquées, comme schématisé sur la figure I-18. Cette dernière représente l'affinité du domaine Z pour le fragment Fc, mesurée par BIAcore, à chaque étape de la minimisation. La mini-protéine résultante a conservé la même affinité que le domaine initial. Un travail similaire a été réalisé sur le domaine B avec une affinité résultante diminuée d'un facteur 10.

Les applications sont nombreuses : développement d'inhibiteur [95], méthodes de diagnostic et d'imagerie, méthode de criblage [96]...Par ailleurs, il a été démontré la possibilité d'utiliser divers anticorps pour le développement de méthodes de diagnostic et d'imagerie [97, 98, 99] Ces exemples, réalisés sur des anticorps, de taille réduite (80 kDa, en comparaison des 150 kDa des IgG), laissent à penser que les « minibody » (dont la taille est de l'ordre de 6 kDa) pourraient être utilisés de façon similaire. De façon évidente, l'obtention de protéines de faible poids moléculaire est plus avantageuse, pour des questions de coût et de facilité de synthèse. A l'exception du cas des « minibody », pour lesquels l'aptitude à reconnaître des structures très variées est mise à profit pour lier toute cible d'intérêt, la minimisation de protéines existantes ne correspond pas, *stricto sensu*, à la conception de nouveaux ligands. Cette approche permet, toutefois, d'envisager l'utilisation de ligands protéiques de tailles importantes, en les rendant accessibles aux différentes techniques de synthèse et d'optimisation.

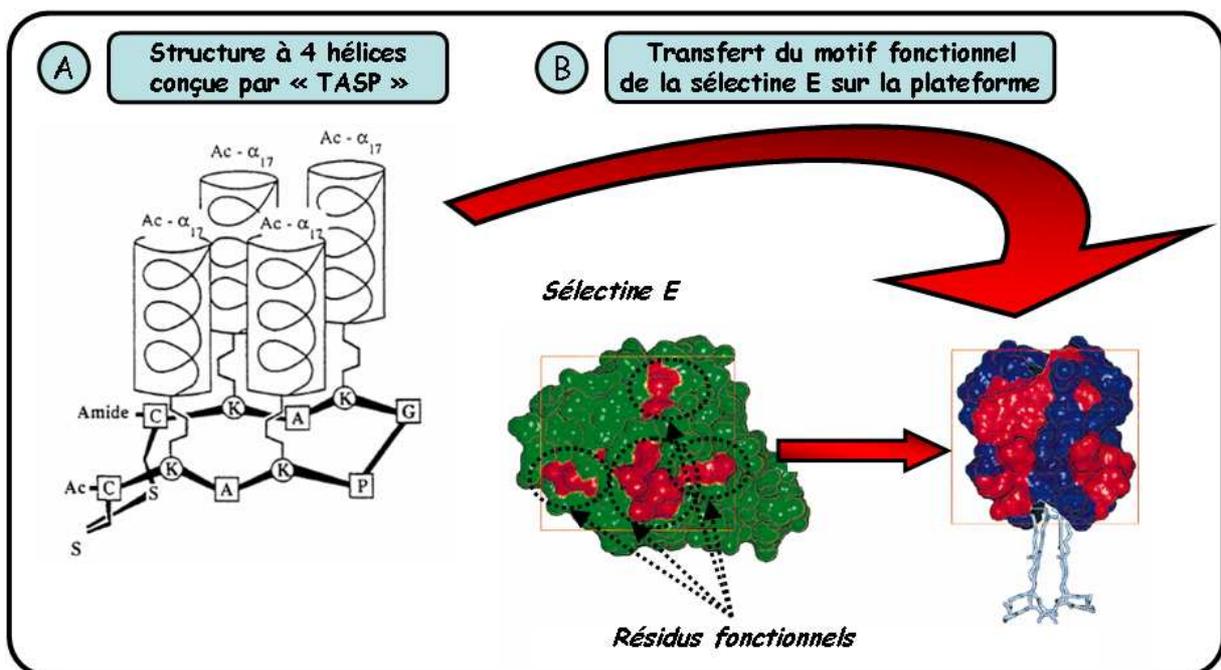


Figure I-21 : Conception de novo d'un mîme de la sélectine E. Les résidus fonctionnels de la sélectine E ont été transférés sur une structure à quatre hélices, conçue de novo. [103, 104]

I.3.2 Synthèse de plateformes *de novo*

La synthèse de plateformes *de novo* correspond à la création de structures protéiques totalement artificielles. Cette opération revient à créer et identifier des séquences de résidus capables de produire un repliement ou une structure secondaire précise. Toutefois, la prédiction de la structure d'une protéine à partir de sa simple séquence primaire demeure à ce jour un défi. La synthèse de plateformes *de novo* s'inspire ou imite des repliements naturels, dans le but de les simplifier.

Le premier concept proposé, de « protéine de synthèse assemblée sur matrice » ou TASP [100, 101], consiste à assembler divers éléments de structure secondaire sur une matrice, comme montré sur la figure I-20, permettant l'obtention des structures 3D voulues. La matrice peptidique permet de réaliser des liaisons spécifiques avec divers peptides de structure secondaire déterminée, afin de générer des structures complexes. Cette stratégie a, par exemple, été appliquée à la construction d'un complexe à 4 hélices, de 17 résidus chacune [102]. Cette structure, représentée figure I-21, A, a permis la

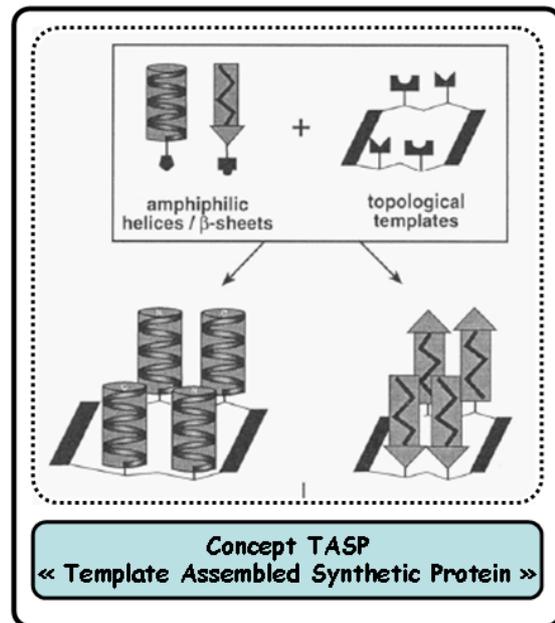


Figure I-20 : Approche de conception de protéine de synthèse assemblée sur matrice (TASP). [101]

conception d'un mime de la sélectine E [103], représenté sur la figure I-21, B. De façon similaire, un mime du CMH-I a été créé, du fait de caractéristiques structurales communes avec la structure à quatre hélices, conçue *de novo* [104]. Ce dernier a été identifié positivement par test ELISA, en utilisant des anticorps spécifiques de molécules du CMH-I. D'autres exemples ont été reportés, démontrant la capacité de cette approche à générer d'autres structures fréquemment observées dans de nombreuses protéines naturelles : hélice-boucle-hélice, doigt à zinc [103, 105].

Plus récemment, une approche combinatoire à la conception de plateformes *de novo* a été proposée. Considérant une protéine de 100 résidus, le nombre de combinaisons, à partir des 20 résidus naturels, correspond à 20^{100} possibilités [106]. Le nombre de solutions générés par cette approche combinatoire doit être limité par l'utilisation d'autres éléments. Les auteurs proposent deux caractéristiques générales des protéines naturelles : la présence d'éléments de structure secondaire abondants, et une distribution non homogène des résidus polaires et apolaires entre la

I.3.3 Utilisation de plateformes naturelles

Une autre alternative réside dans l'utilisation de plateformes naturelles, mettant ainsi à profit la diversité structurale existante dans la nature. Ces informations structurales sont disponibles, en partie, dans les bases de données de structures protéiques (annexe A.2). La conception de ligands protéiques peut être alors envisagée par le transfert d'un motif fonctionnel sur une plateforme naturelle appropriée, de structure connue.

Les premiers exemples de transfert de motifs correspondent à la construction de protéines « chimères », dans lesquelles des segments de séquence entiers sont échangés entre deux protéines appartenant à deux espèces différentes, mais dont le repliement local ou global est similaire. Par exemple, la boucle centrale de la toxine α de serpent a été transférée sur la charybdotoxine, certes, de taille inférieure, mais présentant un repliement en épingle à cheveux identique [111, 112]. La protéine chimère ainsi créée est illustrée sur la figure I-23.

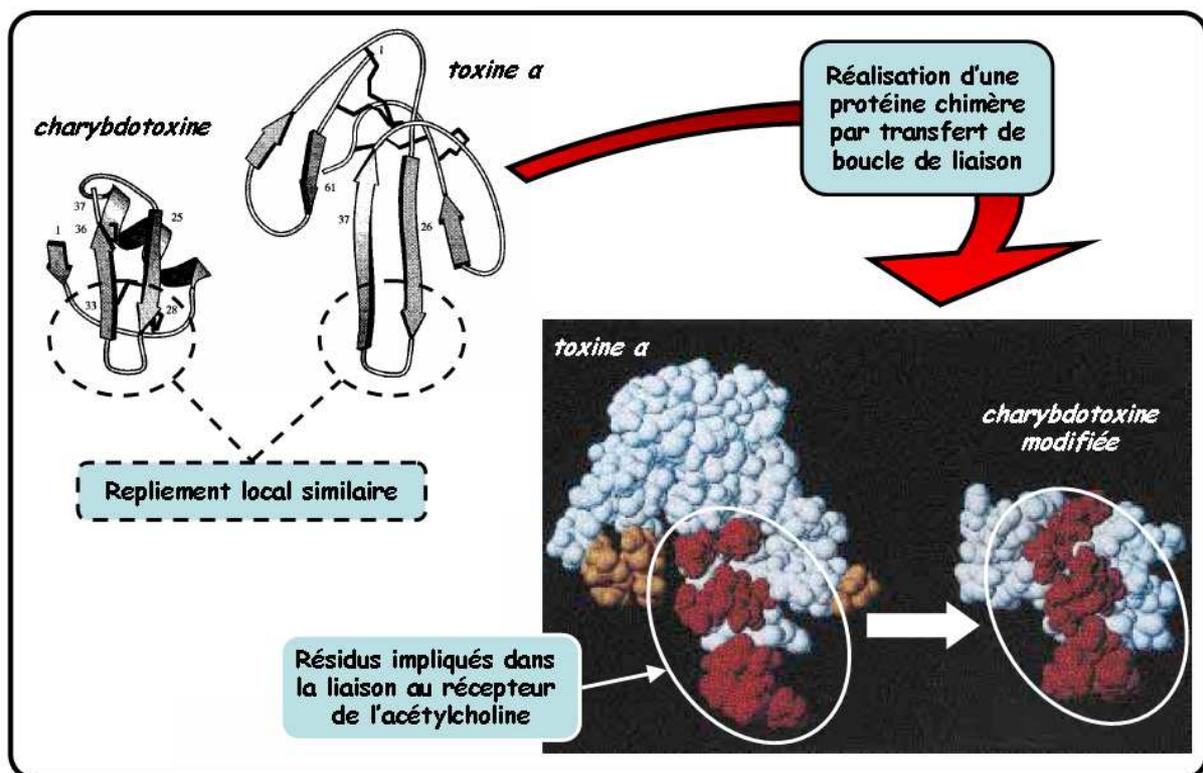


Figure I-23 : Création d'une protéine chimère, capable de lier le récepteur à acétylcholine, après le transfert de la boucle de liaison de la toxine α [111].

La protéine chimère obtenue et la toxine α de serpent lient, toutes deux, le récepteur à acétylcholine avec des affinités respectives de l'ordre de 10^{-5} M et 10^{-9} M. Le transfert de 9 résidus sur une plateforme structurale similaire, a permis de transférer une partie de l'activité biologique.

Le même type d'approche a été réalisé par le transfert d'éléments de structure secondaire sur une plateforme dont le repliement global est différent. La protéine PF4 est constituée de 5 brins β et de 2 hélices α . Les déterminants de l'interaction ne sont portés que par les deux hélices α distantes en séquence. Or, la structure « leucine zipper » de la protéine GCN4, formée de deux hélices uniquement, permet de disposer le motif de liaison dans une topologie similaire. Ainsi, la capacité de lier l'héparine de la protéine PF4 a-t-elle été transférée sur la protéine GCN4 [113], illustrée sur la figure I-24. L'activité de liaison, mesurée par la rétention sur une colonne d'héparine-agarose, a montré un comportement similaire entre la protéine PF4 et la protéine GCN4 modifiée. La même protéine GCN4, a été utilisée pour porter le motif de liaison de l'interleukine 4 pour son récepteur [114], également illustré sur la figure I-23. Les affinités mesurées varient de 2 mM à 5 μ M selon la fraction du site fonctionnel transféré.

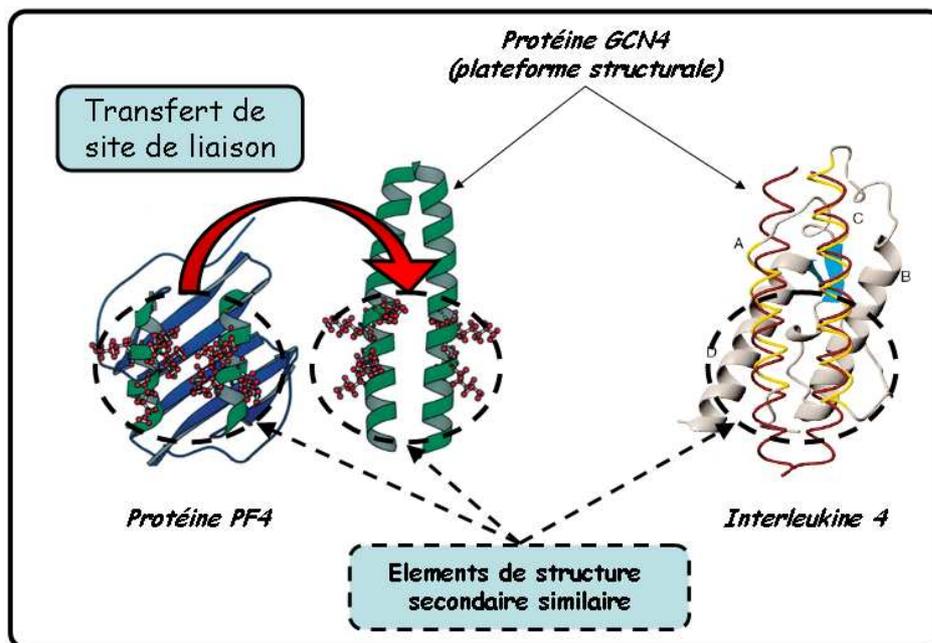


Figure I-24 : Conception de mimes protéiques des protéines PF4 [113] et interleukine 4 [114], en utilisant la même plateforme structurale : la protéine GCN4. Les protéines PF4 et IL-4 présentent des repliements différents, toutefois, les résidus impliqués dans la liaison sont portés par des éléments de structure secondaire identiques.

Ces deux travaux démontrent qu'un même repliement est capable de supporter deux activités biologiques, différentes selon le motif de liaison transféré.

De façon similaire, plusieurs travaux ont été réalisés dans le but de concevoir des protéines liant l'ADN ou diverses surfaces protéiques [115]. L'approche proposée de greffe sur protéine (« protein grafting ») comprend, dans un premier temps, le transfert d'un épitope fonctionnel, présent sur des éléments de structure secondaire type hélice α ou hélice polyPro type II, sur la structure du polypeptide pancréatique aviaire aPP. Cette mini-protéine de 36 résidus, constituée d'une hélice polyPro type II, d'un coude β de type I et d'une hélice α est, malgré sa faible taille, très bien structurée [1PPT, 116]. Dans un second temps, les propriétés de liaisons des protéines

ainsi greffées sont optimisées par la réalisation de banques de phages. La mise en application de cette approche, schématisée figure I-25, a permis la conception de plusieurs ligands d'affinité comprise entre μM et nM , après optimisation : ligands de protéines impliquées dans les processus de cancérogenèse, hDM2 [115] et Bcl2 [117], ligands de sites spécifiques de l'ADN, site QRE [118] et site CRE [119, 120, 121].

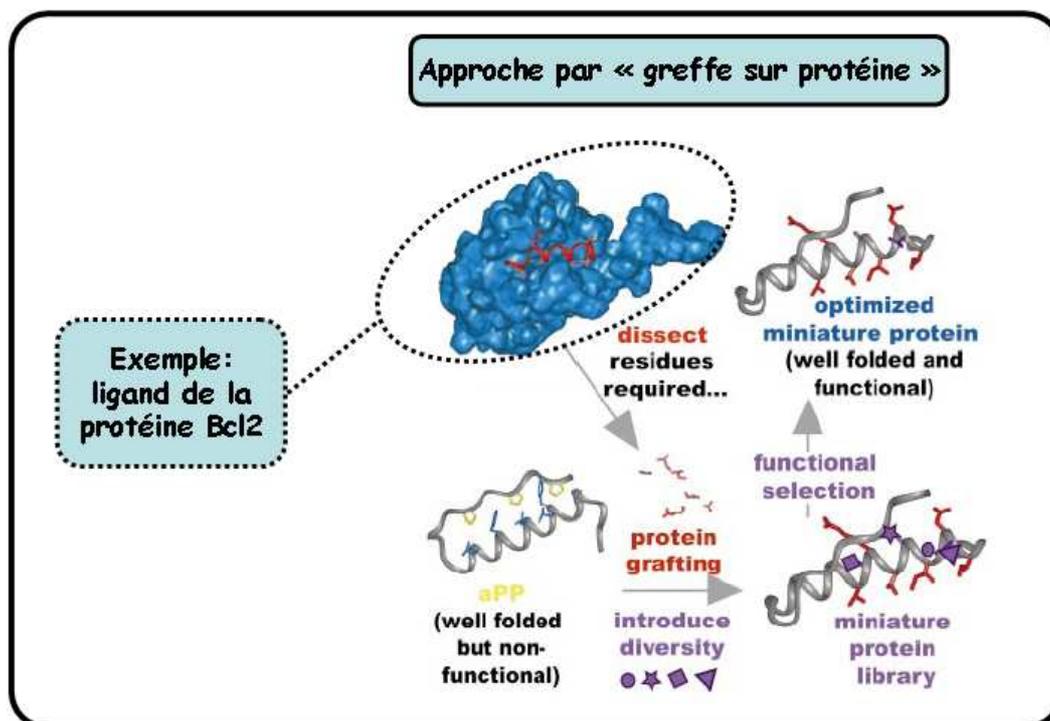


Figure I-25 : Approche de conception par greffe sur protéine. L'exemple montré correspond à la conception d'un ligand de la protéine Bcl2. [117].

Plusieurs exemples, décrits dans la littérature, illustrent le transfert de sites de liaison sur d'autres éléments de structure secondaire divers, boucles [122] ou feuillets β [123, 124, 125, 126]. Ces derniers travaux notamment, concernent l'inhibition de l'entrée du virus HIV-1 dans les cellules du système immunitaire présentant le co-récepteur CD4. Lorsque le virus est à l'état circulant, la protéine gp120 n'est pas exposée en surface et, de ce fait, ne constitue un déterminant antigénique. En présence de la protéine CD4, la protéine gp120 est démasquée. La stratégie envisagée repose sur l'utilisation d'un mime de la protéine CD4 afin de démasquer la protéine gp120 avant l'entrée du HIV-1 dans les cellules immunitaires. Les épitopes de la protéine gp120 ainsi exposées favoriseraient une réponse immunitaire. La protéine CD4 interagit avec la gp120 par l'intermédiaire d'un feuillet composé de deux brins β anti-parallèles, comme indiqué sur la figure I-26, A.

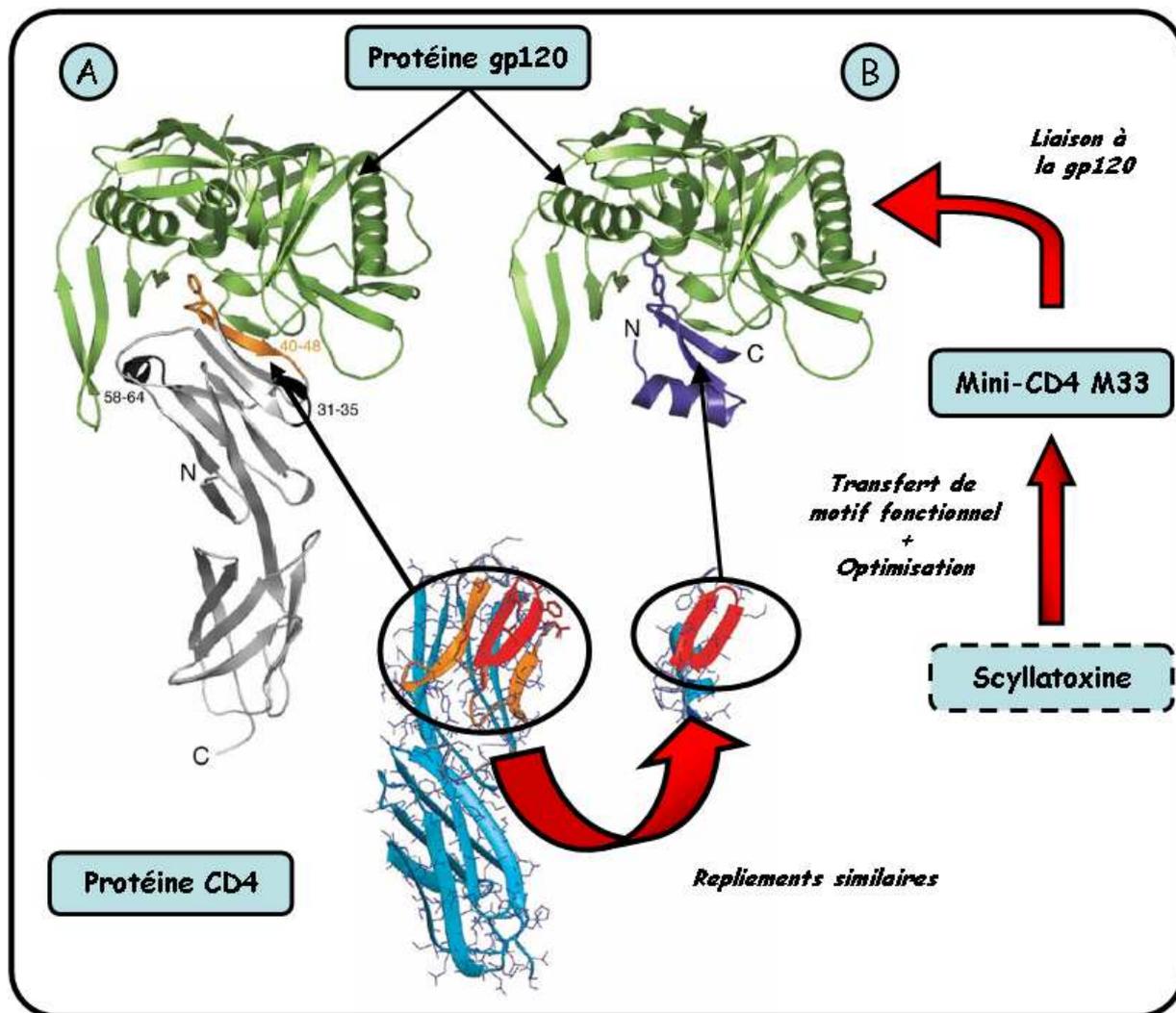


Figure I-26 : Conception de mini-CD4, ligand de la protéine gp120 à partir de la scyllatoxine. La protéine gp120 a été cristallisée en interaction avec la protéine CD4 et CD4M33, ces deux dernières présentant un mode de liaison similaire. [125, 127].

Un élément de structure secondaire identique a été identifié dans la scyllatoxine, petite protéine de 28 résidus dont la structure est stabilisée par 3 ponts disulfures. Les deux structures présentent une grande similarité (RMSD calculé sur la chaîne principale de 1,1 Å). Le transfert des résidus considérés comme important pour la liaison, a permis d'obtenir un mini-CD4, utilisé comme molécule guide [125]. A partir de cette dernière, plusieurs modifications ont été effectuées. L'un des premiers variants de la scyllatoxine, CD4M3, présentait une affinité de 40 μM . Après plusieurs étapes, l'optimisation de la molécule guide initiale a abouti à la protéine CD4M33, d'affinité 7,5 nM pour la gp120, soit une amélioration d'un facteur 5000. Les séquences des différents variants sont montrées sur le tableau I-27. En outre, la structure du complexe gp120-CD4M33 [127] résolue par radiocristallographie a permis de valider le mode de liaison supposé (figure I-26, B).

Protéine		Hélice $\alpha 1$		Brin $\beta 1$		Brin $\beta 2$	Kd
CDR2			QI	KILGN	QGS	FLTKGP	\approx nM
Scyllatoxine	AF	CNLRMCQLS	CRSLG	LLGKC	IGD	KCECVKH	
CD4M3	--	CNLARCQLS	CKSLG	LKGGC	QGS	FCTCG--	40 μ M
CD4M33	--	TpaNLHFCQLR	CKSLG	LLGKC	AGS	BipCACV(NH ₂)	7,5 nM

Tableau I-27 : Séquences des variants de mini-CD4 obtenus à partir de la scyllatoxine et d'un ligand endogène, le CDR2. Les résidus en bleu correspondent aux résidus transférés, du ligand endogène CDR2, sur la structure de la scyllatoxine. Tpa : acide thiopropionique ; Bip : biphénylalanine. [125, 126].

Ce travail, illustre la possibilité de transférer un motif fonctionnel sur une plateforme apte à supporter un motif de liaison, porté par un même élément de structure secondaire, et d'optimiser celle-ci afin d'obtenir un ligand protéique potentiellement utilisable pour une application thérapeutique.

I.3.4 Conclusion

Ces différents travaux démontrent, la capacité d'approches basées sur l'analyse structurale, à concevoir de nouveaux ligands protéiques. Toutefois, ces dernières, sont basées sur le transfert de motifs de liaison sur des éléments de structure secondaire identiques. Or, il est reporté dans la littérature, un exemple de motif fonctionnel porté par des éléments de structure secondaire différents. La topologie du motif RGDF du fibrinogène, présent dans une structure en coude β , est retrouvée identique dans la structure en hélice d'un décapeptide, identifié en tant que ligand, [128, 129], comme le montre la figure I-28. Bien que la superposition des deux éléments de

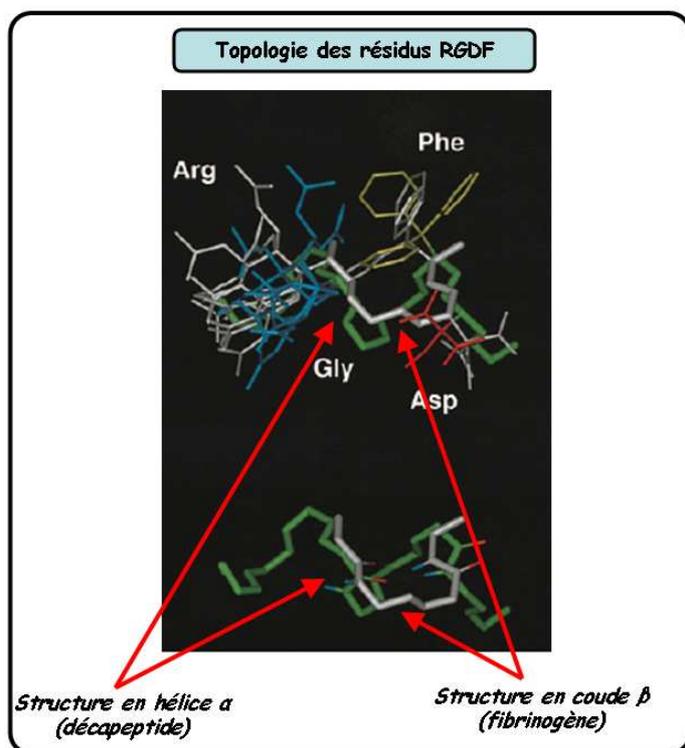


Figure I-28 : Superposition des motifs RGDF (du fibrinogène, en gris) et RGXFD (d'un décapeptide, en vert) via les atomes $C\alpha$, $C\beta$, $C\gamma$ et $C\delta$ du résidu Arg, $C\alpha$, $C\beta$ et $C\gamma$ des résidus Asp et Phe. [128].

structure soit différente, la topologie des résidus R, G, D et F est similaire dans les deux cas. La structure secondaire n'est donc pas essentielle pour présenter un motif de liaison donné : seule la disposition spatiale des chaînes latérales impliquées est déterminante. Toutes les approches de conception de ligands de nature protéique, décrites dans les paragraphes précédents, reposent sur la considération de la seule structure secondaire pour l'identification de plateformes structurales aptes à recueillir un motif de liaison donnée, et ceci quelle que soit l'origine de la plateforme (minimisation, synthèse *de novo*, plateforme naturelle). Nous avons choisi d'orienter le travail réalisé au cours de cette thèse vers la conception de ligands protéiques, à partir de structures 3D de protéines présentes dans la PDB, indépendamment de la structure primaire, secondaire et tertiaire, des plateformes utilisées pour le transfert de motifs fonctionnels.

Chapitre II Analyse des interactions protéine-protéine

« L'important c'est de savoir ce qu'il faut observer » (Edgar Allan Poe)

A l'heure actuelle, malgré le nombre très important de travaux concernant les interactions entre protéines, la prédiction des sites d'interaction ou de l'énergie de liaison demeure un « challenge ». La formation d'un complexe se caractérise par la mise en contact de deux protéines, définissant une région particulière appelée interface, correspondant à l'ensemble des atomes dont l'accessibilité au solvant est modifiée lors de la formation de ce complexe.

L'amélioration et le développement des techniques de résolution de structure ont permis d'obtenir un grand nombre de structures de complexes protéiques. La quantité et la diversité de ces complexes rendent possible l'analyse, la comparaison et la quantification de différents paramètres caractérisant les interfaces protéiques.

La première étude concernant ces interactions a été réalisée sur trois complexes seulement [130], puis sur dix-neuf, quinze ans après [131] ; désormais, plusieurs centaines peuvent être pris en compte, dans le but de répondre à la question : les interfaces protéiques ont-elles des caractéristiques communes ?

Le chapitre suivant présentera ces différents travaux et la description de l'interface telle qu'elle est envisagée actuellement. La mise en application d'une méthode de conception de ligands protéiques nécessite la prise en compte des caractéristiques générales des interactions protéine-protéine : critères généraux, interactions ponctuelles et aspects électrostatiques.

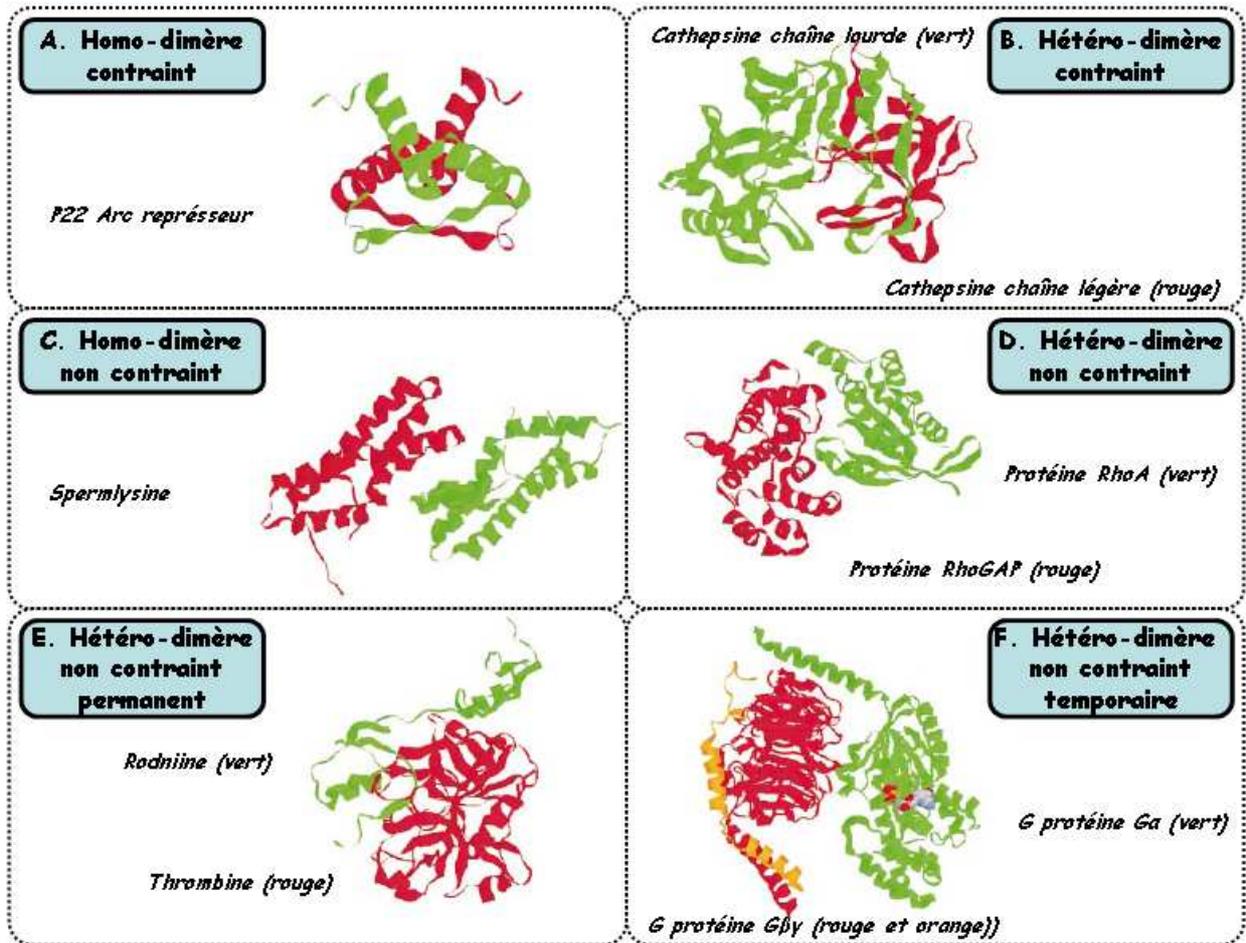


Figure II-1 : Représentation de différents types de complexes selon leurs natures homo ou hétéro dimérique, contrainte ou non, permanente ou temporaire [134]. Seuls les complexes hétéro-dimériques seront traités par la suite.

II.1 Description générale des interfaces

II.1.1 Un ou plusieurs types de complexes ?

Les analyses de complexes cristallographiques ont souvent été réalisées indifféremment de la nature du complexe considéré. Les premières comparaisons différenciant la nature des complexes reposaient, surtout, sur la fonction des protéines impliquées. Plusieurs travaux présentent, notamment, la comparaison de la structure des sites de reconnaissance des complexes protéase-inhibiteur et antigène anticorps [131, 132]. La considération, que les interfaces protéiques puissent différer selon le type de complexe considéré, s'est étendu, par la suite, sur des critères plus généraux que leur seule fonction [133, 134]. Les complexes peuvent être distingués de diverses manières :

- concerner deux protéines identiques ou pas : **hétéro-dimère** ou **homo-dimère** ;
- être de nature **transitoire** (e.g. processus biologique) ou **permanente** (e.g. inhibition) ;
- être de nature « **contrainte** » (obligomère,*) ou « **non-contrainte** ».

Plusieurs exemples de structures de complexes, ainsi distinguées, sont représentées figure II-1. Par la suite, plusieurs cas ne seront pas étudiés : les interfaces concernant les domaines protéiques [133], correspondant plutôt au phénomène de repliement ; les obligomères, dont l'étude nécessite la prise en compte d'aspects dynamiques pour la compréhension des modifications conformationnelles associées à formation du complexe ; les homo-dimères, qui, de façon évidente, ne peuvent s'inscrire dans un objectif de conception de ligand protéique.

* Obligomère

Protéine instable à l'état libre, nécessitant la formation d'un complexe pour acquérir un repliement stable. Ces protéines sont la plupart du temps co-exprimées et co-localisées dans les cellules

II.1.2 Caractéristiques des interfaces protéine-protéine

Le but de l'analyse de complexes protéiques est d'identifier différents critères, caractérisant l'interface, liés à formation de ces mêmes complexes. Toute modification observée, entre les états libre et lié de deux protéines, peut être directement reliée à la formation d'un complexe et non à diverses modifications conformationnelles, à la seule condition que leurs structures respectives ne soient pas modifiées. Ces protéines doivent être alors assimilées à des corps rigides. La comparaison de structures 3D de protéines, entre l'état libre et lié, a été réalisée via le calcul du RMSD sur les atomes Ca ou les chaînes latérales : de faibles variations entre ces deux états ont ainsi été observées [135]. Une partie de ces résultats est présentée dans la figure II-2.

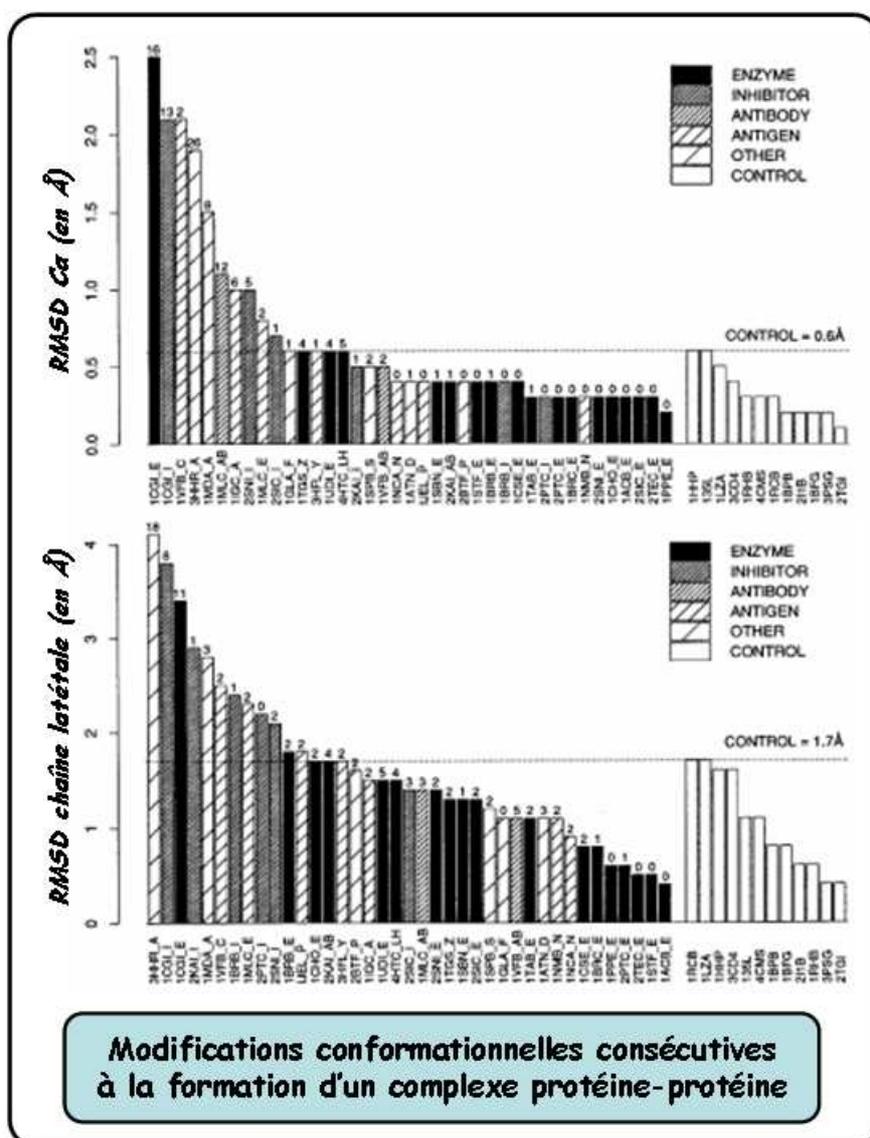


Figure II-2: Valeurs de RMSD calculées entre les structures, à l'état libre ou lié, de complexes cristallographiques enzyme-inhibiteur, antigène-anticorps ou autres, indiqués par leur code PDB et la structure considérée (exemple : 1CGI_E). Les chiffres (au-dessus de chacune des barres) indiquent pour chaque structure le nombre d'atome Ca ou de chaînes latérales dépassant la valeur du seuil. Plus de 75% des structures ont des valeurs de RMSD inférieures à la valeur seuil [135].

Dans plus de 75% des cas étudiés, aucune différence significative n'a été observée. Dans moins de 25% des cas, les modifications, souvent de faible amplitude, correspondent essentiellement à de simples mouvements de chaînes latérales [135]. De plus, ces observations ont été réalisées sur des protéines de taille comprise entre 36 et 489 résidus et impliquées dans divers types de complexe : elles sont, de ce fait, très générales. En revanche, des modifications importantes de conformation ont été observées lorsque la taille de l'interface dépassait 2000 \AA^2 [52]. Au-delà de cette taille, il n'est alors plus possible d'assimiler, de manière systématique, les protéines à des structures rigides.

II.1.2-A Taille de l'interface

La taille de l'interface est mesurée par la variation de surface accessible au solvant entre les états libre et lié des deux partenaires [136,137], schématisé figure II-3.

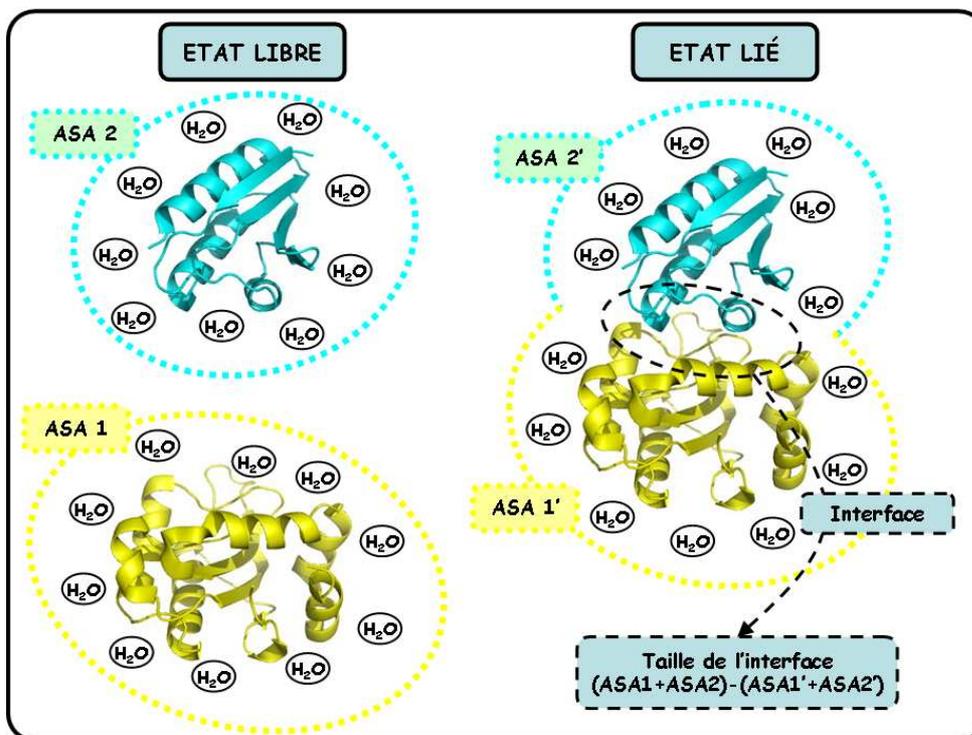


Figure II-3 : Calcul de la taille de l'interface lors de la formation d'un complexe. Les protéines en interaction (en bleu et jaune) sont représentées en mode « cartoon ».

Les travaux, réalisés jusqu'à présent, ont permis de définir une taille relativement homogène des interfaces, dans 70% des complexes considérés, estimée à $1600 \pm 400 \text{ \AA}^2$ (figure II-4, A) [52, 131]. Cette dernière s'applique à divers hétéro-dimères, et ceci malgré des variations de taille importantes, de 2000 à 40000Da (figure II-4, B). Par ailleurs, l'analyse de la structure des interfaces montre que, lorsque la surface calculée est supérieure à cette valeur moyenne, l'interface met en jeu plusieurs régions d'interaction distinctes ou « patch » [138] comme indiqué sur la figure II-5. La partie A montre une représentation visuelle de ces « patch ». La partie B

représente la variation de la taille de l'interface en fonction de leur nombre. Il apparaît que cette valeur, dite standard, inférieure à 2000\AA^2 , ne soit valable que pour une seule et unique région d'interaction.

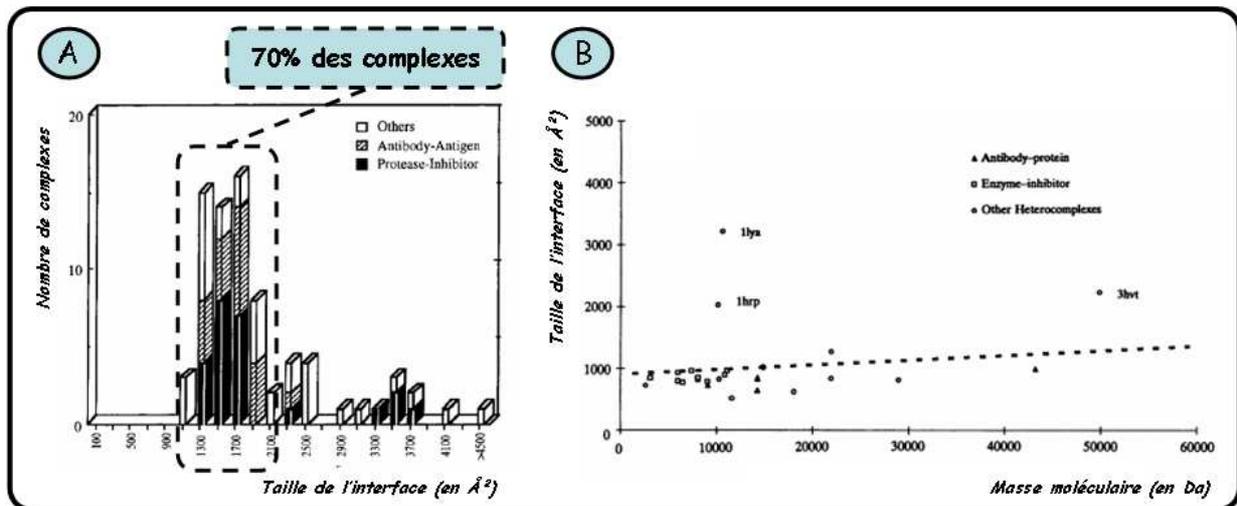


Figure II-4 : La partie A montre la répartition des valeurs de la taille de l'interface de divers complexes, en fonction de leur nature [52]. La partie B, représente la variation taille de la surface (pour chacune de ses faces) en fonction de la taille des protéines impliquées [139].

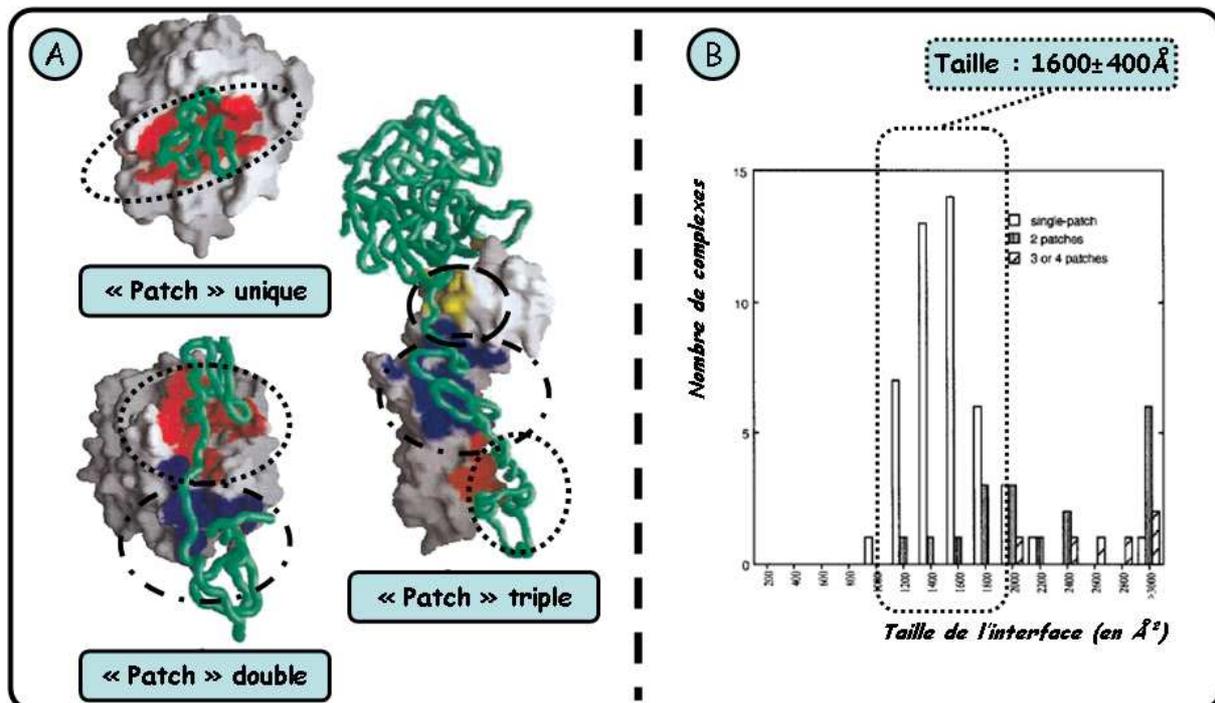


Figure II-5 : La partie A est une représentation visuelle des régions d'interactions, en rouge, bleu et jaune à la surface des protéines. Complexes : trypsine/PTI ; thrombine/ornithodorine ; facteur tissulaire soluble/facteur VIIA. La partie B montre la répartition des complexes, selon leur nombre de « patch » en fonction de l'ASA mesurée. [138].

II.1.2-B Compacité de l'interface

L'analyse de complexes montre que deux protéines en interaction contribuent de façon quasi équivalente à la taille de l'interface. Chaque partenaire participe, en moyenne, à hauteur de

50±3% de la surface totale [52]. L'aspect essentiel réside surtout dans l'importante complémentarité de forme des surfaces en contact [140]. Cette complémentarité se traduit par une forte compacité de l'interface, qui peut être estimée par différentes méthodes : méthode du polyèdre de Voronoï [141], mesure du « gap index » [139, 142], ou méthode statistique basée sur la détermination d'un coefficient de corrélation de forme Sc [140]. Ces trois méthodes sont décrites plus amplement par les figures II-6, II-6bis et II-6ter. Les résultats obtenus par la méthode de Voronoï montrent une compacité peu inférieure ($\approx 4\%$) à celle des résidus présents au cœur des protéines, elle-même peu inférieure ($\approx 5\%$) à la compacité d'acides aminés sous forme cristalline [52]. Les interfaces sont donc des structures très compactes. Le « gap index », qui montre une faible dépendance en fonction du type de complexe considéré [139], sera de ce fait utilisé pour caractériser divers complexes cristallographiques dans les chapitres IV et V.

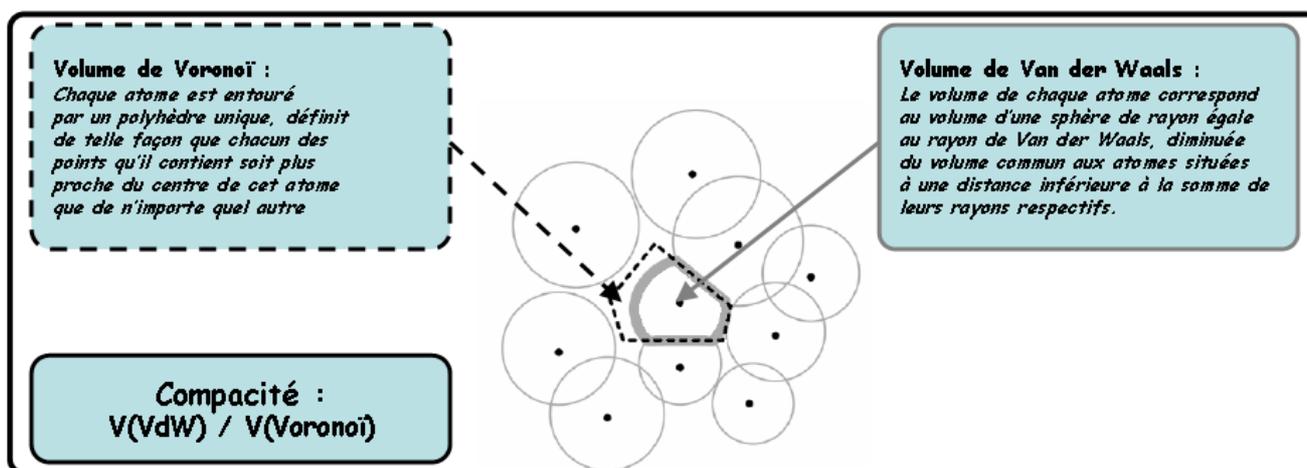


Figure II-6 : Mesure de la compacité des interfaces par la méthode du polyèdre de Voronoï [141].

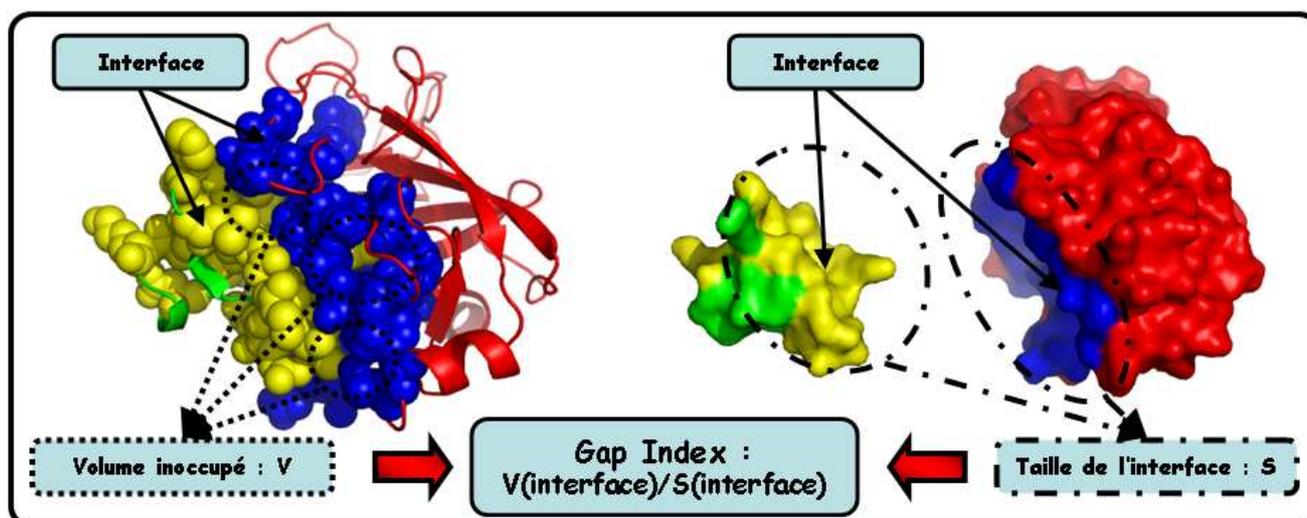


Figure II-6bis : Le calcul du « gap index » est défini par le ratio Surface (à l'interface)/Volume (à l'interface). Le complexe montre l'interaction entre le BPTI (jaune et vert) et la trypsine bovine (bleu et rouge) [143]. Les résidus de l'interface (jaune et bleu) sont représentés en mode « sphere » (gauche) et « surface » (droite); les résidus de surface (vert et rouge) en mode « cartoon » (gauche) et « surface » (droite).

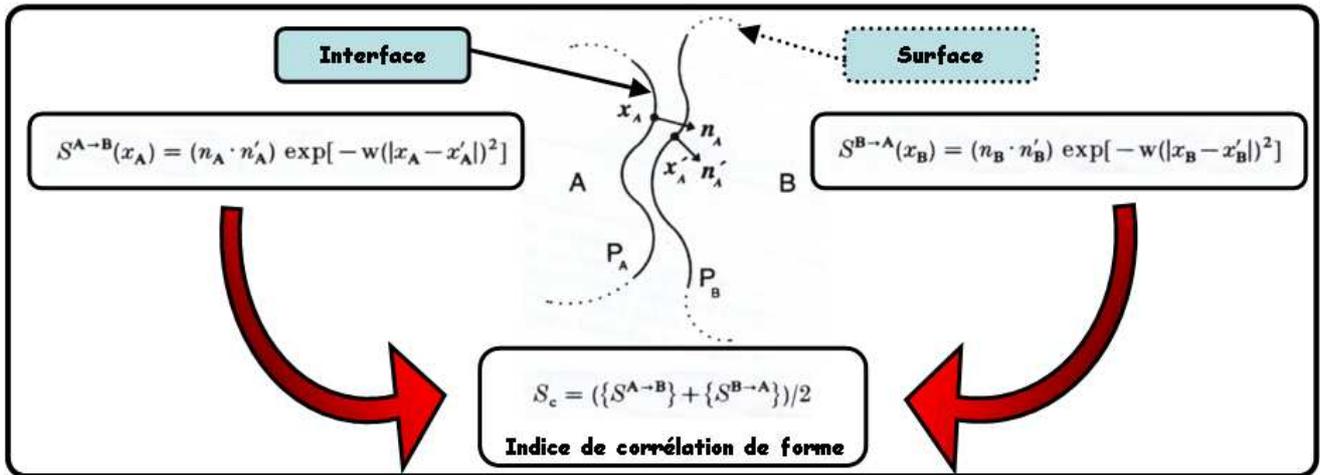


Figure II-6ter : Estimation de la compacité par analyse statistique [140].

II.1.2-C Structure de l'interface

Chaque face de l'interface possède une forme globalement ovoïde [52, 139]. Les atomes qui la composent peuvent être distingués en fonction de la variation de surface accessible, entre état libre et lié, et en fonction de la distance minimum par rapport aux atomes de la protéine partenaire. Trois classes d'atomes distincts sont alors identifiées, et sont schématisées sur les figures II-7 et II-7bis : ceux dont l'ASA est non-nulle et qui établissent (zone B) ou n'établissent pas (zone C) de contact de Van der Waals, et ceux dont l'ASA est nulle et qui établissent des contacts de Van der Waals (zone A) :

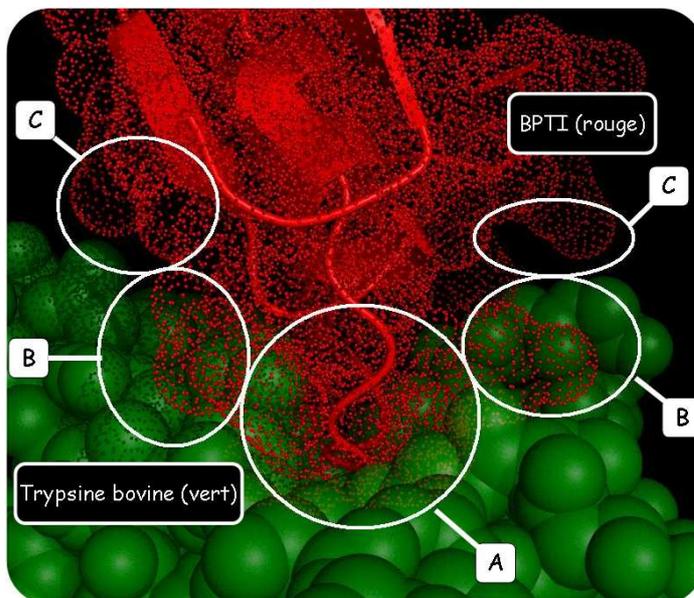
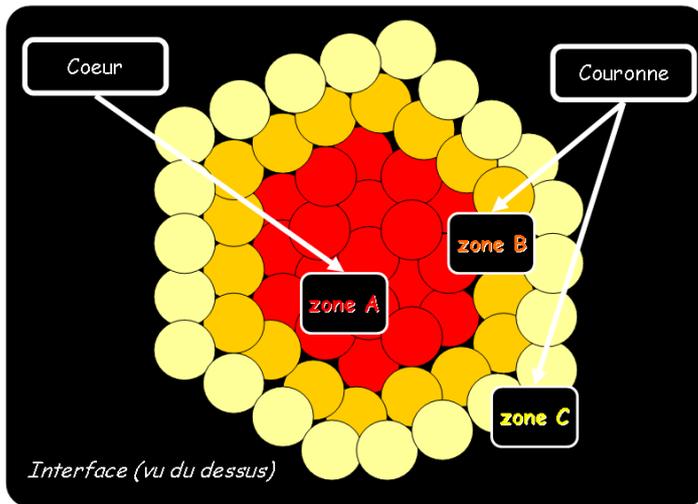


Figure II-7 : Représentation de l'interface entre la trypsine bovine et le BPTI. Le BPTI (en rouge) est représenté en mode « dot » et « cartoon ». La trypsine bovine (en vert) est représentée en mode « sphere ».

Les résidus de classe A forment le cœur de l'interface tandis que les résidus de classe B et C forment une couronne, entourant le cœur. En moyenne, pour des interfaces de taille standard, la moitié des atomes sont des atomes en contact (A et B), et un tiers seulement sont enfouis (A), totalement isolés des molécules de solvant [52]. Il est proposé que les résidus appartenant à la couronne (classe B notamment) protègent le cœur vis-à-vis du solvant, favorisant ainsi la réalisation d'interactions



ponctuelles énergétiquement
favorables [144].

Figure II-7bis : Représentation schématique de l'interface. Les résidus de classe A forment le cœur de l'interface. Ceux de classe B et C forment la couronne. [52]

II.1.3 Composition des interfaces

La composition des interfaces peut être analysée de diverses façons, selon son caractère physico-chimique, les types de résidu présents ou conservés au niveau de celle-ci.

II.1.3-A Composition chimique de l'interface

La composition chimique globale de l'interface est très variable d'un type de complexe à l'autre, et même d'un complexe à l'autre [52, 131], comme indiqué figure II-8.

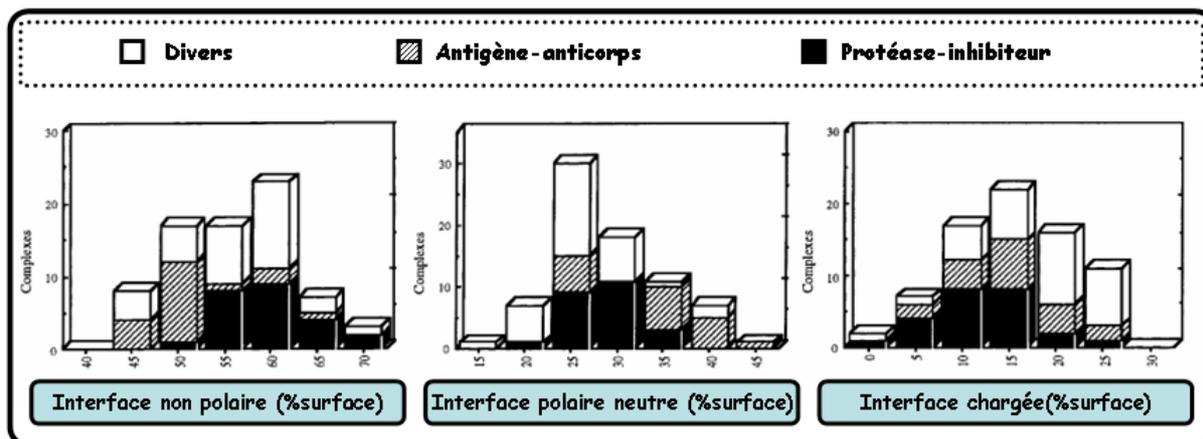


Figure II-8 : Représentation en histogramme de la composition chimique de l'interface, en % de surface, pour différents types de complexes : antigène-anticorps, enzyme-inhibiteur ou autres. Les fonctions chimiques apolaires, polaires neutres et chargées sont différenciées [52].

La surface de l'interface est majoritairement apolaire (en moyenne, $\approx 55\%$, de 45 à 70%) et faiblement chargée (en moyenne, $\approx 15\%$, de 5 à 25%). En outre, des disparités sont observées concernant les complexes antigène-anticorps, pour lesquels la fraction de résidus chargés est augmentée [52]. De part une variabilité importante, ce critère ne permet pas de caractériser, d'un point de vue général, l'interface formée par deux protéines en contact, quelle que soit leur nature.

II.1.3-B Composition en résidus

Plusieurs méthodes ont été proposées pour estimer la composition en résidus de l'interface. Deux d'entre elles sont fréquemment employées, chacune ayant pour but d'évaluer la propension de chaque type de résidu à être présent à l'interface.

Comparaison « composition interface » / « composition surface ». La première méthode est basée sur le calcul de la surface accessible. Pour chaque résidu, la propension est calculée par le logarithme népérien du ratio de la contribution de chaque résidu à l'ASA totale de l'interface sur

la contribution du même résidu à l'ASA de la surface [52, 139, 138]. Les résultats obtenus pour ces différents travaux sont très homogènes, comme le montre le tableau II-9.

	[52]	[139]	[138]	[52]			
	Hétéro-complexes	Hétéro-complexes	Cœur / Couronne	Protéase-inhibiteur	Antigène-anticorps	Complexes divers	
ALA	-	-	- / -	--	--	-	ALA
ARG	+	+	+ / +	+	+	±	ARG
ASN	-	+	- / ±	--	+	±	ASN
ASP	-	-	- / ±	--	+	-	ASP
CYS	++	+	+++ / ++	+++	<i>n.a.</i>	++	CYS
GLN	-	-	- / -	-	-	-	GLN
GLU	-	-	-- / ±	-	--	--	GLU
GLY	±	±	± / +	+	+	-	GLY
HIS	++	+	++ / +	++	0	++	HIS
ILE	++	+	++ / +	++	+	+	ILE
LEU	+	+	+ / +	+	-	+	LEU
LYS	--	-	-- / -	--	--	-	LYS
MET	++	++	+++ / ++	+++	+	++	MET
PHE	++	++	+++ / --	++	++	++	PHE
PRO	-	-	- / -	±	--	-	PRO
SER	-	-	-- / -	-	--	-	SER
THR	-	-	- / -	-	-	-	THR
TRP	++	++	+++ / +	+++	+++	+++	TRP
TYR	++	++	+++ / ++	+	+++	++	TYR
VAL	±	+	± / +	±	--	+	VAL

Tableau II-9 : Tableau récapitulatif des propensions, pour chaque résidu, à être préférentiellement à l'interface. Cette propension est calculée par le logarithme népérien du rapport des valeurs de surface accessible au solvant entre l'interface et la surface des protéines. Les valeurs de propension calculées sont simplifiées par un signe en fonction de leur valeur : pour $p < -1 \rightarrow ---$; de $-1 < p < -0,5 \rightarrow --$; de $-0,5 < p < -0,1 \rightarrow -$; de $-0,1 < p < 0,1 \rightarrow \pm$; de $0,1 < p < 0,5 \rightarrow +$; de $0,5 < p < 1 \rightarrow ++$; et pour $p > 1 \rightarrow +++$. Les variations notables sont colorées en orange, et en rouge lorsque celles-ci sont importantes.

L'analyse des complexes, indifféremment de leur nature (colonnes 2, 3, et 4), montre peu de différences notables. Les différences observées sont notamment dues à la distinction du cœur et de la couronne de l'interface [138] et ne concernent qu'un faible nombre de résidus : Asp, Glu, Gly et surtout Phe. En revanche, la prise en considération du type de complexe, antigène-anticorps, enzyme-inhibiteur ou complexes divers (colonnes 5, 6 et 7 respectivement), donne des résultats plus divergents concernant les résidus Asn, Asp, Gly, His, Leu, Pro et Val dans une large mesure, ainsi que les résidus Arg, Met et Tyr dans une moindre mesure [52].

Comparaison composition interface/composition globale. La seconde méthode rapporte la fréquence (f) de chaque résidu à l'interface à l'abondance moyenne du même résidu, calculée à

partir des protéines contenues dans la Swiss-Prot [133, 145, 146]. La propension par résidu est estimée par l'équation (1) dans le premier travail, et par l'équation (2) dans le second.

(1) $P = \log [f(\text{interface}) / f(\text{Swiss-Prot})]$

(2) $P' = [f(\text{interface}) / f(\text{Swiss-Prot})] - \langle f \rangle$ (fréquence moyenne par résidus, égale à 5%)

Contrairement aux travaux précédents, les résultats sont inhomogènes voir opposés, comme en témoignent les figures II-10 et II-11. Ces différences sont importantes pour les résidus Ala, Arg, Asn, Asp, Gln, Gly, His, Phe, Ser, Thr, Tyr et Trp, d'importance moyenne pour les résidus Ile, Met, Pro et Val. Seuls quatre résidus Cys, Glu, Leu et Lys ne montrent pas de différences de propension dans les deux études. Bien que le principe du calcul soit le même dans ces deux travaux, la méthode de calcul diffère. Il n'est de ce fait pas possible de comparer les valeurs absolues des propensions par résidu observées. Cependant, leur signe ne devrait pas différer. Bien qu'une partie des variations observées puisse être expliquée par la méthode calcul, elle ne peut, à elle, seule rendre compte de l'ensemble des différences. La prise en compte globale de divers complexes, indifféremment de leur type [145], pourrait être à l'origine de ces résultats contradictoires, d'autant plus que les résultats du travail effectué par Ofran et Rost sont concordants avec les résultats précédemment décrits [52, 139, 138].

L'ensemble de ces travaux démontre néanmoins que la composition moyenne de l'interface diffère du reste de la surface [138], schématisée sur la figure II-12.

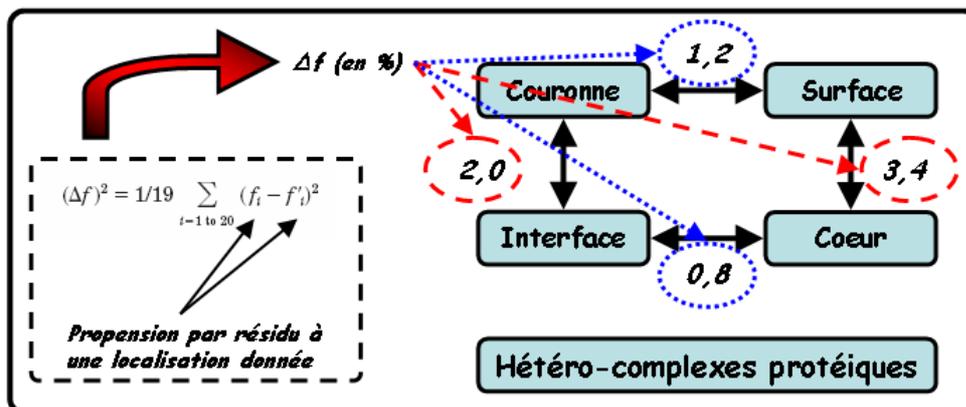


Figure II-12 : Différences de composition selon la localisation des résidus. La composition globale du cœur diffère nettement de celle de la surface des protéines. [138].

La caractérisation de l'interface par la composition en résidus, s'avère néanmoins difficile car le type de complexe et la localisation précise des résidus (cœur ou couronne) sont des paramètres à prendre en compte, comme les résultats, présentés dans le tableau II-9, le démontre.

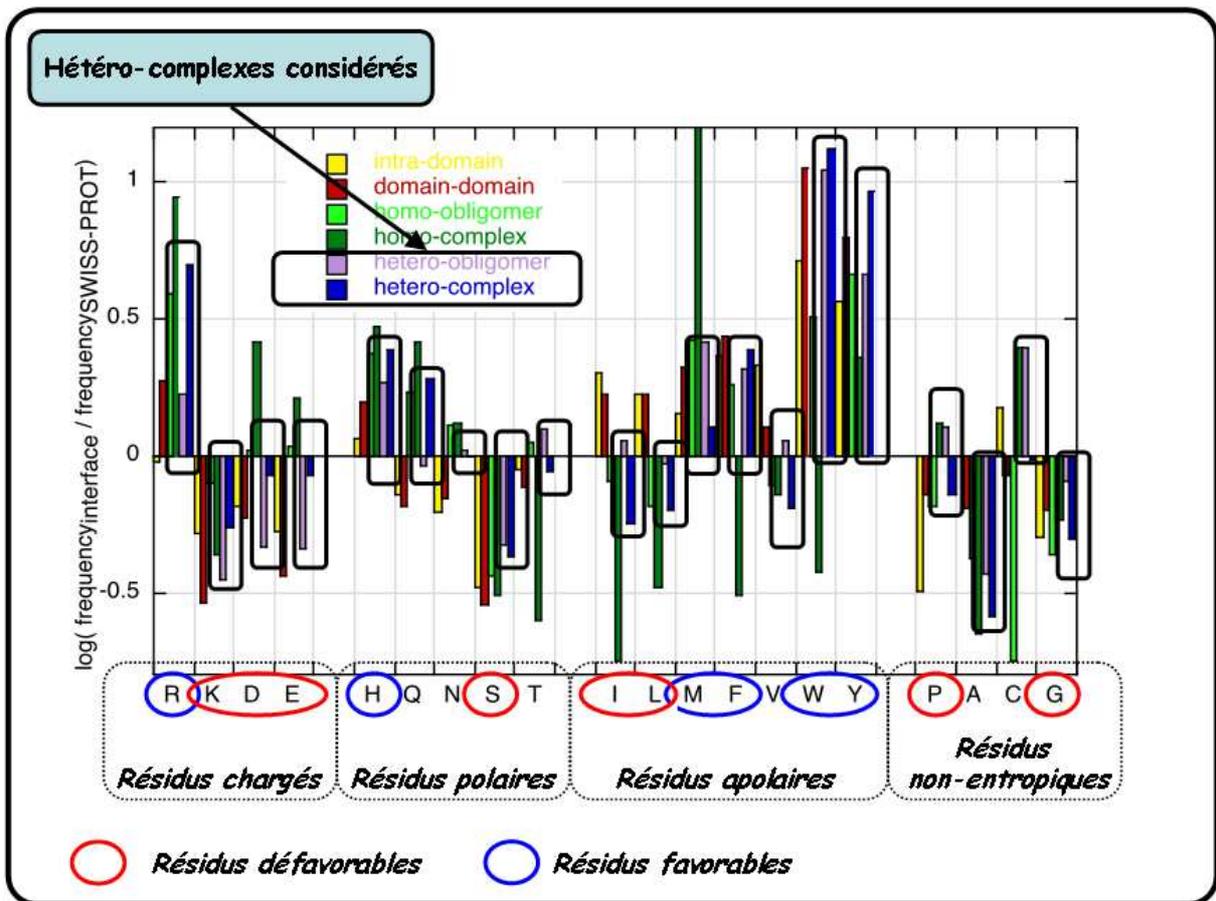


Figure II-10 : Propension des résidus à l'interface, selon le type d'interface considéré. La fréquence de chaque type de résidu à l'interface (équation 1) est rapportée à la fréquence de celui-ci mesurée à partir de la Swiss-Prot [146]. La figure est extraite de l'article [133].

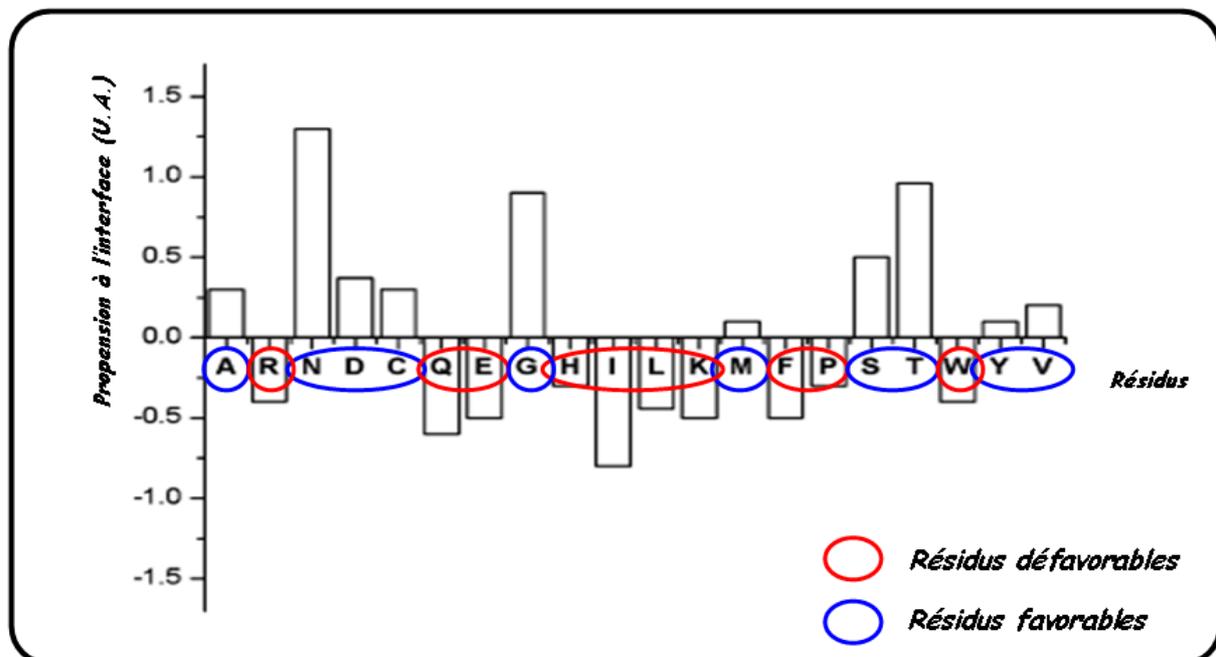


Figure II-11 : Propension des résidus à l'interface. La fréquence de chaque type de résidu à l'interface (équation 2) est rapportée à la fréquence de celui-ci mesurée à partir de la Swiss-Prot [146]. La figure est extraite de l'article [145].

II.1.3-C Conservation des résidus

La conservation de résidus, au niveau de sites actifs, permet de définir de nombreuses familles de protéines, partageant une même fonction [147, 148, 149]. Le maintien d'une fonction semble lié à un ensemble restreint de résidus. Des travaux similaires ont été réalisés, afin d'estimer le degré de conservation de résidus au niveau de sites de liaisons, dans plusieurs familles de protéines identifiées par alignement de leurs séquences primaires [150, 151, 152, 153]. Il a été mis en évidence que ce critère de conservation permettait de distinguer les contacts protéiques, biologiquement relevant, des contacts cristallins, introduits par la méthode de résolution de structure [150]. D'autre part, il semble que les résidus conservés soient regroupés dans l'espace. En revanche, il n'y a pas d'indices quant à leur localisation précise : certains sont proches des sites de liaison, tandis que d'autres participent à la formation du cœur hydrophobe [151]. La conservation de résidus semble participer au maintien de la fonction mais aussi du repliement. Enfin, la conservation de résidus a été comparée, selon leur position à l'interface et à la surface de protéines [152, 153]. Ces deux travaux s'accordent sur le fait que l'interface est plus conservée que le reste de la surface, rejoignant ainsi les observations faites lors de l'analyse de la composition en résidus [138]. Cependant, les résultats diffèrent quant à la possibilité d'utiliser ce critère afin de définir précisément la localisation de sites de liaison. Il apparaît que plusieurs types de résidus sont particulièrement conservés, notamment les résidus polaires ainsi que certains résidus hydrophobe/aromatique tels que Trp, Met et Phe [152, 154]. D'après les auteurs, ces observations permettraient d'améliorer l'identification de sites de liaison. Toutefois, une autre étude estime approximativement à 25%, la probabilité de prédire correctement 50% des résidus de l'interface par ce critère [153].

Ces différentes conclusions peuvent être attribuées à la nature des échantillons utilisés dans ces deux travaux. Les premiers travaux ne prennent en compte que les hétéro-complexes, dont 50% correspondent à des complexes antigène-anticorps. Dans le second, divers complexes sont pris en compte, homo-dimère (65%) et hétéro-dimère (35%), permanent (16%) ou transitoire (19%). L'analyse des résultats de ces travaux suggère que l'on ne peut pas, à l'heure actuelle, compte-tenu de problèmes d'échantillonnage, utiliser ce critère pour identifier et caractériser les interfaces.

II.1.4 Conclusion

L'analyse des hétéro-complexes protéiques permet désormais d'avoir une représentation schématique de l'interface. Leur caractéristique géométrique est standard dans un grand nombre de cas : l'interface est pseudo-circulaire, de taille égale à $1600 \pm 400 \text{ \AA}^2$, et les surfaces mises en contact sont très complémentaires comme en témoigne leur forte compacité. Il s'avère que la complémentarité de forme est un critère général, applicable à toute interface, que nous utiliserons dans notre approche afin de sélectionner les protéines pouvant interagir avec une cible déterminée. Néanmoins, les critères de composition, présentant une grande variabilité, ne peuvent être utilisés afin de discriminer les protéines aptes à interagir avec une cible donnée. Ceux-ci ne peuvent être utilisés dans une approche de conception de ligands, la plus générale possible.

II.2 Considérations énergétiques

II.2.1 Aspects thermodynamiques de l'interaction

La formation d'un complexe protéique en solution s'accompagne d'un gain entropique lié à la libération des molécules d'eau liées aux protéines. Par ailleurs, celle-ci s'accompagne aussi d'une perte de degrés de liberté des partenaires, de la désolvatation de l'interface, de la rupture d'interactions avec le solvant, de modifications éventuelles de conformation ainsi que des changements d'orientation des chaînes latérales de la cible et/ou du ligand [135, 155, 156], se traduisant par un coût énergétique important, schématisé figure II-13, A. Or, l'existence d'un complexe protéique en solution est la conséquence d'un gain d'énergie lié à l'association des deux partenaires protéiques, dû à la réalisation d'interactions dont la contribution énergétique est favorable et à un gain entropique consécutif à l'exclusion de molécules d'eau, indiqué dans le tableau II-13, B.

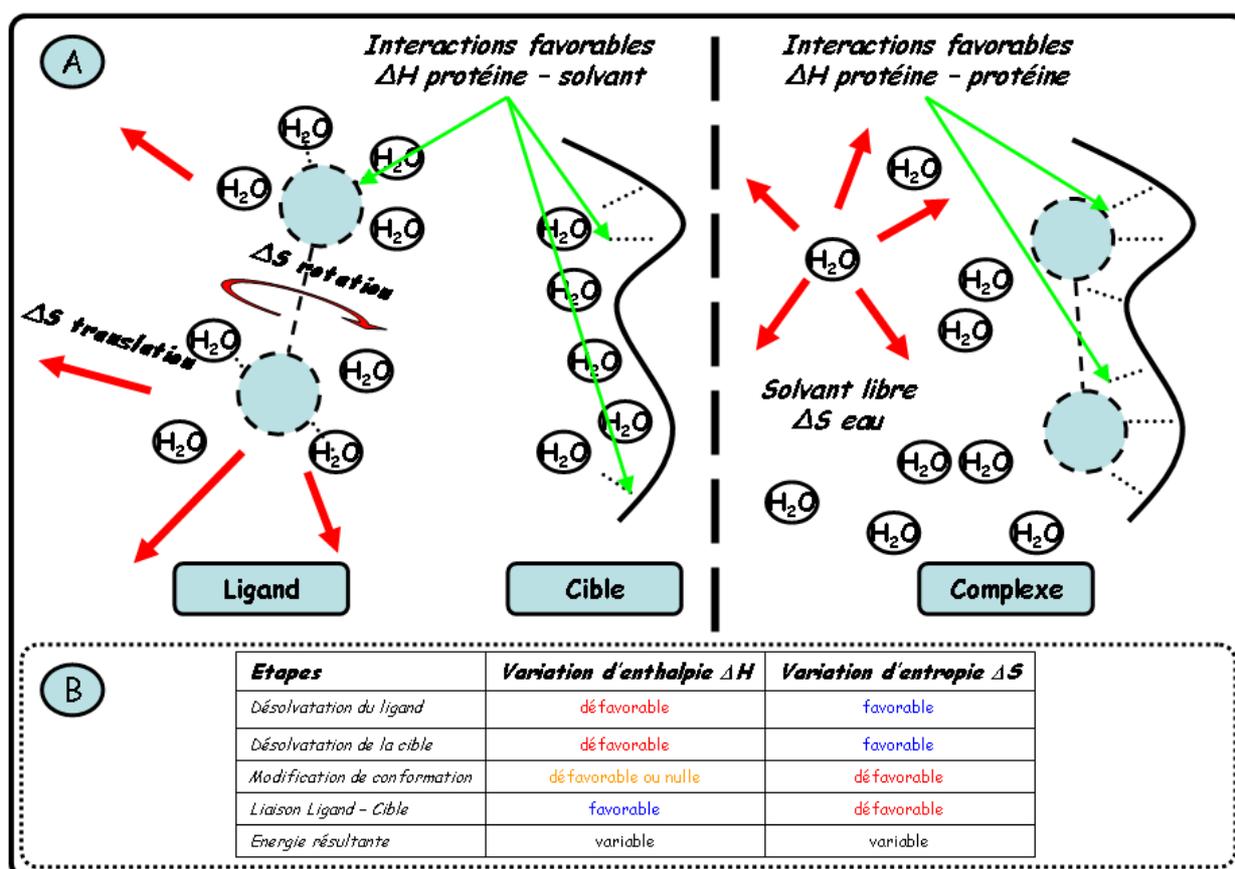


Figure II-13 : Représentation schématique de la formation d'un complexe ligand-cible, et des conséquences énergétiques (enthalpie et entropie) de chaque étape décrivant le phénomène d'association.

