Conception de ligands protéiques artificiels par ingénierie moléculaire in silico

Rym Tlatli Baccouche

▶ **To cite this version:**

Rym Tlatli Baccouche. Conception de ligands protéiques artificiels par ingénierie moléculaire in silico. Médecine humaine et pathologie. Université René Descartes - Paris V; Université de Tunis El Manar, 2012. Français. NNT : 2012PA05P652 . tel-00807525

HAL Id: tel-00807525

https://theses.hal.science/tel-00807525

Submitted on 3 Apr 2013

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.

ECOLE DOCTORALE DU MEDICAMENT MTCE 436
UNIVERSITE DE PARIS DESCARTES,
FACULTE DES SCIENCES DE TUNIS,
UNIVERSITE DE TUNIS EL-MANAR

Thèse présentée pour l'obtention du grade de :
Docteur des Universités Paris Descartes et Tunis El-Manar
Spécialité : Ingénierie des protéines

Conception de ligands protéiques artificiels par ingénierie moléculaire *in silico*

Présentée par
RYM TLATLI BACCOUCHE

Thèse soutenue le 30 Novembre 2012

Membres du Jury :

Pr Christiane Garbay	Président du jury
Pr Philippe Minard	Rapporteur
Pr Nabil Miled	Rapporteur
Pr Vinh Tran	Examinateur
Pr Amel El Gaied	Examinateur
Dr Philippe Cuniasse	Examinateur
Dr Vincent Dive	Directeur de thèse
Dr Mohamed El Ayeb	Directeur de thèse

REMERCIEMENTS

Ce rapport est le fruit d'un travail de recherche mené en cotutelle entre le Commissariat à l'Energie Atomique (CEA) de Saclay et l'Institut Pasteur de Tunis.

Tout d'abord, je tiens à remercier Mr. Philippe Minard et Mr. Nabil Miled qui me font l'honneur de juger ce travail en tant que rapporteurs, en espérant que ce manuscrit de thèse retienne leur intérêt et qu'il leur soit agréable à lire. Je tiens également à exprimer ma reconnaissance à Mr. Vinh Tran et Mme Amel El Gaied qui ont accepté d'évaluer ce travail en tant qu'examinateurs. Je remercie particulièrement Mme Christiane Garbay de me faire l'honneur de présider ma soutenance de thèse malgré ses nombreuses occupations.

Je tiens à exprimer mon immense gratitude à la personne sans laquelle ce travail n'aurait jamais vu le jour, Dr. Vincent Dive, Chef du Service d'ingénierie moléculaire des protéines (SIMOPRO) au CEA, qui m'a accueillie au sein de son service dans le cadre de mon projet de Mastère et qui m'a offert l'opportunité de poursuivre mes travaux de recherche par la réalisation d'une thèse. Il a cru en moi jusqu'au bout malgré toutes les circonstances difficiles liées à cette thèse en cotutelle et s'est toujours battu pour m'offrir des conditions de travail favorables. Je lui exprime donc toute ma reconnaissance pour son soutien permanent et sa générosité tant sur le plan scientifique que sur le plan humain.

Je tiens à exprimer ma profonde gratitude à mon chef spirituel et scientifique, Mr. Philippe Cuniasse qui m'a encadrée, formée et supportée tout au long de ces années de thèse. Il a su me transmettre ses connaissances en bioinformatique structurale et en modélisation moléculaire avec une si grande patience et tant de pédagogie que ce qui n'était pour moi que du chinois est devenu clair et limpide. Il m'a appris à être rigoureuse et méthodique et c'est grâce à son aide précieuse et ses directives pertinentes que j'ai pu mener à bien ce travail de thèse. Merci de ne m'avoir jamais laissé tomber malgré une situation pour le moins difficile.

Je tiens à exprimer ma reconnaissance à mon co-directeur de thèse, Mr. Mohamed El Ayeb, chef du Laboratoire des Venins et Toxines à l'Institut Pasteur de Tunis, qui m'a accueillie au sein de son laboratoire et qui a tout fait pour m'aider à réaliser mes projets scientifiques. Je remercie également tous mes collègues de l'Institut Pasteur de Tunis pour leur aide, leur

soutien et leur collaboration, et en particulier Mme Rym El Fessi qui est devenue une véritable amie.

Je remercie également Mr. Hervé Nozach avec qui et grâce à qui j'ai effectué une grande partie de ce travail de recherche. Il m'a fait bénéficier de son savoir-faire et de sa grande expérience dans le domaine de la biologie moléculaire, mais il a également réalisé l'exploit de me supporter (du moins je l'espère) pendant toutes ces années de collaboration grâce à son calme olympien, sa générosité et sa gentillesse permanente.

Mes remerciements s'adressent à Mr. Nicolas Gilles avec lequel j'ai collaboré pour réaliser la partie sur le criblage de venins de serpents et qui m'a généreusement fourni le matériel nécessaire ainsi que toute son expertise dans ce domaine. Je remercie Mr. Alain Lecoq, Mr. Robert Thai, Mr. Steven Dubois, Mr. Enrico Stura, Mlle Laura Vera, Mr. Fabrice Beau, Mr. Gilles Mourier et Mlle Evelyne Lajeunesse, qui ont largement participé à ce projet et dont la contribution était indispensable pour le mener à bien.

Un grand merci à mes collègues et compagnons de route Catherine, Bertrand, Guillaume, Françoise, Sarah, Laurent, Mehdi, Pascaline, Clémentine et Flore pour tous les échanges scientifiques, et surtout personnels qu'on a pu partager et dont la compagnie m'a souvent fait oublier les 1500 kilomètres qui me séparent de ma famille.

Un énorme merci à mes amis de longue date en Tunisie, Noha, Souzy, Mouna, Inès, Meriem et Anis qui m'ont toujours soutenue et qui ont toujours été présents pour moi dans les moments de joie et de peine.

Je remercie du fond du cœur tous les membres de ma famille, piliers inébranlables de ma vie, pour leur soutien permanent et pour tout l'amour et l'affection qu'ils me portent : ma grand-mère Douja, mes parents Slim et Raja, ma sœur Nadia, son mari Moez, ma petite nièce Lili et bientôt Alia, mes beaux-parents Fadhel et Nabiha, mes belles-sœurs Amel et Najet ainsi que leurs maris et leurs enfants, mes oncles et mes tantes ainsi que mes cousins et cousines.

Enfin pour conclure, je tiens à remercier celui qui a souffert autant et peut être même plus que moi pendant cette thèse mais qui m'a pourtant toujours encouragée à aller jusqu'au bout de mon ambition, celui qui a subi mes grosses déprimes et séché mes larmes quand tout

allait de travers, mais surtout celui qui a su me donner tout l'amour et la compréhension dont j'avais besoin : mon mari alias Kimo.

Un énorme bisou à ma puce Salmouta qui a beaucoup souffert de mon absence pendant ces années de thèse mais qui sera un jour je l'espère fière de sa mère pour tout le travail accompli.

TABLE DES MATIERES

ABREVIATIONS..... 10

INTRODUCTION..... 12

BIBLIOGRAPHIE..... 14

PARTIE A : Les venins d'animaux, un réservoir riche en protéines thérapeutiques 15

I. Composition des venins d'animaux..... 16
II. Peptides et protéines de venins à applications thérapeutiques..... 17
II.1 Les inhibiteurs des canaux calcium..... 17
II.2 Les toxines spécifiques des canaux sodiques 17
II.3 Les bloqueurs des canaux potassium..... 18
II.4 Les toxines inhibant les récepteurs nicotiniques de l'acétylcholine..... 19
II.5 Les toxines utilisées dans le traitement des maladies cardiovasculaires..... 19
II.6 Toxines utilisées dans le traitement du diabète..... 21
III. Les inhibiteurs de protéases dans les venins 22
III.1 Les inhibiteurs de protéases à sérine 22
III.2 Les inhibiteurs de protéases à cystéine..... 23
III.3 Les inhibiteurs de métalloprotéases 23
IV. Etude d'un cas particulier : le venin du serpent *Dendroaspis polylepis* 24
V. Conclusion..... 25

PARTIE B : Conception *in silico* de ligands protéiques 26

CHAPITRE I : Ingénierie des protéines pour le développement de ligands..... 27

I.1 Les protéines thérapeutiques..... 27
I.1.1 Les protéines : de leur connaissance fondamentale aux applications thérapeutiques 27
I.1.2 Limitations des protéines comme outils thérapeutiques..... 32
I.2 Les Anticorps 33
I.2.1 Applications 34
I.2.2 Atouts des anticorps 35
I.2.3 Limitations des anticorps..... 37
I.3 Alternatives aux anticorps : les plates-formes génériques..... 42
I.3.1 VLRs 42
I.3.2 Adnectins 45
I.3.3 Avimers..... 48
I.3.4 Affibody..... 50
I.3.5 Darpins 53
I.3.6 Anticalins 55

I.3.7 Heat Repeat 57
I.4 L'ingénierie *in silico* de protéines fondée sur la structure 59
I.4.1 Amélioration de la spécificité des ligands protéiques 60
I.4.2 Modulation des interactions protéines-protéines 61
I.4.3 Modification de sites de liaison protéiques 61
I.4.4 Transfert de motif fonctionnel sur une protéine hôte 62
I.4.4.1 Identification des résidus du motif fonctionnel 62
I.4.4.2 Sélection des plates-formes protéiques hôtes 65
I.4.4.3 Applications de l'approche de conception de mini-protéines par transfert de motif 66
I.4.4.3.1 Recherche de plates-formes fondée sur la similarité des structures secondaires 66
I.4.4.3.2 Recherche de plates-formes non fondée sur la similarité des structures secondaires 73

CHAPITRE II : Les interactions protéines – protéines 85

II.1 Classification des complexes protéine-protéine naturels 85
II.2 Propriétés structurales des interfaces 86
II.2.1 Taille de l'interface 86
II.2.2 Structure des sites de reconnaissance dans les interfaces 88
II.2.3 Complémentarité géométrique entre les surfaces 91
II.2.4 Changements conformationnels 93
II.3 Composition chimique des interfaces et nature des résidus 94
II.4 Propriétés électrostatiques des interfaces 96
II.4.1 Complémentarité de charge 96
II.4.2 Complémentarité électrostatique 97
II.5 Phénomène d'association 98
II.5.1 Aspects thermodynamiques 98
II.5.2 Interactions protéine-protéine 100
II.6 Etude des interfaces de familles de complexes formés par une cible commune et ses différents ligands 103
II.6.1 Comparaison des formes 105
II.6.2 Comparaison des potentiels électrostatiques 106
II.6.3 Taille et composition des interfaces 107
II.6.4 Complémentarité de charge 108
II.6.5 Liaisons hydrogène 108
II.7 Conclusion 109

CHAPITRE III : Conception de ligands protéiques *in silico* par transfert de sites actifs 112

III.1 Principe de la méthode de conception de ligand protéique 112
III.1.1 Motif fonctionnel 113
III.1.2 Système de référence 113
III.1.3 Protéine hôte 114
III.2 Méthodes de sélection des protéines hôtes 115
III.2.1 STAMPS 115

III.2.1.1 Principe de la recherche de motif 115
III.2.1.2 Filtre stérique 117
III.2.1.3 Autres critères 118
III.2.2 Comparaison avec les autres méthodes décrites dans la littérature 119
III.3 Conclusion 121

CHAPITRE IV : Application : Ciblage du domaine catalytique des MMPs 124

IV.1 Aspects structuraux des MMPs 125
IV.1.1 Organisation des différents domaines des MMPs 125
IV.1.2 Structure tridimensionnelle du domaine catalytique des MMPs 126
IV.2 Les MMPs : des cibles thérapeutiques potentielles 132
IV.2.1 Rôles physiologiques 132
IV.2.2 Implication des MMPs à différents stades d'un processus pathologique 133
IV.2.3 Rôle ambivalent des MMPs 133
IV.3 Les inhibiteurs naturels des MMPs : les TIMPs 135
IV.3.1 Aspects structuraux des TIMPs 135
IV.3.2 Affinité et sélectivité 138
IV.3.3 Mécanisme d'inhibition des MMPs 139
IV.4 Conclusion 142

MATERIELS ET METHODES 144

MATERIEL 145

METHODES 148

I. Recherche de protéines inhibitrices de MMPs dans le venin du serpent *Dendroaspis polylepis* .. 148
I.1 Fractionnement du venin par chromatographie d'échange d'ions 148
I.2 Purification par RP-HPLC des fractions issues de la chromatographie d'échange d'ions 149
I.3 Criblage des fractions sur les MMPs 150
II. Recherche *in silico* de plates-formes par le programme STAMPS 150
II.1 Données nécessaires à la recherche 150
II.2 Paramètres de recherche 150
III. Analyse électrostatique des plates-formes 152
III.1 Calcul du potentiel électrostatique des plates-formes mimes et du ligand de référence 153
III.2 Comparaison des potentiels électrostatiques entre les plates-formes mimes et le ligand de référence 154
III.3 Comparaison des formes des plates-formes mimes et du ligand de référence 155
IV. Procédure de relaxation par dynamique moléculaire et minimisation d'énergie 155
IV.1 Minimisation d'énergie sous contraintes 156
IV.2 Dynamique moléculaire dans le vide 158
IV.3 Dynamique moléculaire en solvant explicite 158
IV.4 Analyse des interfaces plates-formes mimes/MMP-14 159

V. Production des plates-formes par synthèse peptidique 160
V.1 Synthèse chimique sur support solide 160
V.2 Purification du brut de synthèse par RP-HPLC 161
V.3 Formation des ponts disulfures 161
VI. Production des plates-formes par voie recombinante chez *E. coli* 161
VI.1 Construction des vecteurs d'expression 161
VI.2 Production des mini-protéines d'intérêt 162
VI.3 Extraction et purification des protéines de fusion 163
VI.4 Clivage et purification des mini-protéines d'intérêt 163
VII. Caractérisation des mini-protéines produites par synthèse chimique et par voie recombinante 164
VIII. Tests enzymatiques d'inhibition des MMPs 164
IX. Détermination de la structure du complexe mini-protéine/MMP par cristallographie aux rayons X 165
IX.1 Cristallogenèse 165
IX.2 Collecte des données et détermination de la structure tridimensionnelle 166

RESULTATS 167

PARTIE A : Recherche d'inhibiteurs de MMPs dans le venin du serpent *Dendroaspis polylepis* 168
I.1 Fractionnement du venin de serpent *Dendroaspis polylepis* par système bi-dimensionnel 169
I.2 Criblage des venins de serpents sur les MMPs 171

PARTIE B : Conception *in silico* de ligands protéiques inhibiteurs de MMPs 172

CHAPITRE I : Système de référence : TIMP/MMP 173

I.1 Données structurales et fonctionnelles sur les complexes TIMP/MMP 173
I.2 Analyse de l'interface du complexe TIMP-2/MMP-14 (1bqq) 175
I.2.1 Propriétés globales de l'interface 176
I.2.1.1 Variation de surface accessible au solvant/Taille de l'interface 176
I.2.1.2 Compacité de l'interface 177
I.2.1.3 Composition de l'interface 177
I.2.2 Interactions intermoléculaires 178
I.2.2.1 Liaisons hydrogène 178
I.2.2.2 Autres types d'interactions 179
I.2.2.3 Localisation des interactions 180
I.2.3. Propriétés électrostatiques 181
I.3 Conclusion 182

CHAPITRE II : Recherche de plates-formes protéiques par STAMPS 184

II.1 Critères de recherche *in silico* du programme STAMPS 184
II.2 Identification et sélection des plates-formes 185
II.3 Analyse des plates-formes sélectionnées 190
II.3.1 Origines et fonctions des plates-formes **191**

II.3.2 Structures secondaires 192
II.3.3 Les motifs de liaison 195
II.4 Analyse des potentiels électrostatiques 197
II.4.1 Analyse graphique 197
II.4.2 Analyse des indices de similarité de Hodgkin 199
II.5 Analyse des interfaces plateforme/cible 202
II.6 Conclusion 205

CHAPITRE III : Production des ligands conçus *in silico* et caractérisation comme inhibiteurs de MMPs 207

III.1 Production des mini-protéines conçues *in silico* par synthèse chimique 207
III.1.1 Synthèse sur support solide 207
III.1.2 Clivage du produit de synthèse final 211
III.1.3 Purification du produit brut de synthèse par RP-HPLC 211
III.1.4 Formation des ponts disulfures 212
III.2 Production des plates-formes protéiques par voie recombinante 213
III.2.1 Construction des plasmides d'expression 213
III.2.1.1 Essais et mise au point des différentes stratégies d'expression 213
III.2.1.2 Stratégie optimale adoptée pour l'expression des mini-protéines d'intérêt 216
III.2.2 Production et purification des protéines de fusion 218
III.2.3 Clivage des protéines de fusion et purification des mini-protéines d'intérêt 219
III.2.4 Caractérisation des mini-protéines produites par voie recombinante 220
III.2.5 Rendement de production des mini-protéines d'intérêt 221
III.3 Evaluation des mini-protéines comme inhibiteurs de MMPs 224
III.4 Optimisation de certaines mini-protéines identifiées *in silico* 228
III.4.1 Variants de la plate-forme 1edpA pour la position 2 du motif de liaison 228
III.4.2 Variants de la plate-forme 1an1I_m par mutagénèse dirigée 231
III.5 Caractérisation structurale de l'interaction mini-protéine/MMP 236
III.6 Etude de la stabilité des interactions par dynamique moléculaire 241
III.7 Conclusion 245

CONCLUSION ET PERSPECTIVES 247

ANNEXES 260

ARTICLES PUBLIÉS ..... Erreur ! Signet non défini.

REFERENCES BIBLIOGRAPHIQUES ..... Erreur ! Signet non défini.

ABREVIATIONS

Ac2O	Anhydride acétique
ACN	Acétonitrile
ADAM	A Disintegrin And Metalloproteinase
ADAMTS	A Disintegrin And Metalloproteinase with Thrombospondin motifs
AHA	Acetohydroxamic acid
APBS	Adaptative Poisson Boltzmann Solver
CDR	Complementarity Determining Region
CHARMM	Chemistry at HARvard Macromolecular Mechanics
Cl-HOBt	6-Chloro-1-hydroxybenzotriazole
DCCI	N,N'-dicyclohexylcarbodiimide
DCM	Dichlorométhane
DMSO	Diméthylsulfoxyde
DO	Densité Optique
Dpa	Acide N-3-(2,4-dinitrophényl)-L-2,3-diaminopropionique
DTT	Dithiothréitol
ELISA	Enzyme-Linked Immunosorbent Assay
ESI	Electrostatic Similarity Index
FDA	Food and Drug Administration
Fmoc	Fluorenyl-methoxy-carbonyl
GRAVY	Grand average of hydropathicity
HCCA	Acide α-cyano-4-hydroxy-cinnamique
HPLC	High Performance Liquid Chromatography
IC50	Concentration d'Inhibiteur nécessaire pour atteindre 50% d'inhibition
IPTG	Isopropyl-β-D-thiogalactopyranoside
LC/MS	Liquid Chromatography/Mass Spectrometry
MALDI	Matrix Assisted Laser Desorption Ionization
MBHA	4-méthylbenzhydrylamine
Mca	Acide 7-méthoxycoumarin-4-acétique
MES	2-(N-morpholino) ethanesulphonic acid
MMP	Matrix MetalloProteinase ou Métalloprotéase de la matrice
NOE	Nuclear Overhauser Effect

NMP	N-méthyl-2 pyrrolidone
PCR	Polymerase Chain Reaction
PDB	Protein Data Base
PEG	Polyéthylène glycol
PI	Point Isoélectrique
PIPSA	Protein Interaction and Property Similarity Analysis
PyMOL	Python-enhanced MOLecular graphic system
RMN	Résonance Magnétique Nucléaire
RMSD	Root Mean Square Deviation
SSI	Shape Similarity Index
STAMPS	Search for Three-dimensional Atomic Motif in Protein Structure
TACE	Tumor necrosis factor-α Converting Enzyme
TBME	Tert-butyl methyl ether
TEV	Tobacco Etch Virus
TFA	Acide trifluoroacétique
TIMP	Tissue Inhibitor of Matrix MetalloProteinase ou inhibiteur tissulaire de métalloprotéase de la matrice
TIS	Triisopropylsilane
TOF	Time Of Flight

INTRODUCTION

L'ère de la chimiothérapie a été très fructueuse dans le domaine de la médecine durant le 20ème siècle, fournissant de petites molécules médicament pour le traitement de nombreuses maladies infectieuses, troubles métaboliques, maladies cardiaques, neuromédecine, et cancer. Les progrès réalisés ces dernières décennies concernant la connaissance des aspects cellulaires et moléculaires des mécanismes physiologiques et physiopathologiques ont grandement stimulé la poursuite de la recherche de molécules susceptibles d'interagir à haute affinité et avec une grande spécificité avec des cibles d'intérêt thérapeutique. Cependant, le nombre de nouvelles entités chimiques approuvées par an est en constante diminution ces dernières années et des stratégies innovantes pour le développement de molécules thérapeutiques et diagnostiques sont nécessaires.

Dans cette perspective, le développement de méthodes efficaces permettant de concevoir des ligands protéiques artificiels capables d'interagir spécifiquement avec des cibles d'intérêt thérapeutique et/ou diagnostique est devenu un champ de recherche extrêmement actif. De nombreuses approches ont été reportées dans la littérature. L'exploitation de la diversité moléculaire des protéines présentes dans les venins afin de rechercher des protéines naturellement capables d'interagir à haute affinité avec une cible donnée a permis dans certains cas de découvrir des molécules de grand intérêt. Plus récemment, les progrès réalisés dans le domaine des biotechnologies ont permis de produire des bibliothèques de variants protéiques artificiels au sein desquelles des composés de haute affinité ont pu être sélectionnés. Enfin, des ligands protéiques peuvent être conçus grâce à des méthodes *in silico* reposant sur le couplage de méthodes de bioinformatique structurale, de modélisation moléculaire et d'optimisation expérimentale. Des résultats très prometteurs ont été obtenus avec ces différentes approches qui représentent autant d'opportunités de concevoir de nouvelles molécules thérapeutiques de grande valeur. Les approches de criblage conduisent parfois à la découverte de molécules de grand intérêt, mais des méthodes de conception *in silico* efficaces pourraient constituer une voie moins aléatoire. Quelques études récentes utilisant le couplage étroit de méthodes *in silico* visant à découvrir un ou plusieurs ligand(s) protéique(s) présentant une affinité mesurable permettant son optimisation ultérieure par des méthodes expérimentales ont conduit à quelques succès spectaculaires. Ces études ont bien montré la puissance des méthodes d'optimisation

telles que l'évolution moléculaire dirigée une fois les ligands protéiques initiaux d'affinité mesurable obtenus. Les résultats reportés dans ces études montrent également que l'étape limitante reste la conception *in silico* des ligands protéiques initiaux. En effet, le taux de succès (nombre de ligands protéiques conçus montrant effectivement une affinité pour la cible / nombre de ligands conçus *in silico*) est parfois très faible. D'autre part, l'affinité obtenue est souvent plusieurs ordres de grandeur en dessous de l'affinité des ligands protéiques naturels. Cependant, ces études sont encore très peu nombreuses et il est difficile d'évaluer la robustesse des méthodes bioinformatiques mises en œuvre. Il est donc nécessaire de poursuivre leur évaluation et leur développement.

Les travaux réalisés au cours de cette thèse concernent la recherche de ligands protéiques capables de cibler les sites catalytiques de membres de la famille des métalloprotéases matricielles ou MMPs. Pour cela, deux méthodes ont été mises en œuvre. Dans un premier temps, des ligands protéiques qui seraient naturellement capables d'interagir avec les cibles considérées ont été recherchés dans le venin du serpent *Dendroaspis polylepis*. Dans un deuxième temps, des ligands protéiques capables de lier les sites catalytiques des MMPs ont été conçus *in silico* grâce à la méthode de conception de ligands protéiques développée au laboratoire. Cette méthode repose sur le greffage d'un motif fonctionnel sur des petites protéines susceptibles de reproduire la topologie adoptée par ces résidus dans un ligand protéique naturel. Les plates-formes protéiques sont sélectionnées par criblage systématique de la base de données de structures des protéines (PDB) grâce au logiciel STAMPS (Search for Three-dimensionnal Atom Motif in Protein Structure) développé dans l'équipe. La production d'une série de 10 mini-protéines ligands conçues à l'aide de cette approche a été réalisée et leur capacité à cibler les domaines catalytiques de plusieurs MMPs a été déterminée. L'objectif de ces travaux est d'évaluer la robustesse de la méthode de conception *in silico* développée au laboratoire et sa capacité à conduire à une série de protéines ligands de première génération effectivement capables d'interagir avec les cibles considérées. Nous avons également cherché à évaluer la capacité de chaque ligand protéique artificiel à reproduire les interactions intermoléculaires établies par le motif fonctionnel dans le contexte du ligand naturel avec la cible. Ces molécules pourraient être ultérieurement optimisées par des méthodes expérimentales pour obtenir des mini-protéines capables d'inhiber à haute affinité et sélectivement un membre particulier de la famille des MMPs.

BIBLIOGRAPHIE

PARTIE A :

Les venins d'animaux,
un réservoir riche en protéines thérapeutiques

Les animaux venimeux possèdent des glandes exocrines produisant du venin. Les venins d'animaux tels que les serpents, les scorpions, les araignées, les anémones de mer et les cônes marins sont constitués de mélanges complexes contenant un grand nombre de peptides et de protéines caractérisés par des fonctions pharmacologiques diverses et spécifiques. Ces toxines sont conçues pour agir de manière efficace sur des cibles physiologiques clé des vertébrés et des invertébrés, telles que les canaux ioniques, les récepteurs et les enzymes, rendant ces animaux venimeux capables de digérer leurs proies et de se défendre contre les prédateurs. Par ailleurs, les venins contiennent également des inhibiteurs de protéases et des agents stabilisants qui les protègent des conditions externes et internes défavorables, telles qu'une température élevée, et qui les préservent dans les glandes pendant plusieurs semaines. Cette grande diversité de composés et de fonctions caractéristique des venins a conduit au développement de plusieurs médicaments et d'un grand nombre de molécules d'intérêt thérapeutique et diagnostique.

En particulier, différents travaux ont mis en lumière l'existence d'inhibiteurs de protéases dans les venins d'animaux, l'un des plus connus étant celui de l'enzyme de conversion de l'angiotensine ayant conduit à la mise au point d'antihypertenseurs utilisés en clinique humaine depuis plus de 20 ans. Les venins constituent donc une source potentielle de mini-protéines actives sur les MMPs.

I. Composition des venins d'animaux

Les venins contiennent plus de 100 peptides et protéines différents dont les séquences et les structures sont hautement diversifiées. La composition des venins varie considérablement en fonction de l'animal considéré. Dans les venins de serpents, les peptides et les protéines sont essentiellement regroupés en deux familles : les enzymes qui représentent entre 80 à 90% des venins des vipéridés et entre 25 à 70% de ceux des élapidés (protéases à sérine, métalloprotéases à Zn^{2+}, L-acide aminé oxydase, hyaluronidases, acétylcholinestérases, phospholipase A2, phosphodiestérases) et les protéines sans activité enzymatique dont les toxines (cytotoxines, cardiotoxines, neurotoxines postsynaptiques, myotoxines, ohanin, désintégrines, lectines de type C, peptides natriurétiques, protéines sécrétoires riches en cystéines, facteur de croissance des nerfs et de l'endothélium vasculaire, cystatine, inhibiteurs de protéases de type Kunitz et toxines 3 doigts).

La diversité de peptides contenue dans les venins est particulièrement importante dans les venins de cônes (conotoxines et conopeptides), mollusques marins très largement répandus. Le nombre de conopeptides existants est estimé à 50 000 dont seulement moins de 0,1% ont été caractérisés pharmacologiquement. Parmi les conopeptides caractérisés, un nombre très important s'est avéré hautement sélectif de divers canaux ioniques et récepteurs de mammifères, associés à des voies de signalisation de la douleur, tels que les récepteurs nicotiniques de l'acétylcholine, le transporteur de la noradrénaline, les canaux sodium, les canaux calcium, le récepteur N-méthyl-D-aspartate, et récepteur de la neurotensine.

II. Peptides et protéines de venins à applications thérapeutiques

II.1 Les inhibiteurs des canaux calcium

Plusieurs peptides spécifiques des canaux calcium sensibles au voltage (VSCC) sont produits par les cônes, les araignées et les serpents. En particulier, plusieurs travaux ont permis de démontrer que des ω-conotoxines spécifiques des VSCC de type N ($Ca_v2.2$) sont de puissants analgésiques (Malmberg et Yaksh, 1995 ; Wang et al., 2000) grâce à l'inhibition des VSCC de type N qui participent dans les voies de la douleur et qui sont sur-exprimés au niveau de la moelle épinière dans les états de douleur chronique (Cizkova et al., 2002). Des doses sub-nanomolaires de ω-conotoxines (ω-MVIIA ou ω-CVID) administrées par voie intrathécale produisent une analgésie pendant plus de 24 heures, dans des modèles de douleur inflammatoires (Smith et al., 2002) et neuropathiques (Scott et al., 2002). La molécule ω-MVIIA (SNX111/Ziconotide/Prialt ; Elan), représentée dans la figure II-A, a conduit au développement du médicament ziconotide, un puissant analgésique approuvé depuis 2004 aux USA et depuis 2005 dans l'UE pour le traitement des douleurs chroniques sévères (McGivern, 2007). La molécule ω-CVID (AM336 ; Zenyth Therapeutics, anciennement Amrad) est également en cours de développement clinique pour les mêmes applications que le ziconotide, mais semble cependant moins toxique que ce dernier (Kolosov et al., 2010).

II.2 Les toxines spécifiques des canaux sodiques

Les canaux sodiques voltage dépendants jouent un rôle essentiel dans le système

nerveux, et sont impliqués dans divers états pathologiques tels que la douleur, les accidents vasculaires cérébraux et l'épilepsie. Les travaux de l'équipe de Michel Lazdunski ont permis la découverte d'une toxine extraite à partir du venin d'araignée *Psalmopoeus cambridgei*, la psalmotoxin 1 (Figure II-B), bloquant de manière spécifique et puissante les canaux ioniques sensibles à l'acidité ASIC de la famille des canaux sodiques, qui sont exprimés dans le système nerveux central et qui sont impliqués dans les voies de la douleur (Escoubas et al., 2000). Des μ-conotoxines telles que GIIIA (Figure II-C) et GIIIB ont également été identifiées comme étant des inhibiteurs spécifiques des canaux sodiques voltage dépendants (Moczydlowski et al., 1986).

II.3 Les bloqueurs des canaux potassium

La margatoxin (MgTX) issue du venin du scorpion *Centruroides margaritatus* (Figure II-D) est un puissant bloqueur des canaux potassiques voltage dépendants Kv1.1, Kv1.2, et en particulier Kv1.3 (Garcia-Calvo et al., 1993). D'un point de vue structural, c'est un peptide de 39 résidus caractérisé par un repliement de type knottin avec deux feuillets β reliés par 3 ponts disulfures à une hélice α. L'utilisation de cette toxine présente un intérêt thérapeutique essentiel du fait notamment de l'implication des canaux potassiques de type Kv1.2 dans la régulation de la libération de neurotransmetteurs associés au rythme cardiaque, la sécrétion d'insuline, l'excitabilité neuronale, le transport d'électrolyte épithélial la contraction du muscle lisse, la réponse immune et le volume cellulaire. Les canaux Kv1.3 sont exprimés dans les lymphocytes T et B. Dans des conditions pathologiques, l'inhibition des canaux Kv1.3 par la margatoxin entraine l'inhibition de la prolifération de cellules T humaines, ce qui conduit à la suppression de la réponse immune *in vivo*. Plusieurs autres toxines extraites de venin de scorpions, telles que la charybdotoxin (ChTX), ont été caractérisées comme inhibiteurs de types spécifiques de canaux potassiques. Cette toxine isolée du scorpion *Leiurus quinquestriatus* (Figure II-E) décrite à l'origine comme inhibiteur des canaux potassiques activés par le Ca^{2+}, présents dans les cellules musculaires et neuroendocrines (Miller et al., 1985), a également été caractérisée comme inhibiteur des canaux potassiques Kv1.3. Malgré des structures II^{aires} et III^{aires} très similaires, la charybdotoxin et la margatoxin issues de deux espèces différentes de scorpions ne présentent que 45% d'homologie de séquences primaires.

La toxine ShK, un peptide de 35 résidus purifié à partir de l'anémone de mer *Stichodactyla helianthus* (Figure II-F) bloque également les canaux Kv1.1 et Kv1.3. Malgré la présence de 3

ponts disulfures, cette toxine est caractérisée par une structure tertiaire différente des toxines précédentes, avec un repliement de type hélice α analogue au domaine « Crisp ». L'efficacité de cette toxine ShK ainsi que ses analogues spécifiques des canaux Kv1.3 a été démontrée dans des modèles animaux pour le traitement des maladies auto-immunes et notamment de la sclérose multiple (Norton et al., 2004).

II.4 Les toxines inhibant les récepteurs nicotiniques de l'acétylcholine

Cette classe de toxines inhibitrices des récepteurs nicotiniques de l'acétylcholine comprend les α-conotoxines de cônes (McIntosh et al., 1999) ainsi que les α-neurotoxines de serpents (Antil-Delbeke et al., 2001). Ces peptides pourraient représenter une alternative à l'utilisation en chirurgie de petites molécules comme myorelaxants mimant le curare (Tuba et al., 2002). Par ailleurs, la toxine Vc1.1, une α-conotoxine purifiée du venin de *Conus victoriae* (Figure II-G), a récemment été identifiée comme ayant des propriétés analgésiques potentielles (Sandall et al., 2003). La toxine Vc1.1 correspond à un petit peptide de 17 résidus, formant une courte hélice α et structuré grâce à 2 ponts disulfures. Cette toxine est en cours de développement clinique pour le traitement des douleurs neuropathiques (Livett et al., 2006).

II.5 Les toxines utilisées dans le traitement des maladies cardiovasculaires

Le médicament élaboré à partir de peptide de venin qui a rencontré le plus de succès est le captopril (Capoten ; Anakena), un anti-hypertenseur très utilisé en médecine depuis 1982. Ce médicament est développé à partir du Teprotide, un peptide de 9 résidus (Pyr-Trp-Pro-Arg-Pro-Gln-Ile-Pro-Pro) isolé du venin du serpent *Bothrops jararaca* (Ondetti et al., 1971). Le captopril agit en inhibant l'enzyme de conversion de l'angiotensine (ACE), une métalloprotéase à zinc qui catalyse la conversion de l'angiotensine I en angiotensine II qui est un vasoconstricteur associé à l'hypertension. Le captopril est aussi utilisé dans l'insuffisance cardiaque et la prévention après un infarctus du myocarde.

Les peptides issus de venins de serpents ont également été utilisés dans le développement d'agents anti-thrombotiques agissant comme antagonistes compétitifs dans la liaison des ligands (tels que le fibrinogène) à l'intégrine αIIbβ3 sur les plaquettes activées. Le composé Eptifibatide (Integrilin; Millennium Pharmaceuticals) est un anti-agrégant plaquettaire, parmi

les inhibiteurs les plus spécifiques de l'intégrine αIIbβ3. Il consiste en un heptapeptide cyclique synthétique dérivé de la protéine barbourin, une désintégrine présente dans le venin du serpent *Sistrurus miliarius barbouri* (Scarborough et al., 1993). La séquence Lys-Gly-Asp-Trp contenue dans la barbourin, modifiée puis incorporée dans la séquence de l'eptifibatide rend ce dernier hautement sélectif de l'intégrine αIIbβ3. Ce médicament est largement prescrit en clinique dans les syndromes coronariens avec menace d'infarctus.

Un autre anti-agrégant plaquettaire très utilisé est le Tirofiban (Aggrastat ; Medicure Pharma) qui agit comme antagoniste spécifique du récepteur au fibrinogène sur les plaquettes (Hartman et al., 1992). Ce médicament Tirofiban est une version modifiée d'un anti-coagulant présent dans le venin de la vipère *Echis carinatus* ; il est approuvé depuis 1998 pour une utilisation en clinique dans la prévention de la transformation d'une angine pectorale instable en un infarctus du myocarde, lorsqu'il est administré avec de l'héparine et de l'aspirine (PRISM-PLUS, 1998).

La protéine ancrod issue du venin de la vipère *Agkistrodon rhodostoma* ainsi que la protéine batroxobin issue du venin du serpent *Bothrops atrox moojeni* sont des protéases à sérine de 234 et 255 acides aminés respectivement, contenant 6 ponts disulfures, analogues à la thrombine, et ayant un rôle anticoagulant par clivage du fibrinogène sanguin. Ces deux molécules semblent réduire les déficits neurologiques lorsqu'elles sont utilisées dans les stades précoces des accidents vasculaires cérébraux (Samsa et al., 2002). Actuellement, la molécule ancrod n'est pas encore approuvée, cependant elle est en cours d'essais cliniques dans le traitement des accidents vasculaires cérébraux. En revanche, la molécule batroxobin est disponible dans le commerce (Defibrase ® et Reptilase ® ; Pentapharm Ltd) pour le traitement des maladies thrombotiques.

Les venins de serpents contiennent aussi des peptides aux propriétés anti-hémorragiques, tels que le peptide textilinin-1, un inhibiteur de sérine protéase anti-fibrinolytique extrait du venin du serpent *Pseudonaja textilis*. Ce peptide de 59 acides aminés ayant 3 ponts disulfures est caractérisé par un repliement de type Kunitz/BPTI (Figure II-H). Des travaux réalisés sur ce peptide ont permis de montrer ses propriétés anti-hémorragiques, mettant en évidence son potentiel thérapeutique pour une utilisation en chirurgie dans les opérations à cœur ouvert (Filippovich et al., 2002). Ce peptide pourrait être utilisé en remplacement de l'aprotinine (Trasylol ; Bayer), un médicament dérivé du BPTI (Bovine Pancreatic Trypsin Inhibitor) dont

l'utilisation bien que très répandue présente des risques de complications chez les patients (Flight et al., 2009). Actuellement, la compagnie biopharmaceutique QRxPharma Ltd possède 3 candidats au stade de développement préclinique pour des applications anti-hémorragiques en chirurgie et issus de venin de serpent australien : Textilinin-1 (Q8008), Haempatch™ (Q8009) : une protéine analogue au facteur Xa et CoVase™ (V0801) : un cofacteur procoagulant (Earl et al., 2012).

II.6 Toxines utilisées dans le traitement du diabète

Le peptide analogue au glucagon-1 (GLP-1) est une hormone intestinale sécrétée en réponse au repas. Il exerce diverses actions métaboliques favorables dont une stimulation glucose-dépendante de la sécrétion d'insuline, une inhibition de la sécrétion de glucagon, un ralentissement de la vidange gastrique, un effet anorexigène central. Or, la sécrétion de GLP-1 est diminuée chez le patient diabétique de type 2, ce qui rend cette hormone une cible thérapeutique d'intérêt dans le traitement de cette maladie. Cette observation a conduit à la découverte d'un peptide présentant une importante homologie avec l'hormone GLP-1, nommé exendin-4 et extrait du venin du lézard communément appelé monstre de Gila ou *Heloderma suspectum* (Eng et al., 1992). Ce peptide de 39 résidus ne possède aucun pont disulfure et est caractérisé par une structure en longue hélice α (Figure II-I). L'exendin-4 induit la libération d'insuline à travers l'activation du récepteur de GLP-1. Une étude réalisée *in vivo* sur des rats a permis de montrer l'effet bénéfique de l'exendin-4 sur la réduction de l'appétit, la diminution du dépôt de graisse, et l'amélioration des paramètres associés à l'intolérance au glucose (Szayna et al., 2000). La version synthétique de la molécule exendin-4 (Exenatide ; Amylin Pharmaceuticals) a été approuvée en 2005 pour le traitement du diabète de type 2.

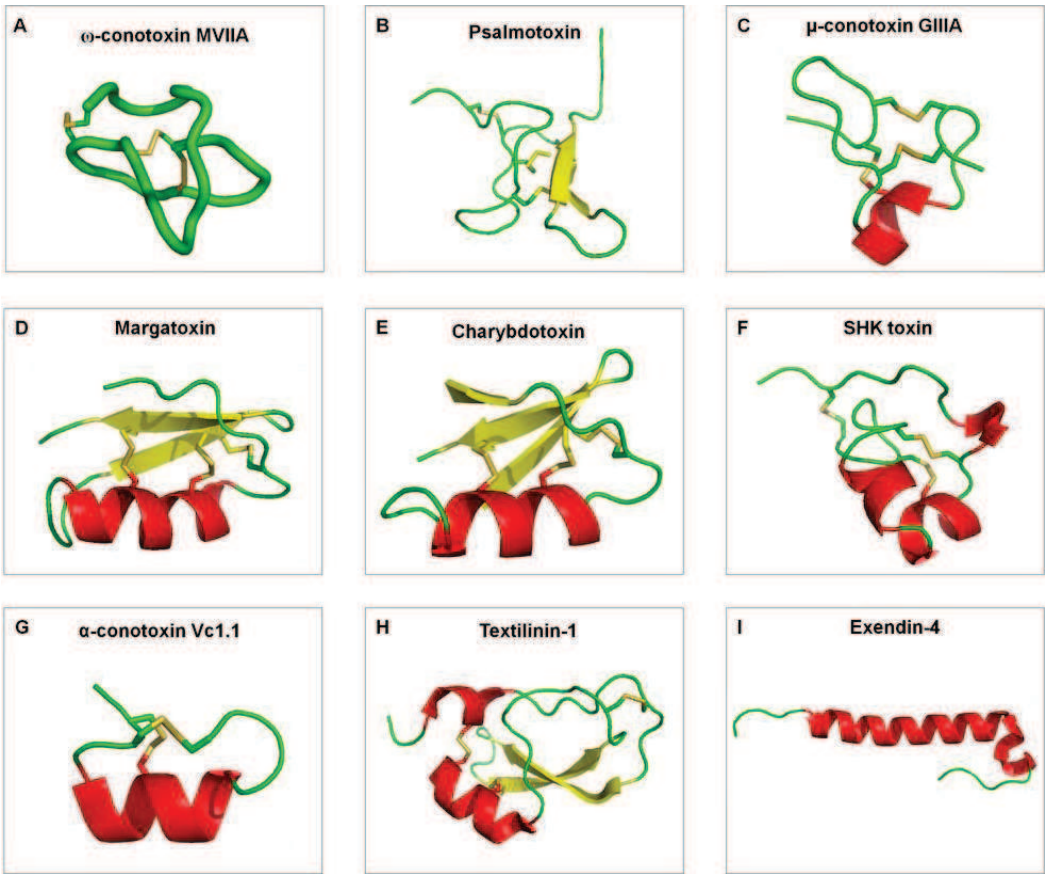


Figure II : Structures 3D de protéines de venins. Les protéines sont représentées en mode cartoon et coloriées selon les structures II^{aires} (hélices en rouge, feuillets en jaune et boucles en vert). Les ponts disulfures sont représentés en bâtons jaunes. A : ω-conotoxin MVIIA (Code PDB 1OMG) ; B : Psalmotoxin (1LMM) ; C : μ-conotoxin GIIIA (1TCG) ; D : Margatoxin (1MTX) ; E : Charybdotoxin (2CRD) ; F : ShK Toxin (1ROO) ; G : α-conotoxin Vc1.1 (2H8S) ; H : Textilinin-1 (3BYB) ; I : Exendin-4 (1JRJ).

III. Les inhibiteurs de protéases dans les venins

III.1 Les inhibiteurs de protéases à sérine

Plusieurs peptides inhibant l'activité protéolytique de la kallikréine, la trypsine, la plasmine et la chymotrypsine ont été trouvés dans les venins de serpents, en particulier ceux de la famille des vipéridés et des élapidés. Le peptide textilinin-1 cité plus haut fait partie de cette classe d'inhibiteurs de sérine protéase. Les travaux de Lu et al. ont permis d'identifier un inhibiteur de sérine protéase dans le venin du serpent *Bungarus fasciatus*, nommé bungaruskunin. Cette protéine de 59 résidus est structurée grâce à la présence de 3 ponts disulfures et montre un repliement de type Kunitz/BPTI ; elle possède une activité inhibitrice contre les enzymes trypsine, chymotrypsine et élastase (Lu et al., 2008).

Plusieurs inhibiteurs de protéase à sérine présentant une importante homologie de séquence avec le BPTI et inhibant l'activité de la trypsine bovine ont également été identifiés dans le venin des serpents *Hemachatus haemachatus* et *Naja nivea* (Hokama et al., 1976) ainsi que dans le venin de la vipère *Vipera russelli* (Takahashi et al., 1974).

III.2 Les inhibiteurs de protéases à cystéine

Le venin du cobra *Naja naja atra* contient un inhibiteur de cystéine protéase, appelé cystatin de cobra, une protéine de 99 acides aminés contenant seulement 1 pont disulfure. La cystatin possède une activité inhibitrice des protéases à cystéine telles que la papaïne, la cathepsine L, la cathepsine B et la cathepsine S avec des Ki compris entre 0,1 et 2,5 nM (Brillard-Bourdet et al., 1998).

III.3 Les inhibiteurs de métalloprotéases

Le fractionnement du sérum du serpent *Bothrops jararaca* a permis d'isoler une protéine de taille importante (322 acides aminés) ayant 6 ponts disulfures, la protéine BJ46a (Valente et al., 2001). L'étude réalisée par Valente et al. a permis de démontrer que la protéine BJ46a est un puissant inhibiteur endogène des métalloprotéases de venin de serpents (SVMPs) telles que les métalloprotéases à zinc atrolysine C et jararhagin. Par ailleurs, la protéine BJ46a peut également être utilisée dans l'étude de métalloprotéases reliées au SVMPs, telles que les MMPs. En effet, les deux familles de métalloprotéases à zinc SVMPs et MMPs, appartenant à la superfamille des metzincines, sont caractérisées par une grande similarité topologique notamment au niveau du site actif, ainsi qu'une importante homologie de séquences (Figure III), avec notamment la présence dans ces deux familles du motif très conservé de liaison au zinc catalytique du site actif (Stocker et al., 1995). De plus, il a été démontré que l'inhibiteur synthétique des MMPs, batimastat ou BB-94, neutralise l'activité d'une SVMP de *Bothrops asper* nommée BaP1 (Escalante et al., 2000), suggérant que les SVMPs et les MMPs sont inhibées selon un mécanisme très similaire. Des inhibiteurs endogènes des SVMPs issus des venins de serpents, tels que la protéine BJ46a, pourraient donc potentiellement être capables d'inhiber nos cibles d'intérêt les MMPs.

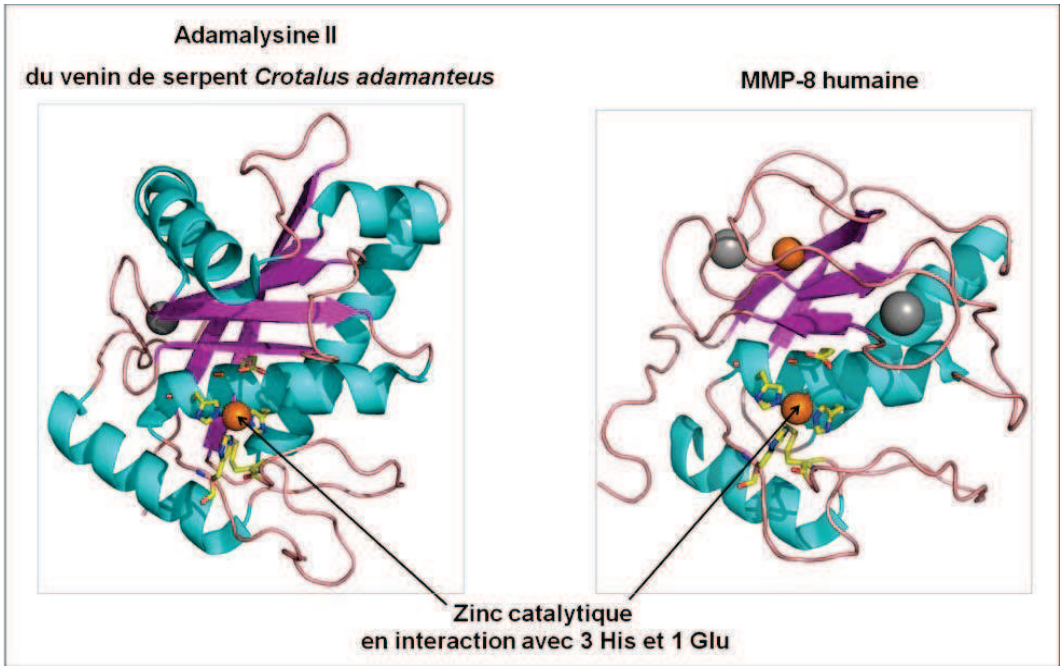


Figure III : Structures 3D de l'adamalysine II issue du venin du serpent *Crotalus adamanteus* (code PDB 1IAG) et de la MMP-8 humaine (code PDB 3DPE). Les protéines sont représentées en mode cartoon et coloriées en fonction des structures II^{aires} : hélices en cyan, feuillets en mauve et boucles en rose. Les ions sont représentés en sphères, de couleur gris pour le calcium et orange pour le zinc. Les résidus strictement conservés du motif de chélation du zinc (3 résidus His et un résidu Glu) ainsi que le résidu Met caractéristique de la famille des metzincines sont représentés en bâtons jaunes.

Les travaux de Huang et ses collaborateurs ont permis d'isoler et de purifier 3 petits peptides à partir du venin du serpent *Trimeresurus mucrosquamatus* : pyroGlu-Lys-Trp, pyroGlu-Asn-Trp et pyroGlu-Gln-Trp (Huang et al., 1998). Des études antérieures sur ces peptides contenant un pyro-glutamate, et retrouvés dans plusieurs espèces de serpents de la famille des vipéridés et des crotalidés, ont montré une activité inhibitrice spécifique des métalloprotéases à zinc présentes dans le venin de plusieurs espèces de serpents (Robeva et al., 1991 ; Francis et Kaiser, 1993).

IV. Etude d'un cas particulier : le venin du serpent *Dendroaspis polylepis*

Le venin auquel nous nous sommes intéressés dans ce travail de recherche est celui du mamba noir *Dendroaspis polylepis*, le plus grand et le plus répandu des serpents venimeux d'Afrique et le deuxième plus grand serpent venimeux au monde après le cobra *Ophiophagus hannah*, mais également parmi les serpents les plus venimeux au monde. Une électrophorèse capillaire du venin de ce serpent a montré la présence d'au moins 70 peptides de poids

moléculaires compris entre 6000 et 9000 Da (Perkins et al., 1993).

Le venin de ce serpent est particulièrement riche en neurotoxines telles que : 1) les dendrotoxines : des peptides de 57 à 60 acides aminés réticulés par 3 ponts disulfures et homologues au BPTI, qui bloquent de manière sélective les canaux potassiques voltage dépendants (Harvey et Robertson, 2004) ; 2) les fasciculines : des peptides d'environ 60 acides aminés, ayant 4 ponts disulfures, qui inhibent l'acétylcholinestérase ; 3) les toxines muscariniques : ces toxines spécifiques de la famille des mambas sont constituées de 64 à 66 résidus formant un repliement de type 3 doigts, et inhibent les récepteurs muscariniques de l'acétylcholine. Le venin du serpent *Dendroaspis polylepis* contient également des cardiotoxines dont la fonction est d'accroître la perméabilité des membranes biologiques, la calciseptine (De Weille et al., 1991) et le peptide FS2 (Albrand et al., 1995) qui sont des bloqueurs de canaux calciques voltage dépendants, ainsi qu'une grande quantité d'acétylcholine (Welsh, 1967) ayant un effet coopératif avec les fasciculines.

Outre ce grand nombre de peptides et de toxines présents dans le venin de *Dendroaspis Polylepis*, un inhibiteur de trypsine homologue au BPTI ayant une taille de 57 résidus et 3 ponts disulfures, nommé toxine Vi2, a également été identifié (Strydom, 1977), démontrant la possibilité de trouver des inhibiteurs de protéases dans le venin de ce serpent.

V. Conclusion

Les études citées ci-dessus montrent que les venins des animaux sont une source importante de peptides et de protéines ayant des tailles, des structures et des séquences totalement différentes et agissant sur diverses cibles d'intérêt thérapeutique. Parmi les différents constituants des venins, des inhibiteurs de métalloprotéases ont été mis en évidence, dont le plus connu est l'inhibiteur de l'enzyme de conversion de l'angiotensine couramment utilisé en clinique humaine en tant qu'antihypertenseur. Le criblage du venin du serpent *Dendroaspis polylepis* auquel nous sous sommes intéressés au cours de ce travail pourrait donc aboutir à l'identification de composés ayant un intérêt thérapeutique ou diagnostique. Une approche expérimentale reposant sur le fractionnement du venin du serpent *Dendroaspis polylepis* suivi d'un criblage des différentes fractions obtenues a été mise au point dans le cadre de ce travail de thèse afin de rechercher dans le venin de ce serpent des mini-protéines actives sur les MMPs.

PARTIE B :

Conception *in silico* de ligands protéiques

CHAPITRE I : Ingénierie des protéines pour le développement de ligands

I.1 Les protéines thérapeutiques

I.1.1 Les protéines : de leur connaissance fondamentale aux applications thérapeutiques

Les protéines ont le rôle le plus dynamique et le plus diversifié parmi les macromolécules du vivant, catalysant des réactions biochimiques, formant des récepteurs et des canaux dans les membranes, transportant les molécules à l'intérieur d'une cellule ou d'un organe à un autre. On estime qu'il existe entre 25000 et 40000 gènes différents chez l'homme, et avec l'épissage alternatif des gènes et les modifications post-traductionnelles des protéines (par clivage, phosphorylation, acylation et glycosylation), le nombre de protéines fonctionnellement distinctes est encore plus élevé. Lorsque l'une de ces protéines contient des mutations ou d'autres anomalies, ou est présente à une concentration anormalement élevée ou faible, il peut en résulter des maladies dont la compréhension des mécanismes représente un immense défi pour la médecine moderne. D'un point de vue thérapeutique, l'exploitation des protéines représente une opportunité pour limiter les maladies ou les diagnostiquer. Depuis 1980, les protéines ont émergé en tant que nouvelle classe de molécules pharmaceutiques. Le secteur des protéines thérapeutiques représente une part significative et croissante du marché pharmaceutique global, avec des ventes totalisant 94 milliards de dollars en 2007. Les protéines thérapeutiques recombinantes (hormis les anticorps) ont enregistré des ventes globales de 61 milliards de dollars en 2009, tandis que les ventes de médicaments à base d'anticorps monoclonaux ont atteint 38 milliards de dollars supplémentaires (Walsh, 2010). Actuellement, plus de 130 protéines ou peptides différents sont approuvés par la FDA pour une utilisation en clinique, et un nombre plus important de protéines est en cours de développement (Leader et al., 2008).

Les protéines thérapeutiques peuvent présenter certains avantages par rapport aux petites molécules-médicaments. Les protéines possèdent à leur surface un ensemble complexe de fonctions chimiques qui peuvent difficilement être mimées par des composés chimiques de faible poids moléculaire. L'action des protéines est le plus souvent très spécifique et on

considère généralement comme peu probable que les protéines thérapeutiques interfèrent avec les processus biologiques normaux et causent des effets défavorables ou toxiques. L'organisme produit naturellement un grand nombre des protéines pouvant être utilisées comme molécules thérapeutiques, qui sont donc moins susceptibles d'induire des réponses immunes. Pour les maladies dans lesquelles un gène est muté ou supprimé, les protéines thérapeutiques peuvent fournir un traitement de remplacement effectif sans avoir recours à la thérapie génique, ce qui n'est pas possible actuellement pour la plupart des maladies génétiques. La durée de développement clinique ainsi que celle de l'acceptation par la FDA (Food and Drug Administration) des protéines thérapeutiques peut être plus rapide que celles des petites molécules-médicaments. Une étude publiée en 2003 a montré que la durée moyenne de développement clinique et d'acceptation par la FDA pour 33 protéines thérapeutiques approuvées entre 1980 et 2002 a été réduite de plus d'une année par rapport à 294 petites molécules-médicaments approuvées pendant la même période (Reichert, 2003). Enfin, les protéines étant uniques en termes de forme et de fonction, les compagnies sont capables d'obtenir, pour les protéines thérapeutiques, une meilleure protection industrielle.

L'utilisation des protéines thérapeutiques a commencé avec l'administration aux diabétiques d'insuline extraite à partir de pancréas bovin et porcin. Puis, le passage aux protéines recombinantes a eu lieu au début des années 1980 avec la production d'insuline humaine (Humulin® crée par Genentech) par des bactéries *E. coli*. Cette insuline recombinante approuvée par la FDA en 1982 a été la 1ère protéine thérapeutique recombinante disponible dans le commerce, et a depuis cette date été utilisée comme thérapie principale dans le traitement du diabète sucré de type I et II afin de remplacer l'insuline naturelle déficiente dans ces types de diabète. Cependant, cette insuline recombinante s'est avérée incapable de mimer de manière adéquate l'insuline endogène concernant deux aspects majeurs : une action rapide après les repas et le maintien prolongé d'un faible taux d'insuline entre les repas. L'insuline a donc été modifiée, par remplacement de deux ou trois acides aminés, afin de créer des analogues à action rapide ou à effet prolongé, et qui reproduisent mieux les différentes propriétés de l'insuline endogène. Plusieurs analogues d'insuline ont été approuvés pour le traitement du diabète sucré, tels que les analogues à action rapide : insuline lispro (Humalog®), insuline aspart (NovoRapid® et Novolog®) et insuline glulisine (Apidra®), ainsi que les analogues à effet prolongé : insuline glargine (Optisulin® et Lantus®) et insuline detemir (Levemir®).

L'importance commerciale des insulines modifiées est illustrée par Lantus®, qui fut l'un des dix produits biopharmaceutiques les plus vendus en 2009 (Walsh, 2010). Ces exemples illustrent bien un des atouts des protéines dans le contexte thérapeutique : leurs propriétés pharmacocinétiques peuvent être optimisées grâce à des modifications ponctuelles de quelques acides aminés sans altérations de leur fonction. La modulation de propriétés pharmacodynamiques de protéines thérapeutiques peut également être réalisée si celles-ci sont produites dans des cellules de mammifères via l'insertion des N-et/ou O-glycosylations. Celles-ci peuvent augmenter la demi-vie dans le sérum et par conséquent le temps d'exposition à la cible. Un exemple de telles modifications de protéines par glycosylation pour une demi-vie prolongée est celui de la darbepoetin-α (Aranesp®), administrée à des patients anémiques afin de stimuler la production de globules rouges dans la moelle osseuse (Carter, 2011). La darbepoetin-α est obtenue à partir d'une molécule de 1ère génération : l'érythropoïétine native, un facteur de croissance important dans la production de globules rouges, qui possède trois N-glycosylations et qui a été modifié par ajout de 2 autres sites de glycosylation afin de créer la darbepoetin-α, un médicament de 2ème génération ayant une demi-vie trois fois plus longue que son analogue natif.