Expérimentalement, la mesure de la constante de dissociation (K_d) d'un complexe permet d'estimer cette variation d'enthalpie libre. D'après la figure II-14, un accroissement de l'affinité

d'un facteur 1000 correspond approximativement à une énergie de 17 kJ.mol⁻¹ (≈4 kcal.mol⁻¹), soit environ l'équivalent de l'énergie d'une seule liaison hydrogène.

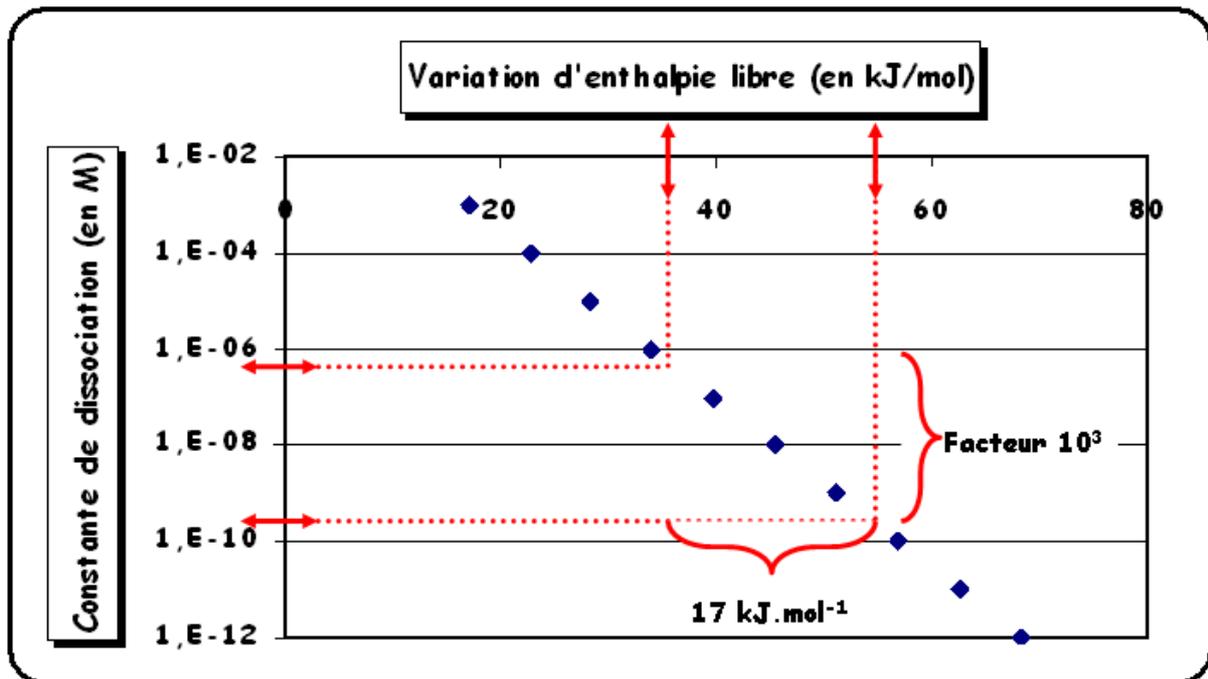


Figure II-14 : Diagramme représentant la variation d'enthalpie libre (exprimée en kJ.mol⁻¹) en fonction de la constante de dissociation (exprimée en M) : $\Delta G_d = -2.303 RT \log K_d$; d'où $K_d = 10e^{(-\Delta G_d/1.36)}$. Les points du graphique sont calculés à partir de cette dernière équation.

De faibles variations d'énergie suffisent à expliquer d'importantes variations d'affinité. De façon similaire, la conséquence sur la variation d'enthalpie libre, consécutive à la modification d'un seul résidu, peut être calculée, permettant une estimation fine des déterminants énergétiques de l'interaction.

II.2.2 Interactions ponctuelles entre atomes non-liés

L'origine de l'énergie des différents résidus à l'interface peut être attribuée à la réalisation d'interactions ponctuelles entre atomes non-liés. Plusieurs types d'interactions électrostatiques sont présentes au niveau des protéines, chacun pouvant avoir un rôle et une importance différentes (tableau II-15) :

Interaction	Charge	Dipôle	Quadrupôle	London	Répulsion e ⁻
Charge	$\pm 1/r$	$\pm \Theta/r^2$	$\pm \Theta/r^3$		
Dipôle		$\pm \Theta/r^3$	$\pm \Theta/r^4$		
Quadrupôle			$\pm \Theta D/r^5$		
London				$-1/r^6$	
Répulsion e ⁻					$+1/r^{12}$

Tableau II-15 : Les différentes interactions électrostatiques présentes dans les protéines. Les symboles r , Θ et D représentent respectivement la dépendance de l'énergie potentielle en fonction de la distance, de l'arrangement spatial et de l'orientation relative des groupements impliqués [157].

II.2.2-A Interactions charge-charge (forces de Coulomb)

Les interactions coulombiennes sont les plus énergétiques parmi les interactions entre atomes non-liés. Au niveau des interfaces, la proximité de deux charges, de signe opposé, permet la réalisation d'un pont salin, ce qui nécessite cependant au préalable la désolvatation des charges mises en jeu. De ce fait, la formation d'un pont salin est souvent énergétiquement neutre voire pénalisante [158]. Pour cette raison, les interactions électrostatiques à courte distance ne peuvent représenter un élément énergétiquement déterminant dans la formation de l'ensemble des complexes protéiques. Ces considérations sont étayées par les analyses de complexes cristallographiques, énumérés dans le tableau II-16.

En moyenne, les interfaces ne comprennent que deux ponts salins, dont 70% n'en contiennent qu'une seule voire même aucune. [159]. Ces derniers semblent toutefois importants pour la spécificité d'une interaction donnée [160, 161]. La description des ponts salins est donc dépendante du complexe considéré et ne peut être réalisée par une analyse statistique, décrivant l'ensemble des complexes.

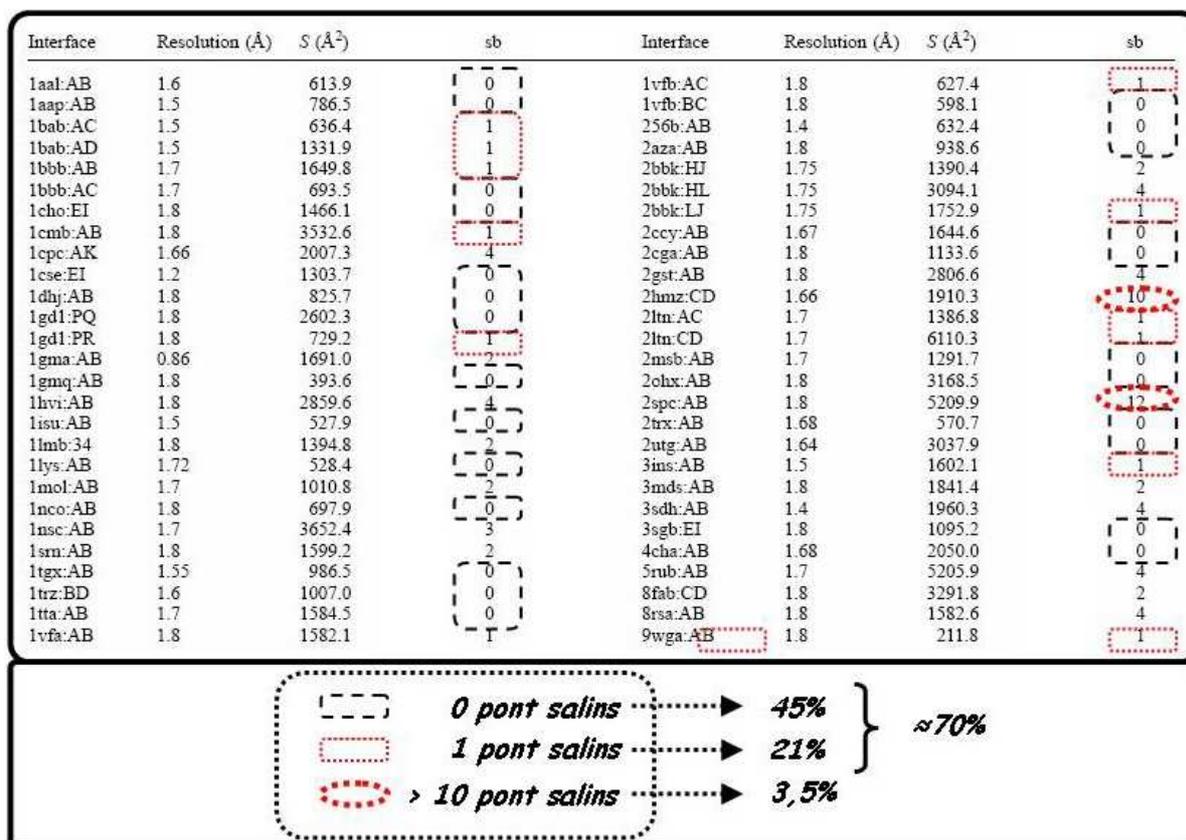


Figure II-16 : Dénombrement des ponts salins présents à l'interface de divers complexes. Ce nombre, indépendant de la taille de l'interface, est en moyenne de deux ponts salins par interface [159].

II.2.2-B Interactions dipôle-dipôle (forces de Van der Waals)

Les interactions entre dipôles correspondent à des forces attractives entre dipôles instantanés, dû au déplacement des électrons (figure II-17). Leur énergie, dépend de la nature du dipôle, permanent ou induit, et décroît de façon importante avec la distance (efficaces entre 0,03 et 0,08 Å).

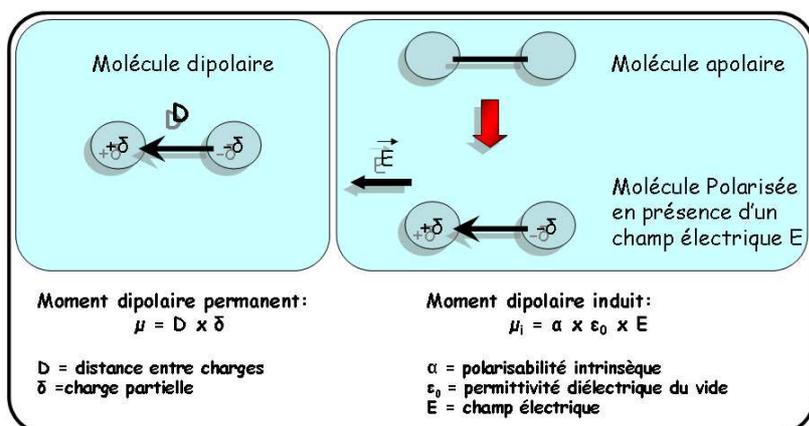


Figure II-17 : Représentation schématique des dipôles, permanent (partie gauche) et induit (partie droite).

Au niveau des interfaces, les forces de Van der Waals expliquent la compacité très forte de l'interface, par la réalisation d'un maximum d'interactions, dont la somme représente un terme énergétique non-négligeable. Plusieurs interactions dipolaires sont particulières aux protéines.

II.2.2-B1 Cas particulier : l'effet hydrophobe (CM, eau-eau)

Dans le cas de composés hydrophobes, donc ne présentant pas de dipôles permanents, les forces de dispersion (dipôle induit/dipôle induit) sont à l'origine d'interactions connues sous le nom d'interactions hydrophobes. Les groupements apolaires vont interagir entre eux, *via* l'établissement d'interactions de Van der Waals, en excluant les molécules polaires et favorisant la réalisation d'interactions dipolaires. Il y a donc deux composantes, l'une enthalpique due aux interactions dipôle-dipôle (eau-eau, et molécules apolaires entre-elles), l'autre entropique par exclusion des molécules de solvant. Le processus est similaire à celui du repliement des protéines mais, dans une moindre mesure, concernant les interfaces protéiques [162].

II.2.2-B2 Cas particulier : la liaison hydrogène

Il s'agit de la combinaison d'une interaction entre deux dipôles appartenant à la même molécule ou deux molécules voisines, un groupement polarisé (O-H, N-H, S-H... représentés par D-H) et un groupement polarisable (atome accepteur porteur d'un doublet non-liant, représenté par A), et d'un terme de transfert de charge. La contribution respective de chaque terme est encore sujette à controverse. Concernant les protéines, ce type d'interaction est responsable de la formation des structures secondaires, par l'intermédiaire des groupements -CO- et -NH- de la chaîne principale. Au niveau des interfaces, différents cas sont observés : interactions chaîne principale-chaîne principale, chaîne principale-chaîne latérale, et chaîne latérale-chaîne latérale. Celles-ci représentent respectivement 23%, 35% et 39% de la totalité des liaisons hydrogène recensées à l'interface, les 3% restant correspondant à des liaisons établies avec différentes molécules non-protéiques, molécules d'eau par exemple. Le nombre moyen est estimé à 10 approximativement [159], mais relativement variable, de 6 à 33 selon l'interaction considérée. En revanche, le nombre de liaisons est corrélé avec la taille de l'interface, montré sur la figure II-18.

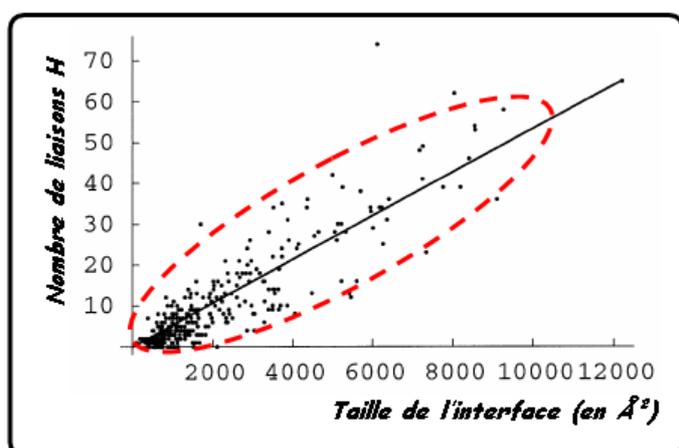


Figure II-18 : Évolution du nombre de liaisons hydrogène en fonction de la taille de l'interface, exprimée en Å^2 [159].

La densité de liaisons hydrogène est relativement constante, à raison de $0,5 \pm 0,3$ liaisons par 100 \AA^2 en moyenne [159, 163]. Par ailleurs, en milieu aqueux, les protéines sont hydratées, à raison de 15 molécules par 1000 \AA^2 dans le cas d'hétéro-dimère [164]. Chaque molécule d'eau est capable de réaliser jusqu'à quatre liaisons hydrogène, de ce fait la réalisation de liaisons hydrogène fortes au niveau de l'interface ne peut se faire que si celle-ci est protégée du solvant. Ces molécules fragilisent un certain nombre de liaisons hydrogène tandis que la formation du complexe protéique permet d'en protéger d'autres du solvant, notamment au cœur de l'interface [144]. L'énergie résultante de la protection de liaisons hydrogène, insuffisamment déshydratées à l'état libre, pourrait représenter un part importante de l'énergie de liaison [165]. Toutes ces observations démontrent que le rôle des liaisons hydrogène est important pour la formation de complexes protéiques, toutefois leur prise en compte s'avère difficile et nécessite de ne pas négliger le rôle du solvant ainsi que la géométrie des liaisons. L'assimilation des protéines à des corps rigides et le traitement du solvant comme un continuum sont alors insuffisants pour traiter correctement cet aspect.

II.2.2-C Interactions charge-quadrupôle

Un quadrupôle est défini par une distribution de charges (au minimum 3), électriquement neutre, dont l'agencement crée deux dipôles permanents, sans toutefois posséder de moment dipolaire comme indiqué par la figure II-19. Chaque distribution possède, en revanche, un moment quadripolaire non-nul, à l'origine de ses propriétés électrostatiques.

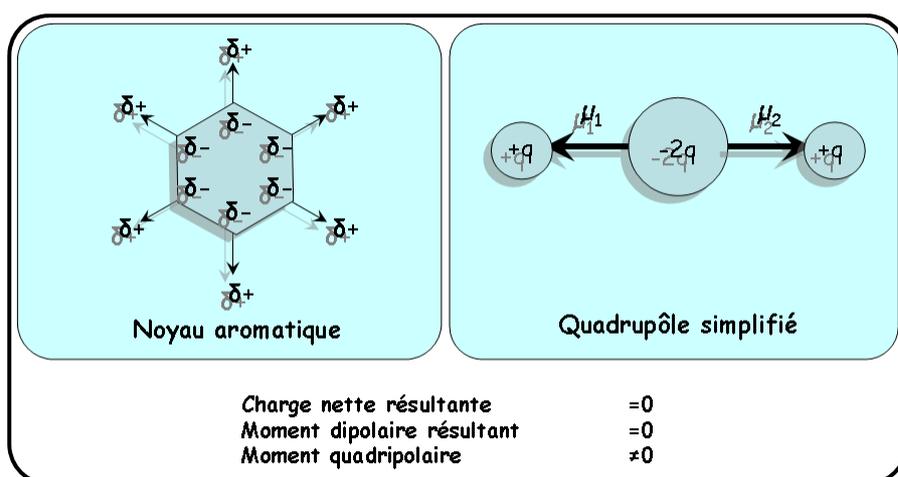


Figure II-19 : Représentation schématique de quadrupôle. Dans les protéines, les résidus aromatiques sont les seuls quadrupôles présents, capables de réaliser des interactions π - π (entre résidus aromatiques) ou cation- π (entre un résidu aromatique et une charge ponctuelle positive).

Dans les protéines, les noyaux aromatiques de résidus Phe, Tyr, Trp et le cas échéant His, constituent les seuls quadrupôles présents. Les résidus aromatiques peuvent interagir avec d'autres résidus aromatiques *via* la réalisation d'interactions π - π ou avec charges positives ponctuelles *via* la réalisation d'interactions cation- π . De façon identique à la formation de ponts

salins, les interactions cation-pi sont peu fréquentes, seule la moitié des complexes analysés présentant au moins une interaction cation- π . [166]. Elles seraient, de la même façon, impliquées dans la spécificité, et donc caractéristiques du complexe considéré.

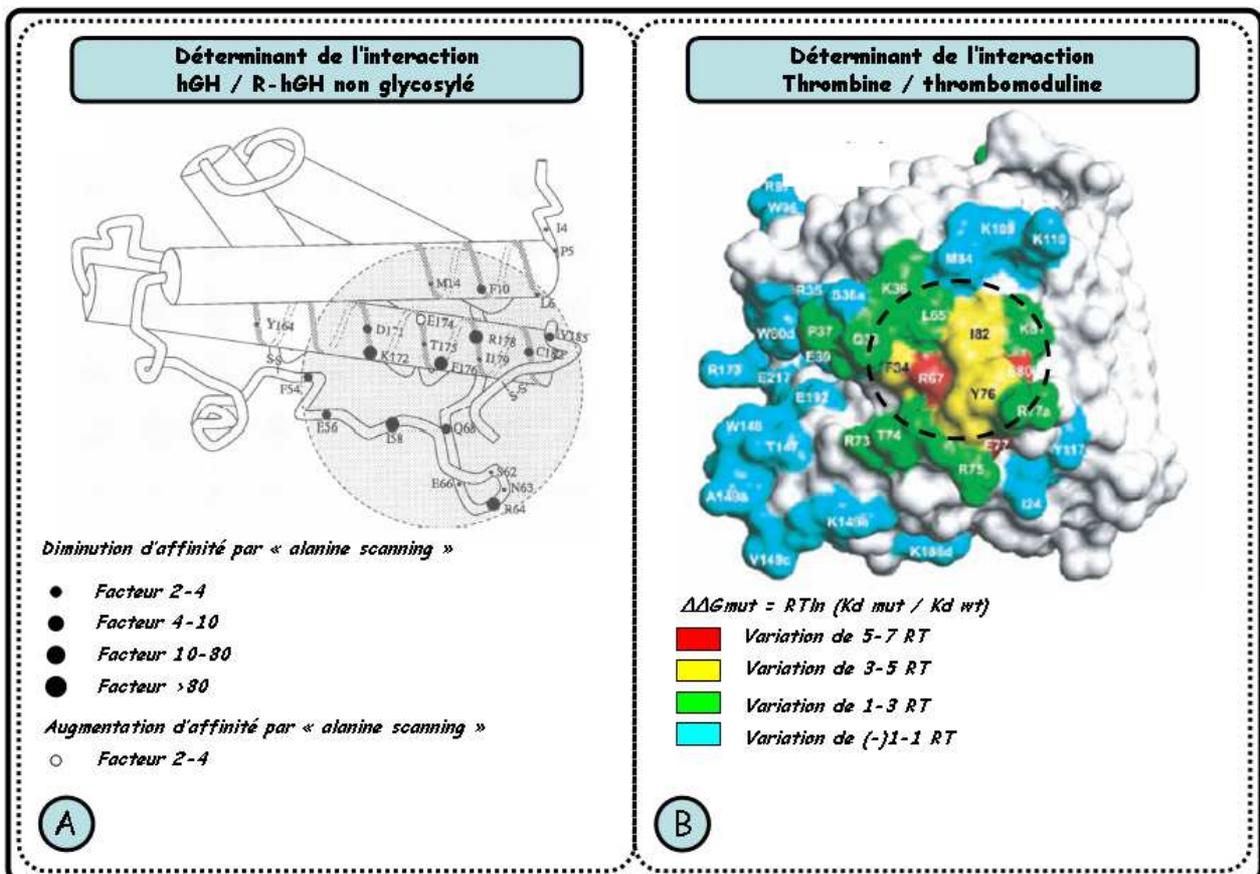


Figure II-20 : Représentation des « hotspots » à la surface des protéines hGH (hormone de croissance humaine), partie A, et thrombine, partie B [167, 169].

II.2.3 Définition de « point chaud de l'interaction » ou « hotspot »

Lors de la formation d'un complexe, le gain d'énergie peut être considéré comme la somme de termes énergétiques discrets (contributions individuelles de chaque résidu). Il est impossible de mesurer expérimentalement, la contribution des éléments appartenant à la chaîne principale de chaque résidu (Ca, C, O et N). Il est toutefois possible d'estimer la contribution énergétique des chaînes latérales. Pour ce faire, l'approche couramment utilisée est « l'alanine scanning », méthode par laquelle chaque position est systématiquement remplacée par un résidu alanine [31]. La mesure de la constante de dissociation du complexe permet de calculer la variation d'énergie libre par résidu, remplacé en alanine [48]. L'énergie libre est alors directement reliée à la présence de fonctions chimiques particulières (chaînes latérales). De cette façon, une valeur d'énergie libre est attribuée à chaque résidu : lorsque cette valeur est supérieure à $8,36 \text{ kJ}\cdot\text{mol}^{-1}$ ($2 \text{ kcal}\cdot\text{mol}^{-1}$), le résidu est considéré comme énergétiquement important et dénommé « point chaud » ou « hotspot » de l'interaction [144]. A titre indicatif, une variation de $5,7 \text{ kJ}\cdot\text{mol}^{-1}$ ($1,4 \text{ kcal}\cdot\text{mol}^{-1}$) modifie l'affinité d'un facteur 10 (figure II-14).

Il a été montré, au niveau des interfaces, que seul un nombre très restreint de résidus énergétiquement important était nécessaire, comme cela a été observé dans les complexes hormone de croissance humaine hGH/récepteur à l'hGH [167, 168] et thrombine/thrombomoduline [169], représentés figure II-20, parties A et B. Les expériences « d'alanine scanning » menées jusqu'à présent ont localisé ces « hotspots » très majoritairement au niveau du cœur de l'interface, plus rarement au niveau de sa périphérie (couronne). La surface accessible au solvant des « hotspots » a été mise en relation avec leur contribution énergétique. Les résidus les plus énergétiquement importants sont, en parallèle, les plus faiblement accessibles au solvant, comme le montre la figure II-21.

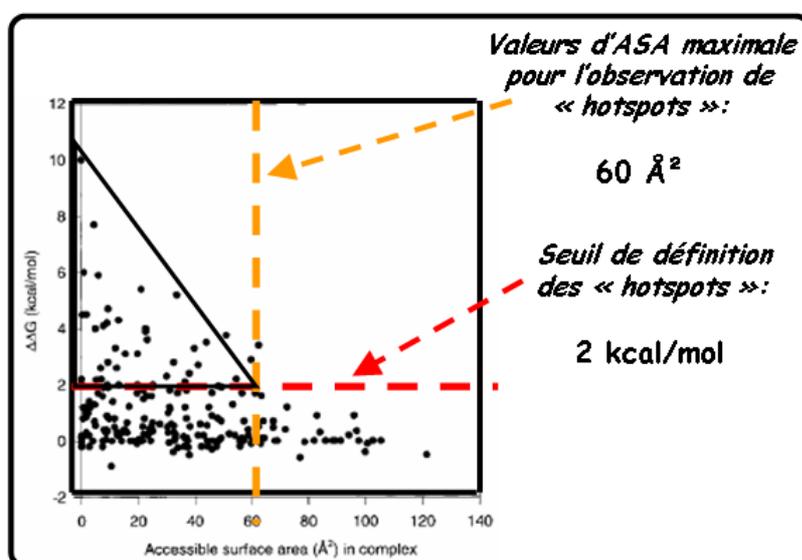


Figure II-21: Répartition des « hotspots » en fonction de leur surface accessible au solvant. [144]

Dans un modèle proposé, cette dernière est supposée protéger le cœur vis-à-vis des molécules de solvant. Ainsi les interactions de contact énergétiquement importantes sont-elles favorisées [144], leur surface accessible étant d'autant plus faible. De plus, l'analyse de la composition en résidus montre certaines particularités. La répartition des hotspots par type de résidus est présentée dans le tableau II-22.

Residue	In database		Contribute ≥ 2 kcal/mol		Enrichment in hot spots
	(Number)	(%)	(Number)	(%)	
Arg	218	9.38	29	13.30	2.47
Asn	99	4.26	5	5.05	0.93
Asp	177	7.61	16	9.04	1.67
Cys	3	0.13	0	0	0
Gln	160	6.88	5	3.13	0.58
Glu	220	9.46	8	3.64	0.68
Gly	28	1.20	1	3.57	0.45
His	50	2.15	4	8.00	1.49
Ile	104	4.47	10	9.62	1.79
Leu	242	10.41	2	0.83	0.01
Lys	143	6.15	9	6.29	1.17
Met	69	2.97	2	2.90	0.54
Phe	166	7.14	5	3.01	0.56
Pro	89	3.83	6	6.74	1.25
Ser	178	7.66	2	1.12	0.21
Thr	131	5.63	2	1.53	0.28
Trp	19	0.82	4	21.05	3.91
Tyr	122	5.25	15	12.30	2.29
Val	107	4.60	0	0	0

Tableau II-22 : Répartition des « hotspots » selon le type de résidu considéré. [144]

D'après les auteurs, trois résidus sont largement conservés en tant que « hotspots » : Trp, Arg et Tyr. Ces résidus sont susceptibles de réaliser différents types d'interactions ponctuelles (ponts salins, liaisons hydrogène, interactions π -cation) expliquant ainsi ces observations. *A contrario*, d'autres sont quasiment absents (<3%) : Leu, Met, Ser, Thr et Val. Globalement, il n'y a pas de préférences de types particuliers de résidus en tant que « hotspots » (hydrophobes, polaires, chargés ou autres). Il apparaît donc relativement difficile d'identifier les « hotspots » à l'interface, sans passer par l'expérimentation. Actuellement, diverses méthodes ont été proposées afin d'identifier les « hotspots » *in silico* [170, 171], permettant ainsi de s'affranchir de lourdes expérimentations. Celles-ci diffèrent notamment par le modèle physique utilisé pour la représentation de l'énergie d'interaction. Le serveur ROBETTA, mis à disposition sur internet [172, 173, *] permet de réaliser un « alanine scanning » *in silico*. Les études effectuées ont permis d'identifier correctement approximativement 80% des « hotspots » et 70% des résidus énergétiquement neutres. Les écarts moyens entre les valeurs prédites et mesurées sont de l'ordre de 4 kJ.mol⁻¹ [171, 173]. Ces résultats, certes très encourageants, semblent toutefois encore insuffisants pour une identification précise de sites fonctionnels.

* <http://rosetta.bakerlab.org>

II.2.4 Conclusion

Malgré certaines tendances claires quant à la présence de chaque type d'interaction au niveau des interfaces, l'identification des déterminants de l'interaction ou « hotspots » ne peut être réalisée de manière systématique : chaque interaction est spécifique et doit être analysée individuellement.

Dans notre approche, une partie de l'énergie de formation d'un complexe devrait être prise en compte par la complémentarité de forme, correspondant à une partie de l'interaction non spécifique (interactions de Van der Waals), associée à un site de liaison spécifique, constitué par un nombre restreint de « hotspots », déterminés par « alanine scanning » et, transférés sur une plateforme structurale appropriée.

II.3 Le potentiel électrostatique dans les interactions protéiques

L'approche développée repose sur le transfert de résidus, énergétiquement importants pour une interaction donnée, sur une plateforme structurale apte à reproduire cette interaction. Toutefois, ces résidus ne sont pas les seuls capables d'établir des interactions, favorables ou défavorables, avec la cible. Une contribution importante est notamment due à la réalisation d'interactions électrostatiques, à longue distance, lesquelles peuvent être attractives ou répulsives. Dès lors, prendre en compte les aspects électrostatiques s'avère nécessaire.

II.3.1 Aspects électrostatiques de l'interaction

Il est désormais admis que les termes électrostatiques soient impliqués dans de nombreux aspects des phénomènes biologiques : désolvatation, variation de pKa des résidus chargés, cinétique d'association [174,175]. La formation d'un complexe protéique peut être décrite simplement par la réalisation de deux étapes successives [176], schématisées figure II-23.

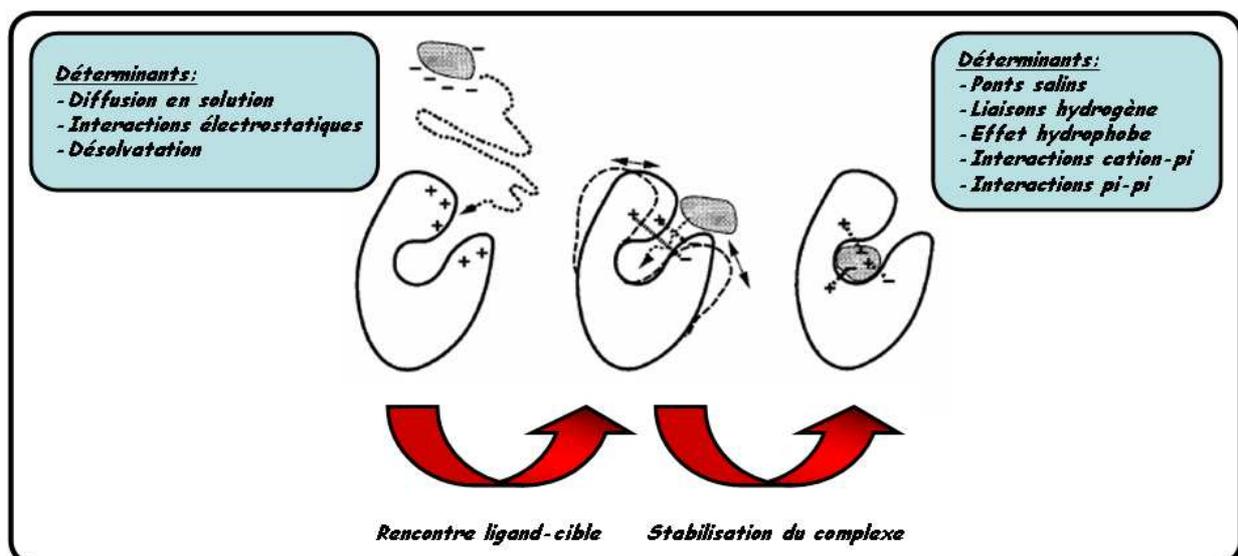


Figure II-23 : Représentation schématique de la formation d'un complexe. Cette dernière nécessite la réalisation de deux étapes successives, une première correspondant à la rencontre des deux protéines, la seconde, à la stabilisation du complexe [177].

La première étape de rencontre correspond à la diffusion des protéines en solution, jusqu'au au positionnement et à l'orientation correct de leurs sites de liaison respectifs. Elle dépend de la réalisation d'interactions longue distance favorables et de la désolvatation des protéines, notamment des résidus chargés [176, 178]. La seconde correspond à la stabilisation du complexe, et repose notamment sur la réalisation d'interactions électrostatiques à courte distance, rendant compte de l'affinité et la spécificité du complexe [160, 179].

Plusieurs travaux ont relié les propriétés électrostatiques avec la fonction des protéines. Par exemple, diverses protéines d'adhérence partageant une même région fonctionnelle ont pu être caractérisées par des critères purement électrostatiques, comme l'indique la figure II-24.

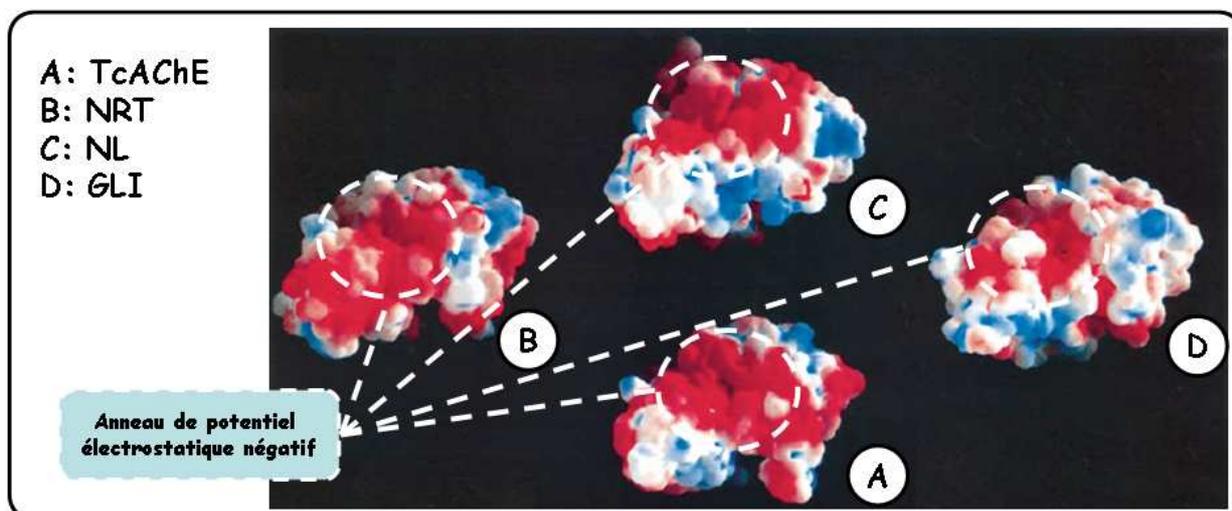


Figure II-24 : Représentation du potentiel électrostatique à la surface des protéines A, B, C et D. TcAChE : *Torpedo californica* acetyl-cholinesterase ; NRT : glycoprotéine transmembranaire (*drosophile*) ; NL : neuroligine (*drosophile*) ; GLI : protéine transmembranaire (*souris*) [180].

Ces protéines diverses présentent des homologies de séquences inférieures à 32% ; pourtant toutes possèdent un motif électrostatique commun, représenté par un anneau de potentiel négatif (en rouge sur la figure, encadré en blanc). Ainsi, une nouvelle famille de protéines a-t-elle été définie, dénommée de ce fait « électrotactines » [180,181].

Plus généralement, il a été observé que la fonction de familles et super-familles d'enzymes était conservée du fait de propriétés électrostatiques similaires [182]. Les enzymes « superoxyde dismutase à Cu/Zn », de différentes espèces (animales ou végétales), possèdent des états de charge compris de -8 à +2 et des identités de séquence variant de 25 à 80%. Pourtant, malgré ces différences, toutes présentent un potentiel électrostatique similaire au niveau de leur site catalytique. Des constatations identiques ont été observées pour les familles du lysozyme (type-c), de la ferritine, de la myoglobine et la super-famille des énoleses [182]. Pour des sites de liaisons comme pour des sites catalytiques, les propriétés électrostatiques semblent conservées afin de maintenir la fonction.

II.3.2 Notion de potentiel électrostatique (*)

La réalisation d'interactions électrostatiques à longue distance est expliquée par la faible dépendance de l'énergie potentielle en fonction de la distance ($1/r^2$) des interactions coulombiennes. La formation de ponts salins est d'ailleurs proposée pour être largement impliquée dans la spécificité d'interaction [160]. Donc, deux protéines en interaction, devraient présenter une distribution de charge au niveau de leur interface similaire ou compatible, justifiant la première hypothèse de complémentarité des charges présentes à l'interface (figure II-25, A). Or, il a été démontré, par l'analyse de complexes cristallographiques, d'une part que le nombre de ponts salins à l'interface était relativement faible, moins de deux dans 70% des cas [159], et d'autre part qu'il n'y avait qu'une très faible corrélation entre la disposition des charges à l'interface [183]. Toutefois, cette dernière étude a mis en évidence une complémentarité importante du potentiel électrostatique des protéines interagissant (figure II-25, B).

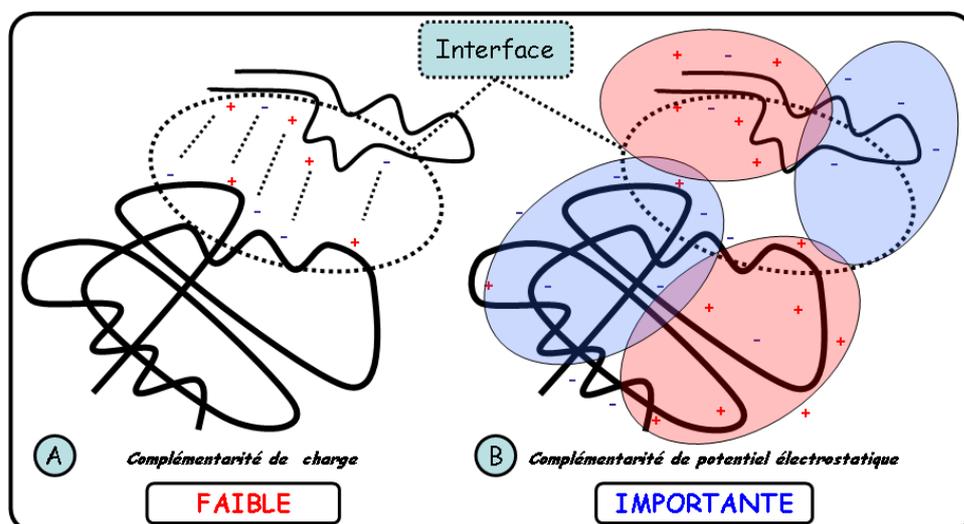


Figure II-25 : Représentation schématique de la complémentarité de charges et de la complémentarité de potentiel à l'interface.

Par ailleurs, il a été montré que cette complémentarité de potentiel électrostatique n'était reliée ni au type de complexe considéré, ni au nombre de ponts salins présents à l'interface. Dès lors ne considérer que les résidus chargés de l'interface et leurs distances respectives n'est pas suffisant pour décrire les propriétés électrostatiques de divers ligands d'une même cible. Il convient d'étudier la résultante globale des charges d'une protéine, représentée par le calcul du potentiel électrostatique. Plusieurs travaux ont été réalisés afin d'estimer la possible utilisation de la complémentarité de potentiel électrostatique pour l'identification de sites catalytique ou de liaison. Ainsi, diverses protéines de la chaîne de transfert d'électron, possédant des structures et des tailles différentes mais exerçant des fonctions analogues, ont-elles été comparées.

* Potentiel électrostatique

Toute distribution de charges crée en tout point de l'espace crée un potentiel électrostatique. Le potentiel est une grandeur scalaire, dérivée du vecteur champ électrique.

La plastocyanine et le cytochrome c_6 ont été orientés de façon à optimiser leur similarité de potentiel électrostatique. Il s'est avéré que les sites présentant la plus forte similarité correspondaient aux sites fonctionnels de ces protéines [184]. Un travail similaire a été effectué concernant la ferredoxine et la flavodoxine [185], aboutissant aux mêmes conclusions.

Un autre exemple décrit l'influence du potentiel électrostatique dans l'étape de rencontre des partenaires protéiques par la comparaison de la triose phosphate isomérase de quatre espèces différentes. Ces dernières montrent une grande hétérogénéité au niveau de leurs charges nettes, de -12 à +12 à pH = 7. Cet enzyme est considéré comme « parfait », à savoir que la vitesse réactionnelle n'est limitée que par la diffusion et ne peut donc pas être accélérée. Les résultats montrent la conservation du potentiel électrostatique autour des sites actifs, ce qui explique probablement l'importance de cette région pour orienter et positionner le substrat par des interactions longue distance. Cette observation pourrait être applicable à tous les enzymes dits « parfaits » [186].

Les mêmes auteurs ont réalisé une étude comparative sur 104 protéines appartenant à la famille des protéines à domaine homologue de la pleckstrine (PH). Malgré de grandes divergences de séquence, les potentiels sont très comparables. La méthode présentée permet d'envisager des comparaisons à très grande échelle dans un temps relativement court, et donc un criblage des banques de structures 3D de protéines [187]. Une étude similaire, réalisée sur la famille des protéines à cuivre bleu, aboutit à des conclusions similaires [188], confortant ainsi l'utilité de la comparaison du potentiel électrostatique de protéines diverses pour l'identification de leur fonction.

II.3.3 Conclusion

La complémentarité de charge à l'interface n'apparaît pas essentielle. En revanche les potentiels électrostatiques, de deux protéines en interaction, sont complémentaires. De nombreux exemples prouvent que la comparaison du potentiel électrostatique permet d'identifier diverses protéines exerçant une même fonction, y compris une fonction de liaison. Nous avons donc estimé que cet aspect électrostatique devait être utilisé, de façon similaire à la complémentarité de forme des interfaces, afin de sélectionner les protéines issues de notre approche de conception les plus favorables à l'interaction avec la cible préalablement choisie.

Le potentiel électrostatique est considéré comme un élément essentiel à la reconnaissance entre différents partenaires. Il est démontré par de nombreux travaux d'une part, que les protéines exerçant une même fonction présentent un potentiel similaire [180, 182, 184, 185] et d'autre part, que les protéines interagissant entre elles présentent un potentiel électrostatique complémentaire [183]. L'hypothèse qui découle logiquement de ces constatations est que différents ligands d'une même cible se doivent de présenter un potentiel électrostatique peu différent. L'estimation de la similarité de potentiel électrostatique avec le ligand de référence correspond au filtre électrostatique utilisé dans notre approche (annexe A.3).

Chapitre III : Conception de ligands protéiques

« A la source de toute connaissance, il y a une idée, une pensée, puis l'expérience vient confirmer l'idée » (Claude Bernard)

En tenant compte des différentes hypothèses et observations, citées précédemment, nous avons élaboré une méthodologie dont la finalité est la conception de ligands de nature protéique. L'approche de conception de ligands développée au cours de ce travail, basée sur la structure, repose sur le transfert d'un motif fonctionnel sur une protéine hôte :

- le motif est constitué de résidus énergétiquement importants (« hotspots ») et transféré en préservant sa topologie. Il peut être indifféremment, continu ou discontinu en séquence ;
- les protéines hôtes sont identifiées par une recherche systématique *in silico* à partir des structures contenues dans la PDB. La recherche de motifs topologiques, similaires au motif de référence, est basée sur le seul agencement tridimensionnel des résidus le composant et de ce fait s'affranchie de la structure secondaire le portant.

Ainsi notre approche est-elle applicable, quel que soit le motif considéré ou la structure qui le supporte.

Les structures identifiées sont ensuite triées selon plusieurs critères déterminants pour l'interaction avec la cible. La complémentarité de forme est prise en compte afin d'éliminer toute structure stériquement incompatible avec la cible du ligand de référence. Les propriétés électrostatiques sont prises en compte *via* la comparaison du potentiel électrostatique des protéines hôtes avec celui du ligand de référence. Les structures aptes à recevoir le motif sont dès lors dénommées plateformes.

Dans une première partie, nous détaillerons l'approche de conception de ligands protéiques envisagée, dans son ensemble. Puis, dans une seconde partie, le système biologique choisi pour la mise en application de l'approche sera présenté.

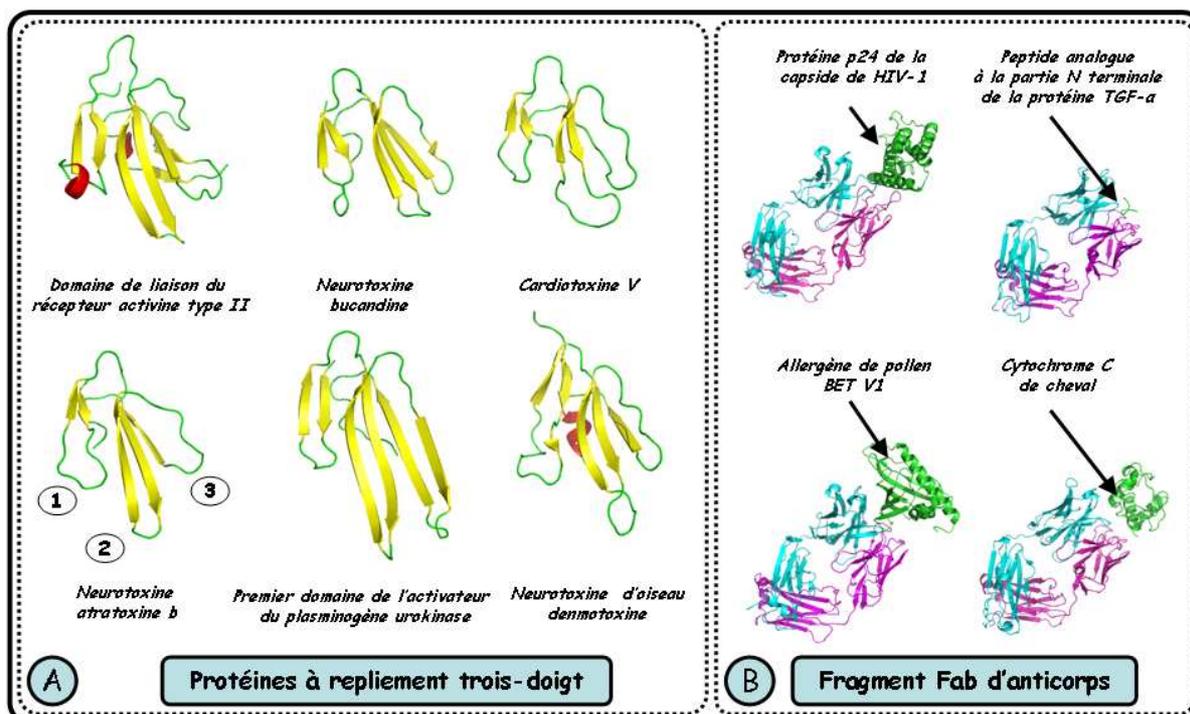


Figure III-1 : Représentation de protéines diverses possédant un même repliement mais liant des cibles moléculaires différentes. La partie A montre diverses protéines possédant un repliement dit à trois doigts. Chacune est représentée en mode « cartoon », les éléments de structure secondaire étant colorés différemment : hélice en rouge, feuillet en jaune et autre en vert. De gauche à droite et de haut en bas, 1BTE [189], 1IJC [68], 1KXI [190], 1VBO [191], 1YWH [192], 2H5F [193]. La partie B montre différents fragments Fab d'anticorps monoclonaux (en cyan et magenta pour les chaînes lourde et légère) en complexe avec leurs cibles respectives (en vert). Tous les complexes sont représentés en mode « cartoon ». De gauche à droite, et de haut en bas, 1AFV [194], 1E4X [195], 1FSK [196] et 1WEJ [197].

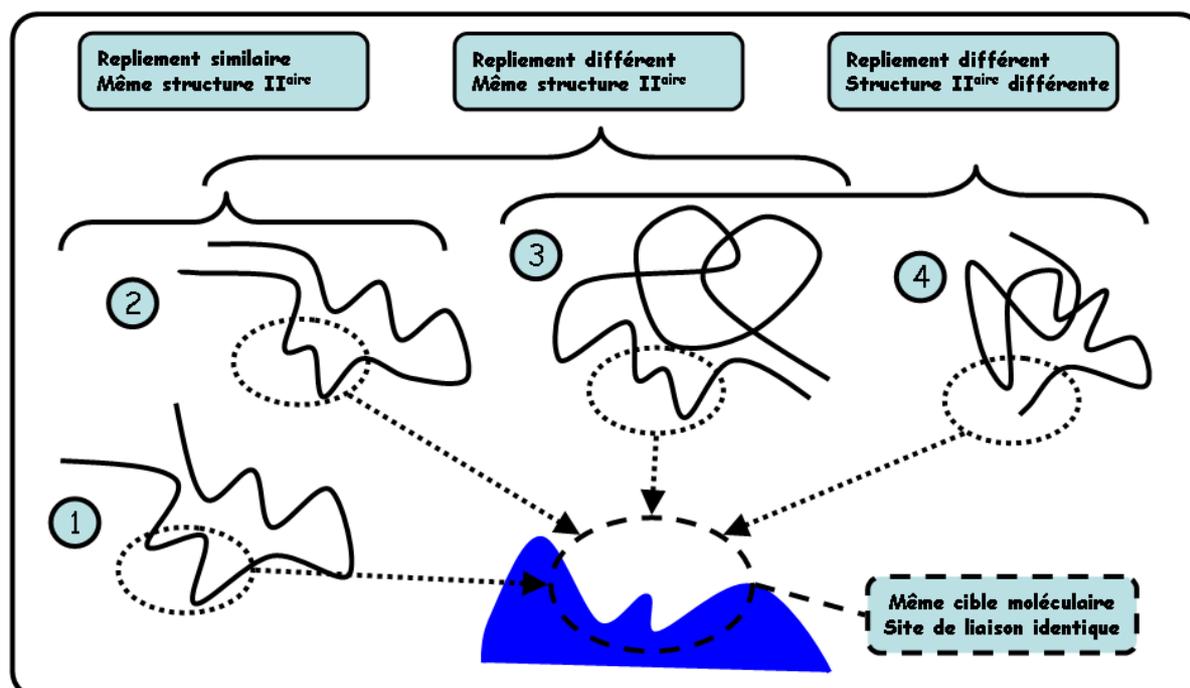


Figure III-1bis : Représentation schématique des différentes structures de ligands capables de lier une même cible. Les inhibiteurs de protéases à serine et les bloqueurs des canaux potassiques sont de parfaits exemples (chapitre III, partie 2 ; chapitre IV).

III.1 Principe de la méthode de conception de ligand protéique

III.1.1 Du motif fonctionnel à la conception de ligands protéiques

Le préalable à notre approche de conception dépend de la définition d'un motif de résidus, nécessaire pour l'interaction avec une cible déterminée. Ce motif est défini comme suit.

III.1.1-A Définition du motif fonctionnel

Il a été proposé, que l'interaction entre deux protéines soit assurée par un nombre restreint de résidus formant un motif fonctionnel [168], constitué de résidus énergétiquement importants. Ceux dont la substitution par « alanine scanning » provoque une variation d'enthalpie libre supérieure à $8,36 \text{ kJ.mol}^{-1}$ (2 kcal.mol^{-1}) sont dénommés « hotspots » de l'interaction [48]. Ces derniers sont, le plus souvent, situés à l'interface et regroupés en son cœur ; ainsi sont-ils protégés du solvant et en situation d'interagir directement par la réalisation d'interactions ponctuelles [144]. Bien que les « hotspots » se trouvent généralement proches dans l'espace les uns des autres, définissant une surface d'interaction, il n'y a pas nécessairement de proximité dans la structure primaire. D'après l'analyse de plusieurs travaux, il semble que les motifs fonctionnels soient souvent discontinus en séquence [49, 167, 169...].

III.1.1-B Rôle de la structure

Le motif fonctionnel, défini précédemment, est supporté par un nombre beaucoup plus important de résidus, formant un édifice structural [168]. Cette observation est corroborée par le fait que différents ligands protéiques, de repliement similaire, peuvent lier des cibles variées. Les anticorps et les protéines trois doigts en sont un exemple, comme indiqué sur la figure III-1, parties A et B. Par ailleurs, divers ligands dont le repliement et/ou les éléments de structure secondaire présents à l'interface diffèrent (schématisés sur la figure III-1bis), peuvent lier une même cible : anticorps poly-clonaux, inhibiteurs des protéases, bloqueurs de canaux K^+ ... Plusieurs exemples de bloqueurs de canaux K^+ et d'inhibiteurs de protéases seront décrits plus avant dans le chapitre IV. Cette observation, que la structure de l'interface n'est pas obligatoirement constituée d'éléments de structure secondaire particuliers, a été confirmée l'analyse systématique de complexes protéiques [53, 198]. La fonction est soutenue par une structure, toutefois cette relation n'est pas univoque.

III.1.1-C Transfert de motif de liaison

Dans leur analyse de l'interaction de l'hormone de croissance humaine avec un anticorps monoclonal MAb3, Jin et Wells ont tenté de répondre à la question suivante. « Est-il possible de transférer les déterminants d'une interaction sur une autre plateforme structurale ?

« Nos résultats suggèrent que l'approche est envisageable mais peu évidente puisque la seule nature chimique des déterminants n'est pas suffisante sans une orientation favorable permettant un ajustement de l'interaction. Nous suggérons que les informations structurales et fonctionnelles sont essentielles à ce but. Bien que l'analyse structurale autorise l'identification des contacts atomiques, elle ne peut caractériser ni leur importance dans l'interaction, ni leur tolérance à des changements minimes » [199].

Le motif de liaison, tel qu'il est défini dans notre approche, correspond à un ensemble de résidus énergétiquement importants (« hotspots » identifiés par « alanine scanning ».) appartenant à un ligand de référence, dont la topologie particulière, permet l'interaction avec la cible. L'approche de conception de ligands développée repose sur le transfert de ce motif de liaison sur une plateforme structurale, sélectionnée par sa capacité à reproduire la topologie fonctionnelle du motif. Celle-ci est schématisée sur la figure III-2.

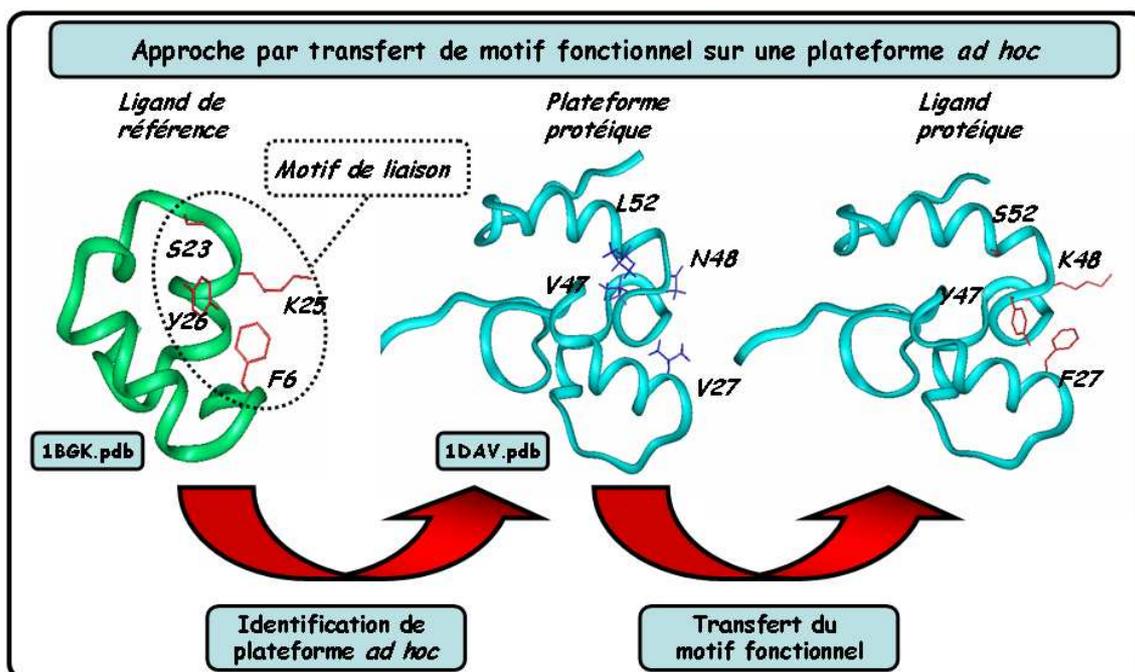


Figure III-2 : Méthode de conception de ligands protéiques par transfert de motif fonctionnel sur une plateforme structurale possédant un motif de topologie équivalente. Les structures 1BGK [200] (en vert) et 1DAV [201] (en cyan) sont représentées en mode « fancy ribbon ». Les motifs de résidus (en rouge et bleu) sont représentés en mode « stick ».

III.1.2 Recherche de structures possédant un motif topologique équivalent

III.1.2-A Recherche de motifs topologiques

L'étape de recherche est basée sur les orientations des chaînes latérales des résidus du motif de référence, déterminées à partir de la structure 3D du ligand de référence. L'orientation d'une chaîne latérale est définie par le vecteur formé des atomes Ca et $C\beta$. Dans l'étape de recherche, le motif est décrit par un ensemble de paires (Ca ; $C\beta$) représentant chaque résidu. La topologie définie par ces paires est utilisée pour rechercher, dans la PDB, toute structure possédant un motif topologique équivalent. Les recherches de tels motifs ont été effectuées avec le programme STAMP (« Search for Tridimensional Atomic Motif in Protein »), dont l'algorithme de recherche présente des similitudes à celui du programme SPASM (« SPatial Arrangements of Side chains and Mains chains ») [202]. STAMP et SPASM sont des programmes de recherches de motifs particuliers dans les protéines. Chaque motif est représenté par une liste d'atomes, défini par leurs coordonnées cartésiennes (figure III-3, A).

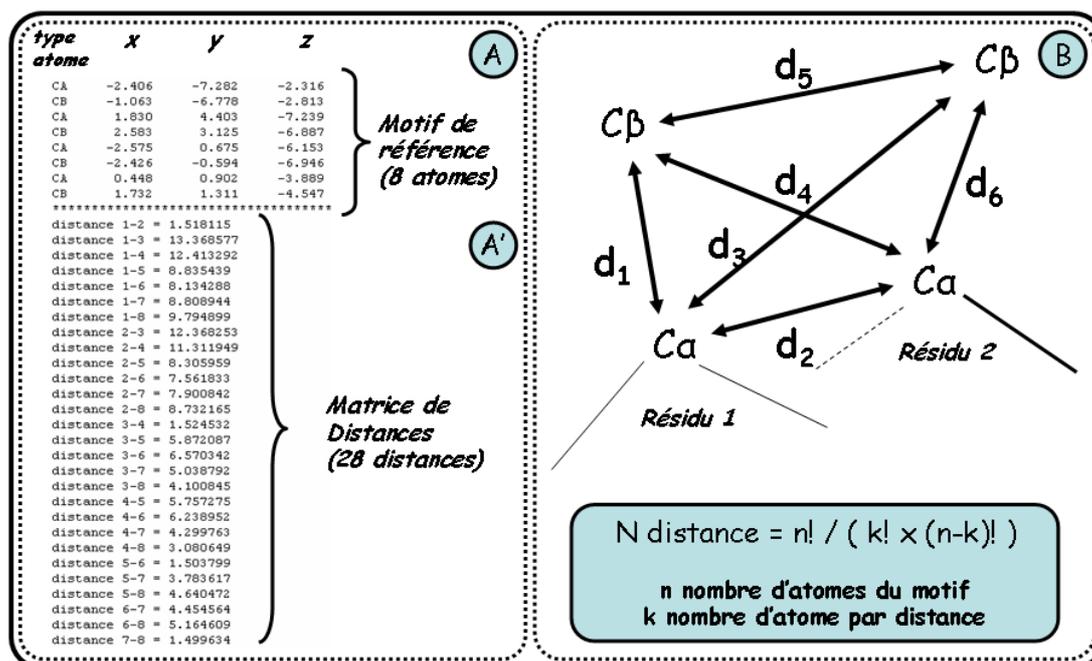


Figure III-3 : Recherche de motifs topologiques par le programme STAMP. Le nombre de distances de la matrice de comparaison est fonction du nombre d'atomes dans le motif par la relation indiquée dans le cadre, partie de droite. Les distances calculées sur les structures de la PDB sont ensuite comparées aux distances de cette matrice de référence.

L'identification de motifs similaires est accomplie par la détermination préalable d'une matrice de distances (figure III-3, A'), calculées entre chaque atome du motif de référence, puis de la recherche, parmi les structures 3D de protéines de la PDB, de toutes les matrices correspondant à la matrice de référence (figure III-3, B).

III.1.2-B Le programme STAMP

Cependant, plusieurs fonctions indispensables à l'approche de conception de ligands protéiques ont été développées. STAMP recherche des motifs atomiques, quels que soient le nombre et le type d'atome de celui-ci, tandis que SPASM ne permet la recherche que de motifs de Ca ou Ca/Cy (barycentre des atomes de la chaîne latérale). La position de ce pseudo-atome Cy est dépendante de la nature du résidu et donc ne peut reproduire, pour chaque type de résidu, une orientation correcte de la chaîne latérale considérée. La figure III-3bis montre les topologies d'un motif de référence représenté par les atomes (Ca ; C β) ou (Ca ; Cy) des résidus qui le composent. Une importante divergence peut être observée selon le résidu considéré. L'orientation d'une chaîne latérale déterminée par le vecteur constitué des atomes (Ca ; C β), permet de reproduire la topologie d'un motif de résidus quelconque (à l'exception des résidus Gly) par l'ensemble des couples (Ca ; C β) des résidus de ce motif.

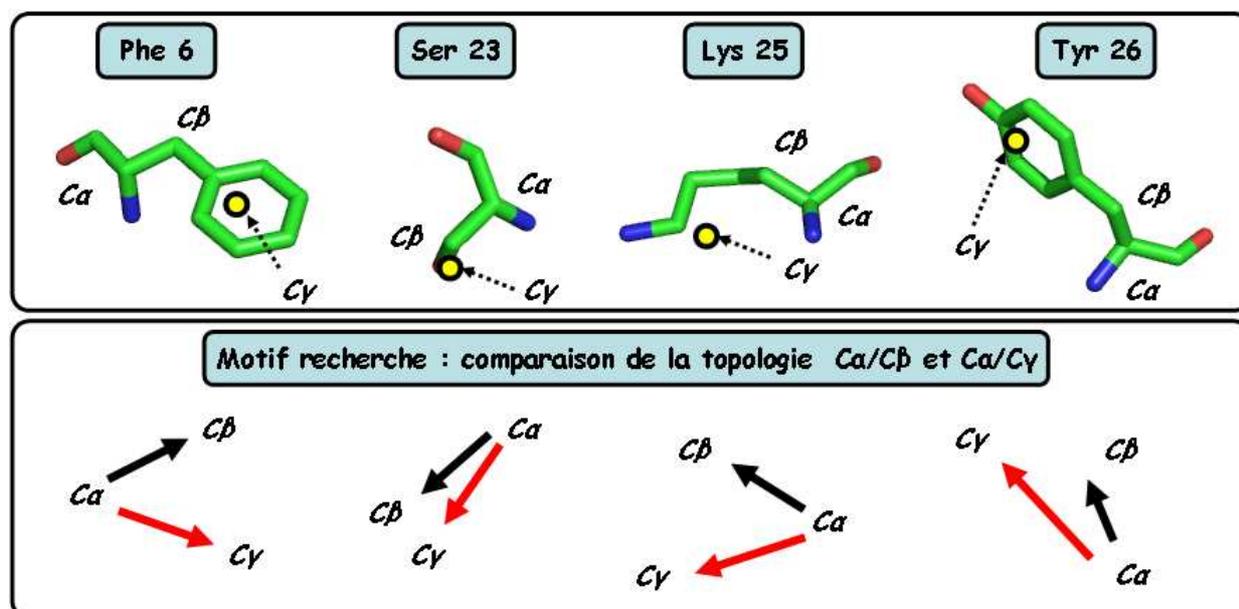


Figure III-3bis : Comparaison de la topologie du motif de recherche à partir de vecteurs Ca/C β (STAMP) et Ca/Cy (SPASM). Le motif de liaison F6/S23/K25/Y26 est représenté dans la partie supérieure, en mode « stick ». Les motifs de recherche sont schématisés dans la partie inférieure.

D'autre part, le programme SPASM effectue les recherches sur des bases de données restreintes : chaque fichier pdb est réduit à une seule structure (la première pour chaque fichier), chaque résidu étant lui même réduit aux coordonnées de ses atomes Ca et Cy. Cette approche diminue considérablement le temps d'exécution du programme. Le programme STAMP effectue la recherche sur l'ensemble des fichiers de la PDB, ainsi que sur l'ensemble des structures contenues dans chaque fichier. La recherche est donc complète puisque chaque fichier est testé dans son intégralité. Le temps d'exécution du programme est alors augmenté, toutefois de façon négligeable par rapport aux avantages procurés. A titre indicatif, pour un

ensemble de critères de recherche identiques, utilisant une station de travail identique, la durée nécessaire est multipliée par un facteur 4 à 10. Malgré cela, la recherche, à partir d'un motif à huit atomes, nécessite moins de deux heures pour trier l'intégralité des 23845 fichiers contenues dans la PDB (version août 2004, annexe A.2).

III.1.2-C Critères de recherche

La recherche de motifs topologiques, décrite précédemment, fait correspondre un ensemble de distances mesurées à un ensemble des distances fixes appartenant au motif de recherche. Deux paramètres caractérisent la topologie du motif et sont imposés par l'utilisateur (figure III-4) :

- l'incertitude maximale sur les distances ou Δd_MAX . Toute distance supérieure à cette valeur seuil est considérée comme inadéquate, excluant de ce fait le motif testé.
- l'écart moyen sur la position des atomes, entre le motif de référence et le motif sélectionné, évalué par le calcul du RMSD. Celui-ci doit être inférieur à une valeur seuil RMSD_MAX.

D'autres critères sont utilisés pour la recherche de motifs mais ne concernent pas la topologie du motif : la taille des protéines recherchées, la nature des résidus ou la taille du motif. Ceux-ci correspondent à des aspects pratiques, liés à la synthèse des protéines structurées (taille, nature des résidus, nombre de résidus à modifier).

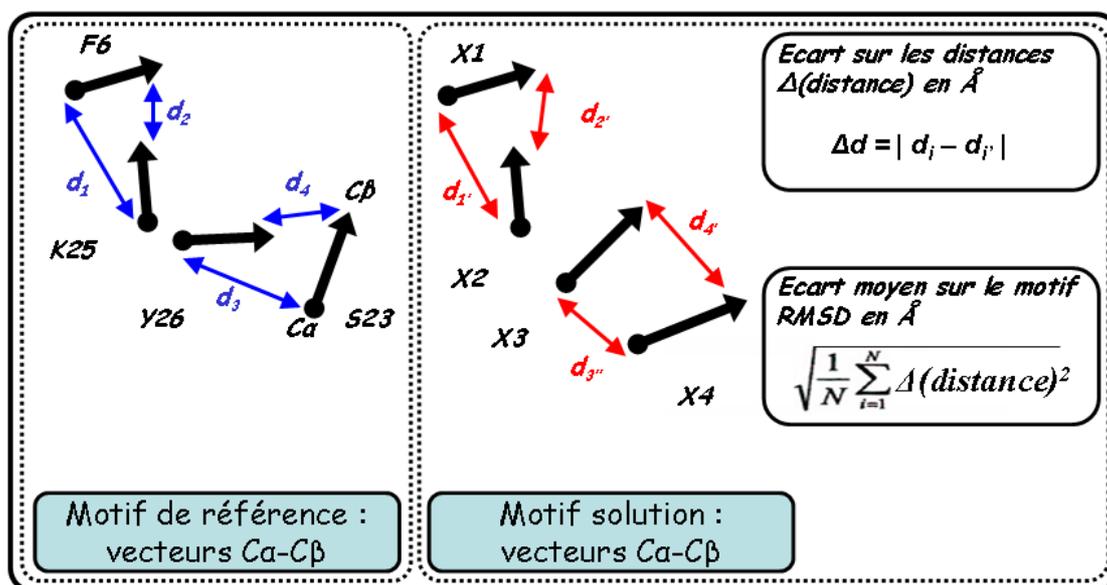


Figure III-4 : Paramètres de recherche du programme STAMP caractérisant la topologie du motif : l'écart maximum sur chacune des distances et l'écart moyen sur le motif, estimé par la valeur du RMSD. L'utilisateur impose une valeur seuil pour chacun de ces paramètres.

III.1.3 Sélection des plateformes

Les solutions obtenues lors de la recherche de motifs topologiques doivent être triées afin de prendre en compte plusieurs paramètres importants pour la réalisation de l'interaction avec la cible mais également des aspects pratiques, concernant la production de ces plateformes protéiques, représentés sur la figure III-5, partie A à E.

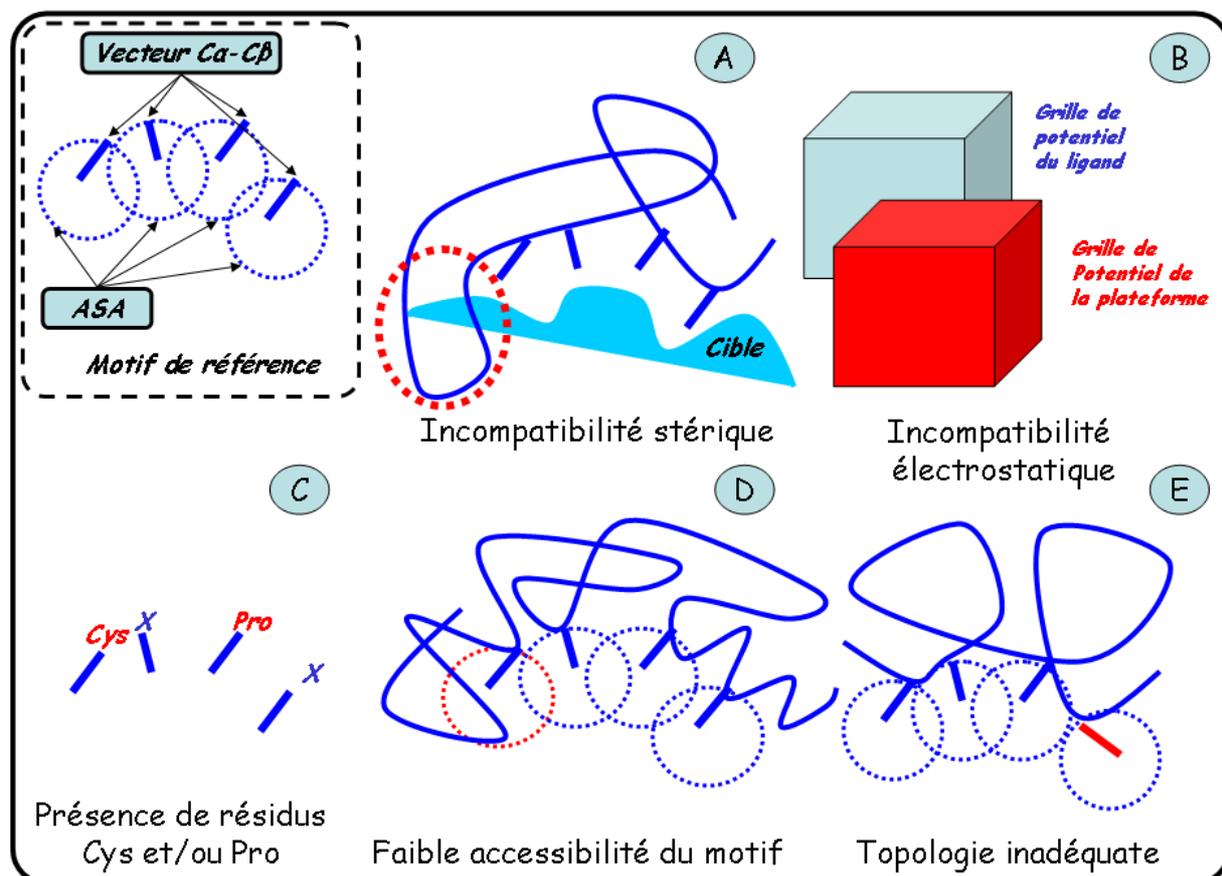


Figure III-5: Les différentes étapes de sélection des solutions issues de la recherche de motifs de topologie équivalente à celle du motif de référence. Les solutions remplissant ces différentes conditions seront utilisées comme plateformes structurales pour la présentation du motif de liaison.

III.1.3-A Critère stérique / Critère de forme

La procédure de recherche de motifs topologiques ne tient pas compte de l'interaction entre les structures identifiées comme solutions et la cible du ligand de référence. La présence d'éléments de structures particuliers au sein des structures identifiées peut occasionner des gênes stériques qui interdiraient le positionnement correct du motif de liaison par rapport à la cible. Toute structure incompatible, d'un point de vue stérique, avec la cible du ligand est éliminée (figure III-5, A). A cette fin, un filtre stérique a été utilisé lors de l'étape de recherche, par l'utilisation d'un critère %R_MAX. Son principe, schématisé sur la figure III-5bis, requiert la structure 3D d'un complexe d'interaction entre le ligand de référence et sa cible.

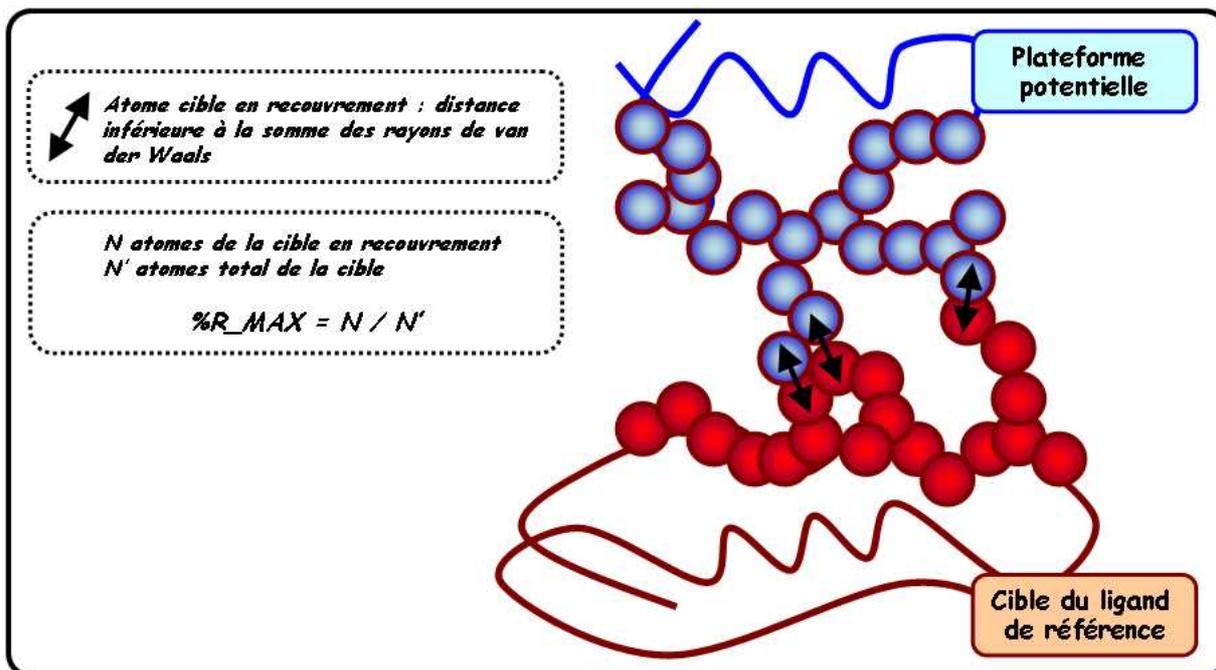


Figure III-5bis : Principe du filtre stérique développé. Les plateformes potentielles (schématisées en bleu) sont positionnées par rapport à la cible (en rouge), pour reproduire la position du motif observée dans le complexe de référence. Ceci permet d'estimer un % de recouvrement défini par le nombre d'atomes de la cible, recouverts par la structure de la plateforme considérée, rapporté au nombre total d'atomes de la cible.

Chaque structure issue de l'étape de recherche est superposée au le ligand de référence via les atomes Ca et C β de leurs motifs respectifs. Le recouvrement de la structure de la cible, par la structure de la plateforme potentielle considérée, est alors estimé par une valeur comprise entre 0 et 100%. Cette dernière correspond au nombre d'atomes de la cible, pour lesquels un atome, au moins, de la cible est situé à une distance inférieure à la somme de leurs rayons de Van der Waals, ramené au nombre total d'atomes de la cible (les atomes d'hydrogène ne sont pas pris en compte pour ce calcul). Par exemple, pour une cible moléculaire contenant 3000 atomes, une valeur de recouvrement de 1% signifie que 30 de ces atomes sont en recouvrement avec des atomes de la structure 3D testée.

III.1.3-B Critère électrostatique

Le calcul du potentiel électrostatique doit tenir compte des protéines, du solvant ainsi que des ions en solution. La loi de Poisson-Boltzmann (annexe A.3) tient compte de ces éléments en combinant les lois de l'électrostatique appliquées à une distribution rigide de charges (loi de Poisson), et une description statistique d'un milieu liquide renfermant des entités chargées (théorie de Debye-Hückel). De ce fait, les méthodes reposant sur la résolution de l'équation de Poisson-Boltzmann se montrent relativement fiables et rapides [174] pour l'étude des protéines en solution. Pour l'étude d'objets ne présentant aucune symétrie, l'utilisation d'une grille de calcul permet de représenter l'ensemble de l'espace par une série de points, épousant au mieux leurs

formes. Les milieux intérieur et extérieur sont assimilés à des continus, représentés par une constante diélectrique (*). Ces dernières sont généralement comprises entre 78 et 80 pour le solvant, et entre 1 et 4 pour l'intérieur des protéines, très hydrophobe donc faiblement polarisable. Le potentiel en un point de la grille dépend du potentiel des 6 points voisins et de la valeur de la constante diélectrique qui leur est appliquée. Les limites de la grille possédant un nombre de voisins réduit nécessite un traitement particulier dénommé conditions aux limites. L'utilisation d'une première grille plus grande que la précédente permet d'utiliser les points correspondants aux limites de la seconde grille plus petite. L'approche décrite permet de créer pour chaque protéine une grille de potentiel électrostatique, représentée sur la figure III-6. Les calculs ont été effectués par le programme APBS [203]. La comparaison de deux protéines, d'un point de vue électrostatique, revient à comparer leurs grilles de potentiel respectives, comme indiqué sur la figure III-6bis. Pour ce faire, le programme PIPSA, version 2.0 (« Protein Interaction Property Similarity Analysis ») [187, 204] a été utilisé. Ce dernier détermine un indice de similarité (ou SI, « Similarity Index »), à partir des structures 3D de deux protéines et de leurs grilles de potentiel respectives.

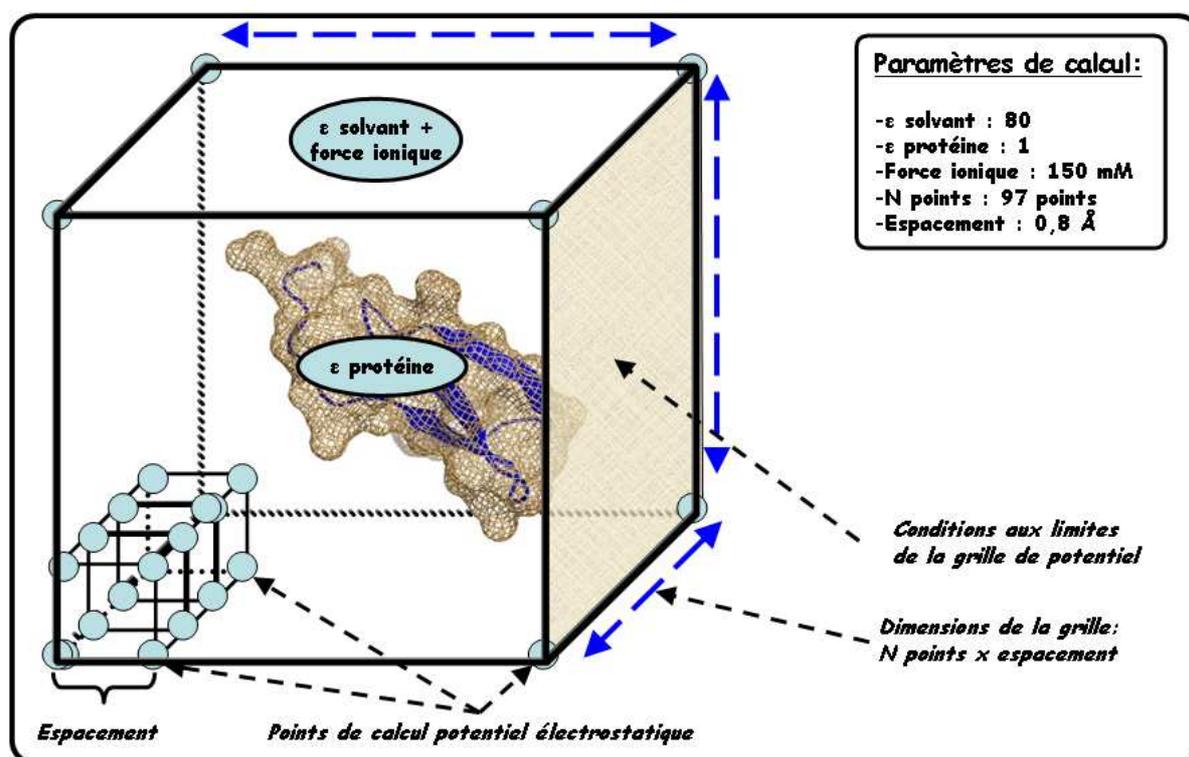


Figure III-6 : Représentation schématique du calcul de grille de potentiel électrostatique, généré par une protéine, par résolution de l'équation de Poisson Boltzmann. Les paramètres utilisés lors des calculs sont indiqués en insert.

*Constante diélectrique

Autrement dénommée permittivité relative, elle se définit comme le rapport entre la permittivité ϵ du milieu considéré et la permittivité du vide ϵ_0 . Elle décrit la réponse d'un milieu donné à un champ électrique. A un niveau microscopique, elle est liée à la polarisabilité électrique des molécules ou atomes constituant le milieu.