Un autre progrès significatif dans le domaine des protéines thérapeutiques s'est produit avec l'avènement des protéines de fusion, par fusion de gènes codant pour deux ou plusieurs protéines différentes. Les protéines de fusion combinent les propriétés de leurs parties composantes. La protéine thérapeutique de fusion la plus réussie à ce jour du point de vue clinique et commercial contient la région Fc (Fragment cristallisable) d'immunoglobulines également destinée à augmenter la demi-vie plasmatique (plusieurs jours à plusieurs semaines) des protéines auxquelles il est lié grâce à la liaison au récepteur FcRn. Les protéines de taille inférieure à 70 kDa sont souvent rapidement éliminées de la circulation par filtration rénale et peuvent avoir des demi-vies de quelques minutes à quelques heures, ce qui dans certains cas peut les rendre inutilisables dans des applications thérapeutiques. Le 1er exemple de protéine de fusion à la région Fc d'IgG est le CD4-Fc (Capon et al., 1989). Jusqu'en 2010, cinq protéines thérapeutiques de fusion Fc ont été approuvées. Le médicament de cette classe qui rencontre le plus de succès est Etanercept (Enbrel®), une protéine de fusion Fc-TNFR 2 (Walsh, 2010); en 2009, Etanercept a été la protéine thérapeutique la plus vendue dans le monde pour le traitement de l'arthrite rhumatoïde et d'autres maladies auto-immunes. Romiplostim (Nplate®),

un agoniste du récepteur de la thrombopoïétine, est la 1ère fusion Fc-peptide (peptibody) à être approuvée en tant que produit thérapeutique humain pour le traitement de la thrombocytopénie (Cines et al., 2008). D'autres exemples de protéines thérapeutiques, leur fonction, ainsi que leur utilisation dans le domaine clinique sont rapportés dans le tableau I.1.

Tableau I.1: Différentes protéines thérapeutiques classées selon leur activité pharmacologique (Carter, 2011).

Protéine	Nom commercial	Fonction	Utilisation clinique
Groupe Ia : Protéines thérapeutiques remplaçant une protéine déficiente ou anormale			
Insuline	Humulin, Novolin	Régule le taux de glucose dans le sang, déplace le potassium dans les cellules	Diabète mellitus, acidocétose diabétique, hyperkaliémie
Hormone de croissance : Somatotropine	Genotropin, Humatrope, Norditropin, NorlVitropin, Nutropin,	Effecteur anabolique et anti-catabolique	Problèmes de croissance, Syndrome de Prader-Willi , Syndrome de Tumer , Syndrome cachectique lié au VIH
Facteur VIII	Bioclate, Helixate, Recombinate, ReFacto	Facteur de coagulation sanguine	Hémophilie A
β-Gluco-cérébrosidase	Cerezyme	Hydrolyse le glucocérébroside en glucose and céramide	Maladie de Gaucher
Groupe Ib : Protéines thérapeutiques augmentant une voie existante			
Hormone folliculos-timulante humaine (FSH)	Gonal-F, Follistim	Augmente l'ovulation	Reproduction assistée
Interféron-β1a (rIFN-β)	Avonex, Rebif	Mécanisme inconnu ; Rôle antiviral et immuno-régulateur	Sclérose en plaques
Groupe Ic : Protéines thérapeutiques conférant une nouvelle fonction ou activité			
Toxine Botulique de type A	Botox	Clive SNAP25 au niveau des jonctions neuromusculaires et empêche la libération d'acétylcholine release, causant une paralysie flaccide	Plusieurs types de dystonie, en particulier dystonie cervicale ; Utilisation en cosmétique
Groupe IIa : Protéines thérapeutiques interférant avec une molécule ou un organisme			
Panitumumab	Vectibix	mAb humain qui se lie à EGFR	Cancer colorectal métastatique
Enfuvirtide	Fuzeon	Peptide de 36 acides aminés inhibant l'entrée du HIV dans les cellules hôtes en se fixant à la protéine gp120/gp41 de l'enveloppe virale	Adultes et enfants (d'au moins 6 ans) atteints d'une infection au VIH à un stade avancé
Groupe IIb : Protéines thérapeutiques délivrant des protéines ou autres composés			
Denileukine diftitox	Ontak	Dirige l'action cytocide de la toxine diphtérique aux cellules exprimant le récepteur de l'IL2	Lymphomes cutanés persistants ou récurrents dont les cellules malignes expriment la sous-unité CD25 du récepteur de l'IL2
Groupe III : Protéines vaccins			
Vaccin HPV	Gardasil	Vaccin recombinant quadrivalent du virus HPV (Types 6, 11, 16, 18), contenant les protéines capsides de 4 souches de virus HPV	Prévention de l'infection par le virus HPV
Antigène de surface de l'hépatite B (HBsAg)	Engerix, Recombivax HB	Protéine non infectieuse à la surface du virus de l'hépatite B	Vaccin contre l'hépatite B
Groupe IV : Protéines diagnostiques			
Protéine recombinante purifiée (DPPD)	DPPD	Protéine non infectieuse de *Mycobacterium tuberculosis*	Diagnostic d'une exposition à la tuberculose

I.1.2 Limitations des protéines comme outils thérapeutiques

Un certain nombre d'inconvénients limitent le développement et le succès des protéines thérapeutiques, tels que la solubilité, la voie d'administration, la distribution et la stabilité. En effet, les protéines sont des molécules pouvant atteindre une taille importante et possédant des propriétés physicochimiques (hydrophiles/hydrophobes) pouvant rendre difficile l'entrée dans les cellules ou dans certains compartiments. La distribution intracellulaire étant à ce jour un objectif encore difficile à atteindre, les protéines thérapeutiques sont donc limitées à une utilisation à la surface cellulaire ou pour des cibles extracellulaires. La demi-vie d'une protéine thérapeutique peut être affectée de manière drastique par les protéases, les produits chimiques et d'autres mécanismes de clairance. Toutefois, comme nous l'illustrent les exemples du § I.1.1, les protéines peuvent faire l'objet d'une modulation des propriétés pharmacodynamiques via la modification de leur surface sans interférer avec les propriétés de liaison à la cible thérapeutique.

Un second inconvénient des protéines thérapeutiques est qu'elles peuvent induire une réponse immunitaire conduisant à la neutralisation de ces protéines et pouvant même causer des réactions nocives chez le patient. Par exemple, une forme tronquée et pégylée de la thrombopoïétine (MGDF-PEG ; MGDF pour Megakaryocyte Growth and Development Factor) a induit une réponse immunitaire chez certains patients qui a conduit à la neutralisation de la thrombopoïétine endogène, entraînant une anémie sévère et persistante à l'opposé de l'effet recherché consistant à augmenter la numération plaquettaire (Li et al., 2001). Il convient de noter que des méthodes sont aujourd'hui en cours de développement pour moduler la réponse immunitaire des protéines thérapeutiques (Baker et Jones, 2007).

Une autre limitation de l'utilisation des protéines thérapeutiques concerne les modifications post-traductionnelles telles que la glycosylation, la phosphorylation et le clivage protéolytique qui sont nécessaires à l'activité physiologique des protéines. Ces modifications nécessitent l'utilisation de cellules spécifiques qui sont capables d'exprimer et de modifier la protéine de manière appropriée. Par ailleurs, le système hôte doit permettre la production à grande échelle de la protéine recombinante afin de satisfaire la demande clinique, tout en maintenant l'activité et la stabilité de la protéine pour une durée de conservation prolongée. Les protéines

thérapeutiques impliquent souvent des coûts de production élevés pouvant limiter l'accessibilité aux patients ainsi que les applications cliniques. Un exemple d'une telle limitation concerne la production par voie recombinante de la β-glucocérébrosidase pour le traitement de la maladie de Gaucher dont le coût est supérieur à 100,000 US$ par patient et par an (Gross, 2002).

Cependant, les possibles inconvénients des protéines dans le contexte thérapeutique n'ont pas freiné les développements dans ce domaine et leur potentiel important a au contraire stimulé les recherches visant à résoudre les inconvénients de cette classe de molécules pour les applications thérapeutiques.

I.2 Les Anticorps

Outre les protéines recombinantes, les entités biopharmaceutiques comptent une fraction sans cesse croissante d'anticorps (immunoglobulines, Ig). Depuis 2007, plus de 40 anticorps monoclonaux par an entrent en phase d'études cliniques. Les anticorps sont donc devenus la classe de médicaments-protéines à l'évolution la plus rapide, avec, au début de l'année 2011, 28 anticorps thérapeutiques approuvés par la FDA sur le marché américain dont 10 anticorps pour le traitement de divers types de cancers, 14 pour le traitement de maladies immunologiques et 4 pour le traitement de maladies inflammatoires ou virales (Reichert, 2011). Deux autres anticorps thérapeutiques ont été approuvés par la FDA pour une commercialisation en dehors des USA. De plus, environ 240 anticorps sont en cours d'essais cliniques (Reichert, 2010 ; Reichert, 2011).
Les cibles naturelles des anticorps (les antigènes) sont des molécules de surface appartenant à des entités intruses (bactéries, virus), incluant des polypeptides et des carbohydrates. Cependant, grâce à leur variabilité illimitée, un nombre infini de molécules organiques sont des antigènes potentiels pouvant être capturés avec une très grande affinité et spécificité.

Les anticorps conventionnels consistent en une paire de chaines lourdes H et une paire de chaines légères L plus courtes, reliées entre elles par un nombre variable de ponts disulfures (Figure I.1). La région Fab (Fragment antigen binding) d'un anticorps consiste en un domaine variable (VH) et un domaine constant (CH) d'une chaine lourde ainsi qu'un domaine variable (VL) et un domaine constant (CL) d'une chaine légère. Au sommet de la molécule, les régions hypervariables déterminant la complémentarité avec l'antigène (CDR) et qui sont au nombre de

trois dans chacun des domaines VH et VL, forment le paratope ou site de reconnaissance de l'antigène. La région Fc d'un anticorps est constituée des domaines constants (CH) de chaines lourdes ; elle permet de maintenir la stabilité et est impliquée dans diverses fonctions du système immunitaire telles que l'activation du système du complément et la liaison aux récepteurs FcγR sur des cellules cibles permettant l'élimination du pathogène par phagocytose (par le biais des macrophages) ou par un phénomène de cytotoxicité (par le biais des lymphocytes) (Nuttall et Walsh, 2008).

En fonction de leur origine, les anticorps peuvent être monoclonaux (mAb) ou polyclonaux (pAb). Les anticorps polyclonaux sont purifiés à partir du sérum d'animaux immunisés ; c'est un ensemble de variants d'anticorps qui se lient à différents épitopes d'un antigène cible, et sont peu spécifiques. Les anticorps monoclonaux sont constitués d'un variant unique d'anticorps dérivé d'une lignée cellulaire unique (hybridome) ; les anticorps monoclonaux se lient à un seul épitope sur un seul antigène, et sont donc plus spécifiques que les anticorps polyclonaux.

I.2.1 Applications

Les anticorps sont souvent utilisés dans les laboratoires de biochimie pour des tests de diagnostique de routine tels que l'ELISA et le Western blot. Dans le cadre médical, ces tests sont généralement utilisés pour la détection de maladies infectieuses et parasitaires. Le test ELISA est également utilisé pour le criblage de routine des aliments et échantillons environnementaux pour détecter une contamination chimique et microbienne. Le test de grossesse est un exemple connu de test diagnostique peu cher et facile à utiliser, qui repose sur l'utilisation d'anticorps pour la détection de l'hormone chorionique gonadotrope HCG (Cabrera, 1969). Le nombre de tests diagnostiques disponibles a rapidement augmenté ces dix dernières années, par exemple, en 2004 la FDA a approuvé le premier test rapide pour la détection du virus de l'immunodéficience humaine (HIV) dans les fluides oraux (Branson, 2004).

Cependant, une des applications les plus développées des anticorps est actuellement leur utilisation comme agents thérapeutiques. Muromonab-CD3 (Orthoclone OKT3), un anticorps anti-CD3, est le 1er anticorps monoclonal murin à avoir été approuvé par la FDA en 1986 en

tant qu'agent thérapeutique pour éviter le rejet d'organe après transplantation (Goldstein et al., 1985). L'administration de ces anticorps murins est cependant soumise au risque d'une réponse immune non désirée (immunogénicité) et d'autres effets secondaires (Waldmann, 2003). Les effets secondaires de ces anticorps devaient être réduits et c'est seulement en 1994 que le 2ème mAb Abciximab (ReoPro), un anticorps chimérique anti-GPIIb/IIIa, a reçu l'approbation de la FDA.

Les anticorps sont également utilisés pour l'administration ciblée des médicaments. A cet effet, les mAb sont marqués par des radio-isotopes ou des médicaments toxiques, combinant la spécificité de l'anticorps avec une toxicité élevée due à leur marquage. Deux radio-immunoconjugués (Ibritumomab tiuxetan (Zevalin) et Tositumomab-131I (Bexxar), deux anticorps anti-CD20) ont été approuvés par la FDA (Wu et Senter, 2005), tandis que le seul conjugué avec une toxine ayant été approuvé par la FDA (Gemtuzumab ozogamicin (Mylotarg), un anti-CD33) a récemment été retiré du marché du fait d'inquiétudes relatives à l'innocuité du produit et aux bienfaits cliniques pour les patients (FDA, 2010). Les radio-immunoconjugués peuvent également être utilisés pour l'imagerie *in vivo* des tumeurs, afin d'évaluer le ciblage ou le succès de la thérapie. La liaison spécifique des anticorps à leurs cibles permet un bon rapport signal/bruit. Mais, dans ce contexte, leur demi-vie plasmatique élevée constitue un inconvénient pour la qualité des images obtenues aux temps courts.

I.2.2 Atouts des anticorps

Il existe plusieurs raisons au succès remarquable des anticorps en tant que classe de médicaments biologiques. Ils peuvent être rapidement générés contre une large gamme de molécules cibles (antigènes ou haptènes), soit par immunisation classique d'animaux suivie par des protocoles pour la préparation d'anticorps monoclonaux, ou plus récemment par sélection *in vitro* à partir de bibliothèques de gènes synthétiques ou clonés (Gebauer et Skerra, 2009).
De plus, ils possèdent des spécificités remarquables pour leurs cibles, avec des affinités souvent dans la gamme du nanomolaire au picomolaire, surpassant ainsi la plupart des médicaments chimiques, et les rendant appropriés pour les applications diagnostiques et thérapeutiques. Par ailleurs, ils ont une structure large et complexe, qui leur permet de se lier aux antigènes et de recruter de façon simultanée des composants du système immunitaire, ce qui constitue une propriété essentielle dans certaines formes d'immunothérapie (Lutterotti et Martin, 2008).

De nos jours, les avancées technologiques ont permis la production hétérologue d'anticorps chimériques en fusionnant des domaines variables murins responsables de l'activité de liaison avec des domaines constants humains, conduisant au développement d'une nouvelle génération d'anticorps thérapeutiques (Neuberger et al., 1985). Ces anticorps chimériques sont d'origine humaine à 70% et possèdent une portion Fc entièrement humaine, ce qui les rend considérablement moins immunogéniques chez l'homme, et leur permet d'interagir avec les cellules effectrices humaines et la cascade du complément. Un exemple d'anticorps chimérique ayant rencontré un grand succès dans la thérapie du cancer est celui utilisé dans le Rituximab (Mabthera), un anti-CD20 destiné au traitement des lymphomes non hodgkiniens à cellules B exprimant l'antigène de surface CD20 (Schulz et al., 2007). Par la suite, il est devenu possible de créer des anticorps humanisés en remplaçant les CDR d'un anticorps entièrement humain par des CDR d'origine murine (Almagro et Fransson, 2008). Ces anticorps humanisés, d'origine humaine à 85-90%, ont moins d'effets secondaires que les anticorps chimériques et une immunogénicité réduite. L'état de l'art actuel porte sur des anticorps entièrement humains grâce au développement des méthodes de sélection *in vitro*, notamment la technique du « phage display » et plus récemment celle du « ribosome display ». Une approche complémentaire pour créer des anticorps entièrement humains consiste à les produire via des souris transgéniques portant un locus Ig humain (Lonberg, 2008a ; Lonberg, 2008b).

Le succès des anticorps en tant qu'agents thérapeutiques s'explique également par l'industrialisation des technologies de production des anticorps permettant de fournir des quantités suffisantes en anticorps, rendant possible l'utilisation de doses plus élevées chez les patients. Par ailleurs, l'expression de fragments d'anticorps chez *E. coli* a connu une avancée technologique permettant une production en fermenteur à haut rendement (gramme/litre) (Carter et al., 1992), dont l'exemple est le médicament Ranibizumab (Lucentis®), un anti-VEGF pour le traitement de la dégénérescence maculaire humide.

Concernant les aspects pharmacocinétiques, les anticorps sont de larges molécules qui sont caractérisées par une très longue demi-vie dans le sérum, allant de plusieurs jours à plusieurs semaines (Chames et al., 2009) ; ils dépassent de loin le seuil de clairance rénale (70 kDa), leur

évitant ainsi d'être éliminés par les glomérules rénaux. De plus, la portion Fc des molécules IgG peut interagir avec divers récepteurs exprimés à la surface de plusieurs types cellulaires, ce qui augmente leur rétention dans la circulation. Par ailleurs, comme décrit au § I.1.1, la portion Fc peut interagir avec le récepteur Fc néonatal (FcRn) exprimé à la surface de plusieurs types cellulaires ce qui permet d'augmenter la demi-vie dans le sérum des IgG et albumines (qui représentent 90% des protéines sériques) d'un jour et jusqu'à plusieurs semaines.

L'efficacité et l'innocuité cliniques des anticorps ont été bien établies, incluant les aspects de spécificité pour l'épitope, l'immunogénicité, la pharmacocinétique et les fonctions effectrices associées à l'immunité, impliquant une acceptation générale par les médecins et les patients.

I.2.3 Limitations des anticorps

L'utilisation croissante des anticorps a permis de mettre plusieurs inconvénients en évidence. Du fait de leur structure compliquée, les anticorps peuvent être soumis à la dégradation, agrégation, modification (comme l'oxydation ou la déamidation) et dénaturation. Leur utilisation est par conséquent limitée à des conditions semblables à leur environnement physiologique. En présence de solvants organiques, ou dans des conditions non physiologiques (température élevée, pH élevé et une concentration élevée en sel), ils perdent généralement leurs fonctions.

Les anticorps n'ont pas encore été produits avec succès dans des hôtes microbiens ; ceci est dû à leur composition compliquée, comprenant quatre chaînes polypeptidiques, des glycosylations au niveau des chaînes lourdes, et au moins un pont disulfure ayant un rôle structural essentiel dans chacun des divers domaines Ig. Ainsi, les anticorps entiers nécessitent une production dans les systèmes d'expression eucaryotes, impliquant des lignées cellulaires de mammifères (cellules d'ovaires de hamster chinois CHO et cellules embryonnaires de rein humain HEK) transfectées de façon stable, dont l'optimisation et la fermentation est laborieuse et coûteuse. De plus, la plupart des études montrent que ces molécules doivent être injectées en grandes quantités afin d'atteindre une efficacité clinique (par exemple : 6 à 12 g/patient pour le Rituximab) (Chames et al., 2009). La production de ces anticorps nécessite donc l'utilisation de très grandes cultures de cellules de mammifères suivies par des étapes de purification

extensives, conduisant à des coûts de production élevés et limitant ainsi l'utilisation de ces médicaments.

De plus, la grande taille caractéristique des anticorps rend difficile la pénétration dans les tumeurs solides. En effet, à l'exception des maladies et cancers hématologiques, les anticorps thérapeutiques doivent pénétrer dans les tissus et la matrice extracellulaire afin d'atteindre leurs cellules cibles. La pénétration tissulaire, en particulier dans le cas des tumeurs solides, est un paramètre essentiel, limitant dans la plupart des cas l'efficacité globale du traitement. Or les larges molécules telles que les anticorps diffusent faiblement, atteignant donc difficilement les tumeurs. En effet, dans des modèles murins de xénogreffe, les mAbs dirigés contre des antigènes tumoraux spécifiques restent en majorité dans le sang et une proportion maximale de 20% de la dose administrée interagit avec la tumeur (Beckman et al., 2007). Ceci représente probablement une des limitations majeures rencontrées par les mAbs. Ainsi, sur neuf anticorps acceptés pour le traitement des tumeurs, seuls trois (Trastuzumab, Cétuximab, Panitumumab) ont pour cible un marqueur exprimé dans les tumeurs solides, ces dernières représentant pourtant 85% des cancers (Chames et al., 2009). Un autre facteur limitant la pénétration des anticorps dans les tumeurs est l'effet de barrière dû aux sites de liaison. En effet, les anticorps très affins se lient fortement à leurs antigènes dès leur première mise en présence à la périphérie de la tumeur, ils ne peuvent pas pénétrer plus en profondeur dans la tumeur jusqu'à ce que toutes les molécules d'antigènes soient saturées à la périphérie. En revanche, des molécules ayant une affinité modérée seront libérées de ces antigènes et pourront alors pénétrer en profondeur dans la tumeur (Fujimori et al., 1990). Adams et ses collaborateurs ont étudié la biodistribution dans une tumeur d'une série de mutants de ScFv (Single chain Fragment variable) radiomarqués ayant des affinités pour leur cible HER-2/*neu* comprises entre 10^{-6} et 10^{-11} M. Les résultats obtenus ont démontré que le mutant ScFv ayant la plus faible affinité entraine une coloration diffuse de la tumeur tandis que le mutant ayant l'affinité la plus élevée a été retenu dans les régions périvasculaires de la tumeur (Adams et al., 2001).

Par ailleurs, lorsqu'ils sont administrés par voie intraveineuse, les anticorps peuvent avoir des demi-vies de plusieurs semaines (Carter, 2006). Une longue demi-vie est requise pour certaines applications car elle augmente l'efficacité et réduit la dose requise. D'un autre côté, pour d'autres applications, des demi-vies courtes sont nécessaires, par exemple lorsqu'une clairance rapide d'immunoconjugués toxiques est requise pour réduire une exposition de tout le corps.

Ceci est également vrai dans le contexte de l'imagerie diagnostique utilisant des radio-isotopes à vie courte comme le technétium 99m où une demi-vie biologique courte contribue à un meilleur rapport signal/bruit. Des radio-conjugués à demi-vie plasmatique courte utilisés en radiothérapie pourront également posséder une moindre radiotoxicité.

Les observations faites au cours de ces dernières années concernant les inconvénients des anticorps ont poussé de nombreuses équipes à explorer des protéines alternatives ayant la capacité de reconnaître les antigènes de façon spécifique et de s'y fixer de manière affine. Ceci a d'abord conduit au développement d'une gamme de différents fragments d'anticorps, principalement le fragment Fab (une chaîne légère et la moitié d'une chaîne lourde, d'une taille d'environ 50 kDa) et le fragment ScFv (deux domaines variables VH et VL, d'une taille d'environ 25 kDa), qui peuvent être simplement préparés en raccourcissant le cadre de lecture des gènes Ig clonés (Figure I.1). Les limitations relatives à la taille sont ainsi dépassées, et ces fragments d'anticorps peuvent être produits dans des hôtes microbiens. Ces fragments étant plus petits, ils peuvent pénétrer plus en profondeur dans les tumeurs ; de plus, les demi-vies peuvent être adaptées (Carter, 2006). Cependant, l'élimination de fragments d'anticorps qui ne sont pas directement impliqués dans la liaison à la cible peut affecter l'affinité de manière négative. De plus, les fragments d'anticorps ont tendance à s'agréger, ce qui les rend difficiles à purifier (Holliger et Hudson, 2005).

Parmi les alternatives aux fragments d'anticorps, on cite également les domaines uniques d'anticorps sdAb, appelés Nanobody et développés par la compagnie Ablynx (Hey et al., 2005). Un sdAb consiste en un domaine variable monomérique d'anticorps. De façon similaire à un anticorps entier, ce domaine est capable de se lier sélectivement à un antigène spécifique. Avec un poids moléculaire de 12 à 15 kDa seulement, les domaines uniques d'anticorps sont plus petits qu'un anticorps entier de 150 à 160 kDa, qu'un fragment Fab et qu'un fragment ScFv. Les premiers sdAb ont été construits à partir du domaine variable de la chaîne lourde d'anticorps isolés à partir de camélidés, et appelés fragments V_HH (Hamers-Casterman et al., 1993), ou isolés à partir de poissons cartilagineux et nommés fragments V_{NAR} (Greenberg et al., 1995) (Figure I.1). Même si la recherche est fondée en grande partie sur l'utilisation de domaines variables de chaînes lourdes, des nanobodies dérivés de domaines variables de chaînes légères ont montré des spécificités de liaison à des épitopes cibles. Ces domaines sont

plus résistants à la température et sont stables en présence de détergents à fortes concentrations en urée. Leur faible masse moléculaire permet une meilleure perméabilité dans les tissus, et une demi-vie dans le plasma plus courte grâce à une élimination rénale rapide. Les sdAb sont capables de se lier à des antigènes cachés qui ne sont pas accessibles à des anticorps entiers, comme les sites actifs d'enzymes, et ceci grâce à leur boucle CDR3 étendue (Lauwereys et al., 1998). Ces domaines ont révélé des affinités nanomolaires et des spécificités de liaison pour des haptènes et plusieurs protéines telles que les facteurs de transcription, les cytokines et les marqueurs tumoraux. Certaines molécules sdAb sont en cours de développement préclinique pour des maladies cardiovasculaires telle que la molécule ALX-0081 développée par la société Ablynx, ainsi que pour l'arthrite rhumatoïde et le psoriasis telle que la molécule Dom-0200/Art621 développée par la société Domantis (Wurch et al., 2008). Un domaine V_HH conjugué à la β-lactamase est également en cours d'essais précliniques pour le traitement de cancers (Cortez-Retamozo et al., 2004).

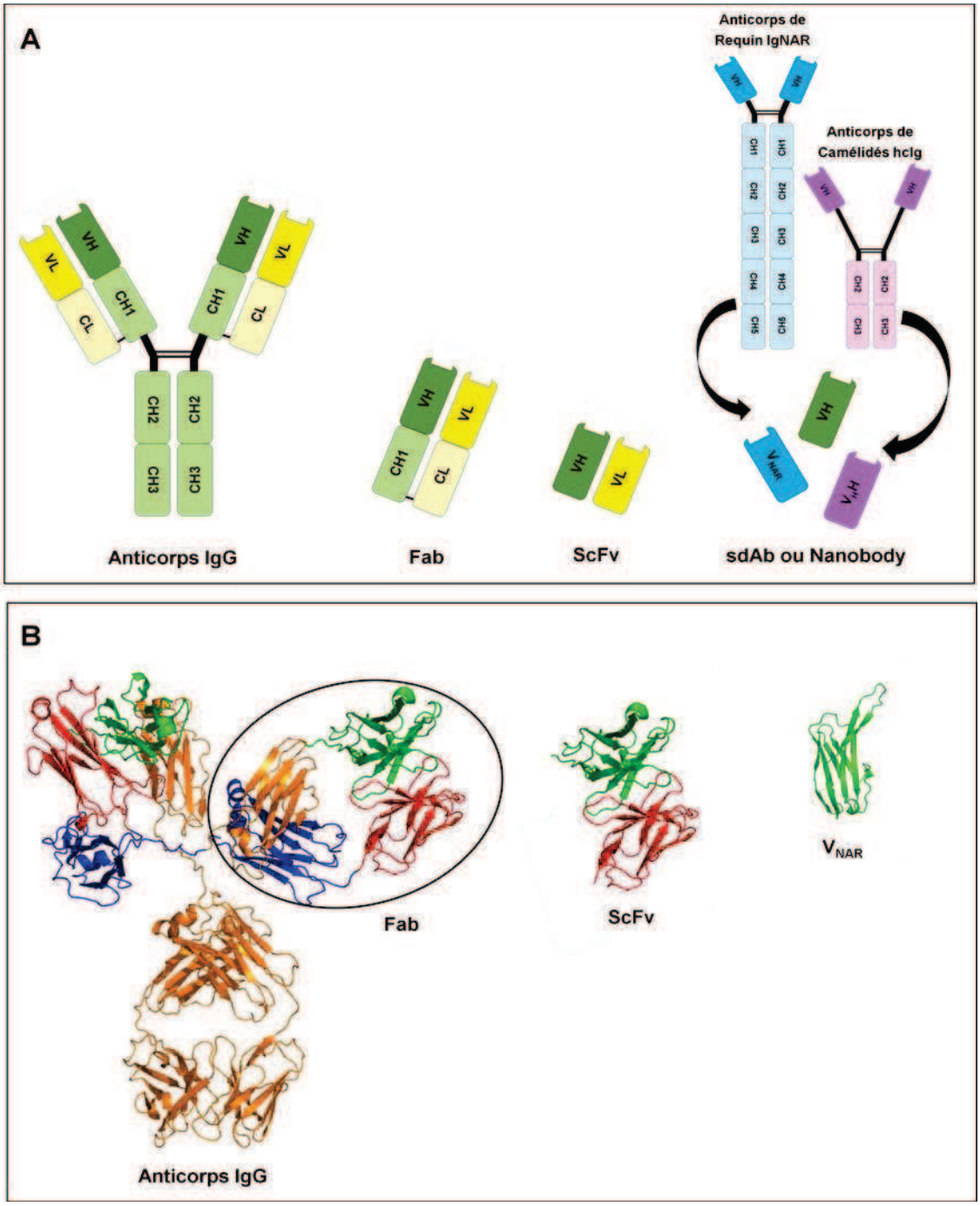


Figure I.1 : Anticorps IgG et fragments d'anticorps. A : Représentation schématique d'un anticorps IgG, d'un fragment Fab, d'un fragment ScFv et d'un domaine sdAb (VH issu d'un anticorps humain, V_HH issu d'un anticorps de camélidés ou V_{NAR} issu d'un anticorps de requins). B : Structure 3D d'un anticorps monoclonal IgG de souris (Code PDB 1IGY) montrant les fragments Fab et ScFv. Le domaine V_{NAR} représenté est issu d'une Ig de requin (Code PDB 1VES).

I.3 Alternatives aux anticorps : les plates-formes génériques

Les limitations des anticorps et des fragments d'anticorps ont grandement stimulé la recherche d'approches alternatives. Une première catégorie de méthodes aujourd'hui bien explorées est fondée sur le transfert d'un site de liaison universel comme celui rencontré dans les anticorps sur une architecture protéique alternative, désignée par le terme de plate-forme. Ce terme a acquis une signification unique dans le contexte de l'ingénierie des protéines. Il décrit un squelette polypeptidique possédant une grande tolérance aux modifications telles que les insertions, délétions ou substitutions multiples. La plate-forme acquiert certaines nouvelles propriétés tandis que son intégrité structurale globale et son comportement physicochimique originel restent conservés. Cette propriété adoptée *de novo* comprend généralement mais non exclusivement la spécificité de liaison pour une molécule cible prédéfinie. Les dix dernières années ont vu la découverte, le développement et l'utilisation de plusieurs de ces plates-formes. Ces plates-formes protéiques alternatives, non apparentées aux anticorps, forment un groupe très hétérogène de molécules ayant des origines, des tailles, des propriétés et des structures très différentes.

I.3.1 VLRs

Les récepteurs variables de lymphocytes (VLRs) sont d'un intérêt particulier du fait de leur utilisation par les lamproies (vertébrés sans mâchoires) à la place des IgG dans la réponse immunitaire adaptative (Caravella et Lugovskoy, 2010). Les VLRs représentent le seul système immunitaire adaptatif connu qui ne soit pas basé sur les immunoglobulines, cependant, leurs propriétés de liaison aux antigènes restent très peu connues.

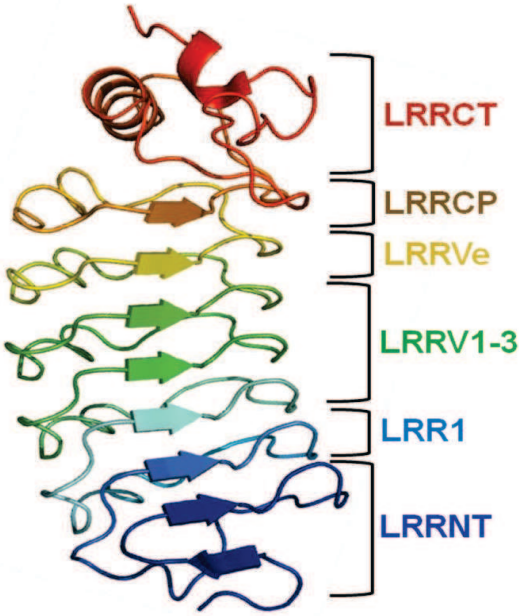


Figure I.2: Structure 3D d'un VLR (Code PDB 2O6S) représenté en mode cartoon en dégradé de couleurs de l'extrémité N-terminale (bleu) à l'extrémité C-terminale (rouge). Les différents modules constituant la molécule de VLR sont indiqués : LRRNT, LRR1, LRRV1 à 3, LRRVe, LRRCP et LRRCT.

Les VLRs sont très différents des immunoglobulines. Ce sont des monomères d'environ 270 acides aminés alors que les immunoglobulines sont des tétramères dont la totalité représente environ 1330 acides aminés. Comme les Ig, ces récepteurs possèdent une partie variable et une partie constante et sont fabriqués par des cellules qui ont une ressemblance morphologique avec les lymphocytes des vertébrés à mâchoire. Les VLRs sont constitués de segments riches en leucine (LRR) hautement diversifiés. Il existe deux types de gènes VLR : les VLRA et les VLRB. Pour former les récepteurs fonctionnels matures, les gènes VLR des lignées germinales subissent des recombinaisons d'ADN où chaque VLR est assemblé à partir de plusieurs cassettes codant pour des LRR. Chaque VLR mature est caractérisé par une région de diversité unique constituée de plusieurs modules LRR ayant des séquences hautement variables : un LRR N-terminal de 27 à 34 résidus (LRRNT), un LRR de 25 résidus (LRR1), de 1 à 9 modules LRR de 24 résidus (LRRVs), un LRRV final (LRRVe), un LRR tronqué de 16 résidus (CP pour « connecting peptide ») et un LRR C-terminal de 48 à 63 résidus (LRRCT). Les LRRNT et LRRCT sont stabilisés par 2 ponts disulfures permettant de protéger les deux extrémités de la région de diversité. L'assemblage des VLRs par insertions séquentielles de cassettes LRR avec des phénomènes fréquents de recombinaisons au niveau des extrémités des modules LRR génère un vaste répertoire de récepteurs estimé à plus de 10^{14} VLRs uniques, ce qui est comparable aux anticorps de mammifères. Ces VLRs représentent donc une excellente

alternative aux anticorps pour des applications biotechnologiques. Tasumi et al. ont récemment isolé des VLRs monoclonaux de lamproies et ont démontré qu'ils peuvent sélectionner des variants qui se lient à différentes cibles avec une grande affinité (Tasumi et al., 2009).

En 2012, les travaux de Lee et ses collaborateurs (Lee et al., 2012) ont porté sur la conception de nouvelles plates-formes basées sur des VLRs. Dans un 1er temps, un module LRRV consensus de 24 résidus a été conçu en analysant les résidus conservés des modules LRR des VLRs grâce à des alignements de séquences de 1000 modules LRR à partir de plusieurs banques de données. La séquence consensus est constituée de 10 résidus parfaitement conservés et 5 résidus variables notés x qui peuvent être modifiés afin de générer des ligands pour diverses cibles. Une plate-forme modèle VLRc-5 a été construite à partir de 5 modules LRRV consensus et des motifs N et C-terminaux des VLRs présents dans la nature. Cette plate-forme a été exprimée sous forme soluble chez *E. coli* avec un niveau d'expression de 2 mg/L. Afin d'augmenter le niveau d'expression, les domaines LRRV2-5, LRRVe, CP et LRRCT de la plate-forme modèle ont été fusionnés aux domaines LRRNT, LRR1 et LRR2 de l'internalin-B, dont la structure en hélice α favorise le repliement. La plate-forme ainsi conçue (Figure I.3) a été nommée Repebody et son niveau d'expression chez *E. coli* est passé à 60 mg/L.

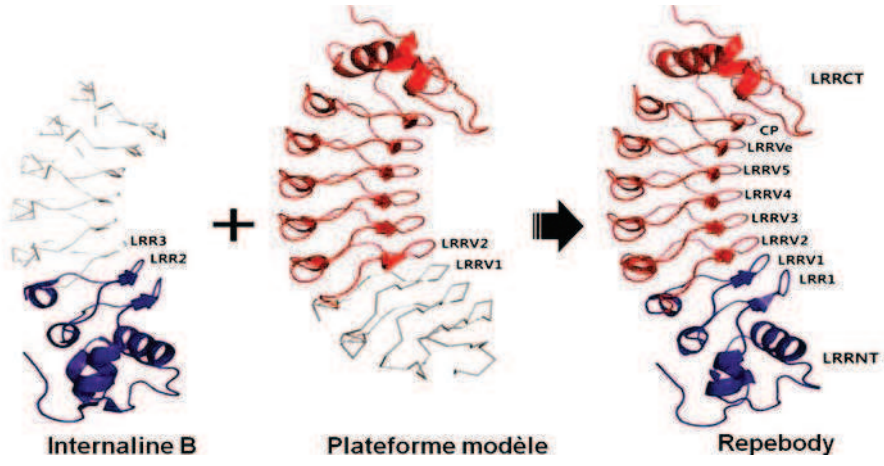


Figure I.3 : Conception de la molécule Repebody à partir de l'internaline B et de la plate-forme modèle VLRc-5 (Lee et al., 2012).

Une approche structurale a permis de concevoir un Repebody spécifique de la protéine de différentiation myéloïde-2 (MD2) (Lee et al., 2012). Un modèle de la plate-forme Repebody a

été superposé à la structure cristallographique de TV3 (protéine hybride comprenant les modules LRR du récepteur Toll-like 4 humain et un VLR de myxine) en complexe avec MD2, et les résidus de Repebody interagissant avec MD2 (11 résidus) ont été identifiés et mutés. La plate-forme MD2-Repebody ainsi conçue a été exprimée chez *E. coli* et a montré une affinité de liaison pour MD2 de 388 nM par analyse SPR. La même approche a été appliquée afin de concevoir un Repebody spécifique du lysozyme d'œuf de poule (HEL) en superposant la structure cristallographique de MD2-Repebody à celle du complexe VLR/HEL (Lee et al., 2012). Neuf résidus ont été identifiés sur MD2-Repebody et mutés, et les domaines CP et LRRCT de MD2-Repebody ont été remplacés par ceux de VLR car ceux-ci participent à l'interaction avec HEL. L'affinité de liaison de HEL-Repebody pour HEL a été estimée à 3,7 µM par analyse calorimétrique (ITC). Afin de démontrer l'applicabilité de Repebody à d'autres cibles protéiques, des Repebodies ont été générés pour l'interleukin-6 à partir d'une bibliothèque de variants de Repebody construite en faisant varier les résidus au niveau des sites hypervariables des modules LRRV. Trois clones montrant les signaux les plus élevés en activité de liaison dans un test ELISA phage ont été isolés, les trois Repebodies sélectionnés ont été exprimés chez *E. coli* et leurs affinités de liaison pour IL-6 déterminées par ITC sont de 48 nM, 89 nM et 117 nM. Ces résultats montrent que la plate-forme Repebody peut être largement utilisée pour générer des ligands ayant des affinités et des spécificités élevées pour une variété de cibles, grâce à une approche combinant la conception fondée sur la structure et la sélection par phage display.

I.3.2 Adnectins

Les Adnectins constituent une nouvelle famille de protéines thérapeutiques, dérivées du 10ème domaine de fibronectine humaine de type III (10Fn3), et conçues pour se lier à des cibles d'intérêt thérapeutique avec une grande affinité et spécificité. La fibronectine est une glycoprotéine extracellulaire ubiquitaire qui joue un rôle majeure dans l'adhérence cellulaire. Elle présente de nombreux sites de liaison pour des protéines de la matrice extracellulaire (comme le collagène, la thrombospondine), des récepteurs membranaires (tels que les intégrines), des protéines du sang circulant (comme la fibrine), des glycosaminoglycanes (comme l'héparine et le chondroïtine-sulfate). D'un point de vue structural, la fibronectine est un long polypeptide à chaîne unique comprenant des domaines de type I, II et III séparés par de

courtes séquences de liaison, avec une interaction limitée entre les domaines adjacents (Figure I.4).

La fibronectine humaine de type III (Fn3), et en particulier, le 10ème domaine de fibronectine humaine de type III (10Fn3) est actuellement l'une des plates-formes les plus largement utilisées. Ce domaine de petite taille (10 kDa) appartient à la classe de protéines plates-formes en sandwich β proche des immunoglobulines.

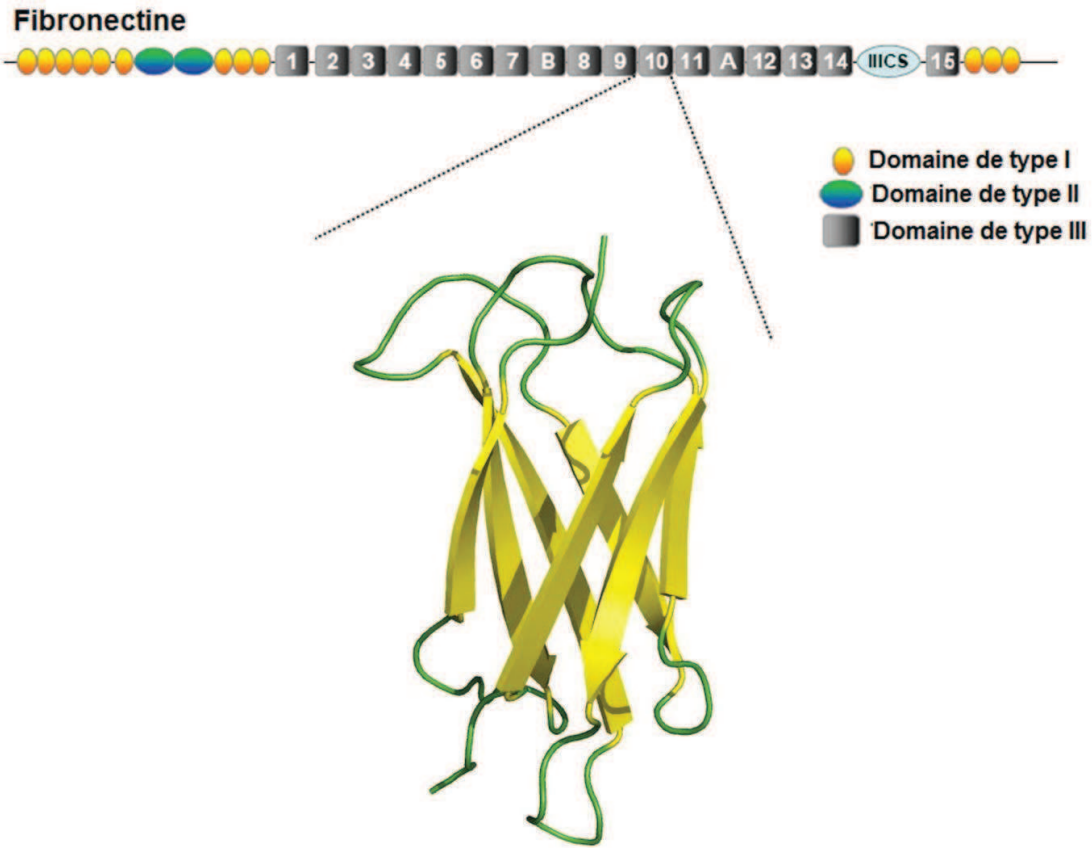


Figure I.4 : Structures des domaines de fibronectine. En haut : Organisation des différents domaines de fibronectine. En bas : Structure 3D du 10ème domaine de fibronectine humaine de type III ou 10Fn3 (Code PDB 1TTG). La protéine est représentée en mode cartoon et coloriée en fonction de la structure secondaire : les boucles en vert et les feuillets en jaune.

La structure en sandwich β du domaine 10Fn3 (Figure I.4) consiste en sept feuillets β antiparallèles avec trois boucles accessibles au solvant, reliant les feuillets à chaque extrémité du domaine (Dickinson et al., 1994). Malgré une absence d'homologie de séquence significative, le domaine 10Fn3 possède une structure similaire aux domaines variables des

anticorps (Main et al., 1992). Ce domaine 10Fn3 est caractérisé par une structure stable grâce à la présence d'un cœur hydrophobe et possède une grande thermostabilité en conditions réductrices (Plaxco et al., 1997). Contrairement aux anticorps dont les feuillets β des domaines variables sont liés par un pont disulfure, le domaine 10Fn3 ne contient pas de ponts disulfures ni de cystéines libres ; il ne nécessite pas de modifications post-traductionnelles et peut donc être produit dans les systèmes bactériens avec des rendements importants.

En vue de concevoir des adnectins à partir de domaines 10Fn3 modifiés, les boucles BC, DE et FG de 10Fn3, dont les structures sont analogues aux CDR H1, H2 et H3 des anticorps, ont été considérées comme candidats évidents pour la diversification afin de générer des surfaces artificielles de liaison aux cibles d'intérêt. Il en est de même pour les trois boucles situées à l'extrémité opposée de 10Fn3 (Bloom et Calabro, 2009). Initialement, des bibliothèques d'adnectins ont été réalisées par Koide et al. en introduisant entre 10 et 20 mutations au niveau de deux des trois boucles BC, DE et FG exposées en surface de la molécule. Le criblage de ces bibliothèques a été réalisé en utilisant la technique du « phage display », et a permis de sélectionner des protéines qui lient l'ubiquitine avec une affinité micromolaire (Koide et al., 1998). Dans une étude plus récente, des bibliothèques d'adnectins ont été réalisées par Xu et al. en faisant varier les trois boucles BC, DE et FG. Le criblage de ces bibliothèques a été réalisé en utilisant la technique de « mRNA display », et a permis de sélectionner des adnectins qui lient le TNF-α avec une affinité sub-nanomolaire après maturation de l'affinité (Xu et al., 2002). Le criblage d'une bibliothèque plus large mais conçue de la même façon par le laboratoire « AdNexus Therapeutics » a permis de sélectionner la 1ère molécule à application thérapeutique : une adnectin anti-angiogénique accrochée à une molécule de PEG, nommée CT-322, qui se lie au récepteur 2 du facteur de croissance endothélial vasculaire VEGF-R2 avec une grande affinité (Kd = 11 nM) et qui bloque son interaction avec VEGF A, C et D, inhibant la formation de nouveaux vaisseaux sanguins dans les tumeurs (Mamluk et al., 2010). Cette molécule est actuellement en phase II d'essais cliniques pour le traitement des glioblastomes multiformes, du cancer du poumon et du cancer colorectal métastatique (Dineen et al., 2008 ; Mamluk et al., 2010).

Des criblages de bibliothèques plus sophistiquées ont permis la construction d'adnectins à domaines multiples, chaque domaine contribuant à la spécificité de liaison pour des cibles différentes. En effet, un des avantages caractéristiques des molécules dérivées des domaines 10Fn3 modifiés est qu'elles se prêtent naturellement à la multimérisation permettant de générer des molécules de liaison multifonctionnelles. Ainsi, un enchaînement d'adnectins génétiquement liées, ayant différentes spécificités, imite l'arrangement naturel des domaines de fibronectine ; les domaines de liaison individuels des adnectins sont donc susceptibles d'avoir un repliement, une stabilité et une fonction indépendants. Un exemple de ce type d'adnectin consiste en une adnectin dont le domaine N-terminal se lie au récepteur du facteur de croissance épidermique avec une affinité de 0,7 nM et une adnectin dont le domaine C-terminal se lie au récepteur du facteur de croissance de type 1 analogue à l'insuline avec une affinité de 0,1 nM, conduisant à des effets antiprolifératifs à travers deux voies séparées (Emanuel et al, 2010). Emanuel et al. ont développé deux adnectins individuelles optimisées, l'une spécifique de EGFR et l'autre de IGFR, et ont construit à partir de ces molécules une protéine unique : une adnectin bispécifique ciblant à la fois EGFR et IGFR (EI-tandem), et ayant un effet antiprolifératif sur des lignées cellulaires cancéreuses humaines (Emanuel et al, 2010).

I.3.3 Avimers

Les avimers (pour avidity *multimers*) sont des protéines artificielles constituées de multimères de domaines A. Les domaines A existent sous forme d'une succession de multiples domaines dans divers récepteurs membranaires. Ces domaines se lient à une centaine de cibles différentes connues, telles que des petites molécules, des protéines et des virus.

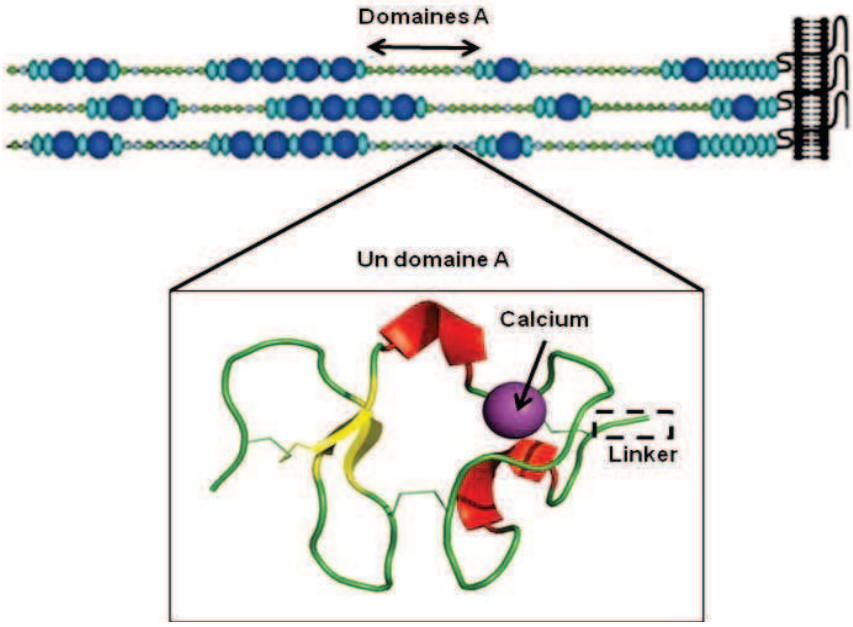


Figure I.5 : Schéma représentant les domaines A. De haut en bas : Organisation des domaines A dans des récepteurs humains ; vue agrandie de la structure 3D d'un domaine A (Code PDB 1AJJ). La protéine est représentée en mode cartoon et coloriée en fonction de la structure secondaire : les boucles en vert, les hélices en rouge et les feuillets en jaune. L'atome de calcium est représenté en sphère mauve. Un linker permet de relier deux domaines A entre eux.

Chaque domaine A comprend une séquence de 30 à 35 acides aminés (environ 4 kDa) ; cette structure est rigide, stabilisée par 3 ponts disulfures et par la présence de 4 résidus liant un ion calcium. Ainsi, un motif conservé de 12 acides aminés seulement est requis pour cette structure commune. Les 28 autres positions non conservées peuvent être comparées aux 25 à 30 positions hypervariables contenues dans les six régions hypervariables d'un anticorps déterminant la complémentarité avec l'antigène (CDR).

Les domaines A sont connectés entre-eux par des peptides-linker d'environ 5 acides aminés. Chaque domaine A peut se lier à un épitope de la protéine cible. La combinaison de domaines se liant à différents épitopes de la même protéine permet d'augmenter l'affinité pour cette protéine, un effet connu sous le terme avidité. Par ailleurs, ces domaines peuvent être dirigés contre des épitopes sur des protéines cibles différentes. Cette approche est similaire à celle utilisée dans le développement d'anticorps monoclonaux bispécifiques. Silverman et al. ont montré que le fait de combiner plusieurs domaines A différents permet de générer des avimers ayant des affinités picomolaires et même sub-picomolaires (Silverman et al., 2005).

Malgré la présence de 6 cystéines dans le domaine A formant 3 ponts disulfures, des avimers

contenant plus de huit domaines A avec un total de 48 cystéines peuvent être produits dans le cytoplasme d'*E. coli* puis correctement repliés, donnant des rendements > 1 g/L. Ces avimers montrent une meilleure stabilité thermique comparés aux anticorps, mais une demi-vie dans le plasma limitée à cause de leur taille plus petite. La demi-vie plasmatique peut être améliorée en les liant à des anticorps. Ainsi, Silverman et al. montrent que la fusion d'un domaine IgG à l'extrémité N-terminale d'un avimer permet d'augmenter la demi-vie de celui-ci dans le plasma (Silverman et al., 2005).

Cinq différents avimers ayant entre 2 et 3 domaines ont été sélectionnés permettant d'inhiber des cibles d'intérêt : IL-6, cMet, CD28, CD40L, BAFF, avec des affinités picomolaires. Par ailleurs, un avimer inhibant IL-6 avec un IC50 de 0,8 pM est biologiquement actif dans deux modèles animaux (Silverman et al., 2005). Un avimer trimérique AMG-220 (c326) ciblant la molécule IL6 et se liant à celle-ci avec une affinité de 25 nM (Silverman et al., 2005) est entré en 2006 en phase I d'évaluation clinique pour le traitement de la maladie de Crohn, mais actuellement aucune information concernant cette étude clinique n'est disponible. Notons que des anticorps inhibant cMet et CD28 ont été difficiles à obtenir, résultant du fait que la nature bivalente d'un anticorps entier conduit à la dimérisation du récepteur et à des effets agonistes indésirables.

I.3.4 Affibody

Les molécules affibody sont à l'origine dérivées du domaine B contenu dans la région de liaison à l'immunoglobuline de la protéine A de Staphylocoque, qui est constituée de cinq domaines homologues de liaison à l'Ig (E, D, A, B et C) capables de se lier au domaine Fc des anticorps de différentes espèces (Moks et al., 1986). Le domaine B est un peptide relativement petit, de 58 acides aminés, sans résidus cystéines. Ce domaine a une structure en 3 hélices α, et une dizaine d'acides aminés situés sur deux hélices α à l'interface du domaine sont impliqués dans l'activité de liaison de la protéine d'origine.

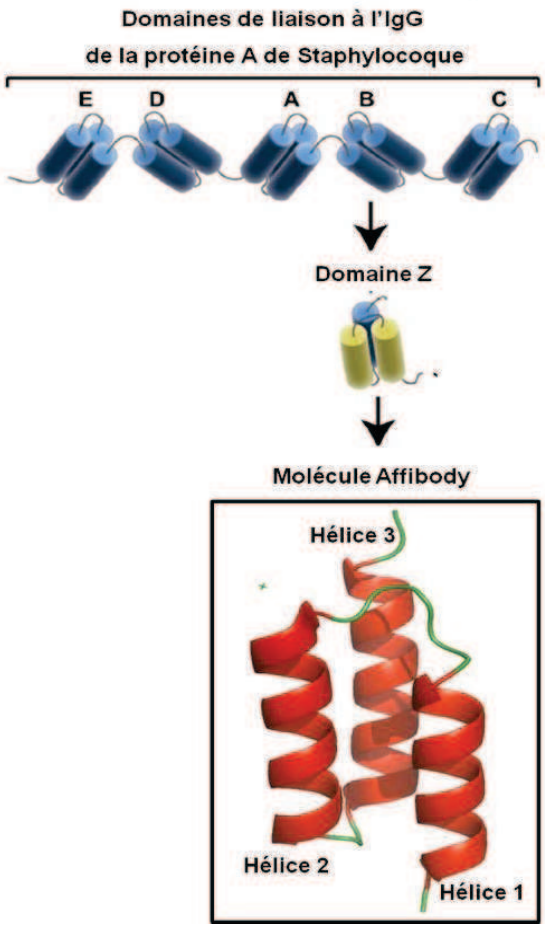


Figure I.6 : Schéma représentant de haut en bas : le domaine B de liaison à l'IgG de la protéine A ; le domaine Z ; une molécule affibody. La molécule affibody (Code PDB 1LP1) est représentée en mode cartoon et coloriée selon la structure secondaire : les boucles en vert et les hélices en rouge.

La cinétique de repliement de ce domaine est l'une des plus rapides ayant été reportée. Les caractéristiques intéressantes telles que la petite taille, la cinétique rapide de repliement, la structure simple mais robuste et ne contenant pas de cystéines, la grande solubilité et une stabilité thermique relativement élevée ont fait de ce domaine une plate-forme de grand intérêt. Le domaine B a par la suite été muté au niveau de positions clé afin d'en améliorer la stabilité chimique, et le variant résultant a été nommé domaine Z. En effet, la séquence Asn28-Gly29 a été remplacée par la séquence Asn28-Ala29 afin de conférer au domaine Z résultant une résistance aux traitements par l'hydroxylamine et le bromure de cyanogène (Nilsson et al., 1987). Ce domaine Z modifié a gardé son affinité pour la région Fc de l'anticorps alors que la très faible affinité pour la région Fab a été complètement perdue. L'exploitation du domaine Z a conduit à la construction de molécules appelées affibody avec une capacité de liaison à n'importe quelle cible donnée.

Contrairement aux anticorps monoclonaux qui doivent être générés par immunisation d'animaux de laboratoire combinée aux techniques d'hybridomes, la sélection de nouvelles protéines d'affinité basées sur des plates-formes non-immunoglobulines est réalisée à l'aide de bibliothèques combinatoires synthétiques et des systèmes de sélection *in vitro* (comme le « phage display »). Les bibliothèques d'affibody sont généralement construites par mutations aléatoires de 13 acides aminés au niveau des hélices 1 et 2 comprenant la surface de liaison au Fc du domaine Z. Grâce à l'utilisation de la technique du « phage display » suivie de criblage contre les cibles désirées, des molécules affibody montrant des activités de liaison spécifiques à une grande variété de protéines différentes (insuline, fibrinogène, transferrine, TNF-α, IL-8, gp120, CD28, albumine sérique humaine, IgA, IgE, IgM, HER2 et EGFR) ont été générées, ayant des affinités comprises entre μM et pM (Nygren, 2008).

Les molécules affibody peuvent être exprimées sous forme soluble et stable à la protéolyse dans diverses cellules hôtes, soit seules ou en fusion avec d'autres partenaires. Elles tolèrent les modifications et montrent une grande stabilité dans des conditions de température élevée (90°C), des conditions acides (pH 2,5) ou basiques (pH 11). La masse moléculaire de cette molécule est de 6 kDa environ, alors qu'un anticorps monoclonal a une masse de 150 kDa et qu'un domaine unique d'anticorps qui représente le plus petit type de fragment d'anticorps liant un antigène a une masse de 12-15 kDa. Grâce à leur petite taille et leur propriété de repliement rapide, les molécules d'affibody peuvent également être produites par synthèse peptidique ce qui rend possible l'incorporation de groupements tels que des sondes fluorescentes ou des groupements chélatants pour la liaison à un atome de métal radioactif. Par ailleurs, la petite taille de ces plates-formes rend possible la combinaison à des protéines de fusion avec différentes propriétés, tout en gardant une taille globale toujours plus petite qu'un anticorps entier. Des fusions de molécules affibody de même spécificité montrent des effets d'avidité lors de la liaison à la cible, et des fusions de molécules affibody de spécificités différentes permettent d'obtenir des protéines d'affinité bi ou multi-spécifiques.

Les molécules affibody ont été exploitées dans divers domaines : diagnostique, thérapeutique, imagerie et biotechnologie (Löfblom et al., 2010). Elles ont été testées à des fins diagnostiques chez des patients ayant un cancer avec des métastases exprimant HER2. L'affibody ABY-025 spécifique de HER2/neu est en cours de développement clinique pour le diagnostic des

tumeurs. Des études récentes indiquent que les molécules d'affibody sont parmi les traceurs les plus prometteurs pour l'imagerie moléculaire spécifique de HER2, comparées aux anticorps et aux fragments d'anticorps (Tolmachev, 2008). En biotechnologie, l'application la plus largement exploitée concerne la bioséparation par affinité. En effet, des molécules affibody avec de nouvelles spécificités ont été évaluées en tant que ligands sur des colonnes de chromatographie par affinité pour la purification de protéines, ciblant une grande variété de protéines différentes telles que l'apolipoprotéine A-1M, la Taq polymérase, la protéine G du virus respiratoire syncytial humain et le facteur VIII recombinant humain.

I.3.5 Darpins

Les DARPins (Designed Ankyrin Repeat Proteins) sont une autre classe de protéines d'origine humaine ayant été utilisées comme molécules affines et thérapeutiques, appartenant à la classe des protéines à répétitions. Les DARPins sont de petites protéines dérivées d'ankyrines naturelles, une classe de protéines qui intervient dans les interactions protéine-protéine de haute affinité. Les répétitions d'ankyrine sont constituées de plusieurs unités répétitives en tandem, de 33 acides aminés chacune.

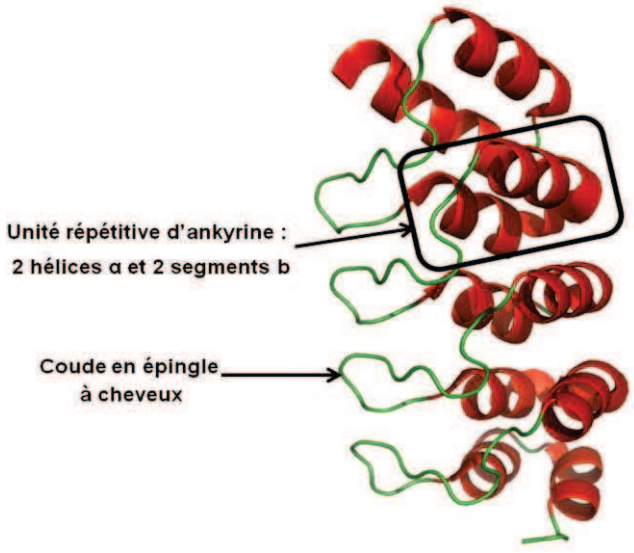


Figure I.7 : Structure 3D d'un DARPin (Code PDB 1SVX). La protéine est représentée en mode cartoon et coloriée en fonction de la structure secondaire : les boucles en vert et les hélices en rouge.

Chaque unité répétitive d'ankyrine est constituée d'un segment β, de deux hélices α et d'un

deuxième segment β, pour former une super-hélice à un tour, de pas à droite. Les unités ankyrine sont reliées par un coude en épingle à cheveux qui unit le segment β C-terminal d'une unité au segment β N-terminal de l'unité suivante. Ces motifs ankyrine sont présents dans un grand nombre de protéines, essentiellement d'origine eucaryote, ayant des fonctions diverses. La plupart des protéines contenant ce motif ont 4 à 6 répétitions. Les DARPins sont généralement constitués de quatre ou cinq unités répétitives d'ankyrine, formant un domaine protéique stable avec une grande surface d'interaction potentielle avec les cibles.

Les DARPins possèdent plusieurs propriétés intéressantes telles que la stabilité thermodynamique grâce à un réseau de ponts salins au niveau de leur surface. De plus, leur stabilité augmente avec le nombre de répétitions. Comme les domaines fibronectine, les DARPins ne possèdent pas de cystéines ni d'oligosaccharides, et peuvent être optimisés par des techniques d'évolution *in vitro*. Les DARPins sont exprimés dans le cytoplasme d'*E. coli* sous forme soluble avec des rendements élevés. Par ailleurs, leur petite taille (entre 14 et 18 kDa) peut les rendre plus adaptés pour pénétrer dans les tumeurs, par rapport à un anticorps conventionnel (Zahnd et al., 2010). Zahnd et al. montrent qu'avec des DARPins ayant une masse de 14,5 kDa et des affinités entre 270 nM et 90 pM pour la cible HER2, l'accumulation dans les tumeurs est importante, alors qu'avec une construction contenant une fusion de deux DARPins (30 kDa), l'accumulation dans les tumeurs est significativement plus faible, similaire à celle observée avec des ScFv (Zahnd et al., 2010).

Les DARPins ont été utilisés comme outils de recherche, agents diagnostiques et thérapeutiques. Le 1er candidat à application thérapeutique, MP0112, est un inhibiteur de VEGF et a été soumis au début de l'année 2010 aux tests cliniques pour le traitement de la dégénérescence maculaire humide et de l'œdème maculaire diabétique. Par ailleurs, un DARPin de grande affinité pour le récepteur du facteur de croissance épidermique humain 2 HER2 (90 pM) a été obtenu en utilisant la technique de présentation à la surface des ribosomes ; Zahnd et al ont démontré que ce DARPin est très sensible pour la détection de HER2 dans des extraits carcinomiques humains (Zahnd et al., 2007).

I.3.6 Anticalins

Les anticalins constituent une classe de protéines modifiées dérivées de lipocalines, une famille de protéines abondantes chez les hommes, les insectes, les bactéries et beaucoup d'autres organismes. Chez l'homme, il existe environ dix lipocalines différentes, la protéine plasmatique de liaison au rétinol étant la plus connue. Les lipocalines sont présentes dans le plasma et autres fluides corporels, et ont pour rôle le transport de vitamines, de stéroïdes ou de composés métaboliques (Flower, 1995).

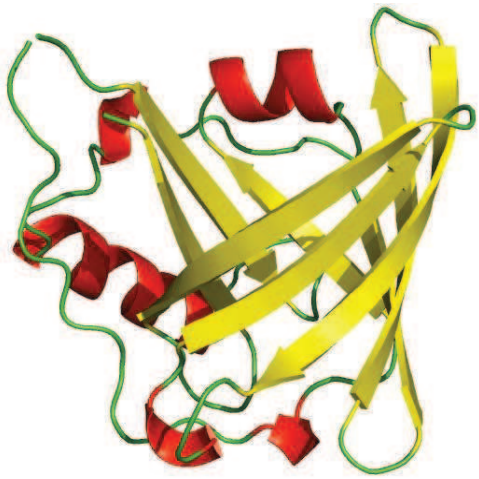


Figure I.8 : Structure 3D d'une molécule d'anticalin (Code PDB 3BX7). La protéine est représentée en mode cartoon et coloriée selon la structure secondaire : boucles en vert, hélices en rouge et feuillets en jaune.

Les lipocalines se présentent toutes sous forme d'un tonneau β ouvert de structure très conservée, constitué de huit feuillets β antiparallèles compactes et rigides qui s'enroulent autour d'un axe central. Au niveau de son extrémité ouverte, cette structure cylindrique supporte quatre boucles hypervariables, qui forment une entrée vers la poche pour la liaison à différentes molécules cibles. L'extrémité opposée du tonneau est fermée par de courtes boucles, et la forte densité des chaînes latérales d'acides aminés forme une région hydrophobe. Malgré une très faible homologie de séquence mutuelle, la structure en tonneau β est bien conservée parmi les 10 à 12 différentes lipocalines humaines. En revanche, la région en boucles autour de la poche de liaison aux cibles montre de grandes différences, en termes de séquence primaire, de longueur et de conformation des quatre segments polypeptidiques, ce qui est en accord avec les diverses spécificités de ligands observées pour cette famille.

Ainsi, de nouvelles protéines ayant des propriétés similaires aux anticorps, les anticalins, ont été générées à partir de lipocalines grâce à la grande plasticité structurale des sites de liaison de ces dernières. Comparées aux anticorps, les anticalins offrent plusieurs avantages grâce à leur petite taille, typiquement entre 160 et 180 résidus, une grande interface potentielle avec leurs cibles, une structure tertiaire robuste et une composition en chaîne polypeptidique unique, ainsi qu'une meilleure pénétration dans les tissus. Elles sont caractérisées par une grande thermostabilité qui rend ces protéines robustes (Schlehuber et Skerra, 2002), elles ne nécessitent pas de glycosylation ni la formation de ponts disulfures et peuvent être facilement produites dans des systèmes d'expression bactériens (Müller et Skerra, 1993). Etant dérivées de lipocalines humaines, un reformatage supplémentaire tel que le greffage de CDR lors de l'humanisation des anticorps n'est pas requis. Par ailleurs, leur propriété de liaison monovalente diminue le risque de liaisons croisées intermoléculaires non désirées.

Une des applications thérapeutiques de l'anticalin est l'utilisation en tant qu'antidote permettant d'éliminer rapidement un composé toxique ou irritant du corps humain. Un variant de l'anticalin dirigé contre la digoxygénine, ayant une affinité améliorée et une constante de dissociation sub-nanomolaire, nommé « Digical », permet d'inverser complètement une intoxication causée par digitalis dans un modèle de cochon d'inde, démontrant ainsi l'efficacité *in vivo* de cette molécule comme antidote (Schlehuber et Skerra, 2005).

La plupart des structures moléculaires actuellement considérées comme cibles d'intérêt dans le domaine biopharmaceutique concerne les protéines extracellulaires ou les récepteurs de surface cellulaire. Une deuxième application thérapeutique des anticalins concerne leur rôle d'antagonistes : ces molécules se lient à des récepteurs cellulaires, bloquant ainsi l'interaction de ces derniers avec leurs molécules de signalisation naturelles. Des bibliothèques d'anticalins construites sur la base de lipocalines humaines (afin de réduire les effets immunogéniques) ont donc été spécifiquement développées pour la reconnaissance d'antigènes protéiques, grâce à des mutations aléatoires au niveau de résidus situés à des positions exposées dans les boucles hypervariables (Skerra et Schlehuber, 2003). A partir de ces bibliothèques, des anticalins ayant de grandes sélectivités et des affinités sub-nanomolaires ont été sélectionnées contre des antigènes protéiques tels que des récepteurs immunologiques comme le domaine extracellulaire de l'antigène No 4 du lymphocyte T cytotoxique CTLA-4 ou des facteurs de croissance comme le facteur de croissance endothélial vasculaire VEGF (Schlehuber et Skerra, 2005). Une

molécule d'anticalin, nommée Prs-050 (Angiocal), ciblant le facteur VEGF-A, est entrée depuis Juin 2010 en phase I d'évaluation clinique dans diverses tumeurs solides.

Les lipocalines naturelles ou génétiquement modifiées sont rapidement éliminées par filtration rénale à cause de leur petite taille (20 kDa), si elles circulent sous forme de protéines monomériques. Dans la perspective d'applications médicales nécessitant un traitement prolongé, la demi-vie des anticalins dans le plasma doit être augmentée, par combinaison à des protéines de fusion comme l'albumine sérique ou par PEGylation (Hohlbaum et Skerra, 2007). Ces techniques de fusion sont réalisables sans altérer l'activité de liaison car les extrémités N et C-terminales des anticalins sont accessibles et éloignées du site de liaison, à la différence des fragments scFv des anticorps dont l'extrémité N-terminale fait partie du paratope.

I.3.7 Heat Repeat

Les répétitions HEAT (abréviation pour Huntingtin – Elongation - A subunit – TOR) sont constituées de la répétition d'un motif structural formant deux hélices α antiparallèles et deux coudes arrangés autour d'un axe commun (Figure I.9) et dont la longueur varie entre 37 et 47 acides aminés.

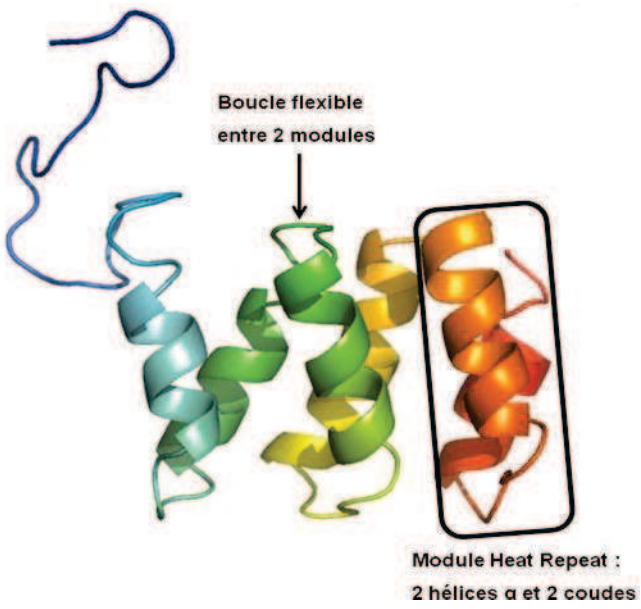


Figure I.9 : Structure 3D d'une protéine de type Heat Repeat, MTH187 (Code PDB 1TE4), constituée de 3 unités répétitives HEAT. La protéine est représentée en mode cartoon en dégradé de couleurs de l'extrémité N-terminale (bleu) à l'extrémité C-terminale.

Le motif est notamment caractérisé par la présence de deux résidus conservés Asp et Arg en positions 19 et 25 respectivement, ainsi que de la présence d'un ensemble de résidus hydrophobes contribuant à la compacité entre les deux hélices d'une même unité. Cependant, les séquences des unités répétitives sont hautement divergentes au sein des protéines "Heat repeat". Les répétitions HEAT existent sous forme d'unités consécutives dont le nombre varie de 3 à 36 et qui sont reliées entre-elles par des boucles flexibles. Elles sont présentes dans un certain nombre de protéines cytoplasmiques, dont les quatre protéines qui ont donné leur nom à cette famille : Huntingtin, facteur d'élongation 3 (EF3), la sous-unité alpha (65 Kd) de la protéine phosphatase 2A (PP2A), et la kinase PI3 de levure TOR1. La famille des protéines "Heat repeat" est divisée en trois sous-groupes sur la base de leurs similarités : IMB, AAA et ADB. Les membres les plus représentatifs de cette famille sont PP65A et Importin β1. Ces protéines "Heat repeat" ont une fonction commune de reconnaissance protéine-protéine dans plusieurs mécanismes cellulaires, en particulier dans les processus de transport cytoplasmique.
Bien que très courantes, ces protéines n'ont pas été utilisées comme plates-formes pour la conception de protéines artificielles du fait de leur très grande variabilité. Récemment, l'équipe de Philippe Minard (Urvoas et al., 2010) a reporté l'utilisation de cette famille de protéines dans la conception d'une séquence consensus pour une nouvelle famille de protéines HEAT repeat

artificielles nommée αRep (alpha-helicoidal repeat proteins). Une bibliothèque de protéines a été construite à partir d'une structure connue de la protéine Mth187 appartenant à une classe spécifique de protéines HEAT repeat thermostables et d'une séquence consensus déterminée par alignement de séquences de protéines HEAT repeat très proches de la protéine Mth187. Cette bibliothèque a été conçue de telle manière que les protéines possèdent un nombre variable d'unités répétitives et des séquences variables au sein de chaque unité répétitive. La variabilité a été introduite au niveau des positions potentielles de liaison aux cibles moléculaires (18, 19, 22, 23 et 26) qui tendent à se regrouper au niveau de la face extérieure de la deuxième hélice du motif. Par ailleurs, la grande surface d'interaction de ces protéines au niveau des hélices du motif permet d'accueillir une grande diversité de chaînes latérales.

I.4 L'ingénierie *in silico* de protéines fondée sur la structure

Les méthodes décrites précédemment possèdent de nombreux atouts et ont déjà démontré leur important potentiel dans la découverte de ligands protéiques. Cependant ces méthodes ne permettent pas d'offrir un contrôle précis sur la position où les ligands protéiques sélectionnés interagissent ni sur le mode de liaison du ligand à la cible. Ceci peut constituer un inconvénient dans le cadre du développement de molécules d'intérêt thérapeutique pour lesquelles la spécificité d'interaction est souvent requise. Au contraire, les méthodes fondées sur la reproduction d'interactions existantes dans un complexe de la cible considérée avec un ligand naturel de celle-ci pourraient permettre un contrôle sur la région où le ligand protéique artificiel interagit à la surface de la cible. Des méthodes d'ingénierie moléculaire exploitant cette idée se sont développées et ont conduit à un certain nombre de succès. Elles sont fondées sur l'exploitation des informations structurales tant pour l'identification des motifs fonctionnels que pour l'identification de structures protéiques servant de plates-formes pour reproduire des interactions importantes dans la liaison des ligands protéiques artificiels à la cible. Elles possèdent généralement un caractère mimétique puisqu'elles visent à reproduire des interactions établies par un ligand naturel de référence ou, dans certains cas, des interactions identifiées par des méthodes de simulations. Les méthodes *in silico* fondées sur l'exploitation des connaissances structurales des protéines ont d'abord été utilisées comme outils afin d'améliorer notre compréhension des propriétés intrinsèques des interactions protéines-

protéines, et de leur spécificité d'interaction et la modulation des interactions naturelles entre partenaires protéiques. Par la suite, les applications de ces méthodes furent étendues à la conception *de novo* d'interactants protéiques souvent fondée sur le transfert de nouveaux sites de liaison sur des architectures protéiques existantes. C'est dans ce cadre que se déroulent les travaux présentés dans ce rapport.

I.4.1 Amélioration de la spécificité des ligands protéiques

Des travaux importants ont été réalisés via l'exploitation de structures 3D et de méthodes de modélisation moléculaire visant l'amélioration de la spécificité de ligands protéiques envers leurs protéines-cibles. Par exemple, Shifman et Mayo ont utilisé ces méthodes pour améliorer la spécificité d'un complexe de la calmoduline avec son peptide cible (Shifman et Mayo, 2002). Alors que la calmoduline native reconnait une variété de différents peptides cibles avec une affinité raisonnable, un variant de calmoduline conçu avec huit mutations montre une affinité pour un peptide cible 86 fois supérieure à l'affinité mesurée envers d'autres peptides partenaires. Les auteurs ont ainsi montré que la conception d'interfaces de protéines par méthodes informatiques a permis d'atteindre des affinités nanomolaires similaires à celles de la calmoduline native, tout en améliorant la spécificité de liaison par optimisation de l'interface de liaison pour un seul peptide cible.

La conception de la spécificité a connu une avancée supplémentaire grâce aux travaux de Reina et al. décrivant la conception de mutants du site de liaison du domaine PDZ (acronyme des premières lettres de 3 protéines à l'origine de la découverte de ce domaine : Post synaptic density protein (PSD95), Drosophila disc large tumor suppressor (Dlg1) et Zonula occludens-1 protein (zo-1)) conduisant à la liaison à de nouvelles séquences peptidiques cibles (Reina et al., 2002). Trois différents mutants de domaine PDZ contenant 6 ou 7 mutations ont été construits à l'aide de méthodes informatiques par modification des interfaces existantes afin de se lier à 3 peptides cibles : une séquence peptidique naturelle présente au niveau de l'extrémité C-terminale d'une molécule similaire à la kinesin, une séquence peptidique chargée et une autre hydrophobe. Les 3 mutants se lient aux peptides cibles avec des affinités plus de deux fois plus importantes que celles du domaine PDZ non modifié.

Des études plus récentes sur la spécificité protéine-protéine ou protéine-peptide ont révélé

l'intérêt de la conception négative, procédé selon lequel des propriétés indésirables sont prises en considération (Bolon et al., 2005). Cette méthode permet d'introduire les mutations qui stabilisent la structure ciblée tout en déstabilisant les structures alternatives. Cette méthode fut appliquée avec succès pour la conception de dimères en torsade d'hélices ou « coiled-coil ». Les travaux de Havranek et Harbury sont réalisés sur le facteur de transcription GCN4, une protéine contenant deux hélices α allongées formant un homodimère par le biais d'interactions hydrophobes (Havranek et Harbury, 2003). Quatre états sont considérés pour cette structure « coiled coil » : 1) l'homodimère ; 2) l'hétérodimère ; 3) l'état non replié ; et 4) l'état agrégé. Des mutations ont été effectuées au niveau de positions bien déterminées des hélices, permettant de favoriser la formation du dimère cible sur les autres états possibles. L'utilité de la conception négative a également été démontrée par les travaux de Joachimiak et al. sur la conception d'une nouvelle interface DNase colicin E7-protéine de l'immunité Im7 (Joachimiak et al., 2006).

I.4.2 Modulation des interactions protéines-protéines

Un exemple de modulation de l'interaction protéine-protéine ayant une application biologique intéressante est celui de la conception d'un variant du facteur de nécrose tumorale (TNF) qui est actuellement développé comme agent thérapeutique pour le traitement de l'arthrite rhumatoïde. La conception du variant de TNF est réalisée en introduisant des mutations à l'interface entre le TNF et son récepteur de façon à empêcher leur interaction (Steed et al., 2003). Les variants de TNF ainsi conçus forment des complexes hétéro-trimériques avec le TNF natif qui sont incapables de se lier aux récepteurs de TNF.

I.4.3 Modification de sites de liaison protéiques

Un certain nombre de travaux de conception de nouvelles fonctions biologiques sont fondés sur l'utilisation de protéines de liaison périplasmiques (PBPs), connues pour être des plates-formes intéressantes du fait de leur capacité à se lier à une très grande variété de petites molécules, telles que les carbohydrates, les acides aminés, les neurotransmetteurs et les oligopeptides (Tam et Saier, 1993). De plus, la liaison au ligand est facilement détectée grâce au changement conformationnel se produisant lors de la liaison. Les régions de liaison de PBPs

ont été modifiées par méthodes de conception informatiques afin de lier des métaux, des neurotransmetteurs, des sucres, des explosifs et des agents neurotoxiques. Par exemple, Marvin et Hellinga ont réalisé la conversion d'un récepteur au maltose (MBP) en un biosenseur capable de lier le zinc (Marvin et Hellinga, 2001). Le récepteur au maltose consiste en une chaîne polypeptidique unique qui se replie pour former deux domaines connectés par une région charnière. Les sites de liaison au maltose ont été identifiés au niveau de l'interface entre les deux domaines. Au niveau de ces sites, les résidus responsables de la liaison au maltose ont été modifiés par mutagénèse dirigée en résidus histidine, permettant ainsi au niveau de chaque site la liaison à un atome de zinc grâce à un réseau tétraédrique constitué de trois résidus histidine et d'une molécule d'eau. Dans ce nouveau biosenseur, la liaison au zinc est contrôlée par un mécanisme de couplage conformationnel associant la liaison au zinc dans la protéine à l'état replié à une émission d'un signal de fluorescence.

I.4.4 Transfert de motif fonctionnel sur une protéine hôte

La conception de mimes fonctionnels de petite taille des larges protéines a longtemps été un défi important. La méthode est basée sur le greffage de sites fonctionnels sur des plates-formes naturelles. Une nouvelle fonction est ainsi conçue par le transfert d'un site actif sur une région structurale précise, permissive envers les mutations et structuralement compatible avec la nouvelle fonction. Cette approche nécessite donc l'étude et l'analyse des interactions dans les complexes naturels protéine-protéine ainsi que la connaissance des principes moléculaires qui les régissent, afin d'identifier les résidus qui interviennent dans la liaison et la stabilisation de ces complexes protéiques.

I.4.4.1 Identification des résidus du motif fonctionnel

Il a été postulé que dans les interfaces protéine-protéine, un nombre réduit de résidus contribuent de manière prépondérante à la variation d'énergie libre de liaison. Pour une interface typique de 1200 à 2000 $Å^2$, moins de 5% des résidus de l'interface contribuent à la liaison avec plus de 2 kcal/mol (Ofran et Rost, 2007). L'examen de la structure 3-D détaillée de complexes protéine-protéine montre que certains résidus sont susceptibles de jouer un rôle prépondérant dans la liaison grâce aux interactions intermoléculaires mises en jeu avec le

partenaire cible (liaisons hydrogène, ponts salins etc..). Cependant, l'analyse structurale seule ne permet pas de déterminer la contribution à la variation d'énergie libre de liaison de chacune des interactions intermoléculaires établies avec le partenaire cible. La mise en évidence des résidus jouant un rôle important dans la liaison peut être réalisée grâce à leur substitution expérimentale en résidus alanine (méthode désignée par le terme « alanine scanning ») et à la mesure de l'effet de ces substitutions sur l'énergie libre d'association des deux partenaires. Plusieurs expériences ont démontré que la plupart des résidus de l'interface peuvent être mutés sans affecter l'affinité de la protéine pour ses partenaires. Les quelques résidus qui après mutation changent l'affinité sont considérés comme prédominant pour l'interaction et sont désignés par l'expression « hot spots ». Ainsi, malgré la taille importante des interfaces de liaison, quelques résidus seulement contribueraient de manière prépondérante à la variation d'énergie libre de liaison.

Les travaux pionniers relatifs à l'identification des « hot spots » de l'interaction protéine-protéine ont été réalisés sur le complexe entre l'hormone de croissance humaine et son récepteur par Clackson et Wells (Clackson et Wells, 1995). La structure cristallographique du complexe de l'hormone de croissance humaine hGH avec le domaine extracellulaire de son récepteur hGHbp montre que sur chacune des protéines, une surface de 1300 $Å^2$ environ est enfouie consécutivement à la liaison des deux partenaires (valeur déterminée par le calcul de la variation d'accessibilité au solvant des deux partenaires entre l'état associé et l'état dissocié), ce qui correspond à 33 résidus sur le récepteur. Ces résidus ont été mutés individuellement en alanine, sauf pour le résidu G168 ainsi que les résidus C108 et C122 engagés dans un pont disulfure. Les résultats montrent que 11 mutations sur les 33 entrainent une perte substantielle dans l'affinité de liaison. La substitution des deux résidus W104 et W169 entraine les diminutions d'affinité les plus élevées (> 4,5 kcal/mol). Une diminution importante de l'affinité (entre 1,5 et 3,5 kcal/mol) est également observée lors de la substitution d'autres résidus hydrophobes (I103, I105, P106 et I165) alors que des effets plus faibles sont observés en substituant des résidus chargés (R43, E44, D126, E127 et D164). Le motif fonctionnel constitué de ces 11 résidus forme un patch au centre de la région de contact caractérisé par un cœur hydrophobe et représente 54% de la surface totale enfouie lors de la liaison. Un travail similaire sur hGH a permis de démontrer que 8 résidus seulement sur 31 participent à l'énergie de liaison à hauteur de 85% (Figure I.10). Les deux motifs fonctionnels du ligand et de la cible ont alors été comparés, ce qui a permis de mettre en évidence une complémentarité entre les

régions énergétiquement importantes du ligand et de la cible ainsi que les régions moins importantes. Les deux partenaires interagissent donc grâce à leurs épitopes fonctionnels complémentaires qui s'assemblent pour former un cœur hydrophobe.

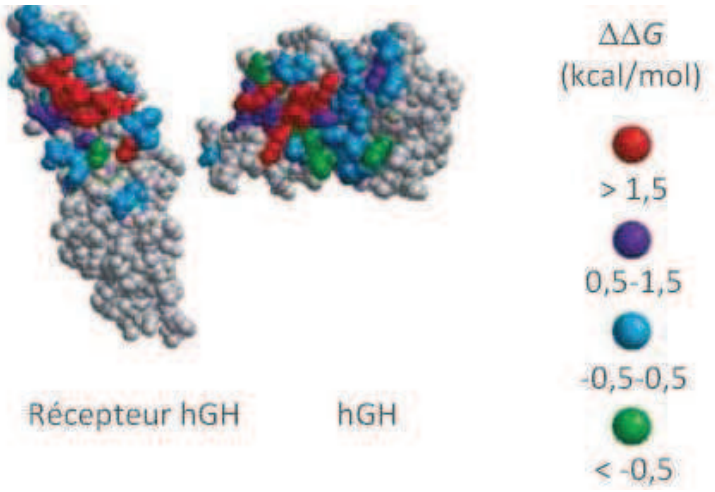


Figure I.10 : Structure du complexe hGH/récepteur hGHbp. Les résidus appartenant aux deux protéines et situés à l'interface du complexe sont coloriés en fonction de leur contribution à l'énergie de liaison, avec en particulier les résidus hot spots, participant à plus de 1,5 kcal/mol dans l'énergie de liaison, en rouge (Clackson et Wells, 1995).

En 1998, le groupe de Bogan et Thorn a réalisé une étude plus approfondie sur les contributions énergétiques de résidus individuels situés au niveau de l'interface dans un complexe protéine-protéine ; cette étude est réalisée par compilation et analyse d'une banque de données de mutants alanine dans des complexes hétérodimériques de protéines pour lesquels des tests d'affinité et des données structurales sont disponibles (Bogan et Thorn, 1998). Dans toutes les interfaces protéiques analysées, les résultats montrent que les « hot spots » de l'énergie de liaison sont situés près du centre de l'interface de l'hétérodimère et sont entourés par des résidus dont les contributions énergétiques à la liaison sont très faibles. Selon ces auteurs, pour qu'un résidu ait un grand impact sur l'énergie libre de liaison, il doit être largement protégé du contact avec le solvant. En effet, leur étude suggère que l'exposition d'un résidu au solvant empêche celui-ci de contribuer de manière importante à l'énergie libre de liaison, l'exclusion du solvant serait donc une condition nécessaire pour qu'un résidu soit considéré comme « hot spot », même si cette condition n'est pas suffisante. L'exclusion des « hot spots » du solvant est réalisée par l'ensemble des contacts périphériques dont la contribution énergétique est très faible mais qui protègent les résidus hydrophobes appartenant aux « hot spots » des molécules d'eau, augmentant ainsi l'affinité d'interaction en ralentissant la dissociation. Les résultats montrent également que la distribution des différents types d'acides aminés présents dans les

« hot spots » n'est pas aléatoire. En effet, dans cette étude, trois acides aminés majoritaires apparaissent dans les hot spots avec une fréquence de plus de 10% : Trp 21%, Arg 13,3% et Tyr 12,3%. De la même façon, certains résidus sont très rarement retrouvés dans les « hot spots » : moins de 3% des résidus Leu, Met, Ser, Thr et Val appartiennent à des « hot spots ». Cette distribution pourrait s'expliquer par la préférence dans les « hot spots » d'acides aminés capables d'établir plusieurs types d'interactions favorables. En effet, les deux résidus aromatiques hydrophobes Trp et Tyr contribuent à l'énergie de liaison grâce à l'effet hydrophobe sans pour autant avoir un coût entropique étant donné le faible nombre de degrés de libertés de rotation des liaisons de la chaine latérale. Ces résidus peuvent également établir des interactions impliquant leur noyau aromatique et constituent des donneurs de liaisons hydrogène. Le résidu Arg chargé positivement a la capacité de former simultanément plusieurs liaisons hydrogène, participer à un pont salin ou à une interaction cation-π.

I.4.4.2 Sélection des plates-formes protéiques hôtes

La mise en évidence d'ensembles restreints d'acides aminés présents à la surface de protéines et pouvant jouer un rôle prépondérant dans l'interaction avec le partenaire protéique est à l'origine de nombreux travaux de conception de ligands protéiques. En effet, ces ligands pourraient être obtenus en transférant ou greffant ces « hot spots » sur des protéines dites « plates-formes » capables de reproduire la topologie fonctionnelle de ce motif. Ceci devrait permettre de reproduire les interactions intermoléculaires observées entre le ligand protéique naturel et la protéine cible.

Les protéines de petite taille, contenant entre 30 et 80 résidus constituent des plates-formes particulièrement attractives pour la conception de protéines. Ces protéines possèdent une structure 3D ainsi qu'une activité biologique bien définies et sont généralement stabilisées par des contraintes structurales telles que des ponts disulfures, des ions métalliques ou des interactions intermoléculaires non covalentes. Les exemples de mini-protéines naturelles pouvant être utilisées comme plates-formes sont nombreux : toxines d'animaux, inhibiteurs de protéases, knottines, défensines, domaine de liaison à la cellulose, phéromones, doigts de zinc, leucine zippers et facteurs de croissance épidermique (Nygren et Uhlen, 1997 ; Vita et al., 1998). Les avantages de l'utilisation de petites plates-formes de présentation dans la conception

de nouvelles mini-protéines fonctionnelles sont nombreux (Stricher et al., 2006). 1) La reproduction d'une activité biologique dans des systèmes de petite taille correspond à une simplification significative par rapport à un interactant protéique naturel de plus grande taille. 2) La caractérisation structurale du nouvel interactant peut être obtenue par RMN grâce à la petite taille et à la stabilité des mini-protéines utilisées. 3) Les mini-protéines conçues peuvent présenter une plus grande résistance à la dégradation protéolytique grâce à leur flexibilité réduite. 4) Ces mini-protéines peuvent également être facilement et rapidement produites par synthèse chimique, permettant ainsi l'insertion d'acides aminés non naturels ou marqués de façon isotopique, de groupes photoactivables, de marqueurs radioactifs, de polymères, de carbohydrates, d'acides gras, de cofacteurs, d'étiquettes fluorescentes, d'étiquettes d'affinité… Ces composés peuvent être incorporés au niveau de positions spécifiques de façon à ne pas perturber la fonction de la mini-protéine conçue. 5) Ces mini-protéines peuvent également être produites en grandes quantités dans des systèmes d'expression biologiques peu coûteux, en vue d'applications industrielles.

I.4.4.3 Applications de l'approche de conception de mini-protéines par transfert de motif

Plusieurs exemples dans la littérature illustrent l'utilisation de l'approche de transfert de sites de liaison sur des plates-formes protéiques. Les premiers travaux réalisés portent sur l'utilisation de plates-formes ayant des éléments de structures secondaires similaires au ligand de référence au niveau de la région portant le motif fonctionnel. Les résultats encourageants obtenus au cours de ces premiers travaux, démontrant l'intérêt des plates-formes protéiques pour la conception de ligands artificiels, ont incité plusieurs équipes à généraliser ces premières approches.

I.4.4.3.1 Recherche de plates-formes fondée sur la similarité des structures secondaires

Les travaux pionniers dans ce domaine ont été réalisés au CEA par le groupe de Vita et al. et concernent le transfert de sites de liaison fonctionnels de lymphocyte humain CD4 sur une petite plate-forme naturelle en vue d'inhiber la liaison de la glycoprotéine gp120 de l'enveloppe du virus HIV-1 au CD4 (Drakopoulou et al., 1998). Le site de liaison du CD4 à la gp120 a été

localisé au niveau de la boucle similaire au CDR2. Un mime de CD4 (CD4M possédant 33 acides aminés) a donc été conçu par transfert des résidus de la boucle similaire au CDR2 sur une boucle de structure équivalente d'une plate-forme possédant trois ponts disulfures, la toxine de scorpion charybdotoxine caractérisée par une structure en α/ββ (Figure I.11). Cette plate-forme CD4M est capable d'inhiber de façon spécifique la liaison du CD4 à la gp120 mais néanmoins avec une faible affinité (IC50 de 20 µM) comparée à celle du CD4 (IC50 de 0,8 nM).

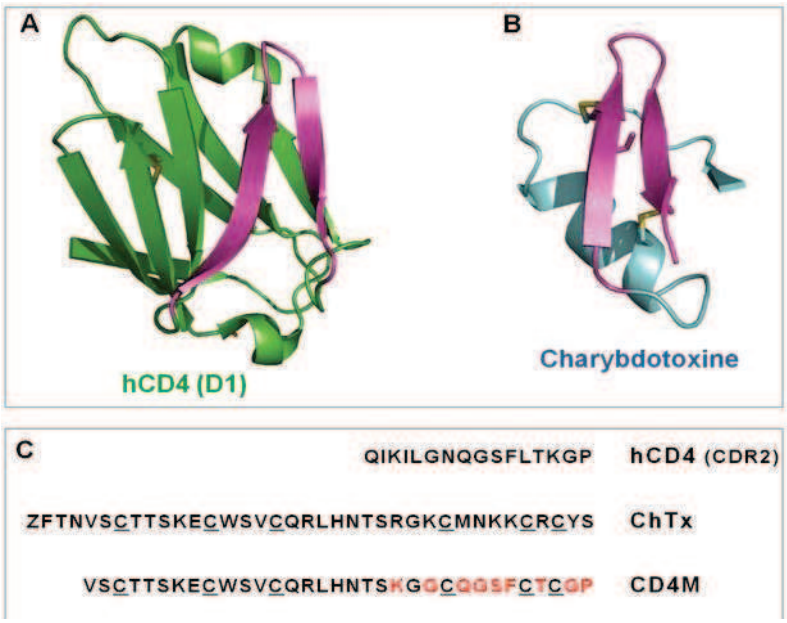


Figure I.11 : Comparaison entre le CD4 et la plate-forme mime. A et B : Structures 3D du domaine D1 du hCD4 (Code PDB 1CD4) et de la plate-forme structurale charybdotoxine (Code PDB 2CRD). Les protéines sont représentées en mode cartoon et la structure secondaire (2 feuillets β antiparallèles) portant le motif fonctionnel d'intérêt est coloriée en rose. C : Séquences du motif fonctionnel d'intérêt (CDR2) du hCD4, de la plate-forme charybdotoxine et de la plate-forme charybdotoxine modifiée CD4M (résidus modifés en rouge) (Drakopoulou et al., 1998).

Des travaux ultérieurs réalisés par le même groupe sur une plate-forme similaire à la charybdotoxine, la scyllatoxine, ont permis de concevoir une deuxième mini-protéine, le CD4M3 (27 acides aminés) par transfert de 9 résidus de CD4 (Figure 1.12 ; Vita et al., 1999). Cette nouvelle plate-forme inhibe la liaison du CD4 à la gp120 avec un IC50 qui est encore au moins 25 fois supérieur à celui de la CD4 native (IC50 de 20 à 350 µM en fonction de la souche virale exprimant les gp120 recombinantes). L'analyse structurale et fonctionnelle réalisée a permis d'identifier cinq mutations supplémentaires qui, une fois réalisées dans la plate-forme, ont conduit à une augmentation de l'affinité pour la gp120 de 40 à 400 fois en fonction de la souche virale (IC50 de 0,1 à 1 µM). Ce mutant CD4M9 inhibe l'infection des

cellules CD4+ par différentes souches virales HIV-1 confirmant l'activité antivirale large spectre de cette plate-forme. Après plusieurs étapes d'optimisation, un nouveau mime de CD4 a été obtenu (Martin et al., 2003). Ce mutant appelé CD4M33 inhibe la liaison du CD4 à deux différents isolats de gp120 avec une affinité similaire au CD4 recombinant (IC50 de 7,5 nM et 4 nM) et bloque l'entrée et l'infection des cellules immunitaires par les souches HIV-1 *in vitro*.

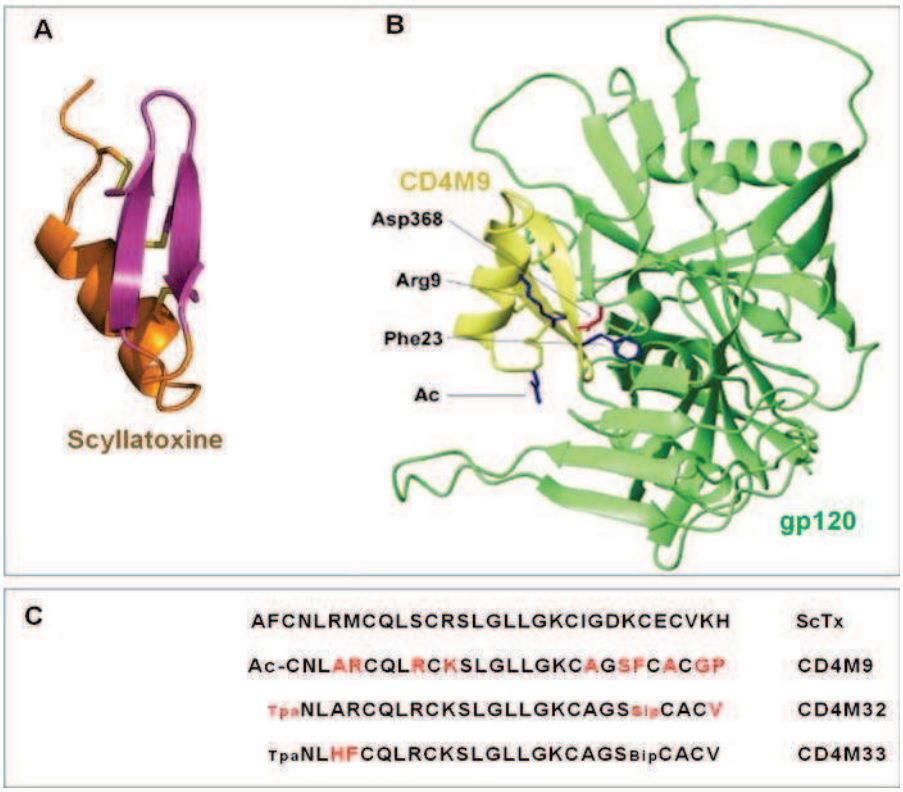


Figure I.12 : Structures 3D et séquences de la plate-forme Scyllatoxine et ses variants. Les protéines sont représentées en mode cartoon. A : Structures 3D de la plate-forme Scyllatoxine (Code PDB 1SCY) en orange, avec la structure secondaire (2 feuillets β antiparallèles) sur laquelle est transféré le motif fonctionnel d'intérêt coloriée en rose. B : Structure 3D d'un modèle de complexe CD4M9 (jaune)/gp120 (vert) avec les résidus essentiels dans l'interaction représentés en bâtons bleu (CD4M9) et rouge (gp120). C : Séquences de la plate-forme Scyllatoxine ainsi que ses variants CD4M9, CD4M32 et CD4M33. Chaque mutation par rapport à la protéine précédente est indiquée en rouge. Ac : groupement acétyl N-terminal ; Tpa : Acide thiopropionique ; Bip : Biphénylalanine. (Martin et al., 2003).

Les charpentes en α/ββ des toxines de scorpion, caractérisées par une hélice α N-terminale à trois pas reliée par trois ponts disulfures à deux feuillets β en épingle à cheveux, ont été utilisées dans d'autres travaux de conception de plates-formes par mimétisme d'épitopes naturels du fait de leur conformation particulièrement stable. Alors que le groupe de Vita et al. a exploité le motif en feuillets β de ces toxines de scorpion, le groupe de Li et al. a exploité le segment en hélice α d'une toxine de scorpion appelée BmBKTx1 (Li et al., 2008). Ce travail décrit la conception d'un mime de la protéine p53 suppresseur de tumeurs afin de bloquer son

interaction avec les protéines oncogènes MDM2 et MDMX qui en régulent négativement l'activité et la stabilité, et dont la surexpression dans plusieurs tumeurs conduit à l'inactivation de p53 et à la survie des tumeurs. La liaison de ces protéines avec p53 se fait au niveau du motif hélicoïdal de séquence $T_{18}F_{19}S_{20}D_{21}L_{22}W_{23}K_{24}L_{25}L_{26}$. En particulier, lors de la liaison au domaine N-terminal de MDM2 ou MDMX, p53 adopte une structure en hélice α où les résidus F_{19}, W_{23} et L_{26} de p53 s'insèrent dans une cavité hydrophobe des protéines oncogènes. Les résidus considérés comme « hot spots » F_{19}, L_{22}, W_{23} et L_{26} ont donc été replacés au niveau des positions topologiquement équivalentes sur la plate-forme BmBKTx1, une protéine synthétique de 31 résidus adoptant une structure hélicoïdale partielle au niveau de la région d'intérêt (Figure I.13).

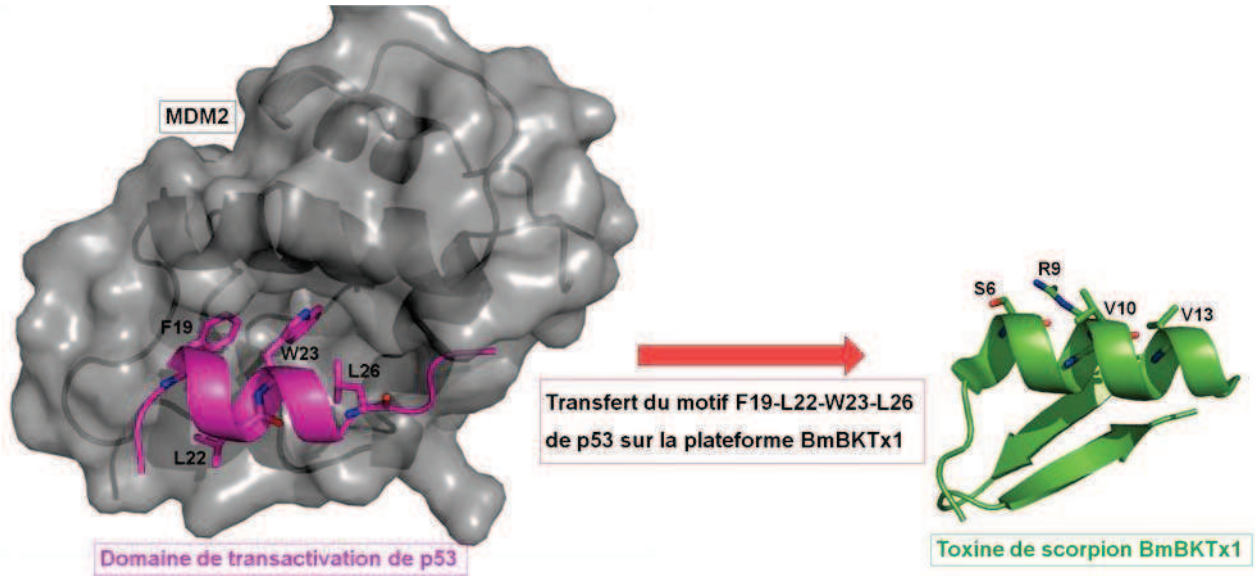


Figure I.13 : Comparaison des structures 3D du domaine de p53 et de la plate-forme BmBKTx1. A gauche : Complexe de MDM2 représenté en mode surface en gris avec le domaine de transactivation de p53 représenté en mode cartoon en rose avec les résidus essentiels à la liaison à MDM2 représentés en mode bâtons et annotés (Code PDB 1YCR). A droite : Structure 3D de la plate-forme BmBKTx1 caractérisée par une hélice α similaire à celle portant le motif d'intérêt dans p53 (Code PDB 1R1G).

Le peptide résultant nommé stoppin-1 adopte également en solution aqueuse une conformation α hélicoïdale partielle identique à celle de BmBKTx1. Des études basées sur les spectres de dichroïsme circulaire montrent que stoppin-1 se lie à des domaines synthétiques de MDM2 (^{syn}MDM2) et MDMX (^{syn}MDMX) impliqués dans la liaison à p53, de manière similaire à p53, par enfouissement du résidu Trp23 dans les cavités hydrophobes de MDM2 et MDMX. L'affinité de stoppin-1 pour ^{syn}MDM2 et ^{syn}MDMX (Kd de 790 nM et 994 nM respectivement) est seulement 6 à 3 fois plus faible que celle de p53 pour ces mêmes ligands

(Kd de 123 nM et 279 nM respectivement), tandis BmBKTx1 n'a aucune affinité pour ces ligands. Ces résultats valident donc l'approche de transfert de résidus qui a permis de concevoir un inhibiteur compétitif des interactions p53-MDM2/MDMX. En revanche, il a été démontré que stoppin-1 est incapable de tuer des cellules tumorales, probablement à cause de son incapacité à traverser la membrane cellulaire. Un nouveau peptide stoppin-2 a été conçu à partir de stoppin-1 par substitution de 5 résidus en résidus Arg formant ainsi un groupe de résidus cationiques destinés à faciliter l'entrée du peptide dans les cellules tumorales. Stoppin-2 possède une affinité pour ^{syn}MDM2 similaire à stoppin-1, en revanche contrairement à stoppin-1, stoppin-2 est capable de réduire drastiquement la viabilité de cellules tumorales.

Une autre exemple de conception de ligand protéique utilisant la similarité de structure secondaire entre un ligand de référence et une protéine plate-forme a permis de concevoir un analogue peptidique de 31 résidus qui mime une protéine plaquettaire de plus grande taille, le facteur anti-héparinique 4 (PF4) (Butcher et al., 1997). La plate-forme structurale choisie pour le transfert de motif de liaison est le peptide « leucine zipper » GCN4, dont la fonction et la séquence sont totalement différentes de la protéine PF4. Cependant, la structure tridimensionnelle de la plate-forme GCN4 consistant en deux hélices α enroulées en hélice l'une autour de l'autre grâce à des résidus leucine, présente des similarités avec la protéine PF4 (Figure I.14). L'activité de liaison à l'héparine de PF4 a été introduite uniquement par incorporation de 4 lysines de PF4 dans des positions équivalentes sur les hélices de la plate-forme GCN4. Des expériences de dichroïsme circulaire et des tests de liaison ont montré que le peptide « leucine zipper » conçu, appelé $PF4_{zip}$, adopte une conformation hélicoïdale stable et montre une activité significative de liaison à l'héparine similaire à PF4.

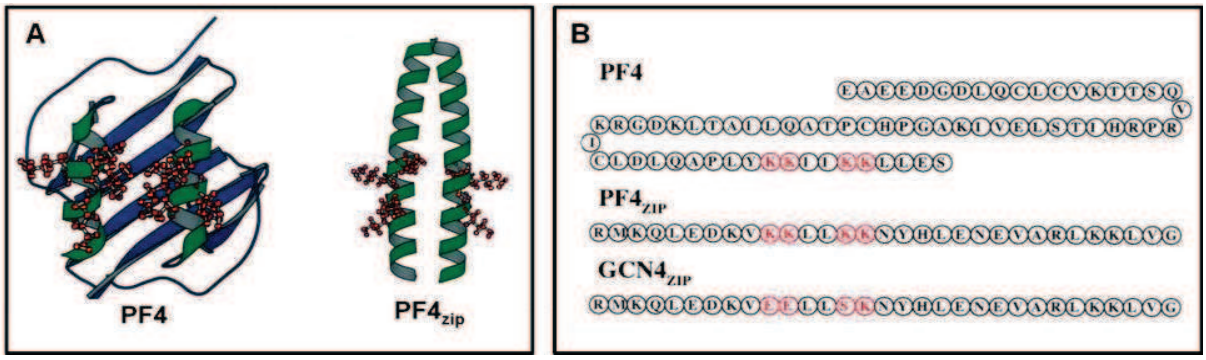


Figure I.14 : A : Structures cristallographiques du dimère PF4 (à gauche) et du domaine « leucine zipper » modifié $PF4_{zip}$ (à droite). Les résidus Lys essentiels dans la liaison à l'héparine sont représentés sur les deux protéines en mode bâtons et boules en rouge. B : Séquences de la protéine naturelle PF4 ainsi que de la plate-forme naturelle $GCN4_{zip}$ dans laquelle 4 lysines ont été introduites pour donner la plate-forme modifiée $PF4_{zip}$ (Butcher et al., 1997).

L'approche de transfert de motifs de liaison utilisant la similarité de structure secondaire entre le ligand de référence et la protéine plate-forme a également été appliquée à la conception d'inhibiteurs de l'interaction entre la glycoprotéine membranaire plaquettaire IIb-IIIa (GPIIb-IIIa) et le fibrinogène. La GPIIb-IIIa est une protéine présente à la surface des plaquettes circulantes et est impliquée dans les phénomènes d'agrégation plaquettaire conduisant à la thrombose, suite à sa liaison avec le fibrinogène. L'inhibition de la liaison entre GPIIb-IIIa et le fibrinogène constitue donc une cible thérapeutique stratégique dans la prévention de la thrombose artérielle. La liaison se produit grâce à la reconnaissance par la sous unité GPIIIa du motif RGD présent dans la séquence du fibrinogène. Afin de déterminer le mode d'interaction entre la GPIIIa et le fibrinogène, un anticorps monoclonal AC7 a été produit contre un peptide synthétique dérivé du fragment de GPIIIa responsable de la liaison au motif RGD. Il a été démontré que l'anticorps AC7 interagissait avec les plaquettes actives par l'intermédiaire de la chaîne lourde du CDR3 (H3) qui possède une séquence RQMIRGYFDV montrant une certaine homologie avec le motif RGDF du fibrinogène. Un décapeptide synthétique correspondant à la séquence RQMIRGYFDV de H3 inhibe l'agrégation plaquettaire et la liaison du fibrinogène aux plaquettes avec un IC50 de 700 µM. Ce résultat suggérait que le motif RGXFD de H3 a une structure qui mime la structure en coude β de type II' de la séquence RGDF du fibrinogène permettant la reconnaissance du récepteur (Jarrin et al., 1994). Le décapeptide a donc été étudié par spectroscopie RMN et les résultats montrent que la conformation d'interaction serait une hélice α ; dans cette conformation, l'arrangement tridimensionnel des résidus R, G, D et F du décapeptide est similaire à celui des résidus R, G, D et F dans une conformation en coude β de type II' présente dans les antagonistes de GPIIb-IIIa. Afin de contraindre la séquence du décapeptide à

une conformation en hélice α, la toxine de scorpion leiurotoxin I, adoptant une structure α/β (une hélice α N-terminale reliée par trois ponts disulfures à deux feuillets β antiparallèles à l'extrémité C-terminale), a été utilisée comme plate-forme. L'hélice α de la leiurotoxin I est alors remplacée par la séquence de H3, légèrement modifiée afin de préserver les ponts disulfures (NQCIRGCFDV). La toxine modifiée H3-leiurotoxin possède la même structure que la toxine native et montre une activité identique à celle du décapeptide. En effet, des essais d'inhibition de l'agrégation plaquettaire montrent que la toxine modifiée possède un IC50 de 200 μM (Mer et al., 1998).

Smith et ses collaborateurs ont effectué des travaux similaires à ceux du groupe de Jarrin fondés sur la création de nouvelles interfaces protéine-protéine grâce à l'utilisation d'une stratégie de transfert de boucles flexibles présentes à la surface de la plupart des protéines et qui sont souvent impliquées dans les phénomènes de liaison entre protéines (Smith et al., 1995). Dans cette stratégie, la séquence en acides aminés d'une boucle flexible possédant une activité biologique portée par une protéine, est utilisée afin de remplacer une boucle présente sur une protéine indépendante. Pour illustrer cette approche, la boucle HCDR3 de l'anticorps monoclonal Fab-9 dirigé contre les intégrines β3 a été greffée à la place d'une boucle de surface présente sur le module EGF de l'activateur tissulaire du plasminogène humain (t-PA). L'anticorps Fab-9 se lie aux intégrines β3 avec une affinité nanomolaire (Smith et al., 1994) grâce à la séquence de liaison SFGRGDIRN contenue dans la boucle HCDR3 stabilisée par deux ponts disulfures, en revanche, des peptides synthétiques contenant cette séquence ont une affinité 100 fois moins importante envers les intégrines β3. Cette séquence doit donc être transférée sur une plate-forme structurée afin de conserver l'affinité pour la cible. Le module EGF a été choisi pour le transfert de cette boucle car les structures de plusieurs protéines contenant ce module sont disponibles ; ces structures montrent que le module EGF contient une boucle exposée caractérisée par une structure en coude β et stabilisée par la présence de deux ponts disulfures. La plate-forme t-PA a donc été mutée par transfert des résidus SFGRGDIRN à la place de la boucle de structure équivalente dans le module EGF. Le variant LG-t-PA ainsi obtenu a gardé une activité enzymatique similaire à la t-PA native envers son substrat synthétique et envers son substrat naturel le plasminogène, et est stimulé de façon normale par le cofacteur physiologique : la fibrine. Ce variant se lie de manière spécifique à l'intégrine plaquettaire αIIbβ3 avec une affinité nanomolaire (Kd environ 0,9 nM) similaire à Fab-9 (Kd

de 5 nM) et à l'intégrine αVβ3 avec un Kd de 1,8 nM similaire au Kd de 1,7 nM obtenu avec Fab-9. Des expériences de compétition avec un peptide synthétique contenant la séquence RGD connue pour bloquer la liaison des ligands aux intégrines confirment que LG-t-PA se lie aux intégrines β3 testées au niveau de leur site de liaison au ligand.

I.4.4.3.2 Recherche de plates-formes non fondée sur la similarité des structures secondaires

Les travaux précédemment décrits reposent sur le transfert d'un motif fonctionnel issu d'un ligand de référence, sur une plate-forme possédant un élément de structure secondaire similaire à celui du ligand de référence dans la région portant le motif fonctionnel. Ces travaux ont permis de valider l'approche de conception de ligand protéique par greffage de motif fonctionnel. Cependant, cette méthode possède des limitations liées à la faible probabilité de trouver une plate-forme adaptée aux caractéristiques de chaque motif fonctionnel ainsi qu'à la contrainte d'avoir un motif fonctionnel continu (les acides aminés impliqués dans le motif fonctionnel sont consécutifs ou voisins dans la séquence primaire). Grâce au développement des méthodes de conception de ligands *in silico*, basées sur l'exploitation systématique des informations structurales des complexes naturels, il est devenu possible de concevoir de nouveaux ligands protéiques à partir de plates-formes dont les structures secondaires et tertiaires ainsi que les séquences sont totalement indépendantes du ligand de référence.

Ce concept se fonde en particulier sur les travaux de DeLano et ses collaborateurs qui ont montré qu'une protéine cible peut utiliser le même ensemble de résidus « hot spots » pour lier plusieurs partenaires différents ayant des structures secondaires différentes (DeLano et al., 2000). L'exemple le plus connu concerne un site de liaison consensus présent dans le fragment Fc de l'immunoglobuline G qui interagit avec quatre protéines (Protéine A, Protéine G, Facteur rhumatoïde, Récepteur Fc néonatal) dont les structures sont radicalement différentes, mais qui possèdent un ensemble de résidus « hot spots » similaires. Par ailleurs, ces auteurs ont démontré qu'un peptide de 13 résidus (Fc-III) sélectionné *in vitro* pour sa capacité à se lier au fragment Fc de l'IgG, mime les interactions établies par les ligands protéiques naturels avec la cible, tout en ayant une structure complètement différente de ces ligands naturels. Le peptide Fc-III se lie au fragment Fc avec une affinité de 25 nM tandis que les protéines A et G ayant des tailles quatre fois plus grandes que le peptide possèdent des affinités pour Fc de

deux fois seulement plus grandes (10 nM). De plus, ce petit peptide couvre une surface d'interaction avec le fragment Fc (650 $Å^2$) presque égale à celle couverte par les protéines A et G de tailles quatre fois plus grandes et par le facteur rhumatoïde 15 fois plus grand (environ 740 $Å^2$ chacun). Ces conclusions se sont trouvées renforcées par des analyses statistiques de complexes protéine-protéine qui ont conduit à la conclusion que des interfaces similaires pouvaient correspondre à des protéines possédant des structures 3D différentes (Keskin et Nussinov, 2005). D'autre part les mêmes auteurs ont pu montrer que des protéines présentant des interfaces similaires pouvaient interagir avec des sites de liaisons de structure 3D différente. Ceci est possible car ces interfaces se caractérisent à l'état lié par une complémentarité imparfaite des deux partenaires et par la présence des molécules d'eau interfaciales qui jouent un rôle important dans la stabilisation des complexes via des interactions intermoléculaires médiées. L'ensemble de ces observations est à l'origine de la proposition de concevoir des ligands protéiques en transférant un ensemble de résidus fonctionnels importants dans la liaison et présents sur un ligand de référence, sur des plates-formes ayant des structures secondaires différentes de ce ligand de référence. Le principe de cette approche pour la conception de ligand protéique a été proposé initialement par le groupe de Luhua Lai en 2000 (Liang et al., 2000) mais c'est seulement en 2006 et 2007 que les premières applications ont pu démontrer la validité de ce concept (Magis et al., 2006 ; Liu et al., 2007).