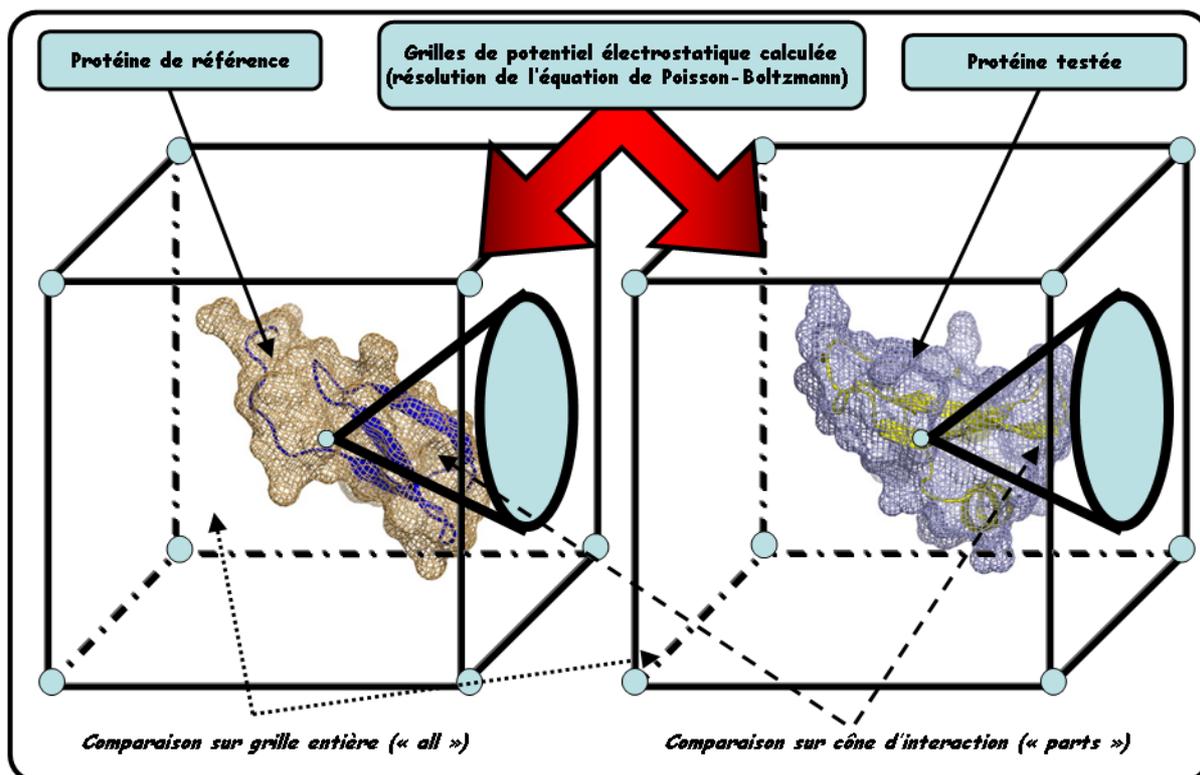


Figure III-6bis : Comparaison des grilles de potentiel électrostatique par le programme PIPSA [187, 204], par le calcul de l'indice de similarité (SI). Ce dernier correspond à un coefficient de corrélation. Le potentiel peut être comparé sur l'ensemble de la grille (« all »), ou sur la région d'interaction seule, définie par un cône (« parts »).

Le SI, calculé sur l'ensemble des grilles (« all ») ou sur la région interface seule (« parts »), correspond à un coefficient de corrélation du potentiel électrostatique. Une valeur de SI égale à 1 signifie que les deux grilles sont identiques, une valeur de -1 que les grilles sont anti-corrélées.

III.1.3-C Critères généraux

Les solutions redondantes sont éliminées. Les structures multiples (notamment résolues par RMN) appartenant à un même fichier sont réduites à une seule. La structure dont le RMSD (calculé par rapport au motif de référence) est le plus faible, est conservée.

Les solutions contenant dans le motif des résidus Cys ou Pro sont éliminées (figure III-5, C). Les résidus cystéine sont connus pour leur importance dans le maintien de la structure 3D par la réalisation de ponts disulfures [62]. Les résidus proline et glycine sont impliqués dans la formation d'éléments de structure secondaire, dont divers coudes, mais défavorables à la formation d'hélices et de feuillets [205, 206, 207, 208]. Les paramètres de Chou et Fasman sont présentés dans le tableau III-7.

Hélices α				Feuillets β				Coude divers			
Glu	1,51	Asp	1,01	Val	1,70	Arg	0,93	Asn	1,56	Trp	0,96
Met	1,45	His	1,00	Ile	1,60	Asn	0,89	Gly	1,56	Arg	0,95
Ala	1,42	Arg	0,98	Tyr	1,47	His	0,87	Pro	1,52	His	0,95
Leu	1,21	Thr	0,83	Phe	1,38	Ala	0,83	Asp	1,46	Glu	0,74
Lys	1,16	Ser	0,77	Trp	1,37	Ser	0,75	Ser	1,43	Ala	0,66
Phe	1,13	Cys	0,70	Leu	1,30	Gly	0,75	Cys	1,19	Met	0,60
Gln	1,11	Tyr	0,69	Cys	1,19	Lys	0,74	Tyr	1,14	Phe	0,60
Trp	1,08	Asn	0,67	Thr	1,19	Pro	0,55	Lys	1,01	Leu	0,59
Ile	1,08	Pro	0,57	Gln	1,10	Asp	0,54	Gln	0,98	Val	0,50
Val	1,06	Gly	0,57	Met	1,05	Glu	0,37	Thr	0,96	Ile	0,47

Tableau III-7 : Paramètres de Chou et Fasman [208] correspondant à la propension pour chaque acide aminé à favoriser ou défavoriser un type particulier d'éléments de structure secondaire (hélice, feuillet ou coude).

Le remplacement des résidus Pro, Gly et Cys peut avoir des conséquences imprévisibles et probablement dommageables pour la structure des plateformes. Toutefois, les résidus Gly permettent d'accéder à un espace conformationnel plus important que les résidus Pro, comme en témoignent leurs diagrammes de Ramachandran respectifs, présentés sur la figure III-7bis. Pour cette raison, le remplacement de résidus Gly peut être envisagé. Toutefois la substitution de plus d'une Gly pourrait être dommageable pour la structure 3D.

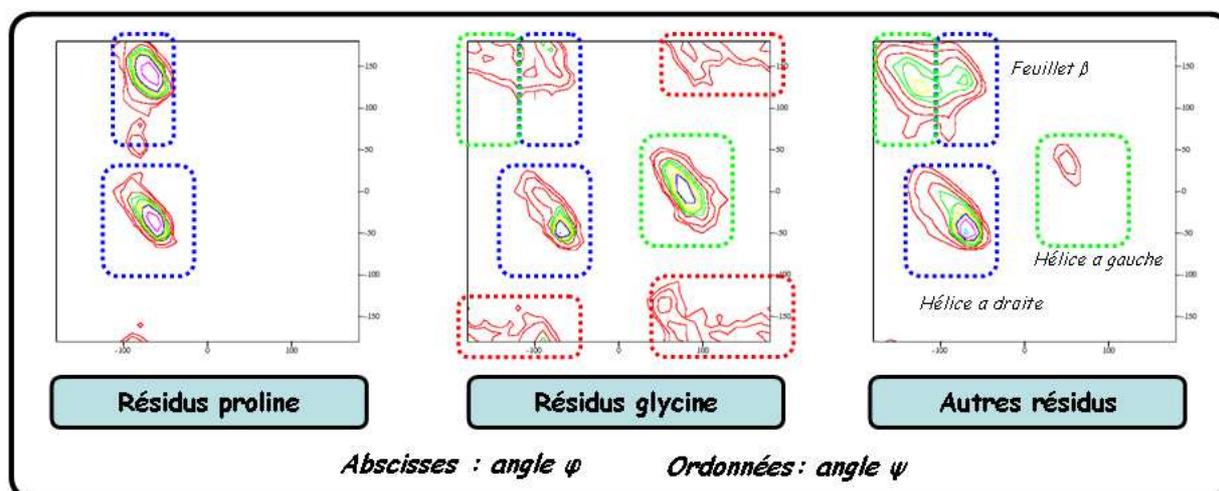


Figure III-7bis : Les diagrammes de Ramachandran des différents résidus. Les zones bleues sont accessibles par l'ensemble des résidus naturels. Les zones vertes sont accessibles par l'ensemble des résidus, à l'exception des résidus Pro. Les zones rouges ne sont accessibles qu'aux résidus Gly. Les diagrammes présentés ont été réalisés par le programme O2D, développé par G.J. Kleywegt [209].

Les solutions dont le motif est inaccessible au solvant sont éliminées (figure III-5, D). Le calcul de la surface accessible au solvant est effectué en utilisant l'algorithme décrit par Lee et Richards [136].

Les motifs dont l'orientation des chaînes latérales est divergente sont éliminés (figure III-5, E). Selon les critères de recherche utilisés, lorsque trop de degrés de libertés sont accordés, le maintien de la topologie n'est plus garanti (figure III-5, E).

Ces dernières étapes de sélection, correspondant à des critères généraux, sont réalisées de façon systématique *in silico* par les programmes MAKEINP et MAKESEL développés au cours de cette thèse (annexe A.4).

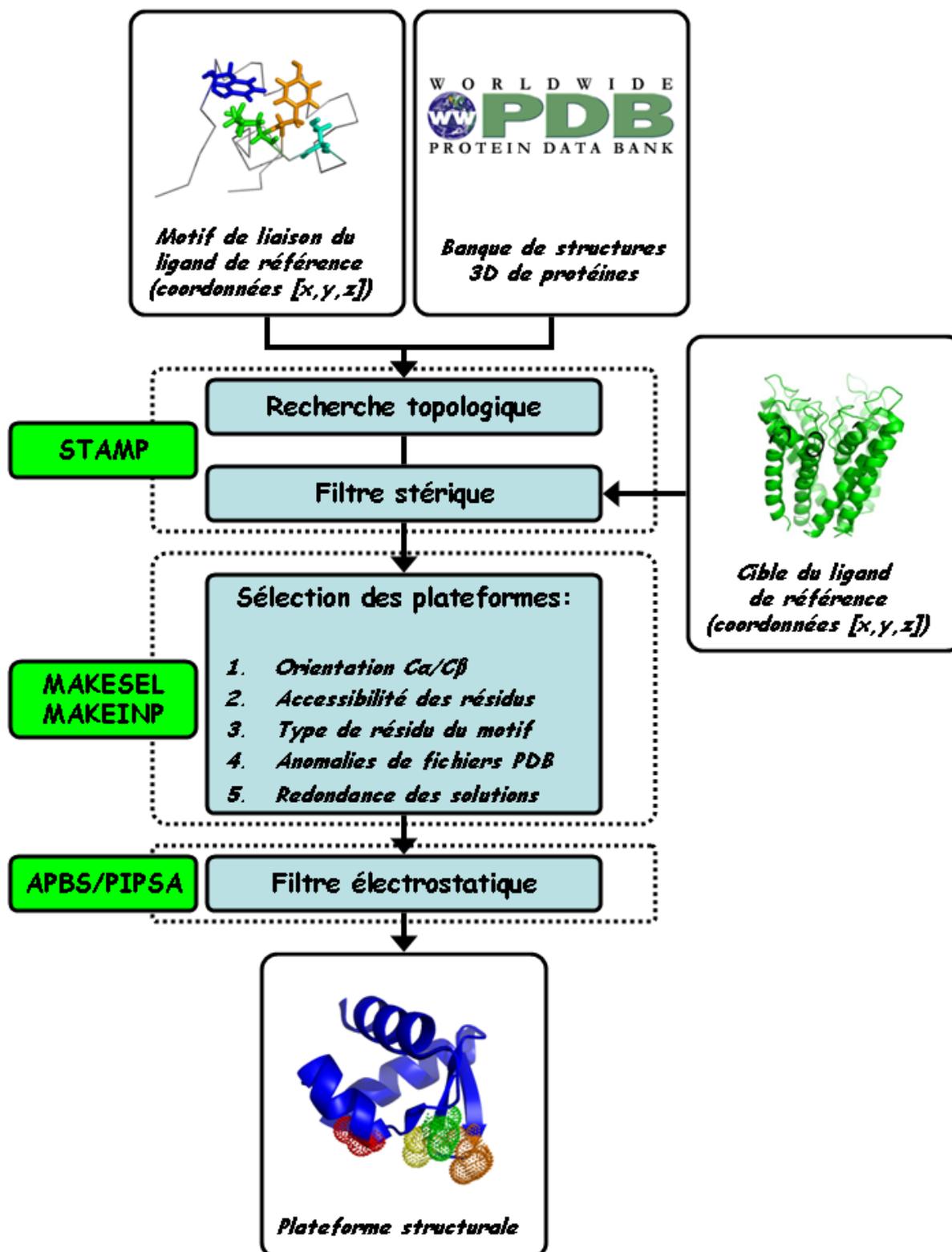


Figure III-8 : Schéma des diverses étapes nécessaires à l'approche de conception de ligand développée. Celle-ci requiert la structure d'un ligand de référence et d'une base de donnée de structures, pour la recherche de motifs topologiques. L'application du filtre stérique nécessite, en outre, la structure de la cible du ligand.

III.1.4 Conclusion

L'approche développée, de transfert de motif fonctionnel sur une plateforme adéquate, est réalisée par plusieurs étapes successives, représentées sur la figure III-8 :

- 1) recherche de motifs topologiques similaires (STAMP)
- 2) application d'un filtre stérique (STAMP)
- 3) application de critères généraux (MAKESEL/MAKEINP) :
 - 3.1) orientation des chaînes latérales
 - 3.2) accessibilité des résidus du motif
 - 3.3) choix du type de résidus
 - 3.4) anomalies de fichiers PDB
 - 3.5) identification des solutions uniques
- 4) application du filtre électrostatique (APBS/PIPSA, version 2.0)

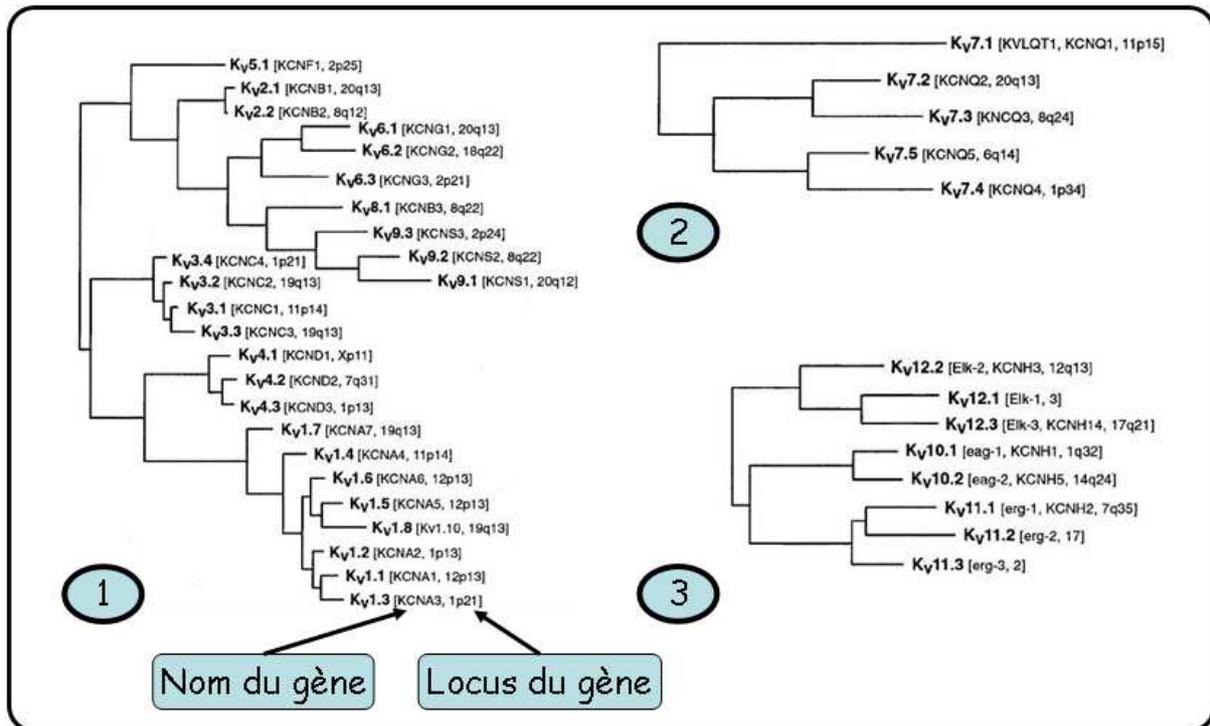


Figure III-9: Représentation sous forme d'arbres phylogénétiques des canaux de la famille Kv [215]. Ceux-ci sont regroupés en trois familles distinctes, d'après l'alignement des séquences des cœurs hydrophobes de chacun par CLUSTAL W (<http://www.ebi.ac.uk/clustalw/>).

III.2 Le système de référence toxine-canal potassique

Les canaux ioniques exercent des rôles majeurs dans le fonctionnement des organismes vivants pour le maintien des conditions cellulaires ou extracellulaires, pour la production d'énergie ou pour la transduction d'informations. Ils constituent, de ce fait, des cibles privilégiées pour de nombreuses toxines, protéiques ou non, synthétisées par différents organismes ou micro-organismes. A l'heure actuelle, plusieurs dizaines de toxines connues pour leurs propriétés de bloqueurs de canaux ioniques, avec une très haute affinité (nM à pM), ont été caractérisées aux niveaux structural et fonctionnel [210, 211, 212, 213, 214]. Les complexes toxine-canal, pour ces raisons, constituent des systèmes intéressants pour la validation de notre méthode.

III.2.1 Les canaux de la famille Kv

L'interaction de la BgK [200, 214] avec les canaux potassiques sensibles au potentiel (famille Kv1), étudiée depuis de nombreuses années au département, a été choisie pour la mise en application de notre approche de conception de ligands protéiques.

III.2.1-A Du gène au canal

Les canaux potassiques sont regroupés en 4 classes par des considérations structurales, en fonction du nombre de segments transmembranaires qu'ils possèdent. La famille des canaux Kv regroupe tous les canaux potassiques à 6 segments transmembranaires sensibles au potentiel. Chaque canal fonctionnel est formé par l'assemblage de quatre monomères, identiques ou pas, codés par différents gènes, présentés sur la figure III-9. Ceux-ci sont soumis désormais à une nomenclature standard (développée par la « HUmAn Genome Organisation » ou HUGO) du fait des nouvelles découvertes de canaux potassiques depuis le séquençage du génome [215]. La classification des gènes tient compte de leur structure (nombre de segments transmembranaires), de leurs caractéristiques fonctionnelles (e.g. sensibilité au potentiel) et de leurs relations phylogénétiques (figure III-9) [215]. A l'heure actuelle, dans l'espèce humaine, 38 gènes codant pour les canaux de cette famille ont été identifiés. Chaque canal est formé par l'assemblage de quatre monomères, identiques ou non, mais appartenant à la même sous-famille (e.g. Kv1, Kv2). L'ensemble des combinaisons possibles pour la sous-famille Kv1 comprenant 8 gènes, correspond à un total de 3024 possibilités (en tenant compte des différentes permutations possibles), témoignant d'une très grande diversité des canaux. Dans notre cas d'étude, le système de référence ne concerne que le canal homo-tétramérique Kv1.2.

III.2.1-B Organisation fonctionnelle des canaux de la famille Kv

Les canaux potassiques de la famille Kv ne sont fonctionnels qu'à l'état de tétramère. Chaque monomère (figure III-10, A), est constitué d'une partie transmembranaire (sous-unité α) et d'une partie intracellulaire (une sous-unité β et un domaine de tétramérisation T1). L'ensemble des quatre sous-unités α forme le pore et confère la dépendance du canal au potentiel. Les sous-unités β sont responsables de l'inactivation et de la modulation de l'activité des sous-unités α . Chaque sous-unité α comprend 6 segments transmembranaires, de S1 à S6, schématisée par la figure III-10, partie A, le pore du canal étant formé par les régions P, comprises entre les segments S5-S6 de chaque sous-unité, schématisé par la figure III-10, partie B. Chaque région P contient plusieurs éléments dont les tourelles (boucles extracellulaires relativement flexibles), l'hélice du pore et le pore lui-même. La partie correspondant aux éléments extracellulaires des segments S5 à S6 est la cible des toxines [216].

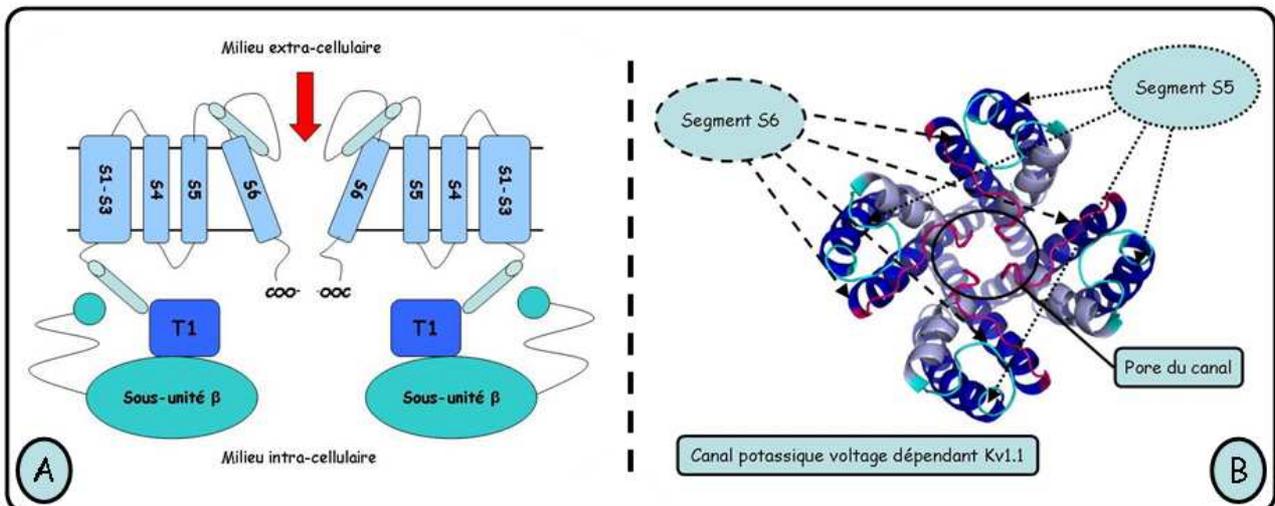


Figure III-10 : A, Représentation simplifiée des canaux de la famille Kv ; en bleu clair, la sous-unité α , en bleu foncé, le domaine T1, et en cyan, la sous-unité β . B, Vu de dessus du pore du canal Kv1.1 ; pour chaque monomère, les segments S5 et S6 sont représentés en bleu foncé, les tourelles en cyan, l'hélice du pore en bleu clair et le pore en magenta. [216].

III.2.1-C Homologie de séquence des canaux potassiques

Les canaux potassiques de la famille Kv1 présentent, entre eux, une très grande homologie de séquence, ainsi qu'avec d'autres canaux appartenant à différentes espèces, mammifères ou bactériennes. La comparaison des séquences primaires des parties extracellulaires de différents canaux (dont les canaux Kv1.1 et Kv1.2 humains), comprise entre les segments S5 et S6, montre une grande variabilité dans la séquence des tourelles mais une très grande conservation de la séquence du pore lui-même. Les séquences sont indiquées dans le tableau III-11. Ainsi, du fait de la forte conservation de séquence du pore et malgré la grande variabilité de séquence des tourelles, il est raisonnable de penser que la structure 3D de la partie extracellulaire des canaux est, elle-même, très conservée.

	-> S5	Tourelle	Hélice du pore	Pore		S6 ->
KcsA	Y L AV L A	ER G AP G A Q L I T Y	PR A L W SV E T A	TT V GY G	DL Y P V T L	W G R L
hKv1.1	SA V Y F A	E A E A ES H F S S I	PD A FW A V V S M	TT V GY G	DM Y P V T I	GG K I
mKv1.1	SA V Y F A	E A E A ES H F S S I	PD A FW A V V S M	TT V GY G	DM Y P V T I	GG K I
hKv1.2	SA V Y F A	E A D E R E S Q F P S I	PD A FW A V V S M	TT V GY G	DM V P T T I	GG K I
hKv1.6	SA V Y F A	E A D D D D S L F P S I	PD A FW A V V T M	TT V GY G	DM Y P M T V	GG K I
hKv1.3	SA V Y F A	E A D D P T S G F S S I	PD A FW A V V T M	TT V GY G	DM H P V T I	GG K I
mKv1.3	SA A Y F A	E A D D P S S G F N S I	PD A FW A V V T M	TT V GY G	DM H P V T I	GG K I
hKv1.4	SA V Y F A	E A D E P T H F Q S I	PD A FW A V V T M	TT V GY G	DM K P I T V	GG K I
hKv1.5	SA V Y F A	E A D N Q G T H F S S I	PD A FW A V V T M	TT V GY G	DM R P I T V	GG K I

Tableau III-11 : Alignement des séquences des parties extracellulaires de différents canaux potassiques (les lettres qui précèdent chaque canal correspondent à m pour mouse, h pour human).

III.2.1-D Structure tridimensionnelle des canaux

Au début de ce travail de thèse, aucune structure 3D de canal appartenant à la famille Kv n'ayant été résolue, les étapes de sélection de structures protéiques aptes à l'interaction, ont été réalisées à partir d'un modèle d'interaction entre le récepteur Kv1.1 et la BgK (modèle affiné par la suite) [217, 218]. La structure du récepteur a, elle-même, été modélisée à partir d'un canal bactérien (KcsA, *Streptomyces Lividans*) de structure connue (1BL8 [219]). Le canal KcsA est un canal potassique possédant deux segments membranaires. Leur structure correspond aux segments S5 et S6 des canaux de la famille Kv. La structure du canal Kv1.2 a été résolue par radiocristallographie (2A79 [220]), permettant de vérifier la structure de notre modèle. La figure III-12 montre la comparaison du modèle du canal Kv1.1 avec la structure cristallographique du canal Kv1.2. La superposition des chaînes principales, *via* leurs atomes Ca, des segments S5-S6 des deux canaux, a montré une importante similarité de structure (RMSD = 1.05Å), validant, *a posteriori*, notre approche basée sur le modèle du récepteur Kv1.1.

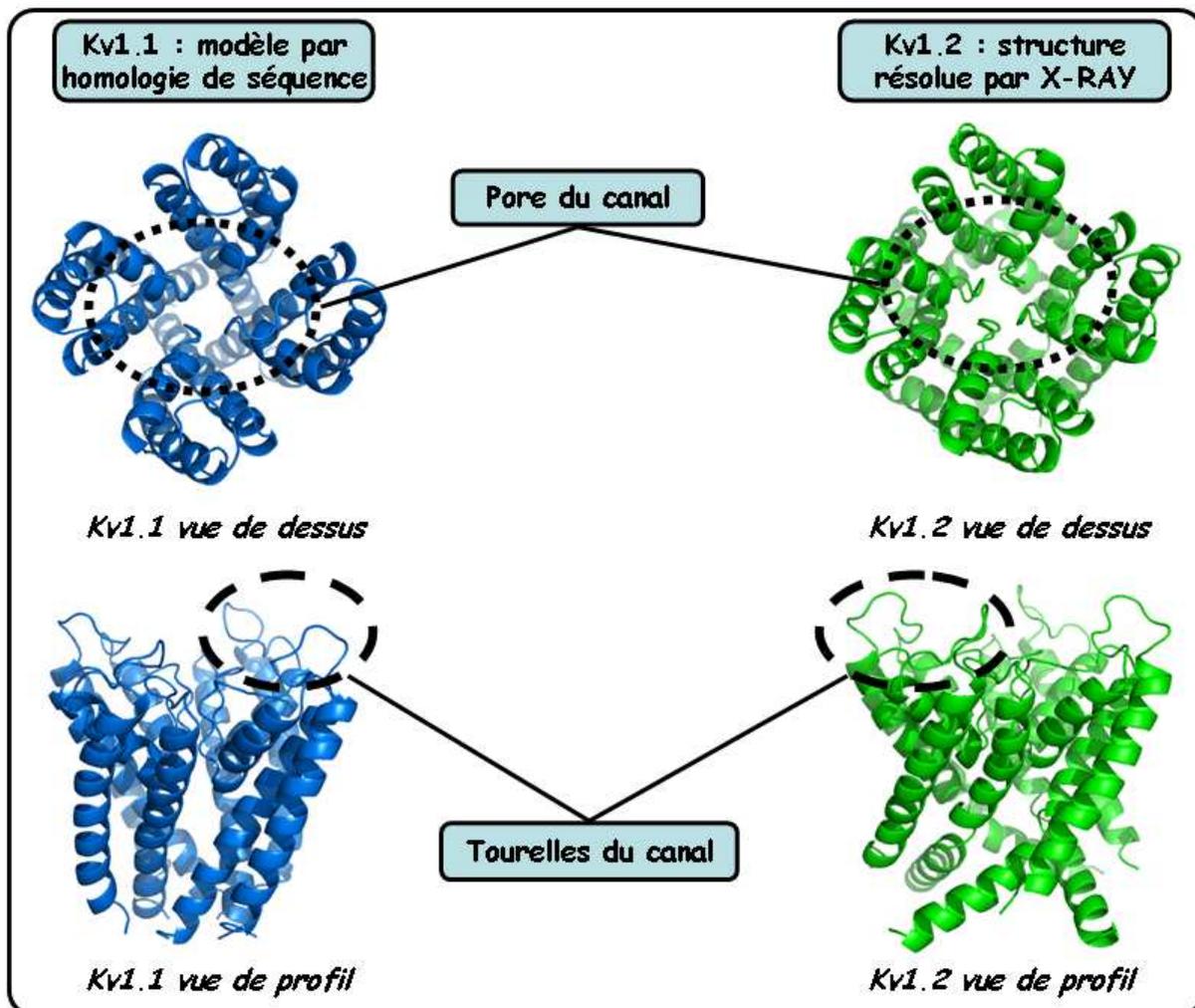


Schéma III-12 : Comparaison du modèle du canal Kv1.1 [217, 218] (en bleu) et de la structure cristallographique du canal Kv1.2 [220] (en vert). Les canaux sont représentés en mode « cartoon ».

III.2.2 Les toxines spécifiques des canaux Kv1

Dans diverses espèces animales, de nombreuses toxines bloquant des canaux ioniques avec une très haute affinité, ont été identifiées et caractérisées durant les vingt dernières années. Les canaux potassiques voltage-dépendant de la famille Kv1 sont les cibles de nombreuses toxines de serpents, scorpions, anémones de mer ou cônes, relativement différentes, en taille, composition ou structure secondaire/tertiaire.

III.2.2-A Mode d'interaction des bloqueurs de canaux potassiques

Malgré cette diversité, de nombreux « bloqueurs » des canaux Kv1 (BgK, ChTx, AgTx2, etc...) possèdent un motif minimum commun de deux résidus, nécessaire à la liaison [200, 216, 217,221]. Ce motif, communément dénommé diade, est composé d'un résidu lysine (K) et d'un résidu aromatique ou hydrophobe, présentés dans une topologie similaire malgré des repliements et des positions dans la séquence différents. Les diades de liaisons des toxines BgK, ChTx et AgTx2 sont représentées sur la figure III-13. Seule leur proximité spatiale et leur topologie déterminent leur fonction.

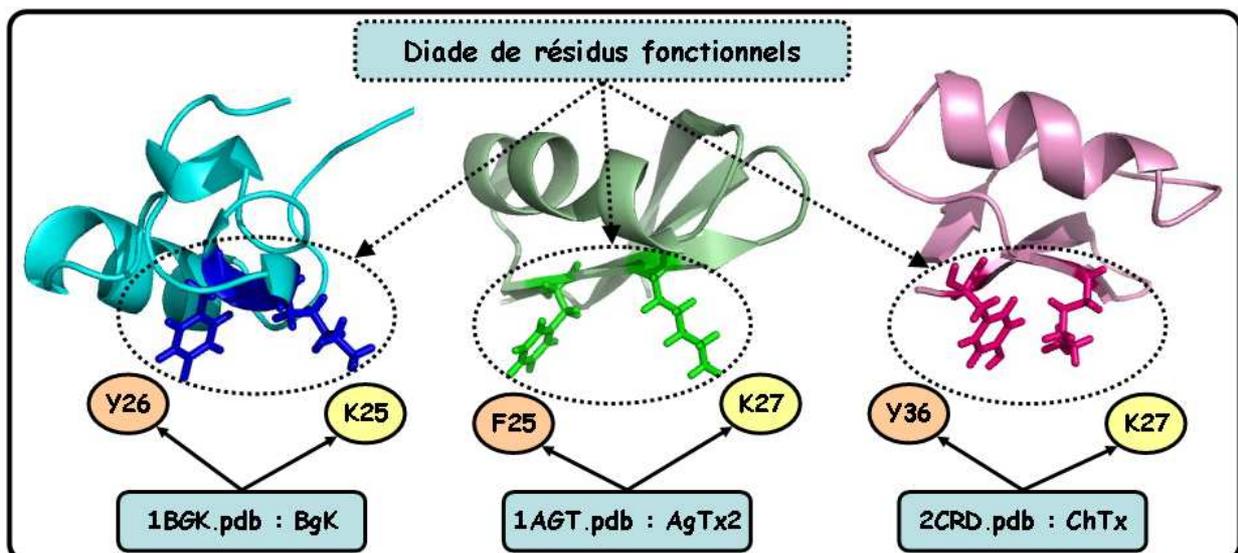


Figure III-13 : représentation de trois toxines, BgK [200] (en bleu), AgTx2 [222] (en vert) et ChTx [223] (en fuschia), liant les canaux potassiques voltage dépendant par leur diade fonctionnelle, identifiée par « alanine scanning ». La diade comprend un résidu hydrophobe (souvent aromatique) et un résidu basique (souvent le résidu Lys). Les toxines sont représentées en mode « cartoon » tandis que chaque diade est représentée en mode « ribbon ».

III.2.2-B La toxine *Bunodosoma granulifera* K⁺ channel voltage-gated ou BgK

La BgK est la première toxine d'anémone de mer isolée et caractérisée [214]. Elle bloque les canaux potassiques de la famille Kv notamment Kv1.1, Kv1.2, Kv1.3 et Kv1.6 [224, 225] avec des affinités variables. Elle a été utilisée comme ligand de référence dans notre approche. Les affinités mesurées [Alessandri-Haber et al., 1999] par des expériences de compétition avec l'a

dendrotoxine radiomarquée (^{125}I -DTX) sont : de **34pM** sur le canal Kv1.1 ; de **66pM** sur le canal Kv1.2 ; de **777pM** sur le canal Kv1.3 ; et de **13pM** sur le canal Kv1.6. La BgK [200] est constituée de 37 résidus, dont 6 cystéines, très conservées parmi les toxines d'anémones de mer liant les canaux, dont les séquences sont indiquées dans le tableau III-14.

Toxines	Séquence des toxines d'anémone de mer
BgK	-VCRDWFKETACRHA ^C SLGNCR ^S TSQKYRAN-CAKT ^C CEL ^C
AeK	-GCKDNFSANTCKHVKANNN ^C -GSQKYATN-CAKT ^C CGK ^C
AsKS	-ACKDNFAAATCKHVKENKN ^C -GSQKYATN-CAKT ^C CGK ^C
HmK	RTCKDLIPVSE ^C ----TDIRCRTSMKYRLNL ^C CRKT ^C CGS ^C
ShK	RSCIDTIPKSR ^C ----TAFQCKHSMKYRLSF ^C CRKT ^C CGT ^C

Tableau III-14: alignement des séquences des toxines courtes d'anémone de mer, bloquant les canaux potassiques. Les ponts disulfures ainsi que la diade K25/Y26 sont très conservés.

La conservation d'un même motif de ponts disulfures et de la même diade n'est en revanche pas corrélée avec le repliement de ces toxines. La conservation de la fonction est donc liée à la présentation d'un même motif, dans une topologie similaire, sur des structures 3D différentes.

III.2.2-C Motif de liaison de la BgK

Diverses expériences « d'alanine scanning », effectuées sur l'ensemble des résidus de la BgK [200], ont permis d'identifier les résidus énergétiquement importants pour l'interaction de la BgK avec les canaux Kv1.1 et Kv1.2, représentés à la surface de la structure 3D de la BgK sur la figure III-15. A partir de ces informations, un motif de liaison pour le canal Kv1.2 a été défini. Ce dernier, discontinu en séquence, permet d'étudier la validité de notre approche, indépendante des structures primaire et secondaire. Les résultats obtenus montrent que **le résidu K25** provoque la perte d'affinité la plus importante, quel que soit le canal considéré (facteur >1000). **Le résidu Y26**, appartenant à la diade entraîne, de la même façon, des pertes d'affinité importantes, notamment sur le canal Kv1.2. Ces deux résidus seront systématiquement présents dans notre motif de liaison. En l'absence de ces deux résidus (mutant K25A/Y26A), l'affinité est diminuée d'un facteur approximatif de 10^4 sur le canal Kv1.1 et d'un facteur approximatif de 10^5 sur le canal Kv1.2. D'autres résidus ont été pris en compte dans ce motif, limité dans un premier temps à 4 résidus. **Le résidu S23** a été choisi pour son importance énergétique, quel que soit le récepteur. **Le résidu F6**, surtout impliqué dans l'interaction avec le canal Kv1.2, a été choisi du fait de son éloignement de la diade dans la séquence primaire, respectant ainsi notre condition sur la discontinuité du motif. Au final, le motif F6/S23/K25/Y26 choisi pour être transféré sur une plateforme structurale adéquate, est présenté sur la structure de la BgK, figure III-15bis.

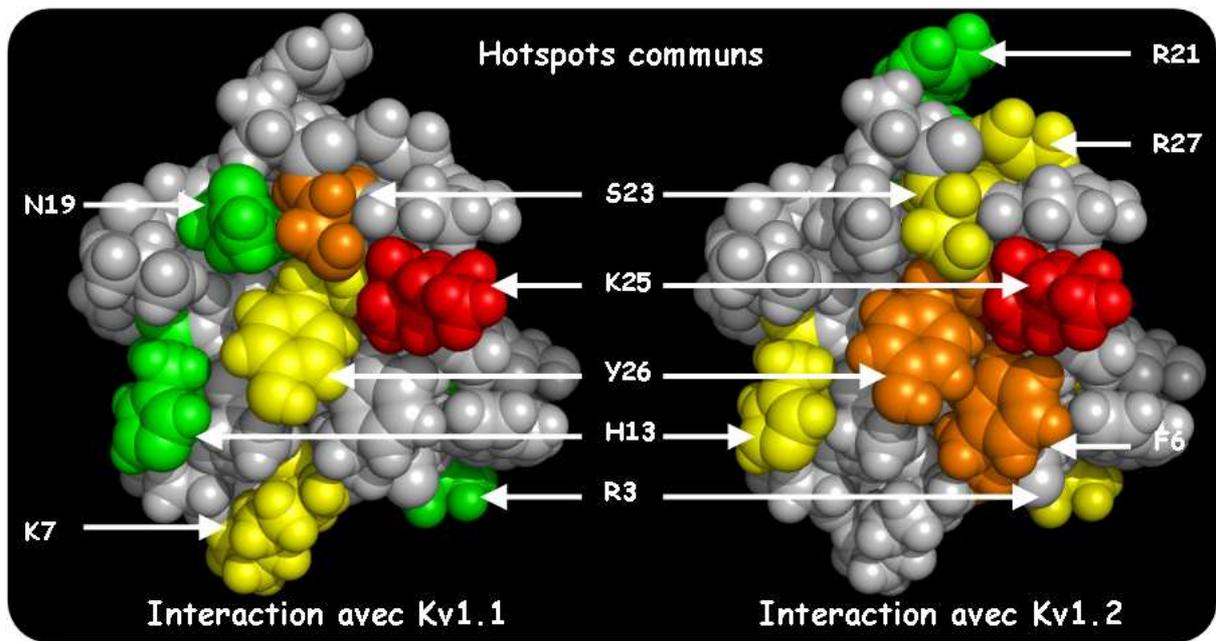


Figure III-15 : Résidus de la BgK, énergétiquement impliqués dans l'interaction avec Kv1.1 (à gauche) et Kv1.2 (à droite). La couleur est fonction du ratio $r = K_i \text{ muté} / K_i \text{ natif}$; en rouge, $r > 1000$; en orange, $1000 > r > 100$; en jaune, $100 > r > 10$; en vert, $10 > r > 5$.

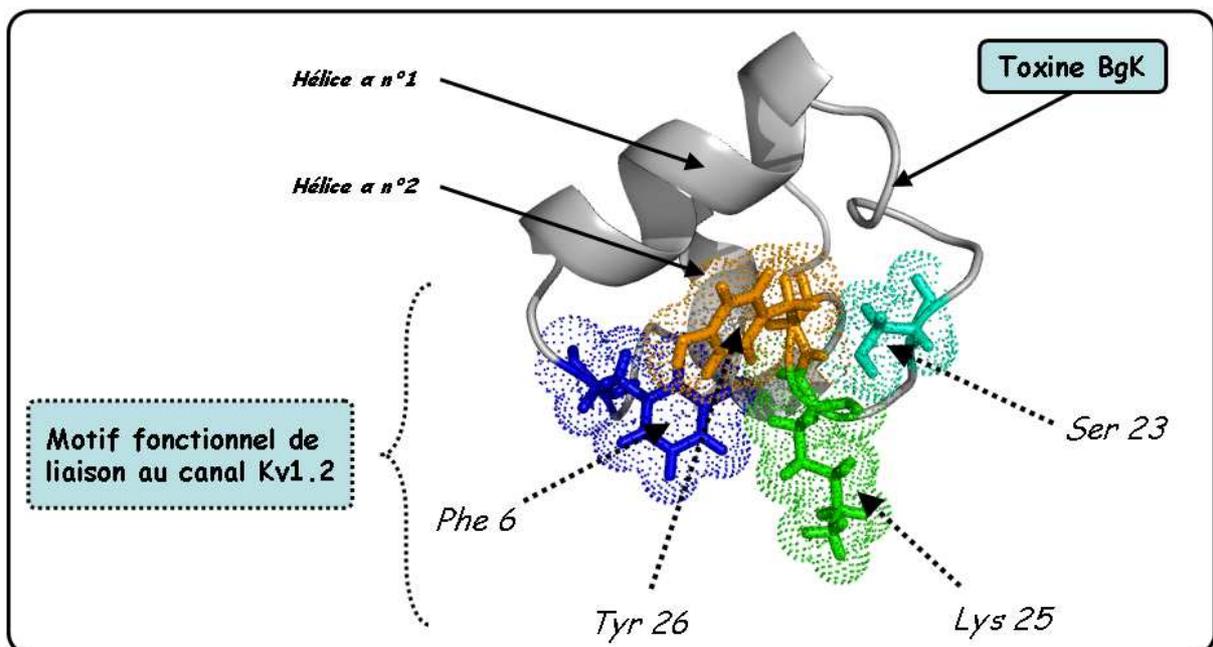


Figure III-15bis : Motif fonctionnel de liaison de la BgK au canal Kv1.2 identifié par alanine scanning. La BgK est représentée en mode « cartoon » en gris. Les résidus du motif sont représentés en mode « ribbon / dot », F6 en bleu, S23 en cyan, K25 en vert et Y26 en orange.

III.2.3 L'interaction entre la BgK et les canaux de la famille Kv

Pour la mise en application de notre méthodologie, nous avons utilisé un modèle d'interaction de la BgK avec le modèle du canal Kv1.1, décrit précédemment.

III.2.3-A Détermination de la position de la BgK dans le canal Kv1.1

Ce modèle, a été déterminé, par l'utilisation de données obtenues par la méthode des cycles thermodynamiques, permettant d'estimer les proximités entre les résidus du ligand (BgK) et ceux du canal (Kv1.1). Cette méthode d'analyse quantifie l'influence d'une substitution, effectuée au niveau d'un partenaire de l'interaction, sur la variation d'affinité provoquée par une seconde substitution, effectuée au niveau de l'autre partenaire. L'approche est schématisée sur la figure III-16.

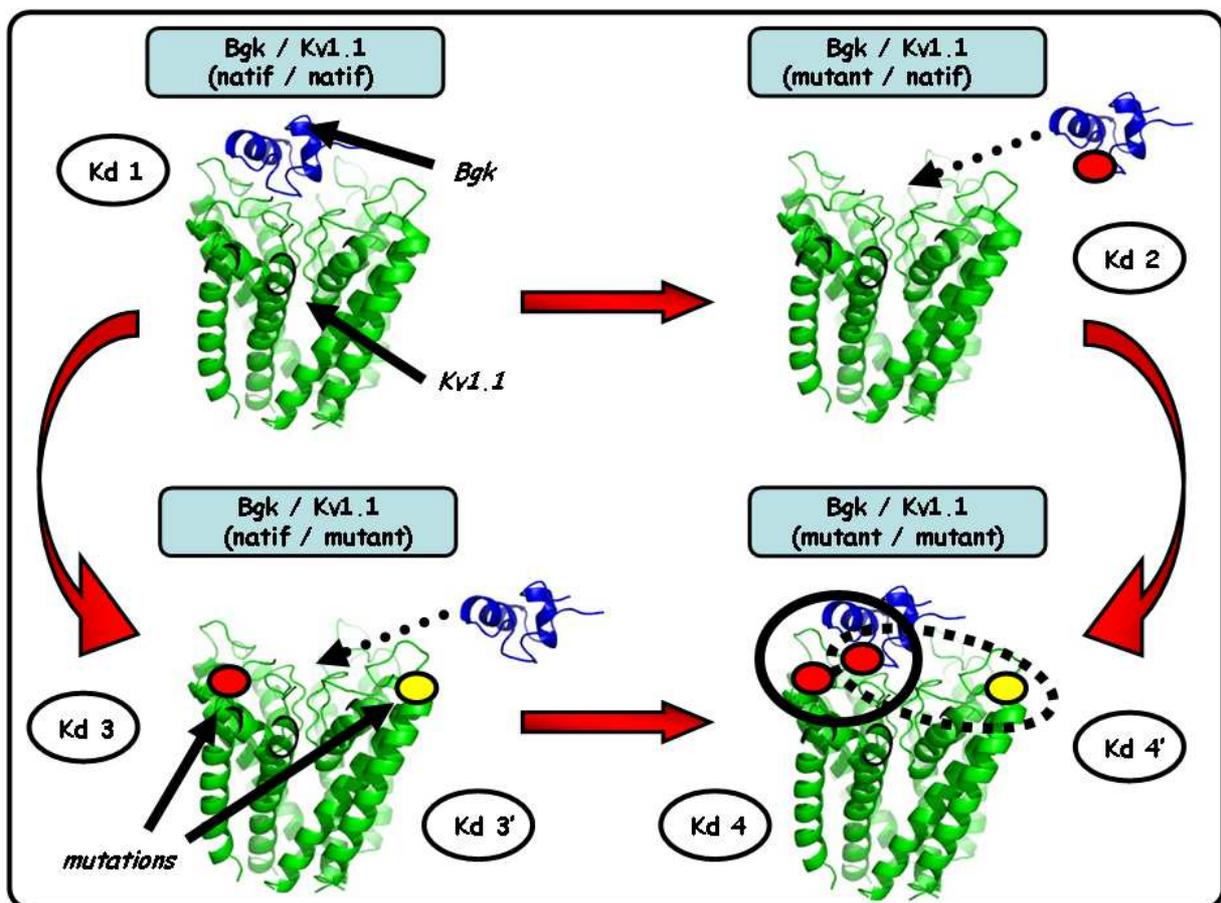


Figure III-16: Principe des cycles thermodynamiques de doubles mutants. Le canal Kv1.1, en vert, et la BgK, en bleu, sont représentés en mode « cartoon ». Les mutations sont représentées en rouge ou jaune, selon l'éloignement du site de liaison.

Pour chaque cycle de double mutation, les valeurs des constantes de dissociation permettent de calculer l'énergie de couplage. Lorsque cette dernière est nulle, les mutations sont considérées comme indépendantes, reflet d'un éloignement spatial entre les positions mutées (cas de la mutation en jaune sur la figure III-16). Lorsque qu'elle n'est pas nulle, les substitutions ont des

effets non additifs. De ce fait les positions mutées sont considérées comme proches ou interagissant entre elles (cas de la mutation en rouge sur figure III-16). L'intensité du coefficient de couplage, transformée en contraintes de distance, a permis de positionner la BgK par rapport au canal Kv1.1.

III.2.3-B Positionnement de la BgK : confirmation du modèle

Au cours de cette thèse, une étude structurale par RMN (2A9H [226]) a permis de déterminer la structure du complexe de la ChTx (ligand des canaux de la famille Kv1) avec un canal KcsA (homologue des canaux potassiques Kv1) modifié. Les modifications ont consisté à insérer les résidus critiques pour la liaison avec les canaux potassiques eucaryotes. La comparaison de notre modèle d'interaction avec la structure de ce complexe (impliquant la ChTx, ligand spécifique du canal Kv1.1 avec une affinité de 104 nM), a été réalisée par superposition de la BgK sur la ChTx par leur diade fonctionnelle respective : K25/Y26 pour la BgK et K27/Y36 pour la ChTx. Les structures superposées sont présentées sur la figure III-17. Malgré des repliements des toxines BgK et ChTx nettement différents, le positionnement de la diade fonctionnelle de la BgK, dans notre modèle d'interaction, est relativement similaire à celle de la ChTx. Notre approche, basée sur le modèle d'interaction BgK-Kv1.1 a été validée, *a posteriori*, par la comparaison de ces deux structures 3D. La modélisation par homologie, nous a donc permis de mettre en application notre méthode de conception en incluant un filtre stérique, afin de sélectionner les structures aptes à porter le motif de liaison F6/S23/K25/Y26 défini précédemment.

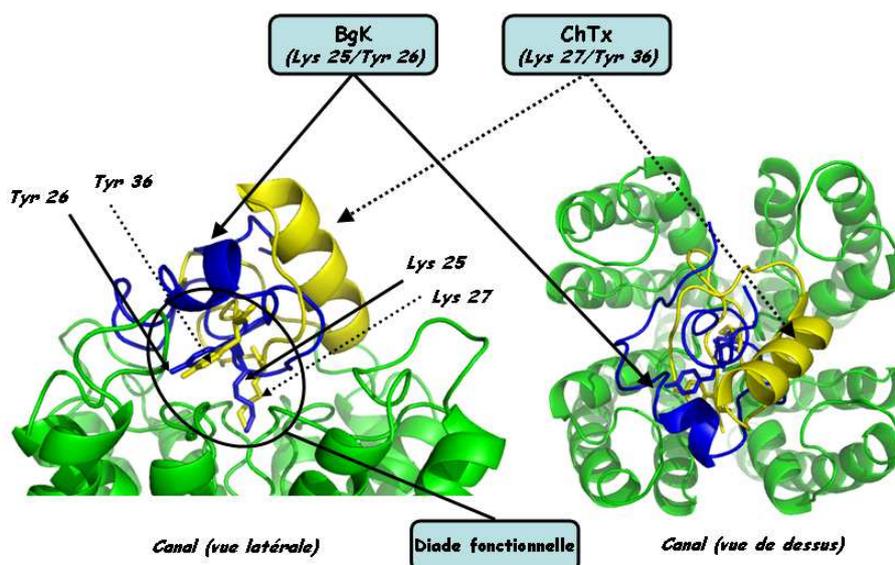


Schéma III-17 : Modèle d'interaction BgK-Kv1.1 comparé à la structure du complexe ChTx-KcsA mutant (2A9H, [226]). Les toxines sont représentées en mode « cartoon », la BgK en bleu et la ChTx en jaune. Les résidus Lys (25 et 27) et Tyr (26 et 36) formant la diade fonctionnelle sont représentés en mode « stick ».

III.2.4 Conclusion

Le système de référence que nous avons choisi correspond à l'interaction de la toxine BgK avec le canal potassique, voltage-dépendant, Kv1.2. Les résidus énergétiquement importants, F6/S23/K25/Y26, constitueront le motif de référence pour la recherche de motifs topologiques similaires par le programme STAMP. Ce dernier a été obtenu à partir d'une structure moyenne, minimisée, de l'ensemble des conformères présents dans le fichier 1BGK.pdb (annexe B.1). Le filtre stérique sera appliqué *via* la structure du modèle d'interaction entre le canal Kv1.1 et la BgK. La similarité des structures des canaux Kv1.1, modélisée à partir du canal KcsA, et Kv1.2, résolue par radiocristallographie, permettra d'appliquer ce filtre indifféremment du canal ciblé.

Chapitre IV : Recherche de similarité d'interaction parmi les ligands d'une même cible. Analyse de complexes protéine-protéine

« Toute connaissance accessible doit être atteinte par des méthodes scientifiques ; et ce que la science ne peut découvrir, l'humanité ne peut pas le connaître » (B. Russell)

La recherche de caractéristiques communes aux interfaces protéine-protéine, fondée sur l'étude de complexes cristallographiques, a suscité de nombreux travaux ayant pour objectif d'identifier des critères permettant de localiser des régions privilégiées d'interaction avec d'autres partenaires protéiques. Ces travaux ont permis de déterminer les valeurs moyennes et les fluctuations de nombreux paramètres, caractéristiques de ces interactions dans les ensembles considérés, constitués de complexes, indistinctement de la nature ou de la fonction des protéines en interaction. Aucun travail spécifique de cette nature n'a été entrepris pour rechercher les similarités d'interactions de plusieurs ligands capables d'interagir avec une même cible à un site identique. Ce travail pourrait permettre de dégager des règles transposables dans le contexte de la conception de ligands protéiques.

Nous avons donc entrepris d'analyser des complexes cristallographiques, mettant en jeu une cible particulière et différents ligands de cette cible. Pour ce faire, nous avons sélectionné dans la PDB, des ensembles de complexes cristallographiques (appelés par la suite « système ») impliquant une même protéine (appelée « cible ») et différents partenaires protéiques (appelés « ligands »). Nous avons recherché si, malgré les disparités de taille, de composition en acides aminés et de structure 3D, ces différents ligands (d'une même cible et d'un même site de liaison) présentaient des caractéristiques physico-chimiques communes.

Le chapitre suivant présentera les différents systèmes utilisés pour la comparaison de ligands protéiques. Puis, les différents complexes seront caractérisés en utilisant plusieurs paramètres globaux (taille de l'interface, composition chimique, complémentarité des surfaces, potentiel électrostatique). Enfin, les interactions intermoléculaires ponctuelles de chaque complexe seront analysées (interactions π - π , cation- π , ponts salins, liaisons hydrogène).

IV.1 Systèmes d'étude

Cette étude a été réalisée dans le contexte plus général du développement de ligands protéique et de son application au développement de ligands des canaux Kv1. Le système de référence (canal Kv1.2-toxine BgK) utilisé dans le cadre de cette étude pour la conception de ligands et la validation de notre méthodologie, est analogue à une interaction « récepteur-inhibiteur », de très haute affinité. Nous avons recherché des systèmes possédant certaines caractéristiques similaires à celles des ligands que nous souhaitons développer, en particulier concernant la taille des ligands impliqués (inférieure à 200 résidus) et la structure des interfaces (une seule zone d'interaction). De plus, parmi les critères de sélection, figure le nombre de complexes cristallographiques disponibles pour chaque système : en effet, l'objectif de ce travail est de comparer les caractéristiques communes et la variabilité des paramètres analysés sur un échantillon de ligands divers, liant avec une même cible. Parmi les différents systèmes envisagés (protéines de structure, protéines du cycle cellulaire, systèmes enzymatiques), nous avons retenu trois systèmes, dont la cible appartient aux groupes des protéases à sérine (E.C. 3.4.21.X).

IV.1.1 Critères de sélection des systèmes

Les critères de sélections, appliqués dans le choix des systèmes, concernent l'identité des cibles dans les différents complexes mais également le mode d'interaction des ligands et leur diversité en termes de composition et de structure primaire et secondaire et tertiaire :

- les cibles contenues dans chacun des fichiers de structure 3D doivent être identiques ou très similaires ;
- les ligands doivent interagir avec la cible dans une région identique, de façon à comparer des interactions similaires ;
- les ligands doivent présenter une diversité, de taille, de structure 3D et de séquence primaire.

Par ailleurs, la qualité et l'intégrité des structures sélectionnées a été prise en compte via plusieurs critères :

- les structures cristallographiques dont la résolution est supérieure à 3,5 Å ont été rejetées ;
- les structures dans lesquelles les positions de certains atomes de la cible ou du ligand n'étaient pas définies (dans la densité électronique) ont été exclues des comparaisons ;
- seuls les complexes hétéro-dimériques ont été analysés.

L'application de l'ensemble de ces critères nous a permis de sélectionner 3 systèmes (annexe B.2) : une série de complexes de la **trypsine bovine**, de la **trypsine porcine** et de la **chymotrypsine bovine**, présentés dans le tableau IV-1.

« Trypsine bovine »			« Trypsine porcine »			« Chymotrypsine bovine »		
Réf.	PDB	Ligand	Réf.	PDB	Ligand	Réf.	PDB	Ligand
[227]	1C9T	Bdellastasin	[228]	1AVX	STI	[229]	1ACB	Eglin C
[230]	1D6R	BBI	[231]	1EJA	Bdellastasin	[232]	1CAO	APPI
[233]	1G9I	Artificial MGI	[234]	1H9H	EETI-II	[232]	1CBW	BPTI
[235]	1OX1	Peptid inhibitor	[236]	1LDT	LDTI	[237]	1CGI	PSTI-3
[238]	1SFI	SSI	[239]	1MCT	MCTI-A	[240]	1CHO	OMTKY-3
[241]	1TAW	APPI	[242]	1TFX	TFPI	[243]	1GLO	PMP-D2V
[244]	1T6S	PPTI	[245]	1TX6	BBI	[243]	1GL1	PMP-C
[246]	2BTC	CPTI-II	[247]	1UHB	Autocatalytical peptid	[248]	1N8O	Ecotin
[143]	2PTC	BPTI	[249]	1Z7K	OMTKY-2			

Tableau IV-1 : Systèmes sélectionnés pour l'analyse des interactions ligand-cible.

IV.1.2 Homogénéité des cibles

IV.1.2-A Homogénéité de séquence

Les structures primaires des cibles ont été comparées par alignement de séquence à l'aide du programme BLAST, [306]. Les complexes cristallographiques dans lesquels la cible comporte une modification importante par rapport à la séquence native de la cible (larges délétions, mutations ponctuelles non-homologues ayant des conséquences importantes sur les propriétés globales) ont été rejetés.

Parmi les cibles du système « trypsine bovine », deux présentent des différences par rapport à la séquence native : 1TAW et 1TGS. La séquence de la trypsine du complexe 1TAW présente une mutation ponctuelle, D165T, se traduisant par une modification de la charge nette globale d'une unité (+7 contre +6 pour la séquence native). La séquence de la trypsine du complexe 1TGS correspond au trypsinogène, précurseur de la trypsine, dont la taille (225 résidus) est supérieure, de deux résidus, à celle de la trypsine (223 résidus). Cette modification de la séquence de la cible ne s'accompagne pas, de modification de la charge nette globale. Les structures 1TAW et 1TGS ont été conservées pour l'analyse des interactions. Les structures des cibles des complexes sélectionnés au sein du système « trypsine porcine » ne montrent aucune différence de séquence. Cependant, les cibles du système « chymotrypsine » présentent des modifications plus nombreuses. La chymotrypsine peut en effet subir des coupures protéolytiques, se traduisant par des séquences primaires discontinues. Les sites de coupure de chaque cible sont présentés dans le tableau IV-2 :

PDB	Nom du ligand	Sites de coupure	Taille (en résidus)	Charge nette globale
1ACB	Eglin C	1-13, 16-146, 149-245	241	+3
1CAO	APPI	1-11, 16-146, 149-245	239	+3
1CBW	BPTI	1-11, 16-146, 149-245	239	+3
1CGI	PSTI-3	1-245	245	+4
1CHO	OMTKY-3	1-10, 16-146, 149-245	238	+3
1GLO	PMP-D2V	1-11, 16-245	241	+3
1GL1	PMP-C	1-12, 16-146, 150-245	239	+3
1N8O	Ecotin	1-11, 16-146, 149-245	239	+3

Tableau IV-2 : Caractéristiques des cibles du système « chymotrypsine bovine ».

L'analyse des séquences montre que ces coupures ne sont pas identiques selon la cible considérée, expliquant que leur taille respective soit variable de 238 à 241 résidus selon la cible. La cible du complexe 1CGI, correspondant au chymotrypsinogène (précurseur de la chymotrypsine), est en revanche continue en séquence et comporte de ce fait 245 résidus. Ces différentes variations

ont, cependant, peu de conséquences sur la charge nette globale, égale a +3 pour chaque cible et +4 pour 1CGI et affecte une région de la cible située à la périphérie du site d'interaction avec les divers ligands.

IV.1.2-B Homogénéité de structure 3D

Pour chacun des systèmes, les structures tertiaires des cibles ont été comparées. Les différences de structures tertiaires des cibles ont été estimées par le calcul du RMSD, après superposition des atomes Ca de leurs chaînes principales respectives, avec une structure à haute résolution de la cible à l'état libre. Les structures de référence sont 1HJ9 [250], 1AKS [251] et 4CHA [252] pour les systèmes « trypsine bovine », « trypsine porcine » et « chymotrypsine bovine » respectivement. Les valeurs de RMSD calculées sur les positions des atomes Ca sont présentées dans le tableau IV-3, pour chacun des systèmes.

« Trypsine bovine »			« Trypsine porcine »			« Chymotrypsine bovine »		
Code PDB	Résolution (en Å)	RMSD (en Å)	Code PDB	Résolution (en Å)	RMSD (en Å)	Code PDB	Résolution (en Å)	RMSD (en Å)
1HJ9	0,95	Réf.	1AKS	1,80	Réf.	4CHA	1,68	Réf.
1C9T	3,30	0,53	1AVX	1,90	0,49	1ACB	2,00	0,51
1D6R	2,30	0,27	1EJA	2,70	0,43	1CA0	2,10	0,56
1G9I	2,20	0,35	1H9H	1,50	0,42	1CBW	2,60	0,63
1SFI	1,65	0,23	1LDT	1,90	0,48	1CGI	2,30	0,97
1OX1	2,00	0,24	1MCT	1,60	0,46	1CHO	1,80	0,37
1TAW	1,80	0,38	1TFX	2,60	0,45	1GL0	3,00	0,57
1TGS	1,80	1,19	1TX6	2,20	0,35	1GL1	2,10	0,44
2BTC	1,50	0,32	1UHB	2,15	0,16	1N8O	2,00	0,89
2PTC	1,90	0,31	1Z7K	1,90	0,54			

Tableau IV-3 : Comparaison des structures tertiaires des cibles présentes dans les différents complexes avec une structure à l'état libre de la cible correspondante. Les RMSD indiqués sont calculés sur les positions des atomes Ca après superposition sur les structures de références, indiquées en gras et en bleu.

La superposition des cibles montre une grande homogénéité des structures 3D, la plupart des variations étant inférieures à 1 Å, à l'exception d'une seule structure : 1TGS (système « trypsine bovine »). L'observation des structures 3D, des cibles du système « trypsine bovine », présentées sur la figure IV-4, illustre l'homogénéité structurale des cibles des différents complexes. Les différences structurales observées affectent presque exclusivement les positions des chaînes latérales. Seule, la structure 1TGS montre une modification locale de structure : la boucle comprise entre les positions 142 à 155 présente une structure différente de celles des autres cibles. Lorsque les atomes Ca appartenant à ce segment ne sont pas pris en compte lors de la superposition, la valeur résultante du RMSD est inférieure à 0,4 Å, témoignant

d'une grande similarité de la structure de cette cible par rapport à la structure de référence 1HJ9. Ces observations nous ont permis de mettre en évidence au sein de chaque système, la similarité des séquences primaires et tertiaires des cibles.

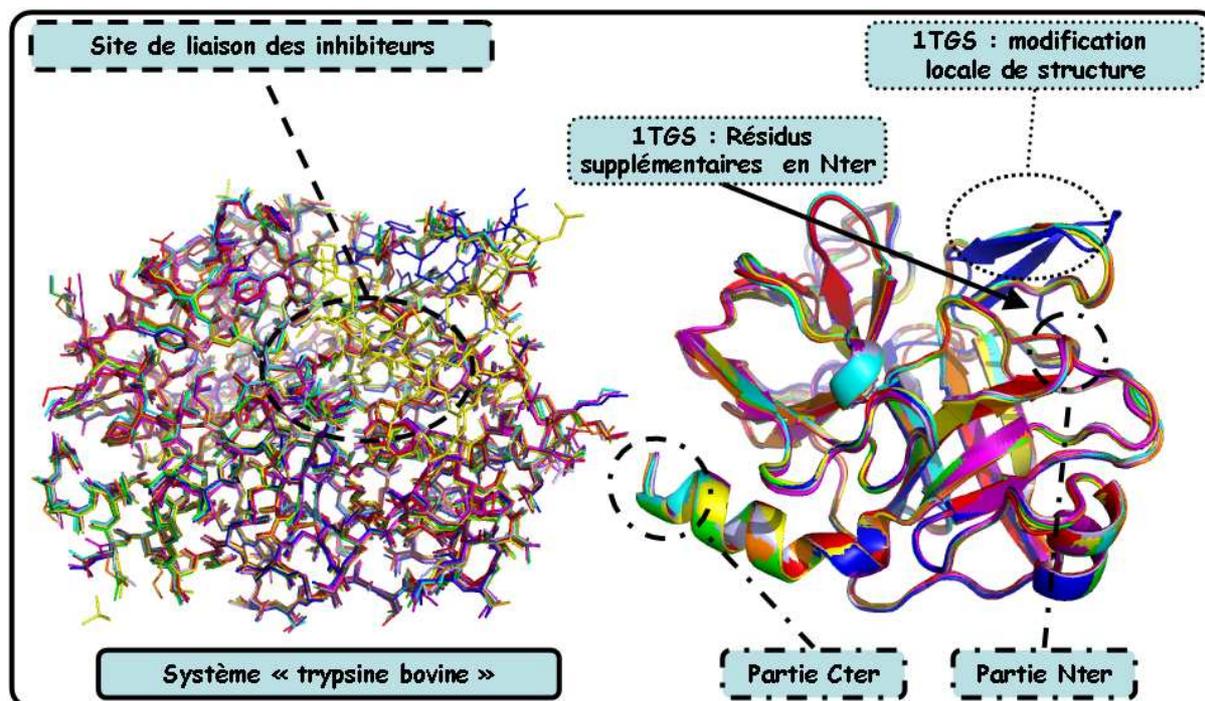
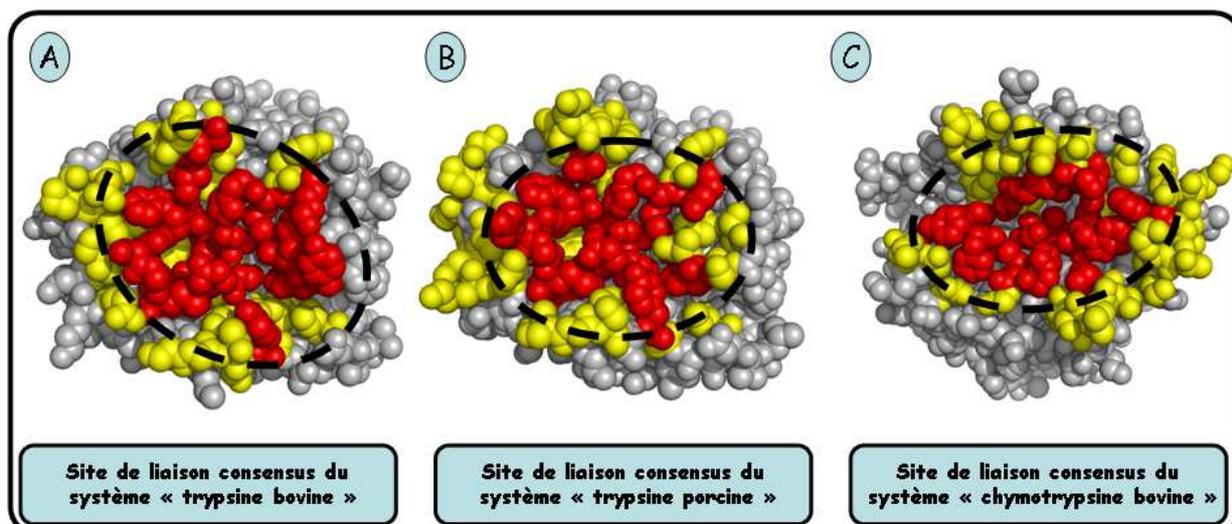


Figure IV-4 : Superposition des structures des cibles des différents complexes du systèmes « trypsine bovine ». Représentation en mode « cartoon ». 1C9T en rouge, 1D6R en orange, 1G9I en jaune, 1OX1 en vert, 15FI en cyan, 1TAW en bleu clair, 1T6S en bleu foncé, 2BTC en magenta et 2PTC en pourpre.

IV.1.3 Le site de liaison

Pour que l'analyse entreprise soit utile dans le contexte du développement de molécules, mimant les propriétés d'un ligand protéique de référence, les différents ligands comparés doivent interagir à un site de liaison identique. Pour chaque système, les sites de liaison de chacun des ligands, ont été déterminés en calculant la variation de surface accessible au solvant (ASA) des résidus de la cible, entre les états libre et lié. Les sites de liaison de chaque système sont représentés sur la figure IV-5.



Figures IV-5 : Représentation des sites de liaisons des différents ligands de chaque système, en mode « sphère ». En A, système « trypsine bovine », en B, système « trypsine porcine » et en C, système « chymotrypsine bovine ». Les résidus en couleurs correspondent à ceux dont l'ASA est modifiée suite à la formation des complexes ligand-cible (en bleu les résidus communs à tous les ligands appartenant à la famille considérée)

Pour chacun des systèmes, un résidu est considéré comme appartenant au site d'interaction lorsqu'une variation de la surface accessible au solvant de ce résidu est observée entre l'état libre et l'état lié dans tous les complexes sélectionnés (coloration rouge sur la figure IV-5, de A à C). Cette figure indique également les résidus de la cible pour lesquels, une variation de surface accessible au solvant entre les états libre et l'état lié, n'est observée que pour certains des complexes considérés. Ceci démontre, dans chaque système, que l'ensemble des ligands interagissent avec la même région. Les résidus, qui ne sont pas systématiquement présents à l'interface avec le ligand mais observés dans certains des complexes, sont situés à la périphérie de la zone pour laquelle les résidus font systématiquement partie de l'interface. Cet observation reflète principalement la disparité des tailles et des structures tridimensionnelles des différents ligands ciblant une région similaire de la cible.

IV.1.4 Diversité des ligands

L'analyse de complexes cristallographiques protéine-protéine entreprise, a pour objectif de rechercher parmi une série de ligands variés sur le plan de leurs tailles, de leurs structures 3D ou de leurs compositions chimiques s'il existe, malgré ces différences, des propriétés communes à ces ligands pouvant expliquer la liaison à la cible. Il est donc nécessaire que les différents ligands sélectionnés possèdent une diversité importante quant aux caractéristiques citées précédemment.

IV.1.4-A Diversité de taille

Les tailles des ligands de chaque système sont indiquées dans le tableau IV-6. Les complexes pour lesquels les ligands protéiques comportent plus de 200 n'ont pas été pris en compte.

« trypsine bovine »		« trypsine porcine »		« chymotrypsine bovine »	
Code PDB	Taille du ligand	Code PDB	Taille du ligand	Code PDB	Taille du ligand
1C9T	59	1AVX	181	1ACB	63
1D6R	71	1EJA	59	1CAO	56
1G9I	22	1H9H	30	1CBW	58
1OX1	11	1LDT	46	1CGI	56
1SFI	14	1MCT	28	1CHO	56
1TAW	58	1TFX	58	1GLO	35
1TGS	56	1TX6	125	1GL1	36
2BTC	29	1UHB	9	1N8O	142
2PTC	58	1Z7K	62		

Tableau IV-6 : Taille des ligands sélectionnés dans chaque système. Les tailles minimales et maximales sont indiquées en bleu et en rouge, respectivement.

Certes, parmi les trois systèmes, la taille des ligands varie de 9 à 181 résidus ; toutefois, la plupart correspondent à diverses « mini-protéines » de taille comprise entre 20 et 100 résidus (3/4 des ligands). Au sein de chacune des familles, il existe donc une variabilité de taille importante. La taille des ligands varie de 11 à 71 résidus, de 9 à 181 résidus et de 35 à 142 résidus pour les systèmes système « trypsine bovine », « trypsine porcine » et « chymotrypsine bovine » respectivement. La disparité des tailles observées, dans les trois systèmes, illustre l'absence de relation entre la taille des ligands et leur capacité à interagir avec une même cible et au même site. En considérant l'ensemble des ligands (quelle que soit leur taille), les variations sont d'autant plus importantes, notamment pour le système « trypsine bovine » : de 3 à 375 résidus. Les tailles minimales et maximales, pour chaque système sont indiquées dans l'annexe B.2, tableau B2-1.

IV.1.4-B Diversité de repliement

Bien qu'il existe de nombreux exemples des protéines possédant des fonctions identiques et des repliements similaires (toxines, anticorps..), aucune relation stricte n'est établie entre le repliement d'une protéine et sa fonction. La comparaison des repliements des différents ligands de chaque système considéré dans le cadre de cette analyse, donne un exemple de variabilité structurale aboutissant à une même « fonction ». Les structures secondaires et leur agencement sont très variables au sein des ligands sélectionnés. Les repliements des différents ligands de la trypsine bovine sont présentés sur la figure IV-7.

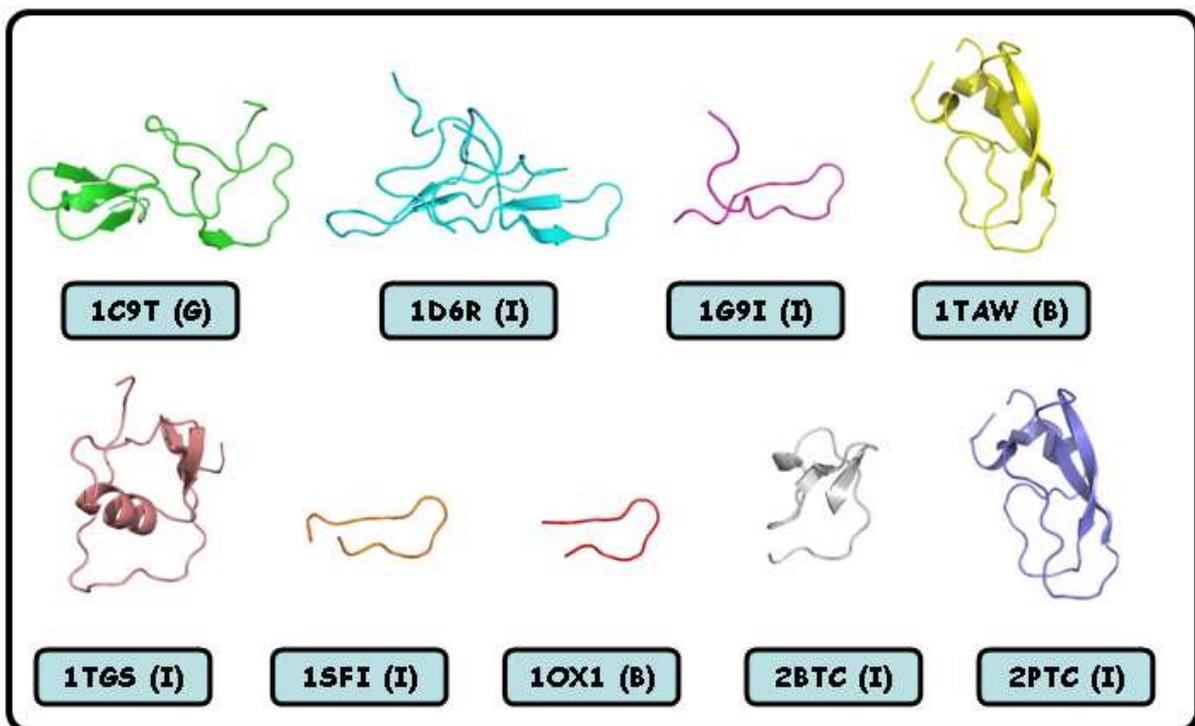


Figure IV-7 : Structures cristallographiques des ligands de la famille trypsine bovine, en représentation « cartoon » de la chaîne principale. Les hélices α sont représentées par des rubans, les brins β par des flèches. Sont indiqués pour chacun, le code PDB ainsi que la dénomination de la chaîne représentée.

Plusieurs ligands sont composés de feuillets uniquement (1C9T, 1D6R et 2BTC); certains comportent un ensemble d'hélices et de feuillets (1TAW, 1TGS et 2PTC) tandis que d'autres ne présentent que des coudes et des boucles (1G9I, 1SFI et 1OX1). Les ligands du système « trypsine porcine », représentés sur la figure IV-8, montrent également une importante diversité, certains d'entre eux composés uniquement de feuillets (1AVX et 1EJA) et d'autre ne présentant que des hélices (1MCT, 1TX6). Plusieurs, en revanche présentent un repliement mixte, composé de feuillets et d'hélices (1H9H, 1LDT, 1Z7K et 1TFX). Enfin le ligand du complexe 1UHB ne contient aucun élément de structure secondaire caractéristique. La figure IV-8 illustre la grande diversité des structures des ligands du système trypsine porcine car, à l'exception des complexes 1H9H et 1MCT, les ligands possèdent tous des repliements différents.

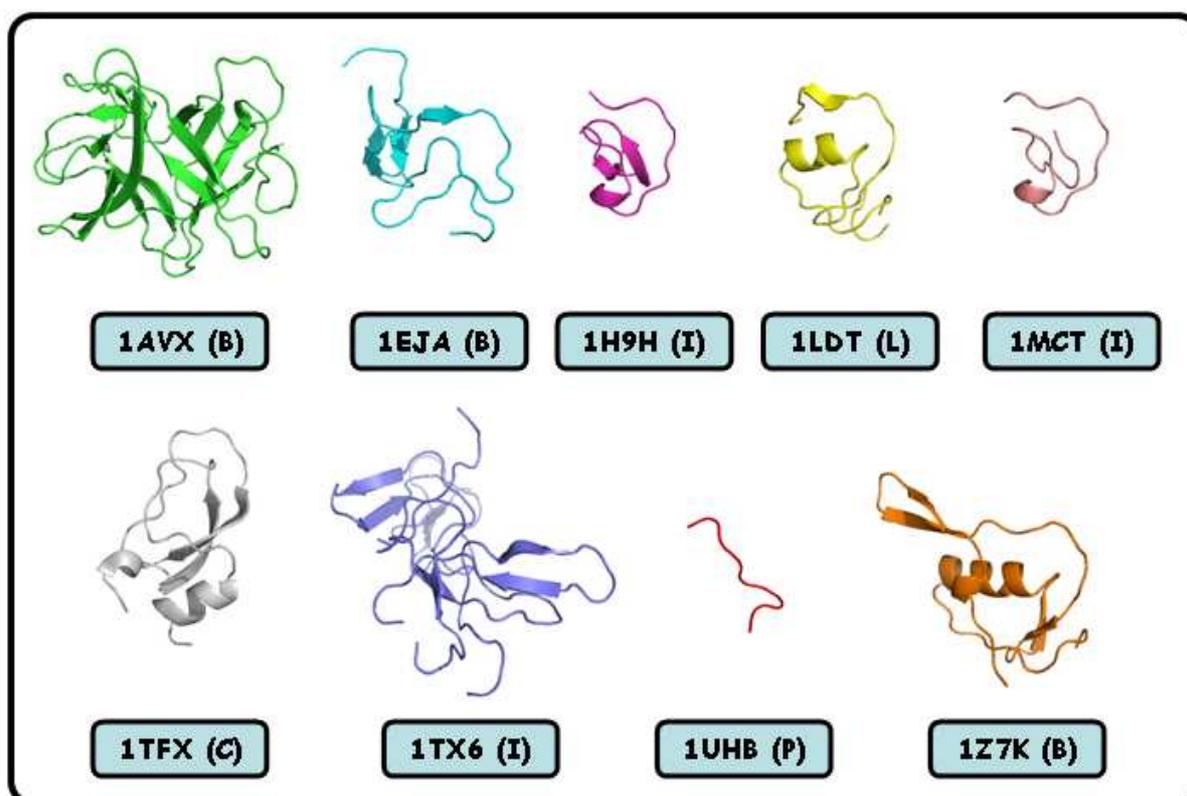


Figure IV-8 : Structures cristallographiques des ligands de la famille trypsine porcine, en représentation « cartoon » de la chaîne principale. Les hélices α sont représentées par des rubans, les brins β par des flèches. Sont indiqués pour chacun, le code PDB ainsi que la dénomination de la chaîne représentée.

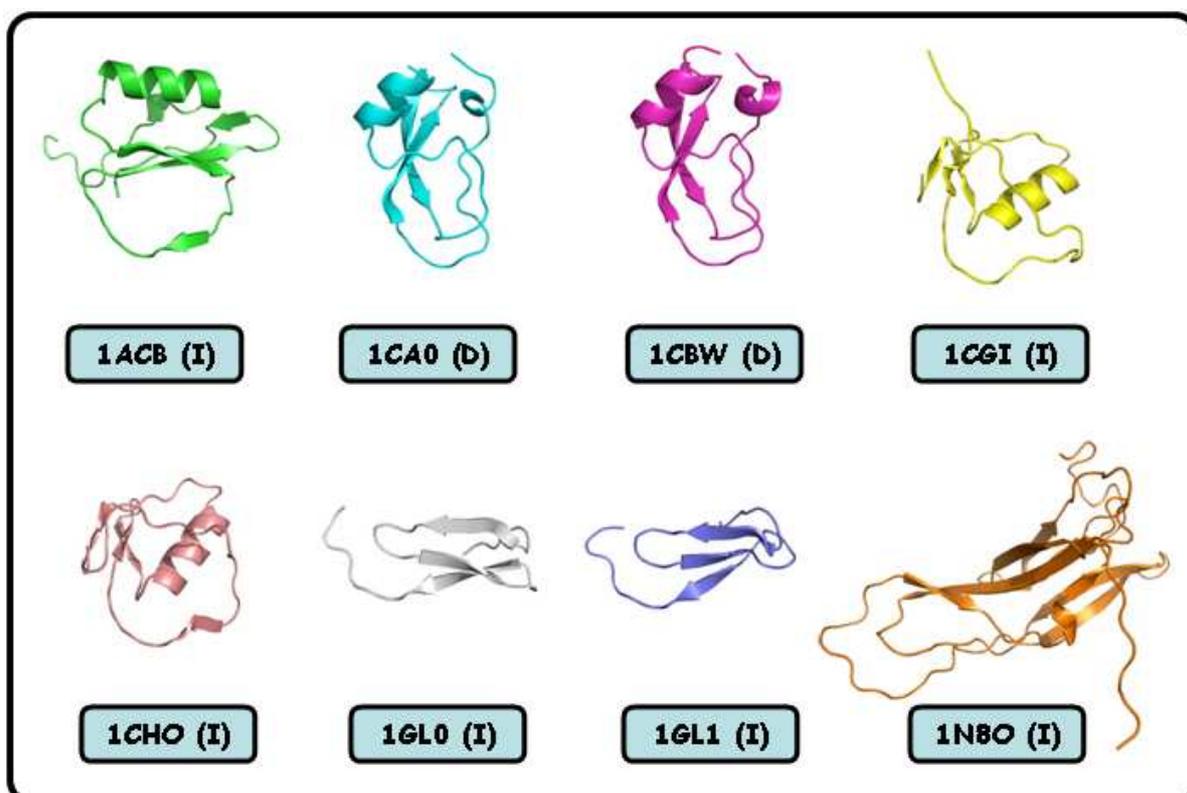


Figure IV-9 : Structures cristallographiques des ligands de la famille chymotrypsine, en représentation « cartoon » de la chaîne principale. Les hélices α sont représentées par des rubans, les brins β par des flèches. Sont indiqués pour chacun, le code PDB ainsi que la dénomination de la chaîne représentée.

Enfin, différents repliements sont également observés parmi les ligands du système « chymotrypsine », représentés sur le figure IV-9 : les ligands composés de feuillets uniquement (1GLO, 1GL1 et 1N8O) et ceux qui sont composés d'association de feuillets et d'hélices (1ACB, 1CAO, 1CBW, 1CGI et 1CHO). La diversité des repliements du système chymotrypsine est plus faible que celles rencontrées dans les deux autres systèmes car l'ensemble des 8 ligands ne correspond qu'à 5 repliements différents.