La première de ces applications, réalisées au sein de notre groupe, avait pour objectif la conception de ligands protéiques capables de bloquer le canal potassique Kv1.2 (Magis et al., 2006). La protéine utilisée pour identifier un motif fonctionnel susceptible d'être greffé sur des plates-formes protéiques était la BgK, une toxine naturelle d'anémone de mer de 37 acides aminés. Cette petite protéine est un inhibiteur naturel de ces canaux potassiques; elle se lie aux canaux Kv1.2 grâce à quatre résidus « hot spot » essentiels F6, S23, K25 et Y26 (motif FSKY) identifiés par « alanine scanning ». Bien que formant un motif compacte, les résidus « hot spot » sont éloignés au niveau de la séquence de la protéine et sont situés sur des éléments structuraux différents. Une recherche systématique de la PDB a été réalisée afin d'identifier des plates-formes protéiques possédant un groupe de quatre résidus adoptant une topologie similaire à celle des résidus du motif FSKY dans la structure du ligand de référence la BgK, et ce indépendamment de leurs structures secondaires. Pour ce faire, le motif fonctionnel de la BgK a été caractérisé par un ensemble de distances interatomiques séparant les Cα et Cβ des

carbones des quatre résidus du motif, calculées à partir de la structure RMN de la BgK. Grâce au logiciel STAMPS (Search for Three-Dimensional Atom Motif in Protein Structure) développé au sein du laboratoire, la PDB a été systématiquement criblée pour identifier des structures protéiques possédant un motif dont les carbones Cα et Cβ sont topologiquement similaires à ceux des résidus FSKY du motif de référence dans la BgK, et ce indépendamment du type de résidus identifiés. Cette recherche fondée sur la topologie du motif fonctionnel est indépendante de la structure secondaire du ligand de référence ainsi que de la plate-forme mime portant le motif. Il convient de noter néanmoins que ce principe de recherche permet d'identifier également des protéines présentant, dans la région du motif, des structures secondaires similaires à celle du motif de référence. Un filtre stérique est appliqué lors de la recherche afin de prendre en considération la complémentarité géométrique des plates-formes avec la cible et d'exclure les solutions qui conduisent à un encombrement stérique empêchant le positionnement correct du motif de liaison par rapport au site actif de la cible. Dans le cas de la conception de ligands des canaux Kv1.2, l'application du filtre stérique a conduit au rejet d'environ 90% des solutions obtenues sur les seuls critères topologiques. Un filtre électrostatique a également permis de sélectionner les plates-formes ayant un potentiel électrostatique, dans la région de liaison, proche de celui du ligand de référence. La recherche topologique *in silico* a permis d'identifier 20 plates-formes de structures secondaires différentes, possédant 4 résidus topologiquement similaires à ceux du motif FSKY et stériquement compatibles pour une interaction avec le canal Kv1.2. Après transfert des résidus du motif FSKY sur les plates-formes identifiées, la comparaison des potentiels électrostatiques a permis de sélectionner 11 plates-formes dont 4 ont été produites et testées. Trois d'entre elles, 1J75_m, 1CKK_m et 1H5O_m inhibent la liaison de BgK au canal Kv1.2 avec des Ki de 1,5, 1,6, et 0,5 µM respectivement. La structure 3D du ligand 1J75_m déterminée par diffraction X a montré que le transfert du motif fonctionnel de la BgK sur cette plate-forme n'entraînait aucune modification de sa structure secondaire et que ce motif de liaison adopte une topologie similaire à celle observée dans le ligand de référence BgK (figure I.15). Bien que peu de candidats ligands aient été produits au cours de cette première étude, le taux de succès (3/4) peut être considéré comme satisfaisant. Ceci peut être en partie lié au fait que des propriétés globales des plates-formes sont prises en compte de manière explicite dans la conception *in silico*. On note cependant que les constantes d'affinité obtenues à l'issue de la seule phase de conception *in silico* sont micromolaires. Bien que ces valeurs initiales puissent permettre d'aborder l'optimisation de ces ligands protéiques artificiels par des méthodes expérimentales,

elles sont cependant largement inférieures à celles de Bgk (environ 50 pM). Ces observations indiquent que l'évaluation et le développement de la méthode doivent être poursuivis.

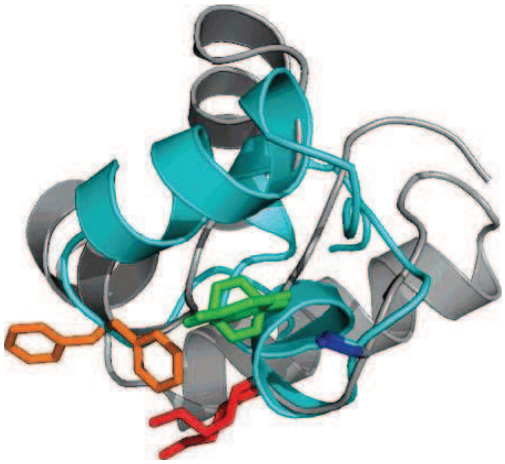


Figure I.15 : Superposition des structures cristallographiques du ligand 1J75_m avec le ligand de référence 1BGK montrant une superposition correcte des motifs de liaison FSKY. Les protéines sont représentées en mode cartoon en gris pour 1J75_m et en cyan pour 1BGK et les résidus du motif FSKY sont représentés en bâtons orange, bleu, rouge et vert respectivement (Magis et al., 2006).

En 2007, Luhua Lai et ses collaborateurs ont reporté l'application de cette méthode pour la conception d'un mime de l'érythropoïétine humaine (hEPO) capable de se lier au récepteur de l'érythropoïétine humaine (EPOR) (Liu et al., 2007). L'objectif de cette étude était la conception d'une protéine capable de mimer l'interaction de l'EPO avec son récepteur EPOR. Les données expérimentales de mutation ainsi que la structure cristallographique du complexe EPO/EPOR (code PDB 1EER) ont permis d'identifier un motif de résidus discontinus Phe48, Asn147 et Arg150 dans la protéine EPO comme étant les résidus essentiels à la liaison avec EPOR (Philo et al., 1996 ; Cheetham et al., 1998). Un criblage de la PDB réalisé afin de trouver une plate-forme protéique hôte capable d'accueillir les résidus essentiels à la liaison de EPO avec EPOR a permis de sélectionner 15 plates-formes satisfaisant les critères topologiques et caractérisées par une bonne complémentarité de forme avec la cible, une grande taille d'interface et une bonne compacité. Des étapes de sélection supplémentaires relatives à des contraintes expérimentales n'ont permis à ces auteurs de retenir qu'une seule plate-forme : la protéine PLCδ1-PH de rat. Des mutants de cette plate-forme protéique ont été conçus par transfert des trois résidus clés de la liaison, et les résultats ont montré que ces mutants possédaient une haute affinité de liaison *in vitro* et *in vivo* pour leur partenaire non naturel EPOR. En particulier, le triple mutant de PLCδ1-PH nommé ERPH1 a l'affinité de liaison pour EPOR la plus élevée avec un Kd de 24 nM obtenu pour le test *in vitro* et un IC50 de 5,7 µM

pour le test cellulaire, alors que la protéine native PLCδ1-PH ne montre pas d'affinité détectable dans les conditions des tests. On peut noter cependant que les structures cristallographiques de la protéine conçue ERPH1 et du complexe ERPH1/EPOR n'ont pas été réalisées empêchant de caractériser les interactions reproduites par ERPH1. Bien que ces résultats soient très encourageants, il est par ailleurs difficile d'évaluer la robustesse de l'approche de conception *in silico* utilisée car une seule protéine a été évaluée expérimentalement au cours de ces travaux.

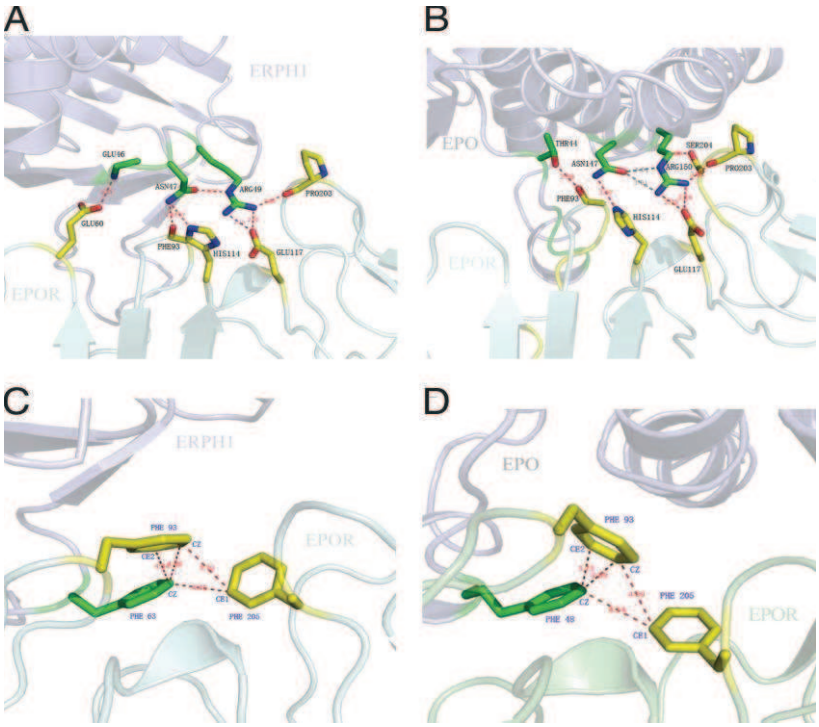


Figure I.16 : Principales interactions intermoléculaires entre le motif de la plate-forme ERPH1 et la cible EPOR, et entre le motif du ligand naturel EPO et EPOR. A et B : Liaisons hydrogène à l'interface des complexes ERPH1/EPOR et EPO/EPOR respectivement. C et D : Interactions hydrophobes à l'interface des complexes ERPH1/EPOR et EPO/EPOR respectivement. Les protéines sont représentées en cartoon et les résidus à l'interface en bâtons (Liu et al., 2007).

Une approche de greffage similaire aux deux approches précédemment décrites a été utilisée par l'équipe de David Baker dans le but de concevoir des protéines qui se lient à une région conservée présente au niveau de la tige de la protéine hémagglutinine (HA) du virus pandémique H1N1 1918 et reconnue par les anticorps neutralisants (Fleishman et al., 2011). Cette stratégie repose sur le greffage des motifs d'acides aminés sur des plates-formes protéiques leur permettant de reproduire des interactions précises avec la cible. Contrairement à la majorité des études de conception de ligands protéiques *in silico* utilisant des données de mutagénèse ou des données structurales pour identifier les résidus à greffer, les motifs et leur

topologies sont issus d'une méthode de simulation. La structure d'un complexe de HA avec le fragment Fab de l'anticorps CR6261 montre que l'anticorps se lie à HA au niveau d'un épitope présent sur la surface de la tige au niveau d'un sillon hydrophobe délimité par deux boucles. Une simulation d'ancrage réalisée à l'aide du programme RosettaDock a été utilisée pour déterminer des résidus « hot spots » ainsi que les positions que pourraient adopter ces résidus pour établir des interactions énergétiquement favorables avec la surface de la cible HA au niveau de l'épitope. Ce procédé a permis d'identifier à la surface de la cible trois positions « hot spot » HS1, HS2 et HS3 pour lesquelles des résidus pourraient établir des interactions d'empilement aromatique, de van der Waals ou des liaisons hydrogène favorables. Une fois les positions les plus favorables pour les résidus « hot spots » déterminées, un ensemble restreint de 865 structures protéiques sélectionnées en vue de faciliter la production des protéines ligands, est criblé afin de rechercher des plates-formes capables d'accueillir les résidus « hot spots » identifiés dans la topologie ad-hoc et possédant une complémentarité de forme avec la région de la tige. Chaque plate-forme est tout d'abord ancrée contre la région de tige grâce au programme PatchDock qui identifie des centaines de modes de liaison pour chaque plate-forme à l'aide d'un modèle gros-grain. Ces modes de liaison sont ensuite raffinés grâce au programme RosettaDock en favorisant les configurations qui maximisent la compatibilité de la chaine principale de la plate-forme avec autant de résidus « hot spots » que possible. Puis en utilisant le programme RosettaScript, les résidus « hot spots » des positions HS1, HS2 et HS3 sont introduits séquentiellement sur la plate-forme, de manière à avoir une similarité topologique entre les résidus « hot spots » issus de l'étape d'ancrage des acides aminés isolés et les résidus correspondants identifiés sur la plate-forme. Pour chaque conformation, la reproduction des interactions entre les résidus transférés sur la plate-forme et la cible HA est évaluée ainsi que l'absence d'encombrement stérique. L'application de cette stratégie a permis d'identifier des plates-formes incorporant les 2 résidus « hot spots » correspondant à HS1 et HS2 et une série d'autres plates-formes incorporant les 3 résidus « hot spots ». Grâce au programme RosettaDesign, des mutations supplémentaires au niveau du reste de la surface de la plate-forme sont suggérées sur la base des modèles de complexes afin de mieux accommoder les résidus « hot spots ». Ces modèles de complexes sont filtrés en fonction de l'énergie de liaison, de la surface enfouie et de la complémentarité de forme. Sur un total de 73 candidats ligands protéiques, produits à la surface de levure et testés pour leur activité de liaison à la région formant la tige de HA, 2 plates-formes HB36 (contenant uniquement les résidus « hot spots » de HS1 et HS2) et HB80 (contenant les 3 résidus « hot spots ») ont montré des affinités de

liaison pour HA de 200 nM et 5 µM respectivement. Les plates-formes natives dont sont issus ces variants ne se lient pas à HA, démontrant ainsi que la liaison se fait par l'intermédiaire de la surface conçue *de novo* de la plate-forme. Une bibliothèque de variants de HB36 et HB80 a été générée par mutagénèse par saturation de sites ou par PCR « error prone » en vue d'une maturation d'affinité ; les résultats de criblage ont permis d'identifier plusieurs variants ayant une affinité plus grande pour la cible par rapport aux plates-formes d'origine, les meilleurs Kd sont obtenus avec les variants HB36.3 et HB80.3 (4 et 3 nM respectivement). Une structure cristallographique du variant HB36.3 en complexe avec HA a confirmé le mode de liaison du ligand avec la cible tel que suggéré par le modèle.

Une approche, présentant des similarités mais également des différences importantes avec les stratégies de conception utilisées dans les exemples de greffage indépendant de la structure secondaire précédemment décrits, a été développée pour concevoir une protéine artificielle capable d'être reconnue par l'anticorps monoclonal neutralisant b12 développé contre la protéine d'enveloppe gp120 du virus HIV (Azoitei et al., 2011). Cet anticorps se lie à un épitope bien conservé de la gp120, au niveau du site de liaison de la protéine CD4 (CD4bs), constitué de 2 segments appelés boucles CD4b (résidus 365 à 372) et ODe (résidus 472 à 476). Ces travaux ont consisté à greffer un épitope composé de deux fragments polypeptidiques de la protéine gp120, reconnu par l'anticorps monoclonal neutralisant b12, sur des plates-formes capables d'accueillir ces fragments. Cette approche requiert, comme dans le cas de la méthode de conception indépendante de la structure secondaire, la recherche de plates-formes possédant des points d'ancrage pour greffer chacune des extrémités des segments polypeptidiques composant l'épitope à transférer. Cette stratégie est résumée sur la figure I.17.

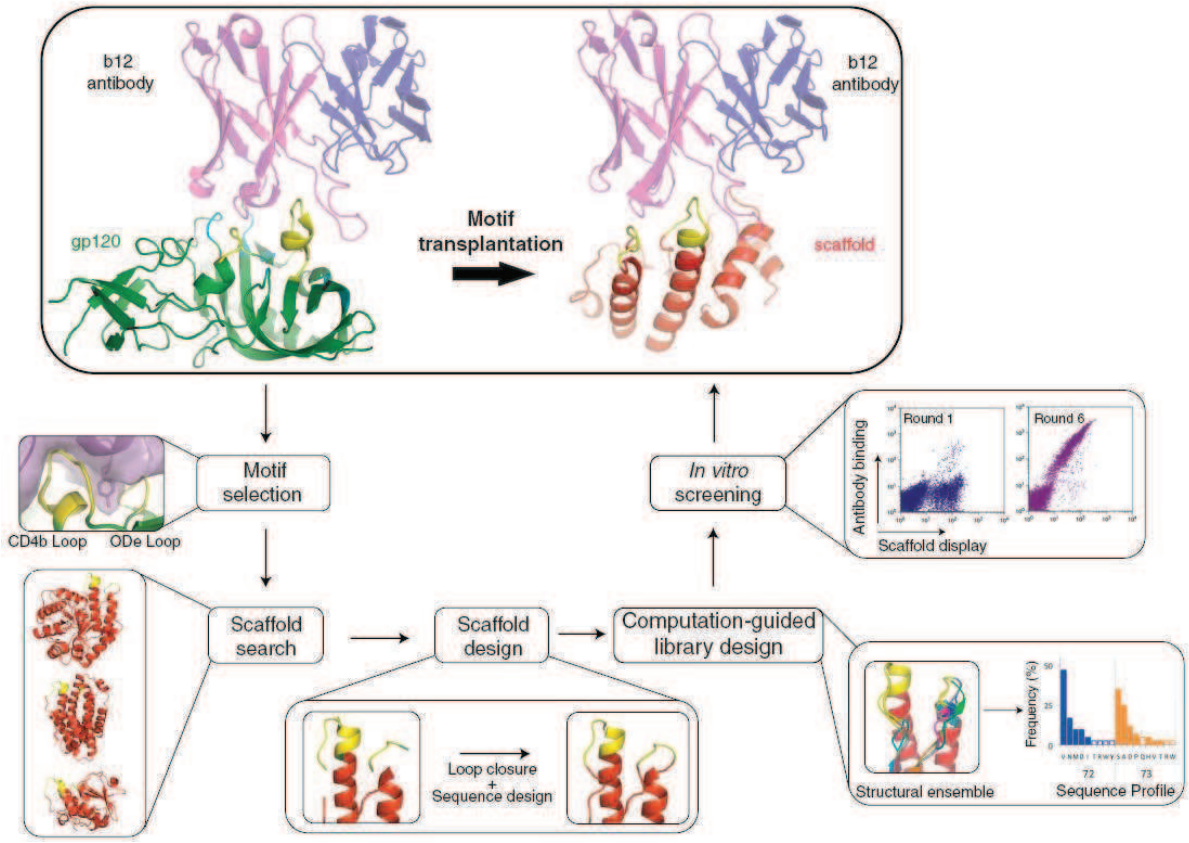


Figure I.17 : Schéma des étapes de conception *in silico* de ligands ciblant l'anticorps b12 par transfert d'un motif discontinu constitué de deux boucles présentes dans la protéine gp120, sur des plates-formes non apparentées (Azoitei et al., 2011).

La recherche de ces plates-formes à été réalisée dans une banque de données PISCES correspondant à une PDB triée contenant des structures non redondantes et de résolution élevée (13337 structures). Ainsi, des plates-formes capables d'accepter la chaîne principale des segments contenant le motif de liaison sont recherchées grâce au programme Multigraft Match implémenté dans la suite logicielle Rosetta. Partant des coordonnées atomiques des motifs contenus dans les deux segments CD4b et ODe de la gp120 en complexe avec b12, le programme Multigraft Match va aligner dans un premier temps le 1er segment à transférer avec tous les résidus des plates-formes contenues dans la PDB. Quatre types d'alignement sont utilisés : une superposition totale des résidus du segment, un alignement du résidu C-terminal du segment uniquement, un alignement du résidu N-terminal du segment uniquement et un alignement simultané des résidus C- et N-terminaux du segment. Lorsque les critères du 1er alignement sont satisfaits, un alignement avec le 2ème segment est effectué par alignement des deux résidus C- et N-terminaux uniquement. Le logiciel Rosetta inclut un filtre géométrique permettant de comparer les RMSD entre le segment de référence et le segment porté par la

plate-forme ainsi qu'un filtre stérique permettant d'éviter les encombrements stériques intramoléculaires. L'étape finale de la recherche consiste à éliminer les plates-formes sélectionnées qui conduiraient à un encombrement stérique intermoléculaire avec la cible b12. En plus des filtres automatiques, une sélection manuelle a permis d'éliminer les solutions pour lesquelles le motif à transférer fait un angle de 90° avec la chaine principale de la plate-forme. Ce protocole a permis de sélectionner 11 plates-formes satisfaisant les critères géométriques et stériques appliqués. La phase suivante de la méthode consiste à remplacer la chaîne principale native des plates-formes sélectionnées par celle des segments du motif puis de construire de nouveaux segments permettant de connecter le motif transféré au squelette de la plate-forme ainsi que les chaînes latérales avoisinant le motif. Ceci est réalisé grâce au programme Multigraft Design. Pour chaque plate-forme identifiée, des bibliothèques de mutants caractérisés par une diversité de structure et de séquence primaire au niveau des segments de raccord sont construites. Sur un total de 62 variants de plates-formes testés pour une expression et purification chez *E. coli*, 25 mutants seulement ont pu être exprimés de façon soluble et purifiés. Ces mutants ont été testés pour leur capacité de liaison à l'anticorps b12 par SPR (résonance plasmonique de surface). Une molécule guide produite, 2bodx_03, différant de la plate-forme native par 39 mutations et 11 délétions, se lie de manière spécifique à b12 mais avec une très faible affinité (Kd de 300 μM). Le taux de succès obtenus (nb de ligands protéiques capables de lier la cible/nombre de ligands conçus) ainsi que les affinités des variants sont donc très faibles. Le criblage par présentation à la surface de levure d'une 2ème bibliothèque de mutants de 2bodx_03 obtenus par mutagénèse aléatoire de la totalité de la séquence protéique a permis d'identifier le mutant 2bodx_43 se liant à b12 avec un Kd de 33 nM. Une 3ème bibliothèque de mutants conçue à partir de 2bodx_43 a permis d'isoler le mutant 2bodx_45 se liant à b12 avec une forte affinité similaire à la gp120 (Kd de 10 nM), alors qu'il ne diffère de 2bodx_43 que par 2 mutations. La superposition des structures cristallographiques des complexes 2bodx_43-b12 et gp120-b12 a montré un degré élevé de mimétisme, notamment au niveau des motifs de liaison de la plate-forme et de la gp120 ainsi que la reproduction des interactions existantes entre l'épitope et le paratope du complexe de référence. Cet exemple montre bien que malgré la faible efficacité de la méthode *in silico* mise en œuvre, le couplage de cette méthode avec des méthodes d'optimisation expérimentales peut être d'une redoutable efficacité.

Cette approche a également été utilisée pour la conception de mime de l'épitope de la

protéine d'enveloppe gp41 du virus HIV reconnu par l'anticorps 2F5 (Azoitei et al., 2012). Pour ce faire, les chaines latérales du motif fonctionnel du virus HIV gp41 reconnu par l'anticorps neutralisant mAb 2F5 ont été transférées sur de nouvelles protéines plates-formes ayant des segments de chaines principales exposés dont les conformations sont similaires à la conformation de la chaine principale de l'épitope lorsque celui-ci est lié à l'anticorps. Ceci a été réalisé par greffage de fragments de chaine principale correspondant à l'épitope à transférer et par greffage de chaines latérales. Des études cristallographiques et de mutagénèse ont permis d'identifier le fragment ^{657}EQELLELDKWASLW^{670} de gp41 comme étant le site de liaison à l'anticorps 2F5 avec les résidus DKW considérés comme critiques dans la liaison et dans la neutralisation du virus (Ofek et al., 2004 ; Julien et al., 2008 ; Bryson et al., 2009). Dans cette étude, des séquences raccourcies de l'épitope entier 657-670 de gp41, centrées sur le motif DKW, ont été choisies pour le transfert afin d'identifier des plates-formes mimant au mieux la conformation du motif d'intérêt. Un criblage d'une partie de la PDB a permis d'identifier des plates-formes avec des chaines principales exposées similaires à celle de l'épitope, puis les chaines latérales des séquences raccourcies de l'épitope ont été transférées sur les plates-formes et enfin des étapes de conception supplémentaires ont été appliquées afin de sélectionner des plates-formes capables d'interagir avec 2F5. Ces étapes ont permis de concevoir 28 plates-formes différentes issues de 28 protéines natives différentes. Des variants de ces plates-formes ont ensuite été conçus grâce au logiciel Rosetta afin d'optimiser la solubilité et la stabilité des plates-formes. Sur les 98 variants conçus, 35 plates-formes se sont avérées solubles et ont pu être purifiées, dont sept se lient à 2F5 avec des Kd < 1 µM. Par ailleurs, une comparaison entre l'affinité de liaison pour 2F5 des épitopes peptidiques raccourcis seuls et celle des plates-formes portant ces épitopes montre que les plates-formes portant les épitopes se lient à 2F5 avec une meilleure affinité que les épitopes seuls correspondants, grâce à des taux de dissociation plus faibles.

Trois plates-formes issues de cette approche de transfert de chaines latérales (SC_2cx5, SC_1wnu et SC_1zko montrant des Kd pour 2F5 de 40,2 nM, 109,4 nM et >2 µM respectivement) ont été choisies pour réaliser la 2ème étape de l'étude. Les 3 protéines natives de ces plates-formes ont servi de support pour le transfert des chaines principales des épitopes précédemment sélectionnés (Figure I.18).

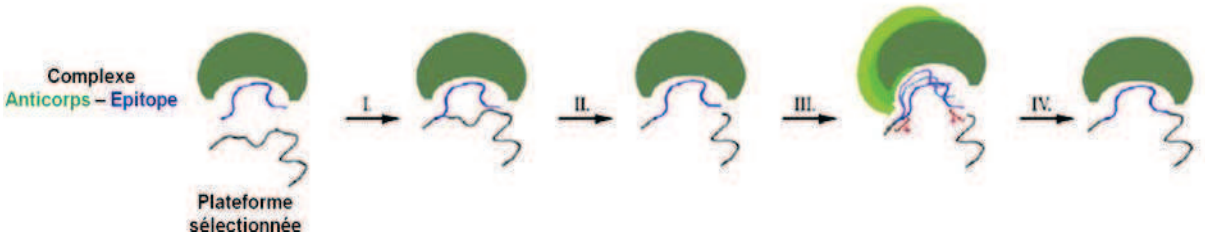


Figure I.18 : Etapes de transfert de chaine principale. I) Alignement de l'épitope (en bleu) sur la plate-forme sélectionnée. II) Délétion de la chaine principale de la plate-forme native correspondant à l'épitope résultant en un polypeptide déconnecté. III) Reconstruction de nouvelles régions aux extrémités de l'épitope (en rouge) afin d'intégrer l'épitope dans la plate-forme. IV) Plate-forme finale portant l'épitope transféré en interaction avec l'anticorps.

Le protocole utilisé est similaire à celui décrit ci-dessus dans le travail de Azoitei et al. portant sur le transfert de chaine principale et des chaines latérales de l'épitope de la gp120 (Azoitei et al., 2011). Un total de 5000 variants ont été conçus à partir de chacune des 3 plates-formes, puis ont été soumis à divers filtres et classés selon des critères énergétiques et géométriques, ainsi que des critères relatifs à la compacité. Trois ou quatre variants de chacune des 3 plates-formes parentes ont été sélectionnés pour une caractérisation expérimentale. Les résultats des tests de liaison ont montré que les trois variants de la plate-forme BB_2cx5 issue de transfert de chaine principale de l'épitope ont des affinités pour 2F5 (Kd de 1,4 nM, 8 nM et 15,5 nM) supérieures à celle de la plate-forme SC_2cx5 issue de transfert de chaines latérales de l'épitope (Kd de 40,2 nM). De même les trois variants de la plate-forme BB_1wnu ont des affinités pour 2F5 (Kd de 8,3 nM, 8,9 nM et 29 nM) supérieures à celle de la plate-forme SC_1wnu (Kd de 91 nM). En revanche, aucun des 3 variants de BB_1zko ne montre d'affinité significative pour 2F5 alors que le variant SC_1zko possède une faible affinité (Kd > 2 µM). Une analyse cristallographique des deux paires de plates-formes SC_2cx5/BB_2cx5_001 et SC_1wnu/BB_1wnu_001 montre que les structures sont en assez bon accord avec les modèles établis *in silico*. Cette étude a donc montré la validité de l'approche de greffage de segment peptidique. Toutefois, on peut noter que les différences d'affinité observées entre les ligands protéiques issus des deux méthodes de greffage restent modestes (de l'ordre d'un facteur 10).

I.5 Conclusion

Les exemples décrits dans le paragraphe I.4.4.3 montrent clairement que l'objectif de développer des ligands protéiques *in silico* par greffage de motifs fonctionnels sur des plates-formes protéiques peut être atteint. Ces motifs fonctionnels sont le plus souvent issus d'études

de mutagénèse dirigée et d'études structurales disponibles pour un nombre croissant de partenaires protéiques naturels. Des interactions énergétiquement importantes peuvent également être identifiées par des méthodes de simulations. Plusieurs méthodes de greffage ont été reportées. Le greffage de chaines latérales sur des plates-formes identifiées par analogie de structure secondaire avec un ligand de référence a tout d'abord été mis en œuvre et a conduit à des succès importants. L'introduction de méthodes de sélection de plates-formes protéiques indépendante de la structure secondaire, bien que nécessitant des outils bioinformatiques performants, a également permis d'obtenir des ligands protéiques d'affinité significative. Les méthodes plus récemment introduites visant à greffer des segments peptidiques correspondants à des épitopes protéiques ont également conduit à des ligands protéiques d'intérêt. En dépit des succès obtenus, le nombre d'études mettant en œuvre ces diverses approches reste faible et les données expérimentales disponibles ne permettent pas encore d'évaluer la robustesse des différentes approches *in silico*.

CHAPITRE II : Les interactions protéines – protéines

La plupart des processus biologiques impliquent une reconnaissance entre des partenaires protéiques via la mise en contact de leurs surfaces. Ceci a lieu par la mise en jeu d'interactions intermoléculaires établies au niveau des interfaces ainsi formées. La connaissance des caractéristiques des interfaces protéine-protéine naturelles et de la nature des interactions résultant de l'association de ces partenaires peut contribuer au succès des approches de conception de ligands protéiques *in silico*. En effet, des critères de sélection et de validation des ligands protéiques artificiels issus de ces méthodes pourraient être définis afin d'augmenter leur taux de succès. Les interfaces protéine-protéine peuvent être analysées selon plusieurs critères : les propriétés structurales, les propriétés chimiques, les propriétés électrostatiques ainsi que les aspects thermodynamiques liés aux phénomènes de solvatation et aux différentes interactions mises en jeu au niveau des interfaces. Dans ce chapitre nous allons passer brièvement en revue les travaux reportés dans la littérature décrivant l'analyse de complexes protéine-protéine naturels et les paramètres permettant de les caractériser.

II.1 Classification des complexes protéine-protéine naturels

La capacité de certaines protéines à former des complexes spécifiques stables est fondamentale dans les processus biologiques. Il existe deux types de complexes selon leur durée de vie :

- les complexes non covalents ou transitoires : les complexes antigène-anticorps (Fab D44.1-lysozyme), enzyme-inhibiteur (Elastase-elafin), récepteur-ligand (Herceptin-EGFR) ou les complexes impliqués dans la transduction de signal (Rho-Rho GAP) appartiennent au groupe des complexes non covalents. De manière générale, les différents composants formant ce type de complexe ne sont pas co-localisés et nécessitent donc d'être stables à l'état libre, de se replier séparément et d'assurer leurs fonctions de manière indépendante avant de s'associer.
- les complexes permanents : dans un complexe permanent, les composants du complexe *in vivo* ne peuvent pas exister sous forme stable lorsqu'ils sont dissociés. Par exemple, les interactions entre les sous-unités d'une protéine oligomérique sont permanentes et ne sont rompues que lorsque la protéine est dénaturée ou dégradée, tel est le cas de l'homo-dimère

répresseur Arc du phage P22.
Il faut cependant garder à l'esprit que l'appartenance d'une protéine à l'une ou l'autre de ces catégories n'est pas toujours stricte et peut dépendre des conditions physiologiques et de l'environnement (Nooren et Thornton, 2003).

L'analyse des interactions protéine-protéine réalisée ci-après ne concernera que les complexes hétéro-dimériques non covalents qui jouent un rôle prépondérant dans divers processus biologiques et qui vont servir de modèles au cours de notre travail, dans un objectif de conception de nouveaux ligands protéiques inhibiteurs de cibles d'intérêt. Les 1ères analyses réalisées sur les interfaces protéine-protéine dans les complexes hétéro-dimériques étaient basées sur un nombre très restreint de structures (Chothia et Janin, 1975 ; Janin et Chothia, 1990). Avec le nombre sans cesse croissant de structures 3D de protéines à l'état libre et à l'état lié disponibles dans la PDB et couvrant une gamme d'activités plus larges que celles précédemment décrites, les propriétés structurales et fonctionnelles des interfaces protéine-protéine sont à présent mieux connues et permettent une meilleure compréhension des phénomènes d'association.

II.2 Propriétés structurales des interfaces

Dans un complexe, les protéines entrent en interaction via une surface de contact appelée interface, qui peut être caractérisée selon plusieurs paramètres.

II.2.1 Taille de l'interface

La taille de l'interface est estimée par certains auteurs comme la surface totale enfouie lors de l'interaction (ΔASA), elle est mesurée par la somme de la surface accessible au solvant (ASA pour Accessible Surface Area) des deux protéines (A et B) moins la surface accessible au solvant du complexe AB (Figure II.1) :

$$\Delta ASA = ASA_A + ASA_B - ASA_{AB}.$$

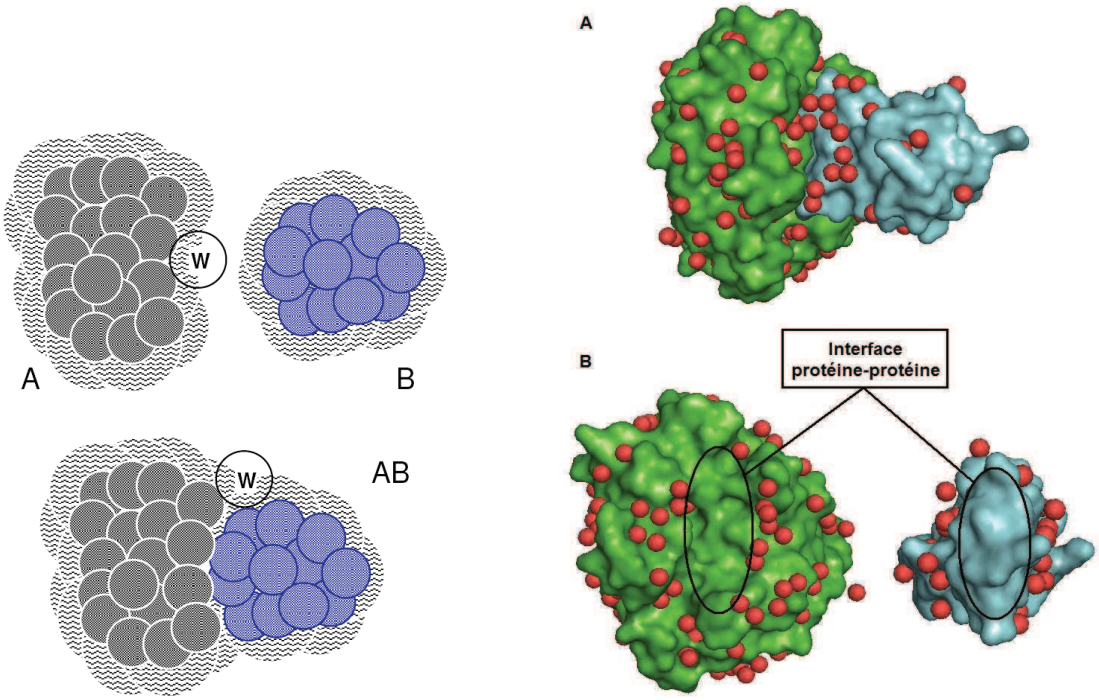


Figure II.1 : Interface et surface accessible au solvant. Gauche : Schéma représentant la surface accessible au solvant (zones hachurées en gris) pour les protéines A et B à l'état libre (haut) et pour les protéines à l'état lié AB (bas). L'interface du complexe correspond à la partie qui n'est plus accessible au solvant (W) après formation du complexe.
Droite : A : Représentation en mode surface de la trypsine (vert) en complexe avec son inhibiteur (cyan) (code PDB 2PTC) montrant la surface accessible au solvant entourée des molécules d'eau (sphères rouges). B : la séparation des monomères montre la surface qui devient enfouie lors de la formation du complexe et qui correspond à l'interface du complexe (entourée par des cercles noirs).

Selon une étude réalisée par Lo Conte et ses collaborateurs sur 75 complexes protéiques non covalents de structures 3D connues (24 complexes protéase-inhibiteur, 19 complexes anticorps-antigène et 32 autres complexes dont 9 enzyme-inhibiteur et 11 impliqués dans la transduction de signal), la plupart des complexes protéiques enfouissent une surface allant de 1200 à 2000 $Å^2$, cette gamme étant considérée comme une taille d'interface standard (70% des interfaces considérées) (Lo Conte et al., 1999). Cependant, il existe parmi ces complexes 3 petites interfaces d'environ 1150 $Å^2$ telles que les complexes CheA-CheY (1140 $Å^2$), cyclophilin-HIV capsid (1170 $Å^2$) et cytochrome peroxydase-cytochrome C (1170 $Å^2$), et 20 larges interfaces d'environ 3300 $Å^2$ telles que les complexes thrombin-hirudin (3350 $Å^2$) ou $G_{t\beta\gamma}$-phosducin (4660 $Å^2$). De manière générale, les interfaces de petite taille sont considérées comme caractéristiques des complexes de basse stabilité et de courte durée de vie, comme c'est le cas pour le complexe cytochrome peroxydase-cytochrome C dont la réaction catalytique est très rapide et ne nécessite donc pas une forte affinité de liaison. Par ailleurs, selon Lo Conte et al., chacune des deux protéines impliquées dans le complexe participe de manière égale dans la taille de l'interface (50±3 %), à l'exception des complexes protéase-inhibiteur qui participent à

la taille de l'interface avec un ratio de 46 et 54 respectivement, dû à une surface convexe de l'inhibiteur se plaçant dans un site actif concave de la protéase.

D'autres auteurs ont reporté des études de complexes protéine-protéine dans lesquels la taille de l'interface, déterminée pour des hétéro-complexes, est définie sur la base du changement de la surface accessible au solvant lors du passage de l'état monomérique à l'état dimérique (Jones et Thornton, 1996). Cette variation correspond donc à la partie d'une protéine qui se trouve restreinte d'accessibilité au solvant consécutivement à la formation du complexe avec son partenaire protéique. La taille moyenne de ces interfaces déterminée pour une série de 27 hétéro-complexes transitoires divers est de 983 $Å^2$. Cependant ces données indiquent également une grande dispersion des tailles d'interface dans cette série de complexes (σ = 582 $Å^2$).

II.2.2 Structure des sites de reconnaissance dans les interfaces

Il est souvent admis que les protéines s'associent par l'intermédiaire de **« patchs » ou sites de reconnaissance** spécifiques sur leurs surfaces. Les interfaces sont en général constituées d'un site de reconnaissance principal de taille standard additionné de sites dont les surfaces sont plus petites, et dans des cas plus rares, le site de reconnaissance principal peut être dupliqué. Une étude réalisée sur 70 complexes protéine-protéine divers (Chakrabarti et Janin, 2002) a montré que la majorité des interfaces (46) possèdent un site de reconnaissance unique, 18 interfaces possèdent 2 sites de reconnaissance et 6 interfaces en possèdent 3 à 4 (Figure II.2).

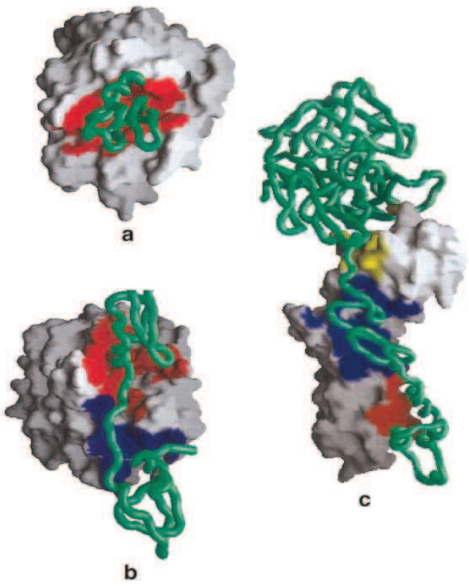


Figure II.2 : « Patchs » de reconnaissance dans les complexes protéase-inhibiteur. Les sites de reconnaissance sont représentés en zones de couleurs différentes sur la surface d'une protéine tandis que la 2ème protéine est représentée en tube. A : Site de reconnaissance unique (rouge) à l'interface entre la trypsine (gris) et PTI (vert) (code PDB 2PTC). B : Ornithodorin (vert) et thrombin (gris) formant une interface avec deux sites de reconnaissance (rouge et bleu) (code PDB 1TOC). C : Facteur VIIA (vert) et facteur tissulaire soluble (gris) formant une interface avec trois sites de reconnaissance (rouge, bleu et jaune) (code PDB 1DAN). (Chakrabarti et Janin, 2002)

Tous les complexes protéase-inhibiteur étudiés ayant une taille d'interface standard possèdent un site de reconnaissance unique tandis que les complexes protéase-inhibiteur de taille plus grande ainsi que ceux impliqués dans la transduction de signal comprennent plusieurs sites de reconnaissance au niveau de leurs surfaces. Ces données indiquent donc une étroite corrélation entre la taille de l'interface et le nombre de sites de reconnaissance. En effet, 43 interfaces parmi les 46 possédant un site de reconnaissance unique ont une taille < 2000 $Å^2$, ce qui correspond à une taille inférieure ou égale à la taille d'interface standard.

Ces auteurs se sont également intéressés à la structure des motifs de reconnaissance. Ceux-ci sont formés d'atomes qui appartiennent à l'interface entre les protéines et dont la surface accessible au solvant (ASA) diminue partiellement ou totalement lors de la formation du complexe. Chaque « patch » est divisé en trois parties :

- les résidus formant le **cœur** du patch (Zone A)
- les résidus entourant le cœur du patch (Zone B)
- les résidus formant le bord ou la **couronne** du patch (Zone C)

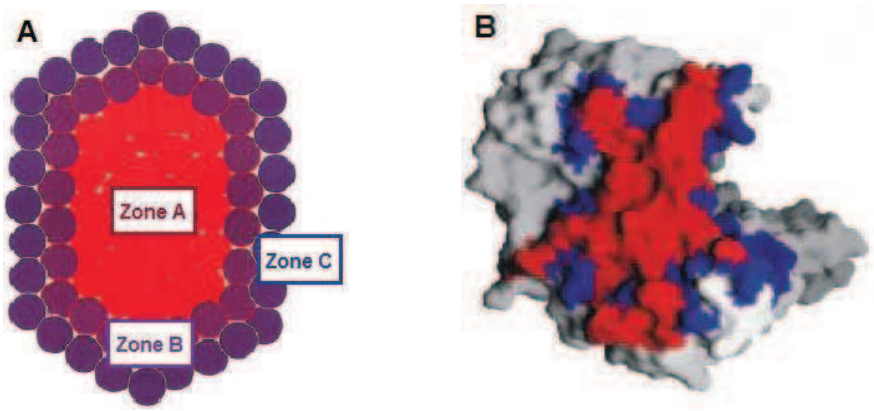


Figure II.3 : Structure des interfaces protéiques. **A** : Représentation schématique d'un patch de reconnaissance au niveau de l'interface montrant la zone A correspondant au cœur et les zones B et C correspondant à la couronne. **B** : Représentation en mode surface de la protéine 2-phospho-D-glycérate hydrolase (code PDB 1EBH) montrant le cœur de l'interface en rouge et la couronne en bleu.

Les atomes formant la zone A sont totalement enfouis dans l'interface (ASA = 0) lors de la formation du complexe tandis que les atomes des zones B et C gardent une accessibilité au solvant partielle, et serviraient à isoler le cœur du solvant. Par ailleurs, les atomes de A et B sont des atomes de contact établissant des interactions de type Van der Waals avec l'autre partenaire du complexe.

Une interface moyenne contient 211 atomes appartenant à 52 résidus, dont la moitié environ sont des atomes de contact (types A et B) et le tiers sont des atomes totalement enfouis (type A), ce qui correspond à environ 13 résidus de contact par protéine dont 9 totalement enfouis (Lo Conte et al., 1999). Les données issues des travaux de Chakrabarti et Janin sont similaires et montrent que toutes les interfaces analysées possèdent au moins un patch de surface, dont la taille moyenne est de 800 $Å^2$ (Chakrabarti et Janin, 2002). Cette interface comporte en moyenne 85 atomes appartenant à 23 résidus dont 12 et 11 forment respectivement le cœur et la couronne du site. La composition de la zone correspondant au cœur du patch est différente de celle correspondant à la couronne ainsi que du reste de la surface de la protéine. Par ailleurs, dans les complexes hétéro-dimériques, les résidus formant le cœur du patch sont plus conservés que ceux formant la couronne. Cette constatation est particulièrement vraie pour les complexes enzyme-inhibiteur dans lesquels les sites actifs des enzymes ont évolué de façon à conserver leurs fonctions spécifiques. En revanche, les résidus formant le cœur et la couronne des interfaces dans les complexes antigène-anticorps sont peu conservés, ce qui s'explique par le fait que les séquences d'anticorps doivent évoluer de façon à se diversifier pour reconnaitre une large variété d'antigènes (Guharoy et Chakrabarti, 2005).

Bogan et Thorn ont proposé un autre modèle décrivant la structure des sites de reconnaissance avec une segmentation cœur/couronne (Bogan et Thorn, 1998). Cette étude fondée sur l'analyse d'une base de données de 2375 mutations en alanine dans des complexes protéine-protéine a conduit à plusieurs conclusions importantes. Il n'existe pas de corrélation entre la surface enfouie par résidu et la variation d'énergie libre de liaison associée à la substitution de ce résidu en alanine. Par ailleurs, la variation d'énergie libre de liaison n'est pas uniformément répartie parmi les résidus formant le site de reconnaissance. Il existe en fait un nombre restreint de résidus pour lesquels la substitution en alanine conduit à une diminution de l'énergie libre de liaison de plus de 2 kcal/mol. Ces résidus correspondent à des « hot spots » au niveau des interfaces et joueraient un rôle essentiel d'un point de vue énergétique dans l'interaction entre les protéines (Voir chapitre précédent). La proposition de ces auteurs est que ces résidus sont généralement enfouis et regroupés au centre des sites de reconnaissance et sont entourés de résidus accessibles qui ont un effet moindre sur l'affinité de liaison (Figure II.4). On voit bien que ce modèle de structure des sites de reconnaissance est très convergent avec celui décrit par Lo Conte et al. ainsi que Chakrabarti et Janin.

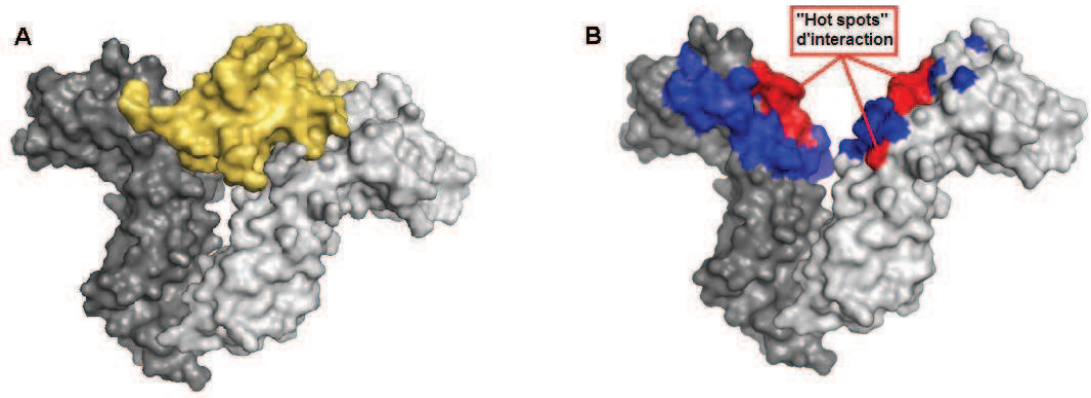


Figure II.4 : Résidus « hot spots » au niveau des interfaces protéine-protéine. A : hormone de croissance humaine (jaune) en complexe avec son récepteur homo-dimérique (gris clair et foncé). B : Interface d'interaction entre les protéines avec en rouge les 6 résidus « hot spots » dont la mutation abolit l'interaction. Les résidus appartenant à la couronne du site de reconnaissance sont représentés en bleu (Ofran et Rost, 2007).

II.2.3 Complémentarité géométrique entre les surfaces

Plusieurs méthodes ont été proposées pour estimer la complémentarité géométrique entre les surfaces en interaction dans un complexe protéine–protéine. **L'indice de vide interfacial** ou « **gap index** » a été proposé pour estimer la complémentarité géométrique de

deux partenaires protéiques impliqués dans un complexe (Jones et Thornton, 1996). Cet indice est défini par la relation suivante :

Gap index (Å) = volume vide entre molécules ($Å^3$) / interface ASA ($Å^2$).

Cet indice est calculé en utilisant la procédure développée par Laskowski (Laskowski, 1991). Plus le « gap index » est petit, plus l'empilement au niveau des interfaces protéine-protéine est compact. Ceci peut conduire à l'optimisation des interactions de Van der Waals et à la limitation de l'accessibilité au solvant. Selon l'étude menée par Jones et Thornton sur une série de complexes protéine-protéine naturels (Jones et Thornton, 1996), les surfaces d'interaction dans les hétéro-complexes sélectionnés pour cette étude se caractérisent par une valeur moyenne du « gap index » de 2,48 Å mais les fluctuations observées sont importantes (σ = 1,02 Å). Ces auteurs ont également montré que la complémentarité était la plus élevée pour les complexes enzyme-inhibiteur (valeur moyenne du « gap index » = 2,20 Å) tandis que les complexes antigène-anticorps sont les moins complémentaires (valeur moyenne du « gap index » = 3,02 Å). Ce dernier résultat peut être expliqué par la présence d'un pourcentage plus élevé en résidus aromatiques au niveau des sites de liaison à l'antigène sur les anticorps (Padlan, 1990) par rapport aux autres interfaces. Ces résidus aromatiques ayant des degrés de liberté conformationnelle moindres par rapport à de plus petits résidus hydrophobes ou polaires, leur prévalence peut conduire à une compacité plus faible au niveau des interfaces. Ces constatations sont également en accord avec les travaux de Lawrence et Colman (Lawrence et Colman, 1993) utilisant le coefficient de corrélation de forme S_c afin d'évaluer la complémentarité géométrique des interfaces protéine-protéine.

La méthode du **polyèdre de Voronoï** constitue une troisième méthode permettant d'évaluer la compacité des interfaces (Richards, 1974). Cette méthode consiste à tracer un polyèdre de Voronoï autour de chaque atome enfoui dans l'interface, à calculer le volume de ce polyèdre pour chaque atome puis à additionner ces volumes. Le volume résultant V est comparé à une valeur de référence V_0 correspondant aux volumes moyens occupés par les atomes de résidus enfouis à l'intérieur de la protéine. Un ratio $V/V_0 > 1$ indique que la compacité au niveau des interfaces est plus faible que celle à l'intérieur des protéines. Lo Conte et ses collaborateurs ont utilisé ce ratio afin de déterminer la compacité des interfaces dans les 75 complexes étudiés et les résultats obtenus ont montré que le ratio V/V_0 est de 1,01 (±0,06)

pour tous les complexes (Lo Conte et al., 1999). Ce résultat montre que les atomes enfouis au niveau des interfaces protéine-protéine sont aussi fortement compactés que ceux au cœur des protéines. Cependant, par cette méthode, aucune différence dans le degré de compacité n'a été observée entre les complexes enzyme-inhibiteur et les complexes anticorps-antigène. Il est à noter également que cette méthode possède un biais car elle ne prend en considération que les atomes totalement enfouis ce qui correspond au 1/3 des atomes de l'interface.

La **planéité** des interfaces dans un ensemble de 70 complexes protéine-protéine a également été étudiée. Chakrabarti et Janin ont quantifié cette caractéristique (Δ) en calculant le RMSD sur la position de tous les atomes présents à l'interface par rapport à un plan des moindres carrés tracé à travers les atomes (Chakrabarti et Janin, 2002). Les résultats ont montré que les interfaces des complexes antigène-anticorps sont plus planes (Δ = 2,5 Å) que la moyenne (Δ = 3 ± 0,6 Å) tandis que les interfaces des complexes protéase-inhibiteur sont moins planes (Δ = 3,5 Å). En effet, dans le cas des complexes protéase-inhibiteur, l'interaction a lieu par le biais de résidus catalytiques qui sont généralement localisés dans des cavités présentes à la surface des enzymes. Ces résultats confirment les observations faites antérieurement sur un ensemble plus restreint d'interfaces protéine-protéine (Jones et Thornton, 1996) où la déviation moyenne des atomes Δ était de 2,80 Å. Toutefois, des fluctuations importantes sont associées à cette valeur (σ = 0,87 Å).

II.2.4 Changements conformationnels

Lors de la formation d'un complexe protéine-protéine, des changements conformationnels au niveau de la structure des protéines peuvent se produire : les protéines se déforment pour s'adapter à leurs partenaires et des mouvements de chaines latérales et/ou de chaine principale peuvent avoir lieu afin de contribuer à la compacité des interfaces et à la formation de liaisons hydrogène. Il est possible d'analyser ces changements conformationnels en comparant les structures cristallographiques des protéines à l'état libre et à l'état lié. Les travaux de Lo Conte et al. ont montré que les protéases, les inhibiteurs, les antigènes et les anticorps associés dans des complexes de taille d'interface standard subissent très peu de modifications conformationnelles (Lo Conte et al., 1999). De faibles modifications conformationnelles peuvent concerner des rotations de chaînes latérales de surface comme c'est

le cas dans le complexe Chymotrypsin-ovomucoid (Fujinaga et al., 1987). Les modifications conformationnelles peuvent également impliquer de faibles mouvements au niveau de la chaîne principale de boucles de surface ou de courts segments de chaine polypeptidique (jusqu'à 1,5 Å). Il existe cependant des exceptions, telles que le complexe trypsinogen-pancreatic trypsin inhibitor (PTI). En effet, dans le trypsinogen à l'état libre, la région formant la poche de spécificité est partiellement désordonnée ; elle devient complètement ordonnée lors de l'association avec le PTI (Bode et al., 1978). Ce changement conformationnel induirait une baisse importante de la stabilité du complexe (diminution du K_d d'un facteur 10^6) par rapport aux complexes formés avec la trypsin qui possèdent une structure entièrement ordonnée, alors que les structures des interfaces sont presque identiques. Contrairement aux complexes de taille d'interface standard subissant très peu de modifications conformationnelles, les complexes dont la taille d'interface est importante peuvent subir d'importants changements conformationnels tels que de larges mouvements au niveau de boucles ou d'hélices α ou des déplacements de domaines entiers. Ce dernier cas est illustré par le facteur d'élongation EFtu de *E. coli* dont les deux sous-unités se déplacent de 18° lors de la formation du complexe EFtu-EFts (Kawashima et al., 1996).

II.3 Composition chimique des interfaces et nature des résidus

L'hydrophobie des sites de reconnaissance est susceptible d'être plus élevée que celle des surfaces qui ne sont pas impliquées dans des interactions hydrophobes. Il a donc été proposé d'évaluer cette hydrophobie en déterminant l'importance relative des différents types acides aminés dans les interfaces de complexes. Toutefois, cette distribution des résidus présents à l'interface doit être comparée à celle des résidus présents sur la surface de la protéine.

Les résidus présents à la surface des protéines peuvent être classés en 3 groupes : les résidus non polaires (contenant des carbones aliphatiques et aromatiques), les résidus polaires neutres (contenant des atomes autres que le carbone, sauf les résidus portant une charge électrique nette) et les résidus chargés. Selon Miller et al., la composition moyenne d'une surface accessible au solvant de petites protéines globulaires est non polaire à 57%, polaire neutre à 24% et chargée à 19% (Miller et al., 1987). Des résultats très similaires ont été obtenus par Lo Conte et ses collaborateurs dans une étude réalisée sur 75 complexes protéiques dans laquelle

ces auteurs ont observé que l'interface moyenne est composée de résidus non polaires à 56%, de résidus neutres polaires à 29% et de résidus chargés à 15% (Lo Conte et al., 1999). Les interfaces des complexes protéase-inhibiteur ont cependant une composition plus hydrophobe que la moyenne (61% de résidus non polaires) et les interfaces des complexes anticorps-antigène ont une composition plus polaire que la moyenne (51% de résidus non polaires). Au total, cette étude a montré que la dispersion autour des valeurs moyennes de chacune des fractions (non-polaire, polaire, chargé) est très importante (Lo Conte et al., 1999).

D'autres études réalisées avec des complexes hétéro-dimériques et homo-dimériques ont montré que les interfaces des complexes hétéro-dimériques étaient plus polaires que les interfaces des complexes homo-dimériques (Jones et Thornton, 1996 ; Nooren et Thornton, 2003). Cette caractéristique reflète probablement la possibilité des protéines impliquées dans les complexes hétéro-dimériques transitoires d'exister sous forme de monomères dans les conditions physiologiques ; une interface relativement polaire est donc énergétiquement plus favorable. En revanche, les protéines impliquées dans les complexes homo-dimériques ne sont stables que sous forme dimérique et leur surface hydrophobe est donc constamment enfouie.

Les 20 acides aminés ne sont pas tous distribués de la même manière au niveau des interfaces et des surfaces totales des complexes (Lo Conte et al., 1999). En effet, les interfaces sont plus riches en résidus aromatiques His, Tyr, Phe et Trp que les surfaces moyennes des protéines (21% contre 8%) et également plus riches en résidus aliphatiques Leu, Ile, Val et Met (17% contre 11%). Ces interfaces possèdent moins de résidus chargés (Asp, Glu et Lys) que la surface exposée au solvant dans le complexe, à l'exception notable du résidu Arg qui représente 10% des résidus présents aux interfaces. L'abondance de ce résidu au niveau des interfaces de protéines pourrait être due au caractère amphiphile de sa chaine latérale qui contribue de manière importante à différents types d'interactions intermoléculaires. Les interfaces des complexes protéase-inhibiteur sont particulièrement riches en Cys et celles des complexes antigène–anticorps sont très riches en résidus Tyr qui contribuent au 1/6 de leurs interfaces.

De façon plus détaillée, la composition des « patchs » de reconnaissance au niveau des interfaces a été déterminée. Une étude réalisée par Chakrabarti et Janin a montré que les atomes totalement enfouis dans l'interface appartenant au cœur du site de reconnaissance (zone A) sont nettement moins polaires (63% de résidus non polaires) que les autres atomes de l'interface qui

ont gardé une accessibilité partielle au solvant (Chakrabarti et Janin, 2002). Le cœur est très riche en résidus aromatiques, notamment Trp et Tyr, et les résidus chargés ainsi que Ser et Thr sont très rares. La couronne semble avoir une composition très pauvre en résidu Phe et plus riche en Asn, Glu et Gly que le cœur. La composition des « hot spots » définis par Bogan et Thorn est en accord avec celle mise en évidence pour le cœur et la couronne des sites de reconnaissance avec une abondance de résidus Trp, Tyr et Arg et une déplétion en résidus Ser et Thr (Bogan et Thorn, 1998). En revanche, les résidus Leu et Val présents en faibles proportions dans le cœur et la couronne sont absents des « hot spots ».

II.4 Propriétés électrostatiques des interfaces

II.4.1 Complémentarité de charge

Les analyses des interactions électrostatiques entre protéines ont longtemps été limitées au dénombrement des résidus chargés et des ponts salins au niveau des interfaces (Janin et Chothia, 1990; Jones et Thornton, 1995). Il était généralement admis que les interfaces protéine-protéine se caractérisaient par une complémentarité de charge qui se manifesterait par la présence d'interactions, appelées ponts salins, établies entre des groupements chargés de signes opposés portés par les chaînes latérales d'acides aminés. Des travaux ont tenté de quantifier cette complémentarité (McCoy et al., 1997). Le calcul de la complémentarité de charge nécessite l'identification de tous les atomes d'une protéine en contact avec des atomes avoisinants de l'autre protéine du complexe. Des charges sont assignées aux atomes en contact et un coefficient de corrélation r entre les listes de charges est calculé. La complémentarité de charge est définie selon McCoy et ses collaborateurs par l'équation : CC = - r. Plus la valeur de CC est positive, plus les charges au niveau de l'interface sont complémentaires et plus cette valeur est négative, plus les charges sont similaires. Deux atomes sont considérés avoisinants sur une interface lorsque la distance qui les sépare est inférieure ou égale à une distance seuil fixée à 3 Å (des distances de 3,5 et 4 Å ont également été testées). Une étude sur 12 interfaces protéine-protéine a conduit à la conclusion qu'aucun des complexes étudiés ne montre de valeur significative de CC pour les 3 distances seuil testées (McCoy et al., 1997). Ainsi, les interfaces protéine-protéine ne montrent pas de complémentarité de charge telle que mesurée

par CC, ce qui abolit l'hypothèse souvent admise dans la littérature que les interactions intermoléculaires ont lieu entre des surfaces avec des charges de surface complémentaires.

II.4.2 Complémentarité électrostatique

La complémentarité électrostatique entre les surfaces a été utilisée pour analyser les interactions entre partenaires protéiques et également dans plusieurs méthodes d'ancrage protéine-protéine. Avec le développement de modèles de continuum électrostatique pour les protéines, il est devenu possible de calculer le potentiel électrostatique des protéines en résolvant numériquement l'équation de Poisson-Boltzmann (Davis et McCammon, 1990). Ce modèle permet de calculer le potentiel électrostatique au voisinage d'une protéine de structure 3D connue en fonction des constantes diélectriques du solvant et du soluté, de la force ionique de la solution, de l'accessibilité des ions à l'intérieur de la protéine et de la distribution de charges de celle-ci. Le potentiel électrostatique résultant est calculé en tous points d'une grille tridimensionnelle contenant la protéine d'intérêt. La projection du potentiel électrostatique ainsi calculé sur la surface de la protéine peut être visualisée à l'aide d'un code de couleurs représentant le signe et l'intensité du potentiel.

L'étude réalisée par McCoy et ses collaborateurs sur 11 complexes protéiques ne montrant pas de complémentarité de charges significative a permis de mettre en évidence une complémentarité des potentiels électrostatiques des interfaces étudiées qui sont complémentaires et anti-corrélés (McCoy et al., 1997). Un exemple de complémentarité électrostatique à l'interface entre deux partenaires d'un complexe hétéro-dimérique est illustré dans la figure II.5 pour le complexe IL-4/IL-4Rα.

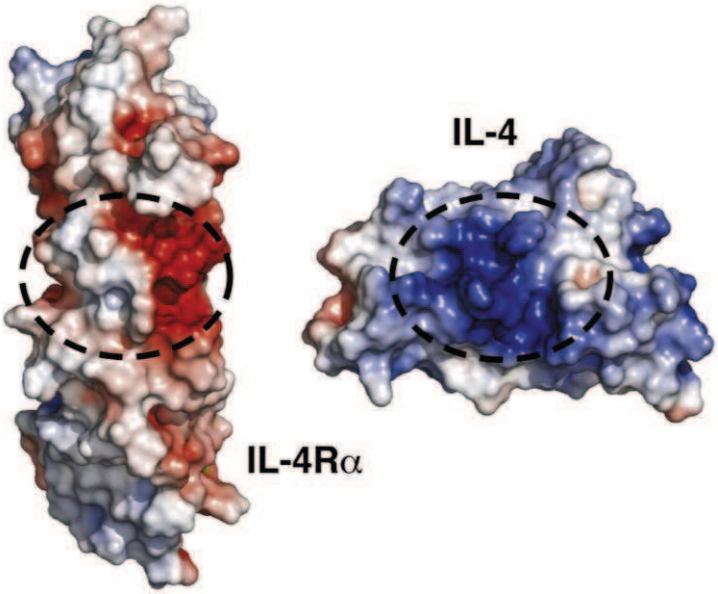


Figure II.5 : Potentiel électrostatique de surface de l'interleukine IL-4 et de son récepteur IL-4Ra (code PDB 1IAR). Les zones de potentiel électrostatique positif sont représentées en bleu et les zones de potentiel électrostatique négatif en rouge. Les zones de contact entre l'interleukine et son récepteur sont indiquées par un cercle et montrent la complémentarité électrostatique entre les interfaces.

Ces résultats contredisent donc ceux de l'étude de Slagle et al. qui suggérait que la complémentarité électrostatique au niveau des interfaces protéine-protéine était liée à la présence et au nombre de ponts salins à l'interface (Slagle et al., 1994). En réalité, les charges non impliquées dans des ponts salins contribuent également de manière significative à la complémentarité électrostatique à l'interface entre les protéines.

Les résultats ci-dessus montrent bien que le potentiel électrostatique calculé à l'aide de l'équation de Poisson-Boltzmann permet une meilleure compréhension des aspects électrostatiques des interactions protéine-protéine. Cette méthode a d'ailleurs permis de mettre en évidence des similarités des propriétés de reconnaissance au sein de différentes sous-familles de protéines (De Rienzo et al., 2000).

II.5 Phénomène d'association

II.5.1 Aspects thermodynamiques

Le phénomène d'association d'un complexe protéine-protéine est caractérisé par sa constante d'équilibre K ou constante de dissociation (unité mol/L). Cette constante est associée

à la variation d'énergie libre standard d'association appelée également énergie libre de Gibbs ΔG° (unité kcal/mol) :

$$\Delta G° = - RT \ln K$$

ΔG° correspond à la variation de la fonction G entre l'état initial du système (les partenaires protéiques à l'état libre) et l'état final (le complexe des deux partenaires formé) dans les conditions standard. R est la constante universelle des gaz parfaits égale à 1,987 $cal \cdot K^{-1} \cdot mol^{-1}$ et T est la température en degrés Kelvin**.**

Pour un complexe entre une protéine et son ligand, l'affinité s'exprime par la valeur de la constante de dissociation K_d : plus cette valeur est faible, plus l'affinité du complexe est élevée.

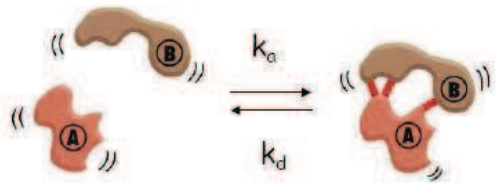


Figure II.6 : Représentation schématique de la formation d'un complexe protéine (A)-ligand (B) avec les constantes d'association K_a et de dissociation K_d.

La variation d'énergie libre d'association correspond à la somme de deux contributions ΔH et ΔS, et est définie par l'équation fondamentale :

$$\Delta G = \Delta H - T \Delta S$$

Où ΔH = variation d'**enthalpie** au cours de la transformation d'association. Elle est notamment associée à la formation de liaisons entre les composantes du système (partenaires protéiques et molécules de solvant mais également aux déformations géométriques nécessaires à la formation du complexe).

Et ΔS = variation d'**entropie** qui caractérise la variation d'ordre du système incluant les deux partenaires et le solvant au cours de la transformation**.**

La désolvatation des surfaces protéiques lors de la formation du complexe contribue à la composante ΔS. L'enfouissement de la surface accessible au solvant lors de la formation d'un complexe implique l'expulsion des molécules d'eau ; les interfaces protéine-protéine étant riches en résidus aromatiques et en résidus aliphatiques, la désolvatation des interfaces contribue de manière positive à l'entropie totale de l'interaction. Deux autres paramètres

contribuent à la variation de l'entropie : la réduction des degrés de liberté de rotation autour d'angles de torsion de la protéine ainsi que la perte des degrés de liberté de rotation et de translation lors de la formation du complexe.

La composante ΔH inclut les modifications des termes d'énergie de liaisons non covalentes qui se produisent entre les partenaires lors de l'association entre les protéines du complexe. Ces interactions intermoléculaires sont représentées à l'aide d'une série de termes incluant les liaisons hydrogène, les interactions de Van der Waals, les interactions électrostatiques, les interactions cation-π etc.

II.5.2 Interactions protéine-protéine

- Interactions de Van der Waals

Les interactions de Van der Waals correspondent à des interactions de faible intensité entre les atomes des interfaces de protéines (énergie entre 0,1 et 1 kJ/mol par paire d'atome) et dont l'énergie décroît de façon importante en fonction de la distance. Ces interactions sont néanmoins d'une importance considérable vu leur nombre important. Le terme de Van der Waals est en général calculé à l'aide d'un potentiel de Lennard-Jones appelé potentiel 6-12. Ce potentiel prend en compte deux effets principaux : les forces répulsives de Pauli, d'origine purement quantique, dominantes à courte distance et qui empêchent l'interpénétration mutuelle des nuages électroniques de deux atomes ; et les forces dispersives de London, dominantes à grande distance, et qui sont attractives et dues à l'interaction de dipôles instantanés entre eux. Les forces répulsives impliquent que l'énergie devient très défavorable à mesure que les atomes se rapprochent les uns des autres.

- Liaisons hydrogène

Les liaisons hydrogène jouent également un rôle important dans la stabilisation des complexes protéine-protéine. La liaison hydrogène peut être vue comme un cas particulier d'interaction de Van der Waals ; c'est une liaison non covalente, de type dipôle-dipôle qui résulte de l'interaction entre un atome d'hydrogène (H), lié à un atome électronégatif (O, N, S)

(donneur) et un deuxième atome électronégatif possédant une paire d'électrons non partagée (accepteur). Bien que de la même famille que les interactions de Van der Waals, les liaisons hydrogène présentent une intensité environ dix fois supérieure à celle de l'interaction de Van der Waals (énergie comprise entre 2 et 8 kcal/mol).

Une étude menée sur 319 interfaces protéiques montre que les interfaces contiennent en moyenne 10,69 liaisons hydrogène (Xu et al., 1997). Ce résultat est en accord avec une précédente analyse basée sur un ensemble plus réduit d'interfaces (15 complexes protéase-inhibiteur et 4 complexes anticorps-antigène) montrant que le nombre de liaisons hydrogène est compris entre 8 et 13 au niveau des interfaces étudiées, avec une moyenne de 10 (Janin et Chothia, 1990). Cependant, les fluctuations associées à ce nombre moyen de liaisons hydrogène par interface sont très élevées (12,35). La corrélation mise en évidence entre la taille de l'interface et le nombre de liaisons hydrogène intermoléculaires est à l'origine de l'utilisation de la densité superficielle de liaisons hydrogène pour comparer les interfaces protéine-protéine.
L'étude réalisée par Xu et al. a également montré que les liaisons hydrogène au niveau des interfaces se font, par ordre de préférence, entre chaines latérales puis entre chaine principale et chaine latérale et enfin entre chaines principales des deux partenaires protéiques (39,3%, 35,4% et 22,6% respectivement) (Xu et al., 1997). L'ordre de fréquence des liaisons hydrogène par type d'atome impliqué est O-N puis O-O et enfin N-N (83,5%, 15,2% et 0,7% respectivement).
Cette étude a également permis de décrire les caractéristiques géométriques des liaisons hydrogène (Xu et al., 1997). Elles sont caractérisées par l'existence d'une distance optimale entre les groupes donneurs et accepteurs estimée à 2,92 Å (la distance maximale observée étant de 3,76 Å et la distance minimale de 1,83 Å), mais aussi d'une orientation optimale des vecteurs de liaison (angle moyen entre le donneur, l'atome d'hydrogène et l'accepteur de 150,7°).

- Ponts salins

On appelle pont salin l'interaction entre un groupement chargé sur une protéine et un groupement de charge opposée d'une autre protéine. L'intensité de ces interactions de nature électrostatique dépend de l'environnement local (type de solvant, constante diélectrique du milieu et pH). Les interactions de type pont salin sont fortes : l'énergie d'une paire d'ions

typique, comme le groupement carboxylate de Glu et le groupement ammonium de Lys dont les centres de charge sont séparés de 4 Å dans un milieu de constante 4, est égale à 86 kJ/mol (énergie plus forte qu'une liaison hydrogène).

D'après l'étude réalisée par Xu et al., il y a environ 2 ponts salins par interface pour une distance ≤ 4 Å entre les atomes (Xu et al., 1997). Sur les 54 interfaces correspondant à des structures de complexes à haute résolution, la moitié ne contient aucun pont salin et deux interfaces en contiennent 10 et 12. La formation des ponts salins se fait préférentiellement entre résidus Arg-Asp ou Arg-Glu.

Cependant, contrairement aux interactions de Van der Waals qui sont faibles mais qui stabilisent significativement les complexes, les interactions de type pont salin ne stabilisent pas fortement les interfaces protéiques car les ions en milieu aqueux sont fortement solvatés et la formation d'un pont salin implique une énergie de désolvatation des chaines latérales chargées impliquées dans le pont salin. Cette contribution compense en grande partie l'énergie de nature électrostatique correspondant à l'interaction directe entre les groupements chargés. Cette constatation est étayée par les résultats obtenus dans une étude de 3 complexes bien caractérisés : barnase-barstar, hormone de croissance humaine-récepteur et neuraminidase N9-anticorps NC41 pour lesquels l'effet électrostatique global est énergétiquement faible dans le 1er complexe et défavorable dans les deux derniers (Sheinerman et al., 2000).

- Interactions cation-π

L'interaction cation-π est une interaction non covalente entre un résidu aromatique (quadrupole) et un cation adjacent (monopole).

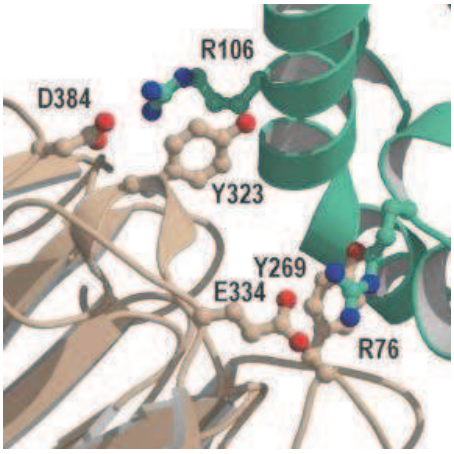


Figure II.7 : Vue représentant l'interface du complexe Ran-RCC1 et montrant des exemples d'interactions cation-π entre des résidus Arg et Tyr représentés en bâtons et sphères (R106-Y323 et R76-Y269). Les résidus Arg forment en plus des ponts salins entre leur groupement guanidinium et les chaines latérales de D384 et E334.

Une étude portant sur l'identification d'interactions cation-π réalisée sur 81 complexes de protéines a montré que la moitié des interfaces contient au moins une interaction cation-π, avec une préférence pour les interactions entre résidus Arg-Trp et Arg-Tyr qui s'explique par la fréquence élevée de ces résidus au niveau des interfaces protéiques (Crowley et Golovin, 2005).

II.6 Etude des interfaces de familles de complexes formés par une cible commune et ses différents ligands

Les études précédemment décrites ont permis de déterminer des valeurs moyennes de paramètres caractéristiques des interactions protéine-protéine dans des ensembles de complexes naturels composés de partenaires différents. Ces études visaient à mettre au point des méthodes de localisation des sites de liaisons via l'analyse des surfaces de protéines. Un autre objectif important de ces travaux était de déterminer les critères les plus pertinents pour mettre au point des méthodes d'ancrage protéine-protéine plus efficaces. Cependant, une analyse des interfaces formées par une cible commune en complexe avec différents partenaires interagissant à un même site de liaison pourrait constituer un cadre de référence pour estimer les fluctuations autorisées pour les différents paramètres relatifs à un site unique. Dans le contexte de la conception *in silico* de nouveaux ligands protéiques, les fluctuations limites observées

pourraient servir à définir des seuils de rejet des candidats ligands. Quelques analyses de familles de complexes impliquant des sites de liaison similaires, chacun d'eux étant en interaction avec différents partenaires, ont été reportées (Keskin et Nussinov, 2007 ; Tyagi et al., 2009 ; Nobeli et al., 2009). Cependant, aucune de ces études n'a déterminé les valeurs des paramètres caractéristiques des interactions protéine-protéine décrites et leurs possibles fluctuations parmi une série de ligands protéiques interagissant à un même site de liaison. Une analyse détaillée de ces paramètres et des interactions intermoléculaires des interfaces de plusieurs complexes ayant une cible protéique commune avait donc été réalisée au sein de notre équipe (Magis et al., soumis). Cette étude a été menée sur 3 ensembles de complexes protéine-protéine naturels correspondant à 9 complexes de trypsine bovine (groupe 1), 9 complexes de trypsine porcine (groupe 2) et 8 complexes de chymotrypsine bovine (groupe 3) (Figure II.8 ; Tableau II.1).

Tableau II.1 : Complexes protéine-protéine impliquant une cible commune et une série de ligand protéiques interagissant au même site de liaison.

Code PDB	Nom du Ligand	Resolution (Å)	Nb AA Taille
Ensemble 1: Trypsine Bovine			
1C9T	Bdellastasin	3.30	53/59
1D6R	Bowman Birk Inhibitor	2.30	57/58
1G9I	Mung Bean Inhibitor	2.20	22/22
1OX1	Synthetic Peptide Inhibitor	2.00	10/11
1SFI	Cyclic Peptide Inhibitor	1.65	14/14
1TAW	APPI	1.80	54/58
1TGS	PSTI	1.80	56/56
2BTC	CPTI-II	1.50	29/29
2PTC	BPTI	1.90	58/58
Ensemble 2: Trypsine Porcine			
1AVX	STI	1.90	172/177
1EJA	Bdellastasin	2.70	53/59
1H9H	EETI-II	1.50	30/36
1LDT	LDTI	1.90	46/46
1MCT	MCTI-A	1.60	28/28
1TFX	TFPI	2.60	58/58
1TX6	BBI	2.20	121
1UHB	Autolysate peptide	2.15	9/9
1YF4	Vasopressin	1.98	10/10
Ensemble 3: Chymotrypsine			
1ACB	Eglin C	2.00	63/70
1CA0	APPI	2.10	54/54
1CBW	BPTI	2.60	58/58
1CGI	PSTI-3	2.30	56/56
1CHO	OMTKY-3	1.80	53/53
1GL0	LCMI-I	3.00	33/35
1GL1	LCMI-II	2.10	34/36
1N8O	Ecotin	2.00	138/142

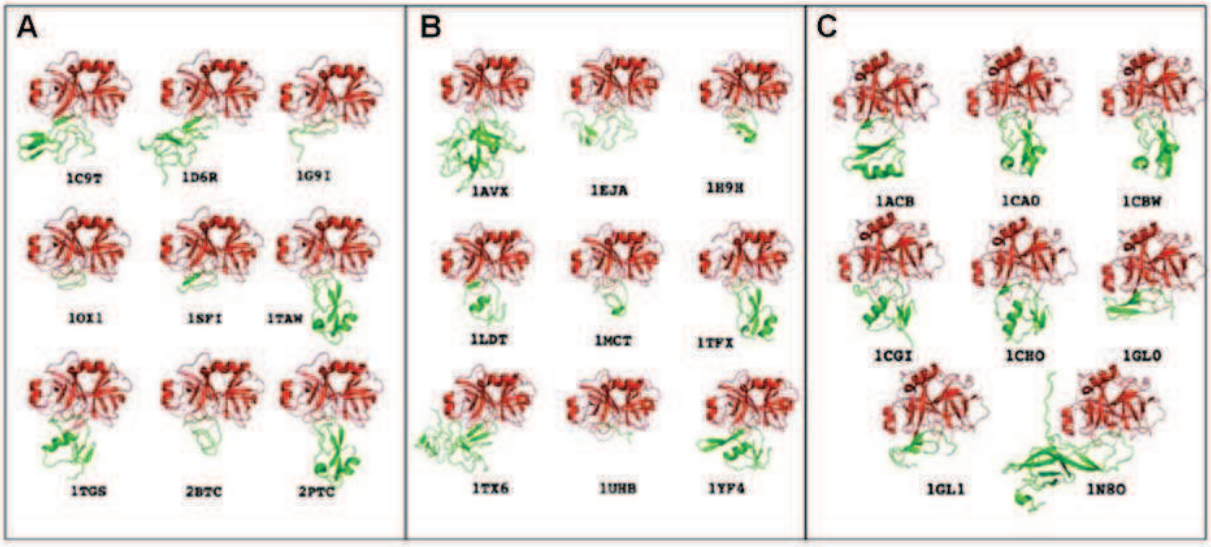


Figure II.9 : Représentation des 3 ensembles de complexes protéine-protéine naturels analysés. Les protéines sont représentées en mode cartoon en rouge pour les 3 cibles protéiques et en vert pour les différents ligands de chaque cible. A : groupe de la trypsine bovine ; B : groupe de la trypsine porcine ; C : groupe de la chymotrypsine bovine.

Ainsi, chacune des 3 cibles protéiques possède entre 8 et 9 ligands naturels différents du point de vue de leurs structures 3D et/ou de leurs séquences primaires en acides aminés. Les différents ligands d'un même ensemble interagissent tous au niveau du même site de la cible commune. Plusieurs paramètres caractérisant l'interface des complexes ont été déterminés : similarité de forme et similarité électrostatique entre les différents ligands d'un même ensemble, la complémentarité de charge entre le ligand et la cible, la densité de surface des liaisons hydrogène, les interactions de type ponts salins, cation-π et π-π entre les ligands et leur cible.

II.6.1 Comparaison des formes

Une comparaison de la forme des différents ligands appartenant à chacun des 3 ensembles a été réalisée en calculant un indice de similarité de forme de Hodgkin (SSI) pour chaque paire de ligands dans un même ensemble (Tableau II.2). Cet indice a permis de comparer la forme globale d'une protéine par rapport à une autre (SSI_all) ou de comparer ces deux protéines au niveau d'une région spécifique correspondant à la région de liaison avec la cible commune (SSI_part). Les résultats ont montré que les protéines ont des formes plus similaires au niveau de la région d'interaction avec la cible que pour ce qui concerne leur forme globale. Cependant, des différences de formes existent malgré tout au niveau de la région

d'interaction, mises en évidence par l'écart important existant entre les valeurs de SSI_part, en particulier au sein du 3ème groupe (SSI_part entre 0,68 et 0,93 pour le groupe trypsine bovine, entre 0,46 et 0,81 pour le groupe trypsine porcine et entre 0,32 et 0,79 pour le groupe chymotrypsine bovine).

Tableau II.2 : Comparaison des formes et des potentiels électrostatiques des différents ligands protéiques au sein des trois familles de complexes protéine-protéine analysées. SSI : Indice de similarité de forme de Hodgkin; ESI : Indice de similarité des potentiels électrostatiques de Hodgkin. _all : pas de restriction spatiale lors de la comparaison ; _part : la comparaison est restreinte à la région d'interaction avec la cible. Seules les valeurs extrêmes sont reportées.

	Ensemble 1 **Trypsine Bovine**	**Ensemble 2** **Trypsine Porcine**	**Ensemble 3** **Chymotrypsine**
	Similarité de forme des ligands protéiques		
SSI_all min /max	+0.24 / +0.77	+0.14 / +0.71	+0.14 / +0.17
SSI_part min /max	+0.68 / +0.93	+0.46 / + 0.81	+0.32/+0.79
	Similarité de forme des cibles protéiques		
SSI_all min /max	+0.80 / +0.89	+0.79 / +0.86	+ 0.75 / +0.83
SSI_part min /max	+0.85 / +0.96	+0.79 / +0.94	+0.73 / +0.90
	Similarité électrostatique des ligands protéiques		
ESI_all min /max	-0.14 / +0.89	-0.48 / +0.78	-0.80 / +0.78
ESI_part min /max	+0.10 / +0.95	-0.29 / +0.91	-0.18 / +0.89
	Similarité électrostatique des cibles protéiques		
ESI_all min /max	+0.87 / + 0.96	+0.76 / +0.92	+0.87 / +0.95
ESI_part min /max	+0.36 / +0.99	+0.51 / + 0.99	+0.84 / +0.99

II.6.2 Comparaison des potentiels électrostatiques

Les potentiels électrostatiques ont également été comparés entre les différents ligands d'un même ensemble (Tableau II.2). Le calcul du potentiel électrostatique a été réalisé à partir d'une grille 3D englobant la protéine entière grâce au logiciel APBS, en résolvant l'équation de Poisson-Boltzmann. L'examen graphique des potentiels électrostatiques de surface des différents ligands d'une même cible a montré que les potentiels étaient similaires au niveau de

la région impliquée dans l'interaction avec la cible, et différents sur le reste de la protéine. Une seule exception a été observée sur un des complexes appartenant au groupe 3, dans lequel le ligand a montré un potentiel électrostatique de surface de signe opposé au niveau de la région de liaison à la cible. La comparaison quantitative des potentiels électrostatiques d'une paire de protéines a été réalisée grâce au logiciel PIPSA en déterminant un indice de similarité électrostatique de Hodgkin (ESI) calculé sur l'ensemble des grilles 3D des deux protéines (ESI_all) ou sur une région bien définie de leur grille, correspondant à la région d'interaction avec leur cible commune (ESI_part). Dans tous les cas étudiés, la similarité électrostatique entre deux protéines est plus élevée au niveau de la région d'interaction avec la cible par rapport à la protéine totale. En effet, la moyenne des valeurs de ESI_all pour les 3 groupes est de 0,43 (groupe 1), 0,21 (groupe 2) et 0,11 (groupe 3) alors que la moyenne des valeurs de ESI_part est de 0,62, 0,42 et 0,37 pour les groupes 1, 2 et 3 respectivement. Cependant, l'écart entre la valeur minimale et la valeur maximale de ESI_part au sein de chaque groupe est important, et certaines valeurs sont négatives. En effet, l'indice ESI_part varie entre +0,10 et +0,95 dans le 1er groupe, entre -0,29 et +0,91 dans le 2ème groupe et entre -0,18 et +0,89 dans le 3ème groupe. Ce résultat suggère que l'indice ESI_part peut être négatif même dans le cas de partenaires protéiques naturels interagissant au même site de liaison d'une protéine cible commune.

II.6.3 Taille et composition des interfaces

La taille totale de l'interface des complexes appartenant aux 3 groupes a été calculée : elle varie de 1265 à 1863 Å^2 au sein du 1er groupe, de 1264 à 2400 Å^2 au sein du 2ème groupe et de 1399 à 2257 Å^2 au sein du 3ème groupe. La dispersion des différentes valeurs au sein de chaque groupe est donc élevée, avec en particulier un facteur proche de 2 entre la plus petite et la plus grande interface dans le 2ème groupe. Il en est de même concernant la taille des interfaces des ligands. Le nombre de résidus à l'interface du ligand est compris entre 10 et 19, 9 et 30, 13 et 24 pour les groupes 1, 2 et 3 respectivement, avec en particulier un facteur 3,3 entre le nombre minimal et le nombre maximal de résidus présents à l'interface des ligands dans le 2ème groupe. La composition chimique des interfaces a montré une grande variabilité entre les différents ligands d'une même cible. En effet, le pourcentage de résidus non polaires à l'interface du ligand est compris entre 29 et 62,5 pour les ligands du 1er groupe, entre 22,2 et 64,5 pour ceux du 2ème groupe et entre 41,7 et 60 pour le 3ème groupe. Ces résultats démontrent

l'absence de conservation de la fraction non polaire au niveau de l'interface des ligands dans chacune de ces familles.

II.6.4 Complémentarité de charge

La charge totale déterminée au niveau de l'interface d'un ligand est considérée comme pouvant jouer un rôle important dans l'interaction avec la cible. Dans cette étude, la comparaison entre les différents ligands d'une même cible a permis de montrer que la charge totale au niveau de l'interface variait considérablement au sein d'un même groupe : elle varie de 0 à +5 dans le 1er groupe, de -1 à +4 dans le 2ème et de -2 à +4 dans le 3ème. Ce résultat suggère donc que l'interaction entre les différents ligands avec la cible commune ne se fait pas de la même manière, chaque complexe étant caractérisé par une charge totale de l'interface bien spécifique. De plus, la complémentarité de charge souvent évaluée dans la caractérisation des interfaces protéine-protéine ne semble pas déterminante dans l'association et l'activité du complexe. Par ailleurs, le nombre de groupements portant une charge nette (+1 ou -1) au niveau de l'interface est également très variable avec des valeurs allant de 1 à 6 pour le groupe 1, de 2 à 9 pour le groupe 2 et de 1 à 6 pour le groupe 3. Ces groupements chargés ne participent pas tous à des interactions intermoléculaires de type ponts salins ou cation-π. En effet, très peu de groupements chargés sont impliqués dans ces interactions dans les complexes du 1er groupe, avec 1 seul pont salin dans chacun des complexes et entre 0 à 2 interactions cation-π, pour un nombre total de groupements chargés compris entre 1 et 6. De même, entre 1 et 3 ponts salins et entre 0 à 2 interactions cation-π ont été identifiés dans les interfaces des complexes du 2ème groupe alors que les ligands sont caractérisés par la présence à leur interface d'un nombre important de groupements chargés compris entre 2 et 9. Dans le groupe 3, 2 complexes ne possèdent ni pont salin ni interaction de type cation-π, 3 complexes possèdent uniquement 1 interaction cation-π tandis qu'un complexe possède 2 ponts salins et 3 interactions cation-π.

II.6.5 Liaisons hydrogène

Une analyse des liaisons hydrogène a été réalisée au niveau des interfaces des complexes des 3 groupes. Dans les groupes 1 et 2, le nombre de liaisons hydrogène varie entre 7 et 11, tandis qu'il varie entre 6 et 9 pour les complexes du groupe 3. La densité de liaisons

hydrogène (le nombre de HB/l'aire d'interface du ligand $Å^2$) est comprise entre 8,0 et 12,6 10^{-3} $Å^{-2}$ pour les complexes du groupe 1, entre 7,7 et 11,1 10^{-3} $Å^{-2}$ pour les complexes du groupe 2 et entre 5,8 et 10,2 10^{-3} $Å^{-2}$ pour les complexes du groupe 3. Ainsi, les interfaces des différents ligands d'une cible commune sont également caractérisées par une grande variabilité du point de vue de leur densité superficielle en liaisons hydrogène. L'analyse détaillée de l'identité des résidus de la cible protéique impliqués dans les interactions intermoléculaires a montré que certaines interactions sont systématiquement conservées dans les différents complexes d'un même groupe, et ce dans les 3 groupes étudiés. Cependant, au sein d'un même groupe, la majorité des liaisons hydrogène établies à l'interface de la cible impliquent des résidus différents d'un complexe à un autre.

De la même manière que pour les groupements chargés qui sont présents à l'interface des ligands mais qui ne sont pas tous impliqués dans des interactions intermoléculaires, seule une faible proportion des donneurs et accepteurs de liaisons hydrogène présents à l'interface de la cible est impliquée dans des liaisons hydrogène. Ces donneurs et accepteurs de liaisons hydrogène présents à l'interface ont été définis comme étant les atomes montrant une variation de surface accessible lors de la liaison supérieure à 0,1 $Å^2$. La proportion des donneurs et accepteurs de liaisons hydrogène présents à l'interface de la cible et impliqués dans des interactions intermoléculaires ou intramoléculaires est comprise entre 30 et 60% pour les 3 groupes étudiés.

II.7 Conclusion

Cette étude, bien que fondée sur un faible nombre de séries de complexes et des protéines cibles voisines, permet de tirer des enseignements utiles pour la conception de ligands protéiques artificiels. Cette étude permet également de définir, pour les paramètres caractéristiques des interactions protéine-protéine, des seuils en deçà desquels il n'est pas possible de rejeter les candidats ligands issus de la conception *in silico*.

Les trois ensembles de complexes analysés fournissent des exemples de la disparité structurale des ligands protéiques naturels interagissant à un site de liaison unique. Cette disparité s'accompagne d'une similarité de forme au niveau de la région d'interaction avec la cible, malgré des structures 3D globales différentes. La similarité de forme entraine une bonne complémentarité des surfaces des différents ligands avec la cible commune au niveau du site de

liaison. Par ailleurs, cette étude montre que, au sein de chaque série, les interactions intermoléculaires observées au site de liaison de la cible se répartissent en deux groupes. Une partie d'entre elles est conservée pour les différents ligands tandis qu'une autre partie est spécifique de chaque ligand au sein de la série. On peut noter que le groupe des interactions conservées pourrait être en relation avec la notion de motif fonctionnel ou « hot spot ». Au total, ces observations illustrent, dans le cas de complexes protéine-protéine naturels la validité de l'hypothèse fondant la méthode de conception *in silico* utilisée au cours de ce travail : des structures 3D différentes peuvent conduire à la reproduction d'interactions similaires et conduire à des interactants protéiques ciblant une région donnée d'une cible protéique. La diversité des structures tridimensionnelles compatibles avec l'interaction de ligands protéiques à un site de liaison donné peut être réalisée grâce à une importante plasticité des interactions intermoléculaires productives conduisant au phénomène d'association.

Concernant les critères généraux caractérisant les interfaces protéine-protéine, nous avons également observé, dans ces trois séries, une variabilité importante entre les différents ligands de la cible commune. Cette variabilité concerne la charge nette et le nombre de groupements chargés présents à l'interface des différents ligands, suggérant que la complémentarité de charge entre les différents ligands et leur cible n'est pas une condition nécessaire pour l'association des deux partenaires. Au sein de chaque série, les potentiels électrostatiques des différents ligands dans la région d'interaction avec la cible possèdent une similitude traduite par les valeurs positives de l'indice ESI_part. Cependant, des fluctuations importantes et quelques exceptions à cette règle ont été notées concernant des paires de ligands caractérisées par des ESI_part < 0. Cette dernière observation est à mettre en relation avec des travaux précédents montrant que deux ligands d'une même cible peuvent posséder une complémentarité électrostatique (caractérisée par les indices de similarité de Hodgkin) avec leur cible commune sans pour autant avoir une similarité électrostatique entre eux (McCoy et al., 1997). Ainsi, dans un contexte de conception *in silico* de ligands protéiques d'une cible d'intérêt, les ligands caractérisés par une faible similarité électrostatique ne doivent pas être rejetés. Par ailleurs, l'analyse des interactions intermoléculaires au niveau des interfaces des complexes des 3 groupes a également mis en évidence des fluctuations du nombre de liaisons hydrogène, de ponts salins, d'interactions cation-π et π-π au sein d'un même groupe. De plus, au sein d'un même groupe, chaque ligand est caractérisé par la présence de liaisons hydrogène spécifiques, établies par le biais de donneurs et d'accepteurs au niveau de l'interface de la cible,

différents d'un complexe à un autre. Cette constatation est à mettre en relation avec l'observation faite dans cette étude que seuls 30% des donneurs et accepteurs à l'interface de la cible sont impliqués dans des liaisons hydrogène, permettant ainsi une certaine adaptabilité de l'interface à des ligands protéiques différents.

CHAPITRE III : Conception de ligands protéiques *in silico* par transfert de sites actifs

Un des défis importants dans la conception de ligands protéiques *in silico* concerne l'introduction de nouvelles fonctions et propriétés dans des plates-formes protéiques existantes. Les protéines réalisent leurs fonctions essentiellement par l'intermédiaire de sites actifs, tandis que le reste de la protéine sert d'architecture pour supporter ce site actif. Une des stratégies appliquées dans la conception de ligands protéiques fonctionnels est le transfert ou greffage de sites actifs sur une protéine hôte non homologue appelée plate-forme. Les sites actifs peuvent être des sites catalytiques dans des enzymes ou des sites de liaison dans des interactions protéine-protéine. Ces sites actifs sont généralement constitués d'un nombre restreint de résidus essentiels et indispensables dans la fonction assurée. Ainsi, dans la chymotrypsine, seuls 3 résidus (Ser-His-Asp) sont essentiels dans la fonction de catalyse. Dans les interactions protéine-protéine, un faible nombre de résidus appelés « hot spots » jouent un rôle prépondérant dans l'énergétique de l'association. Tel est le cas pour l'érythropoïétine humaine (Epo) dans laquelle 3 résidus clé ont été identifiés comme étant responsables de la liaison avec le récepteur EpoR. Plusieurs exemples de travaux portant sur le transfert de résidus « hot spots » sur des plates-formes protéiques afin de concevoir de nouveaux ligands ont été décrits dans le Chapitre I de cette partie. Dans ce chapitre, nous allons décrire la méthode de conception *in silico* de ligands protéiques par transfert de motif fonctionnel, utilisée au cours de ce travail et qui a été développée au sein de notre laboratoire. Cette approche sera comparée aux autres méthodes décrites dans la littérature.

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

L'approche de conception de ligands protéiques développée au sein de notre groupe et utilisée dans ce travail de recherche repose sur le transfert d'un motif fonctionnel, identifié sur un ligand de référence, sur une protéine hôte non homologue, et ce de manière indépendante de la structure secondaire.

III.1.1 Motif fonctionnel

Le motif fonctionnel consiste en un ensemble restreint de résidus. Ces résidus peuvent être sélectionnés sur la base de données expérimentales de mutagénèse ou de substitution en alanine. Ces données permettent d'identifier des résidus énergétiquement prépondérants dans la liaison. Ainsi, ces résidus désignés par le terme « hot spots » par Clackson et Wells ont été définis dans plusieurs travaux comme étant ceux dont la mutation en résidus Ala (« alanine scanning ») entraine une variation de l'énergie libre de liaison supérieure à 2 kcal/mol (Clackson et Wells, 1995 ; Bogan et Thorn, 1998). Par ailleurs, les résidus « hot spots » ont tendance de façon générale à se regrouper au centre de l'interface protéine-protéine et sont entourés de résidus énergétiquement moins importants formant une barrière ou « O-ring » permettant d'isoler les « hot spots » du solvant (Bogan et Thorn, 1998). Les interactions établies par ces résidus « hot spots » concernent des liaisons hydrogène, des interactions de type Van der Waals participant à la compacité de l'interface ainsi que des interactions électrostatiques. Les motifs fonctionnels sont parfois identifiés sur la seule base de données structurales mettant en évidence les interactions intermoléculaires à reproduire ou sur la base de données de simulation. Le motif peut être continu (les résidus se suivent dans la séquence primaire de la protéine) ou discontinu (les résidus sont éloignés les uns des autres dans la séquence primaire).

III.1.2 Système de référence

La conception de ligands protéiques *in silico* a pour objectif le développement de nouveaux ligands protéiques destinés à inhiber ou bloquer la fonction d'une cible d'intérêt pour des applications thérapeutiques, diagnostiques ou biotechnologiques. C'est une approche mimétique dont l'objectif est de reproduire des interactions identifiées par des techniques expérimentales ou dans certains cas par des méthodes de simulation. Toutefois, cette approche nécessite des données structurales caractérisant l'interaction du motif fonctionnel avec la cible protéique. Le point de départ de l'approche de transfert de motif de liaison est donc un complexe constitué de la cible d'intérêt en interaction avec son ligand protéique naturel, pour lequel la structure cristallographique a été décrite.

III.1.3 Protéine hôte

L'étape suivante consiste à rechercher une protéine hôte appelée également plate-forme possédant des résidus capables de reproduire la topologie du motif de référence. Les plates-formes sont identifiées grâce à une recherche systématique *in silico* à partir des structures contenues dans la banque de données de structures tridimensionnelles des protéines (PDB).
La recherche étant basée uniquement sur la topologie du motif fonctionnel, elle s'effectue de manière indépendante de la structure secondaire des protéines hôtes et du ligand de référence. En effet, plusieurs exemples dans la littérature ont décrit l'existence de protéines naturelles ou la conception de plates-formes protéiques ayant des structures secondaires différentes mais réalisant la même fonction grâce à un même ensemble de résidus. Une étude réalisée par Hegyi et Gerstein sur la relation entre les structures et les fonctions des enzymes naturelles a montré que les fonctions glycosidases et carboxylyases étaient associées chacune à 7 structures ou repliements différents et que les hydrolases de monoester phosphorique et les hydrolases de monoester linéaires étaient associées chacune à 6 structures ou repliements différents (Hegyi et Gerstein, 1999). L'exemple le plus connu pour illustrer la diversité de structures pour la réalisation de la même fonction concerne les sérine protéases telles que la chymotrypsine de mammifères, la subtilisine bactérienne et la sérine protéase virale qui possèdent des structures secondaires totalement différentes alors qu'elles ont toutes une fonction de sérine protéase grâce à la même triade catalytique Ser-Asp-His (Figure III.1).

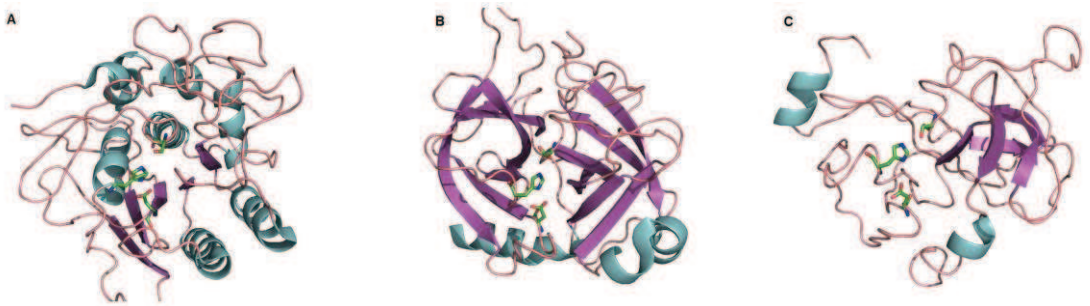


Figure III.1 : Structures tridimensionnelles de trois sérine protéases ayant des structures secondaires différentes mais les mêmes résidus catalytiques. Les protéines sont représentées en cartoon avec des couleurs suivant la structure secondaire (feuillets β en mauve et hélices α en cyan. La triade catalytique Ser-His-Asp est représentée en bâtons verts. A : Subtilisine (Code PDB 1SBT). B : Chymotrypsine (Code PDB 4CHA). C : Sérine protéase virale (Code PDB 1BEF).

III.2 Méthodes de sélection des protéines hôtes

III.2.1 STAMPS

STAMPS (Search for Three-dimensional Atomic Motif in Protein Structrure) est un programme développé au sein de notre laboratoire permettant de rechercher des motifs structuraux particuliers dans les protéines (Debret et al., 2009).

III.2.1.1 Principe de la recherche de motif

Chaque motif est représenté par une liste constituée exclusivement des atomes Cα et Cβ de chaque résidu. Ces atomes Cα et Cβ sont définis par leurs coordonnées cartésiennes.

- Distances interatomiques

STAMPS recherche dans la PDB des structures protéiques possédant des résidus adoptant la topologie du motif fonctionnel de référence. Pour ce faire, le programme calcule au préalable une matrice de distances d1, d2, d3 … entre les atomes Cα et Cβ des résidus composant le motif de référence, par la suite le programme cherche parmi les structures 3D de protéines de la PDB tous les ensembles de résidus décrits par une matrice de distances correspondant à la matrice de référence. La sélection s'effectue en imposant une valeur seuil de variation entre les distances caractérisant chaque résidu du motif de référence et celles caractérisant chaque résidu du motif mime : Δdmax.

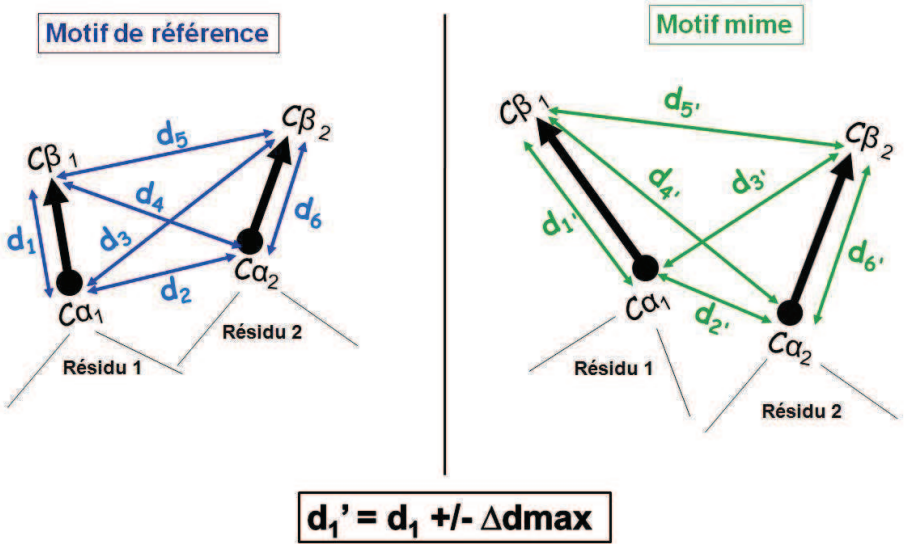


Figure III.2 : Distances interatomiques entre les différents résidus du motif fonctionnel de référence (à gauche) et entre les différents résidus du motif mime (à droite).

- RMSD entre les motifs

Lorsque toutes les distances interatomiques sont satisfaites, les résidus du motif de référence sont superposés aux résidus du motif mime sélectionné et l'écart moyen (RMSD) sur la position des atomes entre le motif de référence et le motif mime est calculé. La plate-forme sélectionnée portant le motif mime est retenue si le RMSD calculé est inférieur à une valeur seuil RMSDmax.

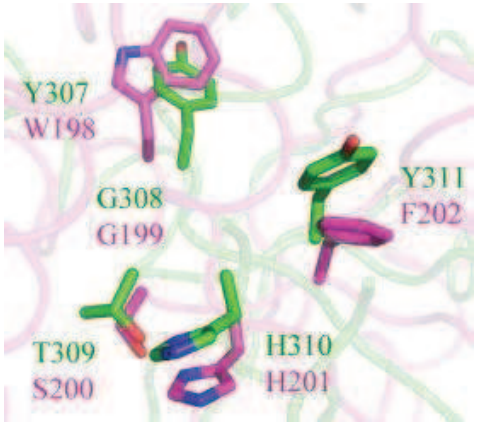


Figure III.3 : Superposition des résidus $Y_{307}T_{309}H_{310}Y_{311}$(bâtons verts) d'un motif mime avec les résidus $W_{198}S_{200}H_{201}F_{202}$(bâtons roses) d'un motif de référence donnant une valeur de RMSD entre les deux motifs.

- Angles entre les vecteurs Cα et Cβ

Le programme compare, après superposition, les angles Θ entre le vecteur Cα et Cβ d'un résidu du motif de référence et le vecteur Cα et Cβ d'un résidu du motif mime, en imposant une valeur seuil Θmax pour ce critère. Par ailleurs, l'écart maximal toléré sur la moyenne des angles Θ entre le motif de référence et le motif mime est fixé à une valeur ΔΘmax au-delà de laquelle le motif mime est rejeté.

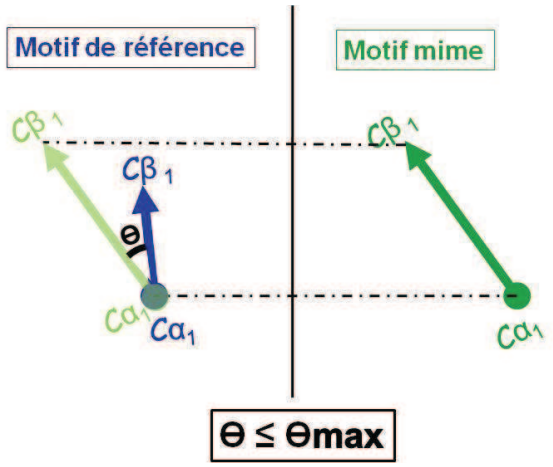


Figure III.4 : Angle Θ entre le vecteur Cα et Cβ d'un résidu du motif de référence et celui du motif mime.

III.2.1.2 Filtre stérique

Les plates-formes sélectionnées selon les critères topologiques relatifs au motif fonctionnel peuvent posséder des éléments structuraux qui entrainent un encombrement stérique entre la plate-forme mime et la protéine cible. Cette gêne stérique empêcherait le positionnement correcte du motif fonctionnel de la plate-forme par rapport à la protéine cible et par conséquent la liaison des deux partenaires. Afin d'éviter ces problèmes d'encombrement stérique, un filtre stérique a été implémenté dans le programme STAMPS. A partir de la structure 3D du complexe entre le ligand de référence et la protéine cible, le motif mime porté par la plate-forme sélectionnée est superposé au motif de référence porté par le ligand de référence dans sa conformation liée avec la protéine cible. Par la suite, les distances interatomiques entre tous les atomes de la plate-forme et tous les atomes de la cible sont calculées. Deux atomes se trouvent dans une situation de recouvrement ou interpénétration lorsque la distance qui les sépare est inférieure à la somme de leurs rayons de Van der Waals (Figure III.5).

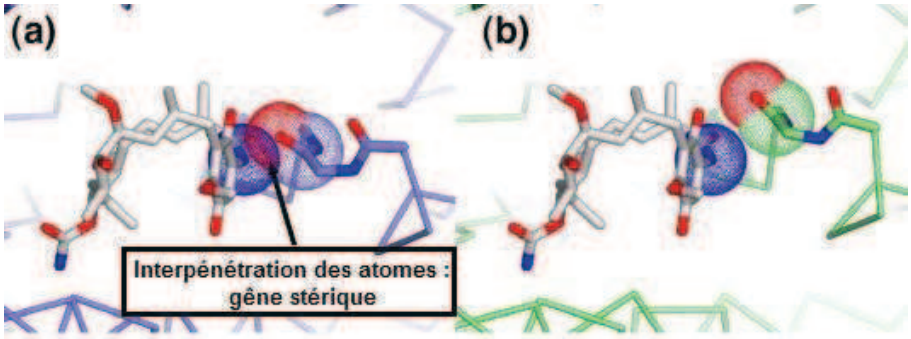


Figure III.5 : Modèle moléculaire d'un complexe protéique illustrant une gêne stérique entre deux résidus. Les sphères de Van der Waals sont représentées en pointillés de couleurs différentes selon le type d'atomes et les résidus sont représentés en bâtons. A : Recouvrement des sphères de VDW de l'azote (bleu) et de l'oxygène (rouge) entrainant un encombrement stérique. B : Léger déplacement du résidu Gly (vert) permettant d'éviter la gêne stérique.

Un score stérique est alors calculé correspondant au rapport entre les atomes de la cible dans une situation de recouvrement stérique et le nombre d'atomes de la cible présents à l'interface. Une plate-forme possédant un score supérieur à une valeur seuil est rejetée. La dernière version de STAMPS possède un mode de recherche dit « flexible » où, à partir des coordonnées du motif en entrée, le programme génère toutes les positions des atomes Cα, Cβ permettant de reproduire la position d'atomes particuliers présent sur la chaîne latérale d'un ou plusieurs acides aminés du motif fonctionnel. Cette fonctionnalité peut entrainer une explosion combinatoire du nombre de motif à rechercher. Ceci a conduit l'équipe à développer un algorithme mixte : une méthode de recherche en profondeur de l'arbre des solutions très efficace pour les motifs comportant un nombre faible de résidus et un algorithme de recherche de clique plus efficace que la première méthode pour les motifs de taille supérieure. Par ailleurs, une version parallèle du code STAMPS a été développée.

III.2.1.3 Autres critères

La recherche d'un motif peut être restreinte à des résidus identiques à ceux du motif de référence, ou à des résidus de propriétés physiques similaires. Les plates-formes examinées peuvent être limitées à des protéines ayant une taille définie par l'utilisateur. Des critères de sélection ou d'exclusion issus de l'analyse des mots clés des fichiers PDB sont possibles.

III.2.2 Comparaison avec les autres méthodes décrites dans la littérature

Plusieurs programmes basés sur la recherche de motifs 3D dans les structures protéiques sont décrits dans la littérature.

Dès 1994, Artymiuk et al. ont décrit un programme appelé **ASSAM** (Amino acid pattern Search for Substructures and Motifs) qui permet de cribler la PDB à la recherche de motifs constitués par les chaines latérales des résidus (Artymiuk et al., 1994 ; Nadzirin et al., 2012). Chaque résidu est représenté par deux pseudo-atomes dont les positions sont définies sur la chaine latérale. Ainsi, la recherche se fait par superposition des chaines latérales des résidus du motif, indépendamment de la position de la chaine principale des résidus. Le motif est constitué de 2 à 12 résidus. Cependant cette méthode utilise une représentation des résidus utilisant deux pseudo-atomes dont la définition est spécifique de chaque type d'acide aminé. Ceci rend donc la recherche de plate-forme dépendante du type d'acide aminé. Cette approche est parfaitement adaptée à l'identification de motifs fonctionnels identiques à celui recherché, mais elle n'est pas adaptée à l'ingénierie car le principe de cette recherche pourrait introduire un biais important.

Le programme **SPASM** (Spatial Arrangements of Side-chains and Main-chain) présente des similitudes avec ASSAM. SPASM a été développé pour rechercher, dans la base de données de protéines, une plate-forme possédant un motif dont l'arrangement spatial est identique à un motif spécifié par l'utilisateur (Kleywegt, 1999). Cependant, comme c'est le cas pour ASSAM, SPASM permet de rechercher des résidus de nature identique ou similaire à ceux contenus dans le motif de référence. Le programme STAMPS a été développé dans notre laboratoire dans la perspective spécifique de l'ingénierie moléculaire. La nature du résidu dans un motif mime peut être différente de celle du résidu correspondant dans le motif de référence. En effet, la recherche dans STAMPS est basée sur des résidus représentés par les atomes Cα et Cβ alors que celle réalisée dans SPASM est basée sur les résidus représentés par les atomes Cα et un pseudo-atome CG localisé au centre de gravité des atomes de la chaine latérale du résidu. Ainsi, ce pseudo-atome est caractéristique de la chaine latérale et dépend donc de la nature du résidu. Cette représentation crée un biais important dans la recherche en vue d'une substitution de l'acide aminé naturellement présent sur la plate-forme par le résidu correspondant dans le motif fonctionnel. Il faut noter cependant que STAMPS, comme SPASM, peut permettre de chercher des motifs de structures secondaires (chaine principale) via la seule prise en compte des atomes Cα uniquement. Par ailleurs, SPASM effectue des recherches à partir de bases de

données restreintes, ce qui n'est pas le cas de STAMPS qui peut parcourir l'intégralité de la PDB.

D'autres programmes ont été développés dans la perspective de l'ingénierie en greffant des motifs fonctionnels sur une plate-forme protéique hôte. **DEZYMER** (Hellinga et Richards, 1991), **GRAFTER** (Hearst et Cohen, 1994), **FITSITE** (Hornischer et Blocker, 1996), le programme développé par Liang et al. (Liang et al., 2000) et **ScaffoldSelection** (Malisi et al., 2009) présentent des similarités entre eux. Ils permettent de greffer des motifs discontinus. La recherche est fondée sur la géométrie des atomes Cα ou Cα et Cβ des résidus du motif fonctionnel. Dans certains cas (DEZYMER, ScaffoldSelection), toutes les combinaisons de Cα, Cβ qui permettent de reproduire les positions des atomes des groupes fonctionnels présents sur les résidus du motif à greffer sont représentées grâce à des bibliothèques de rotamères. Il est à noter que la dernière version du programme STAMPS (3.1.3) inclut une fonction similaire. Les plates-formes identifiées sont ensuite évaluées en termes de géométrie globale, orientation de la chaine latérale, conflits stériques (et accessibilité pour certaines applications, GRAFTER), ou en termes de complémentarité géométrique du modèle de complexe plateforme-protéine cible (programme développé par Liang et al.). Le programme développé par le groupe de Luhua Lai (Liang et al., 2000) est celui dont le principe présente le plus de similitudes avec le programme STAMPS. Les différences résident dans l'ajout d'une étape de mutation des résidus avoisinant le motif transféré sur les plates-formes identifiées afin d'optimiser la complémentarité géométrique.

Récemment, Zhang et Lai ont développé le programme **AutoMatch** (Zhang et Lai, 2012). Contrairement aux approches utilisées dans les programmes cités précédemment (GRAFTER, FITSITE, le programme développé par Liang et al. et ScaffoldSelection), AutoMatch prend explicitement en compte la possibilité de l'existence de faibles différences de conformations au niveau de la chaine principale entre le motif de référence et la plate-forme. Ceci est réalisé grâce à la génération de bibliothèques de conformations de sites actifs. Les jeux de coordonnées des atomes Cα, Cβ des différents conformères ainsi obtenus sont utilisés pour la recherche de protéines plates-formes. Le problème combinatoire complexe dû aux bibliothèques de conformations des résidus du site actif est traité par des étapes de réductions des ensembles. Pour cette raison, le nombre de résidus du site actif est compris entre 3 et 5 au

maximum. L'examen géométrique est réalisé en comparant la géométrie du site actif de référence avec le site actif identifié et en incluant un filtre stérique grâce à un score de collision géométrique. Finalement, les solutions sont évaluées par le calcul de l'énergie de liaison en utilisant un potentiel empirique.

RosettaMatch est un programme développé par le groupe de Baker D. (Zanghellini et al., 2006), utilisé pour la conception d'enzymes artificielles grâce à un protocole similaire à ceux utilisés dans d'autres méthodes précédemment décrites dans la littérature telles que ORBIT (Dahiyat et Mayo, 1997), DEZYMER (Hellinga et Richards, 1991) et plus récemment **OptGraft** (Fazelinia et al., 2009) et **ScaffoldSelection** (Malisi et al., 2009). Le protocole se déroule en 3 étapes : la détermination de paramètres géométriques d'un site actif permettant de décrire les positions spatiales des résidus de ce site par rapport à l'état de transition, la recherche dans une plate-forme protéique de positions où ce site actif peut être placé, et la mutation des résidus entourant ce site actif potentiel afin de stabiliser l'état de transition. De manière similaire à AutoMatch, les différentes conformations des sites actifs sont d'abord générées puis des positions compatibles avec la reproduction de ces géométries sont recherchées dans une plate-forme protéique. Au cours de cette étape, RosettaMatch emploie des algorithmes de hachage géométrique pour la recherche de l'emplacement optimal du site catalytique. Par la suite, le substrat (ou un modèle du substrat à l'état de transition) est positionné dans le site actif identifié. Les résultats sont évalués par des méthodes basées sur le calcul de l'énergie de liaison entre la plate-forme et l'état de transition et sur l'ajustement géométrique des résidus catalytiques dans le site actif.

III.3 Conclusion

L'approche de conception *in silico* de ligands protéiques utilisée dans le cadre de ce travail repose sur le programme STAMPS développé au sein de notre laboratoire. Cette approche permet le greffage d'un motif fonctionnel d'un ligand naturel de référence, pouvant être continu ou discontinu et constitué des résidus jouant un rôle important dans le phénomène d'association entre les partenaires protéiques. L'objectif de ce greffage sur des plates-formes protéiques, généralement de taille plus petite que celle du ligand naturel, est de reproduire la topologie fonctionnelle de ce motif afin de d'établir un jeu d'interactions intermoléculaires particulier. La Figure III.6 résume les différentes étapes de cette approche de conception de

ligands protéiques artificiels.