IV.1.4-C Diversité de séquence

L'analyse des repliements des ligands au sein des trois systèmes a montré l'existence, dans certains cas, de ligands possédant des repliements identiques. Pour éviter la redondance des données dans chaque système, nous avons comparé les structures primaires de ces ligands lorsque leur taille et leur repliement étaient similaires. La comparaison des séquences a été déterminée à l'aide du programme BLAST [306] qui permet de calculer après alignement, leur degré d'homologie et d'identité. Dans la mesure où, des modifications ponctuelles de la séquence et particulièrement pour des résidus en interaction avec la cible, peuvent avoir des conséquences importantes sur certaines propriétés déterminantes pour l'interaction avec la cible, nous avons fixé à 75% le seuil d'homologie au-delà duquel les structures de tailles semblables ne sont pas sélectionnées. L'examen des structures tridimensionnelles des ligands du système « trypsine bovine » schématisées sur la figure IV-7 indique que, seuls les ligands des complexes 1TAW et 2BTC possèdent des tailles et des repliements similaires. Les degrés d'identité et d'homologie obtenus par alignement sont respectivement de 46 et 58 % pour une séquence de 56 résidus alignés. Donc, malgré la similarité des structures 3D entre les ligands des complexes 1TAW et 2BTC, il existe une importante variabilité dans la composition de ces deux protéines. De façon identique, les ligands des complexes 1D6R et 1G9I possèdent un élément de structure 2D similaire, néanmoins leur différence de taille, quel que soit le degré d'homologie de la partie similaire (sur le plan du repliement), rend leur comparaison utile dans le cadre de cette étude. Il en est de même pour les structures des ligands des complexes 1SFI, 1OX1 et 1G9I. De plus, d'après la figure IV-8, seules les structures 1H9H et 1MCT de taille voisine, présentent des structures 3D similaires. L'alignement des séquences de ces deux protéines donne des pourcentages d'identité et d'homologie, respectivement de 55 et 69%, pour 27 résidus alignés. D'après la figure IV-9, il existe, parmi les ligands des complexes du système « chymotrypsine bovine », plusieurs structures présentant des tailles et des repliements similaires. Les résultats de l'alignement de leurs séquences montrent que les ligands présents dans les complexes 1CAO et 1CBW possèdent une identité et une homologie de séquence, respectivement de 46 et 58%, pour

54 résidus alignés. Dans le cas de 1CGI et 1CHO, l'identité et l'homologie de séquence sont respectivement de 39 et 56% pour 41 résidus alignés. Enfin, dans le cas de la comparaison des ligands de 1GLO et 1GL1, l'identité et l'homologie de séquence sont de 42 et 48 %, pour 33 résidus alignés. Donc, malgré la similitude des tailles et structures 3D de ces ligands, il existe une variabilité de composition importante qui rend, la comparaison des interactions, de ces différents ligands, particulièrement intéressante dans le cadre de cette étude.

IV.1.5 Conclusion

Chacun des trois systèmes sélectionnés comporte une cible commune dont nous avons vérifié la similitude de structure primaire, secondaire et tertiaire. Pour chaque système, différents ligands capables d'interagir avec cette cible dans une région identique ont été sélectionnés. Pour chaque famille, ces ligands possèdent une diversité de taille, de composition en acide aminés et/ou de structure tridimensionnelle. Ces trois familles de complexes constituent donc un ensemble de données permettant, sur la base de structures expérimentales issues de l'analyse de cristaux par diffraction X, de rechercher les propriétés conservées et leurs limites pour l'interaction de ligands de tailles, de structure 3D et/ou de composition variables, avec une même cible à un site similaire. Bien que la taille de cet ensemble de données soit réduite et la diversité des cibles faible, cette analyse doit fournir un point de référence pour le développement de notre méthode de conception de ligands protéiques, réalisée au cours de ce travail.

IV.2 Analyse des interfaces ligand-cible

Afin d'analyser et comparer les propriétés d'interaction des ligands de chaque système, les structures cristallographiques des complexes protéiques sélectionnés, ont été préalablement soumises à une procédure de relaxation sous contraintes positionnelles par dynamique moléculaire (annexe B.3). Pour chacune d'entre-elles, cette procédure entraîne des modifications très faibles des positions des atomes lourds (non hydrogène) par rapport à celles définies dans les structures expérimentales ($\text{RMSD} < 0.31 \text{ \AA}$). Cette phase de relaxation a été réalisée afin de reconstruire les atomes d'hydrogène, absents dans la majorité des structures cristallographiques ; reconstruction essentielle, en particulier, pour identifier les liaisons hydrogène et analyser leur géométrie. L'analyse des propriétés globales des interfaces des complexes a ensuite été réalisée au sein de chaque système. Pour cela, plusieurs critères couramment utilisés dans la littérature pour la description d'interfaces protéiques ont été déterminés (taille de l'interface, variation de la surface accessible au solvant consécutive à la liaison, complémentarité des surfaces ligand/cible) et les valeurs de ces paramètres comparées pour les différents ligands, de chaque système (annexe B.3).

IV.2.1 Propriétés globales des interfaces

IV.2.1-A Surface d'interaction

La taille des interfaces peut être caractérisée par le nombre de résidus présents à l'interface mais, également, par la variation de surface accessible au solvant des protéines du complexe entre l'état libre et l'état lié. Les résultats sont indiqués pour les trois systèmes dans le tableau IV-10.

Pour le système « trypsine bovine », le nombre de résidu des ligands présents à l'interface avec la cible varie de 10 à 18 pour une moyenne, sur les différents complexes de ce système, de $14,0 \pm 2,0$ résidus. Dans le cas du système « trypsine porcine », le nombre de résidus des ligands, impliqués dans l'interface, varie de 8 à 27 avec une moyenne de $17,9 \pm 5,5$. Enfin, concernant les ligands du système « chymotrypsine bovine », le nombre de résidus du ligand impliqué dans l'interface avec la cible est de $18,4 \pm 5,5$. Dans ces deux derniers cas, la moyenne et l'écart-type sont donc plus élevées comparativement au système « trypsine bovine ». Cependant, les fluctuations du nombre total de résidus du ligand sont très supérieures à celles du nombre de résidus du ligands, présents à l'interface ; cette observation est illustré par la figure IV-11.

« Trypsine Bovine »	1C9T	1D6R	169I	1OX1	1SFI	1TAW	1T6S	2BTC	2PTC
Nb résidus ligand à l'interface	15	15	14	10	12	12	18	16	14
aire interface totale (en Å ²)	1659	1590	1543	1288	1579	1375	1895	1799	1548
aire interface (ligand) (en Å ²)	954	900	903	797	878	815	1077	1005	891
« Trypsine porcine »	1AVX	1EJA	1H9H	1LDT	1MCT	1TFX	1TX6	1UHB	1Z7K
Nb résidus ligand à l'interface	22	21	16	15	16	15	27	8	21
aire interface totale (en Å ²)	1818	1925	1619	1509	1631	1434	2271	1183	1904
aire interface (ligand) (en Å ²)	1034	1101	961	887	971	854	1272	784	1089
« Chymotrypsine bovine »	1ACB	1CA0	1CBW	1CGI	1CHO	1GLO	1GL1	1N80	
Nb résidus ligand à l'interface	18	13	14	20	16	21	22	23	
aire interface totale (en Å ²)	1612	1344	1514	2157	1595	1974	1825	1792	
aire interface (ligand) (en Å ²)	896	772	885	1205	914	1113	1025	1009	

Tableau IV-10 : Taille des interfaces ligand-cible caractérisées par le nombre de résidus du ligand présent à l'interface et la variation de surface accessible au solvant des deux partenaires ou du ligand seul.

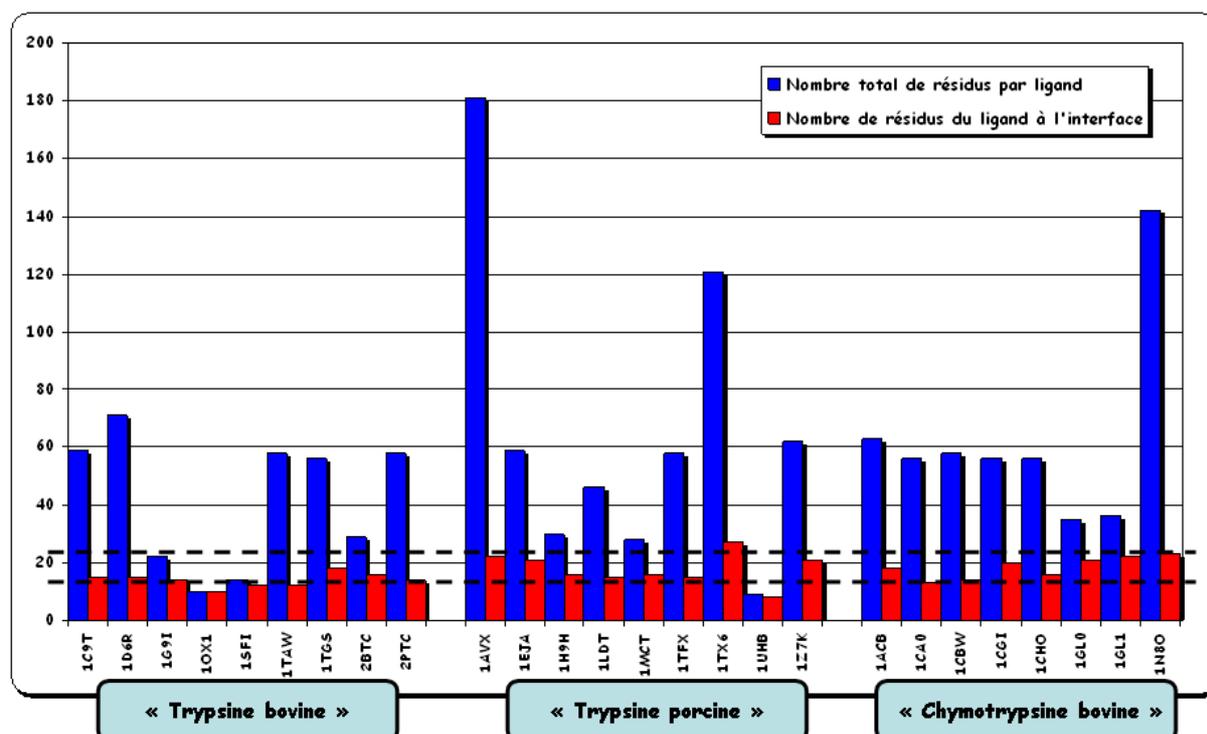


Figure IV-11 : Nombre de résidus des ligands impliqués dans l'interface avec la cible et nombre total de résidus des ligands pour les trois systèmes considérés.

Une observation similaire peut-être faite en considérant les moyennes et fluctuations de la variation de surface accessible au solvant ou de la fraction de celle-ci impliquant des résidus du ligand (tableau IV-10). Les résultats indiquent que, pour les systèmes « trypsine bovine », « trypsine porcine » et « chymotrypsine bovine », les variations totales de surface accessible au solvant sont respectivement 1586 ± 188 , 1699 ± 319 et 1726 ± 262 Å². Pour ces mêmes systèmes, les aires du ligand, impliquées dans l'interface avec la cible, sont respectivement de 913 ± 88 , 994 ± 148

et $977 \pm 138 \text{ \AA}^2$. Les fluctuations de ces deux paramètres sont inférieures à 20 %, autour de la valeur moyenne, pour des tailles des ligands comprises entre 9 et 181 résidus. Les fluctuations de taille d'interface sont donc plus faibles que les fluctuations du nombre de résidus à l'interface, décrit précédemment, reflétant en partie, la dispersion des tailles des résidus. Il n'y a donc pas de corrélation entre l'aire de l'interface et la taille des ligands en interaction. Cette observation rejoint les observations faites sur des complexes protéine-protéine, sélectionnés indépendamment ou non de leur type ou de la fonction des protéines impliquées [52, 131, 132, 133, 134, 139].

IV.2.1-B Complémentarité de structure

La complémentarité de surface est une caractéristique particulièrement importante des interactions protéine-protéine. En effet, une complémentarité importante des surfaces des deux partenaires en présence aura pour conséquence de favoriser l'établissement d'interactions intermoléculaires, dépendantes de la distance entre les groupements impliqués (interactions de van der Waals, liaisons hydrogène, ponts salins) [140, 142]. En revanche, une conséquence de la faible complémentarité de surface entre les partenaires sera une présence accrue de molécules d'eau, piégées à l'interface lors du processus d'association. Ce phénomène, d'autant plus probable que la complémentarité de surface est faible, apporte une contribution entropique défavorable à la liaison.

La complémentarité de l'interface peut être estimée par le calcul volume non occupé entre les partenaires à l'état lié. Cependant, ce volume doit être ramené à l'aire de l'interface, définissant ainsi le « gap index » (chapitre II). Les valeurs de « gap index », calculées pour les différents complexes sont reportées dans le tableau IV-12.

Tryp. Bovine	1C9T	1D6R	1G9I	1OX1	1SFI	1TAW	1TGS	2BTC	2PTC	μ	σ
Gap Index (Å)	1.73	1.69	1.81	1.66	1.76	1.76	1.81	1.86	2.14	1,80	0,14
Tryp. Porcine	1AVX	1EJA	1H9H	1LDT	1MCT	1TFX	1TX6	1UHB	1Z7K	μ	σ
Gap Index (Å)	2.14	1.81	1.81	2.01	1.75	1.82	2.27	1.66	1,74	1,90	0,20
Chym. Bovine	1ACB	1CAO	1CBW	1CGI	1CHO	1GLO	1GL1	1N8O		μ	σ
Gap Index (Å)	1,82	1,88	2,11	1,92	1,75	1,77	1,75	2,01		1,88	0,13

Tableau IV-12 : Valeurs de « gap index », calculées pour les différents complexes des trois systèmes étudiés. La moyenne est indiquée par μ , l'écart-type par σ .

Il a été montré que la valeur de cet indice, calculé pour une série de complexes inhibiteurs enzymes, était similaire à la valeur obtenue pour une série d'hétéro-complexes ou d'homo-dimères

[139]. Les valeurs de « gap index » obtenues sont très homogènes, quel que soit le système considéré : $1,80 \pm 0,14 \text{ \AA}$, $1,90 \pm 0,20 \text{ \AA}$ et $1,88 \pm 0,13 \text{ \AA}$. Ces valeurs indiquent que, dans les différents complexes, les surfaces des partenaires en interaction sont très complémentaires, témoignant d'une grande compacité de l'interface. De plus, la faible dispersion des valeurs de « gap index », observée au sein de chaque système, montre une grande homogénéité de cette propriété parmi les différents complexes. Toutefois, les valeurs de « gap index » calculées pour les trois systèmes sélectionnés sont plus faibles que celles reportées pour une série de ligands enzyme-inhibiteur, publiées dans une étude antérieure [139], dans laquelle figurent certains des complexes utilisés dans le cadre de notre étude. Cette différence (de l'ordre de 20 %) peut être due, en partie, au fait que nous n'utilisons pas le même algorithme pour calculer cet indice. Cependant ce travail nous permet d'analyser, sur notre base de données, les fluctuations de la compacité de l'interface et de calibrer cet outil, pour son application, dans le cadre de notre approche de conception de ligands protéiques.

IV.2.1-C Composition chimique de l'interface

La nature chimique de l'interface dans les différents complexes a été analysée. Une première indication de la nature chimique de l'interface peut être obtenue en déterminant le nombre de résidus de type non polaires, polaires neutres et polaires chargés à l'interface. Les fractions correspondant à ces trois classes sont reportées dans le tableau IV-13.

« Trypsine bovine »	1C9T	1D6R	1G9I	1OX1	1SFI	1TAW	1TGS	2BTC	2PTC
Nb Res Non-Polaires	9	7	8	5	7	9	9	13	9
Nb Res Polaires Neutres	2	6	4	4	3	3	6	0	1
Nb Res Polaires Chargés	4	2	2	1	2	1	3	3	4
« Trypsine porcine »	1AVX	1EJA	1H9H	1LDT	1MCT	1TFX	1TX6	1UHB	1Z7K
Nb Res Non-Polaires	11	12	12	10	12	11	16	3	10
Nb Res Polaires Neutres	2	4	2	0	0	1	5	2	6
Nb Res Polaires Chargés	9	5	2	5	4	3	6	3	5
« Chymotrypsine bovine »	1ACB	1CAO	1CBW	1CGI	1CHO	1GLO	1GL1	1N8O	
Nb Res Non-Polaires	12	10	10	11	9	11	11	13	
Nb Res Polaires Neutres	2	2	1	4	4	8	6	4	
Nb Res Polaires Chargés	4	1	3	5	3	2	5	6	

Tableau IV-13 : Composition en résidus des interfaces des ligands dans les complexes des trois systèmes étudiés (résidus non-polaires : A, C, G, I, L, M, P, V, F, H, Y, W ; résidus polaires neutres : N, Q, S, T ; résidus polaires chargés : K, R, D, E).

Les valeurs calculées indiquent, pour ces trois catégories de résidus, qu'il existe au sein de chaque famille une importante variabilité. Ainsi, le nombre de résidus polaires neutres varient de 0 à 6 pour les ligands des systèmes « trypsine bovine » et « trypsine porcine », tandis qu'il varie de 1 à 8 pour les ligands du système chymotrypsine. Une amplitude de variation similaire est observée pour les résidus de type polaire chargé. Bien que le nombre de résidus non polaires du ligand, présents à l'interface avec la cible, puisse donner une indication de l'hydrophobicité de l'interface, il a été proposé d'évaluer celle-ci en calculant la fraction de l'aire de l'interface impliquant des atomes du ligand appartenant à des groupements aliphatiques et ce, quel que soit le résidu auquel appartient ce groupe [52]. Ces fractions ont été calculées pour les ligands des complexes des trois systèmes et sont reportées dans le tableau IV-14.

« Trypsine bovine »	1C9T	1D6R	1G9I	1OX1	1SFI	1TAW	1TGS	2BTC	2PTC
% résidus non-polaires à l'interface	39	31	34	35	35	32	46.5	57	39
« Trypsine porcine »	1AVX	1EJA	1H9H	1LDT	1MCT	1TFX	1TX6	1UHB	1Z7K
% résidus non-polaires à l'interface	40	39	41	58	40	46	39	17	43
« Chymotrypsine bovine »	1ACB	1CA0	1CBW	1CGI	1CHO	1GLO	1GL1	1N80	
% résidus non-polaires à l'interface	54	33	38	50	49	45	48	38	

Tableau IV-14 : Analyse de la fraction apolaire de l'interface, en % de la surface (coté ligand). Les valeurs inférieures sont indiquées en bleu, les valeurs supérieures en rouge.

La fraction hydrophobe des résidus du ligand présents à l'interface est en moyenne de $38 \pm 8\%$, $40 \pm 1\%$ et $44 \pm 7\%$ pour les systèmes « trypsine bovine », « trypsine porcine » et « chymotrypsine bovine » respectivement. Seul le ligand du complexe 1UHB, du système « trypsine porcine », présente une fraction d'interface hydrophobe significativement inférieure à ces valeurs (17%). Cependant, la dispersion, autour des valeurs moyennes des fractions de surface hydrophobes des ligands ainsi calculées, est beaucoup plus faible que la dispersion du nombre de résidu appartenant à la classe non polaire.

IV.2.2 Aspects électrostatiques

Il a été montré que les interactions électrostatiques jouaient un rôle essentiel pour la reconnaissance, la formation et/ou la stabilisation de complexes protéiques (chapitre II) [174, 175, 176, 178, 179]. Cependant, l'analyse des différents ligands des complexes considérés au sein de chaque système montre une grande hétérogénéité des charges nettes globales. Les différentes valeurs calculées sont présentées dans le tableau IV-15 :

« Trypsine bovine »	1C9T	1D6R	1G9I	1OX1	1SFI	1TAW	1TGS	2BTC	2PTC
Charge nette globale	-2	-5	0	1	1	-3	1	0	6
Charge nette de l'interface	1	2	1	1	1	1	1	2	4
Charge non compensée	3	1	2	2	2	0	2	4	3
« Trypsine porcine »	1AVX	1EJA	1H9H	1LDT	1MCT	1TFX	1TX6	1UHB	1Z7K
Charge nette globale	-10	-3	1	5	1	-3	9	-1	-4
Charge nette de l'interface	0	0	2	4	2	3	0	0	-1
Charge non compensée	9	2	1	5	5	2	4	2	4
« Chymotrypsine bovine »	1ACB	1CAO	1CBW	1CGI	1CHO	1GLO	1GL1	1N8O	
Charge nette globale	1	-3	6	-1	0	-1	2	1	
Charge nette de l'interface	3	1	3	-1	2	-2	1	3	
Charge non compensée	6	1	3	4	3	4	7	4	

Tableau IV-15 : Analyse des ligands des trois systèmes, en termes de nombre de charges nettes, totales ou à l'interface seulement. Les valeurs inférieures et supérieures sont, respectivement en bleu et en rouge.

Sous l'hypothèse que les pK des groupements acido-basiques des acides aminés ne soient pas significativement modifiés au sein des ligands par rapport à ceux adoptés par ces groupes dans les acides aminés à l'état libre, les états de charges de ces ligands à pH 7 varient respectivement de -5 à +6, de -10 à +9 et de -3 à +6 pour les ligands du système trypsine bovine, trypsine porcine et chymotrypsine. Les différents ligands d'une même cible peuvent donc posséder des états de charges très divers. L'analyse de la charge nette globale des résidus des ligands situés à l'interface avec la cible montre également des variations significatives de +1 à +4, de -1 à +4 et de -2 à +3 respectivement pour les complexes des systèmes « trypsine bovine », « trypsine porcine » et « chymotrypsine bovine ». Bien que ces variations apparaissent plus faibles que les variations des états de charge total des ligands, elles peuvent avoir un impact important sur le potentiel électrostatique engendré à l'interface.

Les potentiels électrostatiques des différents ligands ont donc été comparés deux à deux selon le protocole décrit dans son intégralité dans l'annexe A.3. L'analyse graphique du potentiel électrostatique, projeté à la surface de la protéine et dont l'intensité est représentée par un

code de couleur (potentiels négatifs en rouge, potentiels positifs en bleu), permet une comparaison semi-quantitative des ligands de chaque système. Les figures, ainsi obtenues, illustrent les similitudes et les différences entre les potentiels électrostatiques de surface des différents ligands. Sur chacune d'entre-elles, la région correspondant à l'interface avec la cible est encerclée en noir. La figure IV-16 représente le potentiel électrostatique des 9 ligands du système « trypsine bovine ». Certes, dans la région de l'interface, l'intensité et le signe du potentiel électrostatique sont similaires pour les 9 ligands ; cependant, pour les régions distantes du site de liaison avec la cible, les potentiels électrostatiques de surface peuvent présenter des différences très importantes. Ainsi, les comparaisons des potentiels de surface des ligands 1TGS/1D6R et 1TAW/2PTC montrent, pour ses deux ligands, des inversions de signe du potentiel (partie supérieure droite de la surface, sur la figure IV-16). Des observations similaires peuvent être faites sur la figure IV-17, représentant les potentiels électrostatiques de surface des différents ligands du système « trypsine porcine ». Les potentiels de surface des régions interfaces sont très similaires, à l'exception notable de la structure du ligand du complexe 1Z7K. Dans d'autres cas, les régions distantes du site d'interaction montrent des potentiels de surface opposés (1TFX/1TX6, partie supérieure, figure IV-17). Dans le système « chymotrypsine », les structures 1CA0 et 1CHO présentent également des différences significatives dans la partie inférieure droite sur la figure IV-18, alors que les potentiels électrostatiques, dans les régions des ligands correspondant à l'interface avec la cible, sont similaires.

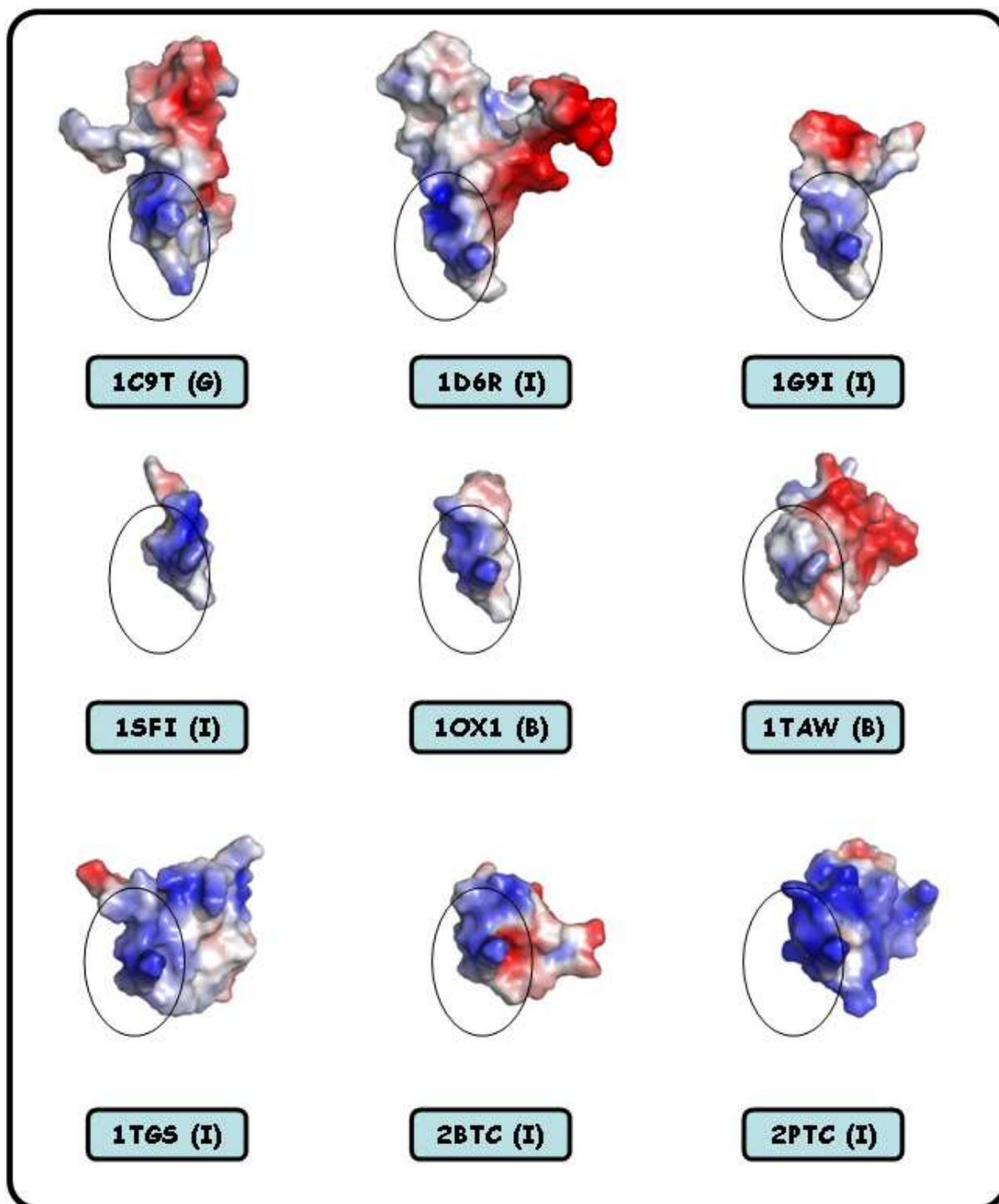


Figure IV-16 : Comparaison des potentiels électrostatiques des ligands du système « trypsine bovine ». Pour chaque ligand, le numéro de chaîne est indiqué entre parenthèses. La région de l'interface est encadrée en noir. Les potentiels sont représentés à la surface accessible au solvant des protéines ($-5k_B T/e$, en rouge, à $+5k_B T/e$, en bleu).

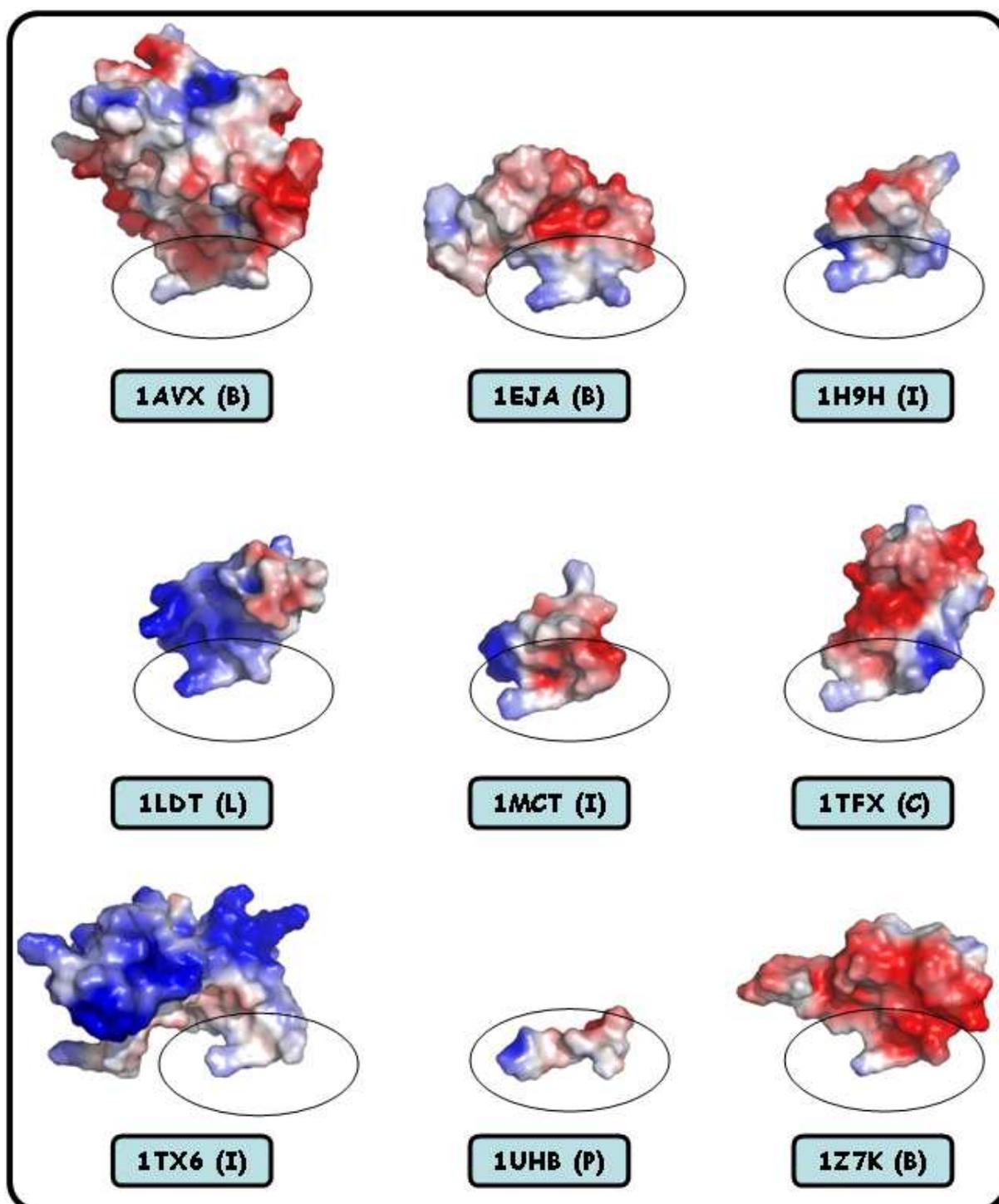


Figure IV-17 : Comparaison des potentiels électrostatiques des ligands du système « trypsine porcine ». Pour chaque ligand, le numéro de chaîne est indiqué entre parenthèses. La région de l'interface est encadrée en noir. Les potentiels sont représentés à la surface accessible au solvant des protéines ($-5k_B T/e$, en rouge, à $+5k_B T/e$, en bleu).

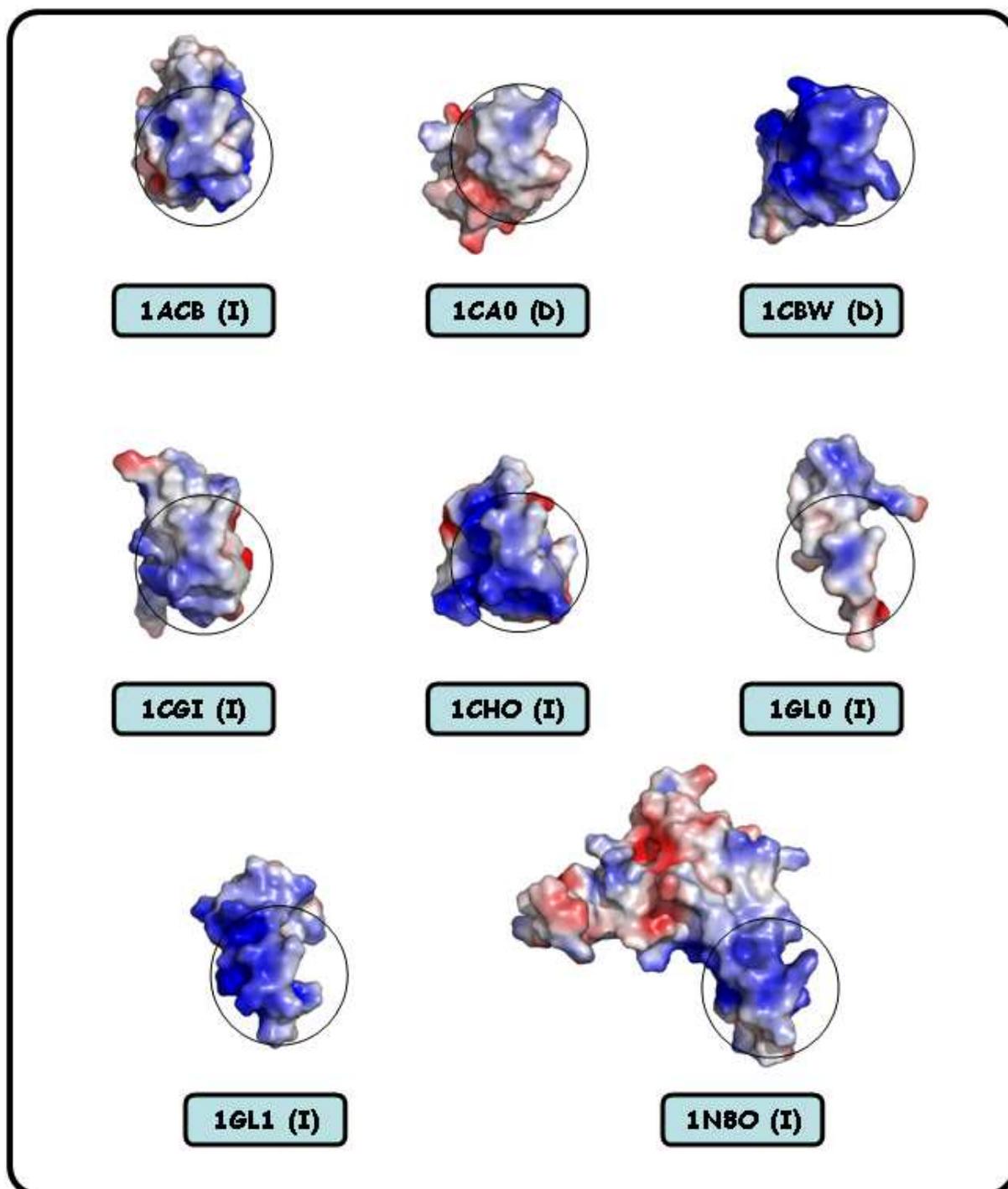


Figure IV-18 : Comparaison des potentiels électrostatiques des ligands du système « chymotrypsine bovine ». Pour chaque ligand, le numéro de chaîne est indiqué entre parenthèses. La région de l'interface est encerclée en noir. Les potentiels sont représentés à la surface accessible au solvant des protéines ($-5k_B T/e$, en rouge, à $+5k_B T/e$, en bleu).

Ces observations nous ont incité à comparer, de manière quantitative, les potentiels électrostatiques dans la région des ligands correspondant à l'interface avec la cible. Ceci peut être réalisé grâce aux indices de similarité de Hodgkin (SI) calculé sur un ensemble restreint de points englobant l'interface (annexe A.3). Les valeurs des indices de similarité, calculés sur ces grilles partielles, sont dénommées SI « part ». Les indices, calculés sur l'ensemble des points de la grille, permettant une comparaison dans toutes les directions de l'espace, sont appelés SI « all ». Les comparaisons, deux à deux, des grilles de potentiel électrostatique, calculées pour les ligands du système « trypsine bovine » et leur caractérisation par le calcul de l'indice de similarité de Hodgkin est reporté dans la figure IV-19.

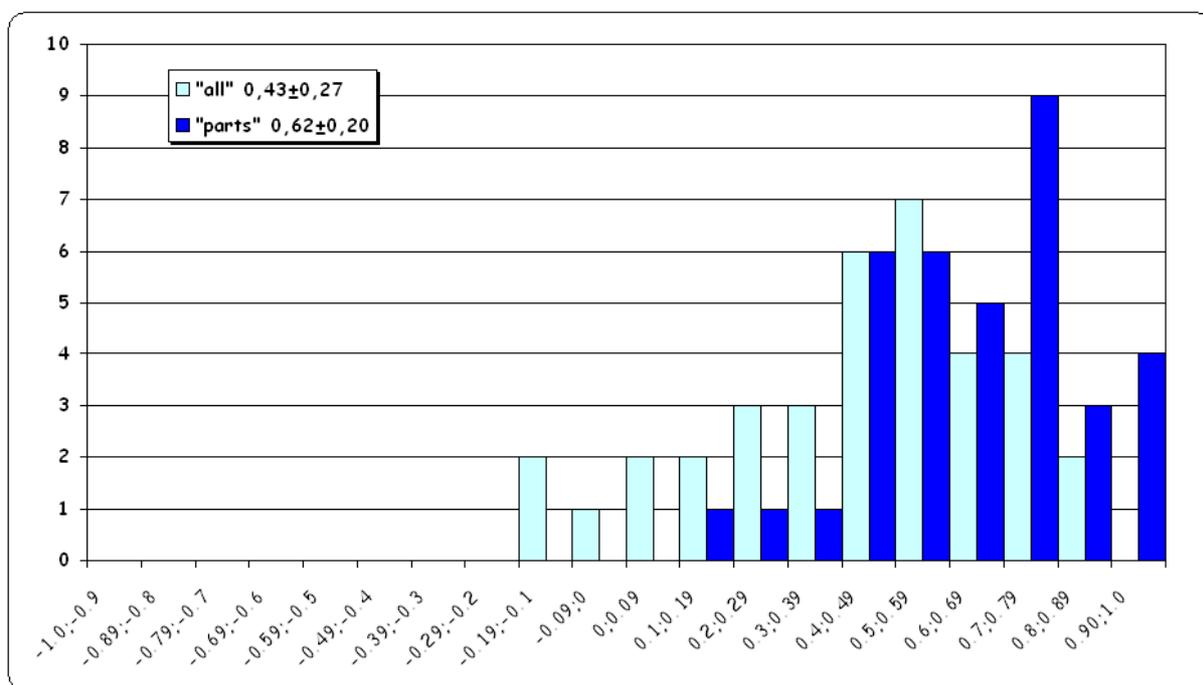


Figure IV-19 : Histogramme de distribution des indices de similarité de Hodgkin calculés sur l'ensemble des points des grilles de potentiel électrostatique (SI « all », bleu clair) et sur un région restreinte des grilles englobant l'interface avec la cible (SI « part », bleu foncé) pour chaque ligand appartenant au système « trypsine bovine ». Les valeurs de SI s'échelonnent de -1 à +1 et sont présentées par intervalles de 0,1 unité.

L'histogramme de distribution de ces comparaisons permet de faire plusieurs observations. La distribution des indices calculés sur l'ensemble des points des grilles de potentiel est centrée sur une valeur supérieure à 0 ($0,43 \pm 0,27$), indiquant que les potentiels électrostatiques des ligands présentent, en moyenne, un caractère corrélé. Toutefois, cette distribution est très large et les extrema correspondent à des valeurs très éloignées (de -0,19 à +0,89). Il semble donc que la largeur de la distribution de ces valeurs reflète la dispersion des charges totales des différents ligands, soulignée précédemment. L'examen des indices de Hodgkin, calculés pour les différents ligands de ce système sur les régions restreintes des grilles de potentiel englobant l'interface avec la cible, suggère une distribution centrée sur une valeur plus élevée que les indices SI « all »

(0,62). De plus, la largeur de la distribution des SI « part » est plus faible que celle observée précédemment (0,20). Les extrema de la distribution des SI « part » indiquent également une plus faible dispersion des valeurs (de +0,2 à 1).

L'analyse de la distribution des indices de similarité pour les ligands du système « trypsine porcine », représentée sur la figure IV-20, permet de faire des observations similaires.

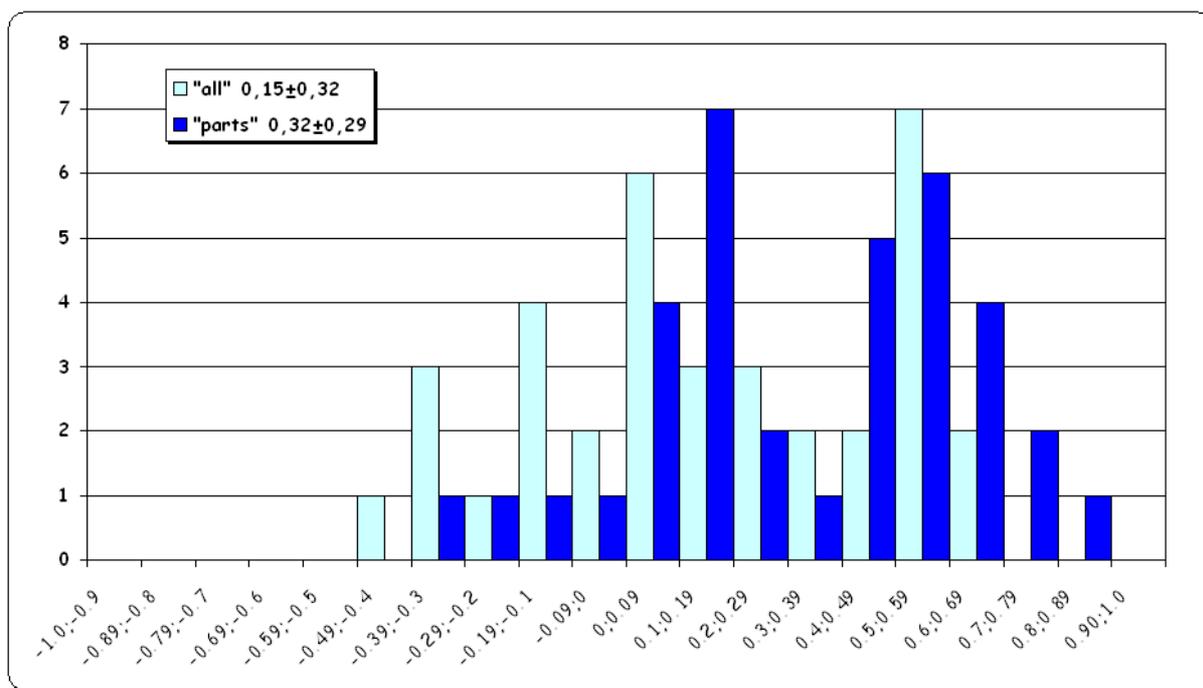


Figure IV-20 : Histogramme de distribution des indices de similarité de Hodgkin calculés sur l'ensemble des points des grilles de potentiel électrostatique (SI « all », bleu clair) et sur un région restreinte des grilles englobant l'interface avec la cible (SI « part », bleu foncé) pour chaque ligand appartenant au système « trypsine porcine ». Les valeurs de SI s'échelonnent de -1 à +1 et sont présentées par intervalles de 0,1 unité.

La figure IV-20 représente la distribution de SI, calculée pour les 36 comparaisons des potentiels électrostatiques des ligands du système « trypsine porcine ». La distribution des SI « all » s'avère très large (écart-type = 0,32) avec une moyenne de 0,15, proche de 0. De nombreuses valeurs négatives sont observées, indiquant des potentiels électrostatiques anti-corrélés et, de nombreuses valeurs positives indiquant des potentiels électrostatiques corrélés. Or, la distribution des SI « part » montre une valeur moyenne positive plus élevée par comparaison aux SI « all » (0,32), suggérant une corrélation plus importante des potentiels électrostatiques dans la région d'interaction avec la cible.

Les distributions des indices de Hodgkin, pour le système « chymotrypsine bovine », sont représentées sur la figure IV-21 :

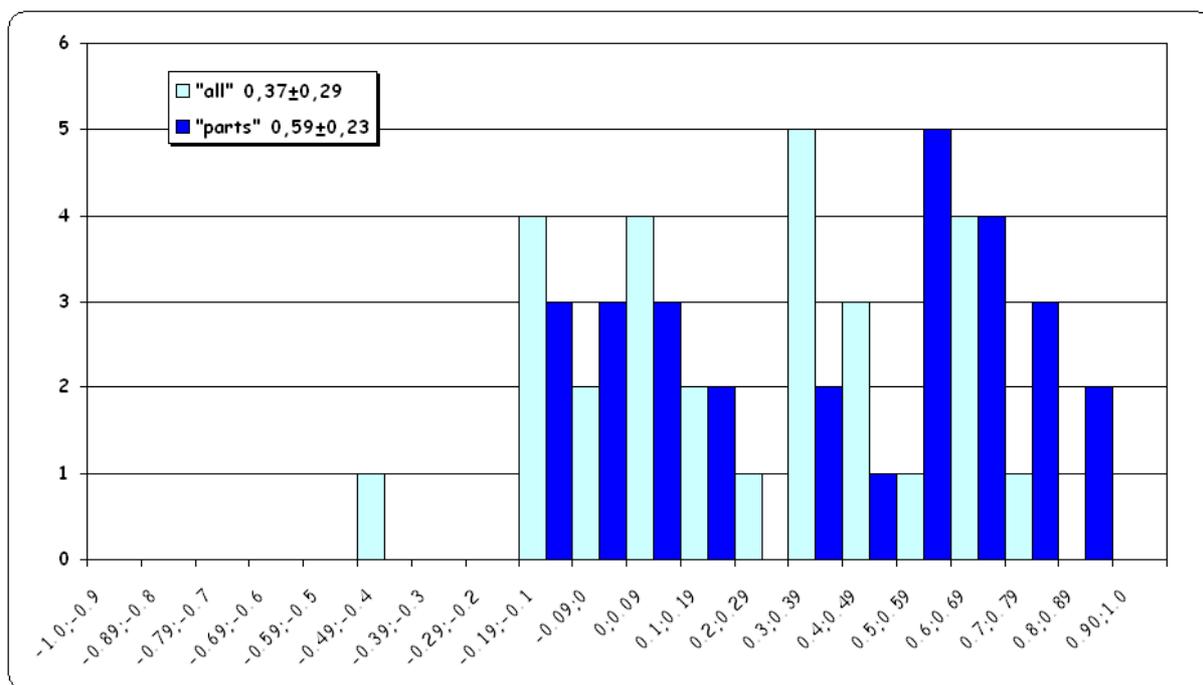


Figure IV-21 : Histogramme de distribution des indices de similarité de Hodgkin calculés sur l'ensemble des points des grilles de potentiel électrostatique (SI « all », bleu clair) et sur un région restreinte des grilles englobant l'interface avec la cible (SI « part », bleu foncé) pour chaque ligand appartenant au système « chymotrypsine bovine ». Les valeurs de SI s'échelonnent de -1 à +1 et sont présentées par intervalles de 0,1 unité.

La distribution des SI « part » indique une corrélation plus forte des potentiels électrostatiques dans les régions des ligands correspondants à l'interface avec la cible. Il faut noter que les distributions des indices de Hodgkin (SI « all » et « part »), dans le cas des systèmes « trypsine porcine » et « chymotrypsine bovine », ne correspondent pas à une distribution unimodale. Ceci peut résulter de la taille réduite des échantillons analysés, des approximations réalisées au cours du calcul du potentiel électrostatique et du mode de sélection des points des grilles correspondants à la région d'interaction avec la cible. Dans le cas du système « chymotrypsine bovine », le caractère bimodale de la distribution est lié aux valeurs obtenues dans les comparaison de 2 parmi les 8 ligands : 1CAO et 1GLO. Les valeurs de SI, calculées par la comparaison de ces deux ligands, sont néanmoins proches : SI « all » = 0,50, SI « part » = 0,67. Cette dernière observation suggère l'existence de plusieurs formes de potentiels électrostatiques, compatibles avec l'association des ligands à une cible particulière. Cependant, une étude de ces similitudes sur un échantillon de complexes de taille plus importante, serait nécessaire pour valider cette hypothèse.

La comparaison des caractéristiques électrostatiques des trois séries de complexes permet de dégager plusieurs observations. Dans plusieurs cas, nous avons montré que les potentiels électrostatiques, de la région d'interaction avec la cible, sont fortement corrélés malgré une différence d'état de charge très importante des ligands dans la région d'interaction. Ainsi, l'indice de similarité (SI « part ») calculé pour le couple 2PTC/1TGS est de 0,63 pour une différence de charge nette globale de ces deux ligands est de 5. De même, les couples 1H9H/1EJA, 1H9H/1LDT, 1H9H/1TFX, 1EJA/1MCT et 1LDT/1MCT présentent des valeurs de SI « part », toutes supérieures à 0,6, pour des différences d'état de charge nette globale de 4. Le couple 1CBW/1CGI représente un exemple extrême de cette situation (SI « part » = 0,45 pour une différence d'état de charge globale de 7 et une différence d'état de charge de 4 dans la région d'interface). Ces différents cas constituent autant d'exemples de l'absence de corrélation entre charge globale et potentiel électrostatique. De plus, parmi les comparaisons des potentiels électrostatiques de différents ligands d'une même cible, effectuées dans le cadre de cette analyse, nous avons identifié plusieurs cas où deux ligands possédaient des potentiels électrostatiques anti-corrélés sur l'ensemble des directions de l'espace, alors qu'une corrélation était observée dans la région d'interaction avec la cible commune (entre parenthèse, les SI « all » sont à gauche, SI « parts » à droite) : 1D6R/1TGS (-0,11 ; 0,48), 1D6R/1G9I (0,19 ; 0,64), 1C9T/2PTC (0,06 ; 0,49), 1LDT/1TFX (0,00 ; 0,44), 1H9H/1TX6 (0,19 ; 0,54), 1EJA/1TX6 (-0,49 ; 0,50), 1EJA/1LDT (0,05 ; 0,50), 1AVX/1TX6 (0,11 ; 0,52), 1ACB/1CGI (-0,17 ; 0,61). Par ailleurs, dans les trois systèmes considérés, les similarités des potentiels électrostatiques mises en évidence par le calcul des indices de Hodgkin sont, en moyenne, plus importantes dans la région d'interaction avec la cible que dans l'ensemble des directions de l'espace.

IV.2.3 Conclusion

L'étude de plusieurs aspects des propriétés électrostatiques des trois systèmes sélectionnés, permet de tirer des enseignements pour la sélection des plateformes dans le cadre du développement de ligands protéiques. Plusieurs ligands sont effectivement capables d'interagir avec une même cible, malgré des charges nettes globales très différentes (jusqu'à +7 dans les cas étudiés). La charge globale des protéines n'est donc pas un critère déterminant pour la capacité à lier une cible donnée. De plus, pour les trois systèmes, le nombre de charges nettes présentes à l'interface, compensées ou non, sont également susceptibles de varier de manière importante. Ce paramètre ne peut donc pas être retenu comme un critère d'évaluation de la capacité d'une protéine à interagir avec sa cible. En revanche, la comparaison quantitative des potentiels électrostatiques des différents ligands a montré que les potentiels de plusieurs ligands d'une même cible pouvaient s'avérer non corrélés voir anti-corrélés. Néanmoins, ces comparaisons ont montré que les potentiels à l'interface avec la cible commune étaient corrélés. Ce dernier indicateur, peut-être retenu comme un critère d'évaluation de la capacité d'une protéine à interagir avec la cible considérée. Il convient toutefois de souligner que ce critère doit être considéré comme un filtre qui permettra d'augmenter la probabilité d'identifier un ligand. Mais, l'étude réalisée ne permet pas de conclure de manière absolue sur la valeur de cet indice et sur la valeur seuil du SI « part » à considérer pour identifier un ligands. En considérant un seuil de similarité de potentiel électrostatique minimum, égal à 0,4 pour les trois systèmes analysés (« trypsine bovine », « trypsine porcine » et « chymotrypsine »), les valeurs de SI, à l'interface supérieures à ce seuil, représentent respectivement : 92%, 50% et 55%. Comparativement les valeurs de SI, calculées sur la totalité des grilles, supérieures à 0,4 représentent respectivement 66%, 30% et 32%.

IV.3 Analyse des interactions intermoléculaires ligands / cible

Les structures 3D des complexes protéiques de chacun des trois systèmes présentés, permettent d'analyser les conséquences de la liaison de chaque ligand vis-à-vis de sa cible, en termes d'interactions ponctuelles : liaisons hydrogène, ponts salins ou interactions coulombiennes défavorables, interactions cation- π et π - π .

IV.3.1 Interactions « π - π » et « cation- π »

Ce type d'interaction peut s'expliquer par la présence préférentielle de certains résidus, dont les résidus Arg, Trp et Tyr [144], au niveau des interfaces, de part leur aptitude à établir de interactions multiples (cation- π , π - π , liaisons hydrogène, interactions apolaires). Les critères d'attribution de ces différentes interactions sont reportés dans l'annexe B.3. Les résultats sont présentés dans le tableau IV-22.

« Trypsine Bovine »	1C9T	1D6R	1G9I	1OX1	1SFI	1TAW	1TGS	2BTC	2PTC
Nb interaction cation- π	1	0	1	0	1	0	0	0	2
Nb interaction π - π	0	0	0	0	0	0	0	0	0
« Trypsine Porcine »	1AVX	1EJA	1H9H	1LDT	1MCT	1TFX	1TX6	1UHB	1Z7K
Nb interaction cation- π	1	1	1	1	1	0	0	0	0
Nb interaction π - π	0	1	0	0	0	0	0	0	0
« Chymotrypsine bovine »	1ACB	1CAO	1CBW	1CGI	1CHO	1GLO	1GL1	1N8O	
Nb interaction cation- π	1	0	1	2	1	0	1	1	
Nb interaction π - π	0	0	0	0	0	0	0	0	

Tableau IV-22 : Interactions π - π et cation- π identifiées dans les complexes des trois systèmes étudiés.

L'analyse des différents systèmes donne des résultats conformes à ce qui a été décrit au chapitre II. Ces interactions, impliquant un résidu aromatique, sont peu fréquentes au niveau des interfaces protéiques, notamment concernant les interactions π - π : sur l'ensemble des complexes analysés, une seule interaction de ce type a été observée (complexe 1D6R). D'autre part, 65% des complexes présentent au moins une interaction cation- π , toutes impliquant un résidu chargé positivement, appartenant au ligand. Cependant, ces interactions ponctuelles ne sont pas systématiquement observées pour les différents ligands d'une même cible, et constituent donc un élément spécifique d'une interaction donnée.

IV.3.2 Interactions coulombiennes

Nous avons vu précédemment, que des propriétés communes des différents ligands d'une même cible, pouvaient être mises en évidence par le calcul d'indices de similarité des potentiels électrostatiques, dans la région d'interaction avec la cible. Ces similarités ne sont pas uniquement dues à la présence de groupements portant des charges nettes à cette interface, mais reflètent l'ensemble de la distribution de charges du ligand considéré. Cependant, ces groupements, portant des charges nettes, peuvent être engagés dans des interactions intermoléculaires impliquant des groupement, de la cible, portant une charge nette de signe opposé (pont salin) ou de même signe (interaction coulombienne répulsive). Ces différentes interactions sont présentées dans le tableau IV-23.

« Trypsine bovine »	1C9T	1D6R	1G9I	1OX1	1SFI	1TAW	1TGS	2BTC	2PTC
Ponts salins	1	1	1	1	1	1	1	1	1
Charge +/- (du ligand)	1/0	1/0	1/0	1/0	1/0	1/0	1/0	1/0	1/0
Interaction répulsive	0	0	0	0	0	0	0	1	0
« Trypsine porcine »	1AVX	1EJA	1H9H	1LDT	1MCT	1TFX	1TX6	1UHB	1Z7K
Ponts salins	1	3	1	1	1	1	2	2	1
Charge +/- (du ligand)	1/0	1/2	1/0	1/0	1/0	1/0	1/1	1/1	1/0
Interaction répulsive	0	0	0	1	0	0	0	0	0
« Chymotrypsine bovine »	1ACB	1CAO	1CBW	1CGI	1CHO	1GLO	1GL1	1N8O	
Ponts salins	0	0	0	1	1	0	0	0	
Charge +/- (du ligand)	0/0	0/0	0/0	0/1	1/0	0/0	0/0	0/0	
Interaction répulsive	0	0	0	0	0	0	0	0	

Tableau IV-23 : Analyses des interactions coulombiennes dans les trois systèmes considérés. Dans le cas de ponts salins, est indiquée, la nature de la charge impliquée dans le pont salin, coté ligand.

Pour le système « trypsine bovine », un seul pont salin est observé, correspondant dans chaque cas, à l'interaction d'un résidu basique du ligand avec le résidu Asp 171 de la trypsine, soulignant l'importance de cette interaction avec la cible. Il a effectivement été montré que la suppression de cette charge positive du résidu Lys 15 en interaction avec le résidu Asp 171, pour l'interaction BPTI-trypsine bovine, engendrait une perte d'affinité, d'un facteur 200 [143].

Dans le cas du système « trypsine porcine », le résidu Asp 171, de façon similaire au système précédent, réalise un pont salin avec un résidu basique de chacun des ligands. Cependant, d'autres ponts salins sont également observés. Les structures 1TX6 et 1UHB présentent deux ponts salins intermoléculaires, l'un impliquant un résidu du ligand chargé positivement et l'autre un résidu chargé négativement. Dans le complexe 1EJA, trois ponts salins ont été identifiés. Pour deux

d'entre eux, les résidus (du ligand) impliqués portent une charge négative alors que le troisième de ces ponts implique un résidu portant une charge positive.

Enfin, concernant les complexes du système « chymotrypsine bovine », la plupart des complexes ne présentent aucun pont salin, à l'exception des complexes 1CGI et 1CHO. Toutefois, leur nature diffère, l'un impliquant une charge négative (1CGI), l'autre une charge positive (1CHO) du ligand.

Par ailleurs, la présence d'interactions électrostatiques répulsives s'avère relativement rare pour l'ensemble des systèmes. En utilisant un critère de distance minimale de 6 Å, seules deux interactions répulsives ont été observées : structure 2BTC entre les résidus Arg 501 (ligand) et Lys 202 (cible) distantes de 5,6 Å, structure 1LDT entre les résidus Lys 11 (ligand) et Lys 43 (cible) distants de 4,96 Å du résidu Lys 43.

Les observations faites suggèrent qu'il n'existe aucune contrainte stricte quant à la conservation du nombre et du type de pont salin. D'autre part, la présence de charges ponctuelles peut avoir des conséquences défavorables, lorsque une charge ponctuelle à l'interface (donc peu accessible au solvant) n'est pas compensée ou lorsque deux charges de même signe sont à proximité. Dans ce dernier cas, même si le coût énergétique d'une telle proximité est difficile à estimer, compte tenu du bilan de solvation/désolvation des charges, la composante coulombienne de l'interaction est susceptible d'apporter une contribution défavorable à la liaison. L'analyse des proximités spatiales de résidus, portant des charges de même signe, révèle que cette situation est toutefois très peu fréquente (avec un seuil de 6 Å). Le nombre réduit de complexes analysés et la faible diversité des fonctions de protéines considérées ne permet pas de conclure à la généralité de cette situation dans les complexes protéines protéines.

IV.3.3 Liaisons hydrogène

Les liaisons hydrogène, établies à l'interface, ont été analysées et regroupées en différentes catégories : nombre total de liaisons, nombre de résidus accepteurs et de résidus donneurs (du ligand) et densité de liaisons hydrogène (pour 100 Å²). Les résultats sont présentés dans le tableau IV-24.

« Trypsine bovine »	1C9T	1D6R	1G9I	1OX1	1SFI	1TAW	1T6S	2BTC	2PTC
Liaisons hydrogène	6	9	9	7	11	6	10	7	10
Donneur/accepteur	2/3	5/3	5/3	4/3	6/5	2/3	4/4	3/3	5/5
Densité de liaison H	0,6	1,0	1,0	0,9	1,3	0,7	0,9	0,7	1,1
« Trypsine porcine »	1AVX	1EJA	1H9H	1LDT	1MCT	1TFX	1TX6	1UHB	1Z7K
Liaisons hydrogène	8	10	8	6	5	6	7	6	11
Donneur/accepteur	2/7	3/3	3/4	3/2	3/1	2/3	3/3	2/4	3/8
Densité de liaison H	0,8	0,9	0,8	0,7	0,5	0,7	0,6	0,8	1,0
« Chymotrypsine bovine »	1ACB	1CAO	1CBW	1CGI	1CHO	1GLO	1GL1	1N8O	
Liaisons hydrogène	9	8	7	11	6	6	8	8	
Donneur/accepteur	5/4	3/2	2/2	5/4	2/3	3/2	5/2	4/3	
Densité de liaison H	1,0	1,0	0,8	0,9	0,7	0,5	0,8	0,8	

Tableau IV-24 : Paramètres globaux analysés pour chaque ligand appartenant aux trois systèmes étudiés.

Pour l'ensemble des systèmes, le nombre de liaison hydrogène varie de manière significative : de 6 à 11 pour les complexes de la trypsine bovine et de la chymotrypsine et de 5 à 11 pour les complexes de la trypsine porcine. Cependant, la surface du ligand, en contact avec la cible, peut présenter des variations non négligeables (tableau IV-10). Pour tenir compte de cette variation dans l'analyse des liaisons hydrogène, il a été proposé de calculer la densité superficielle de liaisons hydrogène en calculant le nombre de ces interaction pour 100 Å² d'interface. Pour les trois systèmes étudiés, les valeurs de densité de liaisons hydrogène sont comprises entre 0,6 et 1,3 pour le système « trypsine bovine », entre 0,5 et 1,0 pour le système « trypsine porcine » et entre 0,5 et 1,0 pour le système « chymotrypsine bovine ». La variabilité de ce paramètre, dans ces trois cas, semble donc relativement importante.

Nous avons également analysé les groupements, de chaque cible, impliqués dans les liaisons hydrogène intermoléculaires avec les différents ligands. Ceux-ci sont représentés dans les tableaux IV-25, IV-26 et IV-27 pour les trois systèmes, « trypsine bovine », « trypsine porcine » et « chymotrypsine bovine » respectivement. Dans chacun d'eux, les liaisons hydrogène impliquant des groupements des chaînes principale (MC) ou latérale (SC) du résidu de la protéine

cible sont différenciées. Les groupements présents à l'interface de la cible et impliqués dans des ponts salins intermoléculaires avec le ligand sont également reportés dans ces tableaux. Nos résultats suggèrent que certains groupements de la cible sont systématiquement impliqués. Ainsi, dans les systèmes « trypsine bovine » et « trypsine porcine », les résidus G175, S177, S192 et G194 de la cible forment une liaison hydrogène, par l'intermédiaire d'un groupement de la chaîne principale. Dans le système « chymotrypsine bovine », la participation des résidus F24, G193, S195 et G216 dans la formation de liaison hydrogène avec des groupements de la chaîne principale, se révèle systématique. Toutefois, pour l'ensemble des systèmes, de nombreuses liaisons hydrogène, impliquant des groupements des chaînes principale et latérale, varient d'un complexe à l'autre. Malgré la faible taille des échantillons analysés, les comparaisons des complexes indiquent que les liaisons hydrogènes, établies lors de l'interaction des différents ligands avec une même cible, ne sont conservées ni en nature (chaînes latérale ou principale, groupements accepteur et donneur), ni en nombre.

Trypsine bovine

	Y22	H23	R24	S28	N79	T80	Q155	D171	Q174	G175	S177	S182	G194	Y195	G196	S195	G194	G194	G194	
1G1T			AC		SC			SB			AC	AC	AC	AC			AC			G194
1U6R					AC(2)		SC	SB		AC	AC	AC	AC	AC(2)			AC	AC(2)	AC(2)	SC
1G8I			AC	AC				SB	SC	AC	AC	AC	AC	AC(2)			AC	AC(2)	AC(2)	
1D9I			AC					SB	SC	AC	AC	AC	AC	AC(2)			AC	AC(2)	AC(2)	
1SFI			AC		AC	AC	SC	SB	SC	AC	AC	AC	AC	AC(2)			AC	AC(2)	AC(2)	SC
1TA'W			AC					SB	SC	AC	AC	AC	AC	AC			AC	AC	AC	
1TGS								SB	SC	AC	AC	AC	AC	AC			AC	AC	AC	
2B7C	SC		AC					SB		AC	AC	AC	AC	AC(2)			AC	AC(2)	AC(2)	
3P7C		SC	AC		AC(2)		SC	SB	SC	AC	AC	AC	AC	AC			AC	AC	AC	

Trypsine porcine

	R24	K43	G78	N79	W121	G128	S129	Y131	D171	Q174	G175	S177	S182	G194	Y195	G196	S195	G194	K200	
1AVX	AC	SC		SC					SB		AC	AC	AC	AC				AC		
1E3A	AC	SB			AC				SB	SC	AC	AC	AC	AC				AC		SB
1HPH	AC						SC		SB	SC	AC	AC	AC	AC(2)						
1LDT	AC								SB	SC	AC	AC	AC	AC						
1WGT	AC								SB	SC	AC	AC	AC	AC						
1TFX	AC								SB	SC	AC	AC	AC	AC						
1TX6	AC		AC	SC					SB	SC	AC	AC	AC						SC	
1UHB		SC/5B							SB	SC	AC	SC(3)								
1Z7K	AC	SC					SC	SC	SB	SC	AC	AC	AC	AC(2)					SC	

Chymotrypsine bovine

	H48	F41	G58	G59	L97	T98	Y146	A149	N150	Y171	K175	K177	S190	G193	S195	S214	G216	S217	S218	
1A'CB		AC	AC				SC							AC	AC	AC	AC	AC(2)		SC
1C49		AC											AC	AC	AC	AC	AC	AC	AC	
1CBW	AC	AC							SC					AC	AC	AC	AC	AC	AC	
1GGI		AC	AC(2)	AC			SC	AC			SC	SB		AC	AC	AC	AC(2)			
1CHD		AC								AC				AC	AC	AC	AC(2)			AC
1G4A		AC												AC	AC	AC	AC(2)			
1G4I		AC	AC											AC	AC	AC	AC(2)			
1N50		AC												AC	AC	AC	AC(2)	AC		SC

Tableaux IV-25, 26 et 27 : Description des différentes interactions (ponts salins et liaisons hydrogène), impliquées dans la formation de chaque complexe de chacun des trois systèmes analysés. Les résidus indiqués sont ceux de la cible considérée (trypsine bovine, trypsine porcine et chymotrypsine bovine).

IV.3.4 Conclusion

L'analyse des interactions ponctuelles, mises en jeu au sein des complexes de chaque système, nous a permis de faire des observations quant à la relation entre la capacité des différents ligands à s'associer à la cible commune et la reproduction d'interactions ponctuelles entre les deux partenaires. Bien que notre échantillon présente une taille faible et une diversité de cibles peu importante, nos résultats démontrent que ces différents ligands ne reproduisent pas les mêmes interactions ponctuelles (π - π , cation- π , ponts salins, liaisons hydrogène). *A contrario*, des combinaisons variées d'interactions ponctuelles peuvent concourir à la formation d'un complexe avec une cible déterminée. D'autre part, les exemples analysés suggèrent que les fluctuations du nombre d'interactions des différents types au sein des ligands de chaque système sont comparables aux observations faites lors de la comparaison de ligands sélectionnés indépendamment ou non de leur fonction [52, 131, 139, 138,159]. Il est à noter que la taille des échantillons et la faible diversité des cibles considérées peuvent être à l'origine de la difficulté à conclure sur la spécificité, du nombre et du type d'interaction, pour une cible donnée. Cet aspect pourrait faire l'objet d'études ultérieures tirant partie de l'accroissement du nombre de complexes protéiques présents dans la PDB (annexe A.2)

L'étude réalisée révèle que le nombre des interactions ponctuelles et les groupements impliqués au niveau de la cible, ne sont conservés pour les différents ligands. La sélection des plateformes, dans l'approche de conception de ligands proposée, ne peut donc se fonder sur la reproduction d'interactions ponctuelles particulières, observées dans le complexe de référence. D'autres critères semblent néanmoins plus pertinents. Le premier concerne la complémentarité des surfaces des partenaires en interaction, conduisant ainsi à l'absence de recouvrement stérique entre eux. Le second correspond à des aspects électrostatiques. Il a effectivement été démontré que, malgré une variabilité du potentiel global généré par les divers ligands, les potentiels électrostatiques dans la région interface présentaient des similarités. Cette similitude reflète les résultats de nombreux travaux consacrés au rôle du potentiel électrostatique dans l'association protéine-protéine [178, 179, 183]. Dans le cadre de la conception de ligands protéiques, la recherche des plateformes présentant un potentiel électrostatique à l'interface similaire à celui du ligand de référence permettrait d'accroître la probabilité d'identifier des plateformes, susceptibles de lier effectivement la cible considérée.

Chapitre V : À la recherche de plateformes protéiques

*« Une condition essentielle de l'hypothèse (scientifique) c'est qu'elle soit aussi probable que possible »
(C. Bernard)*

L'approche proposée, décrite dans le chapitre III, a été appliquée au système BgK-Kv1.2. Le motif de liaison F6/S23/K25/Y26, identifié à partir de « l'alanine scanning » de la BgK, a été utilisé pour rechercher toutes les structures 3D possédant un motif de résidu de topologie équivalent.

Les solutions, obtenues à l'issue de cette recherche, ont ensuite été filtrées, selon des critères stériques et électrostatiques, présentés dans le chapitre III. Les structures 3D résultantes, alors dénommées « plateformes structurales », sont aptes à présenter le motif de liaison F6/S23/K25/Y26, dans sa topologie fonctionnelle.

Le nombre de plateformes protéiques est dépendant des paramètres utilisés lors de la recherche. Or, notre approche est limitée par des contraintes expérimentales, liées à la production et à la caractérisation de ces plateformes. Le nombre initial de solutions doit, de ce fait, être limité. L'étude des critères de recherche s'avère donc essentielle afin de contrôler ce nombre, tout en conservant les solutions topologiquement correctes.

Dans ce chapitre, nous analyserons les différents critères de recherche et leur impact sur le nombre et les caractéristiques des solutions identifiées. Puis, nous détaillerons un exemple de recherche de plateformes protéiques appliqué au système de référence, pour un jeu de paramètres donné. Enfin, les différentes plateformes protéiques conçues au cours de cette thèse seront décrites.

V.1 Influence des paramètres de recherche

Chaque recherche effectuée aboutit à l'identification de N motifs dont la topologie est similaire à celle du motif de référence. Ces derniers sont portés par N' structures 3D, identiques ou non, celles-ci appartenant à N'' fichiers PDB. L'analyse de ces trois paramètres (N , N' et N'' , représentés sur la figure V-1), permet d'estimer l'influence des critères de recherche : 1) taille_MAX, 2) taille du motif, 3) RMSD_MAX et Δd _MAX et 4) %R_MAX (présentés dans le Chapitre III).

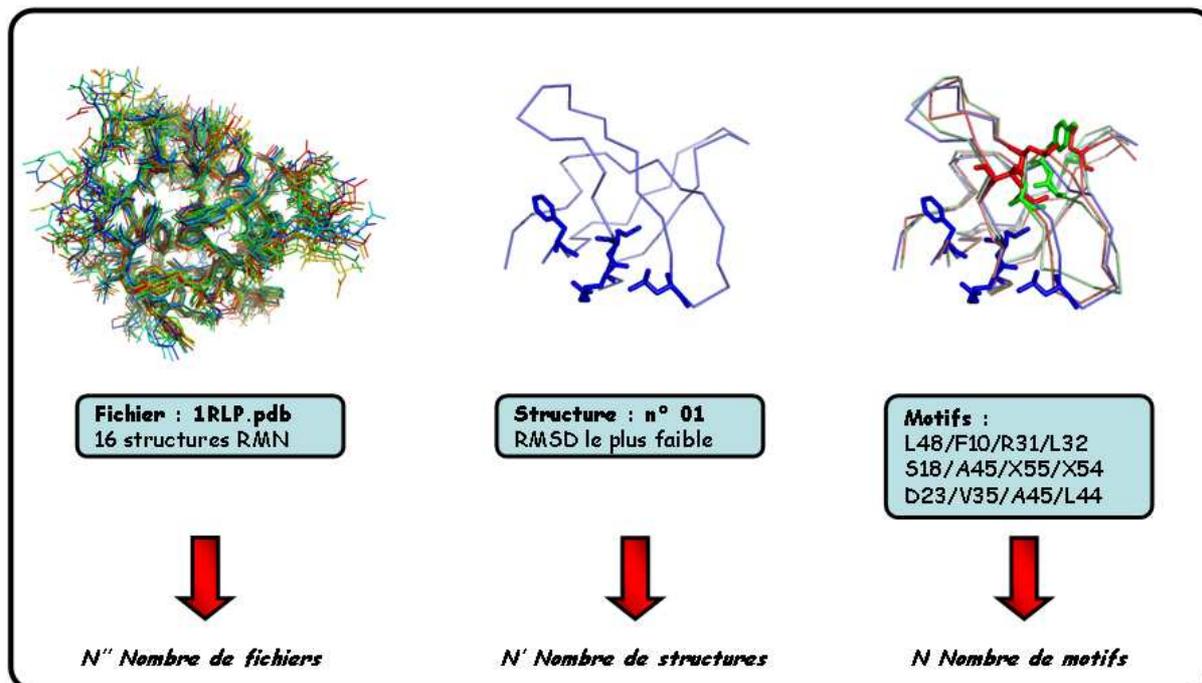


Figure V-1 : Le programme STAMP identifie des motifs de topologie équivalente à celle du motif de référence. Le nombre N correspond au nombre de motifs identifiés, portés par N' structures protéiques de N'' fichiers PDB. Le ratio N/N' est égal au nombre de motifs moyen par structure. Le ratio N/N'' est égal au nombre moyen de structures par fichier. A gauche, les différents conformères présents dans le fichier 1RLP [253] sont représentés en mode « line ». Au milieu, le conformère n°1 (en bleu) est représenté en mode « ribbon ». Le motif (en bleu foncé) est représenté en mode « stick ». A droite, plusieurs motifs, représentés en mode « stick », sont portés par les conformères n°1, 3 et 8 (en bleu, rouge et vert respectivement), représentés en mode « ribbon ».

V.1.1 Influence de la taille des structures

La taille des structures est limitée par la nature des objets recherchés (mini-protéines, inférieures à 100 résidus) d'une part, et par la méthode de synthèse envisagée d'autre part (présentée dans le chapitre VI). Plusieurs recherches ont été effectuées, pour des tailles maximales comprises entre 50 et 110 résidus. Les résultats sont présentés dans le tableau suivant V-2 :

Taille des plateformes (en nombre de résidus)	50	70	90	110
Nombre de fichiers pdb (N'')	51	131	267	428
Nombre de structures 3D (N')	98	328	741	1286
Nombre de motifs topologiques (N)	100	333	772	1357

Figure V-2 : Évolution du nombre de solutions en fonction de la taille des protéines recherchées. Les autres paramètres sont : motif à 8 atomes, RMSD_MAX : 1,0 Å, Δd _MAX : 1,0 Å, pas de filtre stérique.



Plusieurs observations peuvent être faites :

- 1) Le nombre de fichiers N'' et le nombre de structures N' diffèrent, ce qui s'explique notamment par la présence de conformères multiples parmi les structures résolues par RMN.
- 2) Le nombre de structures est peu différent du nombre de motifs, ce qui signifie qu'en moyenne chaque structure ne présente qu'un seul et unique motif.
- 3) En revanche, lorsque la valeur de ce paramètre augmente, le nombre de motifs identifiés par structure augmente également. L'augmentation de la taille des protéines accroît la probabilité d'identifier des motifs multiples au sein d'une même structure 3D. En raison du temps nécessaire à la synthèse chimique des plateformes, nous avons fixé, dans un premier temps, à 70 résidus, la taille maximale de mini-protéines synthétisées. Toutefois, la variation du nombre de solutions étant faible (facteur 4) pour un critère de taille variant de 70 à 110 résidus, ce dernier pourrait être augmenté pour des recherches ultérieures.

V.1.2 Influence de la taille du motif

Le motif de référence, composé de 4 résidus (F6/S23/K25/Y26) est représenté par 4 paires (Ca ; C β), soit 8 atomes. Plusieurs recherches ont été effectuées pour des motifs contenant de 3 résidus (F6/K25/Y26) à 5 résidus (F6/H13/S23/K25/Y26), soit de 6 à 10 atomes (une paire Ca ; C β pour chaque). Les résultats sont synthétisés dans le tableau V-3 :

Taille du motif (en nombre de résidus)	5	4	3
Nombre de fichiers pdb (N'')	35	861	2700
Nombre de structures 3D (N')	79	4941	25348
Nombre de motifs topologiques (N)	79	6476	349864

Tableau V-3 : Évolution du nombre de solutions en fonction du nombre d'atomes du motif de référence. Les autres paramètres sont : taille_MAX : 70 résidus, RMSD_MAX : 1.0 Å, Δd _MAX : 1.0 Å, pas de filtre stérique.

Lorsque la taille du motif diminue, de 5 à 3 résidus :

1) Le nombre de structures par fichier augmente, de 2 à 9. Le paramètre taille_MAX étant fixé à 70 résidus, cet accroissement ne peut être expliqué par la présence de conformères ou de structures multiples au sein des fichiers, mais uniquement par l'identification d'un nombre de solutions supplémentaires.

2) Le nombre de motif par structure varie de 1 à 14, témoignant d'une importante dégénérescence des solutions. L'accroissement du nombre de solutions est particulièrement important en dessous de 4 résidus. Les étapes de sélection s'avèrent alors relativement longues, mais demeurent réalisables.

3) Enfin, le nombre de fichiers de structure, pour un motif de 3 résidus, correspond approximativement à 40% (2700 fichiers sur 7000) de l'ensemble des fichiers contenant, au minimum, une protéine de moins de 70 résidus. Cette fraction importante révèle la dégénérescence des solutions pour un motif de taille inférieure à 4 résidus.

Le nombre de solutions peut, en outre, être modulé par le choix du type de résidu recherché. Par exemple, le choix de motifs incluant un résidu Gly, ne possédant pas C β , entraîne une diminution de la taille du motif, de 8 à 7 atomes. Le fait d'imposer un type particulier de résidu, Gly en l'occurrence, entraîne une diminution importante du nombre de solutions : de 36237 à 2626 sites identifiés. Toutefois, plusieurs tests préliminaires ont montré que le nombre d'atomes et le type de résidu ne sont pas les seuls paramètres à considérer. La topologie spécifique des atomes du motif joue également un rôle important. Compte-tenu des ces résultats, la taille du motif de recherche sera limité à 4 résidus, incluant la possibilité qu'un résidu Gly soit présent. Les motifs utilisés pour la recherche par STAMP sont schématisés sur la figure V-4.

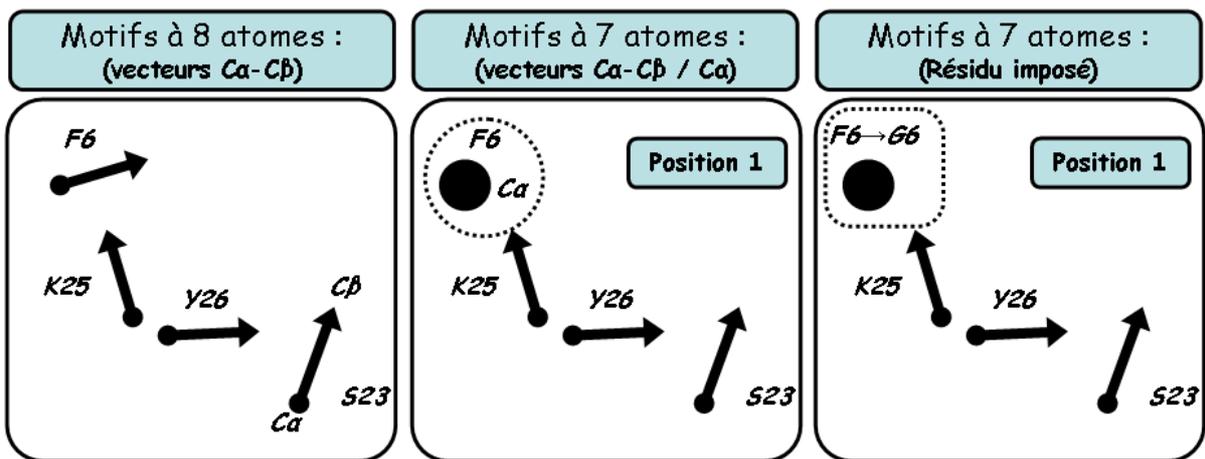


Figure V-4 : Représentation schématique des motifs atomiques pouvant être utilisés pour la recherche de plateformes. Les motifs à 3 résidus étant trop peu restrictifs, et ceux à 5 résidus l'étant trop, seuls des motifs à 4 résidus seront utilisés.

V.1.3 Influence du RMSD_MAX

L'utilisation du RMSD, comme critère de sélection, est nécessaire afin de contrôler l'écart à la topologie du motif de référence. D'après la littérature, deux structures peuvent être considérées comme similaires lorsque la valeur du RMSD, calculée *via* la superposition de leurs

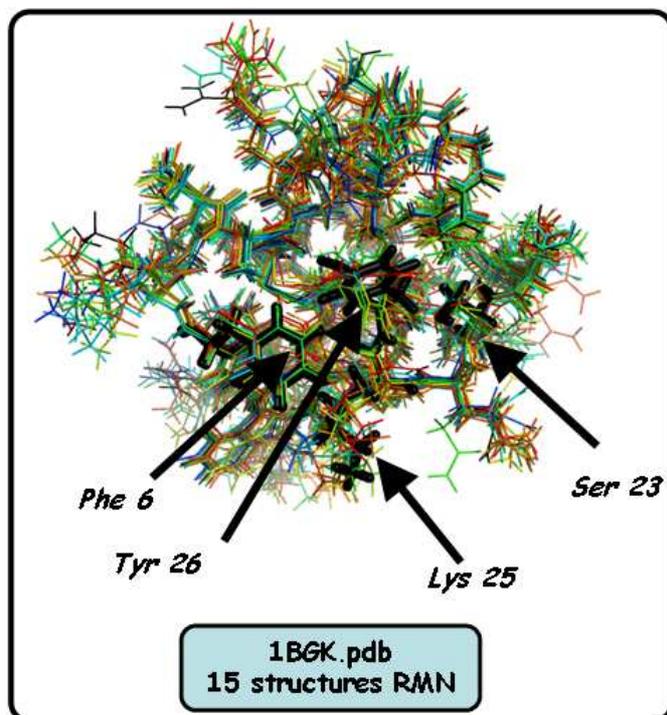


Figure V-5 : Superposition des 15 conformations RMN de la BgK [200] sur le premier conformère (couleur noir) du fichier PDB. La superposition a été effectuée par les atomes Ca. Les structures sont représentées en mode « line ».

atomes Ca, est inférieure à 2 Å. Au delà de cette valeur seuil, diverses solutions présentant des topologies significativement différentes peuvent être conservées. D'autre part, les structures de protéines sont sujettes à une variabilité structurale, pouvant être attribuée à leur flexibilité intrinsèque ou à différents facteurs extrinsèques liés aux méthodes de résolution de structure. A titre d'exemple, la superposition des différents conformères de la BgK, issue de l'analyse RMN, illustrée sur la figure V-5, fournit des valeurs de RMSD s'échelonnant de 0,14 à 0,64 Å, indiquées dans le tableau V-5bis.

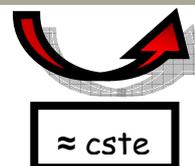
N° du conformère	1	2	3	4	5
RMSD chaîne principale (en Å)	0,00	0,44	0,62	0,14	0,43
RMSD motif (en Å)	0,00	0,29	0,38	0,13	0,32
N° conformère	6	7	8	9	10
RMSD chaîne principale (en Å)	0,56	0,41	0,44	0,26	0,45
RMSD motif (en Å)	0,25	0,21	0,27	0,38	0,33
N° conformère	11	12	13	14	15
RMSD chaîne principale (en Å)	0,64	0,48	0,48	0,42	0,59
RMSD motif (en Å)	0,48	0,38	0,38	0,54	0,39

Tableau V-5bis : Valeurs du RMSD, calculé après superposition des différents conformères de la BgK sur la première conformation du fichier PDB, via leurs atomes Ca.