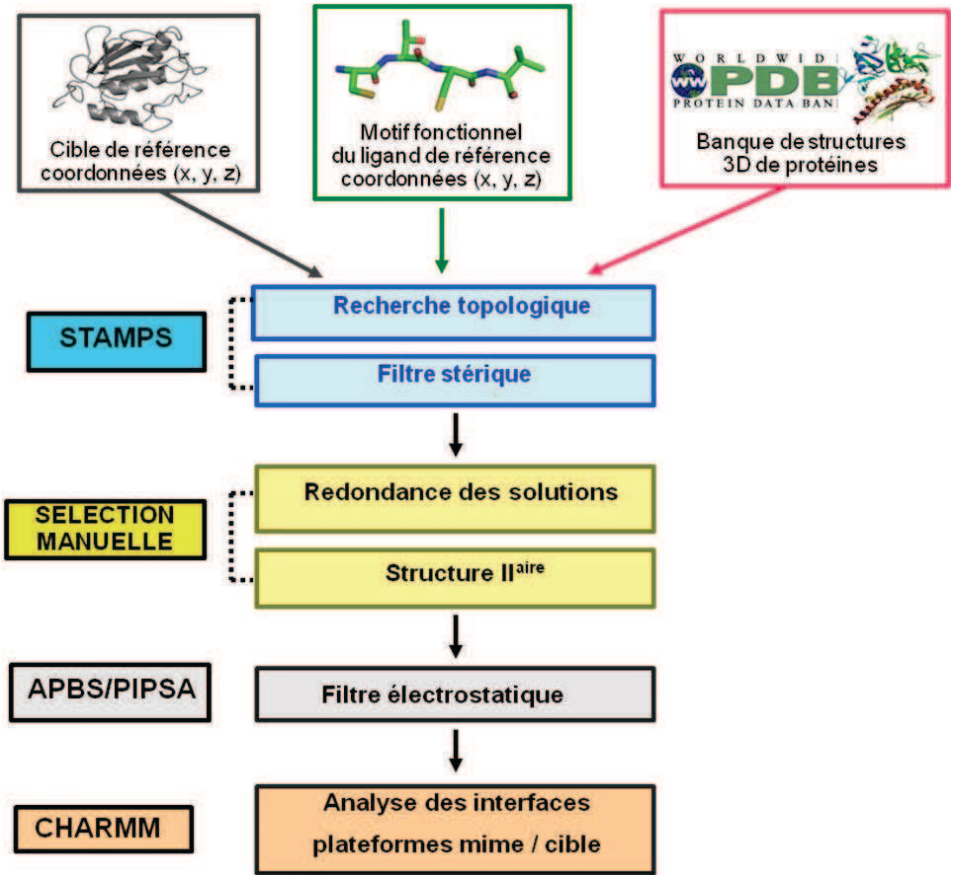


Figure III.6 : Schéma représentant les différentes étapes de conception *in silico* de ligands protéiques destinés à inhiber une cible d'intérêt, ainsi que les différents programmes utilisés.

La recherche des plates-formes hôtes est réalisée sur la totalité des structures contenues dans la PDB (plus de 80000 en 2012), augmentant ainsi la probabilité de trouver des plates-formes adéquates. L'étape de criblage de la PDB réalisée par le programme STAMPS et basée sur les paramètres topologiques du motif fonctionnel à transférer prend également en considération la complémentarité géométrique entre la cible et les plates-formes identifiées par l'application d'un filtre stérique. Les potentiels électrostatiques des plates-formes identifiées à l'issue de cette première phase sont comparés à celui du ligand protéique naturel de référence. Des modèles moléculaires des plates-formes retenues à l'issue de ces étapes, en interaction avec la cible considérée, sont réalisés. La sélection de ces plates-formes inclut la détermination dans les modèles moléculaires obtenus, d'une série de paramètres caractéristiques des interactions protéine-protéine et la vérification de la conformité de ces valeurs à celles reportées dans des études de complexes protéine-protéine naturels. Les plates-formes

satisfaisant ces critères sont alors évaluées expérimentalement pour leur capacité à lier la cible d'intérêt au site de liaison visé.

CHAPITRE IV : Application : Ciblage du domaine catalytique des MMPs

Les métalloprotéases matricielles (MMPs) constituent une famille de protéases à zinc dont 23 membres sont connus chez l'homme à ce jour. Elles sont collectivement capables de dégrader différents composants protéiques de la matrice extracellulaire (ECM). Ces MMPs ont été reconnues depuis des décennies comme cibles thérapeutiques et diagnostiques prometteuses, et ce sur la base de leur implication dans différentes pathologies humaines. L'implication spécifique des MMPs dans les processus pathologiques, et en particulier dans le processus oncogène, a alimenté les efforts fournis depuis une 20^{aine} d'années afin de développer et évaluer des inhibiteurs de MMPs comme potentiels agents thérapeutiques. Cependant, amener un inhibiteur de MMP au stade d'acceptation clinique reste encore un défi. Les inhibiteurs de MMPs naturels et synthétiques à large spectre ayant été développés ont conduit à des résultats décevants dans la plupart des essais cliniques. A ce jour, le seul inhibiteur de MMP approuvé par la FDA est le Periostat, un dérivé de tétracycline possédant une forte activité anti-collagénases et indiqué seulement pour le traitement de la parodontite (Peterson, 2004). L'échec de ces inhibiteurs à large spectre des MMPs a mis en évidence la nécessité de développer des inhibiteurs sélectifs permettant de discriminer les différents membres de la famille. Il existe seulement deux inhibiteurs sélectifs ciblant la MMP-12 et la MMP-13, et ayant des applications importantes, respectivement dans l'athérosclérose (Devel et al., 2006) et l'ostéoarthrite (Engel et al., 2005).

Pour continuer à développer et à évaluer la méthode de conception de ligands protéiques artificiels par greffage de motifs fonctionnels, nous avons choisi de cibler le site catalytique des MMPs. L'objectif de ce travail est donc la conception d'une série de nouveaux ligands protéiques capables d'interagir avec les membres de la famille des MMPs. Cette étude permettra d'estimer le taux de succès de cette approche et devrait fournir une série de molécules de première génération qui pourront être ultérieurement optimisées en vue d'obtenir des molécules plus affines et plus sélectives.

IV.1 Aspects structuraux des MMPs

IV.1.1 Organisation des différents domaines des MMPs

L'organisation des MMPs est résumée sur la figure IV.1. Toutes les MMPs possèdent un domaine minimal commun contenant :

a) Le peptide signal : une séquence signal N-terminale constituée d'une vingtaine de résidus, généralement hydrophobes, permettant de diriger le polypeptide synthétisé vers les voies de sécrétion.

b) Le pro-peptide : un domaine codant pour un pro-peptide composé d'environ 80 résidus contenant le motif strictement conservé PR**C**GXPD avec le résidu cystéine qui coordonne le zinc catalytique afin de maintenir l'enzyme à l'état inactif ; l'activation de la MMP implique le clivage du pro-peptide qui entraine la dissociation de l'interaction zinc/cystéine et le positionnement d'une molécule d'eau comme 4ème ligand du zinc.

c) Le domaine catalytique : ce domaine est composé de 180 résidus environ et est caractérisé par la présence d'une séquence consensus conservée **H**EXG**H**XXGXX**H** contenant trois résidus histidines chélatant le zinc du site actif de l'enzyme. Outre l'atome de zinc catalytique, le domaine catalytique contient un autre atome de zinc dit structural et de un à trois atomes de calcium. Les deux gélatinases (MMP-2 et -9) contiennent dans leurs domaines catalytiques trois motifs répétitifs structurellement proches des domaines de fibronectine de type II, constitués de 175 résidus, et qui interviennent dans l'interaction de ces enzymes avec leur substrat la gélatine (Brinckerhoff et Matrisian, 2002).

d) Le domaine hémopexine : les MMPs ayant uniquement le domaine minimal sont les matrilysines (MMP-7 et -26). Les structures les plus communes pour les MMPs sécrétées ont en plus un domaine hémopexine constitué de 210 résidus et connecté au domaine catalytique par une région charnière flexible. Ce domaine est nécessaire aux collagénases pour la dégradation de la triple hélice de collagène interstitiel (Bode, 1995). De plus, ce domaine apparaît essentiel dans l'activation de la proMMP-2 par la MMP-14 à la surface de

la cellule (Brinckerhoff et Matrisian, 2002; Visse et Nagase, 2003). A la place du domaine hémopexine, la MMP-23 possède un domaine riche en cystéines et prolines ainsi qu'un domaine similaire à celui du récepteur à l'interleukin-1.

Quatre MT-MMPs (MT1-, MT2-, MT3-, et MT5-MMP) ont un domaine transmembranaire de 24 résidus suivi d'une queue cytoplasmique C-terminale de 20 résidus alors que les MT4- et MT6-MMP possèdent un groupement glycosylphosphatidylinositol permettant l'ancrage membranaire de l'enzyme à la surface de la cellule (Brinckerhoff et Matrisian, 2002).

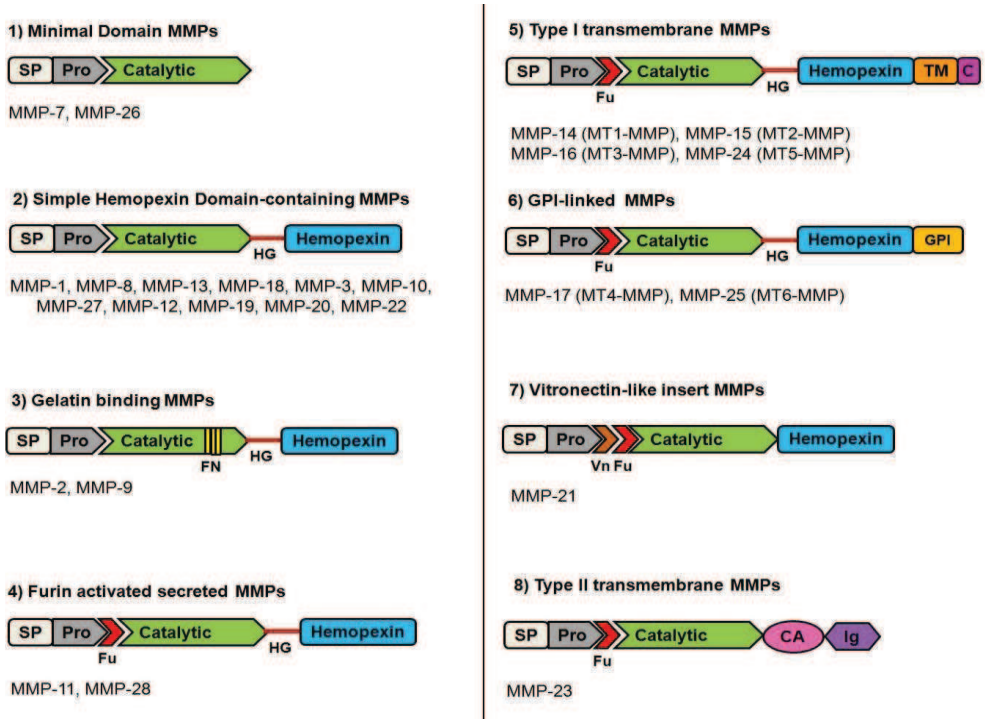


Figure IV.1 : Organisation des domaines des différentes MMPs. Les principaux domaines incluent le peptide signal (SP), le pro-domaine (Pro), le domaine catalytique, le domaine charnière (HG), le domaine hémopexine, et dans certains cas un domaine transmembranaire (TM) et une queue cytoplasmique (C) ou un domaine d'ancrage glycosylphosphatidylinositol (GPI). Un site de clivage à la furine (Fu) et un motif vitronectine (Vn) entre le pro-domaine et le domaine catalytique sont présents dans certaines MMPs. Dans les gélatinases, des répétitions de fibronectine de type II (FN) sont également présentes. Un domaine riche en cystéines (CA) et un domaine semblable à l'immunoglobuline (Ig) sont présents dans la MMP-23.

IV.1.2 Structure tridimensionnelle du domaine catalytique des MMPs

La structure tridimensionnelle de la proMMP-2 résolue en 1999 par le groupe de K. Tryggvason (Morgunova et al., 1999) illustre l'organisation de cette famille de protéines en plusieurs domaines, chacun présentant une structure propre (Figure IV.2).

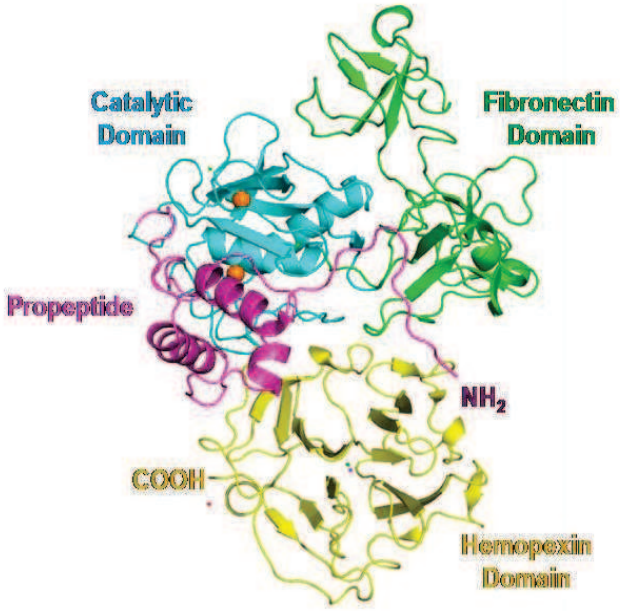


Figure IV.2 : Structure cristallographique de la forme entière de la pro-MMP-2 humaine (code PDB 1CK7) montrant à l'extrémité N-terminale le pro-peptide (rose), le domaine catalytique (cyan), le domaine fibronectine (vert) spécifique aux MMP-2 et -9, et le domaine hémopexine (jaune) à l'extrémité C-terminale ; l'atome de zinc catalytique est signalé par une sphère (orange) au sein du domaine catalytique (Morgunova et al., 1999).

A ce jour, les structures du domaine catalytique des MMP-1, -2, -3, -7, -8, -9, -10, -11, -12, -13, -14 et -16 ont été résolues et déposées dans la PDB. Les domaines catalytiques des différentes MMPs possèdent une forte homologie structurale. En effet, la superposition des chaines principales des domaines catalytiques des différentes MMPs donne un RMSD de 0,7 à 0,8 Å, illustrant la topologie très conservée des sites actifs de ces enzymes. Les domaines catalytiques sont caractérisés par la présence d'un sillon qui s'étend de manière horizontale à la surface de la protéine et qui sépare le sous-domaine inférieur de petite taille du sous-domaine supérieur de taille plus large (Figure IV.3). Ce sillon contient le site actif des MMPs et accueille les différents substrats et inhibiteurs de ceux-ci. Le site actif est organisé en plusieurs sous-sites (S_3, S_2, S_1, S_1', S_2', S_3') dont le **sous-site S_1'** correspondant à une cavité hydrophobe et profonde et contrôlant la spécificité des substrats et inhibiteurs vis-à-vis des MMPs (Figure IV.4).

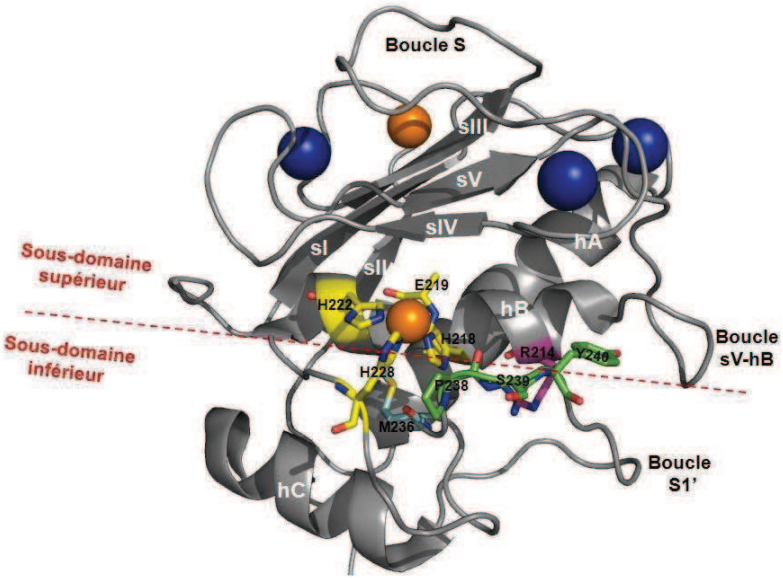


Figure IV.3 : Structure 3D du domaine catalytique de la MMP-1 (Code PDB 1SU3), représentée en mode cartoon, avec les différents éléments structuraux (feuillets β sI à sV, hélices α hA à hC et boucles en gris), les résidus les plus importants pour la fonction représentés en bâtons (les 3 His et le Glu en jaune) et pour la structure (Segment P238-S239-Y240 en vert et R214 en mauve) ainsi que les ions Zn^{2+} (orange) et Ca^{2+} (bleu) représentés par des sphères.

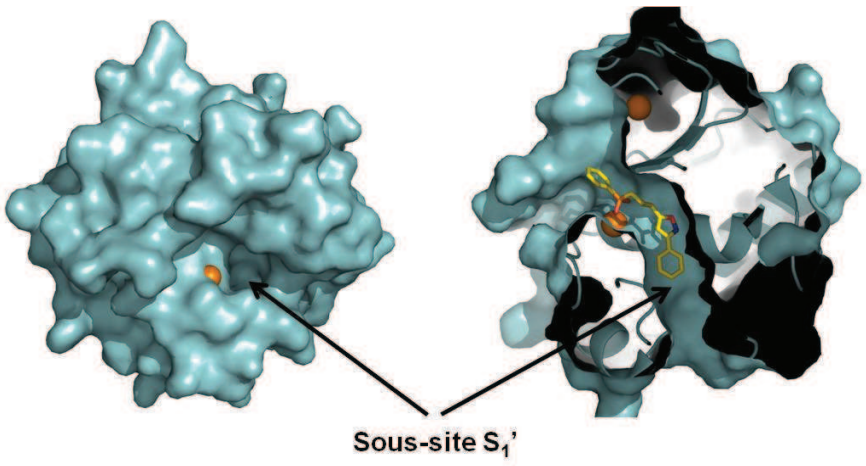


Figure IV.4 : Structure cristallographique du domaine catalytique de la MMP-9 en complexe avec un inhibiteur phosphinique (code PDB 2OVZ) montrant le sous-site S_1'. La MMP-9 est représentée en mode surface de couleur cyan (à gauche) et l'inhibiteur phosphinique est représenté en mode bâtons de couleur jaune (à droite) ; les atomes de Zn sont représentés en sphères de couleur orange.

Le sous-domaine supérieur du domaine catalytique est constitué de 5 feuillets β (sI à sV), 2 hélices α (hA et hB) et 3 boucles en surface (boucle S, boucle sV-hB et boucle de spécificité). L'hélice hB située au niveau du site actif contient deux histidines (His218 et His222) du motif de chélation du zinc catalytique, ainsi que le Glu219 participant à la réaction d'hydrolyse des substrats par les MMPs (nomenclature correspondant à la MMP-1). La 3ème

histidine (His228) liant le Zinc se trouve à la sortie de cette hélice hB. Ces 3 histidines sont strictement conservées à ces positions précises dans toutes les MMPs. En position 236, une Met strictement conservée, caractéristique de la famille des metzincines, forme une structure locale appelée « Met-turn » servant de base afin de supporter la structure autour du zinc catalytique (Bode et al., 1993).
Le sous-domaine inférieur est constitué de l'hélice α hC et de la boucle de spécificité, également appelée boucle S_1'. Cette boucle de spécificité délimite la partie inférieure du sous-site S_1' et le segment Pro238-Ser239-Tyr240 (cette dernière étant strictement conservée dans toutes les MMPs) tapisse la surface extérieure de l'entrée de ce sous-site.

Selon les MMPs, de légères différences existent dans la composition des résidus constitutifs du site actif, notamment au niveau du sous-site S_1'. En particulier, le résidu en position 214 dans la MMP-1 joue un rôle critique dans la taille de la cavité S_1' qui varie en fonction des MMPs. La MMP-1 est caractérisée par une cavité fermée et peu profonde comparée aux autres MMPs (Figure IV.5), du fait de la présence d'un résidu Arg en position 214 au niveau de l'hélice hB qui se projette dans la cavité S_1' en direction du zinc catalytique et qui ferme la cavité. Dans une position équivalente, la MMP-7 possède un résidu Tyr qui a également pour conséquence la fermeture de la cavité. Dans la MT3-MMP (MMP-16), cette cavité est limitée par la présence de la chaine latérale du résidu Met271 au niveau de la boucle de spécificité, qui pointe vers la poche et en réduit la profondeur ; ce résidu étant lui-même bloqué par les résidus Ile263 et Gln269. En revanche, le même résidu Met à cette même position (Met264) dans la MT1-MMP (MMP-14) pointe vers l'extérieur de la cavité donnant lieu à une cavité très profonde et ouverte par le bas (Lang et al., 2004).

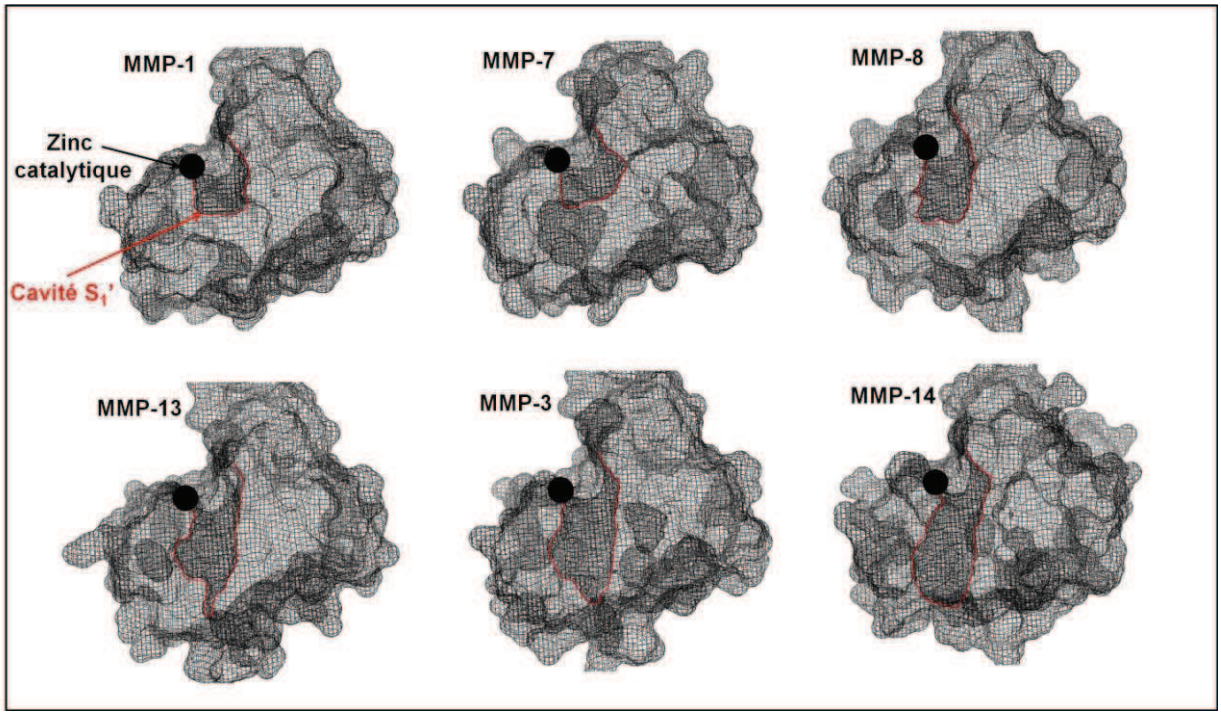


Figure IV.5 : Comparaison entre les cavités S_1' des différentes MMPs. Les MMPs sont représentées en mode « mesh » de couleur gris, les cavités sont délimitées par un contour rouge et le zinc catalytique est représenté en mode sphère de couleur noire. De gauche à droite et de haut en bas : MMP-1 (code PDB 1HFC), MMP-7 (1MMP), MMP-8 (1JAN), MMP-13 (2E2D), MMP-3 (1HFS) et MMP-14 (1BQQ).

Outre les différences qui existent au niveau de la géométrie de la cavité S_1' en fonction des MMPs, des différences existent également au niveau de la boucle de spécificité S_1'. En effet, les séquences et les tailles de cette boucle sont variables selon les MMPs. Les positions de 241 à 248 et la position 250 (numérotation correspondant à la MMP-3) ont une composition très différente et la taille pour cette région non conservée varie entre 7 et 10 résidus (Figure IV.6). Par ailleurs, la boucle de spécificité présente également une très grande flexibilité en solution (Figure IV.7). Ces différences en termes de taille, de séquence mais aussi de mobilité de cette boucle contribuent à la spécificité des différentes MMPs pour leurs substrats et leurs inhibiteurs.

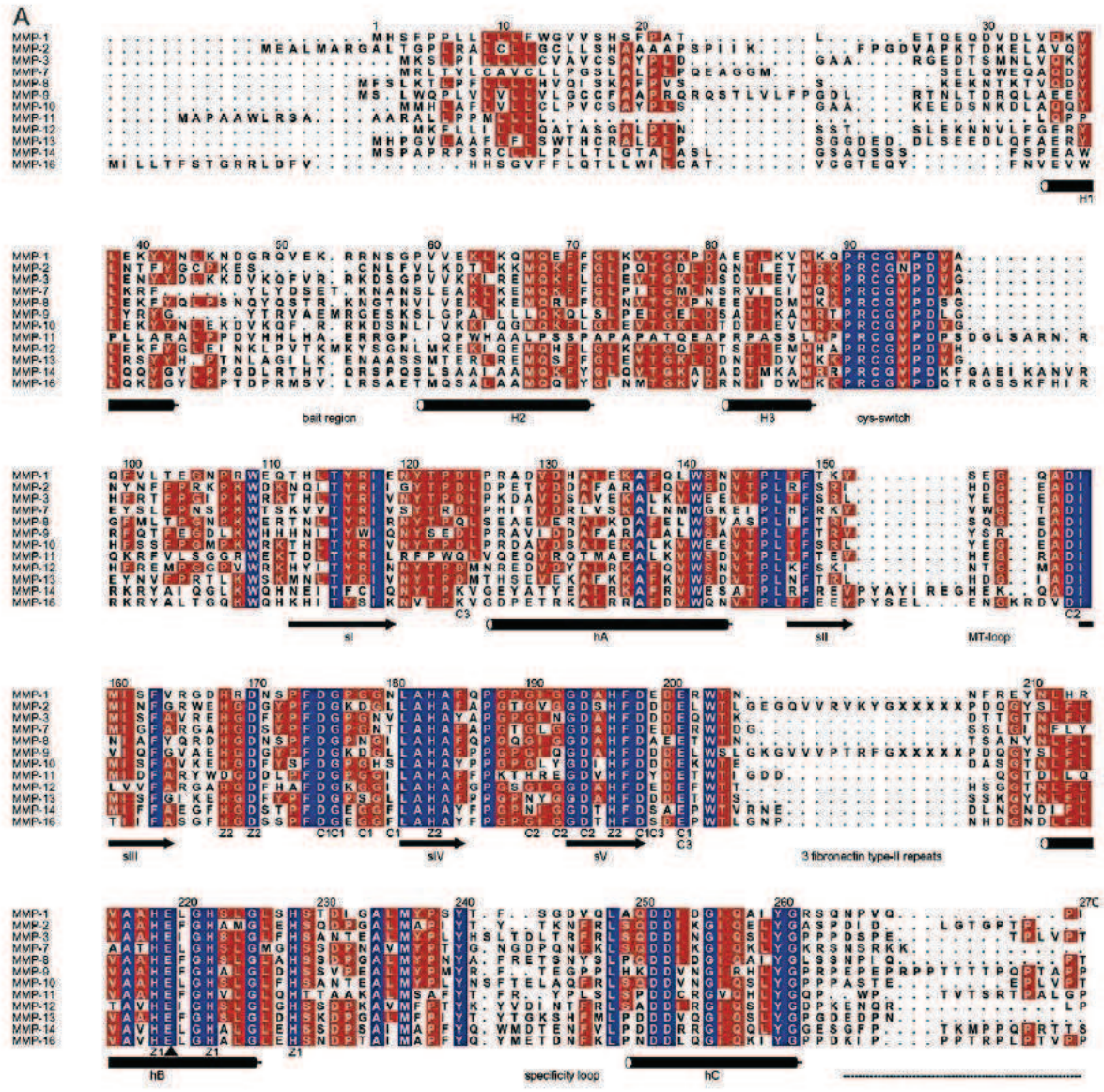


Figure IV.6 : Alignement de séquences correspondant aux pro-domaines et domaines catalytiques des MMP. La numérotation correspond à la pré-proMMP-1. Les différents éléments structuraux sont indiqués en dessous des séquences : les hélices H1 à H3 correspondent au pro-domaine, les feuillets sI à sV et les hélices hA à hC correspondent au domaine catalytique. Les résidus strictement conservés sont indiqués par un fond bleu et les sites hautement conservés par un fond rouge. Z1 et Z2 indiquent les sites de liaison au zinc catalytique et structural respectivement, C1, C2 et C3 se rapportent aux sites de liaison des ions calcium. Le Glu catalytique est indiqué par un triangle. (Maskos, 2005)

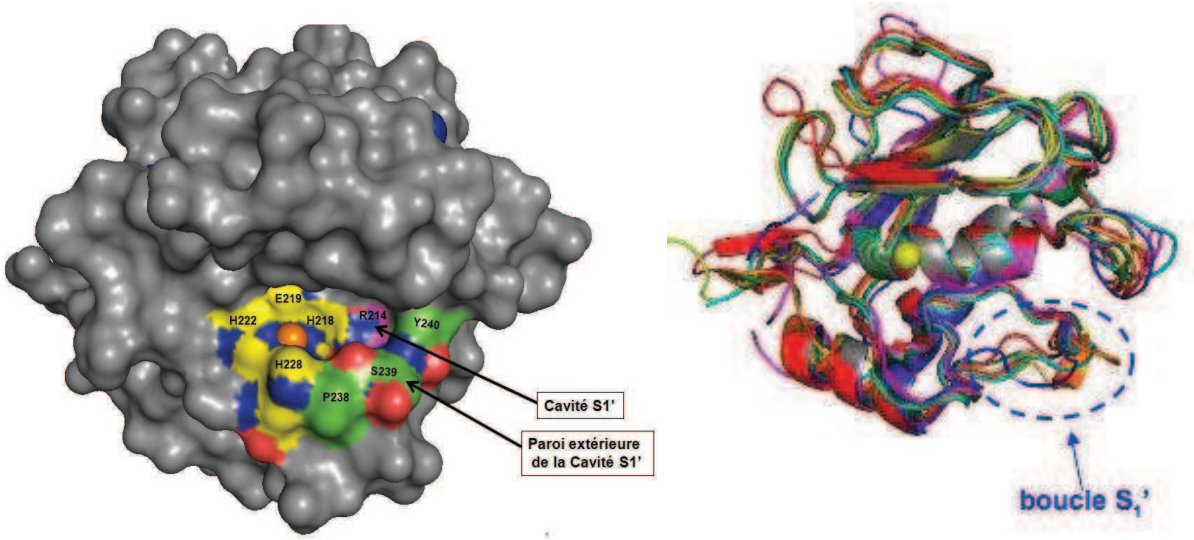


Figure IV.7 : Structures tridimensionnelles du sous-site S_1' du domaine catalytique des MMPs. A gauche : Structure 3D de la MMP-1 représentée en mode surface (gris) montrant la cavité S1' fermée par Arg214 (rose) et la paroi extérieure de la cavité formée par le segment PSY (vert). A droite : Superposition des structures 3D des domaines catalytiques de plusieurs MMPs montrant la grande variabilité de la boucle S_1'.

IV.2 Les MMPs : des cibles thérapeutiques potentielles

IV.2.1 Rôles physiologiques

La plupart des MMPs sont exprimées sous forme de pro-formes inactives (zymogènes) qui doivent être activées pour assurer leurs fonctions physiologiques. Cette activation, impliquant le clivage du pro-peptide, peut se faire soit dans le milieu extracellulaire, soit à la surface des cellules, ou encore dans la cellule au niveau des voies de sécrétion des protéines. Plusieurs protéases, comme la plasmine et les MMPs elles-mêmes, sont capables d'éliminer le pro-peptide et d'activer les MMPs, conduisant à des formes actives de ces enzymes.

Malgré une certaine spécificité des substrats, les spectres protéolytiques des différentes MMPs sont assez larges et se recouvrent, ce qui permet la dégradation de l'ensemble des composants de la matrice extracellulaire. Les MMPs sont donc considérées comme des acteurs majeurs du remodelage tissulaire (Brinckerhoff et Matrisian, 2002 ; Mott et Werb, 2004). Une surexpression de ces enzymes est observée lors de la croissance embryogénique et de la morphogenèse des tissus, évènements fondamentaux qui requièrent la disruption des barrières

de la matrice extracellulaire pour permettre la migration cellulaire et le remodelage du microenvironnement de la matrice. La plupart des gènes de MMP sont aussi surexprimés dans plusieurs processus de la reproduction, tels que le cycle menstruel, l'ovulation et la résorption de l'utérus après l'accouchement. Au stade adulte, il n'existe pas d'expression constitutive des MMPs, sauf au sein de tissus en renouvellement cellulaire permanent (gencive, intestin) ou lors de processus de cicatrisation des blessures.

IV.2.2 Implication des MMPs à différents stades d'un processus pathologique

Si les MMPs ne sont pas exprimées dans la plupart des tissus sains, elles sont en revanche surexprimées lors de différents processus pathologiques, comme le cancer (Egeblad et Werb, 2002 ; Folgueras et al., 2004), la sclérose en plaques (Rosenberg, 2005), l'athérosclérose (Dollery et Libby, 2006) et l'arthrose (Burrage et al., 2006).

Des études récentes ont montré que le rôle des MMPs n'est pas limité à celui qui découle de leur activité de dégradation directe des composants de la matrice extracellulaire (Egeblad et Werb, 2002). Ainsi, les contributions des MMPs dans les différentes étapes de la progression tumorale sont nombreuses. Certaines MMPs sont impliquées dans la perte d'adhérence cellulaire durant les premiers stades du développement d'une tumeur carcinomateuse. D'autres MMPs, intervenant dans l'étape d'invasion locale, ont une activité protéolytique dirigée contre les protéines de la membrane basale et contre le collagène interstitiel présent dans le tissu conjonctif qui encercle les cancers invasifs. Les MMPs interviennent également dans la prolifération des cellules tumorales (la troisième étape de la progression tumorale) faisant intervenir les facteurs de croissance (cytokines). Finalement lors de l'angiogenèse (quatrième étape de la progression tumorale), les MMPs peuvent avoir un effet stimulateur ou répresseur.

IV.2.3 Rôle ambivalent des MMPs

Les MMPs ne sont pas toutes des cibles thérapeutiques. En effet, l'identité et la contribution exacte des MMPs impliquées dans différents processus pathologiques ne sont pas clairement identifiées, et les effets positifs ou négatifs qu'exercent les MMPs peuvent varier en fonction de différents types et différents stades de pathologies.

Ainsi, à travers des études de transcriptomique et d'invalidation de gènes, seulement trois MMPs (MMP-1, MMP-2 et MMP-7) ont été validées comme cibles thérapeutiques potentielles (Overall et Kleifeld, 2006). À l'inverse, d'autres MMPs sont aujourd'hui considérées comme des « anti-cibles » de part leur implication dans des mécanismes de défense de l'hôte. Les expériences menées par le groupe de Overall et Kleifeld ont conduit à la validation des MMP-3, MMP-8 et MMP-9 comme anti-cibles du fait de leurs effets anti-tumorigéniques (Overall et Kleifeld, 2006). De plus, Witty et al. ont également démontré que la progression tumorale dans divers modèles animaux était inversement corrélée avec le niveau d'expression des MMP-3 et MMP-8 (Witty et al., 1995). Ces résultats paraissent paradoxaux en vue des effets pro-tumorigéniques non ambigus des MMP-3 et MMP-8 (Egeblad et Werb, 2002 ; Fingleton, 2003). Au vu de ces rôles ambivalents, ces MMPs ont été classées comme anti-cibles car l'inhibition ou l'inactivation thérapeutique de ces MMPs anti-tumorigéniques peuvent être préjudiciables et même dangereuses pour l'hôte.

La classification des MMPs comme cibles ou anti-cibles thérapeutiques peut devenir encore plus compliquée car une même MMP peut avoir des rôles opposés à différents stades de la progression tumorale. Par exemple, une étude sur un modèle de souris transgénique de carcinome épidermoïde invasif suggère que la MMP-9 stimule la prolifération durant les stades précoces du cancer mais qu'elle limite la progression à des stades plus avancés (Coussens et al., 2000). Une autre étude a montré que la MMP-11 a un rôle stimulateur dans les tumeurs primaires puis agit comme répresseur au cours des invasions distales (Andarawewa et al., 2003).

Par ailleurs, les différentes contributions d'une même MMP peuvent être opposées en fonction du type de pathologie considérée. En effet, une étude réalisée chez des patients atteints de cancer colorectal a démontré que le gène de la MMP-12 joue un rôle important dans l'inhibition de la progression tumorale et que la surexpression de la MMP-12 est étroitement corrélée à un meilleur pronostic (Yang et al., 2001). En revanche, dans un autre type de pathologie tel que l'athérosclérose, la MMP-12 est impliquée dans la rupture de la plaque d'athérome pouvant conduire à un infarctus du myocarde. Il a été démontré que la surexpression de la MMP-12 active conduit à une inflammation et à la réduction du collagène dans les plaques d'athérome chez les lapins atteints d'hypercholestérolémie (Liang et al., 2006). D'autres expériences réalisées sur des souris KO (Knock-Out) ont montré que la délétion de la MMP-12 a pour conséquence des lésions plus petites et plus stables au niveau de

l'artère brachiocéphalique (Johnson et al., 2005) et une dégradation réduite de l'élastine dans l'arc aortique (Luttun et al., 2004), démontrant l'implication de la MMP-12 dans la déstabilisation des plaques d'athérome.

Dans un contexte de conception de nouveaux ligands protéiques inhibiteurs de MMPs, il est donc indispensable de mettre au point des inhibiteurs hautement sélectifs des MMPs afin de cibler spécifiquement les MMPs qui jouent un rôle stimulateur dans le processus pathologique et éviter les MMPs impliquées dans les mécanismes de défense de l'hôte au cours du processus pathologique.

IV.3 Les inhibiteurs naturels des MMPs : les TIMPs

L'activité des MMPs est régulée par des inhibiteurs endogènes présents dans le plasma et les fluides tissulaires comme les α2-macroglobulines, mais également par des inhibiteurs plus spécifiques tels que les TIMPs. Les TIMPs sont des inhibiteurs tissulaires de métalloprotéases qui constituent une famille de quatre protéines TIMP-1, 2, 3, et 4 chez l'homme, se liant fortement et de façon non covalente au site actif des MMPs par formation d'un complexe 1:1 (Brew et al., 2000). Une perturbation de la balance physiologique normale entre les TIMPs et les MMPs est liée au développement des pathologies.

IV.3.1 Aspects structuraux des TIMPs

Tous les TIMPs ont une structure conservée et présentent environ 40 à 50% d'identité de séquence (Greene et al., 1996) (Figure IV.8).

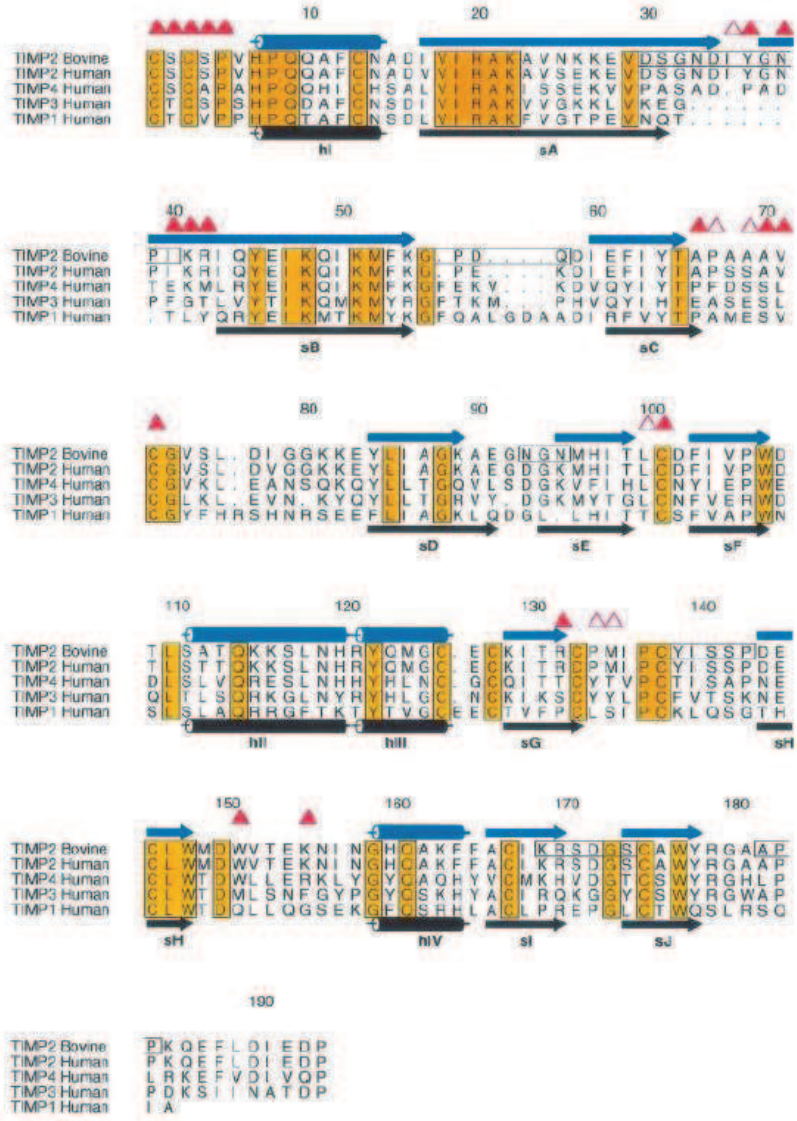


Figure IV.8 : Alignement de séquences des TIMP-1, -2, -3, -4 humains et du TIMP-2 bovin. Les flèches et les cylindres bleus indiquent la présence de feuillets et d'hélices respectivement ; les résidus fortement ou faiblement impliqués dans des contacts intermoléculaires de moins de 4Å de distance sont indiqués par des triangles remplis ou vides respectivement. Les segments du TIMP-2 qui varient de manière importante par rapport au TIMP-1 sont entourés de rectangles et les résidus strictement conservés parmi les TIMPs sont indiqués par des rectangles oranges. (Fernandez-Catalan et al., 1998).

La résolution de la structure cristallographique du complexe TIMP-1/MMP-3 par le groupe de W. Bode (Gomis-Rüth et al., 1997) a permis pour la 1ère fois de montrer la structure tridimensionnelle d'un TIMP et de mettre en évidence les caractéristiques géométriques de l'interaction avec les domaines catalytiques de MMP (Figure IV.9).

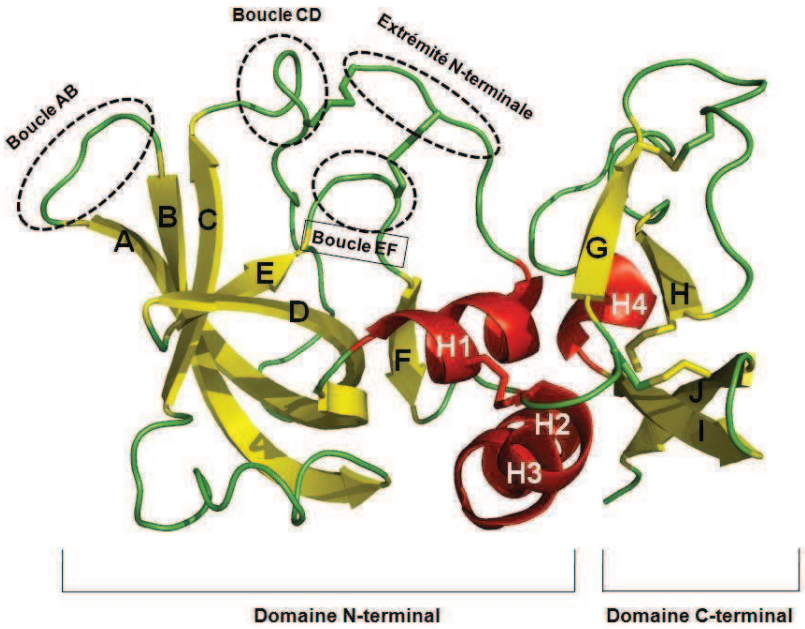


Figure IV.9 : Structure cristallographique du TIMP-1 (Code PDB 1UEA), représenté en mode cartoon et montrant les deux domaines N- et C-terminaux avec les différents éléments structuraux : feuillets (jaune), hélices (rouge), boucles (vert).

Par la suite, le même groupe a résolu la structure cristallographique du complexe TIMP-2/MMP-14 (Fernandez-Catalan et al., 1998), mais à ce jour, aucune structure 3D du complexe de MMP avec le TIMP-3 et le TIMP-4 n'a été décrite. En revanche, plusieurs autres structures 3D existent pour différents complexes du TIMP-1 et du TIMP-2 en interaction avec les MMP-1, -10 et -13 et la Pro-MMP-2.

Les TIMPs, constitués de 184 à 194 acides aminés, sont organisés en deux domaines distincts : un domaine N-terminal d'environ 125 acides aminés et un domaine C-terminal d'environ 65 acides aminés, chacun d'eux est stabilisé par 3 ponts disulfures strictement conservés. Le domaine N-terminal, très conservé, est constitué d'une structure en tonneau β formée de 5 feuillets β (brins de A à F) ainsi que de trois hélices α (H1 à H3). La structure du domaine N-terminal est similaire au repliement de type OB (oligosaccharide/oligonucleotide binding). L'extrémité N-terminale, formée par les 5 premiers résidus, est bordée par les boucles AB, CD et EF (Figure IV.9). Les deux Cys 1 et 3 engagées dans 2 ponts disulfures reliant respectivement les boucles CD et EF permettent de conférer un cadre structural rigide à l'extrémité N-terminale des TIMPs. La comparaison entre les structures 3D des TIMP-1, -2 et -3 (Figure IV.10) montre une parfaite superposition des extrémités N-terminales ainsi que des boucles CD et EF pour les 3 protéines, suggérant que les interactions établies dans ces régions

sont très conservées. En revanche, des différences existent au niveau de la boucle AB dont la longueur diffère en fonction des TIMPs : cette boucle est la plus longue dans le TIMP-2 par rapport à celle des TIMP-1, -3 et -4, avec une insertion de 7 résidus supplémentaires en comparaison avec le TIMP-1 (Figure IV.8 et Figure IV.10). Ce domaine N-terminal est capable de se replier correctement en une structure native stable ainsi que de fonctionner de manière indépendante du domaine C-terminal. Il est à la fois nécessaire et suffisant dans l'inhibition des MMPs (Greene et al., 1996; Apte et al., 1995; Murphy et Willenbrock, 1995). Le domaine C-terminal est formé de 4 feuillets β (G, H, I et J) et d'une hélice α (H4), tandis que l'extrémité C-terminale semble être flexible. Ce domaine C-terminal module l'interaction des TIMPs avec les pro-formes des gélatinases par le biais des domaines hémopexine de ces dernières (Murphy et Willenbrock, 1995) : le TIMP-1 interagit avec la pro-MMP-9, les TIMP-2 et -4 interagissent avec la pro-MMP-2 et le TIMP-3 avec les deux gélatinases.

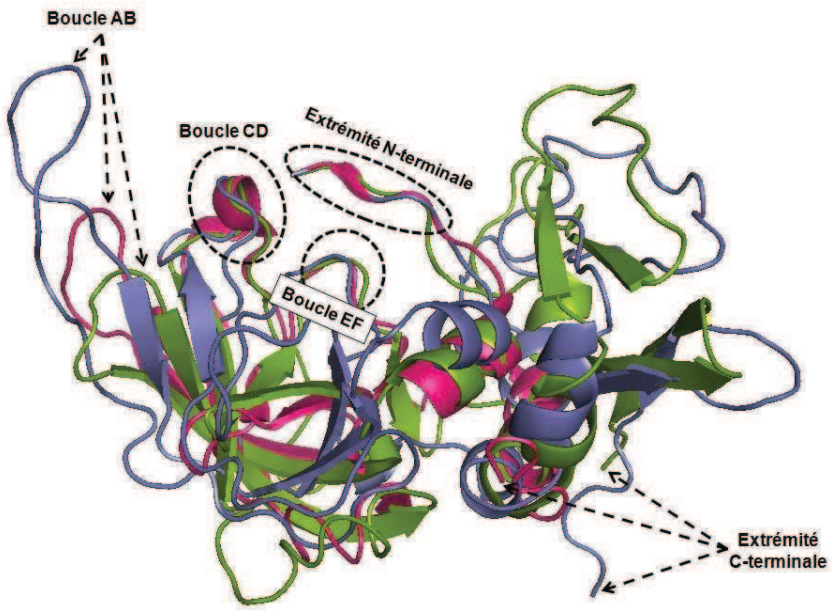


Figure IV.10 : Superposition des structures cristallographiques des TIMPs, représentés en mode cartoon. Les structures des TIMP-1 (Vert), TIMP-2 (Bleu) et N-TIMP-3 (Rose) sont issues respectivement des complexes TIMP-1/MMP-3 (Code PDB 1UEA), TIMP-2/MMP-14 (Code PDB 1BQQ), et N-TIMP-3/TACE (Code PDB 3CKI).

IV.3.2 Affinité et sélectivité

Les TIMPs inhibent toutes les formes actives des MMPs avec une sélectivité faible, et une affinité élevée, à l'exception du TIMP-1 qui possède une affinité très réduite pour la MMP-19, la MT1-MMP (MMP-14), la MT3-MMP (MMP-16) et la MT5-MMP (MMP-24) (Stracke et al., 2000 ; Hernandez-Barrantes et al., 2002 ; Amour et al., 1998). Le TIMP-3 se distingue

des autres TIMPs de part sa capacité à inhiber une gamme plus large de métalloprotéases dont plusieurs membres de la famille des ADAMs (A Disintegrin And Metalloproteinase) et ADAMTS (ADAM with Thrombospondin motifs) (Amour et al., 1998 ; Kashiwagi et al., 2001 ; Wang et al., 2006). Les TIMP-1, -2 et -4 montrent une activité inhibitrice limitée pour la famille des ADAMs. En effet, les TIMPs possèdent une affinité sub-nanomolaire pour toutes les MMPs, déterminée à environ 1 nM (Brew et al., 2000 ; Will et al., 1996) à l'exception de l'interaction de très forte affinité entre le TIMP-2 et la MMP-2 estimée à 1 fM (Hutton et al., 1998) et celle de très faible affinité entre le TIMP-1 et la MMP-14 dont le Ki a été déterminé à environ 200 nM avec le N-TIMP-1 (Lee et al., 2003). Des expériences de mutagénèse réalisées sur le TIMP-1 ont démontré qu'une mutation unique était capable de transformer le TIMP-1 en un inhibiteur actif envers la MMP-14 (Lee et al., 2003). En effet, la mutation T98L au niveau de la boucle EF permet un gain d'affinité d'un facteur 10 avec un Ki déterminé à 11 nM pour la MMP-14. Ceci s'explique probablement par le fait que ce résidu hydrophobe Leu soit présent dans tous les TIMPs à cette position à l'exception du TIMP-1.

IV.3.3 Mécanisme d'inhibition des MMPs

La formation du complexe TIMP/MMP crée une interface d'interaction de 1250 à 1300 $Å^2$ sur chacune des protéines, au niveau de laquelle des contacts plus ou moins forts s'établissent entre les deux partenaires. Il a d'abord été démontré qu'une forme tronquée du TIMP-1 ne possédant que les 2/3 de la chaine polypeptidique en partant de l'extrémité N-terminale (N-TIMP-1) inhibe la MMP-3 avec une affinité légèrement réduite par rapport à la forme entière du TIMP-1 (Murphy et al., 1991 ; Willenbrock et Murphy, 1994). Le résultat de ces travaux suggère que l'interaction spécifique du TIMP avec la MMP se fait essentiellement au niveau de l'extrémité N-terminale. Par la suite, le mécanisme d'inhibition a pu être décrit de manière détaillée grâce à l'étude des structures cristallographiques des complexes TIMP-1/MMP-3 (Gomis-Rüth et al., 1997), TIMP-2/MMP-14 (Fernandez-Catalan et al., 1998), puis de tous les autres complexes TIMP-1 ou -2/MMP.

L'étude des complexes a montré que le TIMP se lie au site actif du domaine catalytique de la MMP par l'intermédiaire du segment N-terminal fixé de manière rigide au reste de la molécule par les deux ponts disulfures engageant les Cys 1 et 3 (Cys 1-Cys 70 et Cys 3-Cys 99 pour le

TIMP-1 ; Cys 1-Cys 72 et Cys 3-Cys 101 pour le TIMP-2). Ce segment N-terminal ainsi qu'une partie de la boucle CD adjacente participent aux ¾ des contacts intermoléculaires entre le TIMP et la MMP (Gomis-Rüth et al., 1997). Le domaine N-terminal étant très conservé, les contacts établis au niveau de cette région sont communs à tous les complexes TIMP/MMP.

- Motif fonctionnel de liaison TIMP/MMP

Ces observations indiquent que les 4 premiers résidus Cys 1 – Thr 2/Ser 2 – Cys 3 – Val 4/Ser 4/Ala 4 se lient au site actif de la MMP au niveau des sous-sites S_1, S_1', S_2' et S_3' respectivement, d'une manière similaire aux résidus P_1, P_1', P_2' et P_3' d'un substrat peptidique. Le résidu Cys 1 joue un rôle primordial dans l'interaction avec les MMPs. En effet, ce résidu joue un rôle essentiel dans la chélation de l'atome de zinc catalytique par l'intermédiaire du groupement α-amino N-terminal et du groupement carbonyl de ce résidu. La liaison du TIMP avec la MMP entraîne l'expulsion de la molécule d'eau qui jouait le rôle de 4ème ligand du zinc pour être remplacée par le résidu Cys 1. Le groupement α-amino N-terminal de Cys 1 semble également établir une liaison hydrogène avec le Glu catalytique. La rupture des diverses interactions établies par le résidu Cys 1, par l'ajout d'un résidu avant la Cys 1 (Wingfield et al., 1999), par carbamylation (Higashi et Miyazaki, 1999), ou par acétylation (Troeberg et al., 2002) du groupement α-amino abolit l'activité inhibitrice des TIMPs pour les MMPs.
Le résidu en position 2 qui, selon les TIMPs, peut être une sérine ou thréonine ne remplit pas complètement la cavité S_1' des MMPs, mais établit une liaison hydrogène avec le Glu catalytique. En particulier, la chaine latérale de Thr2 du TIMP-1 pointe vers le sous-site de spécificité S_1' et son groupement β-méthyl s'insère dans la niche hydrophobe tandis que son groupement β-hydroxyl est dirigé vers l'entrée plus polaire de cette cavité. Dans les substrats des MMPs, le résidu présent dans cette position (P_1') joue un rôle déterminant dans la spécificité des substrats peptidiques. De manière similaire, le résidu 2 des TIMPs joue un rôle prépondérant à la fois dans la spécificité et la sélectivité des TIMPs envers les MMPs. En effet, des mutants de N-TIMP-1 ont été construits avec 14 substitutions différentes à la place de Thr2 et testés pour leur affinité de liaison pour les MMP-1, -2 et 3 (Meng et al., 1999). Les résultats ont montré que la suppression de la chaine latérale du résidu 2 en substituant Thr par Gly produit un variant de TIMP inactif envers les MMPs testées. Les substitutions par des résidus Ala et Asp entrainent une perte importante d'affinité sur toutes les MMPs testées, d'un facteur 70 à 700 avec Ala et d'un facteur 600 à 3000 avec Asp. En revanche, les meilleurs variants

obtenus pour chacune des 3 MMPs ont montré une certaine sélectivité puisque les résidus préférés en position 2 sont différents en fonction de la MMP testée : Val, Leu et Ser pour la MMP-1, -2 et -3 respectivement.

- Régions d'interaction adjacentes au motif fonctionnel

La boucle CD, adjacente aux résidus N-terminaux et délimitée à droite par le 1er pont disulfure, est formée de 4 résidus et se trouve en contact direct avec la MMP. En particulier, les deux résidus en amont du pont disulfure, Val/Leu et Ser/Ala (positions 69 et 68 dans le TIMP-1) interagissent avec les sous-sites non prime S_2 et S_3 des MMPs. L'importance de la boucle CD du TIMP dans l'interaction avec la MMP a été mise en évidence grâce à des expériences de mutagénèse réalisées sur le domaine N-terminal du TIMP-1. Les travaux de Huang et al. ont montré que les mutations individuelles M66A et V69T situées au niveau de la boucle CD (constituée des résidus $M_{66}E_{67}S_{68}V_{69}$) dans le N-TIMP-1 entrainaient une perte d'affinité pour la MMP-3 d'un facteur 20 et 10 respectivement (Huang et al., 1997). Par ailleurs, le résidu Val en position 69 dans le TIMP-1 est un résidu Val ou Leu dans tous les TIMPs, suggérant la présence d'un contact hydrophobe très conservé avec la MMP.

La boucle AB varie en taille et en position selon les TIMPs. Ces boucles jouent donc des rôles différents selon les TIMPs dans les interactions avec les MMPs. Dans le TIMP-2, la boucle AB établit plusieurs contacts avec la boucle MT spécifique des MT-MMPs par le biais des résidus Ile35, Tyr36 et Asn38. Cependant, en dépit de la grande surface d'interaction entre cette boucle et la MMP, la plupart des contacts réalisés ne semblent pas conférer une complémentarité optimale. Seul le résidu Tyr36 semble former une interaction bien définie avec une poche hydrophobe située en dessous de la boucle MT de la MT1-MMP (Williamson et al., 2001). Par ailleurs, les contacts intermoléculaires présents au niveau de l'interface entre la boucle AB du TIMP-2 et la MMP-14 et absents dans le complexe TIMP-1/MMP-3 semblent être compensés au niveau d'autres sites de ce complexe TIMP-1/MMP-3 puisque le nombre total de contacts intermoléculaires à une distance inférieure à 4 Å est le même dans les deux complexes (Fernandez-Catalan et al., 1998).

Le domaine C-terminal contribue très peu dans l'interaction entre le TIMP et le domaine catalytique des MMPs, il participe à hauteur d'environ 15% de l'ensemble des contacts

intermoléculaires et moins de 10% de l'énergie libre de liaison du complexe (Huang et al., 1996).

IV.4 Conclusion

Les différents membres de la famille des MMPs représentent des cibles intéressantes pour l'application de notre approche de conception *in silico* de ligands protéiques artificiels. En effet, la complexité de la famille des MMPs résulte d'un nombre important de membres (23 membres chez l'homme) caractérisés par une grande similarité de structure et de séquence des domaines catalytiques. Bien que des succès importants aient été obtenus concernant l'inhibition sélective de la MMP-12 (Devel et al., 2006) et de la MMP-13 (Engel et al., 2005), l'exploration de nouvelles méthodes permettant d'aborder l'inhibition sélective d'un membre particulier de la famille des MMPs pourrait déboucher sur de nouvelles molécules de grand intérêt thérapeutique ou diagnostique. Dans cette perspective, la conception *in silico* de mini-protéines capables d'interagir au niveau du site catalytique des MMPs et présentant une surface d'interaction importante pourrait constituer une étape importante vers cet objectif. En effet ces molécules pourraient servir de point de départ pour réaliser des bibliothèques de variants sur des positions présentes à l'interface avec les MMPs.