Nous avons donc considéré, la valeur de 0,64 Å, comme une limite en dessous de laquelle les différences ne peuvent être considérées comme significatives. Pour ces raisons, le critère RMSD_MAX doit être compris dans l'intervalle 0,7-2,0 Å. Les résultats de plusieurs recherches, effectuées dans cet intervalle, sont présentés dans le tableau V-6 :

RMSD_MAX (en Å)	0.75	1.00	1.25	1.50	1.75
Nombre de fichiers pdb (N'')	170	861	1200	1291	1298
Nombre de structures 3D (N')	558	4941	8058	8838	8915
Nombre de motifs topologiques (N)	580	6476	13937	16230	16468

Tableau V-6 : Évolution du nombre de solutions en fonction du critère RMSD_MAX. Les autres paramètres sont : motif à 8 atomes, 70 résidus, Δd_{MAX} : 1,5 Å, pas de filtre stérique.



La variation du paramètre RMSD_MAX aboutit aux observations suivantes :

- 1) Le nombre de structures identifiées par fichier est relativement constant ($\approx 6-7$) pour des valeurs de RMSD_MAX supérieures ou égales à 1 Å.
- 2) Le nombre de sites identifiés par structure augmente modérément avec l'augmentation du RMSD_MAX : de 1,30 à 1,85 sites (en moyenne) pour des valeurs de RMSD maximum allant de 1 Å à 1,75 Å. En considérant la valeur seuil supérieur, égale à 2 Å, le nombre de sites topologiques similaires ne dépasse pas 2. Le risque d'avoir plusieurs sites de liaison, fonctionnels (possédant des résidus identiques ou homologues au motif de référence), par protéine apparaît, dès lors, négligeable.
- 3) Le nombre de solution varie peu lorsque la valeur du RMSD_MAX dépasse la valeur du Δd_{MAX} . La valeur du RMSD est un paramètre global, qui tient compte de la topologie globale du motif, mais demeure néanmoins lié au paramètre d'incertitude sur les distances. Plus l'incertitude tolérée est grande et plus la valeur du RMSD le sera aussi. Inversement, si la contrainte sur les distances est forte (écart faible sur les distances), alors les motifs identifiés présenteront nécessairement un RMSD faible.

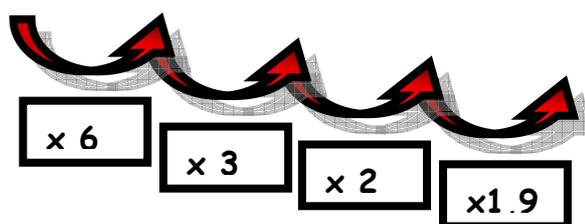
Il apparaît que, utilisée seule, la valeur du paramètre RMSD_MAX n'est pas suffisamment contraignante dans le cas de recherche de motif à 4 résidus : d'une part, le nombre total de sites ainsi que le nombre de sites par structure varient modérément. D'autre part, pour une valeur de RMSD_MAX donnée, le nombre total de sites est conditionné par la valeur de contrainte sur les distances. De ce fait, nous pourrions librement faire varier ce paramètre, dans l'intervalle précédemment justifié.

V.1.4 Influence du Δd_{MAX}

Les recherches ont été effectuées en faisant varier la contrainte de distance de 1 à 2 Å, tout en maintenant les autres paramètres constants. Les résultats sont présentés dans le tableau V-7 :

Δd_{MAX}	1.00	1.25	1.50	1.75	2.00
Nombre de fichiers pdb (N'')	131	428	861	1247	1603
Nombre de structures 3D (N')	328	1752	4941	8462	12239
Nombre de motifs topologiques (N)	333	2018	6476	14559	27574

Tableau V-7 : Évolution du nombre de solutions en fonction du Δd_{MAX} . Les autres paramètres sont : motif à 8 atomes, 70 résidus, $RMSD_{MAX}$: 1,0 Å, pas de filtre stérique.



- 1) Lorsque le critère Δd_{MAX} est augmenté, le nombre de solutions croît de façon importante. De 1 à 2 Å, le nombre de motifs est multiplié par un facteur 80.
- 2) Toutefois, malgré cet accroissement important, le nombre moyen de motifs par structure (entre 1 et 2 Å) varie de 1 à 2,3. La probabilité d'avoir plusieurs sites fonctionnels par structure reste relativement faible.
- 3) Enfin, bien que la valeur Δd_{MAX} dépasse la valeur du $RMSD_{MAX}$ (égale à 1,0 Å), le nombre de solutions s'accroît. Le choix du critère d'incertitude sur les distances s'avère donc très important : lorsque celui-ci est supérieur à la valeur du $RMSD_{MAX}$, ce dernier ne permet plus d'éliminer l'ensemble des solutions dont la topologie est incorrecte. Ce cas est schématisé sur la figure V-8, A. Pour ces raisons, de façon à conserver la topologie d'un motif de vecteurs ($Ca-C\beta$), le critère Δd_{MAX} doit être limité. La distance moyenne des liaisons C-C, estimée à 1,54 Å. Donc, lorsque la valeur du Δd_{MAX} dépasse cette valeur, certains vecteurs $Ca-C\beta$ des motifs solutions peuvent alors être anti-colinéaires avec les vecteurs correspondant du motif de référence, comme le montre la figure V-8, B. Au-delà de 1,5 Å, les recherches par STAMP identifient des solutions de topologie incorrecte, qui ne peuvent être éliminées de façon systématique par le critère $RMSD_{MAX}$. Pour ces raisons, les critères ($RMSD_{MAX} / \Delta d_{MAX}$) ont été choisis en respectant les contraintes suivantes : la valeur du $RMSD_{MAX}$ doit être inférieure ou égale à la valeur Δd_{MAX} ; la valeur ne du Δd_{MAX} ne doit pas dépasser 1,5 Å.

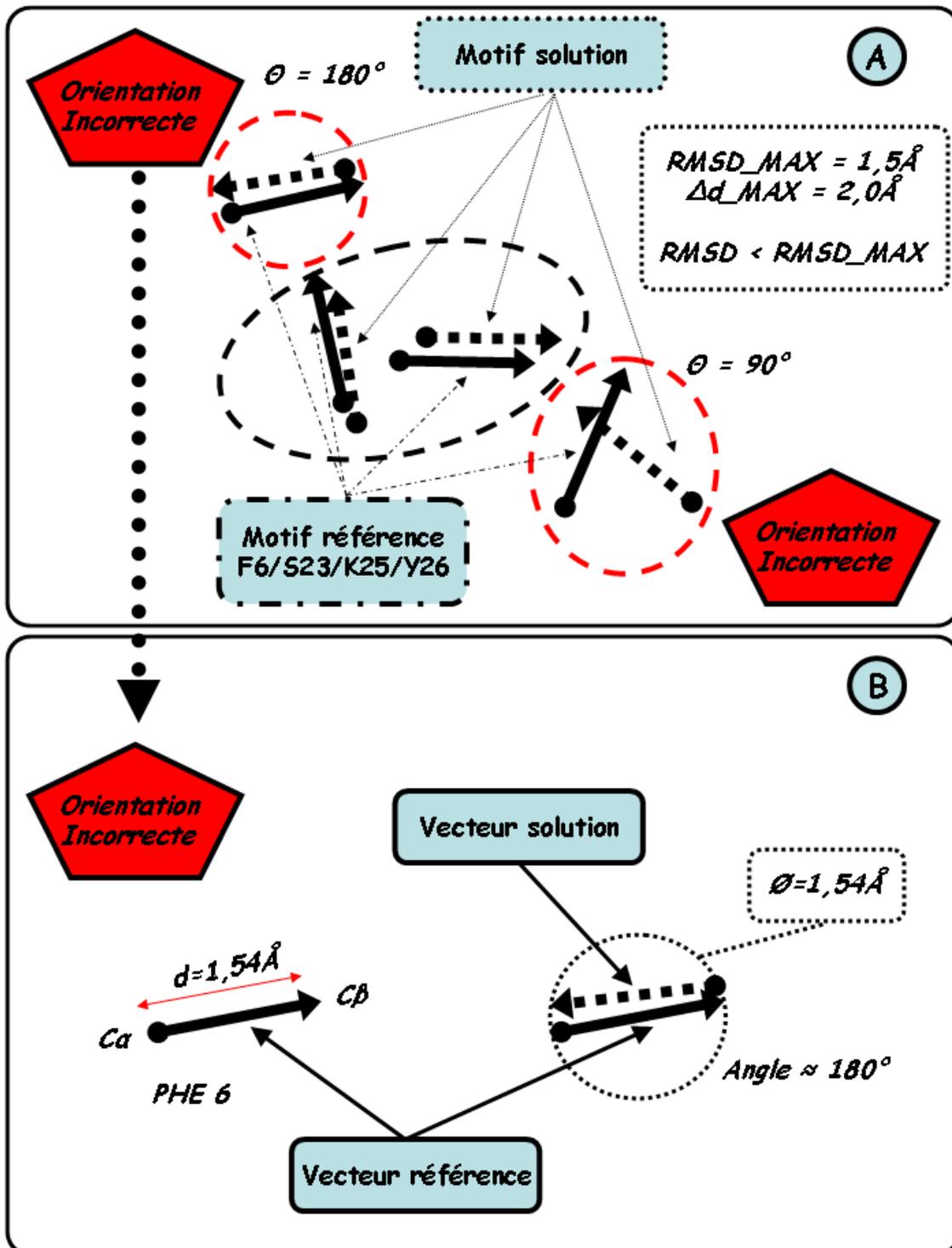


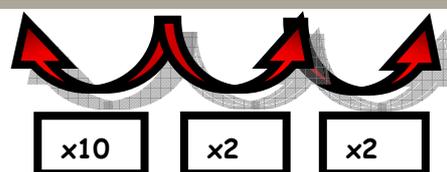
Figure V-8 : Partie A, exemple de recherche de motif par STAMP. Pour une valeur seuil de RMSD faible, mais d'incertitude sur les distances élevée, le motif sera sélectionné par STAMP malgré une topologie incorrecte. Partie B, l'effet de l'incertitude sur les distances dépassant $1,5\text{\AA}$ quant à l'orientation relative des chaînes latérales du motif de référence et des motifs solution.

V.1.5 Influence du filtre stérique

L'utilisation du filtre stérique a été étalonnée avant toute utilisation. La détermination d'un seuil minimum a été effectuée par l'observation, *via* l'utilisation d'un logiciel de visualisation graphique, de structures 3D d'un ensemble de solutions en interaction avec le canal Kv1.1. Aucun recouvrement de chaînes principales n'a été observé pour une valeur de R%_MAX inférieure à 2%, constituant une valeur limite inférieure. L'influence de ce critère est représentée dans le tableau V-9 :

Filtre stérique (en % de recouvrement)	Sans	2%	3%	4%
Nombre de fichiers pdb (N')	508	83	150	233
Nombre de structures 3D (N')	2402	249	467	889
Nombre de motifs topologiques (N)	2735	256	478	932

Tableau V-9: Évolution du nombre de solutions avec ou sans utilisation du filtre stérique. Les autres paramètres sont : motif à 8 atomes, 70 résidus, $\Delta d_MAX : 1,0\text{\AA}$, $RMSD_MAX : 1,0\text{\AA}$.



L'utilisation du filtre stérique n'a aucune influence sur le nombre moyen de motifs par structure ni sur le nombre de structures/conformères par fichiers. Cette observation rend compte du fait que l'utilisation du filtre stérique n'influence, ni la topologie du motif, ni le calcul des distances. Seuls sont éliminés les motifs, dont la structure qui le porte est inadéquate pour l'interaction avec la cible. D'après nos résultats, l'application de ce filtre est très restrictive quant au nombre de solutions conservées : un %R_MAX de 2% (comparativement à l'absence de filtre) entraîne une diminution d'un facteur 10 du nombre total de motifs identifiés. Par ailleurs, l'examen visuel des solutions conservées montre, pour des valeurs de 4%, des recouvrements importants d'atomes des chaînes principales. Cette valeur représente donc une limite supérieure pour l'utilisation de ce critère.

V.1.6 Conclusion : choix des paramètres de recherche

A partir des résultats précédents, nous avons déterminé, pour chacun des paramètres du programme STAMP, un intervalle optimal pour la recherche de motifs topologiques équivalents au motif de référence :

- **RMSD_MAX** compris entre 0,75 et 2,00 Å (mais inférieur au critère Δd_MAX) ;
- **Δd_MAX** compris entre 0,75 et 1,50 Å ;
- **taille des protéines** de 70 résidus (dépend de l'étape de synthèse) ;
- **motif de 4 résidus** (soit 7 ou 8 atomes) ;
- **%R_MAX** inférieur à 4%.

Le choix des paramètres de recherche permet de contrôler, approximativement, le nombre de plateformes identifiées à la fin de l'approche, comme le montre la figure V-10. Ainsi, les contraintes liées à l'expérimentation, peuvent être estimées avant toute recherche.

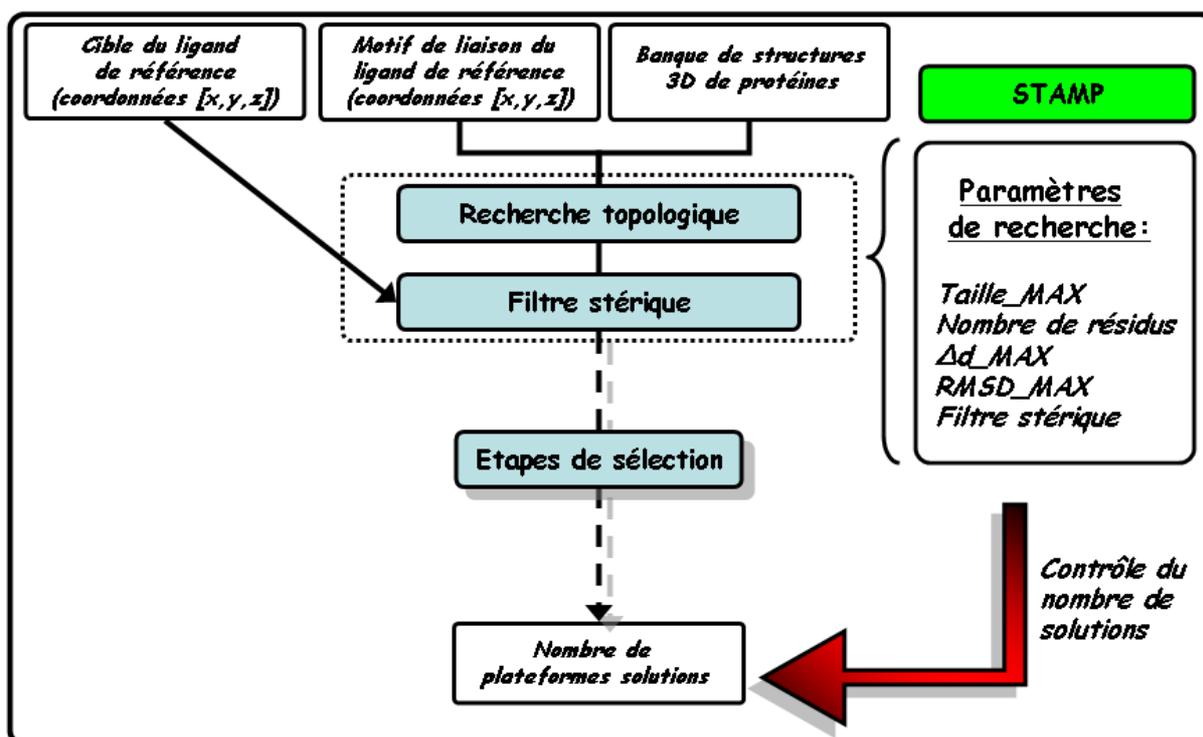


Figure V-10 : Notre approche de conception de ligands. Le choix des paramètres de recherche détermine le nombre de plateformes structurales aptes à porter le motif de liaison dans une topologie similaire à celle du ligand de référence.

V.2 Identification des plateformes protéiques

La partie suivante présentera un exemple de recherche effectuée au cours de cette thèse. Les résultats obtenus, à chaque étape, seront détaillés.

V.2.1 Recherche de motifs topologiques

Pour un motif de référence de 4 résidus (F6/S23/K25/Y26), la recherche de motifs topologiques similaires, incluant au plus un seul résidu glycine, nécessite 5 cycles de recherche, représentée sur la figure V-11. Ceci est lié à la version actuelle du programme STAMP, ne permettant pas la détermination de la position d'un atome C β théorique pour les résidus Gly.

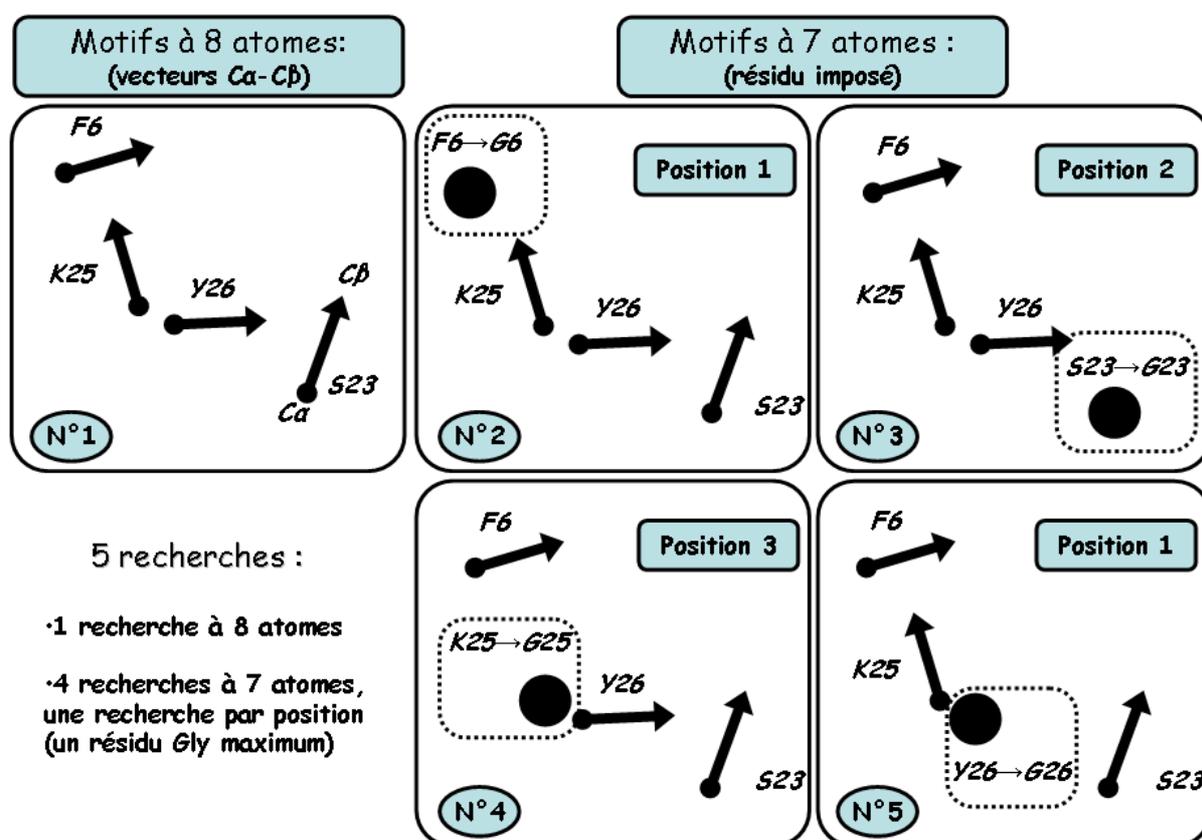


Figure V-11 : Motifs de référence pour une recherche par STAMP, soit sans résidu glycine (motif à 8 atomes) soit en incluant au maximum un résidu glycine aux positions 1, 2, 3 ou 4 (motif à 7 atomes). Les motifs sont représentés sous formes de vecteurs lorsque les atomes Ca et C β sont considérés, ou sous forme de point lorsque seul le Ca est considéré.

Une première recherche est effectuée à partir du motif complet, comprenant 8 atomes, au quel cas aucun résidu glycine ne pourra être présent parmi les motifs solutions. Quatre autres recherches seront nécessaires, selon la position autorisée pour le résidu glycine. Pour chacune de ces recherches, le motif est successivement réduit à 7 atomes par la suppression d'un atome C β à la position souhaitée, en imposant un résidu Gly à la même position (représenté dans les

tableaux par F6→G(Ca) pour la position 1). Les résultats obtenus pour ces cinq recherches sont présentés dans le tableau V-12, en utilisant les paramètres suivant :

RMSD_MAX : 1,0Å / Δd_MAX : 1,3Å / taille_MAX : 70 résidus / %R_MAX : 2%

N°1
N°2
N°3
N°4
N°5

Recherche	F6→G(Ca)	S23→G(Ca)	K25→G(Ca)	Y26→G(Ca)	FSKY	Total
<i>Nombre de motifs solutions (N)</i>	1302	2717	456	1080	2735	8291
<i>Après application du filtre stérique</i>	97	361	60	113	256	887(10%)

Tableau V-12 : Résultats des cinq recherches. FSKY correspond à la prise en compte des 8 atomes du motif, F6→G(Ca) correspond à la recherche de motifs possédant une glycine en position 1, S23→G(Ca) une glycine en position 2, K25→G(Ca) une glycine en position 3, Y26→G(Ca) une glycine en position 4. Le % de solutions identifiées est indiquée dans la dernière colonne en cyan.

Selon le motif utilisé, le nombre de solutions est relativement variable, quand bien même le nombre d'atomes et de résidus est identique. Cette observation corrobore l'affirmation précédente, selon laquelle la topologie spécifique du motif de référence utilisée influe sur le nombre de solutions.

La recherche de motifs topologiques similaires au motif de référence conduit à l'identification finale de 8291 solutions. Ce nombre important peut être expliqué par différentes raisons : 1) aucune sélection des structures 3D adéquates n'a encore été effectuée à cette étape, 2) la recherche n'est réalisée qu'avec seulement 7 ou 8 contraintes de distances, 3) les différents motifs sont des combinaisons de résidus divers parmi les 20 résidus naturels (avec, au plus, un seul résidu Gly), 4) enfin, la recherche est réalisée indifféremment de la localisation des sites, à l'intérieur ou à la surface, des structures protéiques qui les portent. L'utilisation du filtre stérique permet de remédier à cette dernière limitation. Parmi le grand nombre de solutions, seules les structures permettant un positionnement correct du motif au niveau du site de liaison du canal Kv1.2 sont conservées, ce qui implique, de façon implicite, que le motif soit positionné en surface. L'application de ce filtre mène à l'élimination de 90% des solutions. A partir de 23845 fichiers PDB (version août 2004), la recherche a conduit à l'identification de 887 motifs dont la topologie est équivalente à celle du motif F6/S23/K25/Y26.

V.2.2 Sélection *in-silico* des plateformes

Les solutions, issues de la recherche précédente, sont ensuite analysées et filtrées par l'utilisation de plusieurs critères, décrits dans le chapitre III. Ces critères et leur impact sont les suivants.

V.2.2.A Élimination des solutions redondantes

Les solutions redondantes ne sont pas conservées, ce qui inclut les structures résolues par RMN possédant de multiples conformères (jusqu'à 72) ainsi que les structures identiques présentes dans un même fichier. Dans ces cas, seule la structure présentant le RMSD le plus faible a été conservée.

D'après le calcul du ratio N' (nombre de structures) /N'' (nombre de fichiers), chaque fichier solution contient en moyenne cinq structures. L'utilisation de ce critère pourrait donc entraîner une diminution de 80% des solutions.

V.2.2.B Élimination des solutions contenant des résidus cystéine ou proline

Compte-tenu de l'importance des résidus cystéine et proline dans les structures secondaire et tertiaire des protéines (cf. Chapitre III), les motifs possédant un ou plusieurs de ces résidus sont exclus. En supposant que la fréquence de chaque résidu est la même, le nombre de combinaison aléatoires, pour un motif de 4 résidus est de 4845 solutions (pour 20 résidus naturels). En absence de résidus Cys ou Pro, ce nombre n'est plus que de 3060 solutions, soit la suppression approximative de 27% des solutions.

Remarque

Les deux critères précédents, redondance des structures et présence de résidus Cys/Pro, aboutissent conjointement à l'élimination de près de 73% des solutions issues de la recherche précédente. Les résultats sont indiqués dans le tableau V-13.

	F6→G(Cα)	S23→G(Cα)	K25→G(Cα)	Y26→G(Cα)	FSKY	Solutions
Initial	97	361	60	113	256	887
CYS/PRO	57	222	45	63	112	499 (≈56%)
Redondance	46	164	33	69	98	410 (≈46%)
Final	24	107	22	39	48	240 (≈27%)

Tableau V-13 : Résultats obtenus après utilisation du programme MAKEINP, consécutivement aux recherches effectuées par STAMP. Le % de solutions conservées est indiquée en cyan

V.2.2.C Critère de taille (taille minimum autorisée)

La recherche topologique permet de limiter la taille maximum des plateformes recherchées, or nous avons choisit d'exclure les plateformes comprenant moins de 20 résidus, en raison de leur importante flexibilité. De ce fait, la topologie du motif, après transfert sur la plateforme identifiée, risque de ne pas être maintenue. Ainsi, par ce critère, approximativement 5% des solutions sont éliminées.

V.2.2.D Accessibilité du motif

Le motif ne doit être localisé à la surface de la protéine pour permettre l'interaction avec la cible. Pour cette raison, nous avons fixé à 5Å^2 , le seuil de surface accessible, par résidu. Toute solution présentant, au moins, un résidu dont l'ASA est inférieure à ce seuil est exclue des solutions. Ce critère a conduit à l'élimination de 30 à 35% des solutions. Ceci s'explique principalement par le fait que, les résidus du motif identifié pouvant être de type différent de ceux du motif de référence, leur substitution peut aboutir à l'enfouissement total ou partiel de la chaîne latérale considérée. A titre d'indication, la surface accessible au solvant des résidus du motif de référence (F6/S23/K25/Y26) est comprise entre 20 et 50Å^2 .

V.2.2.E Orientation des chaînes latérales des résidus du motif

La méthode proposée repose sur la conservation de la topologie du motif fonctionnel d'un ligand de référence. Or, lorsque les critères de recherche sont augmentés, il se peut que la topologie ne soit plus respectée. Pour cette raison, nous avons déterminé l'orientation relative des vecteurs (Ca-C β) des motifs solutions par rapport aux vecteurs équivalents du motif de référence, après superposition des motifs considérés. Cet angle est fixé à 90° , l'angle de déviation maximum autorisé par résidu. Par ce critère, près de 75% des solutions sont exclues.

Remarque

Les trois critères précédents aboutissent conjointement à l'élimination de près de 80% des solutions pré-filtrées. Les résultats sont indiqués dans le tableau V-14.

	F6→G(C α)	S23→G(C α)	K25→G(C α)	Y26→G(C α)	FSKY	Solutions
Initial	24	107	22	39	48	240
Taille < 20	23	103	19	34	47	226 ($\approx 94\%$)
ASA > 5Å^2	16	72	17	20	25	150 ($\approx 63\%$)
Angle < 90°	8	16	11	5	18	58 ($\approx 24\%$)
Final	6	13	8	3	18	48 ($\approx 20\%$)

Tableau V-14 : Tri des solutions issues des cinq recherches effectuées par STAMP, par les programmes MAKEINP et MAKESEL. Le % de solutions conservées est indiquée dans la dernière colonne en cyan.

V.2.3 Critères additionnels

Cette étape correspond à la vérification de l'intégrité des structures et leur compatibilité pour un transfert de motif :

- 1) Deux solutions appartenant à des fichiers pdb différents peuvent correspondre à la même protéine, expliquant, dans quelque cas, des redondances de solutions : celles-ci seront donc résumées à une seule.
- 2) L'insertion du motif de liaison ne doit pas modifier les propriétés de la plateforme. Ainsi ne seront conservés que les motifs ne contenant pas de résidus impliqués dans une interaction particulière (avec un ion métallique, un cofacteur...).
- 3) Les structures possédant des groupements, dénommés « HETATM » dans les fichiers contenus dans la PDB (sucres, acides aminés non-naturels), ne seront pas non plus conservées, essentiellement pour des problèmes d'absence de paramètres atomiques ou de difficultés de synthèse.

L'application de ces simples critères entraîne l'élimination de 56% des solutions issues des étapes de sélection précédentes. Les résultats sont présentés dans le tableau V-15 suivant :

	F6→G(Cα)	S23→G(Cα)	K25→G(Cα)	Y26→G(Cα)	FSKY	Solutions
Initial	6	13	8	3	18	48
Final	2	5	4	1	9	21 (≈44%)

Tableau V-15 : Tri des solutions, issues des étapes de recherche et de sélection précédentes, par divers critères additionnels.

V.2.4 Conclusion

A la suite de l'ensemble des étapes de recherche et de sélections, 21 plateformes, aptes à porter le motif de liaison F6/S23/K25/Y26, ont été identifiées. L'une de ces solutions correspond au ligand de référence, 1BGK.pdb, ce qui valide, *a posteriori*, notre approche *in silico*. Ces plateformes ne représentent que 0,06% de l'ensemble des fichiers contenus dans la PDB. D'autre part, ce nombre peut être contrôlé par les paramètres de recherche utilisés. Compte-tenu de la croissance importante du nombre de structures 3D de protéines dans la PDB, notre approche permettra, à l'avenir, d'identifier un nombre de solutions d'autant plus important. L'utilisation de critères de sélection, applicable de façon générale à tout système, permet une réduction importante du nombre de solutions par l'exclusion de structures incapables de reproduire l'interaction ciblée. Par conséquent, cette approche reste applicable malgré l'accroissement des informations structurales contenues dans la PDB (annexe A.2).

V.3 Analyse des plateformes

V.3.1 Les plateformes identifiées

La recherche précédemment décrite a permis d'identifier 21 plateformes protéiques. Trois plateformes supplémentaires (1C5A, 1CR8 et 1STU) ont été identifiées par l'augmentation du critère %R_MAX de 2 à 3%. Ces différentes plateformes sont présentées dans le tableau V-16 :

PDB	Réf.	Nom de la protéine	RMSD	N°1	N°2	N°3	N°4	Ch
1BGK	[200]	<i>Bgk (ligand de référence)</i>	0,124	6	23	25	26	*
1BNB	[254]	<i>Bovine Neutrophil Beta-Defensin</i>	0,831	32	1	3	4	*
1CKK	[255]	<i>Calmodulin Dependent Protein Kinase</i>	0,640	26	6	8	9	B
1D6B	[256]	<i>Defensin-Like Peptide 2</i>	0,621	20	34	36	37	A
1EDL	[257]	<i>Staphylococcal Protein AE Domain</i>	0,693	39	4	6	7	*
1H5O	[258]	<i>Crotamine</i>	0,630	15	31	33	34	A
1UXD	[259]	<i>Fructose Repressor DNA-Binding Domain</i>	0,827	28	13	15	16	*
1J75	[260]	<i>DLM1</i>	0,349	132	162	164	165	A
1OQE	[261]	<i>S TALL-1</i>	0,643	21	5	7	8	O
1BOE	[262]	<i>IGF Binding Domain of IGFBP-5</i>	0,565	57	28	30	31	A
1HZ6	[263]	<i>B1 Domain of Protein L</i>	0,739	26	53	55	56	C
1JE9	[264]	<i>NT2</i>	0,648	30	46	48	49	A
1A7F	[265]	<i>Insulin Mutant</i>	0,918	8	1	4	5	A
2PDD	[266]	<i>Domain of Dihydrolipoamide Acetyltransferase</i>	0,767	1	25	27	29	*
2SH1	[267]	<i>Neurotoxin 1 from Sea Anemone</i>	0,770	20	29	46	30	*
1ERF	[268]	<i>Nter fusion peptide of HIV-1 GP41</i>	0,701	22	2	4	5	A
1F7E	[269]	<i>EGF-Like Domain</i>	0,638	87	73	75	76	A
1M3A	[270]	<i>Nter SH3 Domain c-Crk</i>	0,530	169	147	149	150	A
1EGF	[271]	<i>Murine Epidermal Growth Factor</i>	0,667	43	10	12	13	*
1EAY	[272]	<i>CHEY-Binding Domain of CHEA</i>	0,988	215	170	174	171	C
1FGP	[273]	<i>Protein G3P</i>	0,781	21	47	49	50	*
1C5A	[274]	<i>Porcine complement protein C5adesArg</i>	0,786	43	15	17	18	*
1CR8	[275]	<i>LDL receptor-related protein complement repeat 8</i>	0,658	15	39	36	35	*
1STU	[276]	<i>dsRNA binding domain of staufen protein</i>	0,700	60	31	47	46	*

Tableau V-16 : Liste des plateformes protéiques identifiées au cours de ce travail de thèse, à l'occasion de plusieurs recherches. Pour chacune d'entre elles, sont indiqués : le code PDB, le nom de la protéine, le RMSD calculé via la superposition du motif solution sur le motif de référence, les résidus composant le motif ainsi que l'identifiant de la structure 3D. Les marques * indiquent qu'il s'agit d'un conformère obtenu à partir d'un fichier RMN.

L'examen de ces solutions ne montre aucune relation entre les fonctions ces diverses plateformes : toxines et/ou neurotoxines, cytokines, domaines de liaison à l'ADN ou l'ARN,

domaines enzymatiques, protéines du métabolisme, protéines du cycle cellulaire, protéines virales et/ou bactériennes. De ce fait, le développement de ligands protéiques ne serait plus limité à une ou plusieurs familles de protéines, choisies pour leurs caractéristiques fonctionnelles et/ou structurales. Toute protéine capable de reproduire l'interaction, après transfert d'un motif fonctionnelle, peut être identifiée par cette approche. Ainsi, le développement de ligands ne serait plus limité par le nombre molécules guides disponibles.

V.3.2 Diversité structurale des plateformes

V.3.2.A Structures tridimensionnelles des plateformes

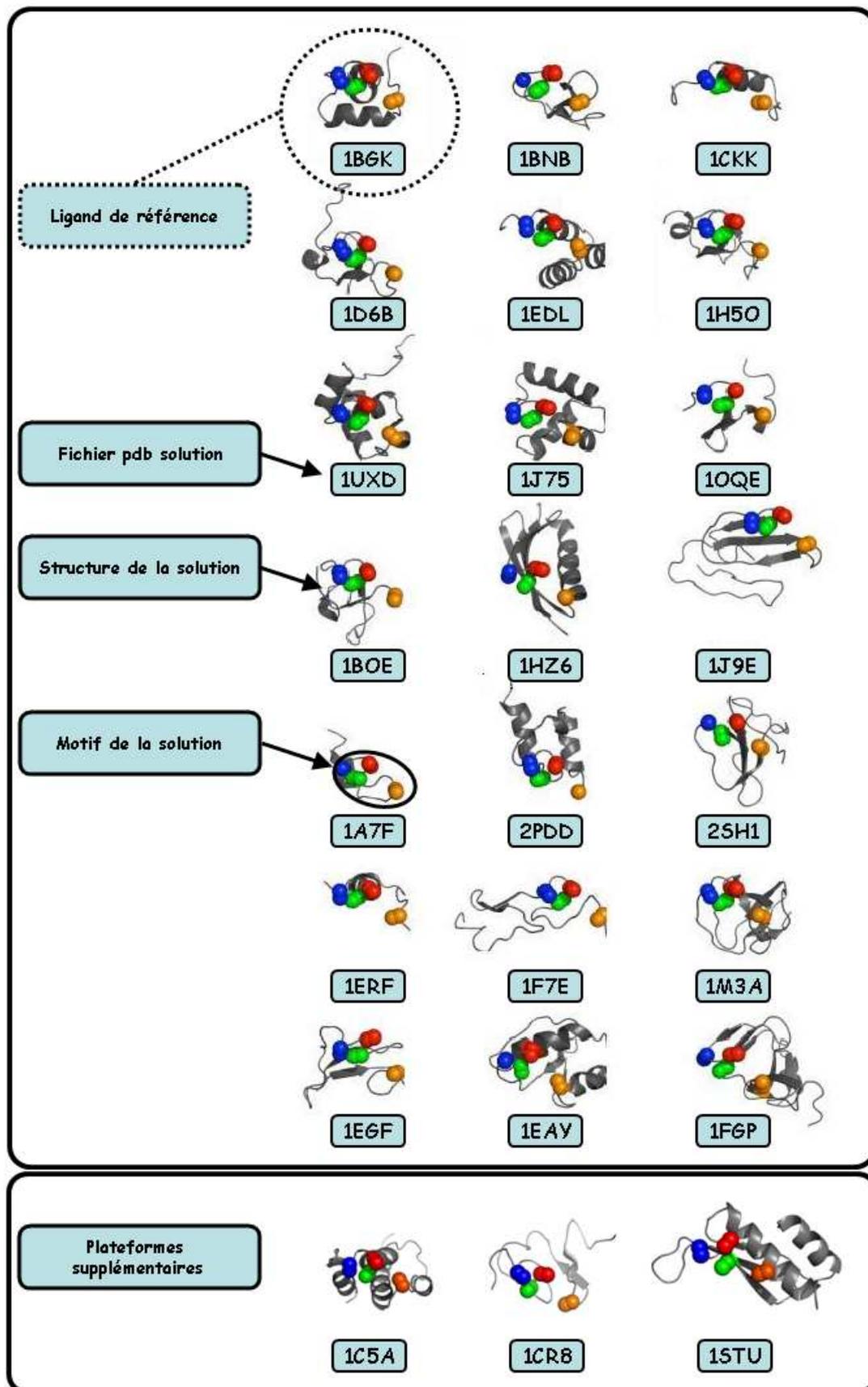
Les plateformes identifiées montrent des repliements divers, regroupés, de façon simplifiée, en trois catégories dans le tableau V-17 suivant :

Repliement	Hélice α	Feuillet β	Mixte (hélice+feuillet)
Code PDB	1BGK, 1CKK, 1EGL, 1UXD, 1A7F, 2PDD, 1ERF, 1C5A	1BNB, 1OQE, 1BOE, 1J9E, 2SH1, 1F7E, 1M3A, 1ERF, 1CR8	1D6B, 1H5O, 1J75, 1HZ6, 1EAY, 1FGP, 1STU

Tableau V-17 : Classification des plateformes identifiées selon leur repliement. Trois catégories sont décrites : structures possédant seulement des hélices α , seulement des feuillets β , ou présentant des structures mixtes (hélice+feuillet).

La plupart des plateformes (2/3) possède des éléments de structure secondaire différents de ceux présents de la BgK, comprenant uniquement des éléments en hélice. D'autre part, les 23 plateformes présentent un repliement bien défini (pas de structures désordonnées), et 14 d'entre-elles possèdent de 1 à 4 ponts disulfures. Ces éléments faciliteront, lors de la synthèse des différentes plateformes, l'obtention de structures correctement repliées.

Les représentations schématiques des structures 3D de ces 23 plateformes sont montrées sur les figures V-17bis et ter. Ces figures démontrent que, malgré une grande diversité de repliements, la topologie des motifs Ca-C β identifiés au sein des différentes plateformes est très similaire à celle du motif de référence F6/S23/K25/Y26 de la BgK. Les valeurs de RMSD, calculées *via* la superposition des atomes Ca et C β du motif sur le motif de référence, sont reportées dans le tableau V-12 pour chaque plateforme. L'approche de recherche systématique de plateformes protéiques à partir d'un motif de liaison donné apparaît effectivement, de part ces observations, indépendante de la structure 3D des plateformes supportant les motifs équivalents à celui-ci.



Figures V-17bis (en haut) et ter (en bas) : Liste des 24 solutions issues des recherches *in silico*. Les structures des solutions sont représentées en mode « cartoon », les motifs en mode « sphere » : la position 1 en orange (F6), la position 2 en bleu (S23), la position 3 en rouge (K25), la position 4 en vert (Y26).

V.3.2.B Topologie du motif et structure secondaire

D'après la structure 3D de la BgK, les résidus S23/K25/Y26 du motif sont localisés au niveau N terminal de la seconde hélice α . Les résidus équivalents, dans certaines plateformes, sont également localisés dans la partie N terminale d'hélices α (1CKK, 1EDL, 1UXD, 1A7F, 1ERF, 1EAY et 1C5A). Par ailleurs, dans la majorité des cas, les résidus équivalents sont localisés au niveau d'éléments structuraux continus différents, boucles, coudes (1BNB, 1D6B, 1H5O, 1J75, 1OQE, 1BOE, 1HZ6, 1J9E, 2PDD, 1F7E, 1FGP et 1CR8). Néanmoins, dans deux cas (1SH1 et 1STU), le motif équivalent est porté par des éléments discontinus en séquence. Concernant la plateforme 1SH1, les résidus équivalents à S23 et Y26 sont localisés en position N terminale du deuxième brin β tandis que le résidu équivalent à K25 est présent au niveau C terminal du troisième brin β . Par ailleurs, le résidu équivalent à F6 est situé en position N terminale du premier brin β . Concernant la plateforme 1STU, les résidus équivalents à K25 et Y26 sont en position N terminale du deuxième brin β , le résidu équivalent à S23 étant en position C terminale du troisième brin. Le résidu équivalent à F6 est situé au milieu de la seconde hélice α . Ces deux motifs, ainsi que le motif de référence, sont schématisés sur la figure V-18.

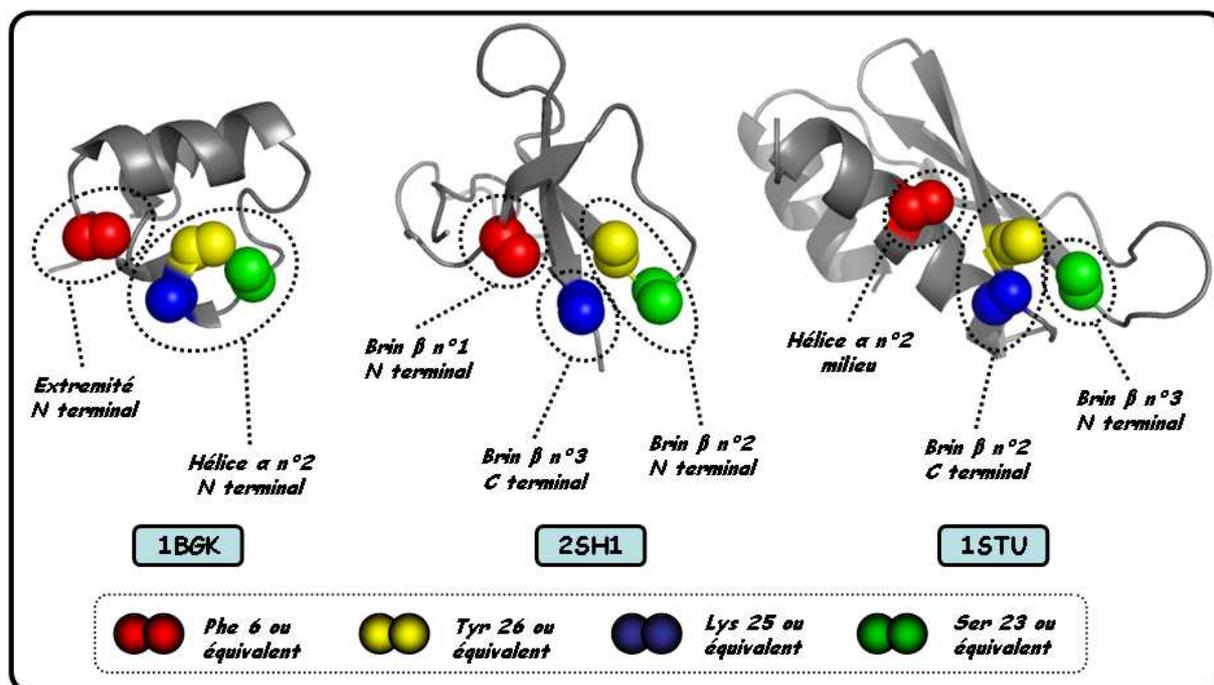


Figure V-18 : Représentation des motifs de liaison FSKY de la BgK et des plateformes 2SH1 et 1STU. Les structures (en gris) sont représentées en mode « cartoon ». Les résidus du motif, schématisés par les atomes Ca et C β , sont représentés en mode « sphere ».

Il apparaît donc, malgré une distribution éclatée, dans la séquence et dans l'espace, des résidus appartenant au motif, une importante similarité de topologie entre les atomes Ca et C β du motif F6/S23/K25/Y26 et ceux des motifs équivalents des plateformes 1SH1 et 1STU. Ces deux exemples démontrent la capacité de l'approche basée sur la recherche systématique des

structures 3D de la PDB, à identifier diverses plateformes protéiques, sans contrainte quant à la structure secondaire portant le motif topologique.

V.3.3 Analyse du potentiel électrostatique

L'analyse de complexes cristallographiques ainsi que de nombreux travaux déjà publiés attestent de l'importance du potentiel électrostatique pour la reconnaissance des protéines entre elles. Il conviendrait néanmoins, pour documenter l'utilisation du potentiel électrostatique comme critère de sélection, de synthétiser et tester l'ensemble des 23 plateformes obtenues.

Le calcul du potentiel électrostatique a été réalisé suivant la procédure précédemment décrite (annexe A.4), et la comparaison des grilles de potentiel effectuée par rapport à celle de la BgK. L'indice de similarité du potentiel électrostatique de la solution 1BGK par rapport à la structure de la BgK de référence s'avère très élevé (0,97). La très faible différence observée s'explique par les légères variations de structures 3D des deux protéines : la structure de la BgK de référence correspond à une structure moyenne minimisée de l'ensemble des conformations RMN (annexe B.1), tandis que la solution identifiée ne correspond qu'à une seule conformation. Cette différence est néanmoins cohérente avec les fluctuations observées lors du calcul des potentiels électrostatiques pour différentes structures d'une même protéine. Les résultats de la comparaison des grilles de potentiel, obtenus pour chacune des plateformes, sont reportés dans le tableau V-19.

Code PDB	SI (« all »)	SI (« parts »)	Code PDB	SI (« all »)	SI (« parts »)
1BGK	0.917	0.970	1J75	0.681	0.720
1H5O	0.820	0.942	1D6B	0.560	0.656
1JE9	0.816	0.941	1EDL	-0.142	0.579
1C5A	0.851	0.932	1STU	0,327	0,446
1UXD	0.857	0.932	25H1	0.258	0.395
1BNB	0.812	0.930	1HZ6	0.146	0.306
1CKK	0.749	0.894	1CR8	-0,129	0,246
1OQE	0.725	0.887	1EGF	-0.013	0.048
1A7F	0.589	0.812	1F7E	-0.374	0.022
1ERF	0.661	0.811	1FGP	-0.436	-0.146
2PDD	0.642	0.776	1M3A	-0.091	-0.272
1BOE	0.376	0.732	1EAY	-0.762	-0.814

Tableau V-19 : Liste des 24 plateformes classées par ordre décroissant de leurs valeurs de SI à l'interface. Les valeurs de SI ont été calculées par comparaison des grilles de potentiel électrostatique de chaque plateforme (après insertion du motif de liaison) par rapport à la grille de potentiel de la BgK. Ces valeurs ont été calculées sur l'ensemble des grilles (« all ») ou sur la zone interface seulement (« parts »).

Parmi ces plateformes, plusieurs ont été synthétisées (séquences présentées dans l'annexe C.1) et seront présentées dans le chapitre suivant. Ces dernières ont été choisies de manière à représenter une hétérogénéité de potentiel électrostatique : les plateformes 1H5O, 1C5A, 1J75 présentant une forte similitude (SI > 0,7) de potentiel comparativement au potentiel de la BgK,

les plateformes 1D6B et 1STU une similitude moyenne ($SI < 0,7$) et les plateformes 1HZ6 et 1CR8, présentant une similitude faible ($SI < 0,4$). Ces 8 plateformes sont entourées de noir dans le tableau V-19. Pour chacune, est indiqué le code PDB de la plateforme identifiée, ainsi que les valeurs de SI calculées après insertion du motif sur la plateforme sur l'ensemble des grilles (« all ») ou sur la région d'interface seule (« parts »).

V.3.4 Conclusion

Au final, l'ensemble de la procédure, de la recherche de motifs topologiques par STAMP, jusqu'à l'application de la totalité des cribles de sélection, y compris l'application du crible électrostatique (avec une valeur seuil de SI 0,4, déterminée chapitre IV), aboutit à la sélection de 13 plateformes. Ces structures sont potentiellement capables de se lier à la cible de référence, à savoir le canal Kv1.2, après introduction du motif de liaison.

	<i>F6→G(Ca)</i>	<i>S23→G(Ca)</i>	<i>K25→G(Ca)</i>	<i>Y26→G(Ca)</i>	FSKY	Total
PDB	n.a.	n.a.	n.a.	n.a.	n.a.	23845
Recherche	1302	2717	456	1080	2735	8291 (35%)
Sélection	2	2	2	1	6	13 (0,05%)

Tableau V-20 : Résultat obtenus pour une recherche, avec un jeu de paramètres donné (RMSD_MAX : 1,0Å / Δd_MAX : 1,3Å / taille_MAX : 70 résidus / %R_MAX : 2%). L'évolution du nombre de structures solution est indiqué, à la suite de chaque étape : recherche puis sélection.

En définitive, l'ensemble de la procédure permet d'identifier 13 structures, correspondant à l'ensemble des critères, caractéristiques de l'interaction (complémentarité de forme, accessibilité du motif, topologie du motif) ou nécessaires à l'obtention des plateformes (taille, présence de résidus défavorables). Ce nombre ne représente que 0,05% de l'ensemble de la PDB : d'une part, il s'avère compatible pour une validation expérimentale, et d'autre part, il permet d'envisager une utilisation de l'approche, malgré l'augmentation du nombre de structures protéiques dans la PDB.

Chapitre VI : Validation expérimentale

« Un fait n'est rien par lui même, il ne vaut que par l'idée qui s'y rattache ou par la preuve qu'il fournit » (C. Bernard)

L'application de l'approche *in-silico*, décrite dans le chapitre précédent, a permis d'identifier 23 plateformes protéiques aptes à porter le motif de liaison au canal Kv1.2 de la BgK. La production de ces plateformes a été entreprise afin de valider notre méthode de conception de ligands.

Pour chaque plateforme, après transfert du motif de liaison F6/S23/K25/Y26, l'identification de ligands a été effectuée par la réalisation de tests de liaison par compétition vis-à-vis d'un analogue radio-marqué de la BgK. L'impact du motif de liaison a été estimé par comparaison de l'affinité de la plateforme modifiée (ou « mt », avec motif de liaison) avec celle de la plateforme native (ou « wt », sans motif de liaison). En outre, de façon à caractériser plus précisément le mode d'interaction des ligands protéiques les plus affins, diverses modifications de séquence ont été effectuées.

Enfin, l'interaction des différentes plateformes, selon le mode supposé (par l'intermédiaire du motif de liaison), a été caractérisée, *in silico*, par la réalisation de modèles moléculaires de complexes entre ces plateformes et le canal Kv1.2. Ces derniers ont été réalisés à partir du modèle moléculaire du complexe BgK-Kv1.1, décrit précédemment. La connaissance de ces modèles nous a alors permis d'analyser les conséquences globales ou microscopiques, consécutives à la fixation des plateformes, au même site de liaison que le ligand de référence (BgK).

Le chapitre VI présentera les résultats expérimentaux (synthèse par biologie moléculaire et par synthèse peptidique, caractérisation biochimique et structurale) concernant 8 plateformes (annexe C.1) parmi les 23 identifiées. Puis, les résultats des tests de liaison des ligands conçus par notre approche *in silico*, seront présentés. Enfin, la caractérisation des modèles moléculaires sera détaillée

VI.1 Production par biologie moléculaire de la plateforme 1J75

La première plateforme synthétisée, le domaine de liaison à l'ADN Z de la protéine DLM1, a été obtenue par des techniques de biologie moléculaire (annexe C.2), en suivant le protocole expérimental décrit dans la littérature [260].

VI.1.1 Plateforme 1J75

Différents variants de la plateforme 1J75 ont été réalisés à partir de la protéine native, illustrée par la figure VI-1.

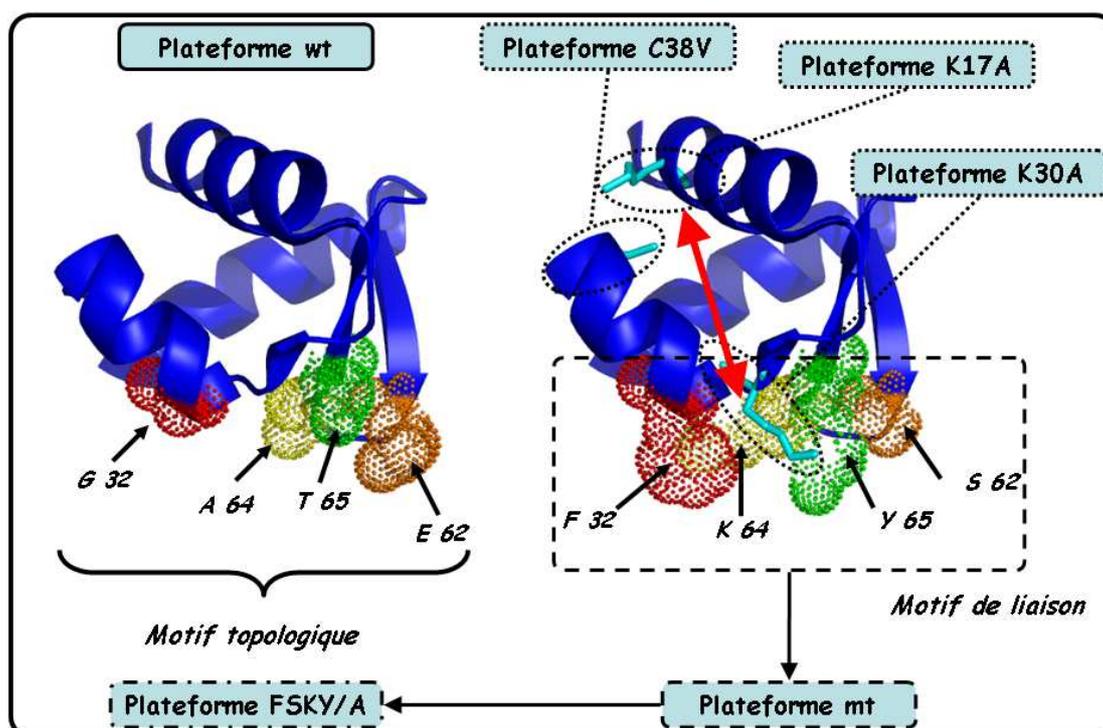


Figure VI-1: Les différents variants de la plateforme 1J75 (en bleu). La structure 3D est représentée en mode « cartoon », les résidus du motif sont représentés en mode « dot ». La position n°1 (F6) du motif en rouge, la position n°2 (S23) en orange, la position n°3 (K25) en jaune et la position n°4 (Y26) en jaune. Les résidus modifiés n'appartenant pas au motif (en cyan) sont représentés en mode « stick ».

Le motif de liaison a été inséré sur la plateforme 1J75 wt, par 4 mutations G32F, E62S, A64K et T65Y, correspondant à la plateforme désignée 1J75 mt par la suite. Plusieurs autres variants protéines de la protéine 1J75mt ont été effectués afin d'évaluer le poids des résidus du motif transféré sur la liaison aux canaux ciblés. Les plateformes 1J75mtF32A et 1J75mtK64A correspondent au remplacement en alanine des positions équivalentes aux résidus F6 et K25 du motif de référence. L'effet de la substitution du résidu K64 a été comparé aux substitutions d'autres résidus lysine en alanine : un résidu spatialement proche du motif de liaison (plateforme 1J75mtK30A), et un second plus éloigné (plateforme 1J75mtK17A). Ces deux résidus sont distants en séquence du résidu K64, équivalent au résidu K25 de la BgK. Enfin, nous avons produit

un variant de 1J75mt pour lequel les 4 résidus du motif ont été substitués en résidus alanine, correspondant à la plateforme 1J75mtFSKY/A. Par ailleurs, pour éviter la formation de dimères liés à la présence d'un résidu cystéine dans la protéine, le résidu C38 a été substitué par un résidu Val (plateforme 1J75mtC38V). Les différentes séquences sont fournies dans l'annexe C.1. La structure de la protéine 1J75mt a été superposée à la structure de la BgK, *via* leurs atomes Ca et C β des résidus du motif, représenté sur la figure VI-2.

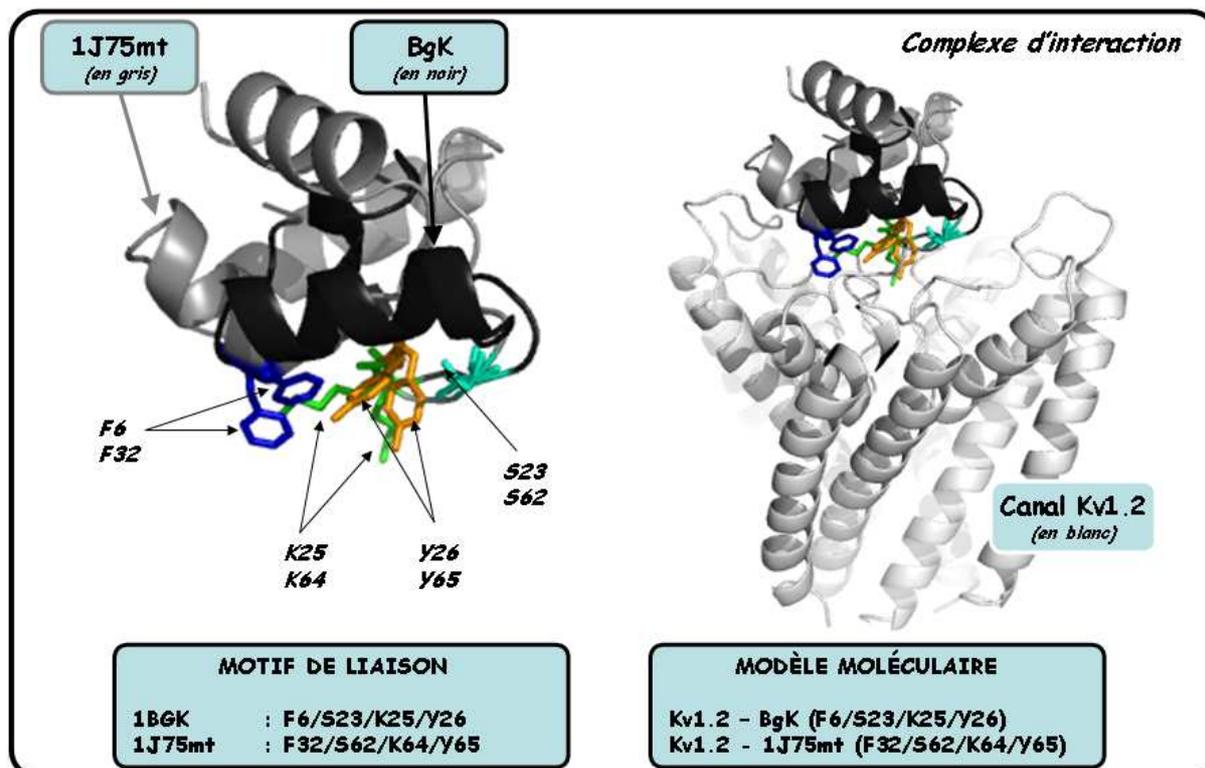


Figure VI-2 : Représentation de la plateforme 1J75mt, superposée à la BgK, par les atomes Ca et C β des résidus du motif de liaison. Les protéines BgK (en noir) et 1J75mt (en gris) sont montrées, en interaction avec le canal Kv1.2 (en blanc), dans le même mode d'interaction par la réalisation de modèles d'interaction. Les structures sont représentées en mode « cartoon », les résidus du motif en mode « stick ». Les résidus du motif sont colorés selon leur type : bleu pour Phe, cyan pour Ser, vert pour Lys et Orange pour Tyr.

Ces deux protéines sont, en outre, placées en interaction avec le canal Kv1.1. D'après l'observation de cette figure, l'orientation des chaînes latérales des résidus du motif sont similaires entre la BgK et la plateforme 1J75mt. Par ailleurs, l'interaction avec le canal Kv1.1 montre un mode d'interaction similaire de ces deux protéines, sans contraintes stériques, validant ainsi les étapes de recherche et de sélection *in silico* précédentes.

VI.1.2 Rendement de production

Le protocole expérimental pour la production des variants de la plateforme 1J75 est décrit en annexe (C.2). Les quantités obtenues diffèrent sensiblement suivant le variant considéré de la protéine 1J75. Les plateformes 1J75mt et 1J75mtC38V donnent les rendements les plus faibles, de l'ordre de 300-500 μ g par production, tandis que la plateforme 1J75mtK30A est produite avec un rendement moyen, par production, de l'ordre de 5mg, soit un rendement 10 à 15 fois supérieur à celui observé pour les plateformes 1J75mt et 1J75mtC38V. Les quantités obtenues ont donc permis de réaliser les tests de liaison pour chaque variant (quantité nécessaire de l'ordre de 1mg). En revanche, la détermination de la structure 3D de ces protéines par radiocristallographie requiert une quantité plus importante, de l'ordre de 2 à 5 mg. Cette étude a donc été mise en œuvre sur la plateforme 1J75mtK30A dont le rendement de production était significativement plus important que celui observé pour 1J75mt.

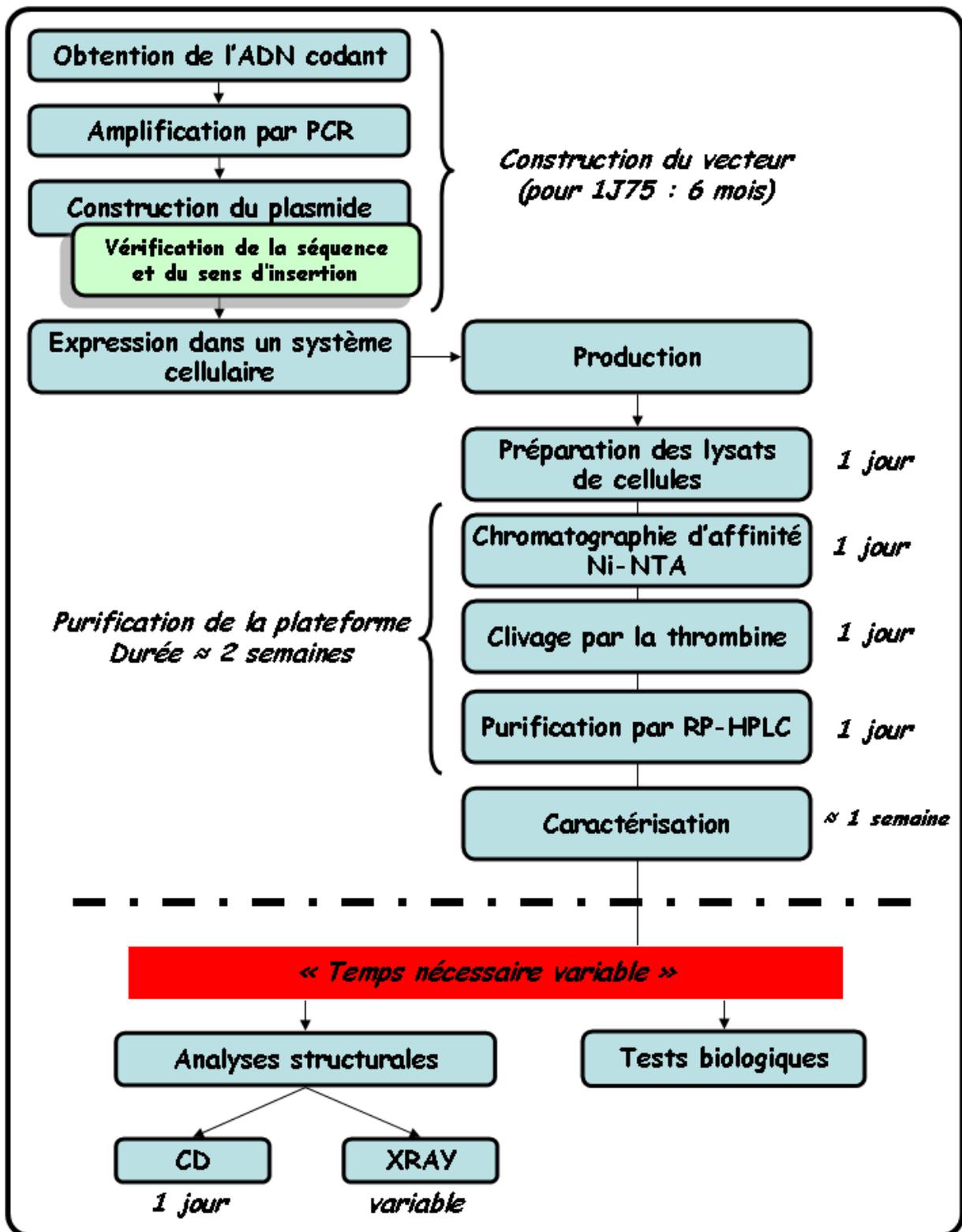


Figure VI-3 : Représentation schématique de l'ensemble des étapes nécessaires pour l'obtention de plateformes par biologie moléculaire.

VI.1.3 Conclusion

Le choix de la biologie moléculaire pour la production de la plateforme 1J75 et de ces variants a été justifié par le fait que l'utilisation des conditions décrites dans la littérature pouvait faciliter l'obtention de ces plateformes protéiques. Néanmoins, pour les différents variants, les conditions ont dû être adaptées et le temps nécessaire à la mise au point des conditions de productions s'est avéré relativement important, notamment pour les différentes étapes de construction du vecteur de la plateforme 1J75 (le processus dans son ensemble est schématisé sur la figure VI-3 et décrit dans l'annexe C.2). L'insertion du motif, par 4 modifications ponctuelles de la séquence native de la protéine DLM1, a donc eu des répercussions significatives, en termes de temps de production. Ces observations nous ont donc incité à produire les autres plateformes par synthèse peptidique. En effet, cette approche permet, lorsque les rendements de synthèse sont favorables, de produire des quantités de l'ordre du mg (permettant la réalisation de tests de liaison) dans un temps relativement court pour des protéines de taille inférieur à 70 résidus.

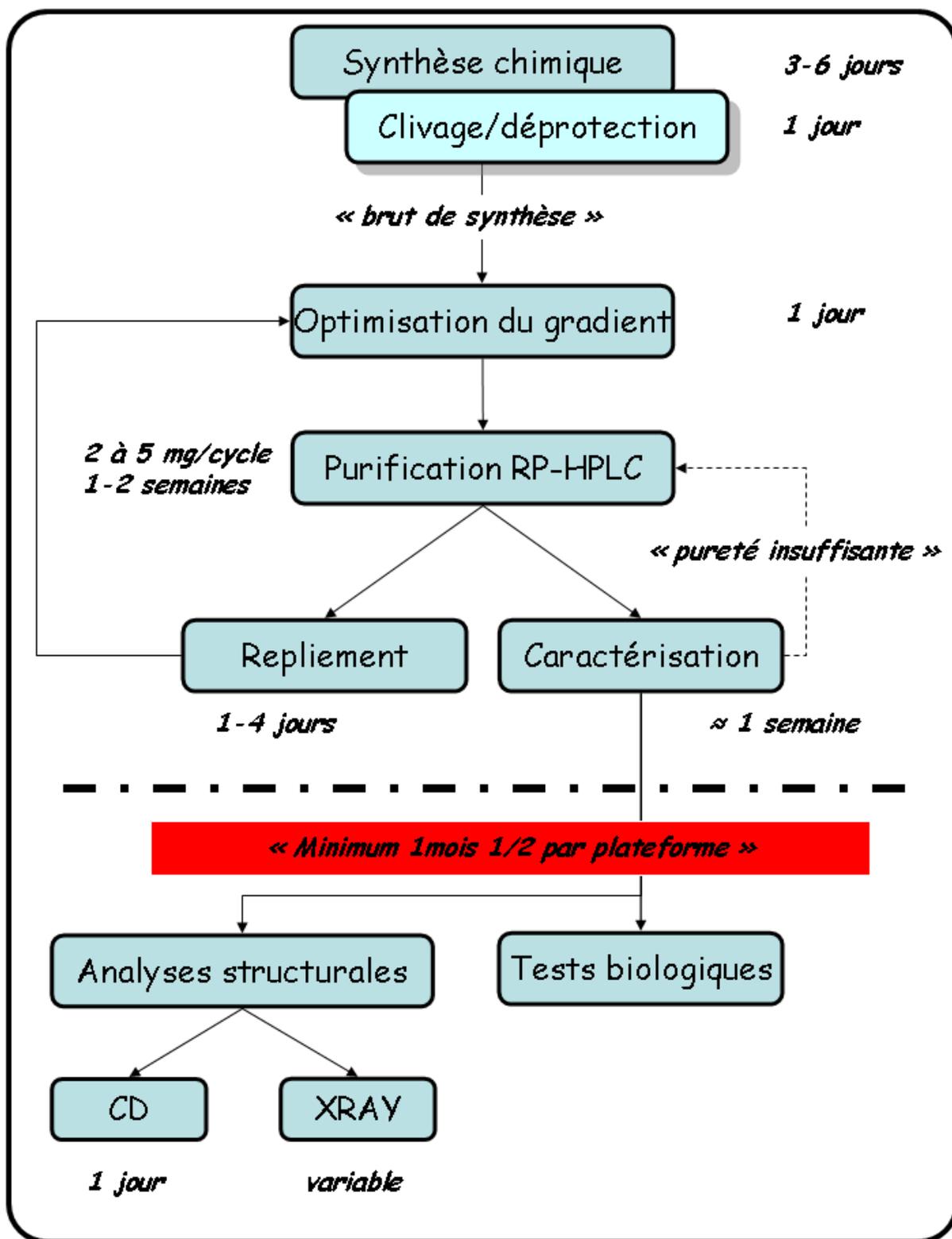


Figure VI-4 : Détail du protocole de production de plateformes par voie chimique, de la synthèse à la caractérisation

VI.2 Production des plateformes par voie chimique

La synthèse chimique (annexe C.3) a été choisie pour la production rapide de plateformes diverses. Elle présente, en effet, plusieurs avantages : un temps de synthèse relativement court dans le cas de mini-protéines de taille inférieure à 70 résidus et la possibilité d'incorporer des résidus non-naturels, lors d'étapes d'optimisation ultérieures. La purification a été effectuée en une seule étape par RP-HPLC permettant de séparer des protéines possédant des propriétés physico-chimiques très variables (taille, charge ou hydrophobicité). Le protocole utilisé est représenté schématiquement sur la figure VI-4. Ainsi, le temps nécessaire à la production d'une plateforme a été estimé à 1 mois et demi, permettant d'envisager la synthèse de plusieurs plateformes et de leurs variants dans le cadre de ce travail de thèse.

VI.2.1 Synthèse chimique sur support solide

La synthèse chimique standard (annexe C3) permet d'obtenir des protéines de tailles comprises entre 50 et 100 résidus avec, dans de nombreux cas, des rendements de synthèse élevés. Cette méthode est donc particulièrement adaptée à la production de mini-protéines. En outre, le rendement de chaque synthèse est dépendant de la séquence primaire et de la composition de la protéine ciblée. De ce fait, peu prévisible, une phase d'optimisation est parfois nécessaire, afin d'obtenir des quantités suffisantes de protéines. Notre objectif étant de tester un maximum de plateformes, les conditions de synthèse ont été optimisées de façon à produire, en une seule synthèse la quantité de protéine nécessaire pour la réalisation de tests de compétition, de l'ordre de 1mg de protéine purifiée (cette quantité permettant d'effectuer des tests de compétition, allant jusqu'à 100 μ M). Au cours de ce travail, plusieurs paramètres ont été étudiés pour l'optimisation des rendements, en particulier le chargement des résines (ou « loading »), le temps de couplage et l'utilisation d'agent chaotropique (LiCl), limitant les phénomènes de collapse de la résine lors de l'élongation des protéines. Les rendements de certaines synthèses ont ainsi été améliorés sensiblement.

VI.2.1-A Rendements de synthèse

En définitive, 14 protéines ont été synthétisées. Leurs séquences sont présentées dans l'annexe C.1 ; pour chaque synthèse, la résine employée et son chargement sont présentés dans l'annexe C.3. Les rendements de synthèse, avant purification sont variables, de 10 à 68%, comme indiqué dans le tableau VI-5 :

Plateforme	Taille	Rendement	Référence
1CKK wt, mt	26	50%	[255]
1HZ6 wt, mt	63	30%	[263]
1STU wt, mt	68	10%	[276]
1C5A mt	66	35%	[274]
1CR8 mt	42	40%	[275]
1D6B mt	42	55%	[256]
1H5O wt, mt, mt2, mt3, mt4	42 (40 pour mt2)	68, 33, 55, 55 et 45%	[258]

Tableau VI-5 : Rendement de synthèse des plateformes.

Pendant le temps imparti pour cette thèse, seules 7 plateformes parmi les 23 plateformes identifiées *in silico* ont pu être synthétisées. La synthèse des 16 plateformes restantes, nécessitant plus de temps, sera effectuée ultérieurement au laboratoire.

VI.2.1-B Importance de la taille des plateformes

Les différentes synthèses ont été réalisées en suivant la stratégie Fmoc et le même protocole expérimental. Leurs rendements peuvent, de ce fait, être comparés. L'analyse du rendement de synthèse des plateformes en fonction de la taille de celles-ci est représenté sur la figure VI-6.

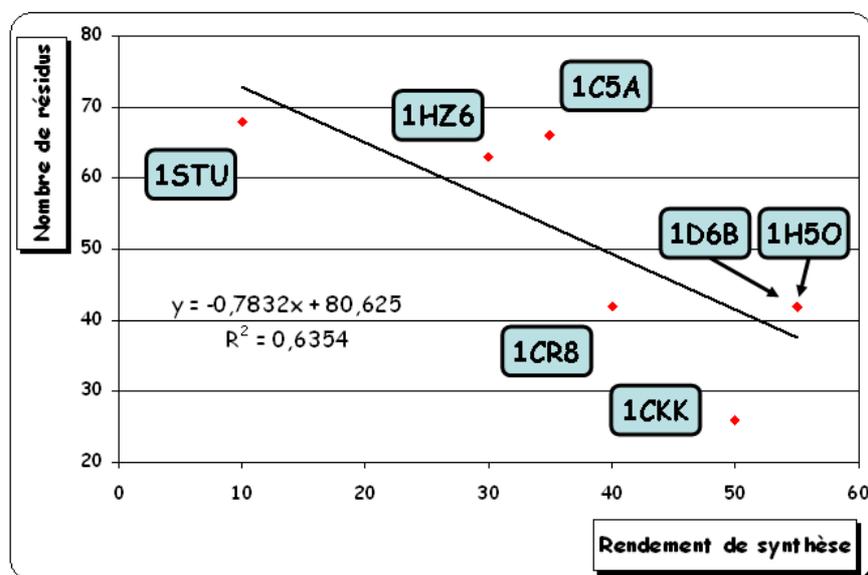


Figure VI-6 : Rendement de synthèse par voie chimique en fonction de la taille des plateformes.

Cette dernière montre qu'il existe une corrélation entre la taille des protéines synthétisées et le rendement final de synthèse. Le rendement par couplage avoisine 99%, soit un rendement théorique global de 50% pour une protéine de 70 résidus. Or, les résultats obtenus, au cours de ce travail montrent qu'au-delà de 70 résidus, les rendements sont très inférieurs aux rendements théoriques (de l'ordre de 10% seulement). La cause de ces observations peut être diverse : présence de zone hydrophobe dans la séquence protéique, gêne stérique du peptide en élongation, condition de synthèse...chaque synthèse est originale. Par ailleurs, les rendements de

purification sont variables suivant les protéines. Dans le cas des protéines à pont disulfures, le rendement de l'étape de repliement peut également avoir un impact important sur les quantités finales produites, du fait de mésappariements possibles (fonctions du nombre de résidus cystéine dans la séquence). Ces différents facteurs, entraînent une importante variabilité des quantités finales produites pour les différentes plateformes, nous incitant à limiter les recherches, *in silico*, de plateformes aux protéines de taille inférieure à 70 résidus.

VI.2.1-C La plateforme 1H50

Plusieurs variants de la plateforme 1H50 ont été produits afin de caractériser le mode de liaison de la protéine à la cible. Les différents variants de la plateforme 1H50 sont schématisés sur la figure VI-7. Le motif de liaison a été inséré sur la plateforme native 1H50 wt, par la modification de 4 résidus : E15F, R31S, R33K et W34Y, correspondant à la plateforme appelée 1H50 mt par la suite. Par ailleurs, les deux premiers résidus de la séquence de 1H50 wt (Tyr1 et Lys2) correspondent aux résidus de la diade, K25 et Y26, de la BgK. La suppression de ces deux premiers résidus a conduit à la plateforme 1H50 mt2. De plus, deux variants portant sur le motif R33/W34 ont été réalisés: la suppression des chaînes latérales des résidus R33 et W34 (plateforme 1H50mt3) et une mutation par inversion de charge K33E (plateforme 1H50mt4).

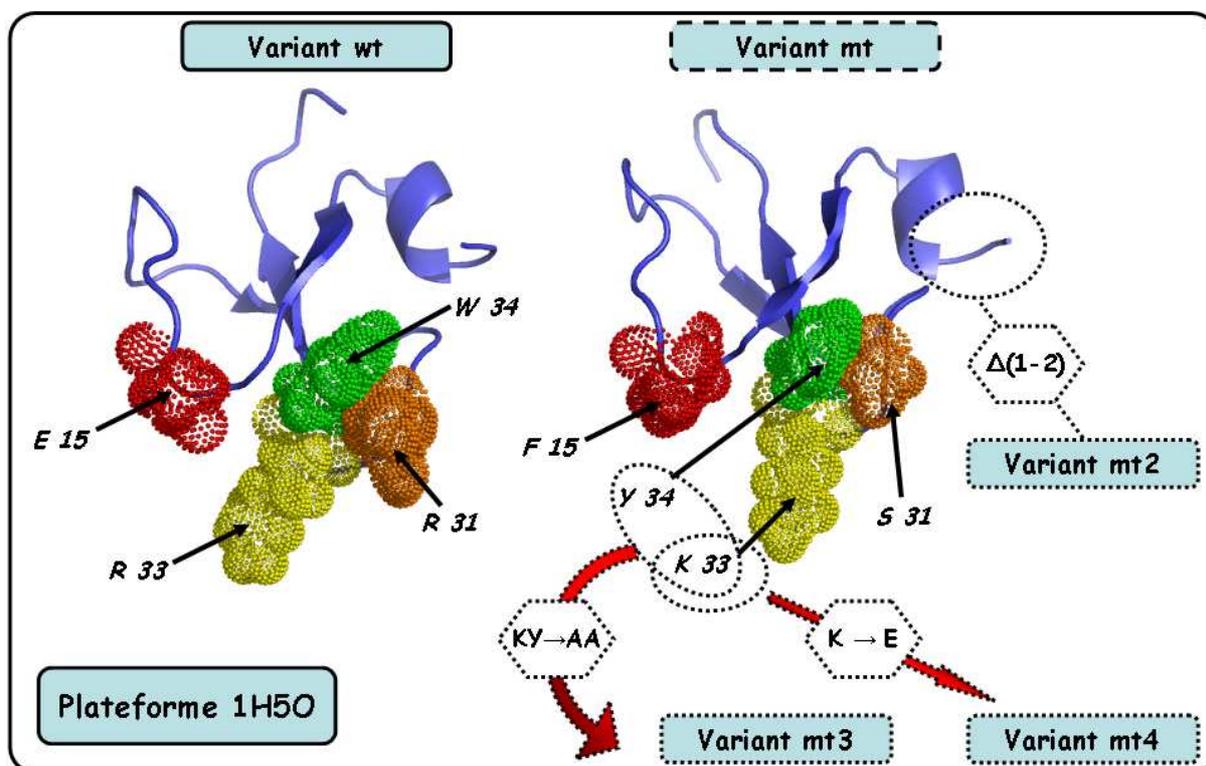


Figure VI-7 : Les différents variants de la plateforme 1H50. La structure 3D de la crotamine (en bleu) est représentée en mode « cartoon ». Les résidus du motif identifié, équivalent au motif de liaison de la BgK sont représentés en mode « dot » : E15 (équivalent à F6) en rouge, E31 (équivalent à S23) en orange, R33 (équivalent à K25), W34 (équivalent à Y26) en vert.

Les synthèses des différentes plateformes modifiées (mt, mt2, mt3, mt4) ont été effectuées en réduisant le chargement de la résine (de 0,45 à 0,25 mmol/g), augmentant ainsi considérablement les rendements (facteur 2-5). Les rendements obtenus pour les différentes versions de la plateforme 1H50 sont relativement homogènes, comme indiqué par l'histogramme VI-8.

Le rendement plus faible de la plateforme 1H50mt est la conséquence de problèmes expérimentaux (présence d'eau lors de la synthèse) et de ce fait n'est pas corrélé à la séquence de cette plateforme. De façon générale, les différences de séquences n'ont, *a priori*, pas modifié de façon importante les profils de synthèse. Cette indication ne permet toutefois pas d'estimer si les propriétés de repliement de chaque version sont identiques.

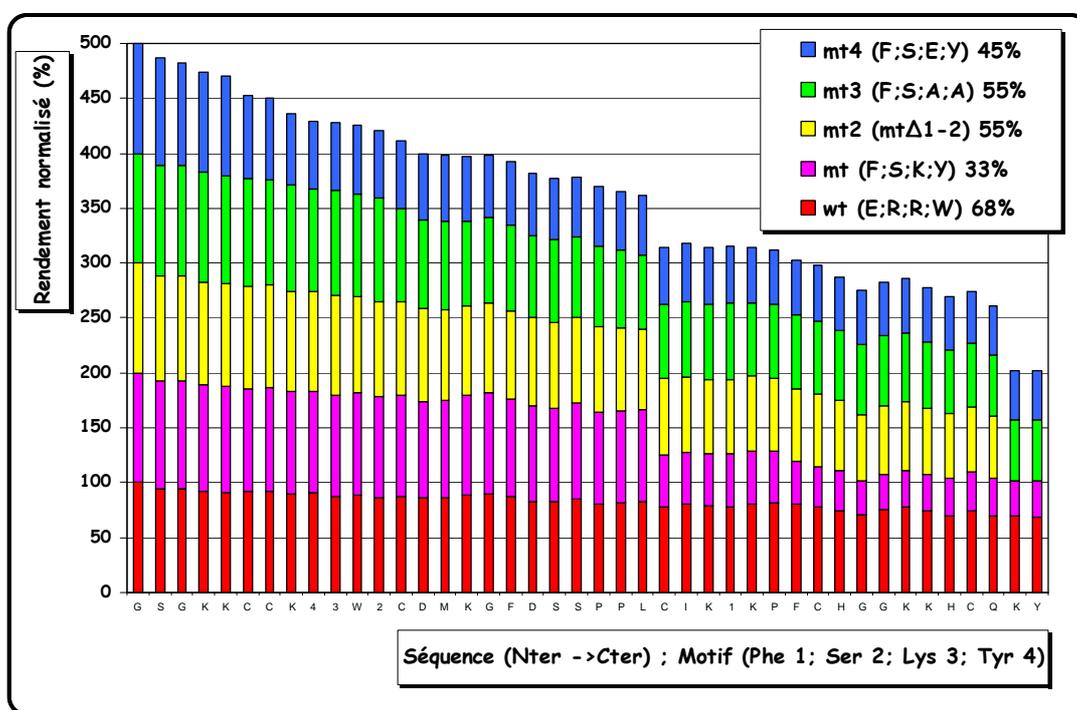


Figure VI-8 : Histogramme représentant les rendements de synthèse des différentes versions de la plateforme 1H50, résidu par résidu. Les rendements ont été normalisés à 100% en début de synthèse.

VI.2.2 Obtention des plateformes

Les conditions expérimentales ainsi que les résultats obtenus, pour l'ensemble des étapes présentées dans cette partie, sont décrites dans l'annexe C.4.