L'objectif de ce travail est de concevoir une série de ligands protéiques capables de cibler le site catalytique des MMPs à l'aide d'une approche de conception *in silico* par greffage de motif fonctionnel. Ce motif fonctionnel peut être déduit d'informations fonctionnelles et structurales disponibles sur les inhibiteurs naturels des MMPs, les TIMPs. Les caractéristiques de ces interactions exploitées pour la conception de ligands protéiques artificiels des MMPs au cours de ce travail sont :

- La conformation des 4 premiers résidus du TIMP qui permettent d'établir un ensemble d'interactions intermoléculaires mettant en jeu la chaine principale de ces quatre résidus.
- L'implication du groupement N-terminal de la plate-forme dans une interaction directe avec l'atome de zinc dit « catalytique ».
- La présence des ponts disulfures à l'extrémité N-terminale conférant à celle-ci une rigidité structurale.

Du point de vue méthodologique, un objectif important de ce travail est de mettre en évidence le taux de succès de l'approche *in silico* et sa capacité à produire des ligands protéiques artificiels capables de cibler un site de liaison précis. En effet, si les méthodes d'optimisation expérimentales ont pu démontrer dans quelques applications leur capacité à conduire à un ligand de haute affinité à partir de composés initiaux d'affinité modeste, l'obtention de ce composé reste l'étape limitante de la démarche. De plus, le fait de pouvoir disposer d'une série de composés initiaux pourrait accroitre considérablement les chances de succès de la méthode. Enfin, cette méthode de conception peut être un outil intéressant pour une meilleure compréhension du phénomène d'association protéine-protéine. De ce point de vue, on peut noter que les quelques études de conception de ligands protéiques montrent que les affinités atteintes par les composés initiaux issus de la phase *in silico* sont le plus souvent faibles (entre 1 μM et 1 mM) suggérant qu'une partie des caractéristiques conduisant à des ligands protéiques de haute affinité n'a pas encore été correctement identifiée. Les travaux d'ingénierie mêlant des aspects structuraux, expérimentaux et la simulation des interactions entre ces partenaires constituent une voie d'intérêt pour progresser dans la compréhension de ces déterminants et ouvrira sans doute la voie à des méthodes d'ingénierie plus efficaces.

MATERIELS ET METHODES

MATERIEL

Tous les solvants et réactifs sont utilisés tels quels sans purification supplémentaire. Les acides aminés utilisés lors des synthèses sont commercialisés chez Novabiochem. Les acides aminés non naturels Fmoc-Homo-L-phenylalanine et Fmoc-L-4,4'-Biphenylalanine sont commercialisés par Advanced ChemTech. L'acide aminé Fmoc-4-(phenoxy)-L-phenylalanine est commercialisé par PolyPeptide GROUP et l'acide aminé Fmoc-cyclohexyl-methoxy-phenylalanine a été synthétisé au laboratoire. Les cartouches utilisées afin de peser les acides aminés sont fournies par Applied Biosystems. La résine utilisée pour les synthèses est de type Rink amide MBHA, commercialisée par Novabiochem. Le TFA, le TIS, le Ac_2O, le DCCI, le méthanol sont des produits Aldrich. Le Cl-HOBt est commercialisé par Molekula, et la collidine par Fluka. Le DCM et le NMP proviennent de Rectapur, la pipéridine provient de SDS. L'acétonitrile pour les purifications en HPLC est commercialisé par VWR PROLABO.

Concernant l'activité de criblage de venin de serpent, le matériel biologique étudié consiste en 1 g de venin du serpent *Dendroaspis polylepis,* commercialisé sous forme lyophilisée par Sigma Aldrich. Le fractionnement du venin de serpent *Dendroaspis polylepis* est réalisé grâce à un système constitué d'une colonne d'échange d'ions à basse pression contenant une phase 15 S (Pharmacia), dont le volume est de 40 ml, et utilisée à un débit de 3 ml/min et une pression de 3-4 bar (45-58 psi). La colonne est reliée à un formateur de gradient avec deux compartiments communiquants pour réaliser les mélanges de tampons A et B, grâce à une pompe péristaltique permettant de régler le débit. La sortie de la colonne est reliée à un collecteur de fractions automatique contenant 95 tubes et permettant de régler le volume souhaité des fractions recueillies. La séparation des différents composés contenus dans le venin se fait grâce à un gradient de tampon B (Acétate d'ammonium 3M, pH 7) dans un tampon A (Acide acétique 10 mM, pH 2). Le rinçage final de la colonne se fait par un tampon C (Acétate d'ammonium 3M, pH 9,5).

La purification par RP-HPLC des fractions mères de venin issues du 1ère fractionnement par chromatographie d'échange d'ions est réalisée en utilisant des colonnes RP-HPLC en phase C18 (Discovery Supelco) : une colonne analytique (Dimensions 250 x 4,6 mm, 10 µm, débit 1 ml/min), une colonne semi-préparative (Dimensions 250 x 10 mm, 5 µm, débit 4,5 ml/min) et une colonne préparative (Dimensions 250 x 21,2 mm, 10 µm, débit 8 ml/min). Les solvants utilisés pour la purification sont le solvant A (H_2O/ACN/TFA : 90/10/0,1%) et le solvant B

(H_2O/ACN/TFA : 10/90/0,1%).

Les synthèses sont réalisées grâce à un synthétiseur automatisé modèle 431A Peptide Synthetizer, et le suivi de synthèse se fait par détection en UV sur le détecteur d'absorbance programmable 785A. Ces appareils sont commercialisés par Applied Biosystems.
Les séparations par HPLC en phase inverse, analytiques, semi-préparatives et préparatives ont été réalisées sur les appareils Thermo Separation Product et Gilson utilisant respectivement des colonnes AIT C18 Kromasil (100 Å, 250x4,6 mm, débit 1 mL/min), AIT C18 Kromasil (100 Å, 250x10 mm, débit 3 mL/min) et AIT C18 Kromasil (100 Å, 250x20 mm, débit 8 mL/min). Les solvants d'élutions sont les suivants : A = H_2O/ACN/TFA : 95%/5%/0,1% et B = ACN/H_2O/TFA : 90%/10%/0,09%. Les produits issus des HPLC sont lyophilisés grâce à un lyophilisateur Bioblock Scientific.

Chaque peptide synthétisé et purifié a été analysé par spectrométrie de masse. Le HCCA contenu dans la matrice organique est commercialisé par Bruker Daltonics. Les peptides synthétiques sont déposés sur une plaque Opti-TOF^{TM} (123 x 81 mm) de 384 puits (Applied Biosystems) puis analysés en masse grâce à un MALDI-TOF/TOF 4800 (Applied Biosystems).

Pour les protéines produites par biologie moléculaire, les amorces sont fournies par Sigma Aldrich. Les amplifications par PCR de l'ADN ont été réalisées grâce à l'ADN polymérase Phusion commercialisée par New England Biolabs (NEB). Les enzymes de restriction FastDigest utilisées lors des différents clonages sont commercialisées par Fermentas et ont été utilisées selon les instructions du fournisseur avec les tampons FastDigest fournis. Les gènes codant pour les mini-protéines d'intérêt sont synthétisés par Geneart A.G. Les clonages sont réalisés dans la souche *E. coli* DH5α fournie par Invitrogen. L'expression des protéines recombinantes est réalisée dans la souche *E. coli* BL21 Star (DE3) fournie par Invitrogen. Le vecteur d'expression utilisé pour la production des protéines recombinantes est le pETG-82A comportant un promoteur T7 inductible à l'IPTG, un gène de résistance à l'ampicilline et un gène codant pour la protéine DsbC qui sert de partenaire de fusion solubilisant. Le milieu de culture LB (Luria Bertani) est commercialisé sous forme lyophilisée par Sigma Aldrich et préparé en dissolvant 25 g de LB dans 1 L d'eau milliQ. Le milieu ZY est préparé en pesant 10 g de N-Z-amine AS (Sigma Aldrich) et 5 g d'extrait de levure (Sigma Aldrich) dans 925 mL d'eau milliQ puis autoclavé. Au moment de l'utilisation, le milieu

complet est reconstitué après autoclavage par ajout de 50 mL d'une solution NPS 20X (0,5M $(NH_4)_2SO_4$; 1M KH_2PO_4 ; 1M Na_2HPO_4 ; pH 6,75), 20 mL de glycérol 50X, 1 mL de $MgSO_4$ à 1M, 1 mL d'ampicilline à 100 mg/mL et 1 mL de chloramphenicol à 30 mg/mL. La TEV protéase a été produite au laboratoire.

Pour les tests enzymatiques, les MMPs testées correspondent aux formes humaines. Les MMP-12h et MMP-9h, domaines catalytiques, sont produits au laboratoire. Les MMP-2, MMP-8 et MMP-13, formes entières, sont commercialisés par R&D Systems. La MMP-14 forme entière est commercialisée par CIRMMP. Le substrat fluorogénique McaMat (Mca-Pro-Leu-Gly-Leu-Dpa-Ala-ArgNH_2) est commercialisé par la société Calcobiochem et solubilisé dans du DMSO. Les tests enzymatiques sont réalisés dans un tampon 50 mM Tris-HCl pH 6,8, 10 mM $CaCl_2$. Le Tris est commercialisé par Merck et le $CaCl_2$ est commercialisé par Sigma. Les tests enzymatiques *in vitro* sont réalisés en plaques 96 puits (plaques 3650, Corning Costar, France). La cinétique de dégradation des substrats est suivie en fluorescence à l'aide d'un spectrophotomètre (Fluoroskan Ascent, ThermolabSystems), équipé d'un agitateur de plaques.
Les protéines natives endothelin 1 et 3 sont fournies par Sigma-Aldrich; la scyllatoxin ainsi que la psalmotoxin sont commercialisées par Peptanova ; la SHK toxin est fournie par Bachem.

Les analyses en LC/MS sont réalisées sur micro-HPLC (pré-colonne Phenomenex Jupiter C18-Max-RP, 4 µm, 80 Å, 2,0x4,0 mm ; colonne Phenomenex Jupiter C18, 4 µm, 80 Å, 2,1x150 mm, débit 0,2 mL/min) couplée à un spectromètre de masse à trappe ionique (Esquire HCT, Bruker Daltonics, Germany). Les solvants d'élution utilisés sont les suivants : A = H_2O/0,1% HCOOH, B = CH_3CN/H_2O/HCOOH : 90%/10%/0,1%.

METHODES

I. Recherche de protéines inhibitrices de MMPs dans le venin du serpent *Dendroaspis polylepis*

Une quantité de 1 g de venin du serpent *Dendroaspis polylepis* est dissoute dans 300 ml d'acide acétique 5 mM à pH 2, puis filtrée sur un filtre 0,45 µm. Le fractionnement de ce venin a été réalisé grâce à une stratégie de fractionnement bi-dimensionnelle.

I.1 Fractionnement du venin par chromatographie d'échange d'ions

- Protocole de fractionnement

La colonne d'échange d'ions utilisée pour le 1er fractionnement du venin du serpent *Dendroaspis polylepis* est équilibrée avec 10 volumes de colonne (400 ml) en Tampon A (Acide acétique 10 mM). Par la suite, l'échantillon de venin est chargé sur la colonne à un débit de 3 ml/min. La colonne ainsi chargée est lavée avec 3 volumes de colonne (120 ml) en Tampon A. L'éluat issu de cette étape de lavage (Dp1A) est gardé pour une analyse ultérieure. La séparation proprement dite se fait grâce à un gradient non linéaire en tampon B dans du tampon A. Pour ce faire, la colonne est placée dans un système semi-automatique avec un formateur de gradient constitué de 2 compartiments cylindriques, celui de droite étant muni d'un agitateur. A chaque palier, le compartiment de droite est rempli par le mélange avec le faible % en B alors que celui de gauche contient le mélange avec le % final de B pour ce palier. Entre les 2 compartiments, un conduit contrôlé par une vanne permet de faire passer la solution la plus concentrée en B vers la solution la moins concentrée.

Le gradient non linéaire utilisé est le suivant :

- 0 à 10% B en 2 volumes
- 10 à 40% B en 20 volumes
- 40 à 100% B en 3 volumes
- 100% B en 1 volume

L'éluat en sortie de colonne est récupéré grâce à un collecteur de fractions, à raison de 6 ml/tube. En fin de fractionnement, la colonne est lavée avec 2 volumes de colonne (80 ml) en tampon C ; cette fraction de rinçage est récupérée dans un tube contenant de l'acide acétique 5 mM à pH 2 afin d'en diminuer le pH.

- Traitement des fractions après chromatographie d'échange d'ions

Les différentes fractions obtenues sont soumises à un dosage UV à 280 nm. Un chromatogramme représentant la DO à 280 nm des différentes fractions est tracé, puis les fractions sont regroupées en fonction des pics obtenus. Les fractions ainsi regroupées sont numérotées (Dp1A, Dp1B, …) puis diluées par 3 dans de l'eau milliQ et mises à lyophiliser.
Les fractions sont ensuite reprises dans 3 à 4 ml d'eau milliQ puis la DO à 280 nm des solutions ainsi obtenues est mesurée et la quantité de protéines totales contenues dans chaque fraction est déterminée.

I.2 Purification par RP-HPLC des fractions issues de la chromatographie d'échange d'ions

Les fractions mères issues du 1er fractionnement par chromatographie d'échange d'ions sont soumises à une 2ème purification par RP-HPLC. En fonction de la quantité de protéines totales contenues dans chaque fraction, des colonnes RP-HPLC analytique, semi-préparative ou préparative ont été utilisées. Une quantité correspondant au 1/3 de la fraction mère est soumise à la purification par RP-HPLC, le reste est gardé pour réaliser les tests d'inhibition sur les MMPs. Après équilibration de la colonne avec le solvant A, l'échantillon est acidifié par du TFA 20% puis injecté sur la colonne. La séparation se fait selon le gradient suivant :

- 15 min en 100% A (dessalage)
- 50 min de 0 à 50% B
- 25 min de 50 à 100% B
- 10 min en 100% B (nettoyage de la colonne)

Le suivi de la purification se fait en mesurant la DO en sortie de colonne à 280 nm et les différents pics de protéines sont récupérés séparément. Les différentes fractions filles obtenues sont numérotées (Par exemple, pour la fraction mère Dp1A, les fractions filles sont notées Dp1A1, Dp1A2 …) puis mises à lyophiliser. Elles sont ensuite reprises dans 3 à 4 ml d'eau milliQ puis la DO à 280 nm des solutions ainsi obtenues est mesurée et la quantité de protéines totales contenues dans chaque fraction est déterminée.

I.3 Criblage des fractions sur les MMPs

Les différentes fractions mères et filles issues du 1er fractionnement par chromatographie d'échange d'ions et du 2ème fractionnement par RP-HPLC respectivement sont soumises à un criblage sur la MMP-12h à la recherche d'un composé inhibiteur de l'activité de cette enzyme. Le protocole est détaillé dans le paragraphe VIII de cette section.

II. Recherche *in silico* de plates-formes par le programme STAMPS

L'objectif de notre travail est de concevoir des ligands protéiques inhibiteurs des MMPs, par greffage d'un motif fonctionnel sur des plates-formes capables de reproduire les caractéristiques structurales d'un motif présent dans des inhibiteurs naturels des MMPs : les TIMPs.

La recherche des plates-formes protéiques a été réalisée par le programme STAMPS version 3.1.1 développé au sein de notre laboratoire. Ce programme permet d'identifier dans la PDB des protéines possédant des résidus topologiquement équivalents aux résidus formant le motif fonctionnel à greffer. La base de données criblée par le programme STAMPS est la version entière de la PDB. Le motif fonctionnel choisi correspond aux 4 résidus N-terminaux Cys_1-Ser_2-Cys_3-Ser_4 du TIMP-2. Les coordonnées x, y, z du motif fonctionnel de référence ainsi que les coordonnées x, y, z de la cible de référence sont issues du fichier PDB 1bqq correspondant au complexe TIMP-2/MMP-14.

II.1 Données nécessaires à la recherche

- Les coordonnées cartésiennes de la cible MMP-14 ;
- l'identité ainsi que les coordonnées cartésiennes des 4 résidus du motif fonctionnel du ligand TIMP-2 : Cys-Ser-Cys-Ser ;
- un lien permettant l'accès à la PDB.

II.2 Paramètres de recherche

Les paramètres utilisés en entrée de STAMPS pour la recherche de plates-formes ayant

aboutit aux solutions qui seront reportées dans ce mémoire sont :

- Motif de liaison :

Le motif de liaison doit être présent parmi les **5** résidus N-terminaux de la séquence protéique. Si les résidus topologiquement équivalents à ceux du motif fonctionnel se trouvent en positions 2-5 de la plate-forme native, la protéine sera produite avec une délétion du résidu en position 1. STAMPS offre la possibilité de restreindre la nature des résidus natifs en position équivalente à l'un ou l'autre des résidus du motif. Plusieurs recherches ont été conduites avec les combinaisons suivantes : le motif fonctionnel sélectionné est constitué de **4 résidus** et doit contenir un résidu **Cys** en position 1 et/ou en position 3 du motif de la chaine principale du ligand artificiel: **C**X**C**X/**C**XXX/XX**C**X.

Le motif de liaison doit également satisfaire des critères géométriques caractéristiques des atomes Cα et Cβ des résidus du motif fonctionnel de référence :

- Distances interatomiques : les distances entre les atomes Cα et Cβ d'une paire de résidus du motif mime (d_1', d_2', d_3' …) sont comparées aux distances entre les atomes Cα et Cβ d'une paire de résidus du motif de référence (d_1, d_2, d_3 …). La différence entre chacune des distances d et d' ne doit pas dépasser une valeur seuil **Δdmax** fixée à **1,85 Å**.

- RMSD : l'écart moyen (RMSD) sur la position des atomes Cα et Cβ entre les 4 résidus du motif de référence et les 4 résidus du motif mime après superposition ne doit pas dépasser une valeur seuil **RMSDmax** fixée à **1,25 Å.**

- Déviations angulaires : l'écart maximum toléré sur les angles Θ entre un vecteur Cα-Cβ d'un résidu du motif de référence et le vecteur Cα-Cβ du résidu correspondant dans le motif mime après superposition est de 130° (**Θmax = 130°)** ; l'écart moyen sur tous les angles Θ entre les 4 résidus du motif de référence et les 4 résidus du motif mime $\Delta\Theta_{max}$ est limité à **90°.**

- Taille des plates-formes : la taille des plates-formes criblées doit être comprise entre **15 et 60 acides aminés**.

- Filtre stérique : un facteur de recouvrement stérique est calculé afin d'évaluer la compatibilité stérique entre la cible MMP-14 et la plate-forme analysée. Ce facteur stérique est défini par : $\frac{1}{n}\sum_j \sum_i (r_i + r_j - d_{ij})^2$ où r_i et r_j sont les rayons des atomes i et j séparés par une distance d_{ij}. Chaque terme est calculé si $d_{ij} < r_i + r_j$. Le facteur stérique correspond donc au nombre d'atomes de la cible dans une situation de recouvrement stérique avec une plate-forme

sur le nombre total des atomes de la cible présents à l'interface. Deux scores stériques sont ainsi calculés. Un score stérique intermoléculaire est évalué par le critère **Steric-inter** qui doit être inférieur à **2,5**. L'encombrement stérique intramoléculaire, résultant du transfert des résidus spécifiques du motif de référence sur la plate-forme analysée, est évalué par le critère **Steric-intra** qui doit être inférieur à **1**.

Tableau II.1: Critères de recherche utilisés dans le programme STAMPS pour la recherche dans la PDB de plates-formes mimes capables d'inhiber les MMPs.

Critères	**Valeurs**
Emplacement du motif :	**5** premiers résidus de l'extrémité N-terminale
Composition du motif :	
- nombre de résidus	**4 résidus**
- nature des résidus	**Cys-X-Cys-X / Cys-X-X-X / X-X-Cys-X**
équivalents	X = n'importe quel acide aminé sur les 20 acides aminés naturels
Topologie du motif :	
- **$RMSD_{max}$ (Å)**	**1,25**
- **Δd_{max} (Å)**	**1,85**
- **Θ_{max} (°)**	**130**
- **$\Delta\Theta_{max}$ (°)**	**90**
Filtre stérique :	
- Steric-inter	**2,5**
- Steric-intra	**1**
Taille des plates-formes (T)	**$15 \leq T \leq 60$ acides aminés**

III. Analyse électrostatique des plates-formes

La compatibilité électrostatique des plates-formes sélectionnées dans la PDB grâce au programme STAMPS est vérifiée.

III.1 Calcul du potentiel électrostatique des plates-formes mimes et du ligand de référence

Le programme STAMPS produit des fichiers de structures de complexes entre le domaine catalytique de la MMP-14 et chacun des ligands satisfaisant les critères définis ci-dessus. Ces structures sont obtenues après superposition des atomes Cα, Cβ des résidus du motif identifié dans la plate-forme sur les atomes correspondants dans le complexe de référence TIMP-2/MMP-14. Après mutation des résidus de la plate-forme topologiquement équivalents à ceux du motif (greffe du motif), les coordonnées cartésiennes de la MMP-14 en complexe avec ce candidat ligand sont sauvegardées. Il faut noter que dans les motifs de liaison des plates-formes, les résidus Cys natifs en position 1 et/ou 3 du motif sont conservés s'ils sont impliqués dans un pont disulfure naturel. Si ce n'est pas le cas, les cystéines sont mutées en serine. Enfin les résidus en positions équivalentes à 2 et 4 dans le motif à greffer sont mutés en serine. Le protocole décrit ci-dessus permet d'obtenir une série de structures de candidats ligands dans lesquelles les motifs greffés sont superposés au motif présent dans la structure de référence MMP-14/TIMP-2. Les positions des atomes d'hydrogène manquants dans la structure sont reconstruites à l'aide du programme CHARMM et les coordonnées cartésiennes des complexes plateforme/MMP-14 sont sauvegardées. L'analyse électrostatique est réalisée à l'aide des coordonnées cartésiennes des candidats ligands protéiques extraits de ces derniers complexes.

Les potentiels électrostatiques des candidats ligands obtenus ainsi que celui du TIMP-2 sont calculés grâce au logiciel **APBS** version 1.3 (Baker et al., 2001), en résolvant numériquement l'équation de Poisson Boltzmann. Le modèle de Poisson Boltzmann permet de décrire le potentiel électrostatique d'une protéine traitée comme un milieu de faible constante diélectrique, et présente dans un milieu aqueux à forte constante diélectrique. Les charges et les rayons des atomes utilisés dans les calculs sont ceux du jeu de paramètres PARSE (Sitkoff et al., 1994). Les constantes diélectriques ε utilisées sont de 1 pour l'intérieur des protéines et de 80 pour le solvant. La force ionique est de 150 mM, le rayon des ions est de 2 Å et le rayon des molécules du solvant est de 1,4 Å. Les dimensions de la grille pour le calcul du potentiel électrostatique sont choisies de façon à englober toutes les plates-formes sélectionnées : 97 points pour chacun des axes x, y et z, avec un espacement de 2,0 Å entre les points pour la grille initiale et de 0,8 Å pour la grille finale.

III.2 Comparaison des potentiels électrostatiques entre les plates-formes mimes et le ligand de référence

Les grilles de potentiel électrostatique générées pour chacune des plates-formes ainsi que pour le ligand de référence ont été comparées quantitativement à l'aide du programme **PIPSA** version 2.0 (Blomberg et al., 1999). Ce programme permet de comparer les potentiels électrostatiques des plates-formes avec celui du TIMP-2 en calculant **l'indice de similarité électrostatique de Hodgkin** $ESI^{H}_{a,b}$ (ESI pour « Electrostatic Similarity Index ») permettant la comparaison de la grille de potentiel électrostatique d'une plate-forme mime avec celle du TIMP-2. Cet indice est défini par :

$$ESI^{H}_{a,b} = 2(M_a, M_b)/(M_a^2 + M_b^2)$$

où M représente le potentiel moléculaire à comparer.
Dans le cas de la comparaison des potentiel électrostatiques Φ_a et Φ_b :

$$(M_a, M_b) = \sum_{i,j,k} \Phi_a(i,j,k)\Phi_b(i,j,k)$$

La valeur de ESI varie de -1 à +1, une valeur de +1 signifiant que la plate-forme et le TIMP-2 ont un potentiel électrostatique identique, tandis qu'une valeur de -1 indique que les potentiels électrostatiques sont anti-corrélés. Un ESI = 0 correspond à des protéines ayant des potentiels électrostatiques non corrélés.

La comparaison entre les grilles de potentiels électrostatiques se fait au niveau d'une région précise de la grille, appelée « peau » (skin). Cette région correspond à un volume formé autour de la protéine, situé à une distance σ de 3 Å de la surface de la protéine et ayant une épaisseur δ de 4 Å (Figure III.1). Ainsi, la comparaison des potentiels électrostatiques de deux protéines se fait à l'intersection de leurs régions « skin ». L'analyse de l'indice ESI peut être réalisée sur la totalité des régions « skin » ce qui donne un indice ESI_all. L'analyse peut également être restreinte à une partie bien définie de la région « skin », correspondant au site de liaison du ligand mime avec la cible, et donnant un ESI_part. Cette région est définie comme étant l'intersection d'une sphère de rayon 10 Å et centrée sur l'atome N du résidu en position 1 du motif de liaison, avec la région « skin » (Figure III.1).

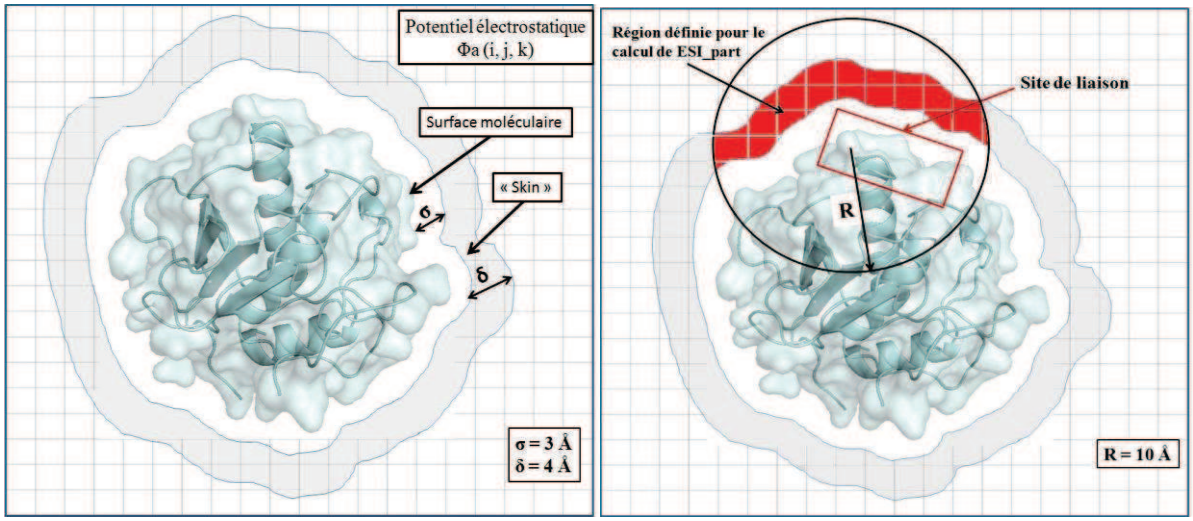


Figure III.1 : Définition des régions utilisées pour le calcul des indices de similarité électrostatique de Hodgkin. La protéine est représentée en mode surface et cartoon. A gauche : L'ensemble de la région « skin » coloriée en gris utilisée pour le calcul des ESI_all. A droite : Le calcul des ESI_part est limité à la région de liaison coloriée en rouge.

III.3 Comparaison des formes des plates-formes mimes et du ligand de référence

Les formes d'une plate-forme mime et du ligand de référence peuvent également être comparées en calculant un **indice de similarité de forme de Hodgkin** $SSI_{a,b}^{H}$ (SSI pour « Shape Similarity Index ») défini par :

$$SSI_{a,b}^{H} = 2N_{iab}/(N_a + N_b)$$

Où N_a, N_b et N_{iab} représentent respectivement le nombre de points présents sur la grille dans la région « skin » autour de la protéine a, le nombre de points présents sur la grille dans la région « skin » autour de la protéine b et le nombre de points présents dans l'intersection des deux grilles a et b. La valeur de cet indice varie de 0, lorsque les formes des protéines sont différentes, à 1 lorsqu'elles sont identiques. De manière similaire aux indices ESI_all et ESI_part, l'indice SSI_all est calculé pour l'ensemble de la région « skin » tandis que l'indice SSI_part est limité à la région de liaison telle que définie pour le ESI_part.

IV. Procédure de relaxation par dynamique moléculaire et minimisation d'énergie

Les structures des complexes plates-formes mimes/MMP-14 issus du programme

STAMPS ont été soumises à une procédure de relaxation par dynamique moléculaire et minimisation d'énergie sous contraintes positionnelles en utilisant le logiciel CHARMM version 35b1 (Brooks et al., 1983). Le champ de force CHARMM version 22 a été utilisé (MacKerell et al., 1998). Après reconstruction des atomes d'hydrogène manquants dans les fichiers PDB, la structure du complexe plate-forme mime/MMP-14 est minimisée en utilisant des contraintes positionnelles. Au cours des calculs de simulations, le groupement N-terminal du TIMP-2 et des candidats ligands a été défini sous forme neutre NH_2 et non sous forme NH_3^+. Ceci est justifié par la proximité de l'atome de zinc catalytique portant une charge nette +2e. Des calculs de simulations de dynamique moléculaire réalisés au cours de ce travail ont montré que l'état de protonation NH_3^+ conduisait à une déformation importante de la structure du complexe et à la rupture d'interactions observées expérimentalement entre le ligand protéique et le domaine catalytique de la MMP-14.

IV.1 Minimisation d'énergie sous contraintes

- Contraintes positionnelles :

- **Motif** : des contraintes harmoniques sont imposées sur les atomes autres que les Cα et les hydrogènes, avec une constante de force égale à 5 kcal/mol/Å^2.
- **Cible et ions** : des contraintes harmoniques sont appliquées aux positions des atomes Cα de la cible ainsi que des ions (zinc et calcium) présents dans la structure de la MMP-14 avec une constante de force égale à 50 kcal/mol/Å^2 à l'exception des résidus de la cible situés à une distance inférieure à 7 Å de la plate-forme et pour lesquels la constante de force imposée est de 0,5 kcal/mol/Å^2. Ce jeu de contraintes positionnelles empêchent la déformation de la structure de la cible au cours de la minimisation mais permettent des ajustements plus importants au voisinage de l'interface avec le candidat ligand.
- **Ligand** : des contraintes harmoniques faibles sont imposées sur les atomes Cα du ligand situés à une distance inférieure à 7 Å de la cible, avec une constante de force égale à 0,5 kcal/mol/Å^2 ; pour les autres atomes (non hydrogène) du candidat ligand situés à une distance inférieure à 7 Å de la cible, la constante de force imposée est de 0,005 kcal/mol/Å^2. Aucune contrainte positionnelle harmonique n'est appliquée aux atomes du candidat ligand situés à une distance supérieure à 7 Å de la cible.

- Contraintes de distances :

Deux types de contraintes de distances interatomiques ont été utilisés. Le potentiel associé est défini grâce à la fonction NOE (Nuclear Overhauser Effect) de CHARMM. La constante de force du potentiel associé (KMIN, KMAX) était de 50 kcal mol^{-1} $Å^{-2}$.

- Motif : des contraintes de distances intermoléculaires sont imposées entre les 4 résidus du motif du ligand mime et certains résidus de la cible qui se trouvent à proximité du motif. Ces contraintes de distance sont égales aux distances entre les atomes des 4 résidus du motif fonctionnel du TIMP-2 et les atomes des résidus de la MMP-14, impliqués dans des liaisons intermoléculaires. Ces contraintes de distance permettent de préserver la géométrie du motif du candidat ligand par rapport au site actif de la cible. Les distances de référence correspondent à celles déterminées dans le complexe TIMP-2/MMP-14. Le fichier de contraintes NOE utilisées pour le motif est reporté ci-dessous :

ASSIgn sele atom @{SEG_LIG} @{RES_M**1**} **N** end sele atom IONS **1 ZN** end - KMIN 50.0 RMIN **4.0** KMAX 50.0 RMAX **4.2** FMAX 2.0
ASSIgn sele atom @{SEG_LIG} @{RES_M**1**} **N** end sele atom @{SEG_TAR} **240 CD** end - KMIN 50.0 RMIN **3.1** KMAX 50.0 RMAX **3.3** FMAX 2.0
ASSIgn sele atom @{SEG_LIG} @{RES_M**1**} **O** end sele atom IONS **1 ZN** end - KMIN 50.0 RMIN **2.7** KMAX 50.0 RMAX **2.9** FMAX 2.0
ASSIgn sele atom @{SEG_LIG} @{RES_M**2**} **O** end sele atom @{SEG_TAR} **240 CD** end - KMIN 50.0 RMIN **3.5** KMAX 50.0 RMAX **3.7** FMAX 2.0
ASSIgn sele atom @{SEG_LIG} @{RES_M**2**} **N** end sele atom @{SEG_TAR} **200 O** end - KMIN 50.0 RMIN **3.5** KMAX 50.0 RMAX **3.7** FMAX 2.0
ASSIgn sele atom @{SEG_LIG} @{RES_M**2**} **O** end sele atom @{SEG_TAR} **199 N** end - KMIN 50.0 RMIN **3.0** KMAX 50.0 RMAX **3.2** FMAX 2.0
ASSIgn sele atom @{SEG_LIG} @{RES_M**3**} **N** end sele atom @{SEG_TAR} **259 O** end - KMIN 50.0 RMIN **2.8** KMAX 50.0 RMAX **3.0** FMAX 2.0
ASSIgn sele atom @{SEG_LIG} @{RES_M**3**} **O** end sele atom @{SEG_TAR} **261 N** end - KMIN 50.0 RMIN **3.0** KMAX 50.0 RMAX **3.2** FMAX 2.0
ASSIgn sele atom @{SEG_LIG} @{RES_M**4**} **N** end sele atom @{SEG_TAR} **197 O** end - KMIN 50.0 RMIN **2.7** KMAX 50.0 RMAX **2.9** FMAX 2.0

- Ligand : des contraintes de distances intramoléculaires, également définies grâce à la fonction NOE de CHARMM sont imposées entre les atomes Cα du candidat ligand (un ensemble de

contraintes de distance $d_{i,j}$ définie entre l'atome Cα du résidu i et l'atome Cα du résidu j, i et j parcourant la séquence primaire du candidat ligand étudié) afin de préserver la structure tridimensionnelle de la plate-forme étudiée tout en autorisant des mouvements d'ensemble du candidat ligand par rapport à la cible. Ces contraintes permettent d'optimiser l'interface du ligand et de la cible au cours de cette phase de relaxation. La structure de référence pour la définition des contraintes de distance est la structure initiale de la plate-forme identifiée dans la PDB.

IV.2 Dynamique moléculaire dans le vide

Chaque structure de complexe candidat ligand/MMP-14 a été soumise à 15 cycles de dynamique moléculaire à 300 K, correspondant à 25 000 pas par cycle et avec un pas d'intégration de 0,0005 ps. Tous les 25 000 pas, l'énergie potentielle de la structure est minimisée et ses coordonnées sont sauvegardées pour une analyse ultérieure. 15 structures sont ainsi obtenues. Ces calculs ont été réalisés dans le vide. Les structures ainsi minimisées sont par la suite analysées afin de déterminer les caractéristiques des interfaces plates-formes mimes/MMP-14.

IV.3 Dynamique moléculaire en solvant explicite

Les calculs ont été réalisés à l'aide du programme CHARMM version C35b1 (Brooks et al., 1983). Le champs de force utilisé est la version 22 (MacKerell et al., 1998). Les coordonnées initiales des complexes sont celles obtenues à l'issue du protocole de conception *in silico*. Les complexes résultant sont immergés dans une boîte d'eau cubique de dimension adaptée à la taille du complexe étudié. Les molécules d'eau en recouvrement avec les atomes de la protéine sont éliminées et l'interface solvant-soluté est optimisée par minimisation d'énergie et dynamique moléculaire. Durant cette phase, une première étape de 300 cycles de minimisation d'énergie est réalisée durant laquelle les atomes du soluté sont fixes. Puis une seconde étape de 300 cycles de minimisation de l'énergie potentielle du système est réalisée en appliquant des contraintes positionnelles harmoniques aux atomes du soluté (k_{harm} = 30 $kcal.mol^{-1}.Å^{-2}$). Ces étapes de minimisation d'énergie sont suivies d'une phase de dynamique moléculaire destinée à thermaliser le système à 298 K. Dans une première étape, cette dynamique est réalisée en présence de contraintes harmoniques (k_{harm} = 5 $kcal.mol^{-1}.Å^{-2}$)

appliquées aux atomes non-hydrogène de la chaine principale, aux ions et aux atomes chélatant les ions en utilisant les coordonnées initiales en référence. Durant cette phase, la température est augmentée graduellement par étapes de 30 K. Puis, lorsque la température de 298 K est atteinte, les contraintes positionnelles sont éliminées et la phase de production est réalisée jusqu'à 1 ns.

Les étapes de dynamique moléculaire sont calculées à Pression et Température constantes. Le pas de temps était de 0,001 ps et les coordonnées sauvegardées tous les 100 pas. Les conditions périodiques aux limites ont été appliquées au système durant le calcul de la trajectoire.

Les interactions électrostatiques ont été traitées en utilisant la méthode du « Particle Mesh Ewald ». Il faut noter que les structures initiales des complexes issus de la conception *in silico* ont été préalablement neutralisées à l'aide d'ions Na^+ ou Cl^-. Le cut-off était de 16 Å.

IV.4 Analyse des interfaces plates-formes mimes/MMP-14

Les structures des complexes plates-formes mimes/MMP-14 sont analysées grâce à une procédure développée au laboratoire et écrite dans le macro-langage du programme CHARMM. La taille de l'interface (en $Å^2$) d'un complexe est calculée comme étant la somme des aires des surfaces totales accessibles des deux protéines séparées moins l'aire de la surface totale accessible du complexe, avec un rayon de sonde de 1,4 Å. Le volume vide ou « gap volume » (en $Å^3$) est calculé comme la différence entre les volumes totaux des deux protéines séparées et celui du complexe. L'indice de volume vide ou « gap volume index » (en Å) est défini comme étant le rapport du volume vide sur la taille de l'interface.

L'attribution des interactions intermoléculaires au niveau des interfaces des complexes se fait selon des critères définis. Pour les liaisons hydrogène (HB), la distance entre le donneur et l'accepteur doit être inférieure à 3,9 Å et la distance entre l'hydrogène et l'accepteur doit être inférieure à 2,5 Å. Les angles donneur-proton H-accepteur (D-H-A), donneur-accepteur-atome lié à l'accepteur (D-A-AA) et accepteur-atome lié à l'accepteur-proton H (A-AA-H) doivent être supérieurs à 90°. Les seuils de distances pour les interactions cation-π (CP) ainsi que pour les interactions π-π (PP) sont fixés à 6 Å. Un pont salin (PS) est attribué lorsque la distance entre un groupement chargé +/- 1 est situé à une distance inférieure à 4 Å d'un groupement portant une charge opposée.

V. Production des plates-formes par synthèse peptidique

V.1 Synthèse chimique sur support solide

Les protéines natives 2it8A, 2uuyB et 2f3cI portant le motif de liaison natif, ainsi que les variants de la protéine native 1edpA, ont été obtenus par synthèse sur support solide grâce à un synthétiseur automatisé (model 433A, Applied Biosystems) en utilisant la stratégie Fmoc/tBu.

Les synthèses sont effectuées en utilisant un protocole standard (décrit en détail dans l'Annexe III) avec 0,1 mmol de résine portant un groupement Fmoc C-terminal, un excès d'un facteur 10 soit 1mmol pour chaque acide aminé-Fmoc (Novabiochem), une activation par le couple DCCI/HOBt et un « capping » des fonctions amines libres par un mélange Ac2O/collidine/NMP.

Les différentes résines utilisées lors de la synthèse peptidique des plates-formes sont présentées dans le tableau V.1.

Tableau V.1 : Résines utilisées lors de la synthèse peptidique des plates-formes.

Plates-formes synthétisées	Type de Résine	Taux de substitution (mmol/g)	Quantité de résine utilisée (mg)
Variants de 1edpA	Fmoc-Rink amid 4-methylbenzhydrylamine (MBHA)	0,66	150
2uuyB	Fmoc-Ala NovaSyn® TGA	0,26	385
2f3cI 2it8A	Fmoc-Rink Amidaminomethyl-polystyrene (AM)	0,62	161

Après lavage de la résine dans du DCM et séchage dans un dessiccateur pendant quelques heures, un clivage dans 10 mL de TFA/TIS/H_2O (95/2,5/2,5 v/v/v) pendant 2 h à température ambiante a permis de libérer le peptide de la résine et d'enlever les groupements protecteurs des chaînes latérales des acides aminés. La résine a ensuite été filtrée et le peptide a été précipité dans 100 mL TBME à froid. Après centrifugation (3000 rpm, 30 min, 4 °C), le culot a été solubilisé dans 30 mL d'une solution d'acide acétique à 10% puis lyophilisé.

V.2 Purification du brut de synthèse par RP-HPLC

Les bruts de synthèse ont été réduits dans un tampon dénaturant (0,5 M Tris pH 8, 6 M guanidine) additionné de 10 mM DTT pendant 90 min à 37 °C. Après acidification à pH 2 avec du TFA/H_2O (50/50 v/v) de la solution de peptides réduits puis filtration sur un filtre 0,45 µm, la purification du peptide d'intérêt a été réalisée par RP-HPLC sur une colonne C18 Discovery® BIO Wide Pore (Supelco) semi-préparative (Diamètre des particules : 5 µm ; taille des pores : 300 Å ; Taille de la colonne : 10 × 250 mm) en utilisant un gradient non linéaire de solvant B (90% ACN, 10% H_2O, 0,09% TFA) dans le solvant A (90% H_2O, 10% ACN, 0,1% TFA) à un débit de 5 mL/min. La détection des peptides a été suivie à 214 nm par un détecteur UV (Merck L4000) équipé d'une cellule de détection de 5 mm. La fraction correspondant au peptide d'intérêt pur a ensuite été congelée dans l'azote liquide puis lyophilisée.

V.3 Formation des ponts disulfures

La formation de ponts disulfures a été réalisée à 4 °C pendant 24 à 48 h, dans le tampon dénaturant additionné du couple oxydo-réducteur glutathion oxydé (GSSG) / glutathion réduit (GSH) à un ratio de 1/10 (0,5 mM GSSG / 5 mM GSH) ainsi que la protéine à une concentration finale de 1 mg/mL. Le suivi de formation des ponts disulfures est réalisé par RP-HPLC analytique en suivant la disparition au cours du temps du pic correspondant à la forme réduite de la protéine. La protéine oxydée est purifiée par RP-HPLC semi-préparative comme décrit précédemment, puis lyophilisée.

VI. Production des plates-formes par voie recombinante chez *E. coli*

VI.1 Construction des vecteurs d'expression

Les gènes synthétiques, optimisés pour une expression recombinante des mini-protéines chez la bactérie *E. coli*, contiennent une séquence non canonique du site de clivage par la TEV protéase (ENLYFQ) suivie par la séquence correspondant à la mini-protéine d'intérêt ayant le motif fonctionnel à greffer à l'extrémité N-terminale, ainsi que les sites de recombinaison Gateway de part et d'autre du gène synthétique.

Ces gènes ainsi construits sont clonés dans le vecteur de destination pETG-82A, contenant la protéine de fusion DsbC ainsi que l'étiquette 6(His), par un clonage LR Gateway™. Les

plasmides d'expression obtenus codent pour une protéine de fusion constituée de :

Protéine DsbC – Etiquette 6(His) - Site de clivage TEV protéase - Mini-protéine d'intérêt

La protéine DsbC est utilisée comme partenaire solubilisant et oxydant permettant la production chez la bactérie *E. coli* de protéines de petites tailles et ayant des ponts disulfures. En effet, la protéine DsbC utilisée en fusion permet d'augmenter la taille de la mini-protéine d'intérêt et est également impliquée dans l'isomérisation des ponts disulfures. L'étiquette 6(His) permet la purification par affinité de la protéine de fusion sur une colonne de nickel et la protéase TEV permet le clivage de la mini-protéine d'intérêt de la protéine de fusion.

Les différents variants des mini-protéines ont été obtenus à partir des plasmides d'expression précédemment décrits en réalisant la mutagénèse dirigée par PCR. Les séquences des plasmides d'expression ont été vérifiées par séquençage (Eurofins MWG).

VI.2 Production des mini-protéines d'intérêt

Pour la production des protéines recombinantes chez *E. coli*, un protocole de co-expression a été mis au point, reposant sur l'utilisation de l'enzyme sulfhydrile oxydase Erv1p afin de catalyser la formation des ponts disulfures des mini-protéines d'intérêt dans le cytoplasme d'*E. coli*.

Pour ce faire, une souche *Escherichia coli* BL21-star a été transformée par un protocole de choc thermique par le plasmide d'expression portant le gène de la mini-protéine d'intérêt ainsi que le plasmide pMJS9 portant le gène codant pour la sulfhydrile oxydase Erv1p, puis mise en culture à 37 °C pendant une nuit dans 10 mL de milieu LB additionné d'ampicilline et de chloramphénicol à des concentrations finales de 100 µg/mL et 30 µg/mL respectivement. Le milieu ZY (925 mL) a été reconstitué en ajoutant 50 mL d'une solution NPS 20X, 20 mL de glycérol 50X, 1 mL de $MgSO_4$ à 1M, 1 mL d'ampicilline à 100 mg/mL et 1 mL de chloramphénicol à 30 mg/mL. Le milieu de culture ainsi préparé a été inoculé avec la pré-culture puis mis en incubation à 30 °C sous agitation à 200 rpm. Lorsque la DO à 600 nm a atteint 0,8, l'expression de la sulfhydrile oxydase a été induite par ajout de 20 mL d'une solution d'arabinose 50X. Après 30 min d'incubation à 20 °C, l'expression de la protéine recombinante a été induite par ajout de 1 mL d'IPTG à 1M et l'incubation a été poursuivie à 20 °C pendant une

nuit.

VI.3 Extraction et purification des protéines de fusion

Les cellules bactériennes ont été récupérées par centrifugation (4500 g, 30 min, 4 °C, rotor JS4.2) et le culot a été ressuspendu dans 50 mL de tampon de lyse (100 mM Tris-HCl pH 8, 150 mM NaCl, 5% glycérol). Les cellules ont ensuite été éclatées dans un système de désintégration de cellules (cell disrupter, Constant System Ltd) permettant de libérer le contenu cytoplasmique des cellules. Le lysat obtenu a été centrifugé (18 000 rpm, 30 min, 4 °C, rotor JA20) et le surnageant contenant la protéine de fusion a été récupéré. La protéine de fusion a par la suite été isolée du mélange cellulaire par purification sur une colonne de nickel HisTrap FF de 5 mL (GE-Healthcare Bio-Sciences) permettant de fixer l'étiquette 6(His) contenue dans la protéine de fusion. La colonne a été préalablement équilibrée avec 25 mL de tampon A (100 mM Tris-HCl pH 8, 150 mM NaCl, 5% glycérol, 40 mM Imidazole). Après chargement du surnageant sur la colonne à un débit de 2 mL/min, la résine a été lavée avec 25 mL de tampon A et la protéine de fusion portant l'étiquette 6(His) a ensuite été éluée par un gradient de tampon contenant une concentration croissante en imidazole permettant de décrocher la protéine de fusion de la colonne par phénomène de compétition : gradient linéaire (0 à 100% B en 30 min) de tampon B (100 mM Tris-HCl pH 8, 150 mM NaCl, 5% glycérol, 500 mM Imidazole) à un débit de 2 mL/min. Un suivi de l'absorbance à 280 nm a été réalisé durant toutes les étapes de chargement du surnageant, de lavage de la colonne et de l'élution de la protéine de fusion. Les différentes fractions contenant la protéine de fusion ont été regroupées.

VI.4 Clivage et purification des mini-protéines d'intérêt

La solution protéique contenant la protéine de fusion a été soumise à une dialyse en boudin (Spectra/Por® Dialysis Membrane ; MWCO : 3500) pendant 3 h contre le tampon C (50 mM Tris-HCl, pH 8), compatible avec l'activité de la TEV protéase. La protéine d'intérêt a ensuite été clivée avec 10% (masse/masse) de TEV protéase pendant une nuit à 4 °C. La solution de clivage a été centrifugée afin d'éliminer le précipité qui se forme après clivage, puis le surnageant est filtré sur des membranes 0,45 µm. Le surnageant issu du clivage, contenant la protéine d'intérêt libre mais également la TEV protéase ainsi que la protéine DsbC portant l'étiquette 6(His), est purifié par RP-HPLC sur une colonne C4 (Vydac 214TP1010, 10 µm, 300

Å, 10x250 mm) en utilisant un gradient non linéaire de solvant B (90% ACN, 10% H_2O, 0,09% TFA) dans le solvant A (95% H_2O, 5% ACN, 0,1% TFA) à un débit de 4 mL/min. Les protéines purifiées ont été lyophilisées puis solubilisées dans le tampon approprié pour les études ultérieures.

VII. Caractérisation des mini-protéines produites par synthèse chimique et par voie recombinante

La pureté des protéines a été contrôlée par RP-HPLC analytique couplée à un spectromètre de masse à trappe ionique à haute résolution (Esquire HCT, Bruker Daltonics) dans les mêmes conditions que celles utilisées lors de la purification. La masse moléculaire exacte des protéines oxydées a été confirmée par spectrométrie de masse MALDI-TOF/TOF 4800 (Applied Biosystems), en mode positif.

L'appariement correct des ponts disulfures a été vérifié pour les plates-formes modifiées 1an1I, 2f3cI, 2it8A et 2uuyB par des tests de liaison sur leur cible naturelle, la trypsine. Une analyse des structures cristallographiques des mutants de la plate-forme 1edpA a permis de confirmer le repliement correct de ceux-ci.

Les concentrations exactes des mini-protéines ont été déterminées par une analyse de la composition en acides aminés (amino acid analyser JLC-500/V AminoTac™, JEOL) ainsi que la mesure de l'absorbance à 280 nm sur un spectromètre Beckmann en utilisant un coefficient d'extinction molaire calculé sur la base de la composition en acides aminés des protéines.

VIII. Tests enzymatiques d'inhibition des MMPs

L'inhibition des MMPs par les mini-protéines produites a été évaluée par des tests enzymatiques de compétition en utilisant le substrat fluorogénique McaMat (Mca-Pro-Leu-Gly-Leu-Dpa-Ala-Arg-NH_2) à une concentration finale de 9 µM et les différentes MMPs humaines (MMP-2h, MMP-8h, MMP-9h, MMP-12h, MMP-13h et MMP-14h) à des concentrations finales nanomolaires. Les concentrations en substrat et en enzymes pour les tests ont été choisies de façon à rester largement en dessous d'une utilisation de 10% du substrat et à observer les taux initiaux.

Les tests enzymatiques sont réalisés à température ambiante, en parallèle sur une plaque de 96 puits, dans un volume final de 50 µl dans le tampon 50 mM Tris-HCl pH 6,8, 10 mM $CaCl_2$.

L'inhibiteur est ajouté dans les puits à quatre concentrations différentes choisies de façon à observer une gamme d'inhibition comprise entre 20 et 80% ; pour chaque concentration en inhibiteur, les tests sont réalisés en double. Après 45 min d'incubation de l'inhibiteur en présence de l'enzyme, sous agitation et à 25 °C, le substrat fluorogénique McaMat est ajouté. L'inhibition est déterminée par mesure de l'augmentation du signal de fluorescence résultant de la dégradation du substrat fluorogénique, en utilisant un spectrophotomètre (Fluoroscan Ascent photon counter) équipé d'un agitateur de plaque. L'excitation se fait à 320 nm et l'émission à 405 nm. Les valeurs de Ki sont déterminées par la méthode proposée par Horovitz et Levitzki (Horovitz et Levitzki, 1987).

Le criblage des venins sur les MMPs est réalisé selon le même protocole que celui décrit ci-dessus. La cible d'intérêt utilisée lors de ces tests de criblage est la MMP-12h. Les tests d'inhibition sont réalisés dans un volume final de 100 µl contenant 90 µl de tampon d'activité des enzymes (50 mM Tris-HCl pH 6,8, 10 mM $CaCl_2$), 10 µl de fraction de venin et la MMP-12h à une concentration nanomolaire. Pour chaque fraction testée, les essais sont réalisés en double. Après incubation de la fraction de venin en présence de l'enzyme, le substrat fluorogénique McaMat est ajouté et l'inhibition est déterminée par mesure de l'augmentation du signal de fluorescence résultant de la dégradation du substrat.

IX. Détermination de la structure du complexe mini-protéine/MMP par cristallographie aux rayons X

IX.1 Cristallogenèse

La solution protéique pour la cristallisation a été préparée à partir du domaine catalytique de la MMP-12 humaine mutant F67D (résidus 106–263) à 643 µM et de l'acide acétohydroxamique AHA (p-aminobenzoyl-Gly-Pro-D-Leu-D-Ala-hydroxamate) à 1 mM, ajouté afin d'éviter l'auto-dégradation de la protéinase pendant la cristallisation, dans un tampon de précipitant (0,1 M glycine pH 8,5 et 27% polyéthylène glycol 10,000). La technique utilisée pour la cristallogenèse est celle de la goutte assise en diffusion de vapeur. Dans cette technique de cristallisation, une goutte contenant 1 µl de solution protéique mère de MMP-12h F67D et de l'inhibiteur AHA et 1µl de tampon de précipitant est déposée au milieu d'une surface concave entourée d'un puits contenant 500 µl de la solution du précipitant à une concentration

deux fois supérieure à celle de la goutte. Par la suite, le système est scellé avec des lamelles et 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 saturation nécessaire au phénomène de cristallisation peut il être atteint. La cristallisation est réalisée à 20 °C pendant 1 à 4 jours d'incubation.

Par la suite, les cristaux sont transférés dans une solution de cryogène (90 mM Tris-HCl pH 8, 27% PEG 8,000, 15% monométhyl éther-PEG 550, 10% glycérol) pendant 5 min pour enlever le AHA, puis trempés pendant 5 min dans un bain contenant 10 µl de la solution de cryogène et 1 µl de solution de mini-protéine (1edpA (1-18) S2ΦΦ et 1edpA (1-18)) à 5 mM dans une solution à 48,8 mM Imidazole. Pour la collecte des données, les cristaux sont congelés dans l'azote liquide.

IX.2 Collecte des données et détermination de la structure tridimensionnelle

Les données sur les complexes MMP-12/inhibiteur ont été collectées à l'ESRF (European Synchrotron Radiation Facility, Grenoble, France) sur la ligne ID14-2 à 100 K. Les cristaux ont diffracté à une résolution élevée de 2,2 Å. Les données ont été traitées par le programme MOSFLM (Leslie, 2006). Les cristaux appartiennent au groupe d'espace tétragonal P21212 avec une molécule dans l'unité asymétrique. L'affinement de la structure, traitée en tant que corps rigide, a été réalisé par le programme REFMAC (Murshudov et al., 1997) en utilisant comme modèle de départ la structure PDB 3LIR (Devel et al., 2010). Le programme CCP4 (Bailey, 1994) a été utilisé pour la détermination des facteurs de structure et la manipulation des coordonnées. L'ajustement final et l'analyse stéréochimique du modèle affiné ont été réalisés à l'aide du programme COOT (Emsley et al., 2010). Les figures ont été réalisées avec le logiciel PyMOL (DeLano, 2002; Schrödinger).

RESULTATS

PARTIE A : Recherche d'inhibiteurs de MMPs dans le venin du serpent *Dendroaspis polylepis*

Un criblage expérimental du venin du mamba noir *Dendroaspis polylepis* a été réalisé afin de rechercher dans ce venin des protéines possédant naturellement la capacité d'inhiber les MMPs. Pour ce faire, nous avons choisi une stratégie de fractionnement bi-dimensionnelle adaptée au grand nombre de protéines généralement présentes dans les venins (jusqu'à un millier). Cette approche a été appliquée à une quantité importante de venin brut afin de pouvoir identifier les toxines actives présentes en faible proportion. Un gramme du venin du serpent *Dendroaspis polylepis* a donc été soumis à un premier fractionnement par chromatographie d'échange d'ions permettant de séparer les toxines en fonction de leur charge. Les fractions mères ainsi obtenues ont été fractionnées par chromatographie en phase inverse (HPLC) afin de séparer les toxines selon leur hydrophobie. Enfin, un criblage des différentes fractions obtenues a été réalisé sur les cibles d'intérêt, les MMPs, grâce à des tests enzymatiques réalisés en parallèle sur des plaques de 96 puits.

I.1 Fractionnement du venin de serpent *Dendroaspis polylepis* par système bi-dimensionnel

Un gramme du venin du serpent *Dendroaspis polylepis* a été soumis à un premier fractionnement par chromatographie d'échange d'ions permettant de séparer les toxines contenues dans le venin en fonction de leur charge, selon le protocole décrit dans la partie Méthodes. La DO à 280 nm des différentes fractions obtenues est mesurée après dilution des fractions, donnant le chromatogramme représenté sur la figure I.1. Les fractions formant un pic bien défini sont regroupées, donnant lieu au final à 12 fractions mères : Dp1A (fraction issue du lavage de la colonne après chargement avec l'échantillon) à Dp1L.

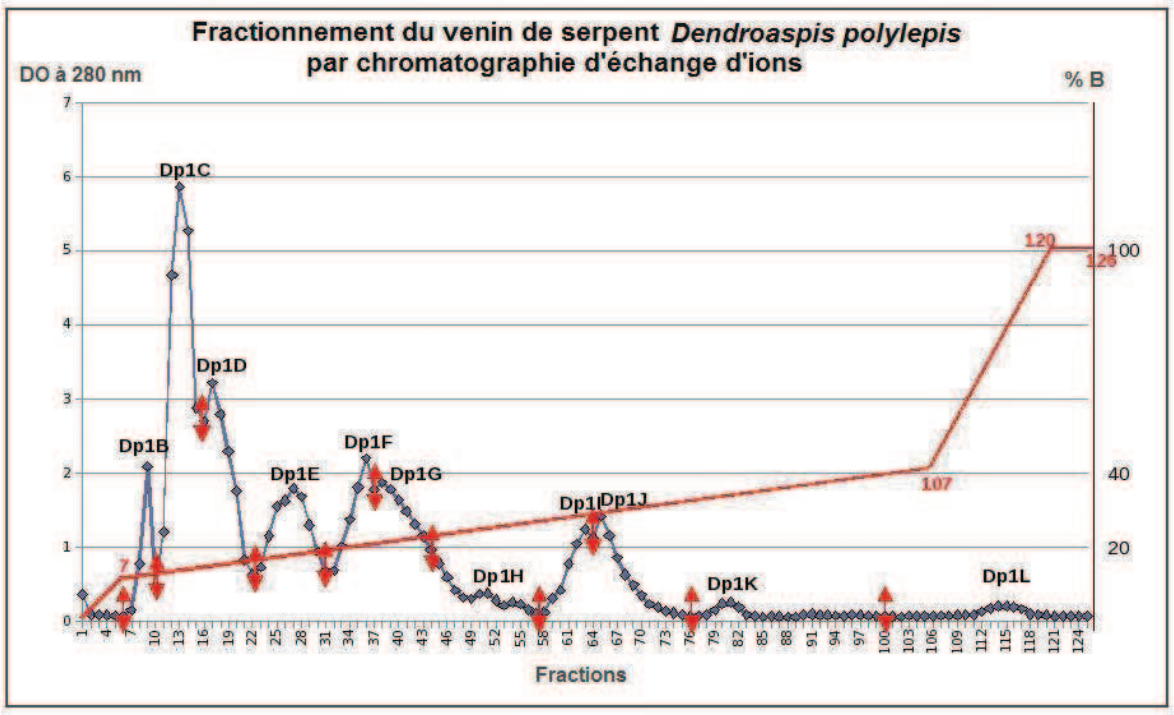


Figure I.1 : Fractionnement du venin de serpent *Dendroaspis polylepis* par chromatographie d'échange d'ions. Le chromatogramme montre la D0 à 280 nm des 124 fractions obtenues (courbe en losanges bleus) ainsi que le gradient non linéaire de tampon B utilisé (tracé rouge). Les doubles flèches rouges délimitent les fractions qui ont été regroupées.