VI.2.2-A Purification des plateformes

De façon à diminuer le temps nécessaire à la purification de chaque brut de synthèse, nous avons opté pour l'utilisation d'une seule méthode, celle-ci devant être rapide et résolutive. La chromatographie en phase liquide à haute performance [277] en phase inverse est une méthode de choix pour l'analyse de mélanges complexes, permettant l'identification, la quantification, et la purification de composés très divers quels que soient leur taille, leur état de charge ou leur nature (peptides, protéines, composés organiques). Elle est donc particulièrement bien adaptée à la purification des différentes plateformes à partir de « bruts de synthèse », mélange de fragments peptidiques dont les propriétés physico-chimiques peuvent être très variables. L'analyse des profils de synthèse permet de déterminer le rendement global (donc la quantité attendue pour chaque plateforme) et d'identifier les différents fragments délétères présents dans le mélange. Le gradient utilisé et le temps d'élution, pour chaque plateforme purifiée, sont présentés dans la partie « matériels et méthodes ».

Les plateformes 1STU wt/mt et 1H50mt2 n'ont pu être purifiées : aucune des différentes fractions collectées lors de la purification, analysées par spectrométrie de masse ne correspondant pas aux masses moléculaires attendus. Ces résultats ont été confortés par la réalisation d'électrophorèse en conditions dénaturantes (SDS-PAGE). De ce fait, la capacité de ces plateformes à lier le canal Kv1.2 n'a pu être évaluée dans le cadre de ce travail. Les autres plateformes ont, en revanche, toutes été obtenues avec une pureté supérieure à 95%.

VI.2.2-B Repliement des plateformes à ponts disulfures

La présence des ponts disulfures étant nécessaire pour conférer à une protéine sa structure tertiaire, la formation de ceux-ci a été effectuée selon le protocole décrit dans la partie « matériels et méthodes ». Le nombre d'appariements possibles est une combinaison du nombre de résidus cystéine de la protéine considérée : $C_p^n = n! / p! - (n-p)!$, avec n le nombre de résidus cystéine, et p le nombre de résidus impliqués pour la formation d'un pont (soit 2). Pour une protéine contenant 6 cystéines (cas de nos plateformes), 15 repliements sont possibles, chacun ayant, en théorie, la même probabilité d'occurrence. Or, l'expérimentation montre que dans la plupart des cas, la forme majoritairement observée correspond à la forme naturelle.

Les plateformes 1C5Amt, 1CR8mt et 1D6Bmt, sous leur forme réduite dans le mélange initial, ont pu être purifiées par RP-HPLC. En revanche, les formes réduites purifiées des plateformes 1C5Amt et 1CR8mt se sont avérées insolubles en milieu basique (pH=7,8), condition nécessaire au repliement des ponts disulfures. La plateforme 1D6Bmt, sous sa forme oxydée, n'a pu être purifiée en raison d'une très faible solubilité en milieu acide. L'utilisation d'agents dénaturants à différentes concentrations (chlorure de guanidinium, de 0,5 à 2M, ou urée, de 4 à 6M) n'a pas permis de remédier à ce problème. Pour ces raisons, la capacité de ces plateformes à lier le canal Kv1.2 n'a pu être évaluée.

La plateforme 1H5O, possédant trois ponts disulfures a été repliée en utilisant un protocole standard [258]. Les chromatogrammes obtenus après 3h d'oxydation sont identiques pour les différentes séquences de 1H5O (native et mutantes). Toutefois, ceux-ci présentent deux pics résolus, pics 1 et 2, correspondant à des protéines de même masse moléculaire. Chaque pic collecté séparément a été réinjecté dans des conditions identiques. Les chromatogrammes étant similaires au précédent, schématisé sur la figure VI-10, témoignent de la présence d'un équilibre dynamique, qui ne peut donc pas être expliqué par la présence de deux repliements distincts, résultant d'un appariement différent des ponts disulfures.

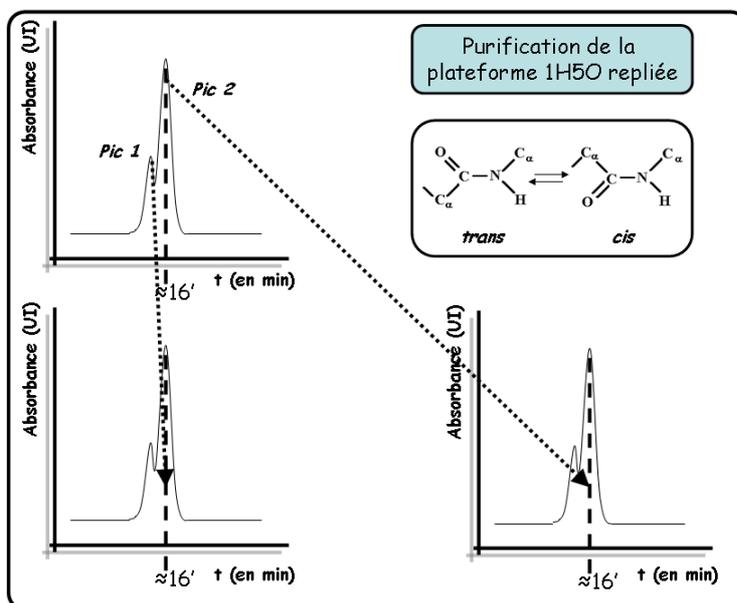


Figure VI-10: Profil d'oxydation des différentes versions de la plateforme 1H5O. Après 3h d'oxydation, deux pics sont obtenus. Chaque pique injecté séparément fournit le même profil, témoignant de la présence d'un équilibre conformationnel. La présence d'un enchaînement Pro-Pro dans la séquence, favorable à une isomérisation pourrait expliquer ce résultat.

En outre, cette protéine possède un enchaînement -Pro-Pro-. Or il a été démontré, dans le cas de peptides riches en résidus proline (en absence d'isomérases) que la conformation de tels enchaînements était l'objet d'un équilibre lié à l'isomérisation cis-trans de ce type de résidu, comme le montre le tableau VI-11 [205].

type	Conformation des résidus Pro 1 et 2			
	cis-cis	cis-trans	trans-cis	trans-trans
α - α	—	—	—	25
α - P_{II}	—	—	—	—
P_{II} - α	—	5	—	129
P_{II} - P_{II}	1	17	9	229
P_{II} -bridge	—	1	19	19
others	—	1	—	7
total	1	24	28	409
%	0.22	5.19	6.06	88.53

Proportion de chaque conformation observée

Tableau VI-11 : Estimation de la proportion de chaque conformation possible pour un enchaînement Pro-Pro, en fonction de la structure secondaire portant chaque résidu [205]

La présence de cet équilibre conformationnel serait donc uniquement liée à la séquence primaire de la plateforme 1H5O et ne serait pas liée à la présence d'appariements divers des ponts disulfures. Par ailleurs, le travail présentant la structure 3D de la plateforme 1H5Owt a été réalisé à partir de protéines synthétisées, purifiées et repliées suivant un protocole similaire au cours de ce travail [258]. Il est donc probable que l'appariement des ponts disulfures, lors des expériences réalisées, soit similaire. La séquence de la protéine 1H5O possède un enchaînement Cys-Cys, interdisant toute attribution des ponts disulfures par la réalisation de divers clivages protéolytiques. Cette opération nécessiterait la mise en œuvre d'expériences de RMN, relativement longues. De ce fait, cette caractérisation n'a pas été développée plus avant.

VI.2.2-C Caractérisation biochimique des plateformes

Les plateformes purifiées ont été caractérisées par RP-HPLC analytique, afin d'estimer la pureté (>95%), par spectrométrie de masse ES-MS et par analyse en composition, en vue de déterminer leur coefficient d'extinction molaire. Les résultats obtenus, lors de la caractérisation de chaque plateforme sont présentés dans l'annexe C.4, tableau C4-2.

VI.2.3 Conclusion

La synthèse de plateformes protéiques par voie chimique s'est avérée plus rapide que la production par biologie moléculaire. Cependant, dans le cas des plateformes 1STU, 1C5A, 1CR8 et 1D6B, les étapes de purification et de repliement n'ont pas permis d'aboutir à la production de ces protéines. Le taux de succès pour la production des plateformes par synthèse peptidique est proche de 50%. Quatre plateformes ont pu être produites pour des caractérisations ultérieures, aussi bien de leur structure que de leur affinité vis-à-vis des canaux potassiques.

VI.3 Caractérisation structurale des plateformes

Notre approche de conception de ligands est basée sur l'utilisation de plateformes protéiques reproduisant la topologie fonctionnelle du motif de liaison F6/S23/K25/Y26 de la BgK. Il est donc nécessaire que la structure des plateformes ne soit pas altérée à la suite des substitutions réalisées (transfert du motif de liaison).

VI.3.1 Caractérisation par dichroïsme circulaire

Le dichroïsme circulaire permet une caractérisation globale et rapide de la structure 3D des plateformes synthétisées (annexe D.1). Cette méthode a, de ce fait, été utilisée afin d'estimer la structure globale des plateformes natives (wt), au regard de leurs structures 3D respectives, présentées dans la PDB. L'influence de l'insertion du motif sur la structure a été évaluée par comparaison des spectres CD des plateformes modifiées et natives.

VI.3.1-A Plateforme 1CKK

La plateforme 1CKK [255] dans sa forme native, présente comme principal élément de structure secondaire une hélice α , représentant approximativement 40% de la structure, représentée par la figure VI-12. Le reste de la structure, comprenant notamment une boucle, apparaît relativement flexible. Les différents spectres obtenus sont présentés sur la figure VI-12bis.

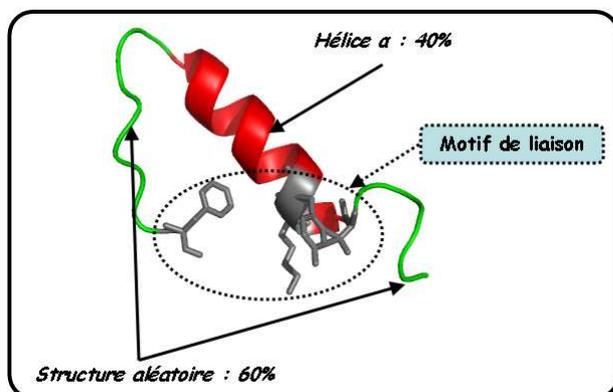


Figure VI-12 : Représentation en mode « cartoon » de la plateforme 1CKKmt. Le motif de liaison F26/S6/K8/Y9 (en gris) est représenté en mode « stick ».

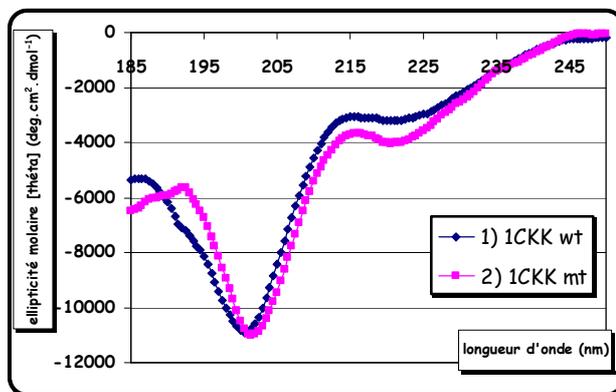


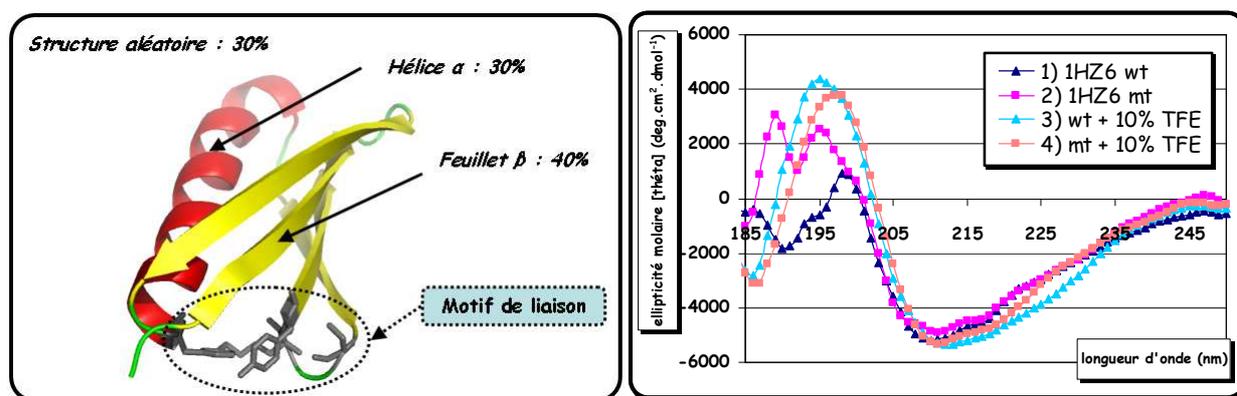
Figure VI-12bis : Spectres CD de la plateforme 1CKK, variants wt et mt

Le spectre 1, correspondant à la plateforme wt, contient plusieurs caractéristiques d'éléments désordonnés (pic à 200nm) et d'éléments en hélice (pic à 220nm). Ces observations sont en accord avec les observations faites à partir de la structure 3D présente dans le fichier 1CKK. La comparaison des spectres obtenus pour les formes wt et mt, montrent certes une augmentation de la proportion d'hélices α , de l'ordre de 20% pour la forme mt, toutefois, la nature des

éléments de structure secondaire sont identiques. Ceci indique que les deux substitutions effectuées (T8K, T9Y) n'ont pas modifié significativement la structure de la plateforme portant le motif de liaison.

VI.3.1-B Plateforme 1HZ6

La plateforme 1HZ6, forme native [263], possède différents éléments de structure secondaire, représentés sur la figure VI-13 : une hélice α (30% de la structure) et un feuillet, composé de quatre brins β antiparallèles (40% de la structure). Les spectres CD sont représentés sur la figure VI-13bis.



Le spectre 1 de la plateforme 1HZ6wt ne permet pas de déduire les éléments de structure secondaire de cette plateforme, toutefois sa forme suggère la présence de multiples éléments (hélice et feuillet).

Le TFE est connu comme étant un agent stabilisant des éléments de structure secondaire et un inducteur de structures en hélice. Ce dernier effet peut être observé pour des structures composées uniquement de feuillets mais seulement pour des quantités de TFE supérieures à 30% [278, 279]. De ce fait, pour plusieurs plateformes décrites par la suite, des ajouts de TFE ont été effectués sans dépasser le taux de 20% en composition volumique, afin de stabiliser les éléments de structure secondaire présents dans ces protéines. En présence de TFE (10%), le spectre CD de 1HZ6 présente plusieurs pics caractéristiques : un pic positif à 195nm caractéristique de feuillets β , et deux pics négatifs autour de 208-210nm et 220nm, caractéristique d'hélice α . Ces observations sont en accord avec la structure pdb de 1HZ6. Par ailleurs, la zone de longueur d'onde comprise entre 185 et 200 nm, en milieu tamponné (Tris,

4mM, pH 6.8), montre des différences notables entre les formes native (spectre 1) et mutante (spectre 2). L'analyse de la forme native [280] a démontré que plusieurs résidus dont le résidu Gly 55, étaient impliqués dans le processus de repliement. La modification du résidu lors de l'introduction du motif, pourrait ainsi entraîner une modification de la stabilité de la protéine. Toutefois, en présence de 10% de TFE, aucune différence significative n'a été observée entre les spectres CD de 1HZ6 wt (spectre 3) et 1HZ6 mt (spectre 4), y compris dans la zone de longueur d'onde 185-200 nm. Cette observation suggère que, malgré l'incorporation du motif et la substitution du résidu Gly 55, les plateformes 1HZ6wt et 1HZ6mt pourraient adopter des structures 3D similaires.

VI.3.1-C Plateforme 1H5O

La structure secondaire de la forme native de 1H5O [258] comprend une hélice α , représentant approximativement 10% de la structure et un feuillet β , composé de trois brins antiparallèles, représentant 25% de la structure ; les structures en feuillets apparaissent donc prépondérantes comme indiqué sur la figure VI-14. Les spectres CD sont présentés sur la figure VI-14bis. Le spectre 1 pour la forme native montre deux pics principaux à 197nm (positif) et 207nm (négatif). Il a été décrit [258] que la présence de pics autour de 195nm et 210-215nm était caractéristique d'éléments en feuillets, mais que le déplacement de ce second pic vers 207nm indiquait la contribution de structures en hélice, corroborant ainsi la structure de la crotamine, déterminée expérimentalement.

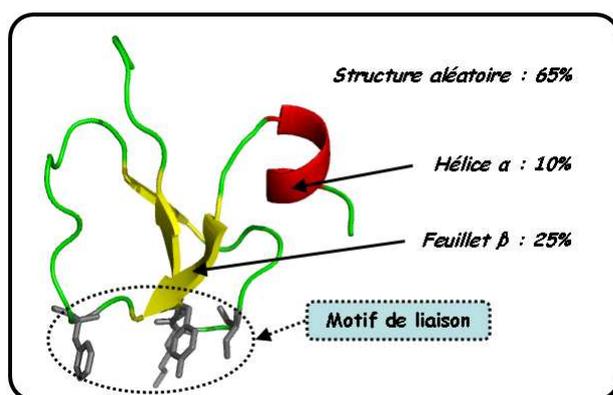


Figure VI-14 : Représentation en mode « cartoon » de la plateforme 1H5Omt. Le motif de liaison F15/S31/K33/Y34 (en gris) est représenté en mode « stick ».

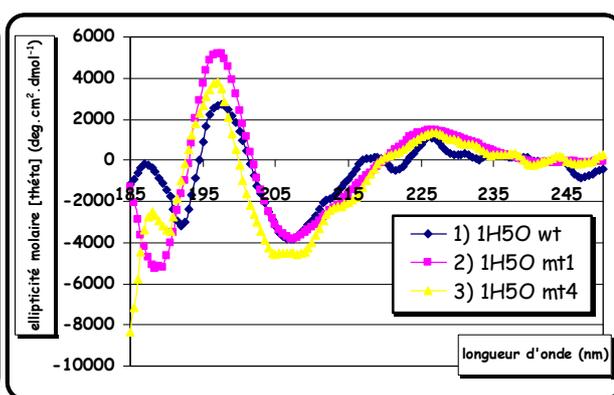


Figure VI-14bis : Spectres CD des variants, wt, mt1 et mt4 de la plateforme 1H5O.

Plusieurs spectres CD ont également été réalisés, sur les mêmes échantillons, après dénaturation thermique et renaturation douce ne montrant aucune différence significative. L'insertion du motif ou sa modification (KY-AA) n'entraîne donc pas de modification du repliement de la

protéine. De plus, le spectre 1 (1H50wt) est relativement similaire au spectre CD (figure VI-15) présenté lors de la détermination de la structure 1H50 [258].

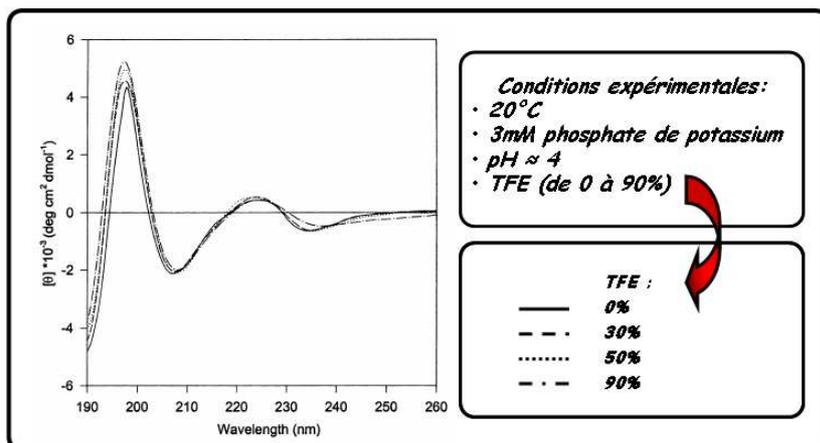


Figure VI-15 : Spectres CD de la crotonamine (plateforme 1H50wt). Ceux-ci sont similaires aux spectres que nous avons obtenus, suggérant un repliement similaire. Par ailleurs, l'addition de TFE (jusqu'à 90%) ne semble pas modifier la structure secondaire.

Ces observations suggèrent que les ponts disulfures formés correspondent effectivement aux ponts disulfures de la plateforme native.

VI.3.1-D Plateforme 1J75

Bien que la structure tridimensionnelle de la plateforme 1J75mtK30A soit présentée par la suite, les spectres de dichroïsme circulaire des formes native et mutante sont présentés. La forme native, décrite dans le fichier pdb 1J75 [260], présente trois hélices α comme seuls éléments de structure secondaire représentant approximativement 50% de la structure, schématisés sur la figure VI-16. Les différents spectres CD obtenus sont présentés sur la figure VI-16bis.

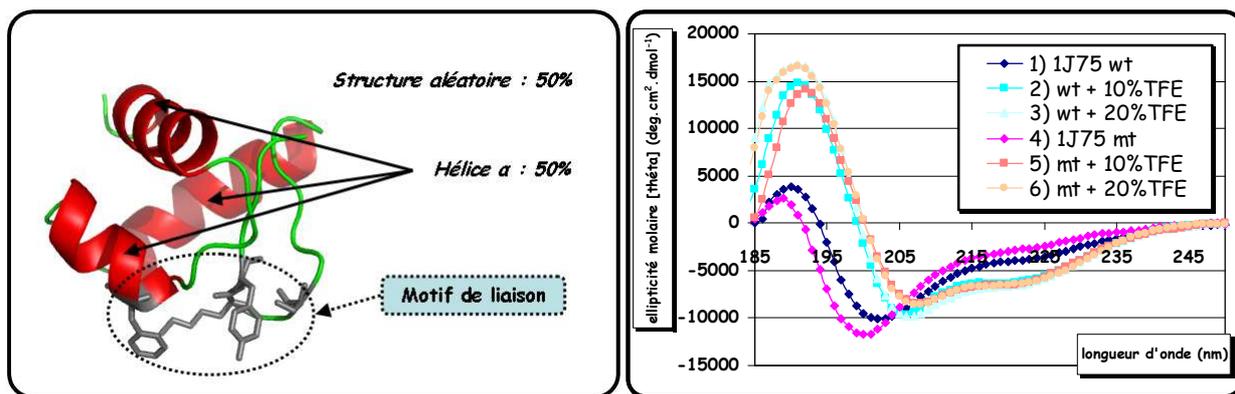


Figure VI-16 : Représentation en mode « cartoon » de la plateforme 1J75mt. Le motif de liaison F32/S62/K64/Y65 (en gris) est représenté en mode « stick ».

Figure VI-16bis : Spectres CD des versions wt et mt de la plateforme 1J75, avec ou sans TFE (10% ou 20%)

Le spectre 1, correspondant à la forme wt, montre un pic négatif à 200nm, caractéristique d'une structure désordonnée, un pic positif à 190nm, de faible amplitude, et une inflexion autour de

220nm, suggérant la présence de structures en hélice. Un spectre a été obtenu pour la forme mutante dans des conditions similaires, montrant un décalage vers les plus faibles longueurs d'ondes (effet bathochrome). En présence de 10% de TFE, le spectre CD montre un signal caractéristique de la présence d'hélice α : un pic positif autour 190nm et deux minimum autour de 205nm et 220nm. Par ailleurs, d'autres expériences ont été menées, après addition de chlorure de potassium, à des concentrations finales de 10, 50 ou 100mM. Dès 50mM de KCl, les pics caractéristiques de structures en hélice sont observés. Les conclusions sont donc identiques après addition de 10% TFE ou de 50mM KCl. L'insertion du motif de liaison n'a donc, *a priori*, pas modifié la structure de la forme native de la plateforme, constituée essentiellement d'hélice α . Les spectres des mutants C38V, K30A, et FSKY/AAAA ne sont pas montrés, néanmoins en présence de tris, 4mM, pH 6,8, le comportement de la forme K30A est similaire à celui de la forme native.

VI.3.2 Détermination de la structure tridimensionnelle de 1J75mt K30A

Les résultats des tests de liaison (présentés ultérieurement) ont montré que la plateforme 1J75mt était capable de lier le canal Kv1.2, en déplaçant l'analogue de la BgK radiomarqué. Ces résultats nous ont incité à déterminer sa structure tridimensionnelle par radiocristallographie (annexe D.2), de façon à caractériser la topologie du motif de liaison, une fois inséré sur la plateforme.

VI.3.2-A Essais de cristallogenèse

Les premiers essais de cristallogenèse ont été effectués pour les plateformes 1J75mt, 1J75mtC38V et 1J75mtK30A. Toutefois, les plateformes 1J75mt C38V et 1J75 mt présentent des rendements de production relativement faible dans le système d'expression (300-500 μ g/production) par rapport à celui observé pour 1J75mtK30A (3-10mg/production), ainsi qu'une solubilité relativement faible (moins de 4 mg/ml) comparativement à celle de la plateforme 1J75mtK30A (supérieure à 17 mg/ml). Ainsi, les études de cristallogenèse n'ont été réalisées que sur la seule plateforme 1J75mtK30A. La cristallisation a été réalisée par la méthode de diffusion de vapeur en goutte assise (annexe D.2), en adaptant les conditions décrites pour l'obtention de la structure 3D de 1J75 [260]. La cristallisation a été obtenue en présence d'ADN Z double brin dans les conditions suivantes : 25% PEG 4K / 0.2M Na/K Phosphate / 0.1M MES pH=6.0. Nos résultats diffèrent de ceux présentés par plusieurs points : les conditions de cristallisation (15% PEG 4K / 0,1 mM MES, pH 6,0 / 0,1M phosphate d'ammonium), le temps nécessaire à l'obtention de cristaux (1 à 2 jours contre 6 à 9 jours) et sur la nature des cristaux obtenus : dans nos conditions expérimentales, deux formes cristallines ont été obtenues, prismatique et aciculaire, dans le même puit de cristallisation. Leurs photos respectives sont montrées sur la figure VI-17.

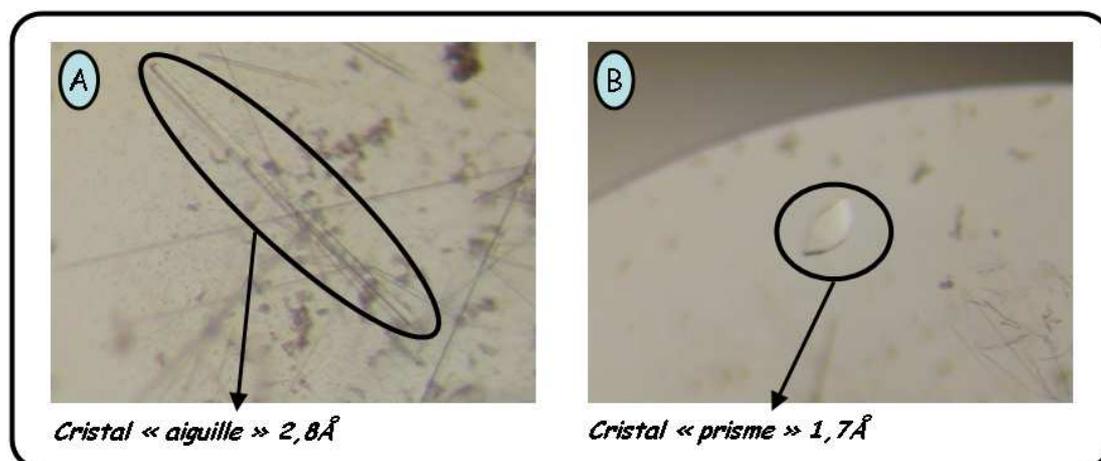


Figure VI-17 : Photographies des cristaux du complexe 1J75mtK30A avec un ADN Z double brin. Les cristaux prélevés sont de forme aciculaire (partie A) et prismatique (partie B).

VI.3.2-B Affinement de la structure

Les deux structures ont été résolues par remplacement moléculaire, à partir de la structure 3D de la protéine 1J75 (matériels et méthodes). Le cristal de structure prismatique, dénommé K30A-2 de résolution la plus basse de 1,7Å a permis l'analyse de la topologie du motif de liaison. Sa structure a été déposée dans la PDB (2HEO). Elle est représentée sur la figure VI-18.

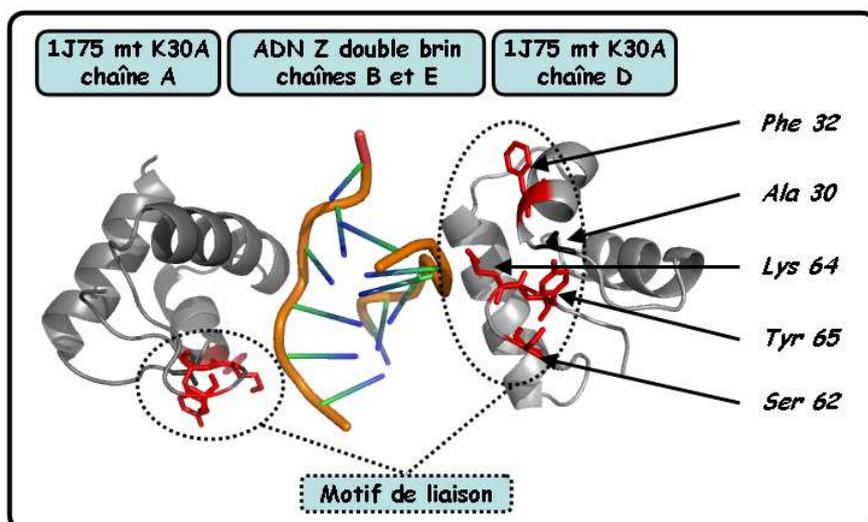


Figure VI-18 : Structure du complexe 1J75mtK30A en interaction avec un double brin d'ADN Z. Les structures protéiques (en gris) et l'ADN Z (en orange) sont représentées en mode « cartoon ». Les résidus du motif de liaison (en rouge) sont représentés en mode « stick ».

La validation de la structure a été obtenue par la réalisation « d'omit-map » présentées sur la figure VI-18bis. Les résultats de l'affinement sont présentés dans le tableau VI-18ter.

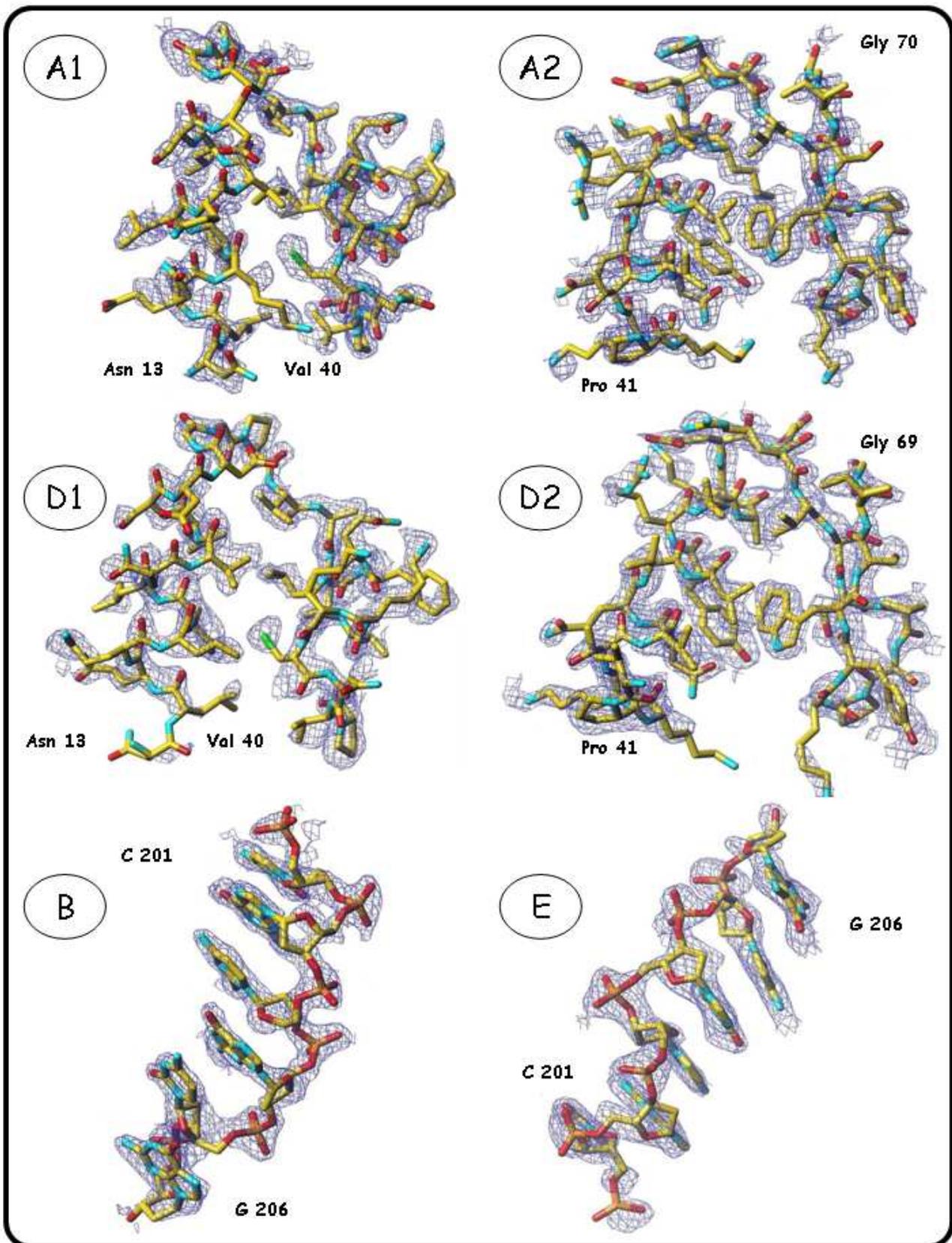


Figure VI-16bis : Les « omit-maps » présentées ont été calculées à partir de cartes de densité électronique $2Fo-Fc$ moyennées (contours à 1σ). A1 et D1 correspondent aux contours des résidus N13 à V40 (chaînes A et D), en absence de modèle pour ces mêmes résidus. A2 et D2 correspondent aux contours des résidus P41 à G69/G70 (chaînes A et D), en absence de modèle pour ces mêmes résidus. B et E correspondent aux brin d'ADNZ (chaînes B et E).

Statistiques après collecte des données

Dénomination	K30A-4	K30A-2
Forme du cristal	aiguille	prisme
Groupe d'espace	P6 ₅ 22	P61
Paramètres de maille(Å) et	a=b=63,15, c=72,97 α=β=90,00, γ=120,00	a=b=79,91, c=55,33 α=β=90,00, γ=120,00
Résolution (Å)	65-2,8	65-1,7
Réflexions mesurées/uniques	13344 / 2239	155589 / 22226
Complétude (%)	95,7 (97,6)	99,9 (100,0)
R _{merge} (%)	9,7 (32,1)	8,2 (41,7)
Rapport I/σ (I)	13,7 (4,8)	16,1 (3,1)
Multiplicité	6,0 (6,4)	7,0 (7,0)
Facteur de température B (Å ²)	55,7	25,8

Statistiques après affinement

Rfacteur (%) / R _{free} (%)	19,6 / 32,3	24,1 / 30,5
RMS bond (Å)	0,023	0,013
RMS angle (Å)	2,200	2,007
RMS chiral (Å)	0,120	0,158
Fom (%)	0,716	0,763
Nombre d'atomes		
Protéine A	460	477
Protéine D		459
ADN B	123	124
ADN E		124
Eau W	49	228
Facteurs de température B moyenné (Å ²)		
Protéine A	33,594	31,956
Protéine D		32,521
ADN B	31,057	30,024
ADN E		31,216
Eau W	30,532	41,758

Tableau VI-16bis : Résultats de l'affinement des structures de K30A-2 et K30A-4 par SCALA (« matériels et méthodes »). Pour le cristal K30-4, l'affinement n'ayant pas été achevé, les valeurs en rouge sont donc montrées, à titre d'information.

VI.3.2-C Description de la structure

Les structures 3D des deux chaînes A et D de la plateforme 1J75mtK30A obtenues (code PDB 2HEO) ont été comparées à celle de la plateforme 1J75wt (code PDB 1J75), via le calcul de RMSD effectué sur les atomes Ca de la chaîne principale ou sur les atomes Ca et Cβ des résidus du motif de liaison. Les résultats sont présentés dans le tableau VI-19.

Variants de 1J75	Chaîne	1J75			1BGK
		Ca (résidus 113-169)	Ca, C β (résidus 113-169)	Ca, C β (motif)	Ca, C β (motif)
1J75mtK30A	A	0,34	0,45	0,59	0,88
1J75mtK30A	D	0,49	0,59	0,54	0,84
1J75 (wt)	A	n.a.	n.a.		0,35

Tableau VI-19 : Comparaisons des structures 3D de 1J75mtK30A, 1J75 et 1BGK. Les RMSD (en Å) sont calculés sur les positions des atomes Ca de tous les résidus de la protéine (Ca des résidus 113-169), les atomes Ca et C β de tous les résidus de la protéine (Ca, C β des résidus 113-169), et les atomes Ca et C β des résidus du motif (Ca, C β motif).

La topologie du motif a été comparée à celle du motif de liaison de la BgK. : les valeurs des RMSD calculées sur les positions des atomes Ca et C β du motif de liaison (BgK et 1J75mtK30A) ou des résidus topologiquement équivalents dans la structure 1J75, sont inférieures à 0,88 Å. Ceci montre que la topologie du motif transféré est similaire, en ce qui concerne l'orientation relative des chaînes latérales, à celle observée dans la structure du ligand de référence. Il est à noter que ces valeurs demeurent inférieures à la valeur du RMSD_MAX (1 Å) utilisée lors de l'étape de recherche topologique. Ces résultats montrent que la structure tridimensionnelle de 1J75mtK30A (chaînes A et D de 2HEO) est très similaire à la structure 1J75 pour ce qui concerne le chaîne principale (RMSD Ca 113-169 < 0.59 Å) mais également les chaînes latérales (RMSD Ca, C β < 0.59 Å). Les mutations réalisées pour l'insertion du motif de liaison de la BgK sur la protéine 1J75 n'ont donc pas entraîné de modification de sa structure 3D. Ceci est également illustré sur la figure VI-20, laquelle représente la superposition des chaînes principales de la protéine 1J75mtK30A (2HEO, chaîne A et D) avec celle de la structure 1J75.

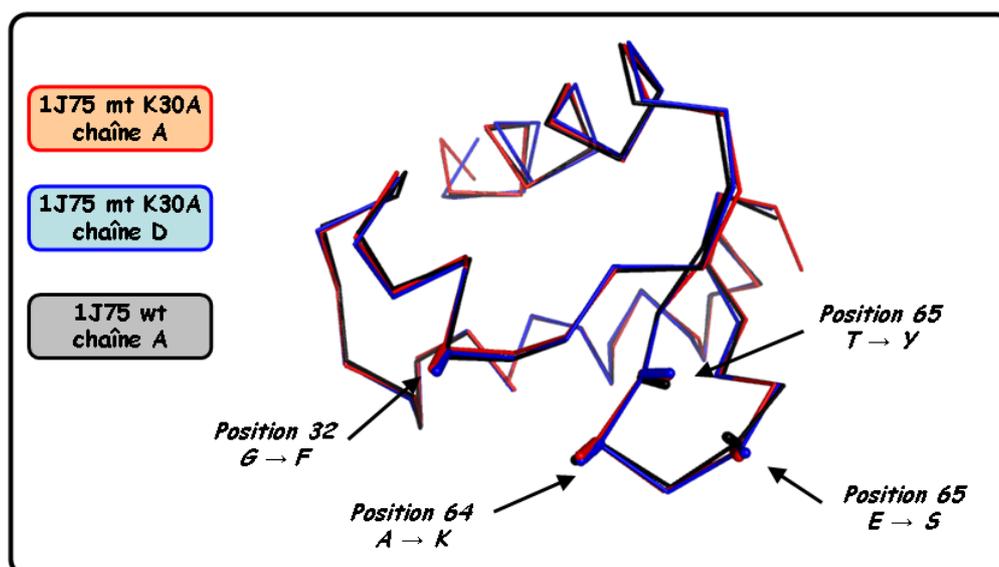


Figure VI-20 : Représentation en mode « ribbon » des plateformes 1J75wt (en noir) et mtK30A (en rouge pour la chaîne A et en bleu pour la chaîne D). Les atomes Ca et C β des chaînes latérales des positions 32, 62, 64 et 65 du motif identifiés sont représentés en mode « stick ». D'après les valeurs de RMSD calculées, il n'y a pas de différences de structures entre chacune de ces plateformes.

L'examen de la structure de 1J75mtK30A révèle cependant que le rotamère adopté par la chaîne latérale F32 dans la structure 1J75mtK30A (configuration \dagger $\chi_1=180^\circ$) est différent de celui adopté dans la structure de la BgK (configuration g^+ , $\chi_1=-60$). Le motif de liaison de la plateforme 1J75mtK30A, superposé à celui de la BgK, est représenté sur la figure VI-21. L'examen de cette structure suggère néanmoins que la configuration *trans* est accessible à la chaîne latérale F32, car aucune gêne stérique ne semble pouvoir empêcher ce changement de configuration.

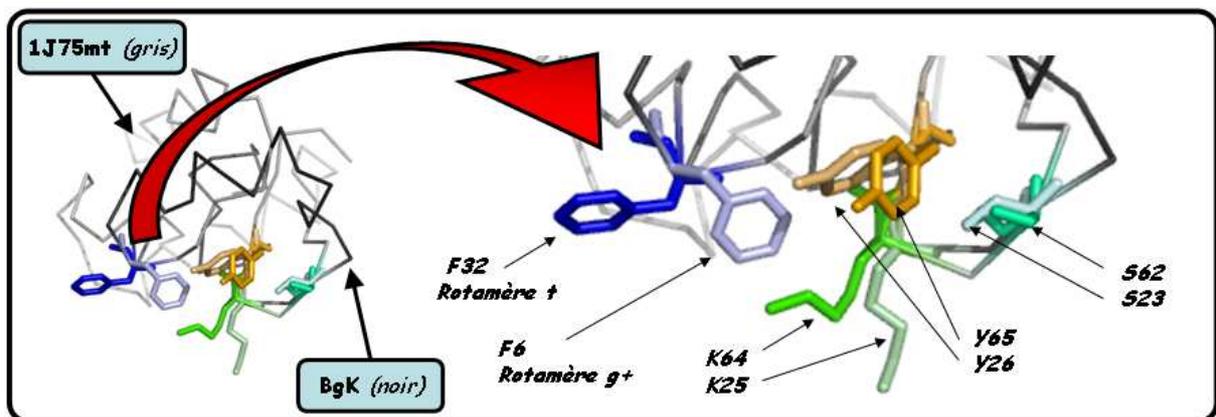


Figure VI.21 : Représentation des structures de la BgK (noir) et de la protéine 1J75mt (gris), superposées. Les protéines sont représentées en mode « ribbon », les résidus en mode « stick » et colorés selon le type de résidu : Phe en bleu, Ser en cyan, Lys en vert et Tyr en orange (les couleurs sont claires pour la BgK).

VI.3.3 Conclusion

Les résultats des expériences réalisées, afin d'évaluer l'impact du transfert du motif de liaison sur la structure 3D des plateformes, ne montrent pas de conséquence significative sur la structure 3D des plateformes produites (dichroïsme circulaire, annexe D.1 ; radiocristallographie, annexe D.2). Si la caractérisation structurale, réalisée dans le cas de la plateforme 1J75mtK30A, permet de définir, sans ambiguïté, sa structure tridimensionnelle, d'autres expériences (radiocristallographie ou RMN) seraient nécessaires pour caractériser la structure 3D des autres plateformes synthétisées.

VI.4 Évaluation expérimentale des ligands protéiques

Le principe général de la méthode des tests de liaison par compétition ainsi que les conditions expérimentales sont décrites dans l'annexe D.3.

VI.4.1 Caractérisation de l'affinité

La liaison des différentes plateformes produites au canal Kv1.2, a été déterminée en réalisant des tests de liaison par compétition *via* le déplacement d'un analogue radio-marqué de la BgK [281] (annexe D.3). L'affinité a été mesurée sur des membranes de cellules, surexprimant le canal Kv1.2. Les résultats obtenus sont présentés dans le tableau VI-22.

Sur Kv1.2	Plateformes protéiques		BgK	
	Ki en μM	nH	Ki en pM	nH
1J75 wt	25,9 \pm 9,5	0,88 \pm 0,13	60 \pm 20	1,21 \pm 0,33
1J75 mt	1,5 \pm 0,7	1,17 \pm 0,45	69 \pm 17	1,17 \pm 0,39
-K30A	2,9 \pm 1,4	1,06 \pm 0,46	77 \pm 74	1,35 \pm 0,40
-K17A	1,7 \pm 0,5	0,70 \pm 0,23	76 \pm 34	1,25 \pm 0,18
-K64A	6,4 \pm 1,8	1,09 \pm 0,45	110 \pm 90	1,08 \pm 0,21
-F32A	3,8 \pm 1,1	0,86 \pm 0,36	94 \pm 17	1,36 \pm 0,14
-C38V	3,8 \pm 0,4	0,87 \pm 0,23	122 \pm 115	1,03 \pm 0,13
-FSKY/A	17,3 \pm 6,0	0,84 \pm 0,22	142 \pm 50	1,20 \pm 0,29
1CKK wt	2,3 \pm 0,5	1,2 \pm 0,2		
1CKK mt	1,6 \pm 0,5	1,3 \pm 0,6		
1HZ6 wt	> 100			
1HZ6 mt	> 100			
1H50 wt	1,1 \pm 0,4	0,93 \pm 0,14		
1H50 mt	0,5 \pm 0,1	0,96 \pm 0,23		
-KY/AA	6,6 \pm 3,2	0,97 \pm 0,23		
-KY/EY	7,5 \pm 4,1	1,55 \pm 0,38		

Tableau VI-22 : Résultats des tests de liaison par compétition. Les valeurs de K_i , pour chaque plateforme, sont obtenues par le déplacement d'un analogue radiomarqué de la BgK vis-à-vis du canal Kv1.2.

Parmi les plateformes testées, seules les plateformes 1HZ6, wt et mt, ne déplacent pas l'analogue radiomarqué, et ce, pour des concentrations allant jusqu'à 100 μM . Les autres plateformes montrent en revanche un déplacement de l'analogue radiomarqué de la BgK, avec des constantes d'inhibition de l'ordre du μM : de 0,5 à 25,9 μM . Ces résultats suggèrent que l'application des critères stériques et de similarité électrostatique avec la BgK, permet de sélectionner des structures protéiques dont les propriétés globales les rendent capables d'interagir avec le canal Kv1.2, au site de liaison de la BgK.

Les tests réalisés permettent également d'évaluer l'effet de l'insertion du motif sur la liaison au canal Kv1.2. L'introduction du motif de liaison de la BgK sur les différentes plateformes produites entraîne, dans chaque cas, une augmentation de l'affinité par rapport à la protéine native. Cette augmentation est cependant variable suivant les plateformes : facteur 17 pour la protéine 1J75mt, facteur 2,2 pour 1H50mt et facteur 1,4 pour 1CKKmt. Sous l'hypothèse que le mode de liaison des plateformes natives (wt) est similaire à celui des plateformes portant le motif transféré (mt), l'interprétation des différences observées doit prendre en compte la nature des résidus présents dans les plateformes natives. En effet, ces résidus sont susceptibles de fournir des contributions défavorables ou favorables à la liaison au canal Kv1.2. Le tableau VI-23 représente les acides aminés natifs présents aux positions substituées lors du transfert de motif.

	Position n°1	Position n°2	Position n°3	Position n°4
BgK	Phe 6	Ser 23	Lys 25	Tyr 26
1CKK	Phe 26	Ser 6	Thr 8	Thr 9
1H50	Glu 15	Arg 31	Arg 33	Trp 34
1J75	Gly 132	Glu 162	Ala 164	Thr 165

Tableau VI-23 : Comparaison des résidus présents dans les plateformes natives aux positions topologiquement équivalentes à celles du motif de liaison de la BgK.

Pour la plateforme 1H50, les résidus natifs dans la position topologiquement équivalente à celle des résidus K25 et Y26 de la BgK, sont respectivement R33 et W34. Pour ces deux positions, les substitutions réalisées lors du transfert du motif n'entraînent donc pas de modification du type de résidu (couple de résidus basique/aromatique). Une observation similaire est faite concernant la plateforme 1CKK, dont les résidus F26 et S6 sont topologiquement équivalents aux résidus F6 et S23 du motif de liaison de la BgK. Ces éléments sont, sans doute, partiellement responsables des faibles variations d'affinités observées pour les plateformes 1CKK et 1H50, entre les formes natives et modifiées, après transfert du motif. Afin d'estimer plus amplement l'influence des résidus du motif transféré sur la liaison au canal Kv1.2., une série de variants des plateformes 1H50 et 1J75 ont été produits.

VI.4.1-A La plateforme 1H50

Nous avons produit deux variants de la plateforme 1H50 mt. Le premier, désigné 1H50-KY/AA, correspond à la substitution des deux résidus, K33 et Y34, de la diade de 1H50mt, en alanine. Le second, désigné 1H50-KY/EY, correspond à une inversion de charge par la substitution du résidu K33 en acide glutamique. Pour ces deux variants, l'affinité mesurée pour le canal Kv1.2 a été diminuée de facteurs 13 et 15 respectivement par rapport à la plateforme 1H50 mt (tableau VI-

22). Ces résultats sont donc en accord avec l'hypothèse que les résidus transférés aux positions 33 et 34 de la plateforme 1H5O soient effectivement impliqués dans l'interaction avec le canal Kv1.2.

VI.4.1-B La plateforme 1J75

Afin de mieux caractériser l'interaction de la plateforme 1J75mt avec le canal Kv1.2, trois variants impliquant le motif de liaison ont été réalisés : 1J75mtK64A, 1J75mtF32A et 1J75mtFSKY/A. La substitution en alanine des résidus équivalant aux résidus F6 et K25 de la BgK, sur la plateforme 1J75mt entraîne respectivement une diminution d'affinité d'un facteur 2,5 et 4,2 pour le canal Kv1.2. La substitution simultanée des quatre résidus du motif (1J75mtFSKY/A) entraîne une diminution de l'affinité de la plateforme par un facteur 11,3 sur Kv1.2.

La caractérisation du mode de liaison de la plateforme 1J75mt a été poursuivie, par la production de deux autres variants, par la substitution de résidus lysine présents à la surface de la plateforme 1J75mt. Le premier de ces résidus (K17) est situé sur la face opposée à la région d'interaction avec le canal Kv1.2 dans le mode de liaison de 1J75mt reproduisant la position du motif FSKY de la BgK et par conséquent distant de l'interface avec le canal. Le second (K30) est localisé à proximité de l'interface avec le canal Kv1. La structure de ces variants est représentée sur la figure VI-24. Le variant 1J75mtK17A ne présente aucune variation significative de l'affinité par rapport à la plateforme 1J75mt (tableau VI-22). Cependant, le variant 1J75mtK30A montre une faible diminution de l'affinité pour le canal Kv1.2, en comparaison à la plateforme 1J75mt (tableau VI-22). Ces résultats sont cohérents avec l'hypothèse que la plateforme 1J75mt interagit avec le canal Kv1.2, par le motif de liaison transféré, en reproduisant la topologie du motif FSKY de la BgK par rapport au canal.

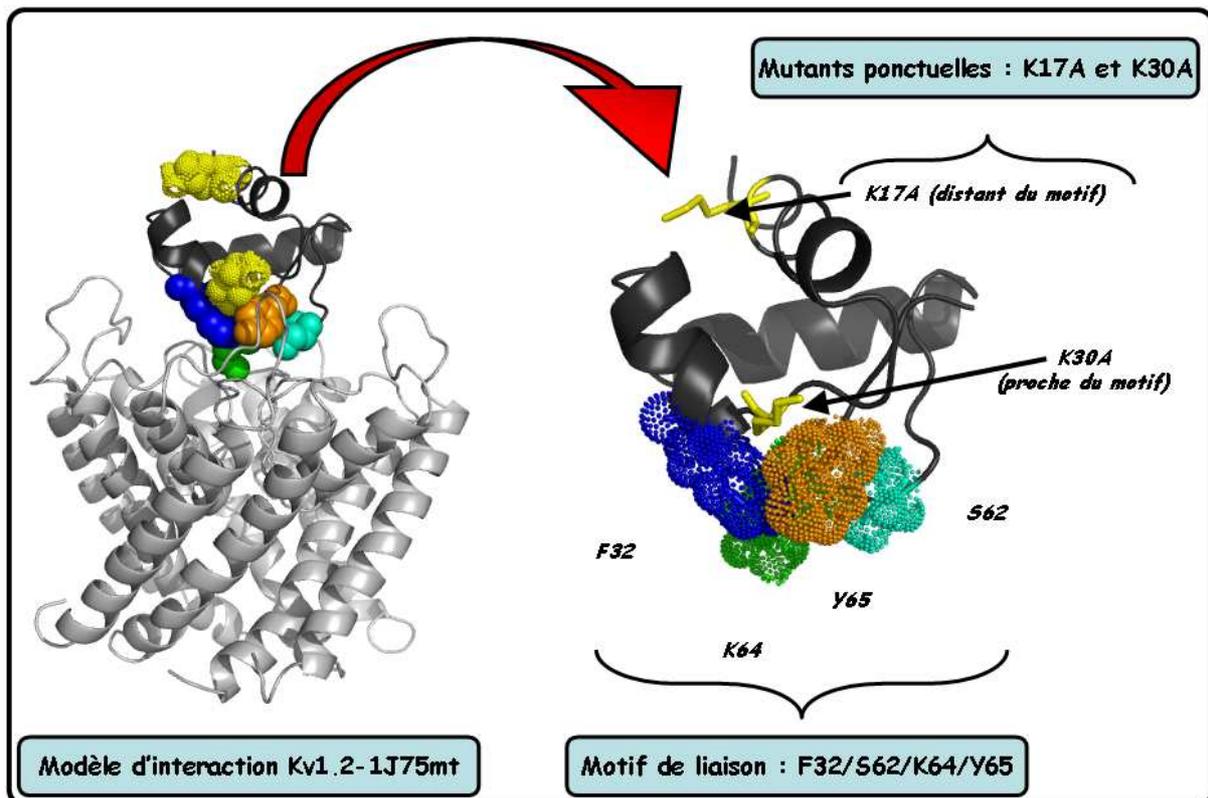


Figure VI-24 : Modèle d'interaction du complexe Kv1.2(en blanc)-1J75mt (en gris). Les positions des mutations K17A et K30A (en jaune) sont indiquées sur la partie gauche. La première mutation est effectuée à distance du site d'interaction supposé, la seconde étant à proximité. Les protéines sont représentées en mode « cartoon », les résidus du motif en mode « sphere » et « dot », les résidus K17 et K30 en mode « dot » et « stick ». Les résidus du motif sont colorés ainsi : F32 en bleu, S62 en cyan, K64 en vert et Y65 en orange.

VI.4.2 Modèles d'interaction ligands/Kv1.2

Les résultats décrits précédemment sont cohérents avec l'hypothèse que les plateformes étudiées interagissent avec le canal Kv1.2 dans un mode de liaison dans lequel les résidus du motif transféré occupent une position similaire à celle occupée par ces mêmes résidus dans l'interaction BgK/Kv1.2. Cependant, l'observation que les plateformes 1HZ6, wt et mt, soient incapables de déplacer l'analogue de la BgK du canal Kv1.2, nous a incité à caractériser en détails les conséquences du mode de liaison des plateformes, notamment en terme d'interaction microscopique. Pour cela, plusieurs modèles moléculaires de complexes des 4 plateformes produites (1J75mt, 1CKKmt, 1H5Omt et 1HZ6mt) avec le canal Kv1.2, ont été réalisés (annexe D.4). L'analyse de ces différents modèles (annexe B.3) permet d'évaluer les implications du mode de liaison supposé en termes de propriétés globales et d'interactions microscopiques.

VI.4.2-A Propriétés globales

Le tableau VI-25 reporte les valeurs des paramètres analysés pour les différents modèles.

PARAMÈTRES GLOBAUX					
LIGAND	1BGK	1J75mt	1H5Omt	1CKKmt	1HZ6mt
Δ ASA interface totale (en \AA^2)	1583	1989	1384	1856	2031
Δ ASA interface ligand (en \AA^2)	782	889	703	991	964
% ligand à l'interface	49,4	44,7	50,8	53,4	47,5
Nb résidus interface	16	24	15	16	23
% résidus non-polaires	37,5	50,0	46,5	62,5	43,5
Volume vide à l'interface (en \AA^3)	1290	1704	1326	1635	1872
Gap index (en \AA)	1,65	1,92	1,89	1,70	1,89

Tableau VI-25 : Paramètres calculés dans les modèles d'interaction de la partie S5-S6 du canal Kv1.2 avec la BgK et les plateformes 1J75mt, 1H5Omt, 1CKKmt et 1HZ6mt. Gap Index: rapport (Volume vide à l'interface / Aire de l'interface). Les valeurs supérieures sont indiquées en rouge, celles inférieures en bleu.

Le premier paramètre caractéristique d'un complexe correspond à la taille de l'interface. La variation de surface accessible au solvant est comprise entre 1384 \AA^2 (pour 1H5Omt) et 2031 \AA^2 (pour 1HZ6mt). Chaque plateforme y contribue de façon variable, de 44,7 à 53,4%. Le nombre de résidus, coté ligand, présents à l'interface est également variable, de 15 (pour 1H5Omt) à 24 résidus (pour 1J75mt). Cependant de telles variations ont été observées lors de l'analyse d'hétéro-complexes [52, 131, 139, 138, 159]. De la même façon, la fraction de résidus non polaires présents à l'interface est également variable de 37,5 pour la BgK à 62,5 % pour 1CKK. Toutefois, la comparaison de plusieurs ligands de protéases à serine (chapitre IV) a démontré que de telles fluctuations peuvent être observées pour des ligands d'une même cible.

La complémentarité de surface des deux partenaires peut-être évaluée par le calcul du volume d'interface, or celui-ci dépend de l'aire de l'interface. L'utilisation du « Gap Index », correspondant à la normalisation du volume d'interface sur l'aire de celle-ci, permet de comparer la complémentarité de surface de plusieurs complexes. Les valeurs de cet indice, calculées pour les 5 complexes, sont comprises entre 1,65 Å et 1,92 Å ce qui montre que la complémentarité de surface entre les différents ligands et le canal Kv1.2 se situe dans la gamme observée dans des hétéro-complexes dont les structures sont accessibles dans la PDB [139] ou aux valeurs calculées dans les protéases à serine étudiées au chapitre IV. Néanmoins, la faible compacité mesurée pour la plateforme 1J75 par rapport à la BgK, pourrait expliquer, en partie, une affinité observée plus faible.

VI.4.2-B Interactions microscopiques

Les modèles de complexes réalisés au cours de ce travail permettent également d'analyser les conséquences du mode de liaison supposé des différentes plateformes, en termes d'interactions microscopiques et de comparer les résultats à ceux observés pour le complexe de référence, BgK/Kv1.2. Ceux-ci sont présentés dans le tableau VI-26.

INTERACTIONS MICROSCOPIQUES					
LIGAND	1BGK	1J75mt	1H50mt	1CKKmt	1HZ6mt
Liaisons hydrogène	10	11	9	5	9
Densité de liaison H	1,3	1,2	1,3	0,5	0,9
Ponts salins	3	5	3	3	4
Charge positive/négative	3/0	3/2	3/0	3/0	1/3
Charge nette globale	5	5	8	6	-1
Charge nette de l'interface	5	3	5	3	-1
Nb charge interface	5	7	7	5	11
Nb charge interface non compensées	2	2	4	2	7

Tableau VI-26 : Paramètres calculés dans les modèles d'interaction de la partie S5-S6 du canal Kv1.2 avec la BgK et les plateformes 1J75mt, 1H50mt, 1CKKmt et 1HZ6mt.

Liaisons hydrogène

Parmi ces interactions, les liaisons hydrogène jouent un rôle prépondérant dans l'association de deux protéines. Leur analyse a été réalisée en utilisant les paramètres du programme HBPLUS (annexe B.3) [282, 283]. Il a été montré, par l'étude de complexes protéiques, que le nombre de liaisons hydrogène dépend de l'aire de l'interface. Nous avons donc calculé, pour chaque complexe, le rapport de nombre de groupements du ligand impliqués comme donneur ou comme accepteur dans une liaison hydrogène sur l'aire de l'interface avec la cible. Les résultats montrent, pour

1J75mt et 1H50mt, des résultats très proches de la valeur calculée pour la BgK (1,3 liaison / 100 Å²). Les densités superficielles de liaison H sont plus faibles pour 1HZ6mt (0,9 liaison / 100 Å²) et pour 1CKKmt (0,5 liaison / 100 Å²). Ces dernières valeurs, bien que plus faibles, restent toutefois dans la gamme des valeurs observées expérimentalement grâce à l'étude de complexes protéine-protéine déterminées par diffraction X [52, 139, 159].

Aspects électrostatiques : le potentiel électrostatique

Les interactions électrostatiques constituent le second type d'interaction ayant un rôle prépondérant dans la reconnaissance et l'association de protéines entre-elles. Leur analyse a été réalisée en utilisant, pour les résidus chargés, le pKa de ces acides aminés à l'état libres et à pH = 7. De cette manière, la charge nette globale des 5 plateformes varie de -1 (1HZ6mt) à +8 (1H50mt), mais s'avère proche pour les protéines 1J75mt, 1BGK et 1CKKmt (+5, +6 et +5 respectivement). Or, comme nous l'avons vu précédemment, dans le cas des protéases à serine (chapitre IV), divers ligands, dont la charge nette est variable, sont capables d'interagir avec une même cible, au même site d'interaction. Cependant, ce même travail a également montré que les potentiels électrostatiques, dans la région de l'interface, peuvent être corrélés, malgré des charges nettes globales très différentes. D'autre part, le degré de similarité du potentiel électrostatique est plus élevé, en considérant la région de l'interface, plutôt que l'ensemble de l'espace. Le potentiel électrostatique de ces 5 protéines a été calculé par la résolution de l'équation de Poisson Boltzmann en utilisant leurs structures 3D respectives, obtenues à l'issue de la phase d'affinement (annexe D.4). La figure VI-21 montre les valeurs de potentiel obtenues, représentées à la surface de ces 5 structures. La région d'interaction avec le canal Kv1.2 est indiquée à l'aide d'un cercle rouge.

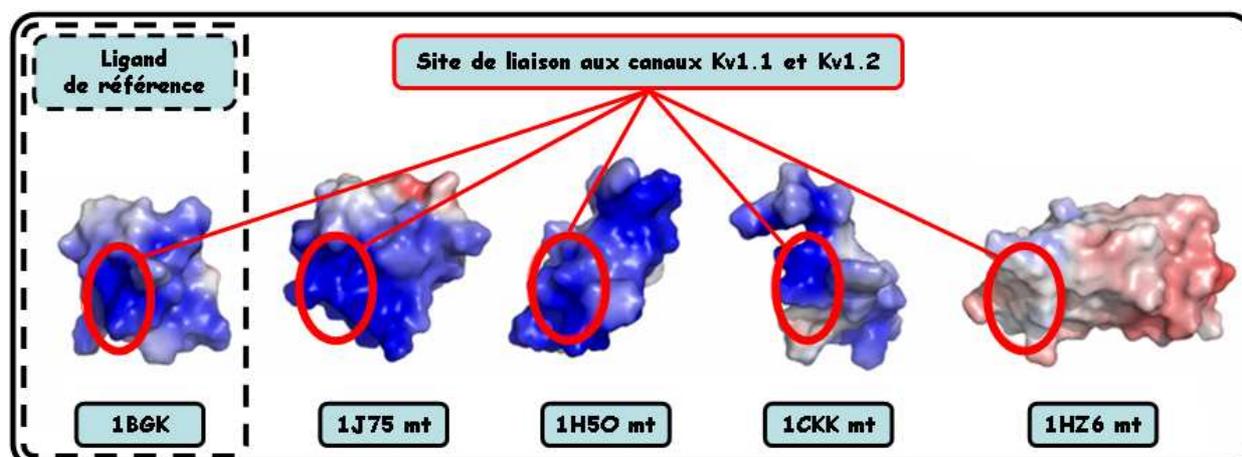


Figure VI-21 : Comparaison des potentiels électrostatiques des différentes plateformes produites, calculés par la résolution de l'équation de Poisson-Boltzmann. Les potentiels sont représentés sur la surface accessible au solvant de la molécule (rouge (-5k_BT/e), bleu (+5k_BT/e)).

Cette figure montre que les plateformes 1J75mt, 1H5Omt et 1CKKmt présentent, à leur surface, dans la région où se situe le motif FSKY, des potentiels électrostatiques similaires à celui calculé pour la BgK dans la même région. Dans le cas de 1HZ6mt, le potentiel dans la région du motif est également positif mais d'amplitude plus faible. Ces observations peuvent être comparées aux indices de similarités calculés au chapitre V pour ces mêmes plateformes, ces derniers étant significativement plus élevés pour 1J75mt, 1CKKmt et 1H5Omt (respectivement 0,72, 0,89 et 0,94) que pour 1HZ6mt (0,31). Il faut cependant noter que, dans le cas des complexes de serine protéases (chapitre IV), les valeurs d'indices de similarité de potentiel électrostatique, dans la région d'interaction, entre certains ligands de la même cible, étaient inférieures à 0,4.

Interactions électrostatiques microscopiques

L'analyse des potentiels, exposée précédemment, ne permet toutefois pas d'appréhender les aspects microscopiques des interactions électrostatiques. Nous avons donc analysé les modèles des complexes des différentes plateformes afin de localiser, à l'interface des deux partenaires protéiques, les groupements possédant des charges nettes positives ou négatives. La première conséquence de la présence de tels groupements à l'interface est l'existence possible d'interactions électrostatiques défavorables à la liaison dues à la présence des groupements portant des charges de même signe et proche dans l'espace. De telles répulsions électrostatiques peuvent être mises en évidence en identifiant les paires de groupements chargés de signe opposé situé de part et d'autre de l'interface à une distance inférieure à 6 Å. Dans les 5 modèles, aucune répulsion de ce type n'a été mise en évidence.

Les groupements chargés peuvent également être responsables de la formation de ponts salins qui constituent un cas particulier d'interaction électrostatique entre deux groupements possédant une charge nette de signe opposé. Le nombre de groupements, impliqués dans des ponts salins à l'interface, entre le canal Kv1.2 et les différents ligands, est compris entre 3 (pour la BgK) et 5 (pour 1J75mt). Certes, aucune contrainte explicite n'a été appliquée au cours de la recherche *in silico* afin d'identifier les plateformes possédant des groupements susceptibles de former des ponts salins. Cependant, la présence de ces groupements pourrait permettre la sélection de plateformes sur la base des critères électrostatiques. En effet, si ces derniers sont déterminés par l'ensemble de la distribution de charge de la protéine considérée, les résidus portant une charge nette, présents à l'interface contribuent de façon importante au potentiel électrostatique, au niveau de la zone d'interaction. De plus, l'analyse détaillée des interfaces montre que les groupements des ligands impliqués dans les ponts salins avec le canal Kv1.2

portent des charges positives pour 1BGK, 1CKKmt et 1H5Omt. De même, à l'interface du complexe 1J75mt-Kv1.2, trois groupements, portant des charges positives impliquées dans des ponts salins, mais également deux groupements, portants de charges négatives, ont été identifiés. La situation est différente pour 1HZ6mt, puisqu'un seul groupement de l'interface portant une charge positive est impliqué dans un pont salin avec le canal, trois autres ponts salins impliquant des résidus de la plateforme portant des charges négatives. L'analyse de la zone d'interaction de la cible, en termes de résidus chargés, permettrait lors de la recherche, d'introduire des contraintes supplémentaires pour la sélection des plateformes.

Nous avons également analysé les 5 modèles de complexes afin de déterminer le nombre total de groupements, portant une charge nette, positive ou négative, présents à l'interface. Ce nombre est très variable : 5 groupements chargés pour les modèles 1BGK et 1CKKmt, 7 pour les modèles 1J75mt et 1H5Omt et 11 pour le modèle 1HZ6mt. De ce fait, il existe à l'interface avec Kv1.2, des groupements possédant des charges nettes qui ne sont pas stabilisées par l'établissement d'un pont salin (nombre de charges nettes à l'interface - nombre de ponts salins). L'analyse des complexes de protéases à sérine réalisée dans le cadre de ce travail (chapitre IV) a également montré la présence, en moyenne, de 3 ± 2 charges non compensées à l'interface (dans 3 cas sur 27 les valeurs sont néanmoins supérieures à cette dernière). Dans le cas des modèles des plateformes produites au cours de ce travail : les structures de 1BGK, 1J75mt et 1CKKmt montrent 2 charges nettes non compensées, 1H5Omt, 4, et 1HZ6mt se distinguent par un nombre plus important, de 7 charges nettes non compensées. Cette observation pourrait être, à l'origine, de l'absence de liaison au canal Kv1.2 observée pour cette dernière plateforme. En effet, lors du processus d'association, les charges présentes, dans la zone d'interaction, sont désolvatées ce qui s'accompagne d'une contribution défavorable à la variation d'énergie libre de liaison. Cette contribution est compensée, dans le cas où la charge nette considérée est impliquée dans une interaction forte (type pont salin), avec la molécule liée. En l'absence de compensation, la présence de charges non compensées à l'interface constitue donc un facteur défavorable à la liaison [284, 285]. Dès lors, la conséquence du positionnement de la plateforme 1HZ6mt par rapport au canal Kv1.2 aurait donc pour conséquence l'enfouissement de 7 charges nettes incapables de compenser la contribution électrostatique de désolvatation par une interaction forte. Ces constatations expliquent, probablement, les résultats obtenus lors des tests de liaison par compétition, à savoir que la plateforme 1HZ6mt ne déplace pas l'analogue de la BgK.

CONCLUSION ET PERSPECTIVES

*« Il ne suffit pas de dire : je me suis trompé ; il faut dire comment on s'est trompé »
(C. Bernard)*

Atouts et limitations de la méthode

Identification du Motif

La méthode de conception de ligands protéiques développée au cours de ce travail repose sur le transfert d'un motif de résidus, jouant un rôle prépondérant dans une interaction donnée (ligand-cible), sur une plateforme capable de reproduire la topologie fonctionnelle de ce motif. L'identification du motif repose sur un travail « d'alanine scanning » réalisé antérieurement sur la BgK. Ce travail a permis d'identifier plusieurs résidus dont le remplacement, par un résidu alanine, entraîne une augmentation significative de la variation d'énergie libre standard associée à la liaison aux canaux Kv1. Formellement, sur le plan thermodynamique, cette variation résulte d'une déstabilisation de l'ensemble des conformations accessibles au système à l'état lié par rapport à l'état libre. Les résultats obtenus par cette approche sont généralement utilisés en référence à un complexe, entre les deux partenaires, comme l'indication que les résidus concernés du ligand établissent des interactions spécifiques avec la cible, et qui se trouveraient rompues lors du remplacement du résidu considéré en alanine. Cependant, la source des variations observées dans l'énergie de liaison des variants par rapport à la protéine native est complexe et l'interprétation indiquée précédemment n'est pas univoque : les variations observées peuvent résulter de plusieurs phénomènes. Les phénomènes impliquant directement le résidu substitué sont la diminution des interactions de Van der Waals, la perte d'une interaction électrostatique (dans le cas où celle-ci apporterait une contribution favorable à la liaison comparée au bilan de solvation des charges), la suppression d'une liaison hydrogène ou l'exposition au solvant de surface non polaire. Toutefois, la variation observée peut également résulter d'autres phénomènes comme un changement conformationnel local, une déstabilisation de la structure, une augmentation du désordre du ligand à l'état lié, de l'agrégation ou d'un dépliement plus global. Beaucoup de ces effets sont très difficiles à mettre en évidence expérimentalement [286]. Il apparaît donc que l'interprétation des résultats issus de « l'alanine scanning » est délicate, particulièrement concernant leur aspect quantitatif. Ce point doit être pris en considération dans les approches de conception de ligand protéiques par transfert de motif, la sélection du motif à transférer devant se faire sur la base de données expérimentales et/ou structurales complémentaires.

Dans le cas étudié, la participation des résidus K25 et Y26 de la BgK et leur rôle dans l'interaction avec les canaux Kv1 ont été mis en évidence par des travaux « d'alanine scanning ».

Cette observation a été corroborée par l'identification de résidus équivalents à la paires K25, Y26 dans plusieurs autres toxines d'anémone de mer mais également de scorpion possédant des repliements différents de celui adopté par la BgK et capable de lier ces mêmes canaux [200, 216]. Le mode d'interaction de ces toxines avec les canaux Kv1 a été déduit d'études de modélisation moléculaire utilisant des données expérimentales [217, 218] et a pu être caractérisé expérimentalement dans le cas de l'interaction de la charybdotoxine avec le canal potassique KcsA, modifié par transfert de résidus des canaux Kv humains, impliqués dans l'interaction avec les toxines [226]. Ces résultats confortent l'interprétation des résultats « d'alanine scanning » de la BgK et montrent le rôle de ces résidus dans l'interaction avec les canaux Kv1, en particulier l'interaction du groupement ammonium du résidu lysine de la diade avec le pore du canal, permettant d'établir des interactions avec plusieurs groupements carbonyles. Cependant, comme nous l'avons déjà souligné, les aspects quantitatifs des expériences « d'alanine scanning » peuvent difficilement être interprétés en termes d'interaction localisée. Ainsi, la substitution en alanine des résidus transférés sur la nouvelle plateforme peuvent ne pas reproduire quantitativement les observations faites sur la plateforme de référence. Ceci est effectivement observé sur les différentes plateformes étudiées, ou les variations d'affinité entraînées par le remplacement en alanine de certains résidus du motif sont significativement plus faibles que celles observées sur la plateforme de référence. L'application de notre approche à d'autres interactions pourra permettre de documenter cette question afin d'étudier la validité et les limites de la méthode « d'alanine scanning » pour l'étude des interactions protéine-protéine.

Diversité de Solutions

La diversité des solutions obtenues au cours de ce travail est une caractéristique fondamentale de cette approche de conception de ligands. L'incertitude, propre aux outils et aux méthodes utilisés dans les différentes étapes, entraîne un inévitable taux de résultats incorrects (des plateformes se révélant incapables de lier la cible considérée ou possédant des affinités trop faibles). Ceci est particulièrement vrai pour les méthodes fondées sur la structure, qui utilisent les outils et concepts de la modélisation moléculaire. Cette remarque n'est pas restreinte à la méthode développée au cours de ce travail, mais est générale aux approches fondées sur l'analyse des structures tridimensionnelles et leur utilisation *via* des champs de force empiriques qui permettent difficilement d'estimer les grandeurs thermodynamiques pertinentes pour évaluer des différences de stabilité. C'est d'ailleurs la raison pour laquelle nous avons utilisé, dans le cours de ce travail, des méthodes de modélisation moléculaires sous contraintes positionnelles dans le seul objectif de mettre en évidence les conséquences des modes de liaison proposés sur les

caractéristiques globales des interfaces et les interactions microscopiques entre les deux partenaires. Par ailleurs, d'autres étapes nécessaires à la production effective des ligands, conçus *in silico*, peuvent se révéler difficiles à franchir. Parmi celles-ci, figure la synthèse chimique des composés, dont les rendements de synthèse ne sont pas systématiquement satisfaisants, les étapes de purification, de repliement, et de formation des ponts disulfures : ces étapes différentes peuvent constituer autant de sources d'échecs interdisant la production le ligand protéique, empêchant ainsi la caractérisation de l'interaction. Enfin, les conséquences du transfert du motif sur le processus de repliement et la structure tridimensionnelle adoptée par la plateforme ainsi modifiée ne peuvent être prédites avec certitude.

L'ensemble de ces considérations nous ont amené à prendre en charge les incertitudes et le taux potentiel de solutions ne conduisant pas à la production de molécules capables de lier la cible considérée, en inscrivant la diversité des solutions comme un objectif de la méthode à développer. Pour cette raison, la recherche de plateforme a été menée par une analyse topologique exhaustive de toutes les structures contenues dans la base de données (PDB), permettant d'exploiter, de façon systématique, son répertoire structural. Notre méthode tirera dans l'avenir profit de l'accroissement constant du nombre de structures dans cette base de données et de la variété des topologies qui y sont décrites. Pour indication, le nombre de structures contenues dans la PDB était approximativement de 19000 structures au début de ce travail et dépasse désormais les 40000. Cette recherche systématique a nécessité le développement d'un outil bioinformatique adapté (programme STAMP). Cet outil permet l'examen topologique de plusieurs dizaines de milliers de structures de protéines à l'aide d'une méthode suffisamment rapide pour permettre de parcourir la base de données dans son intégralité. L'objectif de la méthode est de produire une diversité contrôlée de solutions identifiées (afin de produire ces protéines expérimentalement) par application de critère raisonnable mais robuste sans surestimation de la performance des méthodes *in silico*. L'approche est donc globalement récurrente, de manière à contrôler le flux de sortie (entrée en production) et intégrant implicitement le rendement global propre des phases expérimentales (synthèse, purification, repliement, caractérisation). A la fin de chaque cycle, les contraintes et critères de tolérance (nombre de résidus pris en compte dans le motif, tolérance topologique, taille des structures....) peuvent être modifiées afin d'obtenir une nouvelle liste de solutions (figure CP-1). La phase *in silico* de la méthode développée constitue le premier volet d'une approche globale aboutissant à la conception de ligands protéiques. L'application de cette approche au développement de ligands des canaux Kv1.2 a permis d'obtenir plusieurs ligands protéiques capables d'interagir au même

site de liaison que le ligand de référence. Seules les plateformes 1J75mt, 1H50mt, 1CKKmt et 1H50mt ont été produites et leur capacité à lier le canal, Kv1.2 évaluées dans le cadre de ce travail. Plusieurs autres plateformes obtenues par cette approche *in silico* restent donc à évaluer expérimentalement.

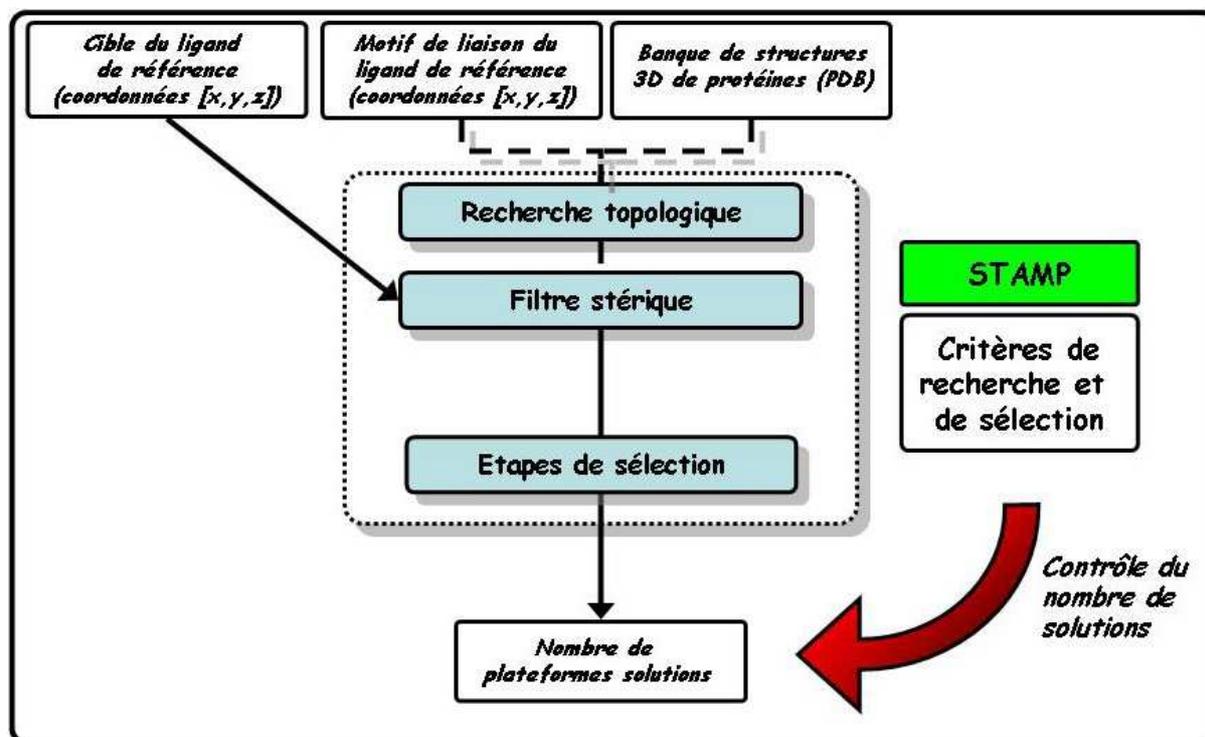


Figure CP-1 : Schéma de l'ensemble de la méthode développée. La modification des critères de recherche et de sélection permet de moduler le nombre de solution final. Ainsi est-il possible de contrôler la diversité de solutions recherchées.

Ainsi, 7 autres plateformes sont aptes à reproduire la topologie du motif de liaison de la BgK possédant des indices de similarités des potentiels électrostatique (SI) supérieurs à 0.7 : 1BNBmt, 1UXDmt, 1OQEmt, 1BOEmt, 1JAEmt, 1A7Fmt, 2PDDmt et 1ERFmt (tableau V-19, chapitre V). Ceci montre que, malgré l'ensemble des critères appliqués au cours de la recherche (présence d'un motif topologiquement équivalent au motif FSKY, taille de la plateforme inférieure à 70 résidus, mode de liaison n'impliquant pas de recouvrement stérique entre la plateforme et la cible, potentiel électrostatique similaire à celui du ligand de référence dans la région d'interaction), nous avons abouti à une diversité de solutions importante. Il convient de noter que la plupart des approches mises en œuvre pour la conception de ligand par transfert de « hotspots » n'utilisaient pas de recherche exhaustive mais exploitaient les propriétés d'une plateforme particulière (chapitre I, partie III). La sélection de la plateforme est souvent réalisée de manière empirique et motivée par la présence d'un élément de structure secondaire similaire à celui portant les éléments du motif à transférer [111-125]. Ces résultats ont montré qu'il était possible, après optimisation, de concevoir des ligands de haute affinité en transférant

un motif de liaison sur une plateforme protéique capable de restituer la topologie fonctionnelle de ce motif. Ces motifs peuvent adopter des topologies extrêmement diverses. La recherche exhaustive dans la PDB de plateformes possédant la topologie requise grâce aux outils bioinformatiques développés au laboratoire rendra la méthode de développement de ligands par transfert de « hotspots » applicable à un plus grand nombre de systèmes.

Il faut noter que la diversité des solutions identifiées par notre approche *in silico* dépend de nombreux paramètres. Nous avons déjà évoqué l'influence des valeurs maximales de paramètres caractérisant la similarité topologique entre le motif présent sur le ligand de référence et le motif identifié. Toutefois, la diversité des solutions dépend également étroitement du nombre de résidus composant le motif. Cependant, ce paramètre n'est pas indépendant de la topologie particulière dans laquelle se trouve le motif, ni de son caractère discontinu dans la séquence primaire. En effet, si le motif recherché est constitué de résidus adjacents décrivant une portion d'hélice alpha, le nombre de solutions identifiées (sans tenir compte d'éventuelles incompatibilités stériques avec la cible) sera très important. La classification des repliements des protéines de la PDB [287] indique que, près de 70% des protéines contenues dans la base de donnée de structure que nous analysons, possède des éléments en hélice α (structures contenant principalement des hélices, ou structures mixtes, hélices et feuillets). Toutefois, si le motif recherché ne correspond pas à une topologie fréquemment rencontrée dans les structures de la PDB, alors le nombre de solution retournée sera faible. Le cas des motifs discontinus en séquence primaire, peut également aboutir à un nombre plus réduit de solutions topologiques convenables, bien que celui-ci dépend également de la topologie particulière adoptée par les résidus du motif. Ainsi, les différents paramètres influant sur la diversité de solutions obtenues, apparaissent interdépendants. Un des objectifs de la poursuite du développement de cette approche sera donc d'étudier ces dépendances, *in silico*, afin d'appréhender au mieux les limites de la méthode.

Structure secondaire et transfert de motifs discontinus

Le deuxième aspect contribuant à l'obtention d'une diversité de composés est lié à l'absence de contraintes sur la structure secondaire de la plateforme recherchée. En effet, comme cela a été montré dans le chapitre V (partie III), il est possible de reproduire, de manière satisfaisante, la topologie du motif de liaison FSKY de la BgK avec des structures secondaires qui sont sans relation avec celles de la structure du ligand de référence. Dans cette recherche, les résidus SKY se trouvent dans l'extrémité N-terminale d'une hélice alpha, comme c'est le cas dans le ligand de référence, dans seulement 30 % des cas (6 plateformes sur 20). Les 14 autres

plateformes n'ont donc pas été identifiées par la reproduction d'une structure secondaire, similaire à celle portant les résidus du motif de référence. L'absence de contrainte sur la structure secondaire participe donc à la diversité des solutions identifiées. Par ailleurs, notre approche permet d'envisager le transfert de motifs discontinus. En effet, les résidus SKY sont présents dans l'extrémité N-terminale de la seconde hélice de la BgK alors que le résidu F du motif se trouve 16 résidus en amont dans la structure primaire. Notre approche permet, sans aucune contrainte sur la structure secondaire et la structure primaire d'identifier des plateformes reproduisant la topologie du motif.

Un exemple de l'avantage que peut présenter notre approche est donné par le cas de l'interaction du CD4 avec la protéine d'enveloppe gp120 du virus HIV-1. La résolution de la structure d'un complexe gp120/CD4/anticorps monoclonal par radiocristallographie [127] a permis d'élucider les détails structuraux de cette interaction. Cette structure a révélé que le site de liaison du CD4 sur la gp120 était constitué d'une large dépression de 800 Å² et mis en évidence le rôle prépondérant dans l'interaction, d'une boucle appelée CDR2-like constituée des résidus 36 à 47 du CD4. L'examen de cette structure a également révélé le rôle du résidu Arg59 du CD4 qui établit plusieurs interactions avec des résidus de la gp120. L'importance de ces résidus dans la liaison à la gp120 a été démontrée par des études de mutagenèse [288, 289, 290, 291, 292, 293]. Le ligand réalisé pour inhiber la liaison CD4/gp120 a été construit en transférant certains des résidus présents sur la face exposée du feuillet beta de la séquence 36 à 47 du CD4 sur la scyllatoxine, une petite toxine animale de 31 résidus. Le résidu R59 du CD4 n'a pas pu être transféré dans cette construction. Bien que ces travaux aient permis d'obtenir, après optimisation, un inhibiteur puissant de l'interaction gp120/CD4, notre approche permettrait de prendre en compte explicitement le résidu 59 du CD4 dans la recherche de plateforme et pourrait conduire à de nouveaux inhibiteurs de l'interaction CD4/gp120.

Considérations énergétiques

La mise en application de notre approche a abouti à la conception de plusieurs ligands protéiques, possédant des affinités comprises entre 0,5 et 1,6 μM pour le canal Kv1.2. L'affinité la plus importante s'avère donc de l'ordre de 10⁴ plus faible que celle de la Bgk. Cette valeur appelle plusieurs remarques. La première remarque, concerne le fait que notre méthode comporte des incertitudes sur le plan de la reproduction des interactions réellement mises en jeu entre le ligand de référence et la cible. En l'absence de structure du complexe entre les ligands conçus et

le canal Kv1.2, nous ne pouvons pas vérifier la topologie d'interaction exacte des différents ligands avec la cible. Cette information serait particulièrement utile à l'interprétation des observations faites sur la plateforme 1J75mt où nous avons pu mettre en évidence que, dans la structure obtenue par radiocristallographie, le résidu topologiquement équivalent à F6 dans le motif de la BgK adoptait une conformation différente de celle observée dans cette structure à l'état libre et dans le modèle de complexe BgK/Kv1.2. Par ailleurs, en l'absence de données thermodynamiques sur l'interaction de référence et sur l'interaction des différents ligands conçus, il n'est pas possible de connaître l'origine des variations d'énergie libre associées en terme de contribution relative enthalpie/entropie. Ces informations sont l'étape initiale, vers la compréhension des mécanismes à l'œuvre, lors du processus d'association et permettraient d'avancer dans la comparaison quantitative de liaison des ligands conçus dans le cadre de ce travail.

Il faut cependant noter, que la différence d'affinité observée entre le ligand de référence et les trois ligands conçus, ne représente qu'une augmentation de 23 kJ.mol^{-1} de la variation d'énergie libre standard du système par rapport à l'interaction de référence. De nombreux travaux ont été consacrés à l'estimation du coup énergétique des interactions ponctuelles. En particulier, l'énergie libre de stabilisation correspondant à la formation d'une liaison hydrogène est de $-2,09$ à $20,9 \text{ kJ.mol}^{-1}$ ($-0,5$ à $-5,0 \text{ kcal.mol}^{-1}$) [294]. L'incidence de la formation d'une seule liaison hydrogène sur les constantes d'association peut donc être voisine d'un facteur 5000. La différence observée pour les constantes de dissociation des ligands conçus par notre approche par rapport à l'interaction du ligand de référence avec les canaux Kv1.2, pourrait donc être réduite grâce à un travail d'optimisation permettant d'établir un nombre limité d'interactions ponctuelles.

Perspectives : vers une optimisation des ligands...

L'optimisation des plateformes peut être réalisée en utilisant les observations faites sur les modèles moléculaires réalisés au cours de cette étude, afin de mieux caractériser les conséquences du mode de liaison de ces ligands reproduisant l'interaction du motif transféré avec le canal Kv1.2. Ces modèles nous ont permis d'analyser les propriétés des interfaces de chacun des ligands avec la cible et les interactions microscopiques entre les deux partenaires. Pour les quatre modèles, les paramètres caractérisant la taille des interfaces, la complémentarité des surfaces et leur composition, la densité superficielle de liaison hydrogène, le nombre de charges nettes des interfaces et le nombre de pont salins intermoléculaires, ont

été calculés. Bien que des différences dans les valeurs de ces paramètres soient observées entre les quatre plateformes produites et le complexe de référence BgK/Kv1.2, ces valeurs se situent, néanmoins, dans les limites observées dans l'analyse de complexes issus d'études de radiocristallographie (chapitre IV) [52, 131, 139]. En particulier, un nombre important de liaisons hydrogène et de ponts salins est observé dans la plupart des complexes, alors qu'aucune contrainte spécifique n'a été imposée lors du processus de sélection pour reproduire ces interactions. Ceci suggère, du moins pour partie, que les critères globaux appliqués, ont des conséquences en termes d'interactions microscopiques. L'application de critère de similarité des potentiels électrostatiques a permis, dans certains cas, de sélectionner des plateformes portant des charges nettes topologiquement équivalentes à celles observées dans le complexe de référence. La comparaison des résidus du canal Kv1.2 impliqués dans ces interactions microscopiques, illustre les points convergents et divergents pour les différents complexes. Le tableau CP-2 reporte les résidus du canal Kv1.2 impliqués dans l'interaction avec le ligand dans la formation de ponts salins et/ou de liaisons hydrogène pour les modèles des quatre plateformes produites et pour le complexe de référence Kv1.2/BgK.

Kv1.2	E353 (C)	D355 (A)	D355 (B)	D355 (C)	D355 (D)	R354 (B)	R354 (C)	R354 (D)
1BGK			SB	SB		SC		
1J75mt	SB		SB	SB	SB/SC	SB/SC	SC(2)	
1H50mt		SB		SB				
1CKKmt				SB/SC/MC				
1HZ6mt			SB	SC	SC	SB/SC	SB	SB/SC
Kv1.2	S356 (B)	Q357 (A)	Q357 (B)	Q357 (C)	F358 (A)	D363 (A)	Y377 (A)	Y377 (B)
1BGK	MC	SC(2)	MC(2)/SC				MC	MC
1J75mt		SC		SC(2)				MC
1H50mt		MC+SC	SC		MC			MC
1CKKmt		SC	SC			SB		
1HZ6mt			SC				MC	MC
Kv1.2	Y377 (C)	G378 (A)	G378 (B)	D379 (A)	D379 (B)	D379 (C)	M380 (B)	
1BGK	MC	MC						
1J75mt			MC	MC		SC		
1H50mt		MC	MC	MC			MC	
1CKKmt				MC/SB	SB	SB		
1HZ6mt	MC			MC				

Tableau CP-2 : Représentation des interactions (liaisons hydrogène et ponts salins) réalisées entre les différentes plateformes et le canal Kv1.2 dans les modèles moléculaires. Les résidus du canal Kv1.2 impliqués sont indiqués, ainsi que leur numéro de chaîne respectif (présenté entre parenthèses).

Dans les quatre modèles, des résidus du canal Kv1.2 peuvent être impliqués dans des interactions similaires à celles mises en évidence dans le complexe de référence avec la BgK. Cependant, dans la plupart des cas, les interactions établies par les ligands, mettent en jeu des résidus différents

de la cible. Des observations identiques ont été réalisées par l'examen de complexes cristallographiques, reporté dans le chapitre IV. Il n'est donc pas nécessaire d'utiliser des contraintes explicites pour pouvoir reproduire le nombre d'interactions (liaisons hydrogène et ponts salins) comparables à ce qui est observé dans le complexe de référence lors du processus de sélection. La densité superficielle de groupements, susceptibles de participer à diverses liaisons à la surface des protéines, permet à la cible de trouver des partenaires pour établir ces interactions. Ceci est parfaitement illustré par l'observation que certains résidus du canal Kv1.2 sont capables d'établir des liaisons hydrogène *via* les groupements donneurs ou accepteurs présents sur les chaînes principales et ce, malgré les contraintes positionnelles appliquées sur celles-ci. D'autre part, la plasticité des protéines autorise des modifications conformationnelles des chaînes latérales permettant l'établissement d'interactions intermoléculaires qui doivent répondre à des critères géométriques stricts, en particulier concernant les liaisons hydrogène. D'après le tableau CP-2, de nombreuses chaînes latérales de résidus du canal Kv1.2 sont impliquées dans des interactions avec le ligand alors que ces mêmes chaînes ne le sont pas dans le modèle d'interaction avec la BgK.

Un aspect important de ces modèles et des interactions microscopiques entre les partenaires, concerne la présence de groupement du ligand possédant des charges nettes non compensées à l'interface avec la cible. En effet, comme nous l'avons montré au chapitre VI, dans les quatre modèles de complexes avec les ligands produits, nous avons pu identifier entre 2 (1J75mt, 1CKKmt) et 7 (1HZ6mt) charges nettes non compensées par la formation de ponts salins à l'interface avec la cible. Deux groupements portant des charges nettes non compensées ont été mis en évidence à l'interface BgK/Kv1.2. Au cours du processus d'association entre le ligand et la cible, les groupements portant des charges nettes présents à l'interface doivent être désolvatés. La contribution à la variation d'énergie libre du système liée à cette désolvatation est défavorable à la liaison. [284, 285]. Ceci laisse supposer, que la présence de ces charges nettes à l'interface, pourrait être à l'origine de l'absence de liaison observée pour 1HZ6mt (>100 μ M). Ainsi, les modèles d'interactions réalisés pourraient permettre une première phase d'optimisation par la substitution des résidus portant des charges nettes non appariées dans l'interaction avec le canal Kv1.2. Une étude théorique de l'interaction barnase-barstar a permis de montrer qu'il était possible d'optimiser la distribution de charge de la barstar pour rendre plus favorable la contribution de l'électrostatique à la liaison avec la barnase [295]. Un travail expérimental concernant la stabilisation du répresseur Arc a également été réalisé et dans ce cas, une diminution de 21,32 kJ.mol⁻¹ de la constante de dissociation. Le repliement de cette protéine a

été obtenu grâce à la substitution de résidus polaires ou chargé par des résidus non polaires [296]. Une telle approche permettrait d'optimiser nos plateformes afin d'augmenter leurs affinités respectives pour le canal Kv1.2. Parmi les plateformes identifiées, trois d'entre elles présentent des affinités micromolaires pour le canal Kv1.2 : 1CKK, 1H5O et 1J75. Les affinités observées sont significativement plus faibles que celle de la BgK pour le canal Kv1.2 (environ 50 pM). Toutefois, si une approche d'optimisation électrostatique comme celle réalisée dans le cas du répresseur Arc permettait d'obtenir une diminution de la variation d'énergie libre de liaison de $20,9 \text{ kJ.mol}^{-1}$, les ligands ainsi optimisés posséderaient une affinité sub-nanomolaire.

Des approches plus systématiques pour l'optimisation des ligands conçus peuvent être envisagées. En effet, la méthode de conception de ligands protéiques développée aboutit à identifier des plateformes protéiques capables de reproduire la topologie du motif de liaison transféré : F6/S23/K25/Y26. La présentation des chaînes latérales des résidus du motif sur la plateforme sélectionnée, doit permettre de reproduire les interactions ponctuelles observées pour ces résidus dans le cas du complexe de référence, BgK/Kv1.2. D'autre part, la présence du motif sur la plateforme protéique, doit guider l'interaction avec la cible, dans une orientation correcte. La procédure de recherche et de sélection *in silico* aboutit donc à des propositions de ligands, et conjointement, pour chacun d'eux, à un modèle décrivant l'interaction du ligand avec la cible. Ce modèle peut être exploité, afin d'identifier un certain nombre de positions à modifier. Il est également possible, d'envisager des approches combinatoires pour la réalisation de ces variants sur un nombre limité de positions choisies, en exploitant les modèles d'interaction issus de notre approche. Cette phase d'optimisation peut jouer un rôle très important dans l'obtention d'un ligand d'intérêt, comme cela a été le cas dans de nombreuses études. Ainsi, dans le cas du développement de mimes de CD4, les premiers variants possédaient des affinités très modestes pour la cible [125]. Cependant, les variants réalisés sur la base de l'analyse structurale ont permis d'obtenir des affinités importantes, de l'ordre du nanomolaire [126].

ANNEXES

A.1 Bioinformatique et Modélisation Moléculaire

Par le passé, l'évolution des différentes sciences s'est toujours faite de façon conjointe, l'amélioration de l'une favorisant l'amélioration des autres. Ceci est d'autant plus vrai pour des domaines interdisciplinaires tels que la bioinformatique et la modélisation moléculaire. Quelques dates ont marqué l'évolution de ces domaines, expliquant pourquoi le développement de ces méthodes est si récent ; leur utilisation s'est réellement développée depuis seulement une trentaine d'années, justifiant les efforts encore à faire dans ces domaines.

1926 : première cristallisation d'une protéine : l'uréase

1934 : premier pattern de diffraction du cristal de la pepsine (Bernal et Crowfoot 1934)

1951 : description des éléments de structure secondaire, hélice α et feuillet β .

1953 : détermination de la structure en double hélice de l'ADN par radiocristallographie par Watson et Crick, qui recevront le prix Nobel pour cette découverte.

1954 : développement de la méthode des « atomes lourds » pour la résolution des phases en radiocristallographie.

1955 : premier séquençage d'une protéine (l'insuline bovine) par Sanger, qui lui vaudra le prix Nobel

1958 : le ministère de la défense américaine crée ARPA (Advanced Research Project Agency), dont le but est de renforcer les développements scientifiques susceptibles d'être utilisés à des fins militaires.

1969 : l'ARPA crée l'ARPANET, ancêtre de l'internet.

1964 : le code génétique est déchiffré (Nirenberg et al. 1964)

1965 : première structure de protéine (lysozyme) résolue par radiocristallographie.

1972 : première protéine recombinante par P. Berg et son groupe ; mise au point d'un nouveau moyen de communication : l'email

1977 : description d'une base de données informatisée de structure de protéines : la PDB ; mis au point d'une méthode de séquençage de l'ADN par Maxam, Gilbert et Sanger ; Internet devient une réalité grâce au protocole de transfert de données TCP/IP

1980 : premier séquençage complet du génome d'un organisme (le bactériophage ϕ X174), qui vaudra à Sanger, Gilbert et Berg le prix Nobel ; description de la RMN multidimensionnelle pour la résolution de structures de protéines.

1981 : IBM introduit le PC (Personal Computer)

1985 : description de la PCR par Mullis et son groupe, il recevra le prix Nobel en 1993.

1986 : création de la SWISS-PROT, banque d'informations sur les protéines.

1988 : création du NCBI ; début du projet « Human Genome » ; l'algorithme FASTA pour les comparaisons de séquence est publié

1990 : le programme BLAST, pour l'alignement de séquence, est implémenté ; le protocole HTTP est mis au point ainsi que le langage HTML, posant les bases du www.

1991 : le CERN annonce la création du www (World Wide Web) ; création de Linux par Linus Thorvald.

1994 : la compagnie Netscape Communication est fondée et distribue la première version commerciale de navigateur sur internet.

1996 : création de Prosite, une base de données associant à la fois des considérations biologiques et des éléments de structure (séquences, domaines, etc...)

2001 : le génome humain est publié.

2002 : le CIB (Comité International de Bioéthique) interdit tout brevetage du génome humain

La bioinformatique s'appuie sur la gratuité des informations et des moyens d'étude, de façon à promouvoir d'autres découvertes scientifiques et ainsi participer à l'amélioration et au développement de nos sociétés.

A.2 La « Protein Data Bank » ou PDB

A.2.1 Historique

La PDB (ou « Protein Data Bank ») est la seule base de données mondiale de structures 3D de macromolécules biologiques, protéines et/ou acides nucléiques. Elle fut établie en 1971 au « Laboratoire National Brookhaven », et contenait initialement seulement 7 structures. Cette dernière fut officiellement présentée en 1977 [297]. En 1998, le « Research Collaboratory for Structural Bioinformatics », ou RCSB [298] devint responsable de la gestion de la PDB avec, pour but, de créer une ressource basée sur les technologies les plus modernes, facilitant l'utilisation et l'analyse des données structurales permettant ainsi de favoriser la recherche biologique. La RCSB PDB (*) fournit désormais à la fois un ensemble d'outils et de ressources permettant l'étude des structures des macromolécules biologiques et leurs relations avec la séquence, la fonction ou les aspects pathologiques. En 2003, la RCSB PDB intégrait le « WorldWide PDB », ou wwPDB [299], organisation constituée de trois membres : RCSB PDB pour les USA, MSD EBI pour l'Europe et PDBj pour le Japon, comme indiqué sur la figure A2-1. L'objectif de la wwPDB est le maintien d'une archive PDB, unique et uniforme, pour les informations structurales des macromolécules, gratuite et officiellement disponible à l'ensemble de la communauté [300].

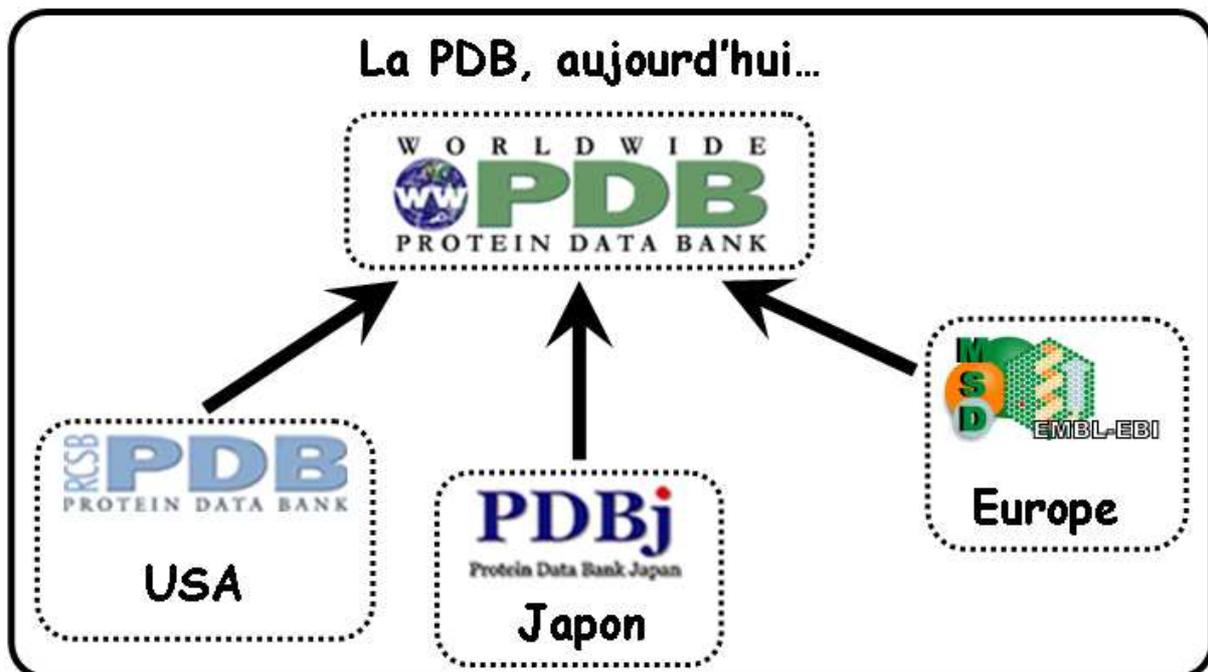


Figure A2-1 : L'organigramme de l'organisation wwPDB.

* <http://www.pdb.org/>

A.2.2 Évolution et perspectives

L'utilisation de la PDB dans les différents projets de recherche en biologie est désormais quasi incontournable, grâce notamment à l'évolution des techniques de résolution de structure des macromolécules biologiques comme la RMN et la radiocristallographie (représentant 99,5% des structures de la PDB), expliquant l'essor de domaines tels que la biologie structurale ou la génomique structurale. Ces récentes évolutions permettent d'une part, la résolution d'un très grand nombre de structures et d'autre part, la résolution de structures de plus en plus complexes (protéines de grandes tailles, complexes protéiques, complexes protéine-acides nucléiques...), laissant présumer, à l'avenir, de l'utilisation croissante de la PDB. La croissance de la PDB, en nombre de fichiers disponibles, est quasi exponentielle, comme en témoigne la figure A2-2.

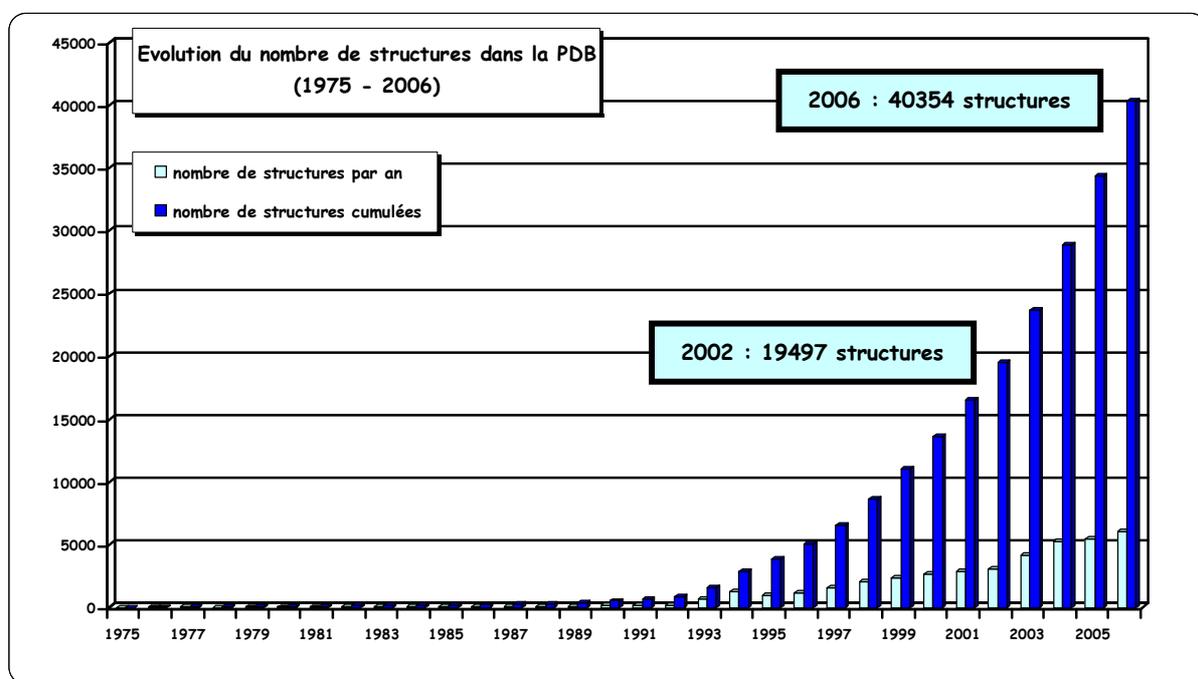


Figure A2-2 : Évolution du nombre de structures résolues de la PDB entre 1975 et 2006. Entre 2002 (début du projet de thèse) et 2006 (fin du projet), le nombre de structures de la PDB a presque doublé.

Ceci explique le besoin de développer des outils bioinformatiques, de plus en plus puissants, pour le traitement et l'utilisation des informations disponibles.

A.3 L'équation de Poisson-Boltzmann

A.3.1 Établissement de l'équation de Poisson-Boltzmann

Afin d'interpréter les interactions entre des molécules chargées (les protéines dans notre cas), il est nécessaire de connaître le potentiel autour de chacune d'elles ; pour ce faire, nous devons tenir compte des protéines et du solvant (notamment des ions en solution) [301]. Dans le cadre de cette thèse, le potentiel électrostatique de différentes protéines a été calculé par la résolution de l'équation de Poisson-Boltzmann. Cette dernière est établie comme suit.

A.3.2-A La loi de Poisson

La loi de Poisson permet de calculer le potentiel électrostatique d'une distribution complexe de charges : les charges ponctuelles sont remplacées par une densité de charges, ρ , dans un milieu dont la constante diélectrique, ϵ , est considérée comme invariante dans l'espace (ϵ_0 étant la permittivité diélectrique du vide, égale à $10^{-9}/36\pi$). La notion de densité de charges découle du théorème de Gauss. A chaque élément dS d'une surface, on associe un vecteur unitaire u_n dirigé suivant la normale à la surface, de l'intérieur vers l'extérieur. Au champ électrique E produit par les charges à l'intérieur de la surface, on associe une autre grandeur ou flux électrique, Φ , défini par :

$$\Phi = \int \mathbf{E} \cdot \mathbf{u}_n \, dS \text{ encore égal à } \rho / \epsilon_0 \epsilon$$

Le théorème de Gauss, appliqué à un élément de volume infiniment petit, permet de définir la loi de Poisson (forme différentielle du théorème) reliant le potentiel électrostatique, Ψ , à la densité de charges, ρ , les autres paramètres étant constants :

$$\Delta\Psi = -\rho / \epsilon_0 \epsilon$$

De nombreuses solutions mathématiques, complexes, permettent de résoudre cette équation ; néanmoins les progrès informatiques permettent de fournir des solutions numériques à celle-ci.

A.3.2-B La théorie de Debye-Hückel

La théorie de Debye-Hückel permet de prendre en compte la distribution des ions dans un milieu mais repose néanmoins sur plusieurs hypothèses : autour d'un ion donné j , la distribution des charges est supposée de symétrie sphérique, la solution doit respecter l'électroneutralité, et la densité de charges varie selon une statistique de Boltzmann. La densité électronique créée par les ions en solution prend alors la forme suivante :

$$\rho(\text{ion}) = \sum e z_i n_i^0 \exp(-z_i e \Psi(r) / kT)$$

Pour des concentrations en ions faibles, $z_i e \Psi(r) / kT \ll 1$, l'équation se réduit à :

$$\rho(\text{ion}) = K^2 \Psi \text{ avec } K^2 = e^2 \sum z_i n_i^0 / \epsilon_0 \epsilon kT$$

A.3.2-C La loi de Poisson Boltzmann

En combinant les lois de l'électrostatique (loi de Poisson), pour une distribution rigide de charges (approximation des corps rigides), et une description statistique d'un milieu liquide renfermant des entités chargées (théorie de Debye-Hückel), on obtient l'équation complète de la loi de Poisson-Boltzmann, sous forme réduite ou générale selon la concentration en sels de la solution :

$$\epsilon_0 \epsilon \nabla^2 (\Psi(r)) = -\rho + K^2 \Psi \text{ (forme réduite)}$$

$$\epsilon_0 \epsilon \nabla^2 (\Psi(r)) = -\rho + \sum q_i c_i \exp(e\Psi(r) / kT) \text{ (forme générale)}$$

Dans ce modèle, le solvant est considéré comme un continuum, représenté par la valeur de la constante diélectrique du milieu considéré (intérieur des protéines, solvant).

A.3.3 Résolution de l'équation de Poisson-Boltzmann

Les potentiels électrostatiques reportés dans ce manuscrit de thèse ont été calculés par résolution de l'équation Poisson-Boltzmann à l'aide du programme APBS [203]. Dans le chapitre IV, les calculs ont été effectués à partir des structures des ligands, après superposition (dans un même système) des atomes Ca de leurs cibles. Dans le chapitre VI, les structures des plateformes ont été superposées à la BgK, par l'intermédiaire des atomes Ca et C β des résidus du motif F6/S23/K25/Y26. Les charges et rayons atomiques utilisés pour le calcul sont les paramètres PARSE [302]. La taille des grilles et l'espacement des points de calcul ont été choisis de façon à englober l'intégralité des structures protéiques : les axes x, y et z contenant chacun 97 points, avec un espacement de 2,0 Å (pour la première grille) et fonction de la taille des protéines pour la grille de focalisation (de 0,8 à 1,2 Å). Le calcul des grilles a été réalisé par la méthode de focalisation, par résolution de la forme non-linéaire de l'équation de Poisson-Boltzmann dans les conditions suivantes : force ionique 150 mM (ions mono-chargés, +1 et -1, à une concentration de 0,075 mM, et de rayon atomique 2,0 Å), rayon des molécules de solvant 1,4 Å, constantes diélectriques du solvant et des protéines de 1 et 80 respectivement. Le calcul des grilles finales a été effectué en utilisant les conditions aux limites multiples de Debye-Hückel, déterminées à partir des premières grilles calculées. Les grilles de potentiel (au format UHBD [282]) ainsi obtenues ont pu être comparées par l'utilisation du programme PIPSA, version 2.0 [187]. Le paramètre « skin », correspondant à l'épaisseur des régions des grilles analysées a été fixé à 4 Å. Cette région a été définie à une distance de 3 Å des surfaces moléculaires, correspondant au paramètre « probe ». La similarité des potentiels électrostatiques et des

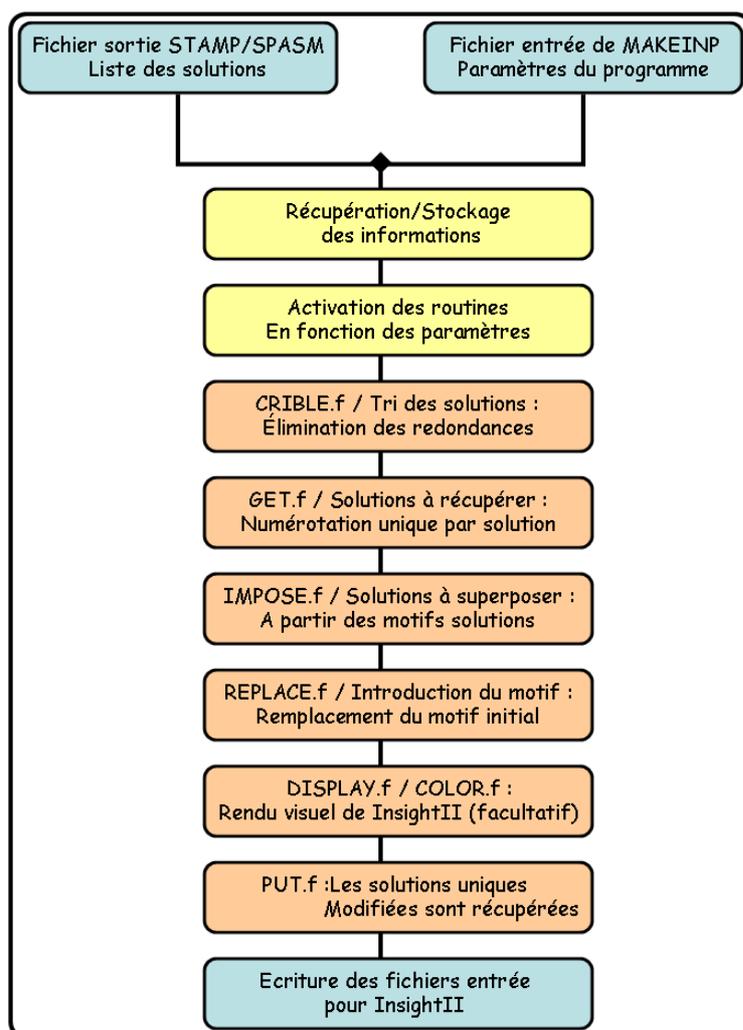
formes des protéines comparées a été calculé à l'intersection de leurs régions « skin » respectives par un indice de Hodgkin [204].

A.4 Les programmes de sélection de plateformes

La sélection des plateformes a été effectuée *in-silico*, à partir de programmes développés au laboratoire. Deux programmes ont été écrits à cette fin : les programmes MAKEINP et MAKESEL. Tous deux ont été codés en langage Fortran 77, puis compilés à l'aide du compilateur g77 du GNU (GNU is Not Unix).

A.4.1 Étape préliminaire : MAKEINP / InsightII

Le programme MAKEINP suit un algorithme simple et linéaire (figure A4-1) permettant de générer, à partir d'une liste de solutions (issues de programmes de recherche de motifs, SPASM ou STAMP), un ensemble de fichiers « entrée » pour le programme InsightII. Dans chacun de ces fichiers, seules les solutions non-redondantes ne contenant aucun résidu Cys ou Pro sont conservées. Parmi les structures redondantes, celle dont la valeur du RMSD, calculée par rapport



au motif de référence, est conservée. Chaque solution est alors désignée par un identifiant unique à 8 caractères :

- code PDB (4),
- n° de la solution pour ce fichier (2),
- n° du conformère ou de la structure (2).

Ces fichiers sont alors utilisés pour la superposition des plateformes sur le motif de référence, et l'introduction du motif de référence sur les différentes structures.

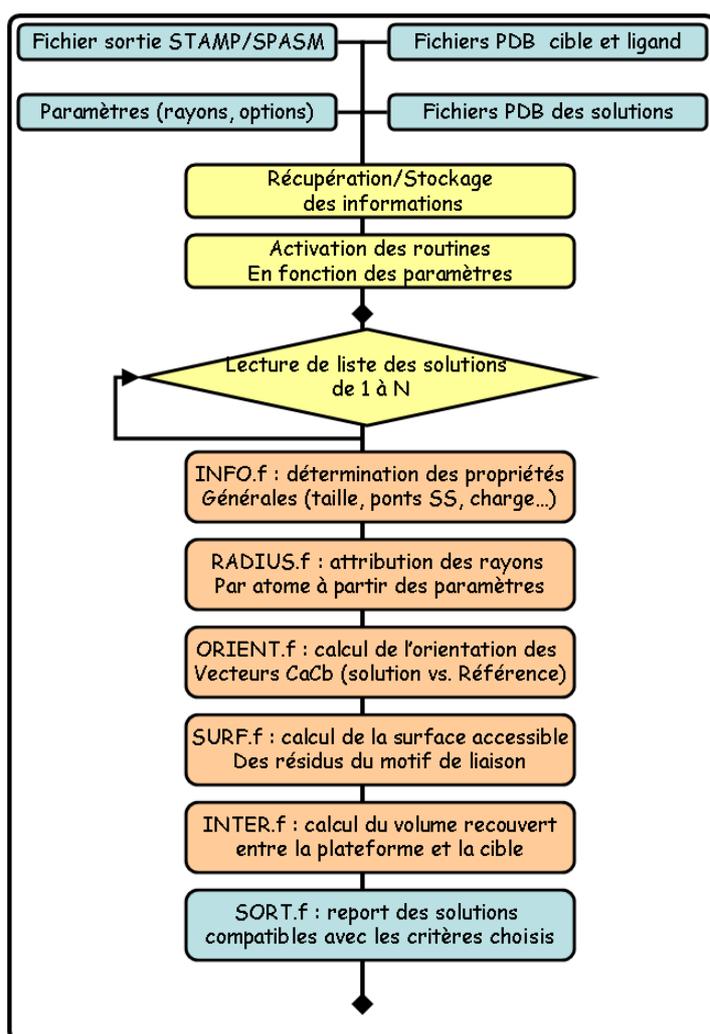
Le temps nécessaire à l'exécution du programme MAKEINP est de ce fait relativement court, et surtout limité par les étapes d'ouverture et de fermeture des fichiers pdb.

Figure A4-1 : Algorithme du programme MAKEINP

En revanche, la superposition et la modification de multiples structures (l'introduction du motif nécessite la réalisation de 4 modifications de séquence) est relativement longue, selon le nombre de solutions obtenues par le programme de recherche de motifs topologiques et de la taille des fichiers PDB (approximativement, de l'ordre de 24h pour 10000 solutions).

A.4.2 Étape de sélection : le programme MAKESEL

Le programme MAKESEL est plus complexe et nécessite au préalable de définir des paramètres atomiques pour les calculs de surface accessible (selon l'algorithme de Lee et Richards [136]) et



de volume de recouvrement (décrit dans le chapitre III). La détermination des solutions est fonction des paramètres choisis par l'opérateur : taille des protéines (inférieur et supérieur), surface accessible du motif, % de recouvrement, état de charge, résidus à l'interfaces, présence de résidus Pro, Cys ou autres, présence de ponts disulfures, orientation des chaînes latérales du motif. La liste des solutions est alors présente dans un fichier texte, reportant l'ensemble des caractéristique de chaque structure : composition en résidus, en type de résidus, nombre de ponts disulfures, taille, charge, charge à l'interface, surface accessible.

Figure A4-2 : Algorithme du programme MAKESEL.

B.1 Affinement de la structure de la BgK

La structure du motif transféré (F6/23/K25/Y26) a été obtenue à partir de la structure 3D de la BgK (code PDB 1BGK), déposée dans la PDB. Une structure moyenne a été calculée à partir des coordonnées atomiques des 15 conformères déposés dans le fichier 1BGK. La structure résultante a ensuite été minimisée avec la version 27 du programme CHARMM [303, 304]. Pour ce faire, le champ de force de CHARMM, version 22, a été utilisé. La minimisation d'énergie a été effectuée par cinq cycles, de 1000 pas chacun, en utilisant la méthode de Newton-Raphson, avec des contraintes harmoniques sur la position des atomes lourds (C, H, O et N). Pendant la minimisation, les contraintes harmoniques appliquées aux atomes lourds de la chaîne principale ont été modifiées de 100 à 20 kcal.mol⁻¹ Å⁻², par pas de 20 kcal.mol⁻¹ Å⁻², et 10 à 2 kcal.mol⁻¹ Å⁻², par pas de 2 kcal mol⁻¹ Å⁻² pour les atomes lourds des chaînes latérales. La structure obtenue a ensuite été soumise à une minimisation de 10000 pas, avec des contraintes positionnelles et harmoniques pour des atomes lourds de la protéines de 5 et 0.5 kcal.mol⁻¹ Å⁻² respectivement. La structure finale, moyennée et minimisée, a été superposée sur les 15 différents conformères de la BgK, par les atomes C α et C β des résidus du motif. Pour l'ensemble des conformères, toutes les valeurs de RMSD calculées ont été inférieures à 0,42Å, témoignant ainsi du maintien de la topologie du motif.

B.2 Systèmes de complexes cristallographiques

B.2.1 Recherche de complexes cristallographiques

Les systèmes présentés dans le chapitre IV sont composés de structures issues de la PDB. Plusieurs voies ont été utilisées afin d'identifier dans la PDB des systèmes protéiques, composés d'une même cible et de divers ligands, nous permettant ainsi de comparer leur mode d'interaction. Pour que les comparaisons effectuées ultérieurement soient représentatives, il convient de disposer d'échantillons suffisamment grands, or nous sommes limités par le nombre de complexes présents actuellement dans la PDB. Pour chaque système, nous avons fixé à 5 le nombre minimum de ligands, différents, représentant au minimum 10 comparaisons de ligands, pris deux-à-deux. Les différentes approches envisagées sont les suivantes :

B.2.1-A Recherche par PDBsum (*)

Cet outil bioinformatique est mis à disposition pour la recherche des fichiers contenus dans la PDB, triés selon différents critères de recherche (espèce, code PDB, classe d'enzyme...). Dès lors, chercher toutes les structures contenant une enzyme précise ou une classe d'enzymes devient aisément réalisable. Néanmoins, plusieurs contraintes sont inhérentes à cette approche : les options de recherche ne peuvent être associées entre elles, obligeant le cas échéant de multiples recherches, et les recherches par types de protéines ne concernent que les enzymes.

B.2.1-B Recherche par BLAST (**)

L'utilisation de BLAST [305, 306], dans notre optique de recherche de systèmes de complexes, nécessite au préalable la connaissance de la séquence primaire de la cible : il faut donc connaître la cible recherchée, contrairement à la recherche par PDBsum. A partir de la séquence, le programme BLAST recherche, dans l'ensemble des bases de données ou d'une seulement (la PDB notamment), toutes les structures primaires identiques ou homologues. La liste des solutions pouvant être relativement imposante, il est possible d'utiliser différents critères pour diminuer le nombre de solutions (ceux-ci peuvent être consultés sur le site internet du programme BLAST). Au final, une liste de solutions est proposée présentant les solutions par ordre croissant du score de similarité des séquences, liste pouvant aussi être présentée sous différents formats (arbres d'évolution, séquences alignées...) pour des facilités de recherche.

* www.ncbi.nih.gov/BLAST/

** www.ebi.ac.uk/thornton-srv/databases/pdbsum/

Contrairement à PDBsum, il n'y a aucune limite quant à la nature de la protéine recherchée (enzyme ou autre) puisque la recherche dépend uniquement de la séquence primaire.

B.2.1-C Recherche par STAMP

De part la polyvalence du programme STAMP, nous avons pu l'utiliser pour cette tâche. La recherche par STAMP s'appuie sur les structures primaire et tertiaire de la cible : il faut donc déterminer un motif de résidus unique, dans une topologie propre à la cible, à partir duquel la recherche dans l'intégralité de la PDB sera effectuée. Connaissant la taille de la protéine recherchée, il est possible d'utiliser le critère de taille du programme STAMP, réduisant ainsi le nombre de fichiers testés et le nombre de solutions potentielles. Si le motif est choisi de façon adéquate (par exemple, la triade catalytique Asp-Ser-His pour la recherche de protéases à serine), la liste de solutions ne contiendra que les fichiers incluant la structure de la cible recherchée, sous forme de complexe ou sous forme libre. Pour différencier les différentes protéases, il suffira d'ajouter au motif un résidu spécifiquement présent dans la cible recherchée.

B.2.1-D Résultats

L'utilisation des différents outils de façon complémentaire permet un gain de temps conséquent, la méthode employée a donc été la suivante :

- 1) **Sélection de systèmes** de protéines (PDBsum ; ex : serine protéase)
- 2) **Détermination de motifs spécifiques** de chaque cible (BLAST)
- 3) **Recherche des fichiers** pdb contenant la cible (STAMP)

En utilisant ces différents moyens, nous avons réussi à déterminer plusieurs systèmes de complexes cristallographiques, de protéases à serine (étudiées par la suite), de protéines du cycle cellulaire (cdk2, cdc42, ras, rac), de protéines de structure (actine, caténine), et d'enzymes diverses (amylase, lysozyme). A partir de la PDB, version août 2004, 19 systèmes de ligands ont été identifiés. Ces derniers sont présentés dans le tableau B2-1.

Système comparatif (protéine-protéine)	Nombre de complexes	Fichiers pdb conservés	Taille des ligands (min)	Taille des ligands (max)
Protéases à serine				
Trypsine bovine	34	9	3	375
Trypsine porcine	18	9	10	182
Trypsine rat	16	1	56	378
Élastase	10	3	4	61
Thrombine humaine	24	4	10	480
Thrombine bovine	7	4	11	142
Chymotrypsine bovine	23	8	3	138
Subtilisine (BPN')	26	3	4	107
Subtilisine (Carlsberg)	6	4	51	123
Protéines du cycle cellulaire				
RAS humaine (p21)	9	2	87	965
P21-RAC	9	2	135	383
CDK-2	13	1	79	261
CDC-42	12	3	44	353
Hétéro-complexes divers				
β -caténine	10	4	15	151
hGH	7	0	203	238
Actine	17	2	125	458
Lysozyme	19	2	135	429
α -amylase	5	2	74	223
α -amylase (T. molitor)	3	3	32	198

Tableau B2-1: Liste des systèmes de complexes protéiques utilisables pour la comparaison des interfaces et la caractérisation des mode de liaison de leurs ligands respectifs. Les systèmes conservés pour différentes comparaisons (chapitre IV, annexeB.2) sont indiqués en gras et en bleu.

B.2.2 Sélection des complexes cristallographiques

La recherche dans la PDB de systèmes de complexes, dont les ligands interagissent avec une même cible, dans le but de les comparer entre eux, implique que tous les complexes du système identifié doivent posséder des cibles strictement identiques et des ligands relativement différents. A partir de la liste initiale des fichiers pdb contenus dans chaque système, différentes sélections seront nécessaires, concernant notamment la qualité et l'intégrité des structures :

- seuls les complexes cristallographiques de résolution égale ou supérieure à 3.5Å ont été conservés ;
- seuls les ligands de taille inférieure à 200 résidus ont été considérés ; les tailles, minimale et maximale, observées dans chaque système sont notées dans le tableau B2-1 ;
- seuls les ligands de nature protéique ont été considérés ;
- seuls les complexes binaires, « cible-ligand » ont été pris en compte ;

- seuls les ligands dont la structure du ligand est entièrement résolue ont été utilisés pour les comparaisons. Lorsque la structure est incomplète, différents cas ont été envisagés : lorsqu'il existe, dans la PDB, un fichier correspondant à une structure identique à celle du ligand et complètement résolue, la structure incomplète est reconstruite à partir de celle-ci ; lorsque les délétions dans la structure sont de faible importance, ceux-ci sont reconstruits dans InsightII (Accelrys). Les structures ainsi obtenues ont été préalablement minimisées avant toute comparaison.

En définitive, parmi les 19 systèmes sélectionnés, seuls 3 d'entre eux ont été conservés (en bleu dans le tableau B2-1), contenant chacun au minimum 5 ligands : « trypsine bovine », trypsine porcine » et « chymotrypsine bovine ». Ces systèmes sont amplement décrits dans le chapitre IV.

B.2.3 Comparaison des potentiels électrostatiques

Les potentiels électrostatiques des différents ligands de chaque système (« trypsine bovine », « trypsine porcine » et « chymotrypsine bovine ») ont été comparés, deux à deux, en suivant la procédure décrite dans l'annexe A.3. Les résultats obtenus sont présentés dans les tableaux B2-2, B2-3 et B2-4 pour les trois systèmes.

	1c9t	1d6r	1g9i	1ox1	1sfi	1taw	1tgs	2btc	2ptc
1c9t									
1d6r	0,71 :0,72								
1g9i	0,56 :0,77	0,20 :0,65							
1ox1	0,38 :0,61	0,14 :0,42	0,83 :0,90						
1sfi	0,59 :0,78	0,56 :0,82	0,79 :0,95	0,67 :0,81					
1taw	0,53 :0,42	0,71 :0,63	0,33 :0,52	0,23 :0,41	0,45 :0,59				
1tgs	0,29 :0,70	-0,11 :0,50	0,67 :0,90	0,90 :0,95	0,65 :0,84	0,10 :0,50			
2btc	0,31 :0,43	0,54 :0,72	0,47 :0,70	0,68 :0,71	0,58 :0,76	0,75 :0,70	0,41 :0,64		
2ptc	0,06 :0,48	0,07 :0,38	0,49 :0,59	0,44 :0,53	0,50 :0,59	-0,14 :0,10	0,46 :0,63	-0,03 :0,26	

Tableau B2-2 : Comparaison des grilles de potentiels des ligands du système « trypsine bovine ». Dans chaque case, est indiquée, à gauche, la comparaison du potentiel global, à droite, celle du potentiel dans la région de l'interface seulement.

	1avx	1eja	1h9h	1ldt	1mct	1tfx	1tx6	1uhb	1z7k
1avx									
1eja	0,10 :0,45								
1h9h	-0,10 :0,02	0,39 :0,75							
1ldt	-0,19 :0,11	0,05 :0,50	0,52 :0,82						
1mct	-0,28 :0,18	0,36 :0,65	0,43 :0,72	0,54 :0,61					
1tfx	0,51 :0,14	0,61 :0,45	0,56 :0,64	0,00 :0,44	-0,03 :0,16				
1tx6	0,11 :0,52	-0,49 :0,50	0,19 :0,54	0,52 :0,38	0,02 :0,19	-0,31 :0,67			
1uhb	0,07 :0,01	0,55 :0,46	0,26 :0,27	0,07 :0,06	0,65 :0,58	-0,32 :0,30	-0,30 :0,16		
1z7k	0,52 :0,52	0,20 :0,21	0,03 :0,01	-0,15 :0,03	0,44 :0,49	-0,06 :0,27	-0,11 :0,14	0,24 :0,17	

Tableau B2-3 : Comparaison des grilles de potentiels des ligands du système « trypsine porcine ». Dans chaque case, est indiquée, à gauche, la comparaison du potentiel global, à droite, celle du potentiel dans la région de l'interface seulement.

	1acb	1ca0	1cbw	1cgi	1cho	1gl0	1gl1	1n8o
1acb								
1ca0	-0,10 ; -0,05							
1cbw	0,41 ; 0,56	-0,14 ; 0,06						
1cgi	-0,17 ; 0,62	0,31 ; -0,02	0,34 ; 0,45					
1cho	0,33 ; 0,68	0,31 ; 0,03	0,36 ; 0,79	0,65 ; 0,69				
1gl0	-0,40 ; -0,18	0,50 ; 0,67	-0,15 ; -0,06	0,04 ; -0,16	-0,08 ; -0,10			
1gl1	0,40 ; 0,57	0,19 ; 0,15	0,79 ; 0,89	0,49 ; 0,57	0,68 ; 0,83	0,00 ; 0,01		
1n8o	0,15 ; 0,53	0,02 ; 0,38	0,69 ; 0,73	0,21 ; 0,33	0,09 ; 0,52	-0,01 ; 0,12	0,67 ; 0,73	

Tableau B2-4 : Comparaison des grilles de potentiels des ligands du système « chymotrypsine bovine ». Dans chaque case, est indiquée, à gauche, la comparaison du potentiel global, à droite, celle du potentiel dans la région de l'interface seulement.

B.3 Analyse d'interfaces protéine-protéine

Diverses interfaces protéiques ont été analysées selon des paramètres globaux ou des aspects microscopiques (complexes cristallographiques dans le chapitre IV et modèles moléculaires dans le chapitre VI). Les analyses ont été effectuées avec un programme, écrit avec le macro-langage de CHARMM [303, 304]. La taille des interfaces a été calculée comme la différence de la surface accessible totale des deux partenaires, à l'état libre, et de la surface accessible du complexe considéré. La surface accessible a été calculée selon l'algorithme décrit par Lee et Richards [136], en utilisant une sonde moléculaire de rayon 1,4 Å. Les volumes vides à l'interface ont été calculés par la fonction COOR VOL de CHARMM, comme la différence entre le volume total du complexe est le volume des protéines, prises séparément. Le « gap index » [139, 142] est défini comme le ratio du volume vide à l'interface et de la surface de l'interface, calculés précédemment.

Les critères utilisés pour identifier les liaisons hydrogène sont ceux du programme HBPLUS [282, 283], les distances entre atomes, donneur (D) et accepteur (A), devant être inférieures à 3,9 Å, et les distances, entre atomes d'hydrogène (H) et accepteur, inférieures à 2,5 Å. De plus, l'angle formé par les atomes D-H-A, D-A-A₍₋₁₎ et H-A-A₍₋₁₎, doit être inférieur à 90°.

Les résidus sont considérés être dans un état de protonation identique à celui des mêmes résidus, libres, à pH=7,0. Les ponts salins ont été décrits par tout couple de charges opposées distantes de moins de 3,0 Å, et les « choc » électrostatiques par tout couple de charges identiques distantes de moins de 6,0 Å.

C.1 Séquence des plateformes synthétisées

Les résidus en rouge représentent le motif de liaison FSKY transféré. Les résidus en vert représentent les variants effectués afin de valider le mode d'interaction. Les résidus en jaune correspondent aux résidus du motif, identiques aux résidus du motif de liaison sur la plateforme native.

C.1.1 Plateformes obtenues par biologie moléculaire

1J75WT : GSHMLSTGDNLEQKILQVLSDDGGPVKIGQLVKKCQVPKKTLNQVLYRLKKEDRVSS PEPATWSIGG
1J75MT : GSHMLSTGDNLEQKILQVLSDDGGPVKIFQLVKKCQVPKKTLNQVLYRLKKEDRVSS PSPKYWSIGG
1J75MT C38V : GSHMLSTGDNLEQKILQVLSDDGGPVKIFQLVKKVQVPKKTLNQVLYRLKKEDRVSS PSPKYWSIGG
1J75MT K30A : GSHMLSTGDNLEQKILQVLSDDGGPVAFIFQLVKKCQVPKKTLNQVLYRLKKEDRVSS PSPKYWSIGG
1J75MT K17A : GSHMLSTGDNLEQAIFLQVLSDDGGPVKIFQLVKKCQVPKKTLNQVLYRLKKEDRVSS PSPKYWSIGG

C.1.2 Plateformes obtenues par synthèse chimique

1HZ6WT : EEVTIKANLIFANGSTQTAEFKGTFEKATSEAYAYADTLKKDNGEWTVDVADKGYTLNIKFAG
1HZ6MT : EEVTIKANLIFANGSTQTAEFKGTFEKATSEAYAYADTLKKDNGEWTVDVA~~SKY~~TLNIKFAG

1CKKWT : VKLIPSWTTVILVKSMLRKR~~SFGNPF~~
1CKKMT : VKLIP~~SWKY~~VILVKSMLRKR~~SFGNPF~~

1STUWT : PISQVHEIGIKRNMTVHFVLRREEGPAHMKNFITACIVGSIVTEGEGNGKKVSKKRAAEKMLVELQKL
1STUMT : PISQVHEIGIKRNMTVHFVLRREEGPAHMK~~S~~FITACIVGSIVTEG~~YK~~NGKKVSKKRAAE~~F~~MLVELQKL

1H5OWT : YKQCHKKGGHCFPKFKICLPPSSDFGKMDCRWRWKCKKGGSG
1H5OMT : YKQCHKKGGHCFPKFKICLPPSSDFGKMDC~~SWKY~~KCKKGGSG
1H5OMT2 : ~~xx~~QCHKKGGHCFPKFKICLPPSSDFGKMDC~~SWKY~~KCKKGGSG
1H5OMT3 : YKQCHKKGGHCFPKFKICLPPSSDFGKMDC~~SWAA~~KCKKGGSG
1H5OMT4 : YKQCHKKGGHCFPKFKICLPPSSDFGKMDC~~SWKY~~KCKKGGSG

1D6BMT : IMFFEMQACWSHSGVCRDK~~ERNCKPMAWTYCESRKY~~KCCEY

1C5AMT : MLQKKIEEEAAKY~~SAKY~~KKCCYDGAYRNDETCEERAARIK~~SGPKCVKAFKDCCYIANQVRAEQSHKNIQLG~~

1CR8MT : PGGCHTDEFQCR~~EDGLC~~IPLRWRCDGDTDCMD~~YKDE~~SCEGV

C.2 Synthèse par biologie moléculaire

C.2.1 Protocole expérimental : construction du vecteur d'expression

La plateforme 1J75 (domaine DLM1 de liaison de l'ADN Z) a été obtenue par des techniques de biologie moléculaire, en adaptant le protocole décrit dans la littérature [260, 307]. Un gène de synthèse codant pour la plateforme 1J75mt, contenant notre motif de liaison, a été obtenu par Genotech, cloné dans un plasmide pcDNA3.1. L'utilisation des enzymes de restriction HindIII et NdeI ont permis d'isoler le fragment d'ADN correspondant, de le purifier puis de l'insérer dans un vecteur d'expression pET28a (Novagen). La construction correspond à l'enchaînement d'une étiquette histidine (« His tag », (His)₆), d'un site de clivage à la thrombine (-Leu-Val-Pro-Arg-Gly-Ser-His-Met), et de la séquence de la protéine 1J75mt (résidus 8 à 70). Le vecteur d'expression est décrit par la figure C1-1.

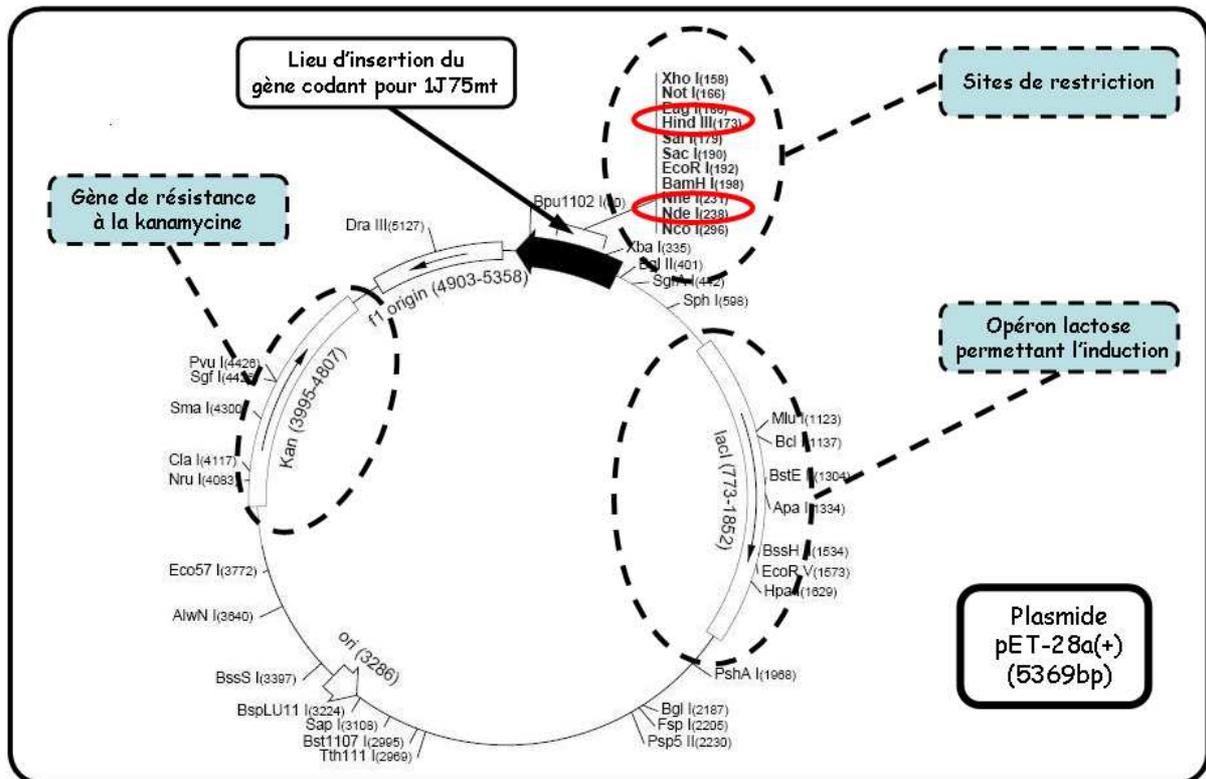


Figure C2-1 : Vecteur d'expression plasmidique utilisé pour l'obtention de la plateforme 1J75mt recombinante.

Les différents variants de la plateforme 1J75 ont été obtenus à partir de la construction précédemment décrite, par mutagenèse dirigée par PCR (Stratagene, « Quick Change Site-Directed Mutagenesis Kit »). Chaque construction ainsi obtenue a été séquencée, par un analyseur ABI PRISM 310 (Applied Biosystems), de façon à vérifier le sens d'insertion ainsi que la séquence codante pour la plateforme : aucune erreur n'a été observée dans les séquences codantes et la présence des mutations insérées a été confirmée.

C.2.2 Protocole expérimental : production, purification

La production des plateformes a été réalisée dans 500 ml de milieu LB, contenant 30 µg/ml de kanamycine,ensemencée avec 5 ml d'une culture d'Escherichia Coli, souche BL21-DE3, transformée avec le plasmide pET28a-1J75, et incubée à 37°C pendant une nuit. L'expression des protéines recombinantes a été induite par addition d'IPTG, pour une absorbance à 600nm égale à 0,8. L'incubation a été poursuivie pendant 3h à 37°C, puis les cellules furent récupérées par centrifugation. Le culot a été alors resuspendu dans un tampon contenant 50 mM Tris-HCl, pH 8,0 / 300 mM NaCl / 10 mM Imidazole / 5mM β-mercaptoéthanol / 0,1 mM PMSF / 20 µg/ml benzonase, puis les cellules furent lysées par une « french press ». Le lysat a été centrifugé (16000 rpm / 1h), puis le surnageant récupéré et incubé avec 4 ml de résine Ni-NTA pendant 1h à 4°C. La résine, déposée dans une colonne, a été lavée avec 50 ml de tampon B (50 mM Tris-HCl, pH=8,0 / 1 M NaCl / 10 mM Imidazole / 5 mM β-mercaptoéthanol). Les protéines de fusion (« His-tag »-1J75) ont été éluées avec 20 ml de tampon C (50 mM Tris-HCl, pH=8,0 / 300 mM NaCl / 200 mM Imidazole / 5 mM β-mercaptoéthanol) et dialysée, une nuit, avec un tampon D (20 mM Tris-HCl, pH=8,0 / 150 mM NaCl / 2 mM DTT). La protéine a été obtenu après clivage à la thrombine (1 unité thrombine / 100 µg protéine) pendant 24h à 22°C. Puis, 10 ml de tampon de Na₂HPO₄/NaH₂PO₄, 500mM, pH=4,2 ont été ajoutés et la solution saturée en Argon, avant addition de TCEP (1 équivalent). Après 2h d'incubation, les plateformes 1J75 ont été purifiées par RP-HPLC sur une colonne C18 préparative Vydac par un gradient linéaire de 92 min, allant de 20 à 100% d'éluant B (60% acétonitrile / 0,085% TFA) dans l'éluant A (0,1% TFA), à un débit de 10 ml/min. Les protéines obtenues ont été lyophilisées et solubilisée dans l'eau pour des analyses ultérieures.

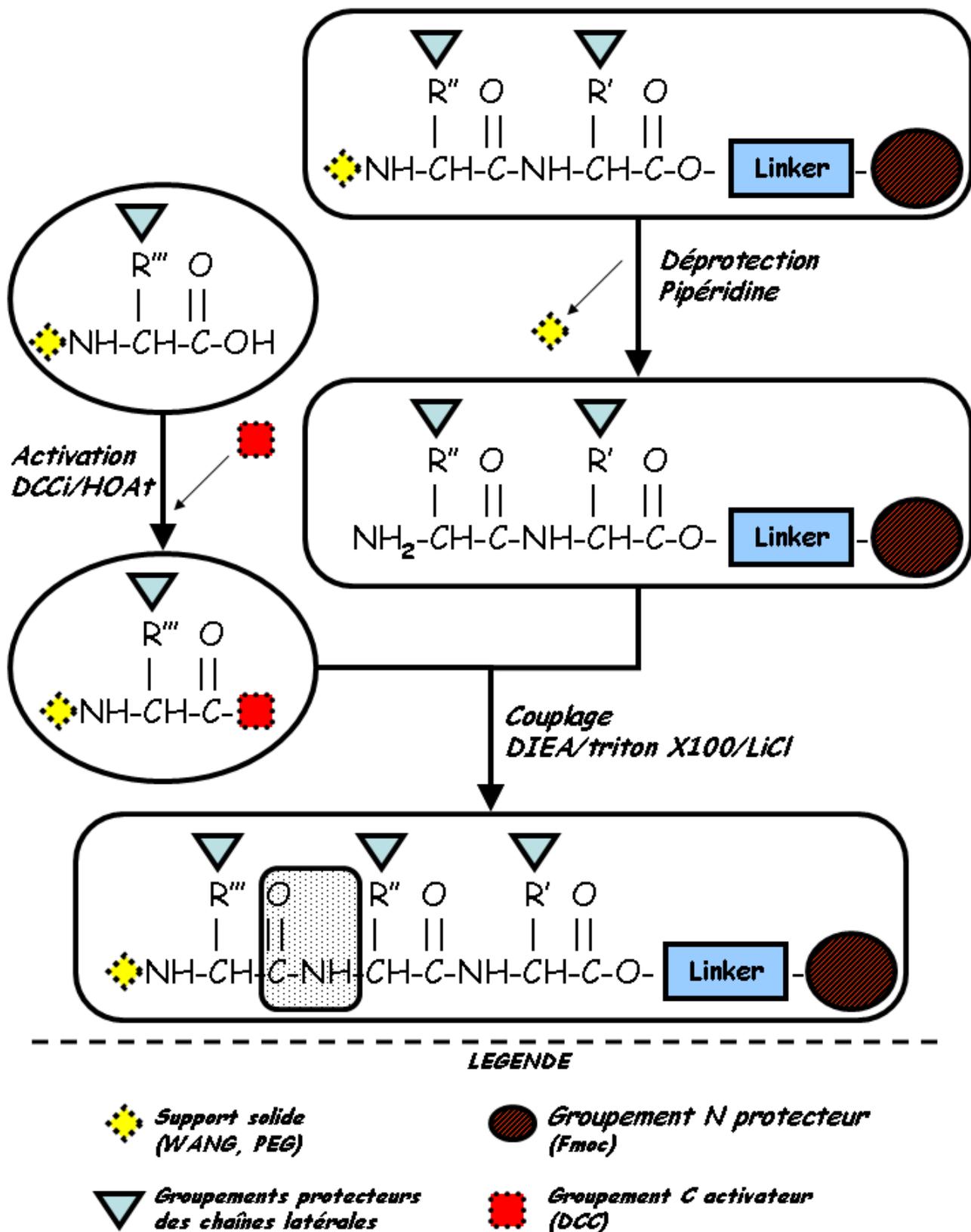


Figure C3-1 : Schéma général de la synthèse par voie chimique.

C.3 Synthèse par voie chimique

C.3.1 Synthèse chimique sur support solide : rappel

Les principes de la synthèse chimique de peptides en phase solide remontent au début des années 60 par Merrifield [14]. Depuis, cette méthode a été grandement optimisée en termes de tailles de protéines accessibles à la synthèse, bien que son principe demeure identique. A l'heure actuelle, la taille maximale d'une protéine obtenue par cette méthode est de 136 résidus pour une seule synthèse [308]. Il est possible d'augmenter cette limite par l'association de différents fragments synthétisés séparément. Ainsi, la GFP (« Green Fluorescent Protein »), protéine de 238 résidus a été obtenue [309]. En utilisation « routinière », cette méthode permet d'obtenir des protéines de tailles comprises entre 50 et 100 résidus.

La synthèse chimique, décrite sur la figure C3-1, sur support solide est réalisée sur des billes de polymères, auxquelles sont attachées de façon covalente, par l'intermédiaire d'un « linker », l'acide aminé C terminal, (N). Pour le déroulement du procédé l'ensemble des chaînes latérales des acides aminés seront protégées par un groupement protecteur, évitant que celles-ci puissent réagir. La fonction amine, quant à elle, est protégée par un groupement protecteur (Boc ou Fmoc), qui sera clivé pour permettre la création d'une liaison peptidique entre le résidu N et N+1 (étape de couplage), le groupement protecteur de la fonction amine de l'acide aminé N+1 restant intact. La fonction acide carboxylique de l'acide aminé N+1 est ensuite activée (rendue plus réactive par formation d'un anhydride, typiquement le DCCI), lui permettant de réagir avec la fonction amine de l'acide aminé N. L'addition d'esters activés (HOAt, HOBt) permet d'accroître la réactivité tout en limitant les réactions secondaires.

En fin de couplage, certaines chaînes en élongation n'ont pas réagi, la fonction amine de l'acide aminé N est donc encore libre ; pour éviter que celle-ci réagisse ultérieurement, aboutissant à des protéines délétères (intra-chaîne), une étape d'acétylation ou « capping » de la fonction N terminale est réalisée.

Le groupement Fmoc de l'acide aminé N+1 est ensuite clivé par la pipéridine, permettant le couplage avec l'acide aminé N+2. Ces réactions peuvent être réalisées de façon itérative aboutissant à l'élongation du polymère de nature protéique. Au final, le polymère est clivé de la résine (billes) et ses chaînes latérales déprotégées, en milieu acide durant une seule étape de clivage/déprotection.

C.3.2 Protocole expérimental

Les plateformes ont toutes été obtenues sur un synthétiseur automatisé (model 433A, Applied Biosystems, Foster City, USA) en utilisant la stratégie Fmoc/tBu [310]. La résine et le chargement de celle-ci, utilisés pour chaque synthèse, sont présentés dans le tableau C3-2.

Plateforme	Résine	Chargement (mmol)
1H50 wt	Fmoc-gly-Wang	0.610
1H50 mt, mt3, mt4	Fmoc-gly-PAL-PEG-PS	0.250
1HZ6 wt, mt	Fmoc-gly-Wang	0.610
1CKK wt, mt	Fmoc-phe-Wang	0.640
1FOE wt, mt	Fmoc-phe-Wang	0.640
1STU wt, mt	Fmoc-leu-Wang	0.840
1CR8 mt	Fmol-val-PAL-PEG-PS	0.220
1C5A mt	Fmoc-gly-Wang	0.610
1D6B mt	Fmoc-tyr-Wang	0.450

Tableau C3-2 : Conditions utilisées lors de la synthèse chimique pour chaque plateforme et ses différentes versions.

Les synthèses ont été effectuées en utilisant un protocole standard : 0.1 mmol pour chaque acide aminé (Fmoc) C terminal préchargé, un excès d'un facteur 10 pour chaque acides aminés (Fmoc) utilisé pour la synthèse, l'activation par le couple DCCI / HOAt, et le couplage par un mélange 0,1 mmol DIEA / 2% (v/v) Triton X100 / 0,5M LiCl / NMP. Le rendement de couplage est estimé, lors de l'étape de clivage du groupement Fmoc de l'acide aminé juste couplé, par la mesure de l'absorbance de celui-ci à 301 nm. Le rendement de synthèse totale est calculé par le ratio de l'absorbance mesurée pour premier résidu sur celle dernier résidu. Par la suite, après séchage sous vide de la résine, les protéines ont été obtenues à la suite d'une étape de clivage de la résine et de déprotection des chaînes latérales par addition d'une solution de TFA / TIS / eau (95 / 2,5 / 2,5, v/v/v), pendant 2 heures, à température ambiante (25°C). La résine a été ensuite filtrée et les protéines précipitées dans 100 ml de TBME à froid (4°C). Après centrifugation (3000 rpm / 15 min / 4°C), le culot a été solubilisé dans une solution d'acide acétique 20% (20 ml). Les protéines ont été ensuite obtenues après congélation et lyophilisation, sous vide.

C.4 Obtention des protéines synthétisées par voie chimique

C.4.1 Purification par chromatographie en phase liquide à haute performance

La purification des plateformes issues de la synthèse chimique a été effectuée en une seule étape par chromatographie en phase liquide à haute performance en phase inverse (RP-HPLC). La chromatographie en phase liquide à haute performance [277] est une méthode de choix pour l'analyse de mélanges complexes, permettant l'identification, la quantification, et la purification de composés possédant des propriétés physico-chimiques diverses. Lors de cette thèse, toutes les purifications ont été effectuées par la méthode des gradients. La purification a été réalisée sur une colonne semi-préparative Vydac C18 (granulométrie : 5 μm ; porosité : 300 \AA ; taille de la colonne 10x250 mm), en utilisant des gradients non-linéaires d'éluant B (90% d'acétonitrile dans l'eau, 0.085% TFA) dans un éluant A (10% d'acétonitrile dans l'eau, 0.1% TFA) à un débit de 3ml/min. Les protéines ont été purifiées, à raison de 1 à 5 mg de brut de synthèse. Ces quantités ont été estimées par spectroscopie UV en utilisant les coefficients d'extinction molaire théoriques, calculés pour chaque plateforme (*). La détection des produits a été suivie à 220 nm sur un détecteur Merck L4000 UV, équipé d'une cellule de détection de 5 mm. Les conditions de purification sont indiquées dans le tableau C4-1.

Plateformes	Gradient de purification	T _{elut.} en min
1H5Owt (red)	0-4' 5-15%/4-34' 15-25%	24'
1H5Owt (ox)	0-4' 5-15%/4-34' 15-25%	16'
1H5Omt1 (red)	0-4' 5-15%/4-34' 15-25%	23'
1H5Omt1 (ox)	0-4' 5-15%/4-34' 15-25%	17'
1H5Omt3 (red)	0-10' 5%/10-14' 5-15%/14-40' 15-30%	28'
1H5Omt3 (ox)	0-10' 5%/10-14' 5-15%/14-40' 15-30%	26'
1H5Omt4 (red)	0-10' 5%/10-14' 5-15%/14-40' 15-30%	27'
1H5Omt4 (ox)	0-10' 5%/10-14' 5-15%/14-40' 15-30%	26'
1CKK wt	0-5' 0-30%/5-25' 30-50%	16'
1CKK mt	0-5' 0-30%/5-25' 30-50%	14'
1HZ6 wt	0-10' 5%/10-14' 5-25%/14-42' 25-35%	42'
1HZ6 mt	0-10' 5%/10-14' 5-25%/14-42' 25-35%	39'

Tableau C4-1: Purification par RP-HPLC des différentes plateformes synthétisées par voie chimique. Sont indiqués pour chaque plateforme le gradient utilisé et le temps d'éluion. Le % correspond à la proportion d'éluant B dans le mélange :

A : 90% eau 10% AcN 0,1% TFA

B : 90% Acn 10% eau 0,085% TFA.

* <http://www.expasy.ch/tools/protparam.html>

C4.2 Repliement des protéines (formation de ponts disulfures)

La formation des ponts disulfures a été réalisée, suivant un protocole standard, à partir des protéines, préalablement purifiées, à l'état réduit, par l'utilisation de solvants dégazés et par la présence de TCEP (agent réducteur). Le couple oxydo-réducteur glutathion oxydé (GSSG) / glutathion réduit (GSH) à été utilisé, avec un ratio de 1/10 (0,5 mM GSSG / 5 mM GSH), dans une solution tampon $\text{Na}_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$, 50 mM, pH=7,8, préalablement dégazée, et une concentration finale en protéines de 1 mg/ml. Différentes conditions ont été testées, comme la modification du ratio GSSG/GSH (1/10, 1/1 et 10/1), la dénaturation totale des protéines au préalable en présence d'agents dénaturant et réducteur (6M GuCl / 10 mM DTT) à 37°C pendant 1h, la modification de la solubilité (glycérol, 20%) l'utilisation d'une autre solution tampon (Tris-HCl, 50 mM, pH=8,0), ou la modification de la température durant l'expérimentation (4°C, pendant 72h). Néanmoins, l'utilisation de ces différentes conditions expérimentales pour le repliement des plateformes ont fourni des résultats satisfaisant et ont été choisie pour les différentes expérimentations. L'oxydation est évaluée par la disparition de la forme réduite de la protéine, en différents temps (après 1h, 3h et 24h pour les expériences à 25°C), suivie par chromatographie analytique, après acidification au TFA 50%, pH=2,0. Les fractions correspondant aux protéines repliées ont ensuite été caractérisées.

C.4.3 Caractérisation biochimique des plateformes

Pour chacune des plateformes purifiées au préalable, l'ensemble des caractérisations suivantes a été effectué, permettant de connaître la pureté (par RP-HPLC analytique), la masse moléculaire (par ES-MS), le coefficient d'extinction molaire (par analyse en composition). Les résultats sont présentés dans le tableau C4-2.

C.4.3.A Caractérisation de la pureté : chromatographie analytique

Les plateformes purifiées ont été analysées par chromatographie RP-HPLC sur une colonne Vydac C18 (granulométrie : 5 μm ; porosité : 300 Å ; taille de la colonne 4,6x150 mm), dans les mêmes conditions que celles utilisées lors de la purification. La pureté des échantillons a été vérifiée par l'intégration des différents pics du chromatogramme, le pic principal devant correspondre à 95% (au minimum) de l'aire totale intégrée. Le débit a été modifié de façon à conserver, approximativement, la même vitesse linéaire soit 1 ml/min pour les différentes analyses. Du fait d'une même vitesse linéaire, les temps d'élution, pour chaque plateforme, sont donc approximativement les même que ceux observés lors de la purification.

C.4.3-B Caractérisation de la masse moléculaire : spectrométrie de masse

La caractérisation des différentes masses moléculaires de nos plateformes a été réalisée par les laboratoires ATHERIS (Genève, Suisse), par spectrométrie de masse électro spray en mode positif. Par ailleurs, l'utilisation de la spectrométrie de masse a permis de vérifier la présence ou l'absence de composés autres dans les échantillons adressés en analyse.

C.4.3-C Détermination du coefficient d'extinction molaire

Chaque plateforme purifiée a été soumise à une analyse en composition, permettant de vérifier la composition en acides aminés et de calculer leur coefficient d'extinction molaire expérimentalement. Ces valeurs (tableau C4-2) ont été utilisées pour la détermination de la concentration des solutions utilisées ultérieurement par spectroscopie UV-visible.

Plateformes	M _{exp} en Da	M _{theo} en Da	ε ₍₂₈₀₎ theo.	ε ₍₂₈₀₎ exp.
1H5Owt (red)	4889.8	4889.8	12865	n.c.
1H5Owt (ox)	4883.0	4883.8	13020	12490
1H5Omt1 (red)	4787.8	4787.7	8855	n.c.
1H5Omt1 (ox)	4781.0	4781.7	9055	8480
1H5Omt3 (red)	4638.6	4638.5	7365	n.c.
1H5Omt3 (ox)	4632.4	4632.5	7940	6990
1H5Omt4 (red)	4787.3	4788.6	8855	n.c.
1H5Omt4 (ox)	4782.4	4782.6	9100	8480
1CKK wt	3018.5	3018.7	7800	5500
1CKK mt	3107.6	3017.8	6480	6990
1HZ6 wt	6882.8	6883.5	11960	9970
1HZ6 mt	6925.6	6926.7	12730	9970

Tableau C4-2 : Caractérisation des différentes plateformes obtenues par voie chimique. La masse moléculaire a été mesurée par spectrométrie de masse, ES-MS. Les coefficients d'extinction molaire ont été estimés par une analyse en composition.

D.1 Caractérisation par dichroïsme circulaire

Notre approche reposant sur des considérations structurales et des modifications de plateformes natives, il convient, avant d'effectuer tout test de liaisons, de vérifier que les formes natives et mutantes des plateformes ont des structures similaires, en accord avec les structures utilisées lors des travaux réalisés *in-silico*. L'approche la plus simple à mettre en œuvre dans cette optique, permettant une caractérisation structurale globale, repose sur l'utilisation du dichroïsme circulaire.

D.1.1 Dichroïsme circulaire : théorie (Lowry, 1935)

Le dichroïsme circulaire est une méthode spectroscopique qui repose sur la capacité, que possède les structures optiquement actives (dont les éléments de structure secondaire), d'absorber de façon inégale la lumière polarisée circulairement à droite de la lumière polarisée circulairement à gauche. L'absorption préférentielle d'éléments chiraux pour l'une de ces deux polarisations aboutira à une déviation du champ électrique résultant, permettant le tracé d'un spectre caractéristique de ces éléments. Dans le cas de protéines, cette méthode permet d'étudier la structure secondaire en solution (hélice α et brin β notamment, comme indiqué sur la figure D1-1.

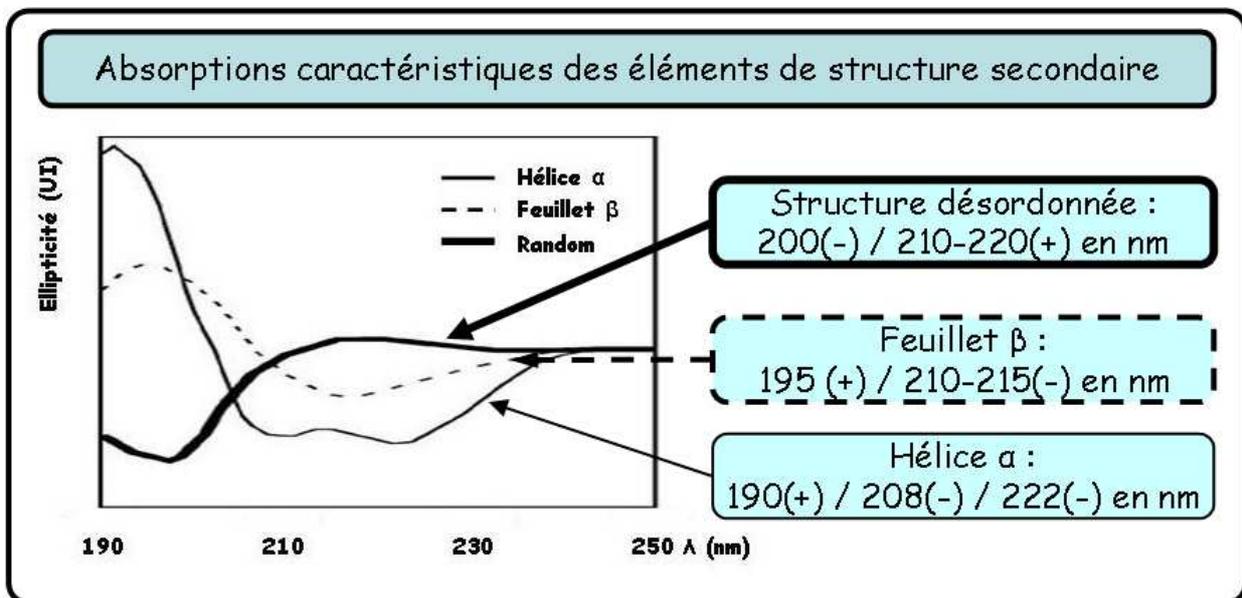


Figure D1-1 : spectres caractéristiques d'éléments de structures secondaires, présentes dans les protéines.

Ainsi, les changements dus aux conditions expérimentales peuvent être observés (changement de structure en fonction de la température, du pH, de la présence de détergents ou de sels, modifications de structure).

D.1.2 Protocole expérimental

Les spectres CD, dans l'UV lointain, ont été obtenus sur un dichrographe CDMax (Jobin Yvon, Longjumeau, France), à 21°C, de 180 à 250 nm, dans une cuve en quartz de 1 cm. Les protéines ont été solubilisées, pour atteindre une concentration finale allant de 5 à 20 μM , dans une solution tampon Tris-HCl, 4 mM, pH=6,8. Chaque spectre représente la moyenne, lissée, de cinq spectres, obtenus avec un temps d'intégration de 1 seconde tous les 1 nm. Des expériences supplémentaires ont été réalisées avec addition de 5 à 20% de TFE, ou de 10 à 100 mM KCl. Pour la plateforme 1H5O (native et mutants), les spectres ont été obtenus sur un dichrographe JASCO J-815 (Jasco France, Nantes, France), à 20°C, de 180 à 250 nm, dans une cuve en quartz de 0.1cm. Chaque spectre a été obtenu avec un temps d'intégration de 1 seconde tous les 0.2 nm. Des expériences de dénaturation thermique ont été réalisées, par pas de 5°C, de 20 à 90°C (5°C/min).

D.2 Détermination de structure par radiocristallographie

D.2.1 Cristallogenèse : technique de la goutte assise en diffusion de vapeur

Dans cette technique de cristallisation, une goutte d'un mélange de la solution de protéine et de l'agent précipitant est placée sur un support de surface concave, disposé dans un puits dans lequel la solution du précipitant est ajoutée à une concentration supérieure à celle de la goutte. Par la suite, le système est fermé avec une lamelle, et scellé avec de la graisse de silicone pour assurer l'étanchéité. Une diffusion, de vapeur d'eau ou de solutions volatiles, va s'établir et amener lentement à l'équilibre les concentrations en agent précipitant entre la goutte et le réservoir. Ainsi, l'état de sursaturation nécessaire au phénomène de cristallisation, peut-il être atteint.

D.2.2 Cristallisation et collecte des données

L'oligonucléotide 5'TCGCGCG (Genosphere Biotechnologies) a été solubilisé dans une solution de 10 mM HEPES, pH 7,5 / 20 mM NaCl. La protéine 1J75mtK30A a été solubilisée dans une solution de 5 mM HEPES, pH 7,5 / 0,5M NaCl / 5 mM DTT. La protéine 1J75mtK30A et l'ADN ont été mélangés afin d'obtenir une solution équimolaire, de concentration égale à 1,15 mM. La cristallisation a été réalisée à 17°C par la technique de la « goutte assise » par diffusion de vapeur dans les conditions suivantes : 25% PEG 4K / 0.2 M Na/K phosphate / 0.1 M MES, pH=6,0. Dans ces conditions, deux formes cristallines (aciculaire et prismatique) ont été obtenues en 1 à 2 jours d'incubation. Les deux cristaux ont été préalablement congelés dans de l'éthane liquide, après trempage dans une solution cryo-protectrice de 24% PEG 4K / 20% éthylène glycol. Pour chaque cristal, les données ont été collectées à l'ESRF (« European Synchrotron Radiation Facility ») sur la ligne ID29, sur un détecteur Quantum 315r à la longueur d'onde de 0.976 Å. Les données ont été ensuite indexées par le programme MOSFLM [311] de CCP4 [312], indiquant des différences de résolution pour les deux formes cristallines : 1,7 Å pour la forme prismatique et 2,8 Å pour la forme aciculaire.

D.2.3 Détermination de la structure tridimensionnelle

La structure du complexe a été résolue par remplacement moléculaire, effectué par le programme MOLREP [313] de CCP4, en utilisant les données de plus haute résolution (cristal prismatique) et la structure du complexe DLM1:ADN Z comme modèle [260]. Après plusieurs recherches indépendantes pour les positions de l'ADN et de la protéine, le remplacement a fourni une solution compatible avec deux protéines et deux brins d'ADN par unité asymétrique : le

groupe d'espace énantiomérique P61 ainsi a-t'il été clairement identifié. La validité de cette solution a été validée par la reconstitution de l'appariement de Watson et Crick des paires de bases de l'ADN Z dans l'unité asymétrique, par l'estimation du % de solvant dans ce groupe d'espace avec le complexe (programme MATTHEWS [314] de CCP4), par une diminution rapide du R_{factor} et du R_{free} après un cycle préalable d'affinement, en mode « rigid body » (programme REFMAC [315] de CCP4). L'affinement a été poursuivi par le programme REFMAC avec l'introduction successive de molécules d'eau et de paramètres TLS. Les quatre molécules ont été initialement contraintes par l'utilisation de symétries non-cristallographiques et de paramètres TLS, appliqués sur chaque molécule indépendamment. Les résultats de l'affinement ont été validés par le calcul systématique « d'omit-maps », par la suppression de parties du modèle utilisé pour le remplacement moléculaire. Ces dernières ont été réalisées à l'aide du programme TURBO-FRODO [316].

D.3 Tests de liaison par compétition

D.3.1 Définitions

Liaison spécifique : quantité de radioligand lié spécifiquement sur des sites de liaison, obtenue par soustraction de la liaison non-spécifique à la liaison totale. La liaison spécifique est saturable.

Liaison non spécifique : quantité de radioligand fixé non spécifiquement (par adsorption sur le matériel : membranes, filtres, plastiques, protéines..), obtenue après incubation du radioligand en présence d'un excès (x1000 en général) du même ligand non radioactif. En pratique, la liaison non spécifique n'est pas saturable et est proportionnelle à la concentration de radioligand.

Liaison totale : quantité totale de radioactivité liée spécifiquement (sites de liaison) et non spécifiquement (adsorption sur membranes, plastiques, filtres...)

Ligand : toute molécule capable de se lier à un site de liaison

IC₅₀ : concentration en compétiteur aboutissant à une diminution de 50% de la liaison du radioligand pour son site.

D.3.2 Mesure de la constante d'inhibition par compétition : rappel

La constante d'inhibition K_i est mesurée sur une préparation de membrane, surexprimant le canal Kv1.2 en présence d'un analogue radiomarqué de la BgK, par iodation. L'ajout dans la solution de concentrations croissantes pour chaque plateforme, aboutit, s'il y a compétition avec l'analogue, au déplacement de celui-ci. La mesure effectuée correspond à la liaison totale, il convient donc de soustraire à cette valeur la valeur de la liaison non-spécifique mesurée préalablement par addition de BgK, non radiomarquée. Par cette méthode, en présence de compétiteur, une courbe de déplacement du ligand marque est obtenue, montrant la radioactivité mesurée (lorsque le marquage est effectué par un radioélément) à différentes concentrations de compétiteur, égale à 100% en absence de compétiteur du ligand marqué à son site de liaison. A partir de la courbe obtenue, la valeur de l'IC₅₀ est obtenue, définie comme la concentration du compétiteur permettant l'inhibition de 50% de la liaison spécifique du ligand marqué. La valeur de la constante d'inhibition K_i est calculée par la relation de Cheng et Prusoff, correspondant à l'affinité relative du compétiteur pour le site de liaison :

$$K_i = IC_{50} / (1 + [Bgk^*]/K_d)$$

Les constantes d'affinité reportées dans le chapitre VI, partie IV, ont ainsi été calculées, correspondant aux valeurs de K_i , obtenues à partir de l'équation précédente.

D.3.3 Protocole expérimental

Les tests de liaisons ont été effectués sur des préparations membranaires, provenant de cellules tsA surexprimant le canal Kv1.2. Le radioligand utilisé est un analogue de la Bgk, ¹²⁵I-BgK(W5Y/Y26F), présentant les mêmes effets biologique que la BgK native [281]. Tous les tests ont été réalisés à température ambiante, dans un milieu contenant Tris-HCl, pH 7.4, 20 mM / NaCl, 100 mM / KCl, 5 mM / BSA, 0.1%, ainsi que le radioligand, à une concentration de 50 à 100 pM. Après une courte période d'incubation, les échantillons ont été filtrés sur des filtres Whatman GF/C, préalablement imbibés avec une solution de polyéthylène imine 0.5% (w/v), puis rincés à trois reprises avec un tampon contenant Tris-HCl, pH 7,4, 150 mM / NaCl, 150 mM à froid. Les résultats ont ensuite été analysés comme décrits dans la littérature [281].

D.4 Réalisation de modèles moléculaires

La structure des complexes entre les segments S5-S6 du canal Kv1.2 et nos différentes plateformes (1CKK, 1H5O, 1HZ6 et 1J75) ainsi que la BgK, ont été construits à partir de la structure du complexe Kv1.1-BgK, calculée à partir de cycles de double-mutants [218]. La structure cristallographique du canal Kv1.2 a été superposée à la partie correspondante du canal Kv1.1, dans le modèle d'interaction Kv1.1-BgK. La structure du complexe Kv1.2-BgK a alors été relaxée à partir d'un protocole de dynamique moléculaire et de minimisation d'énergie, avec contraintes positionnelles, en utilisant la version 27 du programme CHARMM [303, 304]. Le champ de force de CHARMM, version 22, a été utilisé lors des calculs. Dans ce protocole, les contraintes positionnelles ont été appliquées aux atomes Ca des segments S5-S6 du canal Kv1.2, avec une constante de force de $50 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ appliquée aux résidus localisés à une distance supérieure à 7 \AA du ligand, de $5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ pour ceux localisés à une distance inférieure à 7 \AA , et une contrainte harmonique avec une constante de force de $0,005 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ pour les atomes, non hydrogène, autres que Ca, localisés à une distance inférieure à 7 \AA . En revanche, aucune contrainte de position n'a été appliquée sauf concernant les atomes lourds des chaînes latérales du motif F6/S23/K25/Y26 dans le complexe Kv1.2-BgK, auxquels une contrainte de force de $5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ a été utilisée pendant l'affinement du complexe. De plus, un ensemble de contraintes de distances Ca-Ca intramoléculaires a été appliqué à la structure du ligand. Cette procédure permet un positionnement des plateformes par rapport à la structure du canal Kv1.2, la relaxation des chaînes latérales des ligands et du canal Kv1.2, tout en assurant la préservation de la structure globale des protéines et de la topologie du motif de liaison en interaction avec le canal, de façon similaire à celle observée dans le modèle Kv1.2-BgK. La première étape de la procédure consiste à neutraliser électriquement le complexe, par addition d'atomes Na ou Cl, selon la charge globale du complexe. Ces contre-ions sont ajoutés successivement à des positions correspondant à des minimum d'énergie du système, jusqu'à atteindre une charge globale nulle. Les cinq complexes réalisés (1BGK, 1CKK, 1H5O, 1HZ6 et 1J75) ont alors été soumis à une dynamique moléculaire à 300K de 750000 pas, en utilisant l'algorithme de Verlet, avec un temps d'intégration, par pas, de 0,0005 ps. Trente structures, isolées tous les 25000 pas, ont été soumises à 2500 pas de minimisation d'énergie, puis analysées, afin d'estimer la convergence de l'énergie globale du système. Ces calculs ont été réalisés en représentant les interactions non-liées par une fonction de Lennard Jones ; les interactions coulombiennes par un terme électrostatique, avec un « cut-off » de 15 \AA . La constante diélectrique a été fixée à 1 pour l'ensemble des calculs, réalisés dans le vide.

Ces étapes apparaissent nécessaires afin de permettre l'ajustement de chacun des complexes modélisés. Ces structures finales minimisées représentent chaque système dans un état stable, permettant ultérieurement l'analyse de chaque complexe. La qualité de la structure de chaque plateforme a été évaluée par le calcul du RMSD entre leur structure finale (après simulation de dynamique moléculaire et minimisation) et leur structure initiale (complexe cristallographique). Dans chacun des cas, aucune modification significative de la valeur de RMSD n'a été observée, toutes étant inférieures à 0,28 Å (au maximum). Les valeurs moyennes de RMSD calculées, sur les structures de la cible et du ligand sont respectivement de $0,11 \pm 0,04$ Å et de $0,21 \pm 0,04$ Å. De part ces résultats, aucune modification importante de structure n'a été effectuée, seuls des mouvements de chaînes latérales peuvent expliquer des variations de RMSD aussi faibles.

RÉFÉRENCES BIBLIOGRAPHIQUES

« Souvent, il n'y a rien dessus, tout est en dessous » (Paracelse)

-
- 1 Tsinopoulos, C., Mc Carthy, I. (2002). An evolutionary classification of the strategies for drug discovery. *Conference Proceedings, University of Cambridge*, <http://www.ifm.eng.cam.ac.uk/mcn/proceedings.htm>.
 - 2 La Revue Prescrire (2003). Coût de recherche et développement du médicament : la grande illusion. *La Revue Prescrire*, 23(244), 782-787.
 - 3 DiMasi, J.A., Hansen, R.W., Grabowski, H.G. (2003). The price of innovation: new estimates of drug development costs. *Journal of Health Economics*, 22, 151-185.
 - 4 Jung, G. Combinatorial Chemistry: synthesis, analysis, screening. Tiebes, D., Chapitre 1 : Combinatorial Chemistry, p1-30. Wiley-VCH, 1999.
 - 5 Smith, G.P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on virion surface. *Science*, 228, 1315-1317.
 - 6 Yuan, X., Kurek, I., English, J., Keenan, R. (2005). Laboratory-directed protein evolution. *Microbiology and Molecular Biology Reviews*, 69(3), 373-392.
 - 7 Pollock, S., Safer, H.M. (2001). Bioinformatics in the drug discovery process. *Annual Reports in Medicinal Chemistry*, section V-Topics in Biology, 1-10.
 - 8 Egner, U., Krätzschmar, J., Kreft, B., Pohlenz, H.D., Schneider, M. (2005). The target discovery process. *Chembiochem*, 6(3), 468-479.
 - 9 Gallop, M.A., Barrett, R.W., Dower, W.J., Fodor, S.P.A., Gordon, E.M. (1994). Applications of combinatorial technologies to drug discovery. I. Background and peptide combinatorial libraries. *Journal of Medicinal Chemistry*, 37(9), 1233-1251.
 - 10 Warr, W.A. (1996). Combinatorial chemistry and molecular diversity. An overview. *Fourth International Conference on Chemical Structures*, June 4, 1996, Noordwijkerhout, Nederland.
 - 11 Bures, M.G., Martin, Y.C. (1998). Computational methods in molecular diversity in combinatorial chemistry. *Current Opinion in Chemical Biology*, 2(3), 376-380.
 - 12 Rodrigues, A.D. (1997). Preclinical drug metabolism in the age of high-throughput screening: an industrial perspective. *Pharmaceutical Research*, 14(11), 1504-1510.
 - 13 Sprous, D.G., Lewis, D.R., Leonard, J.M., Heritage, T., Burkett, S.N., Baker, D.S., Clark, R.D. (2004). OptiDock: virtual HTS of combinatorial libraries by efficient sampling binding modes in product space. *Journal of Combinatorial Chemistry*, 6(4), 530-539.
 - 14 Merrifield, R.B. (1963). The solid phase synthesis. 1. Synthesis of a tetrapeptide. *Journal of American Chemical Society*, 85(14), 2149-2154.
 - 15 Schreiber, S.L. (2000). Target-oriented synthesis and diversity oriented organic synthesis in drug discovery. *Science*, 287, 1964-1969.
 - 16 Spring, D.R. (2003). Diversity-oriented synthesis ; a challenge for synthetic chemists. *Organic & Biomolecular Chemistry*, 1(22), 3867-3870.
 - 17 Smith, G.P., Petrenko, V.A. (1997). Phage display. *Chemical Reviews*, 97(2), 391-410.
 - 18 Cortese, R., Monaci, P., Nicosia, A., Luzzago, A., Felici, F., Galfre, G., Pessi, A., Tramontano, A., Sollazzo, M. (1995). Identification of biologically active peptides using random libraries displayed on phage. *Current Opinion in Biotechnology*, 6(1), 73-80.
 - 19 Sams-Dodd, F. (2005). Target-based drug discovery : is something wrong?. *Drug Discovery Today*, 10(2), 139-147.
 - 20 Lewis, R.A. (1991). Rational methods for site-directed drug design: novel approaches for the discovery of potential ligands. *Biochemical Society Transactions*, 19(4), 883-887.
 - 21 Kuntz, I.D. (1992). Structure-based strategies for drug design and discovery. *Science*, 257, 1078-1082.
 - 22 Rowland, R.S. (2005). Using X-ray crystallography in drug discovery. *Current Opinion in Drug Discovery & Development*, 5, 613-619.
 - 23 Marshall, G.R. (1987). Computer-aided drug design. *Annual Review of Pharmacology and Toxicology*, 27, 193-213.
 - 24 Ehrlich, P. (1909). "Titre en allemand". *Deutsch Chem. Ges.*, 42, 17.

-
- 25 **Gund, P.** (1977). Three-dimensional pharmacophoric pattern searching. *Progress in Molecular & Subcellular Biology*, 5, 117-143.
- 26 **Cunningham, B.C., Wells, J.A.** (1993). Comparison of a structural and a functional epitope. *Journal of Molecular Biology*, 234, 554-563.
- 27 **Martin, Y.C.** (1992). 3D database searching in drug design. *Journal of Medicinal Chemistry*, 35(12), 2145-2154.
- 28 **Greene, J., Kahn, S., Savoj, H., Sprague, P., Teig, S.** (1994). Chemical function queries for 3D database search. *Journal of Chemical Information and Computer Science*, 34(6), 1297-1308.
- 29 **Munshi, S., Chen, Z., Li, Y., Olsen, D.B., Fraley, M.E., Frungate, R.W., Kuo, L.C.** (1998). Rapid X-ray diffraction analysis of HIV-1 protease inhibitor complexes : inhibitor exchange in single crystals of the bound enzyme. *Acta Crystallographica Section D*, 54(5), 1053-1060.
- 30 **Kitzinger, C., Benzinger, T.H.** (1960). Principle and method of headburst microcalorimetry and the determination of free energy, enthalpy, and entropy changes. *Methods of Biochemical Analysis*, 8, 309-360.
- 31 **Wells, J.A.** (1991). Systematic mutational analyses of protein-protein interfaces. *Methods in Enzymology*, 202, 390-411.
- 32 **Martin, Y.C., Danaher, E.B., May, C.S., Weininger, D., Van Drie, J.H.** (1989). Strategies in drug design based on 3D-structures of ligands. *Progress in Clinical and Biological Research*, 291, 177-181.
- 33 **Luban, J.** (1996). Absconding with the chaperon : essential cyclophilin-Gag interaction in HIV-virions. *Cell*, 87(7), 1157-1159.
- 34 **Guichou, J.-F., Viaud, J., Mettling, C., Subra, G., Lin, Y.L., Chavanieu, A.** (2006). Structure-based design, synthesis, and biological evaluation of novel inhibitors of human cyclophilin A. *Journal of Medicinal Chemistry*, 49(3), 900-910.
- 35 **Chothia, C., Lesk, A.M.** (1986). The relation between the divergence of sequences and structure in proteins. *EMBO Journal*, 5(4), 823-826.
- 36 **Blundell, T.L., Sibanda, B.L., Sternberg, M.J.E., and Thornton, J.M.** (1987) Knowledge-Based Prediction of Protein Structures and the Design of Novel Molecules. *Nature*, 326, 347-352.
- 37 **Zielenkiewicz, P., Rabczenko, A.** (1984). Protein-protein recognition: method for finding complementarity surfaces of interacting protein. *Journal of Theoretical Biology*, 111(1), 17-30.
- 38 **Santavy, M., Kypr, J.** (1984). A fast computer algorithm for finding an optimum geometrical interaction of two macromolecules. *Journal of Molecular Graphics*, 2, 47-49.
- 39 **Smith G.R., Sternberg M.J.E.** (2002). Prediction of protein interactions by docking methods. *Current Opinion in Structural Biology*, 12, 28-35.
- 40 **Shoichet, B.K., Kuntz, I.D.** (1991). Protein docking and complementarity. *Journal of Molecular Biology*, 221, 327-346.
- 41 **Mendes, J., Guerois, R., Serrano, L.** (2002). Energy estimation in protein design. *Current Opinion in Structural Biology*, 12(4), 441-446.
- 42 **Goodford, P.J.** (1985). A computational procedure for determining energetically binding sites on biologically important macromolecules. *Journal of Medicinal Chemistry*, 28, 849-857.
- 43 **Miranker, A., Karplus, M.** (1991). Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins*, 11(1), 29-34.
- 44 **Eisen, M.B., Wiley, D.C., Karplus, M., Hubbard, R.E.** (1994). HOOK : a program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site. *Proteins*, 19(3), 199-221.
- 45 **von Itzstein, M., Wu, W.Y., Kok, G.B., Pegg, M.S., Dyason, J.C., Jin, B., Van Phan, T., Smythe, M.L., White, H.F., Oliver, S.W., Colman, P.M., Varghese, J.N., Ryan,**

-
- D.M., Woods, J.M., Bethell, R.C., Hotham, V.J., Cameron, J.M. & Penn, C.R. (1993). Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature*, 363, 418-423
- 46 Caflich, A., Wälchli, R., Ehrhardt, C. (1998). Computer-aided design of thrombin inhibitors. *News in Physiological Sciences*, 13, 182-189.
- 47 Willett, P. (2005). Searching techniques for databases of two- and three-dimensional chemical structures. *Journal of Medicinal Chemistry*, 48(13), 4183-4199.
- 48 Clackson, T., Wells, J.A. (1995). A hot spot of binding energy in a hormone-receptor interface. *Science*, 267, 383-386.
- 49 Clackson, T., Ultsch, M.H., Wells, J.A., de Vos, A.M. (1998). Structural and functional analysis of the 1:1 growth hormone:receptor complex reveals the molecular basis for receptor affinity. *Journal of Molecular Biology*, 277(5), 1111-1128.
- 50 Toogood, P.L. (2001). Inhibition of protein-protein association by small molecules : approaches and progress. *Journal of Medicinal Chemistry*, 45(8), 1543-1558.
- 51 Berg, T. (2003). Modulation of protein-protein interactions with small organic molecules. *Angewandte Chemie*, 42(22), 2462-2481.
- 52 Lo Conte, L., Chothia, C., Janin, J. (1999). The atomic structure of protein-protein recognition sites. *Journal of Molecular Biology*, 285, 2177-2198.
- 53 De Lano, W.L., Ultsch, M.H., De Vos, A.M., Wells, J.A. (2000). Convergent solutions to binding at a protein-protein interface. *Science*, 287, 1279-1283.
- 54 Olson, G.L., Bolin, D.R., Bonner, M.P., Bös, M., Cook, C.M., Fry, D.C., Graves, B.J., Hatada, M., Hill, D.E., Kahn, M., Madison, V.S., Rusiecki, V.K., Sarabu, R., Sepinwall, J., Vincent, G.P., Voss, M.E. (1993). Concepts and progress in the development of peptides mimetics. *Journal of Medicinal Chemistry*, 36(21), 3039-3049.
- 55 Pauletti, G.M., Aubé, J., Siahaan, T.J., Gangwar, S., Borchardt, R.T. (1997). Improvement of oral peptide bioavailability: peptidomimetics and prodrug strategies. *Advanced Drug Delivery Reviews*, 27(2-3), 235-256.
- 56 Adessi, C., Soto, C. (2002). Converting a peptide into a drug : strategies to improve stability and bioavailability. *Current Medicinal Chemistry*, 9(9), 963-978.
- 57 Hruby, V.J., Al-Obeidi, F., Kazmierski, W. (1990). Emerging approaches in the molecular design of receptor-selective peptide ligands: conformational, topographical and dynamic considerations. *Biochemical Journal*, 268(2), 249-262.
- 58 Temussi, P.A., Picone, D., Castiglione-Morelli, M.A., Motta, A., Tancredi, T. (1989). Bioactive conformation of linear peptides in solution : an elusive goal?. *Biopolymers*, 28(1), 91-107.
- 59 Saragovi, H.U., Greene, M.I., Chrusciel, R.A., Kahn, M. (1992). Loops and secondary structure mimetics: development and applications in basic science and rational drug design. *Biotechnology*, 10(7), 773-778.
- 60 Hruby, V.J. (2001). Design in topographical space of peptide and peptidomimetic ligands that affect behavior. A chemist's glimpse at the mind-body protein. *Accounts of Chemical Research*, 34(5), 389-397.
- 61 Sia, S.K., Carr, P.A., Cochran, A.G., Malashkevich, V.N., Kim, P.S. (2002). Short constrained peptides that inhibit HIV-1 entry. *Proceedings of the National Academy of Science*, 99(23), 14663-14669.
- 62 Polticelli, F., Raybaudi-Massilia, G., Ascenzi, P. (2001). Structural determinants of mini-protein stability. *Biochemistry and Molecular Biology Education*, 29, 16-20.
- 63 Ménez, A., Bontems, F., Roumestand, C., Gilquin, B., Toma, F. (1992). Structural basis for functional diversity of animal toxins. *Proceedings of the Royal Society of Edimburgh*, 99B, 83-103.

-
- 64 Laskowski, M.Jr, Kato, I. (1980). Protein inhibitors of proteinases. *Annual Review of Biochemistry*, 49, 593-626.
- 65 Le Nguyen, D., Heitz, A., Chiche, L., Castro, B., Boige grain, R.A., Faval, A., Coletti-Previero, M.A. (1990). Molecular recognition between serine proteases and new bioactive microproteins with a knotted structure. *Biochimie*, 72(6-7), 431-435.
- 66 Klug, A., Schwabe, J.W. (1995). Proteins motifs 5. Zinc fingers. *FASEB Journal*, 9(8), 597-604.
- 67 Lanschultz, W.H., Johnson, P.F., McKnight, S.L. (1988). The leucine zipper : a hypothetical common to a new class of DNA binding proteins. *Science*, 240, 1759-1764.
- 68 Torres, A.M., Kini, R.M., Selvanayagam, N., Kuchel, P.W. (2001). NMR structure of buccandin, a neurotoxin from the venom of the Malayan krait (*Bungarus candidus*). *Biochemical Journal*, 360(3), 539-548.
- 69 Parkin, S., Rupp, B., Hope, H. (1996). Structure of bovine pancreatic trypsin inhibitor at 125K definition of carboxy-terminal residues Gly57 and Ala58. *Acta Crystallographica Section D*, 52(1), 18-29.
- 70 Barbault, F., Landon, C., Guenneugues, M., Meyer, J.P., Schott, V., Dimarcq, J.L., Vovelle, F. (2003). Solution structure of Alo-3 : a new knottin-type antifungal peptide from the insect *Acrocinus longimanus*. *Biochemistry*, 42(49), 14434-14442.
- 71 Narayan, V.A., Kriwacki, R.W., Caradonna, J.P. (1997). Structures of zinc finger domains from transcription factor Sp1. insights into sequence-specific protein-DNA recognition. *Journal of Biological Chemistry*, 272(12), 7801-7809.
- 72 O'Shea, E.K., Klemm, J.D., Kim, P.S., Alber, T. (1991). X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science*, 254, 539-544.
- 73 Pissarra, P.N. (2004). Changes in the business of culture. *Nature Biotechnology*, 22(11), 1355-1356.
- 74 Weng, Z., De Lisi, C. (2002). Proteins therapeutics : promises and challenges for the 21st century. *Trends in Biotechnology*, 20(1), 29-35.
- 75 Kurtzman, A.L., Govindarajan, S., Vahle, K., Jones, J.T., Heinrichs, V., Patten, P.A. (2001). Advances in directed protein evolution by recursive genetic recombination : applications to therapeutic proteins. *Current Opinion in Biotechnology*, 12(4), 361-370.
- 76 Vasserot, A.P., Dickinson, C.D., Tang, Y., Huse, W.D., Manchester, K.S., Watkins, J.D. (2003). Optimization of protein therapeutics by directed evolution. *Drug Discovery Today*, 8(3), 118-126.
- 77 Marshall, S.A., Lazar, G.A., Chirino, A.J., Desjarlais, J.R. (2003). Rational design and engineering of therapeutic proteins. *Drug Discovery Today*, 8(5), 212-221.
- 78 Schellekens, H. (2002). Immunogenicity of therapeutic proteins : clinical implications and future prospects. *Clinical Therapeutics*, 24(11), 1720-1740.
- 79 Schellekens, H. (2003). Immunogenicity of therapeutic proteins. *Nephrology, Dialysis, Transplantation*, 18(7), 1257-1259.
- 80 Chirino, A.J., Ary, M.L., Marshall, S.A. (2004). Minimizing the immunogenicity of protein therapeutics. *Drug Discovery Today*, 9(2), 82-90.
- 81 Tangri, S., Mothe, B.R., Eisenbraun, J., Sidney, J., Southwood, S., Briggs, K., Zinckgraf, J., Bilsel, P., Newman, M., Chesnut, R., Licalsi, C., Sette, A. (2005). Rationally engineered therapeutic proteins with reduced immunogenicity. *Journal of Immunology*, 174(6), 3187-3196.
- 82 Levy, F., Johnsson, N., Rumenapf, T., Varshavsky, A. (1996). Using ubiquitin to follow the metabolite fate of a protein. *Proceedings of the National Academy of Science*, 93(10), 4907-4912.
- 83 Varshavsky, A. (1996). The N-end rule : functions, mysteries, uses. *Proceedings of the National Academy of Science*, 93(22), 12142-12149.

-
- 84 Kohler, G., Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature*, 256, 495-497.
- 85 Lazar, G.A., Marshall, S.A., Plecs, J.J., Mayo, S.L., Desjarlais, J.R. (2003). Designing proteins for therapeutic applications. *Current Opinion in Structural Biology*, 13(4), 513-518.
- 86 Martin, L., Vita, C. (2000). Engineering novel bioactive mini-proteins from small size natural and de novo designed scaffolds. *Current Protein and Peptide Science*, 1(4), 403-430.
- 87 Nygren, P.A., Uhlén, M. (1997). Scaffolds for engineering novel binding sites in proteins. *Current Opinion in Structural Biology*, 7(4), 463-469.
- 88 Cunningham, B.C., Wells, J.A. (1997). Minimized proteins. *Current Opinion in Structural Biology*, 7(4), 457-462.
- 89 Struthers, M.D., Cheng, R.P., Imperiali, B. (1996). Design of a monomeric 23-residue polypeptide with defined tertiary structure. *Science*, 271, 342-345.
- 90 Braisted, A.C., Wells, J.A. (1996). Minimizing a binding domain from protein A. *Biochemistry*, 93(12), 5688-5692.
- 91 Starovasnik, M.A., Braisted, A.C., Wells, J.A. (1997). Structural mimicry of a native protein by a minimized binding domain. *Proceedings of the National Academy of Sciences*, 94(19), 10080-10085.
- 92 Nedwiedek, M.N., Hecht, M.H. (1997). Minimized protein structures : a little goes a long way. *Proceedings of the National Academy of Sciences*, 94(19), 10010-10011.
- 93 Pessi, A., Bianchi, E., Crameri, A., Venturini, S., Tramontano, A., Sollazzo, M. (1993). A designed metal-binding protein with a novel fold. *Nature*, 362, 367-369.
- 94 Quiocho, F.A. (1993). Making of the minibody. *Nature*, 362, 293-294.
- 95 Martin, F., Toniatti, C., Salvati, A.L., Venturini, S., Ciliberto, G., Cortese, R., Sollazzo, M. (1994). The affinity selection of a minibody polypeptide inhibitor of human interleukin-6. *EMBO Journal*, 13(22), 5303-5309.
- 96 Dimasi, N., Martin, F., Volpari, C., Brunetti, M., Biasiol, G., Altamura, S., Cortese, R., De Francesco, R., Steinkuhler, C., Sollazzo, M. (1997). Characterization of engineered hepatitis C virus NS3 protease inhibitors affinity selected from human pancreatic trypsin inhibitor and minibody repertoires. *Journal of Virology*, 71(10), 7461-7469.
- 97 Hu, S., Shively, L., Raubitschek, A., Sherman, M., Williams, L.E., Wong, J.Y., Shively, J.E., Wu, A.M. (1996). Minibody : a novel engineered anti-carcinoembryonic antigen antibody fragment (single-chain Fv-CH3) which exhibits rapid, high-level targeting of xenografts. *Cancer Research*, 56(13), 3055-3061.
- 98 Carter, P., Merchant, A.M. (1997). Engineering for imaging and therapy. *Current Opinion in Biotechnology*, 8(4), 449-454.
- 99 Wu, A.M., Yazaki, P.J., Tsai, S., Nguyen, K., Anderson, A.L., McCarthy, D.W., Welch, M.J., Shively, J.E., Williams, L.E., Raubitschek, A.A., Wong, J.Y., Toyokuni, T., Phelps, M.E., Gambhir, S.S. (2000). High-resolution microPET imaging of carcinoembryonic antigen-positive xenografts by using a copper-64-labeled engineered antibody fragment. *Proceedings of the National Academy of Sciences*, 97(15), 8495-8500.
- 100 Mutter. (1988). Nature's rules and chemist's tools : a way for creating novel proteins. *Trends in Biochemical Science*, 13(7), 260-265.
- 101 Tuchscherer, G., Mutter, M. (1996). Template assisted protein *de novo* design. *Pure & Applied Chemistry*, 68(11), 2153-2162.
- 102 Mutter, M., Hersperger, R., Gubernator, K., Muller, K. (1989). The construction of new proteins : V. A template-assembled synthetic protein (TASP) containing a 4-helix bundle and beta-barrel-like structure. *Proteins*, 5(1), 13-21.
- 103 Fernandez-Carneado, J., Grell, D., Durieux, P., Hauert, J., Kovacsics, T., Tuchscherer, G. (2000). Surface grafting onto template-assembled synthetic protein scaffolds in molecular recognition. *Biopolymers*, 55(6), 451-458.

-
- 104 Tuchscherer, G., Servis, G., Corradin, G., Blum, U., Rivier, J., Mutter, M. (1992). Total chemical synthesis, characterization, and immunological properties of an MHC class I model using the TASP concept for protein *de novo* design. *Protein Science*, 1(10), 1377-1386.
- 105 Tuchscherer, G., Scheibler, L., Dumy, P., Mutter, M. (1998). Protein design: on the threshold of functional properties. *Biopolymers*, 47(1), 63-73.
- 106 Beasley, J.R., Hecht, M.H. (1997). Protein design : the choice of *de novo* sequences. *Journal of Biological Chemistry*, 272(4), 2031-2034.
- 107 West, M.W., Wang, W., Patterson, J., Mancias, J.D., Beasley, J.R., Hecht, M.H. (1999). De novo amyloid proteins from designed combinatorial libraries. *Proceedings of the National Academy of Sciences*, 96(20), 11211-11216.
- 108 Wei, Y., Liu, T., Sazinsky, S.L., Moffet, D.A., Pelczer, I., Hecht, M.H. (2003). Stably folded *de novo* proteins from a designed combinatorial library. *Protein Science*, 12(1), 92-102.
- 109 Wei, Y., Hecht, M.H. (2004). Enzyme-like proteins from an unselected library of designed amino acid sequences. *Protein Engineering, Design & Selection*, 17(1), 67-75.
- 110 Das, A., Trammell, S.A., Y., Hecht, M.H. (2006). Electrochemical and ligand binding studies of a *de novo* heme protein. *Biophysical Chemistry*, 123(2-3), 102-112.
- 111 Drakopoulou, E., Zinn-Justin, S., Guennegues, M., Gilquin, B., Ménez, A., Vita, C. (1996). Changing the structural context of a functional β -hairpin. Synthesis and characterisation of a chimera containing the curamimetic loop of a snake toxin in the scorpion α/β scaffold. *Journal of Biological Chemistry*, 271(20), 11979-11987.
- 112 Zinn-Justin, S., Guennegues, Drakopoulou, E., Gilquin, B., Vita, C., Ménez, A. (1996). Transfer of a β -hairpin from a functional site of snake curamimetic to a α/β scaffold of scorpion toxins : three-dimensional solution structure of the chimeric protein. *Biochemistry*, 35(26), 8535-8543.
- 113 Butcher, D.J., Kowalska, M.A., Li, S., Luo, Z., Shan, S., Lu, Z., Niewiarowski, S., Huang, Z. (1997). A natural motif approach to protein design: a synthetic leucine zipper peptide mimics the biological function of platelet factor 4 protein. *FEBS Letters*, 409, 183-187.
- 114 Domingues, H., Cregut, D., Sebald, W., Oschkinat, H., Serrano, L. (1999). Rational design of a CGN4-derived mimetic of interleukin-4. *Nature Structural Biology*, 6(7), 652-656.
- 115 Kritzer, J.A., Zutshi, R., Mingtatt, C., Ran F.A., Webman, R., Wongjirad, T.M., Schepartz, A. (2006). Miniature protein inhibitors of the p53-hDM2 interaction. *ChemBioChem*, 7(1), 29-31.
- 116 Blundell, T.L., Pitts, J.E., Tickle, I.J., Wood, S.P., Wu, C.W. (1981). X-ray analysis (1.4-angstrom resolution) of avian pancreatic polypeptide : small globular protein hormone. *Proceedings of the National Academy of Sciences*, 78(7), 4175-4179.
- 117 Chin, J.W., Schepartz, A. (2001). Design evolution of a miniature Bcl-2 binding protein. *Angewandte Chemie (International Edition in English)*, 40(20), 3806-3809.
- 118 Montclare, J.K., Schepartz, A. (2003). Miniature homeodomains : high specificity without an N-terminal arm. *Journal of American Chemical Society*, 125(12), 3416-3417.
- 119 Zondlo, N.J., Schepartz, A. (1999). Highly specific DNA recognition by a designed miniature protein. *Journal of American Chemical Society*, 121(29), 6938-6939.
- 120 Chin, J.W., Grotzfeld, R.M., Fabian, M.A., Schepartz, A. (2001). Methodology for optimizing functional miniature proteins based on avian pancreatic polypeptide using phage display. *Bioorganic & Medicinal Chemistry Letters*, 11(12), 1501-1505.
- 121 Chin, J.W., Schepartz, A. (2001). Concerted evolution of structure and function in a miniature protein. *Journal of American Chemical Society*, 123(12), 2929-2930.
- 122 Smith, D.A., Tachias, K., Madison, E.L. (1995). Protein loop grafting to construct a variant of tissue-type plasminogen activator that binds platelet integrin $\alpha_{IIb}\beta_3$ *. *Journal of Biological Chemistry*, 270(51), 30486-30490.

-
- 123 Vita, C., Roumestand, C., Toma, F., Ménez, A. (1995). Scorpion toxins as natural scaffolds for protein engineering. *Proceedings of the National Academy of Sciences*, 92(14), 6404-6408.
- 124 Vita, C., Vizzanova, J., Drakopoulou, E., Zinn-Justin, S., Gilquin, B., Ménez, A. (1998). Novel miniproteins engineered by the transfer of active sites to small natural scaffolds. *Biopolymers*, 47(1), 93-100.
- 125 Vita, C., Drakopoulou, E., Vizzanova, J., Rochette, S., Martin, L., Ménez, A., Roumestand, C., Yang, Y.S., Ylisastigui, L., Benjouad, A., Gluckman, J.C. (1999). Rational engineering of a miniprotein that reproduces the core of the CD4 site interacting with HIV-1 envelope glycoprotein. *Proceedings of the National Academy of Science*, 96(23), 13091-13096.
- 126 Martin, L., Stricher, F., Missé, D., Sironi, F., Pugnère, M., Barthe, P., Prado-Gotor, R., Freulon, I., Magne, X., Roumestand, C., Ménez, A., Lusso, P., Veas, F., Vita, C. (2003). Rational design of a CD4 mimic that inhibits HIV-1 entry and exposes cryptic neutralization epitopes. *Nature Biotechnology*, 21, 71-76.
- 127 Huang, C.C., Stricher, F., Martin, L., Decker, J.M., Majeed, S., Barthe, P., Hendrickson, W.A., Robinson, J., Roumestand, C., Sodroski, J., Wyatt, R., Shaw, G.M., Vita, C., Kwong, P.D. (2005). Scorpion-toxin mimics of CD4 in complex with human immunodeficiency virus gp120 crystal structures, molecular mimicry, and neutralization breadth. *Structure*, 13(5), 755-768.
- 128 Mer, G., Kellenberger, E., Lefèvre, J.F. (1998). α -helix mimicry of a β -turn. *Journal of Molecular Biology*, 281, 235-240.
- 129 Kellenberger, E., Mer, G., Kellenberger, C., Marguerie, G., Lefèvre, J.F. (1999). Solution structure of a conformationally constrained Arg-Gly-Asp-like motif inserted into the α/β scaffold of leurotoxin I. *European Journal of Biochemistry*, 260(3), 810-817.
- 130 Chothia, C., Janin, J. (1975). Principles of protein-protein recognition. *Nature*, 256, 705-708.
- 131 Janin, J., Chothia, C. (1990). The structure of protein-protein recognition sites. *Journal of Biological Chemistry*, 265(27), 16027-16030.
- 132 Jackson, R.M. (1999). Comparison of protein-protein interactions in serine protease-inhibitor and antibody-antigen complexes: implications for the protein docking problem. *Protein Science*, 8(3), 603-613.
- 133 Ofra, Y., Rost, B. (2003). Analysing six types of protein-protein interfaces. *Journal of Molecular Biology*, 325(2), 377-387.
- 134 Nooren, I.M.A., Thornton, J.M. (2003). Diversity of protein-protein interactions. *EMBO Journal*, 22(14), 3486-3492.
- 135 Betts, M.J., Sternberg, M.J.E. (1999). An analysis of conformational changes in protein-protein association: implications predictive docking. *Protein Engineering*, 12(4), 271-283.
- 136 Lee, B., Richards, F.M. (1971). The interpretation of protein structures: Estimation of static accessibility. *Journal of Molecular Biology*, 55, 379-400.
- 137 Connolly, M.L. (1983). Solvent-accessible surfaces of proteins and nucleic acids. *Science*, 221(1), 709-713.
- 138 Chakabarti, P., Janin, J. (2002). Dissecting protein-protein recognition sites. *Proteins*, 47(3), 334-343.
- 139 Jones, S., Thornton, J.M. (1996). Principles of protein-protein interactions. *Proceedings of the National Academy of Sciences*, 93(1), 13-20.
- 140 Lawrence, M.C., Colman, P.M. (1993). Shape complementarity at protein/protein interfaces. *Journal of Molecular Biology*, 234(4), 946-950.
- 141 Richards, F.M. (1974). The interpretation of protein structures: total volume, group volume distributions and packing density. *Journal of Molecular Biology*, 82(1), 1-14.

-
- 142 Laskowski, R.A. (1995). SURFNET: a program for visualizing molecular surfaces, cavities and intermolecular interactions. *Journal of Molecular Graphics*, 13(5), 307-308.
- 143 Castro, M.J., Anderson, S. (1996). Alanine point-mutations in the reactive region of bovine pancreatic trypsin inhibitor. Effects on the kinetics and thermodynamics of binding to β -trypsin and α -chymotrypsin. *Biochemistry*, 35, 11435-11446.
- 144 Bogan, A.A., Thorn, K.S. (1998). Anatomy of hot spots in protein interfaces. *Journal of Molecular Biology*, 280(1), 1-9.
- 145 Fernandez, A., Scott L.R., Scheraga H.A. (2003). Amino-acid residues at protein-protein interfaces: why is propensity so different from relative abundance? *Journal of Physical Chemistry*, 107, 9929-9932.
- 146 Bairoch, A., Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement in TrEMBL in 2000. *Nucleic Acids Research*, 28(1), 45-48.
- 147 Grishin, N.V., Phillips, M.A (1994). The subunit interfaces of oligomeric enzymes are conserved to a similar extent to the overall protein sequences. *Protein Science*, 3(12), 2455-2458.
- 148 Ouzounis, C., Perez-Irratxeta, C., Sander, C., Valencia, A. (1998). Are binding residues conserved? *Pacific Symposium on Biocomputing*, 401-412.
- 149 Bartlett, G.J., Porter, C.T., Borkakoti, N., Thornton, J.M. (2002). Analysis of catalytic residues in enzyme active sites. *Journal of Molecular Biology*, 324(1), 105-121.
- 150 Valdar, W.S., Thornton, J.M. (2001). Conservation helps to identify biologically relevant crystal contacts. *Journal of Molecular Biology*, 313(2), 399-416.
- 151 Shueler-Furman, O., Baker, D. (2003). Conserved residue clustering and protein structure prediction. *Protein*, 52(2), 225-235.
- 152 Ma, B., Elkayam, T., Wolfson, H., Nussinov, R. (2003). Protein-protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proceedings of the National Academy of Sciences*, 100(10), 5772-5777.
- 153 Caffrey, D.R., Somaroo, S., Hughes, J.D., Mintseris, J., Huang, E.S. (2004). Are protein-protein interfaces more conserved in sequence than the rest of the protein surface. *Protein Science*, 13(1), 190-202.
- 154 Hu, Z., Ma, B., Wolfson, H., Nussinov, R. (2000). Conservation of polar residues as hot spots at protein interface. *Proteins*, 39(4), 331-342.
- 155 Stanfield, R.L., Wilson, I.A. (1994). Antigen-induced conformational changes in antibodies : a problem for structural prediction and design. *Trends in Biotechnology*, 12(7), 275-279.
- 156 Gerstein, M., Lesk, A.M., Chothia, C. (1994). Structural mechanisms for domain movements in proteins. *Biochemistry*, 33(22), 6739-6749.
- 157 Burley, S.K., Petsko, G.A. (1988). Weakly polar interactions in proteins. *Advances in Protein Chemistry*, 35, 125-189.
- 158 Sheinerman, F.B., Norel, R., Honig, B. (2000). Electrostatic aspects of protein-protein interactions. *Current Opinion in Structural Biology*, 10, 153-159.
- 159 Xu, D., Tsai, C.J., Nussinov, R. (1997). Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein Engineering*, 10(9), 999-1012.
- 160 Wells, J.A., Powers, D.B., Bott, R.R., Graycar, T.P., Estell, D.A. (1987). Designing substrate specificity by protein engineering of electrostatic interactions. *Proceedings of the National Academy of Sciences*, 84(5), 1219-1223.
- 161 Kumar, S., Nussinov, R. (2002). Close-range electrostatic interactions in proteins. *ChemBioChem*, 3(7), 604-617.
- 162 Tsai, C.J., Lin, S.L., Wolfson, H.J., Nussinov, R. (1997). Studies of protein-protein interfaces: a statistical analysis of the hydrophobic effect. *Protein Science*, 6(1), 53-64.

-
- 163 Janin, J. (1999). Wet and dry interfaces: the role of solvent in protein-protein and protein-DNA recognition. *Structure*, 7(12), R277-R279.
- 164 Rodier, F., Bahadur, R.P., Chakrabarti, P., Janin, J. (2005). Hydration of protein-protein interfaces. *Proteins: Structure, Function and Bioinformatics*, 60(1), 36-45.
- 165 Fernandez, A., Scheraga, H.A. (2003). Insufficiently dehydrated hydrogen bonds as determinants of protein interactions. *Proceedings of the National Academy of Sciences*, 100(1), 113-118.
- 166 Crowley, P.B., Golovin, A. (2005). Cation- π interactions in protein-protein interfaces. *Proteins*, 59(2), 231-239.
- 167 Cunningham, B.C., Wells, J.A. (1991). Rational design of receptor-specific variants of human growth hormone. *Proceedings of the National Academy of Sciences*, 88(8), 3407-3411.
- 168 Cunningham, B.C., Wells, J.A. (1993). Comparison of a structural and a functional epitope. *Journal of Molecular Biology*, 237(4), 554-563.
- 169 Pineda, A.O., Cantwell, A.M., Bush, L.A., Rose, T., Di Cera, E. (2002). The thrombin epitope recognizing thrombomodulin is a highly cooperative hot spot in exosite I. *Journal of Biological Chemistry*, 277(35), 32015-32019.
- 170 Massova, I., Kollman, P.A. (1999). Computational alanine scanning to probe protein-protein interactions: a novel approach to evaluate binding free energies. *Journal of American Chemical Society*, 121(36), 8133-8143.
- 171 Kortemme, T., Baker, D. (2002). A simple physical model for binding energy hot spots in protein-protein complexes. *Proceedings of the National Academy of Sciences*, 99(22), 14116-14121.
- 172 Kim, D.E., Chivian, D., Baker, D. (2004). Protein prediction and analysis using the Robetta server. *Nucleic Acids Research*, 32, W526-W531.
- 173 Kortemme, T., Kim, D.E., Baker, D. (2004). Computational alanine scanning of protein-protein interfaces. *Science's SKTE*, 2004(219), p12.
- 174 Honig, B., Nicholls, A. (1995). Classical electrostatics in biology and chemistry. *Science*, 268, 1144-1149.
- 175 Gilson, M.K. (1995). Theory of electrostatic interactions in macromolecules. *Current Opinion in Structural Biology*, 5(2), 216-223.
- 176 Janin, J. (1997). The kinetics of protein-protein recognition. *Proteins*, 28(2), 153-161.
- 177 Wade, R.C., Gabdoulina, R.R., Lüdemann, S.K., Lounnas, V. (1998). Electrostatic steering and ionic tethering in enzyme-ligand binding : insights from simulations. *Proceedings of the National Academy of Science*, 95(11), 5942-5949.
- 178 Selzer, T., Schreiber, G. (2001). New insights into the mechanism of protein-protein association. *Proteins*, 45(3), 190-198.
- 179 Sheinerman, F.B., Honig, B. (2002). On the role of electrostatic interactions in the design of protein-protein interfaces. *Journal of Molecular Biology*, 318(1), 161-177.
- 180 Botti, S.A., Felder, C.E., Sussman, J.L., Silman, I. (1998). Electrotactins: a class of adhesion proteins with conserved electrostatic and structural motifs. *Protein Engineering*, 11(6), 415-420.
- 181 Felder, C.E., Botti, S.A., Lifson, S., Silman, I., Sussman, J.L. (1997). External and internal electrostatic potentials of cholinesterase models. *Journal of Molecular Graphics & Modelling*, 15(5), 335-337.
- 182 Livesay, D.R., Jambeck, P., Rojnuckarin, A., Subramaniam, S. (2003). Conservation of electrostatic within enzyme families and superfamilies. *Biochemistry*, 42(12), 3464-3473.
- 183 Mc Coy, A.J., Chandana Epa, V., Colman, P.M. (1997). Electrostatic complementarity at protein/protein interfaces. *Journal of Molecular Biology*, 268(2), 570-584.
- 184 Ullmann, G.M., Hauswald, M., Jensen, A., Kostic, N.M., Knapp, E.W. (1997). Comparison of the physiologically equivalent proteins cytochrome c_6 and plastocyanin on the basis

of their electrostatic potentials. Tryptophan 63 in cytochrome c_6 may be isofunctional with tyrosine 83 in plastocyanin. *Biochemistry*, 36, 16187-16196.

185 Ullmann, G.M., Hauswald, M., Jensen, A., Knapp, E.W. (2000). Structural alignment of ferredoxin and flavodoxin based on electrostatic potentials: implication for their interactions with photosystem I and ferredoxin-NADP reductase. *Protein: Structure, Function, and Genetics*, 38, 301-309.

186 Wade, R.C., Gabdoulline, R.R., Luty, B.A. (1998). Species dependence of enzyme-substrate encounter rates for triose phosphate isomerase. *Proteins: Structure, Function, and Genetics*, 31, 406-416.

187 Blomberg, N., Gabdoulline, R.R., Nilges, M., Wade, R.C. (1999). Classification of protein sequences by homology modelling and quantitative analysis of electrostatic similarity. *Proteins: Structure, Function, and Genetics*, 37, 379-387.

188 De Rienzo, F., Gabdoulline, R.R., Menziani, M.C., Wade, R.C. (2000). Blue copper proteins: a comparative analysis of their molecular interaction properties. *Protein Science*, 9, 1439-1454.

189 Greenwald, J., Fischer, W.H., Vale, W.W., Choe, S. (1999). Three-finger toxin fold for the extracellular ligand-binding domain of the type II activin receptor serine kinase. *Nature Structural Biology*, 6(1), 18-22.

190 Sun, Y.J., Wu, W.G., Chiang, C.M., Hsin, A.Y., Hsiao, C.D. (1997). Crystal structure of cardiotoxin V from Taiwan cobra venom : pH-dependent conformational change and a novel membrane-binding motif identified in the three-finger loops of P-type cardiotoxin. *Biochemistry*, 36(9), 2403-2413.

191 Lou, X., Liu, Q., Tu, X., Wang, J., Teng, M., Niu, L., Schuller, D.J., Huang, Q., Hao, Q. (2004). The atomic resolution crystal structure of atratoxin determined by single wavelength anomalous diffraction phasing. *Journal of Biological Chemistry*, 279(37), 39094-39104.

192 Llinas, P., Le Du, M.H., Gardsvoll, H., Dano, K., Ploug, M., Gilquin, B., Stura, F.A., Menez, A. (2005). Crystal structure of the human urokinase plasminogen activator bound to an antagonist peptide. *EMBO Journal*, 24(9), 1655-1663.

193 Pawlak, J., Mackessy, S.P., Fry, B.G., Bhatia, M., Mourier, G., Fruchart-Gaillart, C., Servent, D., Menez, R., Stura, F., Menez, A., Kini, R.M. (2006). Denmotoxin, a three-finger toxin from the Colubrid snake *Boiga dendrophila* (Mangrove catsnake) with bird-specific activity. *Journal of Biological Chemistry*, 281(39), 29030-29041.

194 Momany, C., Kovari, L.C., Prongay, A.J., Keller, W., Gitti, R.K., Lee, B.M., Gorbalenya, A.E., Tong, L., McClure, J., Ehrlich, L.S., Summers, M.F., Carter, C., Rossmann, M.G. (1996). Crystal structure of dimeric HIV-1 capsid protein. *Nature Structural Biology*, 3(9), 763-770.

195 Hahn, M., Winkler, D., Welfle, K., Misselwitz, R., Welfle, H., Wessner, H., Zahn, G., Scholz, C., Seifert, M., Harkins, R., Schneider-Mergener, J., Hohne, W. (2001). Cross-reactive binding of a cyclic peptides to an anti-TGF alpha antibody Fab fragment : an X-ray structural and thermodynamic analysis. *Journal of Molecular Biology*, 314(2), 293-309.

196 Mirza, O., Henriksen, A., Ipsen, H., Larsen, J.N., Wissenbach, M., Spangfort, M.D., Gajhede, M. (2000). Dominant epitopes and allergic cross-reactivity : complex formation between a Fab fragment of a monoclonal murine IgG antibody and the major allergen from birch pollen Bet v1. *Journal of Immunology*, 165(1), 331-338.

197 Mylvaganam, S.E., Paterson, Y., Getzoff, E.D. (1998). Structural basis for the binding of an anti-cytochrome c antibody to its antigen : crystal structures of FabE8-cytochrome c complex to 1.8 Å resolution and FabE8 to 2.26 Å resolution. *Journal of Molecular Biology*, 281(1), 301-322.

-
- 198 Tsai, C.J., Xu, D., Nussinov, R. (1997). Structural motifs at protein-protein interfaces: protein cores versus two-state and three-state model complexes. *Protein Science*, 6(9), 1793-1805.
- 199 Jin, L., Wells, J.A. (1994). Dissecting the energetics of an antibody-antigen interface by alanine shaving and molecular grafting. *Protein Science*, 3(12), 2351-2357.
- 200 Dauplais, M., Lecoq, A., Song, J., Cotton, J., Jamin, N., Gilquin, B., Roumestand, C., Vita, C., De Medeiros, C.L.C., Rowan, E.G., Harvey, A.L., Ménez, A. (1997). On the convergent evolution of animal toxins. Conservation of a diad of functional residues in potassium channel-blocking toxins with unrelated structures. *Journal of Biological Chemistry*, 272(7), 4302-4309.
- 201 Lytle, B.L., Volkman, B.F., Westler, W.M., Heckman, M.P., Wu, J.H. (2001). Solution structure of a type I dockerin domain, a novel prokaryotic, extracellular calcium-binding domain. *Journal of Molecular Biology*, 307(3), 745-753.
- 202 Kleywegt, G.J. (1997). Recognition of spatial motifs in protein structures. *Journal of Molecular Biology*, 285(4), 1887-1897.
- 203 Baker, N.A., Sept, D., Holst, M.J., Mc Cammon, J.A. (2001). Electrostatics of nanosystems : application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences*, 98(18), 10037-10041.
- 204 Wade, R.C., Gabdoulhine, R.R., De Rienzo, F. (2001). Protein interaction property similarity analysis. *International Journal of Quantum Chemistry*, 83, 122-127.
- 205 Rai, R., Aravinda, S., Kanagarajadurai, K., Raghothama, S., Shamala, N., Balaram, P. (2006). Diproline templates as folding nuclei in designed peptides. Conformational analysis of synthetic peptide helices containing amino terminal Pro-Pro segments. *Journal of American Chemical Society*, 128(24), 7916-7928.
- 206 Pahlke, D., Freund, C., Leitner, D., Labudde, D. (2005). Statistically significant dependence of Xaa-Pro peptide bond conformation on secondary structure and amino acid sequence. *BMC structural Biology*, 5, 1-8.
- 207 Yuan, X., Werner, J.M., Knott, V., Handford, P.A., Campbell, I.D., Downing, K. (1998). Effects of proline cis-trans isomerization on TB domain secondary structure. *Protein Science*, 7(10), 2127-2135.
- 208 Chou, P.Y., Fasman, G.D. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. *Advances in Enzymology and Related Areas of Molecular Biology*, 47, 45-148.
- 209 Kleywegt, G.J. (1997). Les amis d'O. *CCP4/ESF-EACBM Newsletter on Protein Crystallography*, 34, 5-8.
- 210 Harvey, A.L., Vatanpour, H., Rowan, E.G., Pinkasfeld, S., Vita, C., Ménez, A., Martin-Eauclaire, M.F. (1995). Structure-activity studies on scorpion toxins that block potassium channels. *Toxicon*, 33(4), 425-436.
- 211 Huys, I., Dyason, K., Waelkens, E., Verdonck, F., van Zyl, J., du Plessis, J., Müller, G.J., van der Walt, J., Clynen, E., Schoofs, L., Tytgat, J. (2002). Purification, characterization and biosynthesis of parabutoxin 3, a component of *Parabuthus transvaalicus* venom. *European Journal of Biochemistry*, 269(7), 1854-1865.
- 212 Jacobsen, R., Koch, E.D., Lange-Malecki, B., Stocker, M., Verhey, J., Van Wagoner, R.M., Vyazovkina, A., Olivera, B.M., Terlau, H. (2000). Single amino acid substitution in κ -conotoxin PVIIA disrupt interaction with the *shaker* K⁺ channel. *Journal of Biological Chemistry*, 275(32), 24639-24644.
- 213 Schweitz, H., Bruhn, T., Guillemare, E., Moinier, D., Lancelin, J.M., Béress, L., Lazdunski, M. (1995). Two different classes of sea anemone toxins for voltage-sensitive K⁺ channels. *Journal of Biological Chemistry*, 270(42), 25121-25126.

- 214 Aneiros, A., Garcia, I., Martinez, J.R., Harvey, A.L., Anderson, A.J., Marshall, D.L., Engstrom, A., Hellman, U., Karlsson, E. (1993). A potassium channel toxin from the secretion of the sea anemone *bunodosoma granulifera*. Isolation, amino acid sequence and biological activity. *Biochimica et Biophysica Acta*, 1157(1), 86-92.
- 215 Gutman, G.A., Chandy, K.G., Adelman, J.P., Aiyar, J., Bayliss, D.A., Clapham, D.E., Covarriubias, M., Desir, G.V., Furiuchi, K., Ganetzky, B., Garcia, M.L., Grissmer, S., Jan, L.Y., Karschin, A., Kim, D., Kuperschmidt, S., Kurachi, Y., Lazdunski, M., Lesage, F., Lester, H.A., McKinnon, D., Nichols, C.G., O'Kelly, I., Robbins, J., Robertson, G.A., Rudy, B., Sanguinetti, M., Seino, S., Stuehmer, W., Tamkun, M.M., Vandenberg, C.A., Wei, A., Wulff, H., Wymore, R.S. (2003). International Union of Pharmacology. XLI. Compendium of voltage-gated ion channels: potassium channels. *Pharmacological Reviews*, 55(4), 583-586.
- 216 Gasparini, S., Gilquin, B., Ménez, A. (2004). Comparison of sea anemone and scorpion toxins binding to Kv1 channels: an example of convergent evolution. *Toxicon*, 43(8), 901-908.
- 217 Gilquin, B., Racapé, J., Wrisch, A., Visan, V., Lecoq, A., Grissmer, S., Ménez, A., Gasparini, S. (2002). Structure of the BgK-Kv1.1 complex based on distance restraints identified by double mutant cycles. *Journal of Biological Chemistry*, 277(40), 37406-37413.
- 218 Gilquin, B., Braud, S., Eriksson, M.A., Roux, B., Bailey, T.D., Priest, B.T., Garcia, M.L., Ménez, A., Gasparini, S. (2005). A variable residue in the pore of Kv1 channels is critical for the high affinity of blockers from sea anemones and scorpions. *Journal of Biological Chemistry*, 280(29), 27093-27102.
- 219 Doyle, D.A., Cabral, J.M., Pfuetzner, R.A., Kuo, A., Gulbis, J.M., Cohen, S.L., Chait, B.T., MacKinnon, R. (1998). The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. *Science*, 280, 69-77.
- 220 Long, S.B., Campbell, E.B., Mac Kinnon, R. (2005). Crystal structure of a mammalian voltage-dependent shaker family K⁺ channel. *Science*, 309(5736), 903-908.
- 221 Gasparini, S., Danse, J.M., Lecoq, A., Pinkasfeld, S., Zinn-Justin, S., Young, L.C., de Medeiros, C.C., Rowan, E.G., Harvey, A.L., Ménez, A. (1998). Delineation of the functional site of alpha-dendrotoxin. The functional topographies of dendrotoxins are different but share a core with those of other Kv1 potassium channel-blocking toxins. *Journal of Biological Chemistry*, 273(39), 25393-15403.
- 222 Krezel, A.M., Kasibhatla, C., Hidalgo, P., Mac Kinnon, R., Wagner, G. (1995). Solution structure of the potassium channel inhibitor agitoxin 2 : caliper for probing channel geometry. *Protein Science*, 4(8), 1478-1489.
- 223 Bontems, F., Gilquin, B., Roumestand, C., Ménez, A., Toma, F. (1992). Analysis of side-chain organization on a refined model of charybdotoxin : structural and functional implications. *Biochemistry*, 31(34), 7756-7764.
- 224 Cotton, J., Crest, M., Bouet, F., Alessandri, N., Gola, M., Forest, E., Karlsson, E., Castañeda, O., Harvey, A.L., Vita, C., Ménez, A. (1997). A potassium channel acid sequence, disulfide-bridge assignment, chemical synthesis, and biological activity. *European Journal of Biochemistry*, 244(1), 192-202.
- 225 Alessandri-Haber, N., Lecoq, A., Gasparini, S., Grangier-Macmath, G., Jacquet, G., Harvey, A.L., De Medeiros, C., Gola, M., Ménez, A., Crest, M. (1999). Mapping the functional anatomy of BgK on Kv1.1, Kv1.2 and Kv1.3. Clues to design analogs with enhanced selectivity. *Journal of Biological Chemistry*, 274(50), 35653-35661.
- 226 Yu, L., Sun, C., Song, D., Xu, N., Gunasekera, A., Hadjuk, P.J., Olejniczak, E.T. (2005). Nuclear magnetic resonance structural studies of a potassium channel-charybdotoxin complex. *Biochemistry*, 44, 15834-15841.

- 227 Rester, U., Bode, W., Moser, M., Parry, M.A., Huber, R., Auerswald, E. (1999). Structure of the complex of the antistatin-type inhibitor bdellastasin with trypsin and modelling of the bdellastasin-microplasmin system. *Journal of Molecular Biology*, 293, 93-106.
- 228 Song, H.Y., Suh, S.W. (1998). Kunitz-type soybean trypsin inhibitor revisited: refined structure of its complex with porcine trypsin reveals an insight into the interaction between a homologous inhibitor from *Erythrina caffra* and tissue-type plasminogen activator. *Journal of Molecular Biology*, 275, 347-363.
- 229 Frigerio, F., Coda, A., Pugliese, L., Lionetti, C., Menegatti, E., Amiconi, G., Schnebli, H.P., Ascenzi, P., Bolognesi, M. (1992). Crystal and molecular structure of the bovine α -chymotrypsin-eglin c complex at 2.0Å resolution. *Journal of Molecular Biology*, 225, 107-123.
- 230 Koepke, J., Ermler, U., Warkentin, E., Wenzl, G., Flecker, P. (2000). Crystal structure of cancer chemopreventive Bowman-Birk inhibitor in ternary complex with bovine trypsin at 2.3Å resolution. Structural basis of Janus-faced serine protease inhibitor specificity. *Journal of Molecular Biology*, 298, 477-491.
- 231 Rester, U., Moser, M., Huber, R., Bode, W. (2000). L-isoaspartate 115 of porcine β -trypsin promotes crystallisation of its complex with bdellastasin. *Acta Crystallographica Section D*, 56, 581-588.
- 232 Scheidig, A.J., Hynes, T.R., Pelletier, L.A., Wells, J.A., Kossiakoff, A.A. (1997). Crystal structure of bovine chymotrypsin and trypsin complexed to the inhibitor domain of Alzheimer's amyloid β -protein precursor (APPI) and basic pancreatic trypsin inhibitor: engineering of inhibitors with altered specificities. *Protein Science*, 6, 1806-1824.
- 233 Zhu, G., Huang, Q., Zhu, Y., Li, Y., Chi, C., Tang, Y. (2001). X-RAY structure study on an artificial mung bean inhibitor complex with bovine β -trypsin in neat cyclohexane. *Biochimica et Biophysica Acta*, 1546, 98-106.
- 234 Kraetzner, R., Wentzel, A., Kolmar, H., Uson, I. (To be published). Complex of Eeti-II mutant with porcine trypsin.
- 235 Wu, G., Huang, Y., Zhu, G., Huang, Q., Tang, Y., Miyake, H., Kusunoli, M. (To be published). Crystal structure of the bovine trypsin complex with a synthetic 11 peptide inhibitor.
- 236 Stubbs, M.T., Morenweiser, R., Stürzebecher, J., Bauer, M., Bode, W., Huber, R., Piechotka, G.P., Matschiner, G., Sommerhoff, C.P., Fritz, H., Auerswald, E.A. (1997). The three-dimensional structure of recombinant leech-derived trypsin inhibitor in complex with trypsin. *Journal of Biological Chemistry*, 272(32), 19931-19937.
- 237 Hecht, H.J., Szardenings, M., Collins, J., Schomburg, D. (1991). Three-dimensional structure of the complexes between bovine chymotrypsinogen A and two recombinant variants of human pancreatic secretory trypsin inhibitor (kazal-type). *Journal of Molecular Biology*, 220, 711-722.
- 238 Lockett, S., Garcia, R.S., Barker, J.J., Konarev, A.V., Shewry, P.R., Clarke, A.R., Brady, R.L. (1999). High resolution structure of a potent cyclic protease inhibitor from sunflower seeds. *Journal of Molecular Biology*, 290(2), 525-533.
- 239 Huang, Q., Liu, S., Tang, Y. (1993). The refined 1.6 angstroms resolution crystal structure of the complex formed between porcine beta-trypsin and mcti-a, a trypsin inhibitor of squash family. *Journal of Molecular Biology*, 229, 1022-1036.
- 240 Fujinaga, M., Sielecki, A.R., Read, R.J., Ardelt, W., Laskowski, M.Jr., James, M.N. (1987). Crystal and molecular structures of the complex of α -chymotrypsin with its inhibitor turkey ovomucoid third domain at 1.8Å resolution. *Journal of Molecular Biology*, 195, 397-418.
- 241 Scheidig, A.J., Hynes, T.R., Pelletier, L.A., Wells, J.A., Kossiakoff, A.A. (1997). Crystal structures of bovine chymotrypsin and trypsin complexed to the inhibitor domain of Alzheimer's amyloid β -protein precursor (APPI) and basic pancreatic trypsin inhibitor (BPTI): engineering of inhibitors with altered specificities. *Protein Science*, 6, 1806-1824.

- 242 Burgering, M.J., Orbons, L.P., Van der Doelen, A., Mulders, J., Theunissen, H.J., Grootenhuys, P.D., Bode, W., Huber, R., Stubbs, M.T. (1997). The second kunitz domain of human tissue factor pathway inhibitor: cloning, structure determination and interaction with factor Xa. *Journal of Molecular Biology*, 269, 395-407.
- 243 Roussel, A., Mathieu, M., Dobbs, A., Luu, B., Cambillau, C., Kellenberger, C. (2001). Complexation of two proteic insect inhibitors to the active site of chymotrypsin suggests decoupled roles for binding and selectivity. *Journal of Biological Chemistry*, 276(42), 38893-38898.
- 244 Zhu, G., Huang, Q., Zhu, Y., Li, Y., Chi, C., Tang, Y. (2001). Three-dimensional structure of the complex between pancreatic secretory trypsin inhibitor (Kazal type) and trypsinogen at 1.8Å resolution. Structure solution, crystallographic refinement and preliminary structural interpretation. *Journal of Molecular Biology*, 162, 839-868.
- 245 Park, E.Y., Kim, J.A., Kim, H.W., Kim, Y.S., Song, H.K. (2004). Crystal structure of the Bowman-Birk inhibitor from barley seeds in ternary complex with porcine trypsin. *Journal of Molecular Biology*, 343(1), 173-186.
- 246 Helland, R., Berglund, G.I., Otlewski, J., Apostoluk, W., Andersen, O.A., Willassen, N.P., Smalås, A.O. (1998). High resolution structures of three new trypsin-squash-inhibitor complexes: a detailed comparison with other trypsin and their complexes. *Acta Crystallographica Section D*, 55, 139-148.
- 247 Ibrahim, B.S., Shamaladevi, N., Pattabhi, V. (2004). Trypsin activity reduced by an autocatalytically produced nonapeptide. *Journal of Biomolecular Structure & Dynamics*, 21(6), 737-744.
- 248 Cambillau, C., Spinelli, S., Lauwereys, M. (To be published). Crystal structure of a complex between bovine chymotrypsin and ecotin at 2.0Å resolution.
- 249 Ibrahim, B.S., Pattabhi, V. (2004). Crystal structure of trypsin-turkey egg white inhibitor complex. *Biochemical and Biophysical Research Communications*, 213(1), 8-16.
- 250 Leiros, H.K., Mc Sweeney, S.M., Smalås, A.O. (2001). Atomic resolution structures of trypsin provide insight into structural radiation damage. *Acta crystallographica, Section D*, 57(4), 488-497.
- 251 Johnson, A., Krishnaswamy, S., Sundaram, P.V., Pattabhi, V. (1997). The first structure at 1.8Å resolution of an active autolysate form of porcine alpha-trypsin. *Acta crystallographica, Section D*, 53(3), 311-315.
- 252 Tsukada, H., Blow, D.M. (1985). Structure of alpha-chymotrypsin refined at 1.68 Å resolution. *Journal of Molecular Biology*, 184(4), 703-711.
- 253 Feng, S., Chen, J.K., Yu, H., Simon, J.A., Schreiber, S.L. (1994). Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions. *Science*, 266, 1241-1247.
- 254 Zimmermann, G.R., Legault, P., Selsted, M.E. (1995). Solution structure of bovine neutrophil beta-defensin-12: the peptide fold of the beta defensins is identical to that of the classical defensins. *Biochemistry*, 34(41), 13663-13671.
- 255 Osawa, M., Tokumitsu, H., Swindells, M.B., Kurihara, H., Orita, M., Shibamura, T., Furuya, T., Ikura, M. (1999). A novel target recognition revealed by calmodulin in complex with Ca²⁺-calmodulin-dependent kinase kinase. *Structural Biology*, 6(9), 819-824.
- 256 Torres, A.M., De Plater, G.M., Doverskog, M., Birinyi-Strachan, L.C., Nicholson, G.M., Gallagher, C.H., Kuchel, P.W. (2000). Defensin-like peptide-2 from platypus venom: member of a class of peptides with a distinct structural fold. *Biochemical Journal*, 348, 649-656.
- 257 Starovasnik, M.A., Skelton, N.J., O'Connell, M.P., Kelley, R.F., Reilly, D., Fairbrother, W.J. (1996). Solution structure of the E-domain of staphylococcal protein A. *Biochemistry*, 35(48), 15558-15569.

- 258 Nicastro, G., Franzoni, L., De Chiara, C., Mancin, A.C., Giglio, J.R., Spisni, A. (2003). Solution structure of crotamine, a Na⁺ channel affecting toxin from *Crotalus durissus terrificus*. *European Journal of Biochemistry*, 270(9), 1969-1979.
- 259 Penin, F., Geourjon, C., Montserret, R., Bockmann, A., Lesage, A., Yang, Y.S., Bonod-Bidaud, C., Cortay, J.C., Negre, D., Cozzone, A.J., Deleage, G. (1997). Three-dimensional structure of the DNA-binding domain of the fructose repressor from *Escherichia coli* by ¹H and ¹⁵N NMR. *Journal of Molecular Biology*, 270(3), 496-510.
- 260 Schwartz, T., Behlke, J., Lowenhaupt, K., Heinemann, U., Rich, A. (2001). Structure of the DLM1-Z-DNA complex reveals a conserved family of Z-DNA-binding proteins. *Nature Structural Biology*, 8(9), 761-765.
- 261 Liu, Y., Hong, X., Kappler, J., Jiang, L., Zhang, R., Xu, L., Pan, C.H., Martin, W.E., Murphy, R.C., Shu, H.B., Dai, S., Zhang, G. (2003). Ligand-receptor binding revealed by the TNF family member TALL-1. *Nature*, 423, 49-56.
- 262 Kalus, W., Zweckstetter, M., Renner, C., Sanchez, Y., Georgescu, J., Grol, L., Demuth, D., Schumacher, R., Donv, C., Lang, K., Holak, T.A. (1998). Structure of the IGF-binding domain of the insulin-like growth factor-binding protein-5 (IGFBP-5) : implications for IGF and IGF-I receptor interactions. *EMBO Journal*, 17(22), 6558-6572.
- 263 O'Neill, J.W., Kim, D.E., Baker, D., Zhang, K.Y. (2001). Structures of the B1 domain of protein L from *peptostreptococcus magnus* with a tyrosine to tryptophan substitution. *Acta Crystallographica Section D Biological Crystallography*, 57(Pt 4), 480-487.
- 264 Cheng, Y., Meng, Q., Wang, W., Wang, J. (2002). Structure-function relationship of three neurotoxins from the venom of *Naja kaouthia* : a comparison between the NMR-derived structure of NT2 with its homologues, NT1 and NT3. *Biochimica et Biophysica Acta*, 1594(2), 353-363.
- 265 Ludvigsen, S., Olsen, H.B., Kaarsholm, N.C. (1998). A structural switch in a mutant insulin exposes key residues for receptor binding. *Journal of Molecular Biology*, 279(1), 1-7.
- 266 Kalia, Y.N., Brocklehurst, S.M., Hipps, D.S., Appella, E., Sakaguchi, K., Perham, R.N. (1993). The high-resolution structure of the peripheral subunit-binding domain of dihydrolipoamide acetyltransferase from the pyruvate dehydrogenase multienzyme complex of *Bacillus stearothermophilus*. *Journal of Molecular Biology*, 230(1), 323-341.
- 267 Fogh, R.H., Kem, W.R., Norton, R.S. (1990). Solution structure of neurotoxin I from the sea anemone *Stichodactyla helianthus*. A nuclear magnetic resonance. *Journal of Biological Chemistry*, 265(22), 13016-13028.
- 268 Gordon, L.M., Mobley, P.W., Pilpa, R., Sherman, M.A., Waring, A.J. (2002). Conformational mapping of the N-terminal peptide of HIV-1 gp41 in membrane environments using ¹³C-enhanced Fourier transform infrared spectroscopy. *Biochimica et Biophysica Acta*, 1559(2), 96-120.
- 269 Kao, Y.H., Lee, G.F., Wang, Y., Starovasnik, M.A., Kelley, R.F., Spellman, M.W., Lerner, L. (1999). The effect of O-fucosylation on the first EGF-like domain from human blood coagulation factor VII. *Biochemistry*, 38(22), 7097-7110.
- 270 Schumann, F.H., Varadan, R., Tayakuniyil, P.P., Hall, J.B., Camarero, J.A., Fushman, D. (to be published). Solution structure of a circular form of the truncated N-terminal SH3 domain from oncogene protein c-Crk.
- 271 Montelione, G.T., Wuthrich, K., Burgess, A.W., Nice, E.C., Wagner, G., Gibson, K.D., Scheraga, H.A. (1992). Solution structure of murine epidermal growth factor determined by NMR spectroscopy and refined by energy minimization with restraints. *Biochemistry*, 31(1), 236-249.
- 272 McEvoy, M.M., Hausrath, A.C., Randolph, G.B., Remington, S.J., Dahlquist, F.W. (1998). Two binding modes reveal flexibility in kinase/response regulator interactions in the

-
- bacterial chemotaxis pathway. *Proceedings of the National Academy of Science*, 95(13), 7333-7338.
- 273 Holliger, P., Riechmann, L. (1997). A conserved infection pathway for filamentous bacteriophages is suggested by the structure of the membrane penetration domain of the minor coat protein g3p from phage fd. *Structure*, 5(2), 265-275.
- 274 Williamson, M.P., Madison, V.S. (1990). Three-dimensional structure of porcine C5a_{desArg} from ¹H nuclear magnetic resonance data. *Biochemistry*, 29(12), 2895-2905.
- 275 Huang, W., Dolmer, K., Gettins, P.G. (1999). NMR solution structure of complement-like repeat CR8 from the low density lipoprotein receptor-related protein. *Journal of Biological Chemistry*, 274(20), 14130-14136.
- 276 Bycroft, M., Grünert, S., Murzin, A.G., Proctor, M., St Johnston, D. (1995). NMR solution of a dsRNA binding domain from drosophila staufer protein reveals homology to the N-terminal domain of ribosomal protein S5. *EMBO Journal*, 14(14), 3563-3571.
- 277 Kirkland, J.J. (1968). A high performance ultraviolet photometer detector for use with efficient liquid chromatographic columns. *Analytical Chemistry*, 40, 391.
- 278 Jayaraman, G., Kumar, T.K., Arunkumar, A.I., Yu, C. (1996). 2,2,2-trifluoroethanol induces helical conformation in an all beta-sheet protein. *Biochemical and Biophysical Research Communication*, 222(1), 33-37.
- 279 Arunkumar, A.I., Kumar, T.K., Jayaraman, G., Samuel, D., Yu, C. (1996). Induction of helical conformation in all beta-sheet proteins by trifluoroethanol. *Journal of Biomolecular Structure & Dynamics*, 14(3), 381-385.
- 280 O'Neill, J.W., Kim, D.E., Johnsen, K., Baker, D., Zhang, K.Y. (2001). Single-site mutations induce 3D domain swapping in the B1 domain of protein L from *Peptostreptococcus magnus*. *Structure*, 9(11), 1017-1027.
- 281 Racapé, J., Lecoq, A., Romi-Lebrun, R., Liu, J., Kohler, M., Garcia, M.L., Ménez, A., Gasparini, S. (2002). Characterization of a novel radiolabeled peptide selective for a subpopulation of a voltage-gated potassium channels in mammalian brain. *Journal of Biological Chemistry*, 277(6), 3886-2893.
- 282 Baker, E.N., Hubbard, R.E. (1984). Hydrogen bonding in globular proteins. *Progress in Biophysics & Molecular Biology*, 44, 97-179.
- 283 Mc Donald, I.K., Thornton, J.M. (1994). Satisfying hydrogen bonding potential in proteins. *Journal of Molecular Biology*, 238(5), 777-793.
- 284 Barlow, D.J., Thornton, J.M. (1983). Ion-pairs in proteins. *Journal of Molecular Biology*, 168(4), 867-885.
- 285 Rashin, A.A., Honig, B. (1984). On the environment of ionizable groups in globular proteins. *Journal of Molecular Biology*, 173(4), 515-521.
- 286 De Lano, W.L. (2002). Unraveling hot spots in binding interfaces: progress and challenges. *Current Opinion in Structural Biology*, 12, 14-20.
- 287 Pearl, F.M., Bennett, C.F., Bray, J.E., Harrison, A.P., Martin, N., Shepherd, A., Sillitoe, I., Thornton, J., Orengo, C.A. (1968). The CATH database : an extended protein family resource for structural and functional genomics. *Nucleic Acids Research*, 31(1), 452-455.
- 288 Wyatt, R., Kwong, P.D., Desjardins, E., Sweet, R.W., Robinson, J., Hendrickson, W.A., Sodroski, J.G. (1998). The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature*, 393, 705-711.
- 289 Arthos, J., Deen, K.C., Chaikin, M.A., Fornwald, J.A., Sathe, G., Sattentau, Q.J., Clapham, P.R., Weiss, R.A., Mc Dougal, J.S., Pietropaolo, C. (1989). Identification of the residues in human CD4 critical for the binding of HIV. *Cell*, 57(3), 469-481.
- 290 Ashkenazi, A., Presta, L.G., Marsters, S.A., Camerato, T.R., Rosenthal, K.A., Fendley, B.M., Capon, D.J. (1990). Mapping the CD4 binding site for human immunodeficiency

virus by alanine-scanning mutagenesis. *Proceedings of the National Academy of Sciences*, 87(18), 7150-7154.

291 Moebius, U., Clayton, L.K., Abraham, S., Harrison, S.C. & Reinherz, E.L. (1992). The human immunodeficiency virus gp120 binding site on CD4 : delineation by quantitative equilibrium and kinetic binding studies of mutants in conjunction with high-resolution CD4 atomic structure. *Journal of Experimental Medicine*, 176(2), 507-517.

292 Sweet, R.W., Truneh, A., Hendrickson, W.A. (1991). CD4 : its structure, role in immune function and AIDS pathogenesis, and potential as a pharmacological target. *Current Opinion in Biotechnology*, 2(4), 622-633.

293 Ryu, S.E., Truneh, A., Sweet, R.W., Hendrickson W.A. (1994). Structure of an HIV and MHC binding fragment from human CD4 as refined in two crystal lattices. *Structure*, 2(1), 59-74.

294 Fersht, A.R., Shi, J.P., Knill-Jones, J., Lowe, D.M., Wilkinson, A.J., Blow, D.M., Brick, P., Carter, P., Waye, M.M., Winter, G. (1985). Hydrogen bonding and biological specificity analysed by protein engineering. *Nature*, 314, 235-238.

295 Lee, L.P., Tidor, B. (2000). Optimization of binding electrostatics: charge complementarity in the barnase-barstar protein complex. *Protein Science*, 10, 362-377.

296 Hendsch, Z.S., Jonsson, T., Sauer, R.T., Tidor, B. (1996). Protein stabilization by removal of unsatisfied polar groups : computational approaches and experimental tests. *Biochemistry*, 35(24), 7621-7625.

297 Bernstein, F.C., Koetzle, T.F., Williams, G.J.B, Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M. (1977). The Protein Data Bank: a compute-based archival file for macromolecular structures. *Journal of Molecular Biology*, 112, 535-542.

298 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 Research*, 28(1), 235-242.

299 Berman, H.M., Henrick, K., Nakamura, H. (2003). Announcing the worldwide Protein Data Bank. *Nature Structural Biology*, 10(12), 980.

300 Westbrook, J., Feng, Z., Jain, S., Bhat, T.N., Thanki, N., Ravichandran, V., Gilliland, G.L., Bluhm, W.F., Weissig, H., Greer, D.S., Bourne, P.E., Berman, H.M. (2002). The Protein Data Bank: unifying the archive. *Nucleic Acids Research*, 30(1), 245-248.

301 Daune, M. Biophysique en mouvement. Partie 4, Chapitre 1-4, 340-348. *InterEditions*, 1993.

302 Sitkoff, D., Sharp, K.A., Honig, B. (1994). Accurate calculation of hydration free energies using macroscopic solvent models. *Journal of Physical Chemistry*, 98(7), 1978-1988.

303 Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M. (1982). CHARMM: A program for macromolecular energy minimization and dynamics calculations. *Journal of Computational Chemistry*, 4(2), 187-217.

304 MacKerell Jr., A.D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnik, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher III, W.E., Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., Karplus, M. (1998). All-atom empirical potential for molecular modelling and dynamics studies of proteins. *Journal of Physical Chemistry B*, 102, 3586-3616.

305 Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990). "Basic local alignment search tool". *Journal of Molecular Biology*, 215, 403-410.

306 Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17), 3349-3402.

-
- 307 Schwartz, T., Lowenhaupt, K., Kim, Y.G., Li, L., Brown II, B.A., Herbert., A., Rich, A. (1999). Proteolytic dissection of Zab, the Z-DNA-binding domain of human ADAR1. *Journal of Biological Chemistry*, 274(5), 2899-2906.
- 308 Inui, T., Nakao, M., Nishio, H., Nischiuchi, Y., Kojima, S., Muramatsu, T., Kimura, T. (2000). Solution synthesis and biological activity of human pleiotrophin, a novel heparin-binding neurotrophic factor consisting of 136 amino acid residues with five disulfide bonds. *Journal of Peptide Research*, 55(5), 384-397.
- 309 Sakakibara, S. (1995). Synthesis of large peptides in solution. *Biopolymers*, 37(1), 17-28.
- 310 Kenner, G.W. McDermott, J.R. and Sheppard, R.C. (1971). Safety catch principle in solid phasesynthesis. *Chemical Communications*, 12, 636-637.
- 311 Leslie, A.G. (1992). Joint CCP4 and ESF-EAMCB. *Newsletter on Protein Crystallography*, No. 26.
- 312 Collaborative Computational Project, number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallographica, section D.*, 50, 760-763.
- 313 Vagin, A., Teplyakov, A. (1997). MOLREP : an automated program for molecular replacement. *Journal of Applied Crystallography*, 30(6), 1022-1025.
- 314 Matthews, B.W. (1968). Solvent content of protein crystals. *Journal of Molecular Biology*, 33, 491-497.
- 315 Murshudov, G.N., Vagin, A.A., Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum -likelihood method. *Acta Crystallographica, section D.*, 53, 240-255.
- 316 Roussel, A., Cambillau, C. (1989). Silicon graphics geometry partner directory. *Silicon Graphics*, 77-78.