Par la suite, les 12 fractions mères ainsi obtenues ont été soumises à une deuxième purification par chromatographie en phase inverse (RP-HPLC) afin de séparer les toxines selon leur hydrophobie. Un exemple de purification par RP-HPLC de la fraction mère Dp1D est reporté sur la figure I.2. La fraction Dp1D a été reprise dans 4 ml d'eau milliQ ; le dosage UV à 280 nm a permis de déterminer la quantité totale de protéines présentes dans cette fraction à 138 mg. Un volume de 1 ml de cette fraction a été injecté sur une colonne préparative et la séparation a été réalisée à un débit de 8 ml/min. La purification de cette fraction mère par RP-HPLC a permis d'obtenir 10 fractions filles : Dp1D1 à Dp1D10.

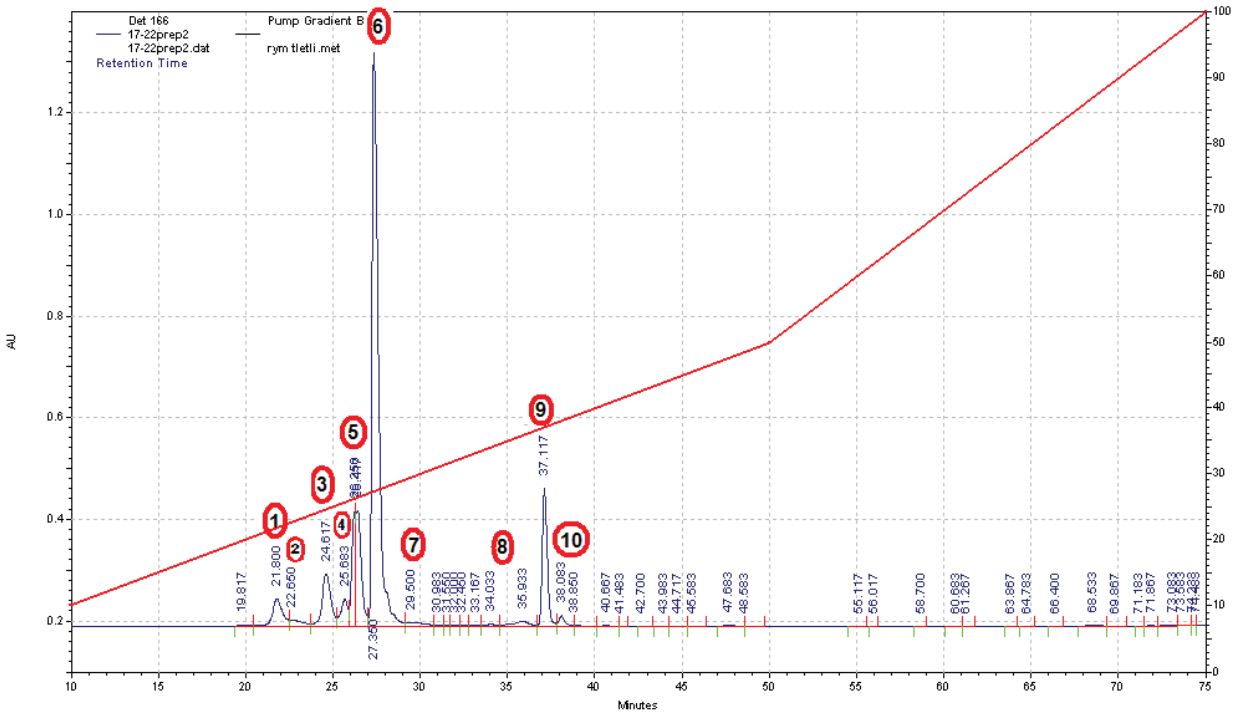


Figure I.2 : Purification par RP-HPLC de la fraction mère Dp1D issue du fractionnement par chromatographie d'échange d'ions. Le tracé bleu représente le suivi de la DO à 280 nm. Les chiffres en bleu au dessus des pics correspondent aux temps de rétention. Le tracé rouge correspond au gradient non linéaire en solvant B utilisé pour la purification.

I.2 Criblage des venins de serpents sur les MMPs

Les fractions mères du venin de *Dendroaspis polylepis* obtenues par chromatographie d'échange d'ions (12 fractions au total : Dp1A à Dp1L) ont été soumises à un criblage sur la cible d'intérêt, la MMP-12h, selon le protocole décrit dans la partie Méthodes.

Deux fractions mères parmi celles testées ont montré une activité significative sur la MMP-12 : Dp1A et Dp1D. Afin de vérifier que les résultats obtenus ne sont pas dus à des faux positifs, les tests ont été refaits en utilisant deux volumes différents des fractions mères positives : 10 µl et 20 µl. Les résultats ont permis de confirmer que les fractions mères Dp1A et Dp1D inhibent la MMP-12 étant donné que le signal de fluorescence obtenu augmente lorsque le volume de la fraction testée augmente.

Le deuxième fractionnement par RP-HPLC des fractions mères Dp1A et Dp1D a donné lieu à 11 et 10 fractions filles respectivement. Ces fractions filles ont donc été testées sur la MMP-12 afin de confirmer la présence d'activité inhibitrice de MMPs. Cependant, les résultats obtenus ont montré qu'aucune activité significative n'a été détectée dans ces fractions.

PARTIE B : Conception *in silico* de ligands protéiques inhibiteurs de MMPs

CHAPITRE I : Système de référence : TIMP/MMP

L'approche de conception *in silico* décrite au cours de ce travail a été appliquée au développement de nouveaux ligands protéiques destinés à inhiber les membres de la famille des MMPs. Le système de référence utilisé dans cette étude est un complexe formé par une MMP avec son inhibiteur protéique naturel le TIMP. Dans cette partie, une analyse détaillée du complexe TIMP-2/MMP-14 choisi comme référence a été réalisée afin d'identifier le motif fonctionnel du TIMP essentiel dans l'inhibition de la MMP, les caractéristiques générales de l'interface des deux partenaires ainsi que les interactions intermoléculaires établies au niveau de cette interface.

I.1 Données structurales et fonctionnelles sur les complexes TIMP/MMP

Plusieurs structures tridimensionnelles de complexes TIMP/MMP ont été résolues et leurs coordonnées cartésiennes sont accessibles dans la PDB (Tableau I.1). Ces structures concernent les TIMP-1 et -2 avec différentes MMPs, tandis que la structure cristallographique du N-TIMP-3 a été résolue en complexe avec l'enzyme TACE (Tumor Necrosis factor-α-Converting Enzyme ; ADAM-17). En revanche, aucune structure 3D n'a été décrite pour le TIMP-4.

Tableau I.1 : Données structurales sur les complexes TIMP/MMP déposés dans la PDB.

Complexe TIMP/MMP	Code PDB	Méthode	Taille de l'interface ($Å^2$)	Référence
N-TIMP-1/MMP-1	2J0T	Rayon X (Résolution 2,54 Å)	1770	Iyer et al., 2007
TIMP-1/MMP-3	1UEA	Rayon X (Résolution 2,80 Å)	2910	Gomis-Rüth et al., 1997
N-TIMP-1/MMP-3	1OO9	RMN		Arumugam et Van Doren, 2003
TIMP-1/MMP-10	3V96	Rayon X (Résolution 1,90 Å)	3160	Batra et al., 2012
TIMP-1/MT1-MMP	3MA2	Rayon X (Résolution 2,05 Å)	1600	Grossman et al., 2010
TIMP-2/Pro-MMP-2	1GXD	Rayon X (Résolution 3,10 Å)	2490	Morgunova et al., 2002
TIMP-2/MMP-13	2E2D	Rayon X (Résolution 2 Å)	2640	Maskos et al., 2007
TIMP-2/MT1-MMP	1BQQ/ 1BUV	Rayon X (Résolution 2,75 Å)	2940	Fernandez-Catalan et al., 1998

L'interaction des TIMPs avec le site catalytique des MMPs se fait essentiellement par le biais d'un motif constitué de 4 résidus présents au niveau de l'extrémité N-terminale de la protéine. Dans ce motif, deux résidus cystéine en position 1 et 3 sont strictement conservés parmi les TIMPs et sont engagés dans deux ponts disulfures. Les deux autres résidus du motif varient en fonction des TIMPs, comme indiqué dans le tableau I.2. Parmi les différents TIMPs chez l'homme, le résidu en position 2 peut donc être une sérine ou une thréonine et le résidu 4 peut être une sérine, une valine ou une alanine.

Tableau I.2 : Les motifs N-terminaux des 4 TIMPs humains.

TIMP	Motif fonctionnel de liaison
TIMP-1	C_1-T_2-C_3-V_4
TIMP-2	C_1-S_2-C_3-S_4
TIMP-3	C_1-T_2-C_3-S_4
TIMP-4	C_1-S_2-C_3-A_4

I.2 Analyse de l'interface du complexe TIMP-2/MMP-14 (1bqq)

Le complexe TIMP-2/MMP-14 a été choisi dans ce projet comme complexe de référence. Il correspond à la structure entière du TIMP-2 liée au domaine catalytique de la MMP-14 (MMP-14dc) (Figure I.1).

Figure I.1 : Structure tridimensionnelle du complexe TIMP-2/MMP-14 (code PDB 1bqq). Le domaine catalytique de la MMP-14 (gris) ainsi que la forme entière du TIMP-2 (vert) sont représentés en mode cartoon, le motif de liaison du TIMP-2 est représenté en bâtons et le zinc catalytique de la MMP-14 est représenté en sphère orange.

Une analyse de l'interface de ce complexe a été réalisée afin de déterminer les propriétés structurales, chimiques et électrostatiques de l'interface et d'identifier les interactions établies entre ce TIMP et la MMP.

Pour ce faire, la structure cristallographique du complexe TIMP-2/MMP-14 a d'abord été

soumise à une procédure de relaxation sous contrainte positionnelle par dynamique moléculaire grâce au logiciel CHARMM à l'aide du protocole décrit dans la section Méthodes. Au cours de ces calculs, le groupement N-terminal du TIMP-2 a été défini sous forme neutre NH_2 et non sous forme protonée standard (Voir Méthode). Ce protocole conduit à une très bonne préservation de la structure 3D des deux partenaires. En effet, la structure finale issue de ce protocole se caractérise par un écart moyen (RMSD) sur les positions des atomes Cα de 0,64 Å pour le TIMP-2 et de 0,10 Å pour les atomes Cα de la MMP-14 par rapport à la structure expérimentale après superposition.

Cette structure a été utilisée pour analyser l'interface entre les deux partenaires naturels TIMP-2/MMP-14. Cette analyse est fondée sur des critères couramment utilisés dans la littérature (voir Bibliographie Chapitre II.) relatifs aux propriétés globales des interfaces protéine-protéine telles que la taille, la géométrie et la composition chimique. L'analyse concerne également les interactions intermoléculaires dans le complexe regroupant les liaisons hydrogène, les ponts salins, les interactions cation- π et les interactions π- π.

I.2.1 Propriétés globales de l'interface

I.2.1.1 Variation de surface accessible au solvant/Taille de l'interface

Comme nous l'avons vu dans la partie B, Chapitre II de la bibliographie, il existe dans la littérature deux définitions de l'interface protéine-protéine. Certains travaux reportent la variation totale de surface accessible au solvant. Celle-ci est calculée comme la différence entre la surface accessible au solvant du complexe et la somme des surfaces accessibles au solvant des deux partenaires isolés. Toutefois, certaines études utilisent la définition relative à un seul des partenaires, c'est à dire la variation de surface accessible au solvant d'un des partenaires consécutif à son association.

L'analyse de la structure du complexe TIMP-2/MMP-14 montre que la variation totale de surface accessible au solvant consécutive à la liaison des deux partenaires est de 3190 $Å^2$. Sur une surface totale de 9976 $Å^2$, le TIMP-2 possède une interface avec la cible de **1643 $Å^2$**, ce qui représente 51,5% de la variation totale de surface accessible au solvant consécutive à la liaison. De même, le domaine catalytique de la MMP-14 possède une interface de 1546 $Å^2$, ce qui correspond à 48,5% de la variation totale de surface accessible au solvant consécutive à la

l'association des deux partenaires. Une telle différence, observée dans certains complexes protéine-protéine naturels est liée à la concavité de l'interface et à la méthode de calcul de celle-ci. Les deux partenaires contribuent cependant de manière presque égale à la variation totale de surface accessible au solvant consécutive à leur association.

I.2.1.2 Compacité de l'interface

Le volume vide présent au niveau de l'interface entre le ligand et la cible renseigne sur la compacité de l'interface et la complémentarité entre les deux surfaces. Plus ce volume est petit, plus l'interface est compacte, ce qui a pour conséquence une meilleure complémentarité et une distance réduite entre les surfaces, favorisant ainsi les interactions entre les deux partenaires. Dans le cas du complexe entre le TIMP-2 et la MMP-14, le volume vide entre les protéines est de 3207 $Å^3$. L'indice de volume vide ou « Gap Index » est calculé à partir du volume vide entre les molécules, selon la formule suivante : Gap index (Å) = volume vide entre molécules ($Å^3$) / aire de l'interface ($Å^2$). Le « Gap Index » calculé pour le **ligand** est de **1,95 Å**, et celui calculé pour la **cible** est de **2,07 Å**. Cette valeur est plus faible que la moyenne correspondante reportée dans la littérature pour les complexes enzymes-inhibiteurs (2,20 Å ; Jones et Thornton, 1996), indiquant que les surfaces entre le TIMP-2 et la MMP-14 sont très complémentaires. On peut noter de ce point de vue que dans la structure cristallographique du complexe TIMP-2/MMP-14 (1bqq), il existe très peu de molécules d'eau interfaciales.

I.2.1.3 Composition de l'interface

Sur un total de 184 résidus, **36 résidus** du ligand naturel TIMP-2 font partie de l'interface avec la MMP-14. Ces résidus sont en majorité apolaires (23 résidus) dont 3 résidus aromatiques. Parmi les 13 résidus polaires, 7 sont des résidus polaires chargés et 6 sont des résidus polaires neutres. Ces valeurs sont en accord avec les valeurs rapportées dans la littérature concernant la composition des interfaces dans les complexes hétéro-dimériques décrites comme étant non polaires à 56-57%, avec en particulier des interfaces plus hydrophobes dans les complexes protéase-inhibiteur constituées de 61% de résidus non polaires. Les acides aminés les plus représentés au niveau de l'interface du TIMP-2 sont l'alanine (5), la cystéine (4), l'isoleucine (4) et la lysine (3), tandis que les acides aminés

phenylalanine, leucine et glutamine sont absents de l'interface.

I.2.2 Interactions intermoléculaires

Quatre types d'interactions intermoléculaires sont recherchés au niveau de l'interface du complexe de référence : les liaisons hydrogène, les ponts salins, les interactions cation-π et les interactions π-π.

I.2.2.1 Liaisons hydrogène

Pour une distance seuil fixée à 3,9 Å entre l'accepteur et le donneur d'hydrogène, 17 liaisons hydrogène (HB) ont pu être identifiées dans le complexe TIMP-2/MMP-14. Ce nombre correspond à la moyenne de liaisons hydrogène observée dans les données de la littérature (Xu et al., 1997 ; Janin et Chothia, 1990). Le nombre de liaison hydrogène rapporté à la variation totale de surface accessible au solvant consécutive à la liaison est cependant faible, avec 0,53 HB/100 $Å^2$, du fait de la taille importante de l'interface. En considérant l'interface du TIMP-2, la densité superficielle de liaison hydrogène (nb de HB/100 $Å^2$) est de 1,03 HB/100 $Å^2$.
L'analyse détaillée des liaisons hydrogène intermoléculaires est reportée dans le tableau I.3.

Tableau I.3 : Liaisons hydrogène (HB) établies entre le TIMP-2 et la MMP-14 dans la structure du complexe issu du 15ème cycle de relaxation. Les donneurs (D) et les accepteurs (A) sont identifiés par le nom du résidu, la position du résidu dans la protéine et l'atome impliqué dans la liaison. La géométrie de la liaison est déterminée par la distance Donneur-Accepteur (D-A), la distance proton H-Accepteur (H-A) ainsi que l'angle formé par Donneur-proton H-Accepteur (DHA).

HB id	TIMP-2				MMP-14				d D-A (Å)	d H-A (Å)	Angle DHA (°)
1	D:	CYS	1	N	A:	ALA	200	O	2,81	1,83	162,6
2	D:	SER	2	N	A:	ALA	200	O	3,71	2,9	138,9
3	D:	SER	2	OG	A:	GLU	240	OE1	2,63	1,64	172,5
4	A:	SER	2	O	D:	LEU	199	N	2,98	2	167,1
5	D:	CYS	3	N	A:	PRO	259	O	2,83	2	138,9
6	A:	CYS	3	O	D:	TYR	261	N	3,06	2,09	163,1
7	D:	SER	4	N	A:	GLY	197	O	2,82	1,88	153
8	D:	TYR	36	OH	A:	ASP	212	OD2	2,62	1,73	149,4
9	A:	GLY	37	O	D:	ASN	208	ND2	2,89	2,07	136,6
10	A:	LYS	41	O	D:	SER	189	OG	2,67	1,7	175,6
11	D:	ARG	42	NH1	A:	SER	189	O	2,94	1,96	161,1
12	A:	ALA	70	O	D:	ALA	202	N	3,57	2,57	169,3
13	A:	VAL	71	O	D:	THR	190	OG1	2,8	1,89	154,1
14	D:	ARG	132	NH1	A:	LEU	228	O	2,75	1,74	167,5
15	D:	ARG	132	NH1	A:	ASN	229	O	2,73	1,89	138,6
16	D:	TRP	151	NE1	A:	TYR	261	O	3,1	2,16	159,2
17	D:	LYS	155	NZ	A:	GLN	262	OE1	2,7	1,66	167,9

Sur les 17 liaisons hydrogène, 7 liaisons impliquent des atomes de la chaine principale du TIMP-2 dont 4 donneurs N et 3 accepteurs O. Parmi ces 7 liaisons, 6 impliquent les 4 résidus en position N-terminale du TIMP-2. Pour l'ensemble des interactions, les distances DA sont comprises entre 2,62 et 3,71 Å et les angles formés par les vecteurs DH et HA sont compris entre 138 et 175°.

I.2.2.2 Autres types d'interactions

Les ponts salins sont peu fréquents dans les interfaces des complexes hétéro-dimériques (Xu et al., 1997). L'interface du complexe de référence ne déroge pas à cette observation puisqu'aucun pont salin n'est observé à l'interface des deux partenaires. De même, aucune

interaction cation-π et π-π n'est observée entre le TIMP-2 et le domaine catalytique de la MMP-14 dans le complexe examiné.

I.2.2.3 Localisation des interactions

Les interactions intermoléculaires (liaisons hydrogène) identifiées sur le complexe TIMP-2/MMP-14 issu du 15ème cycle de relaxation sont visualisées grâce au logiciel PyMol (Figure I.2). Le domaine catalytique de la MMP-14 est représenté en mode surface (cyan) et les résidus impliqués dans des interactions intermoléculaires sont indiqués en rouge. Cette figure montre que les interactions ont essentiellement lieu au niveau de deux zones : le sillon du site catalytique et une zone latérale située au niveau des boucles II-III et IV-V. Les points de contacts forment des zones discontinues, mais sont néanmoins tous localisés de part et d'autre d'une ligne horizontale située au niveau du sillon catalytique et séparant le sous-domaine supérieur du sous-domaine inférieur. De la même manière, les résidus du TIMP-2 (mode surface en vert) impliqués dans les différentes interactions sont regroupés en trois zones distinctes : une zone formée par l'extrémité N-terminale (résidus 1 à 4), une zone formée par la boucle AB du TIMP-2 (résidus 36, 37, 41 et 42) et une zone située dans le domaine C-terminal de la molécule (résidus 132, 151, 155).

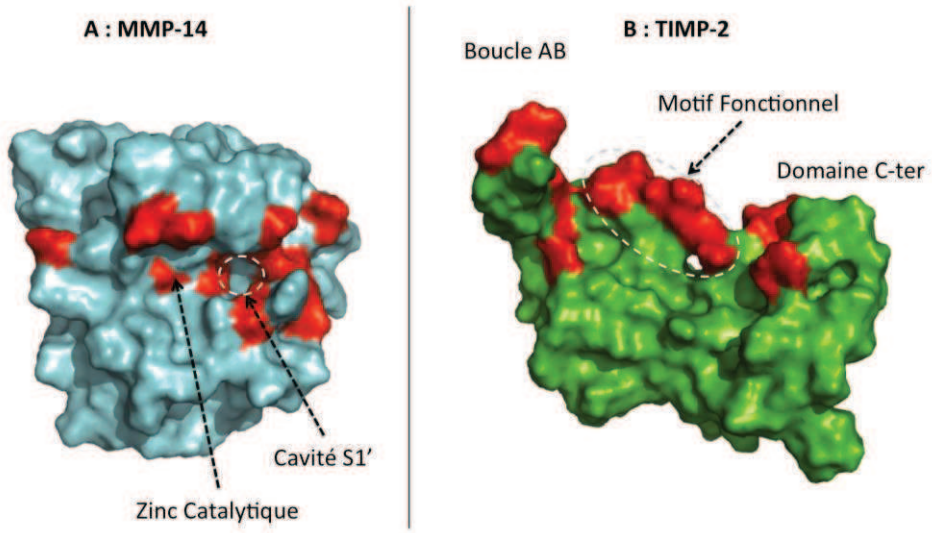


Figure I.2 : Interactions intermoléculaires entre la MMP-14dc et le TIMP-2. Les protéines sont représentées en mode surface (MMP-14 en cyan et TIMP-2 en vert) et les résidus établissant des interactions intermoléculaires de type liaisons hydrogène sont coloriés en rouge.

Parmi les trois régions identifiées, le motif fonctionnel du TIMP-2 constitué des 4 résidus N-

terminaux présente la plus forte densité de liaisons hydrogène (7 interactions sur 17). Le détail de ces interactions reporté dans le tableau I.3 est illustré sur la figure I.3. Toutes les interactions établies au niveau du motif se font par le biais de la chaine principale.

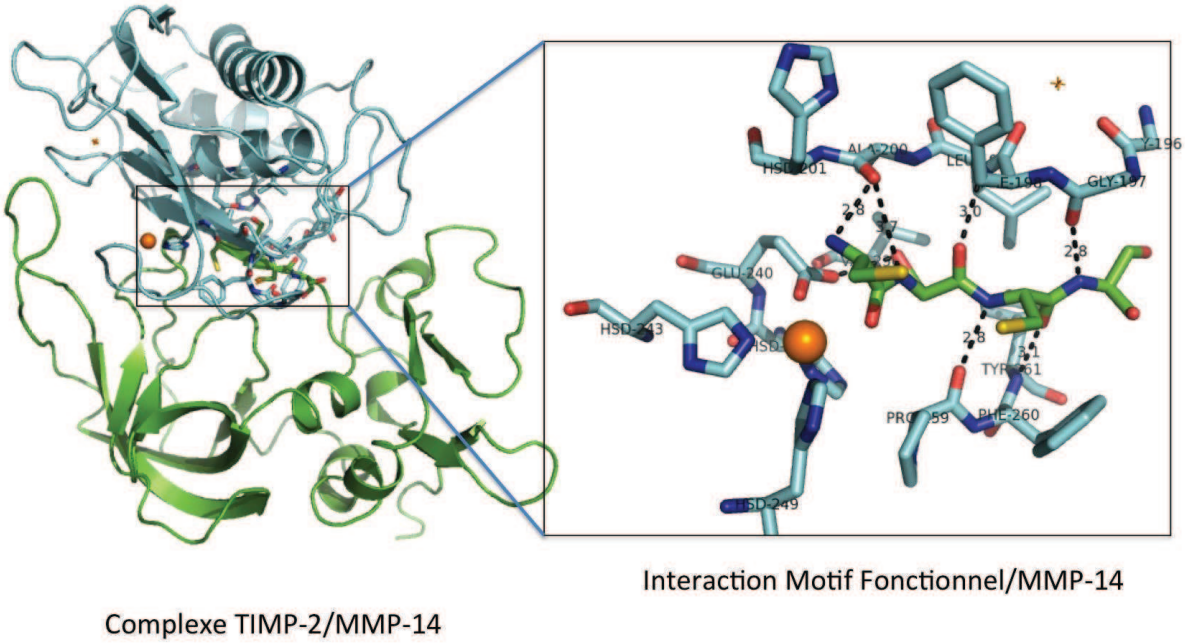


Figure I.3 : Interactions intermoléculaires établies au niveau du motif fonctionnel du TIMP-2. Le TIMP-2 (vert) et la MMP-14 (cyan) sont représentés en mode cartoon, les résidus du ligand et de la cible impliqués dans des interactions intermoléculaires sont représentés en mode bâtons et l'atome de zinc catalytique en mode sphère coloriée en orange. Les distances intermoléculaires sont représentées en pointillés noirs.

I.2.3. Propriétés électrostatiques

Les potentiels électrostatiques du domaine catalytique de la MMP-14 et de son inhibiteur TIMP-2 à l'état libre ont été calculés à l'aide de l'équation de Poisson Boltzmann. La valeur de ce potentiel à la surface de chaque partenaire est représentée sur la Figure I.4. Le potentiel électrostatique dans le sillon de la MMP-14 est faible à l'exception de la région des sous-sites S_1' à S_4' qui montre un potentiel de surface négatif. Concernant le TIMP-2, ces calculs montrent que le potentiel électrostatique de surface du TIMP-2 de la région de la boucle AB est proche de 0. Dans la région du domaine C-terminal du TIMP-2 impliquée dans l'interaction avec la cible, le potentiel électrostatique est nettement positif. On peut noter que cette région interagit avec la région du domaine catalytique de la MMP-14 qui possède le potentiel de surface le plus négatif, illustrant la complémentarité électrostatique des deux partenaires dans ces régions. Enfin, dans la région correspondant aux résidus 1-4 du TIMP-2 le

potentiel de surface apparait négatif mais d'intensité faible. Ces observations suggèrent que les liaisons hydrogène identifiées dans la région 1-4 pourraient constituer la contribution essentielle à la liaison des résidus du motif 1-4 à la cible.

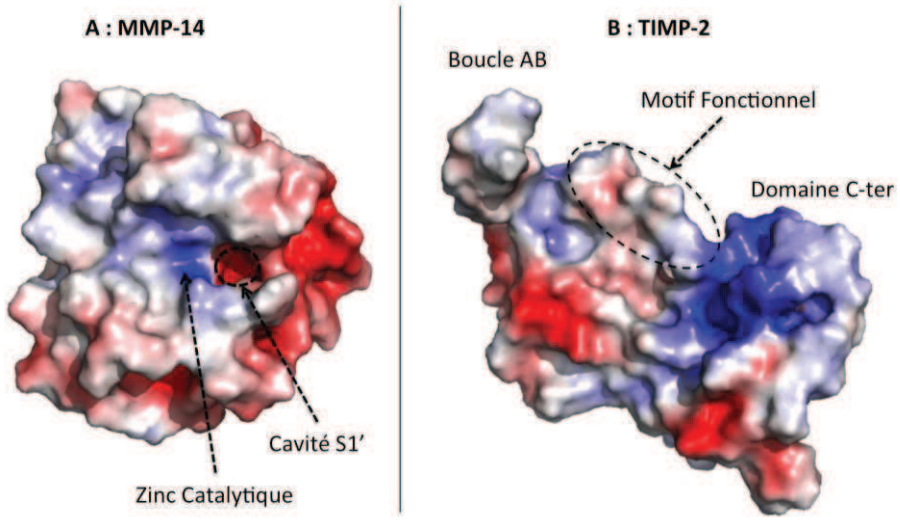


Figure I.4: Potentiels électrostatiques des partenaires MMP-14dc et TIMP-2 calculés à l'aide de l'équation de Poisson Boltzman. Les protéines sont représentées en mode surface. Les valeurs du potentiel électrostatique sont indiquées à l'aide d'un code de couleur. Bleu : E_elec = +3 k_BT/e ; Rouge : E_elec = -3 k_BT/e.

I.3 Conclusion

L'analyse de la structure du complexe TIMP-2/MMP-14 a montré que l'inhibiteur protéique naturel interagit avec sa cible via trois sites principaux établissant des liaisons hydrogène intermoléculaires. Parmi ces sites, le motif constitué des 4 résidus N-terminaux Cys1-Ser2-Cys3-Ser4 (en fonction des TIMPs, le résidu 2 = Ser/Thr et le résidu 4 = Ser/Val/Ala) de l'inhibiteur établit des liaisons hydrogène qui impliquent très majoritairement des groupements présents sur la chaine principale de l'inhibiteur. De plus, ces groupements se trouvent principalement en interaction avec des groupements localisés sur la chaine principale de la cible. Les TIMPs possèdent d'autres caractéristiques faisant du segment 1-4 un motif d'intérêt à greffer pour la conception de ligands protéiques artificiels. Leur résidu N-terminal est engagé dans plusieurs interactions au cœur du site catalytique. Dans les TIMPs, la présence des ponts disulfure impliquant les résidus 1 et 3 confère un cadre structural rigide favorisant la stabilité des interactions engagées dans le site actif des MMPs par le segment 1-4. Par ailleurs, les données structurales disponibles décrivant l'interaction de TIMPs avec des domaines catalytiques de MMPs et la très grande conservation structurale de ces domaines au sein de la

famille suggèrent que le mode d'interaction adopté par le motif 1-4 du TIMP-2 en complexe avec la MMP-14 n'est pas restreint à ces deux partenaires. Nous avons donc sélectionné ce motif pour la conception de ligands protéiques de première génération capables d'interagir avec les domaines catalytiques de MMPs. Les données structurales utilisées pour définir la topologie de ce motif sont celles issues du complexe TIMP-2/MMP-14 (1bqq). Au cours de ce travail, cette structure est considérée comme le prototype de l'interaction TIMP/MMP pour ce qui concerne le positionnement du segment 1-4 du TIMP dans le domaine catalytique des MMPs.

CHAPITRE II : Recherche de plates-formes protéiques par STAMPS

Le programme STAMPS a été utilisé pour cribler la totalité des structures déposées dans la version complète de la PDB afin de rechercher des plates-formes protéiques capables de reproduire la topologie fonctionnelle du segment 1-4 du TIMP-2. A cette fin, STAMPS recherche des groupes de 4 résidus possédant une topologie similaire à celle adoptée par le motif 1-4 du TIMP-2 dans le complexe TIMP-2/MMP-14 (1bqq). Les plates-formes identifiées sont ensuite soumises à une série de critères de validation/sélection.

II.1 Critères de recherche *in silico* du programme STAMPS

- Emplacement du motif : le motif de liaison mime doit être présent à l'extrémité N-terminale de la protéine, parmi les **5 résidus N-terminaux**. Ceci autorise, par exemple, l'identification d'un motif équivalent au motif 1-4 du TIMP qui se situerait en position 2-5 de la plate-forme. Si une telle plate-forme était identifiée, elle serait produite en supprimant le résidu en position 1 laissant le résidu 2 en position N-terminale du candidat ligand.

- Composition du motif : le motif mime doit contenir **4 résidus** avec au moins un résidu **Cys en position 1 et/ou 3**. Deux raisons ont conduit à ce choix : d'une part, des recherches préliminaires à l'aide de STAMPS ont été réalisées en n'excluant aucun des types de résidus en position équivalente à celle des 4 résidus composant le motif 1-4 du TIMP-2. Ces recherches ont conduit à l'identification d'un très grand nombre de plates-formes, dont le post-traitement n'est pas possible. Pour restreindre ce nombre, nous avons imposé la nature du résidu équivalent au résidu 1 (Cys) et/ou la nature du résidu équivalent au résidu 3 (Cys). Par ailleurs, les cystéines engagées dans un pont disulfure constituent un élément de rigidification de structure pouvant stabiliser les interactions avec le partenaire. Ceci a pu être réalisé en procédant à trois recherches distinctes avec STAMPS : une recherche en imposant la présence de deux cystéines dans le motif, en position 1 et 3 ($Cys - X - Cys - X$), une recherche en imposant uniquement la présence de la cystéine en position 1 ($Cys - X - \overline{Cys} - X$) et une recherche en imposant uniquement la présence de la cystéine en position 3 ($\overline{Cys} - X - Cys - X$) ; X étant un résidu quelconque.

- Topologie du motif : le motif mime doit posséder une topologie identique au motif de référence afin de reproduire les interactions intermoléculaires décrites dans le complexe de référence, et ce indépendamment de la nature des résidus du motif mime. La recherche topologique est basée sur la comparaison entre les atomes Cα et Cβ du motif de référence et ceux du motif mime, grâce à l'utilisation des critères définis dans la section Méthodes : **$RMSD_{max}$ (1,25 Å)**, **Δd_{max} (1,85 Å)**, **Θ_{max} (130°)** et **$\Delta\Theta_{max}$ (90°)**.

- Filtre stérique : les protéines sélectionnées doivent être stériquement compatibles avec la cible. Les résidus de la plate-forme mime situés au niveau de l'interface doivent se trouver à une certaine distance des résidus de la cible afin d'éviter une gêne ou un recouvrement stérique. La sélection des plates-formes se fait grâce à un score d'encombrement stérique intermoléculaire **Steric-inter (2,5)** ainsi qu'un score d'encombrement stérique intramoléculaire **Steric-intra (1)** (voir Méthodes).

- Taille des protéines : l'objectif de ce travail est la conception de protéines ligands de faible poids moléculaire ou mini-protéines. Une taille minimum a également été définie pour éviter la sélection de peptides faiblement structurés. Les protéines sélectionnées doivent avoir une taille comprise **entre 15 et 60 acides aminés**.

II.2 Identification et sélection des plates-formes

- Identification des plates-formes par STAMPS

Les deux recherches distinctes menées à l'aide de STAMPS ($Cys - X - \overline{Cys} - X$ et $\overline{Cys} - X - Cys - X$) ont permis d'identifier **219 plates-formes** protéiques satisfaisant les contraintes structurales et les valeurs seuil des paramètres définis ci-dessus. Toutefois, dans ces deux groupes, il existe certaines redondances. Le principe de cette recherche entraine que les plates-formes possédant naturellement deux cystéines en positions équivalentes aux résidus 1 et 3 du motif ($Cys - X - Cys - X$) sont identifiées dans les deux recherches. Elles sont au nombre de 5. Si l'on retranche aux groupes 2 et 3 les protéines possédant deux cystéines en positions équivalentes aux résidus 1 et 3 du motif, le groupe $Cys - X - \overline{Cys} - X$ (pas de résidu

Cys en position 3) contient 193 solutions et le groupe $\overline{Cys} - X - Cys - X$ (pas de résidu Cys en position 1) contient 16 solutions. La présence simultanée des deux résidus Cys dans le motif est donc un critère très restrictif.

Type de motif	$Cys - X - Cys - X$	$Cys - X - \overline{Cys} - X$	$\overline{Cys} - X - Cys - X$
Solutions identifiées	5	16	193

Les valeurs des paramètres correspondant aux critères utilisés pour sélectionner ces plates-formes, qui sont retournées pour chacune d'entre elles, permettent de caractériser l'écart géométrique entre le motif identifié et le motif 1-4 du TIMP-2. Toutefois, certaines caractéristiques des plates-formes, importantes pour le greffage, sont difficiles à apprécier au vu de ces seuls paramètres. L'analyse graphique des structures tridimensionnelles peut être utile et permettre d'éliminer certaines plates-formes retenues à l'issue de la première étape.

- Sélection graphique des plates-formes

- Redondance des solutions : la recherche effectuée par le programme STAMPS permet de cribler l'ensemble des structures contenues dans la PDB entière, soit plus de 82000 structures (version 07/2012). Cette base de données contient des redondances correspondant soit à des fichiers PDB contenant plusieurs structures identiques d'une même protéine (rayon-X), soit à plusieurs structures (conformères) d'une même protéine résultant du processus de reconstruction des structures à partir de données de résonance magnétique nucléaire (RMN). Dans ce dernier cas, STAMPS élimine toutes les solutions redondantes et une seule structure par fichier PDB est conservée, correspondant à la structure dont le RMSD sur les coordonnées des atomes du motif est le plus bas. Cette redondance n'apparaît pas dans les fichiers de sortie. En revanche la version actuelle de STAMPS (3.1.3) permet d'examiner également systématiquement les différentes molécules décrites dans les fichiers de coordonnées cartésiennes de structures résolues par diffraction des rayons-X et qui correspondent aux différentes molécules présentes dans l'unité asymétrique. Un exemple de ce type de redondance est constitué par le fichier PDB 1p0t retrouvé parmi les solutions de la recherche $\overline{Cys} - X - Cys - X$ qui contient 60 chaines correspondant à une séquence unique.

- Absence de ponts disulfures dans le motif : parmi les résultats obtenus à l'issue des recherches $Cys - X - \overline{Cys} - X$ et $\overline{Cys} - X - Cys - X$, la présence d'un ou deux résidu(s) Cys dans le motif de la plate-forme engagés dans des ponts disulfure est une condition souhaitable afin de mieux préserver la conformation et la stabilité du motif greffé. La version actuelle du programme STAMPS (3.1.3) ne permettant pas de réaliser cette sélection de manière automatique, les plates-formes contenant des cystéines libres parmi les résidus composant le motif équivalents aux résidus 1-4 du TIMP-2 sont éliminées à l'issue d'une analyse graphique.

Cette étude a permis de montrer que, de manière générale, les plates-formes dont le motif contient des résidus Cys libres sont des protéines ne possédant aucun pont disulfure. Parmi ces protéines, on distingue 3 sous-groupes selon la structuration de la chaine peptidique, illustrés dans la figure II.1 : A) les protéines n'ayant aucune structure secondaire régulière ; B) les protéines possédant un élément de structure secondaire régulier, principalement une longue hélice α ; C) les protéines parfaitement structurées avec plusieurs éléments de structures secondaires réguliers. Ces plates-formes protéiques non stabilisées par des ponts disulfures peuvent conduire à un contrôle plus réduit de la conformation «fonctionnelle » du motif à greffer. Par ailleurs, la présence de cystéines libres peut entrainer des difficultés au cours des phases de production et de repliement. Si de telles plates-formes étaient sélectionnées, il serait préférable de muter ces résidus cystéines libres.

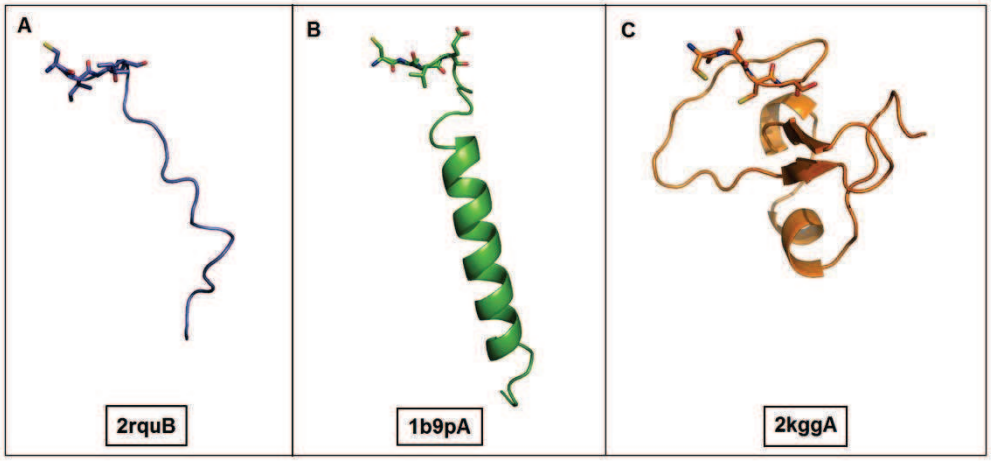


Figure II.1 : Différents degrés de structuration des plates-formes sans ponts disulfures. Les protéines sont représentées en mode cartoon et le motif de liaison en bâtons. A : Protéine sans aucune structure secondaire régulière (2rquB) ; B : Protéine avec un élément de structure secondaire régulière de type hélice α (1b9pA) ; C : Protéine avec plusieurs éléments de structures secondaires régulières de type hélices α et feuillets β (2kggA).

- Orientation du motif de liaison : ce critère est également déterminant puisque ce motif peut être présent à l'extrémité N-terminale mais orienté de manière opposée au motif de référence. Dans ce cas, les résidus 1, 2, 3 et 4 du motif de référence sont superposés respectivement aux résidus 4, 3, 2 et 1 du motif mime. Par conséquent, le motif mime est présent parmi les 5 premiers résidus N-terminaux de la plate-forme, sa topologie est correcte, mais le résidu N-terminal libre est dirigé à l'opposé du zinc catalytique avec lequel il doit interagir (Figure II.2).

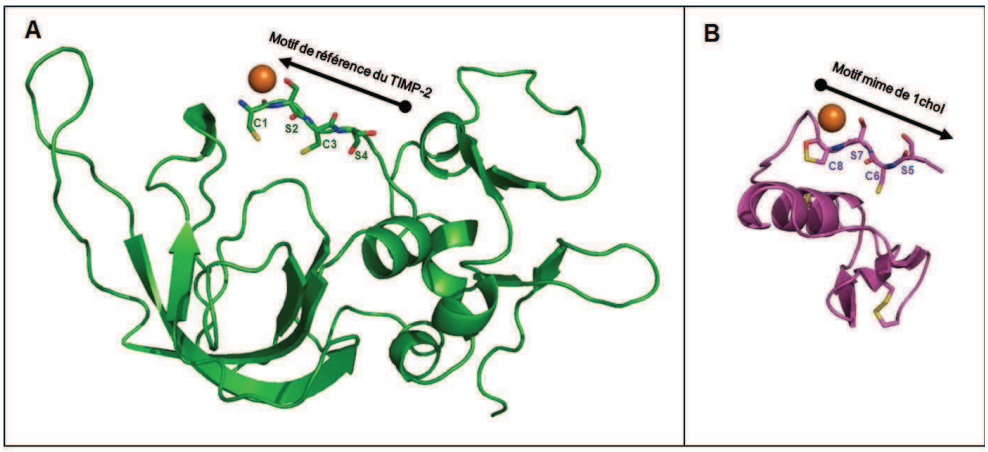


Figure II.2 : Orientations du motif mime par rapport au motif de référence. Les protéines sont représentées en mode cartoon, et le motif de liaison en mode bâtons. A : Ligand de référence TIMP-2 dont l'extrémité N-terminale du motif est orientée vers le zinc catalytique (sphère orange). B : Plate-forme mime (1cholI) dont l'extrémité N-terminale du motif est orientée à l'opposé du zinc catalytique.

Ces différentes étapes de sélection ont conduit à l'élimination de 161 plates-formes. Les 58 plates-formes restantes ont été triées selon des critères liés à leur structure 3D et à leur interface avec la cible.

Critères de sélection	Nombre de solutions **éliminées**	Nombre de solutions **retenues**
Redondance des fichiers PDB	106	113
Absence de ponts disulfures dans le motif	25	88
Orientation incorrecte du motif	30	58

- Structure 3D des protéines : des protéines différentes ayant des séquences primaires différentes peuvent avoir le même type de repliement et de structure secondaire. L'exemple type observé dans cette analyse concerne l'endothéline et la sarafotoxine, deux protéines

appartenant à la famille des endothélines, dont les séquences sont différentes mais possédant la même structure secondaire caractérisée par la présence d'une hélice α située en dessous du brin N-terminal et reliée à celui-ci par deux ponts disulfures (Figure II.3). De manière similaire aux solutions redondantes, pour chaque type de repliement, la plate-forme dont le motif possède le meilleur RMSD sur les positions des atomes du motif à greffer est retenue.

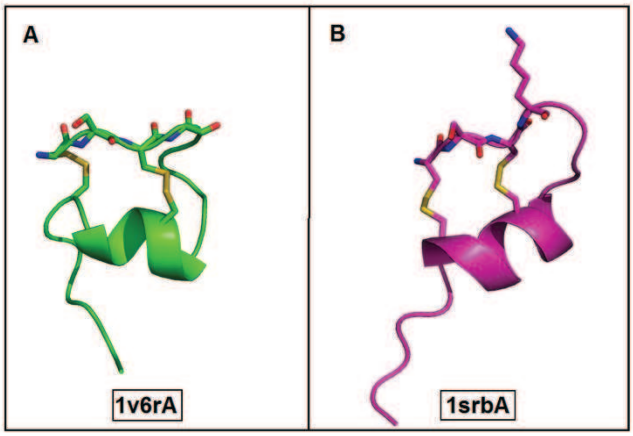


Figure II.3 : Repliement identique de deux protéines distinctes. Les protéines sont représentées en mode cartoon et les motifs en mode bâtons. A : Endothéline-1 (1v6rA) ; B : Sarafotoxine (1srbA).

- Aire de l'interface plateforme/cible trop faible : l'examen graphique des modèles moléculaires de complexes plateforme/cible permet de visualiser l'interface formée entre les plates-formes et la cible. Certaines plates-formes possèdent une large surface moléculaire mais forment cependant une interface trop réduite avec la cible. Ceci peut rendre l'interface ainsi formée difficile à exploiter pour les phases ultérieures d'optimisation (Figure II.4). Ces plates-formes dont l'aire de l'interface avec la cible est trop réduite sont donc éliminées.

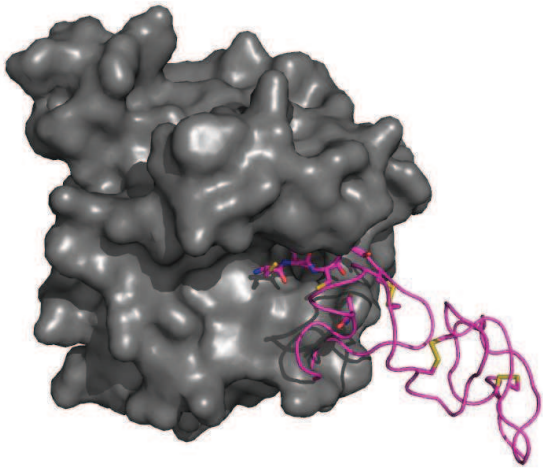


Figure II.4 : Interface non optimale entre une plate-forme (1tleA) représentée en mode cartoon (rose) et la cible MMP-14 (1bqq) représentée en mode surface (gris).

A la suite de la prise en compte de ces différents critères, 10 plates-formes ont été retenues pour des investigations *in silico* permettant de valider leur choix sur la base de critères utilisés pour caractériser les interactions dans les complexes protéine-protéine naturels. Nous avons déterminé un ensemble de critères topologiques, structuraux et électrostatiques et réalisé une analyse détaillée de l'interface de ces plates-formes en interaction avec la cible d'intérêt dans le mode de liaison imposé par la reproduction des interactions du motif fonctionnel greffé avec la MMP-14.

II.3 Analyse des plates-formes sélectionnées

Les 10 plates-formes sélectionnées grâce au programme STAMPS sont caractérisées par des critères topologiques et stériques ainsi que des critères de taille différents, rapportés dans le tableau II.1.

Tableau II.1 : Ensemble des 10 plates-formes protéiques identifiées *in silico* par STAMPS et sélectionnées graphiquement par PyMol.

Code PDB	Nom de la protéine	Motif natif	RMSD	ΔΘ	Steric inter/intra	Nb de résidus	Nb de ponts SS
2uuyB	Tick Derived Tryptase Inhibitor (TDPI)	C_{24}-T_{25}-V_{26}-P_{27}	0,36	9,7	1,05/0,25	52	4
2f3cI	Thrombin Inhibitor Infestin	C_6-A_7-C_8-P_9	0,48	16,7	1,18/0,30	46	3
1cz6A	Androctonin	S_2-V_3-C_4-R_5	0,54	17,8	1,76/0,17	25	2
1edpA	Endothelin 1	C_1-S_2-C_3-S_4	0,61	21,4	2,12/0,11	17	2
2it8A	Trypsin Inhibitor II	G_1-V_2-C_3-P_4	0,71	16,4	1,17/0,30	30	3
1an1I	Leech Derived Tryptase Inhibitor (LDTI)	K_2-V_3-C_4-A_5	1,05	47,4	2,27/0,28	40	3
1lmmA	Psalmotoxin 1	E_1-D_2-C_3-I_4	1,13	53,5	1,97/0,24	40	3
1osxA	Extracellular domain of BlyS Receptor 3 (BR3)	T_{17}-P_{18}-C_{19}-V_{20}	1,13	51,8	1,21/0,23	26	2
1scyA	Scyllatoxin	A_1-F_2-C_3-N_4	1,14	48,2	0,49/0,16	31	3
1rooA	SHK toxin	R_1-S_2-C_3-I_4	1,21	51,9	0,80/0,13	35	3

II.3.1 Origines et fonctions des plates-formes

Toutes les plates-formes retenues ont des fonctions natives sans relation entre elles et proviennent d'organismes différents tels que les arachnides (araignées : 1lmmA, scorpions : 1scyA et 1cz6A, tiques : 2uuyB), les échinodermes (étoile de mer : 1rooA), les annélides (sangsue médicinale : 1an1I), les insectes (triatome : 2f3cI), les plantes (citrouille : 2it8A) ou l'homme (1edpA, 1osxA).

Les protéines d'origine humaine identifiées, 1edpA et 1osxA, sont impliquées respectivement dans le système vasculaire (vasoconstricteur) et dans le système immunitaire (survie des lymphocytes B). Trois toxines sont présentes parmi les plates-formes retenues : la psalmotoxin 1 (1lmmA), la toxine SHK (1rooA) et la scyllatoxin (1scyA). Ces trois toxines sont des bloqueurs spécifiques des canaux Na+ (1lmmA) ou K+ (1rooA et 1scyA). L'androctonin (1cz6A) est un peptide antimicrobien de scorpion. Parmi les protéines identifiées, on distingue

un groupe de 4 protéines appartenant à la famille des inhibiteurs de protéases à sérine, dont des inhibiteurs de trypsine (2it8A), de tryptase (2uuyB et 1an1I) et de thrombine (2f3cI).

Les fonctions natives des plates-formes protéiques identifiées n'ont donc aucun lien entre elles et ne correspondent pas à des inhibiteurs naturels de haute affinité connus des MMPs. On peut noter que les structures de TIMPs n'ont pas été identifiées au cours de ces recherches car la taille maximale des protéines plates-formes était fixée à 60 résidus. Cependant, des recherches menées sans limiter la taille maximale des plates-formes ont permis d'identifier toutes les structures PDB correspondant à des TIMPs (Résultats non reportés).

II.3.2 Structures secondaires

Nous avons déterminé les structures secondaires et tertiaires des plates-formes retenues. Le type de repliement caractéristique de chaque plate-forme ainsi que la composition des plates-formes en éléments de structures secondaires sont détaillés dans le tableau II.2.

Tableau II.2 : Structures secondaires des 10 plates-formes protéiques identifiées *in silico* par STAMPS.

Code PDB	Type de Repliement	Hélice α	Feuillet β
2uuyB	Kunitz/BPTI-like	1 hélice α (9% ; 5 résidus)	4 feuillets β (30% ; 16 résidus)
2f3cI	Kazal	1 hélice α (20% ; 11 résidus)	4 feuillets β (21% ; 12 résidus)
1cz6A	Antimicrobial β-hairpin	x	2 feuillets β (32% ; 8 résidus)
1edpA	Endothelin like	1 hélice α (14% ; 3 résidus)	x
2it8A	Knottins	1 hélice α (10% ; 3 résidus)	3 feuillets β (13% ; 4 résidus)
1an1I	Kazal	1 hélice α (10% ; 5 résidus)	4 feuillets β (26% ; 12 résidus)
1lmmA	Knottins	x	3 feuillets β (22% ; 9 résidus)
1osxA	TNFR-like	x	3 feuillets β (14% ; 9 résidus)
1scyA	Knottins	1 hélice α (32% ; 10 résidus)	2 feuillets β (25% ; 8 résidus)
1rooA	Crisp domain-like	3 hélices α (40% ; 14 résidus)	x
TIMP-2	OB	6 hélices α (13% ; 25 résidus)	17 feuillets β (26% ; 48 résidus)

Les plates-formes protéiques ont des structures tridimensionnelles différentes, avec des tailles variant entre 17 et 52 acides aminés et un nombre de ponts disulfures compris entre 1 et 5 par protéine. Ainsi, on distingue les plates-formes possédant un seul type d'élément de structure secondaire régulier, soit de type hélice α (2 solutions), soit de type feuillet β (3 solutions), et les plates-formes ayant des structures secondaires mixtes d'hélices α / feuillets β (5 solutions). Dans les structures mixtes, on note une prépondérance des feuillets β avec en moyenne un nombre de 3 feuillets et une proportion de 22% par protéine, tandis qu'en moyenne une seule hélice α est présente par protéine avec une proportion de 15%.

L'agencement de ces structures secondaires au sein de la structure 3D de la protéine est différent d'une plate-forme à une autre créant une grande diversité de repliements (Figure II.5). Ces plates-formes diffèrent également des ligands de référence, par la taille, le nombre de ponts disulfures, la composition en éléments de structures secondaires et la structure 3D globale. En effet, les TIMP-1 et -2, dont les structures 3D existent pour la protéine entière, ont une taille de 184 acides aminés, 6 ponts disulfures et une structure secondaire composée de 6 hélices α et 12

(TIMP-1 dans la structure 1uea) ou 17 (TIMP-2 dans la structure 1bqq) feuillets β.

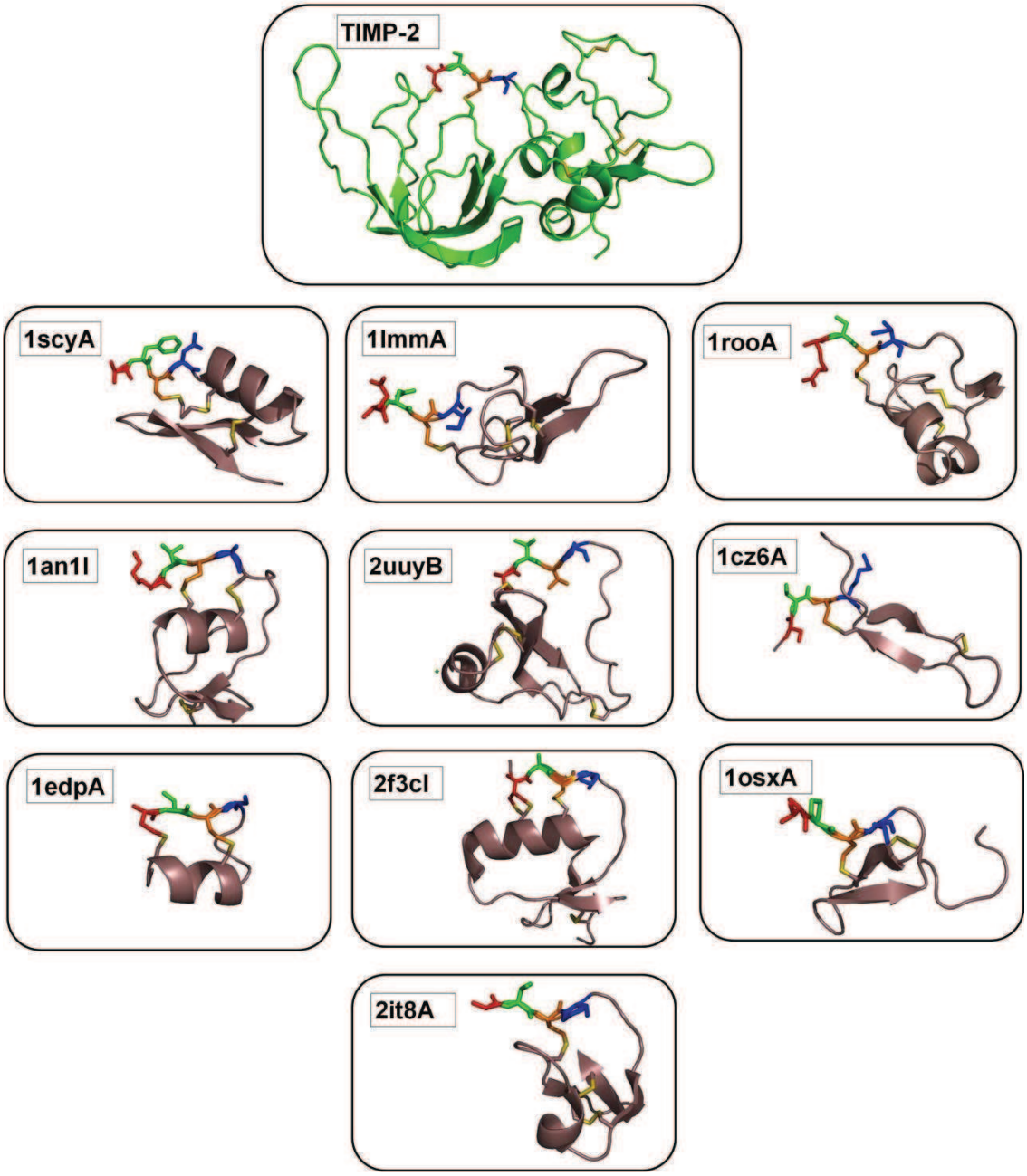


Figure II.5 : Structures 3D des 10 plates-formes protéiques retenues et du ligand de référence. Les protéines sont représentées en mode cartoon, en mauve pour les plates-formes et en vert pour le ligand de référence TIMP-2. Les ponts disulfures sont représentés en mode bâtons jaunes et les motifs de liaison sont représentés en mode bâtons de différentes couleurs : rouge pour le 1er résidu du motif, vert pour le 2ème, orange pour le 3ème et bleu pour le 4ème.

Ces résultats montrent que le protocole de recherche de plates-formes protéiques utilisant le programme STAMPS a permis de sélectionner des protéines de structures tertiaires très différentes qui sont néanmoins capables de supporter le motif 1-4 du TIMP-2 et de reproduire

sa topologie fonctionnelle. De plus, les premières observations faites à l'issue des différentes étapes de sélection suggèrent que, tout en reproduisant la topologie fonctionnelle du motif, ces différentes plates-formes sont susceptibles de présenter une complémentarité stérique et une taille d'interface comparable à celles observées dans des complexes protéine-protéine naturels. Toutefois, le critère utilisé par STAMPS pour caractériser la complémentarité stérique des deux partenaires n'est pas suffisamment précis pour permettre une comparaison aux observations faites dans des complexes protéine-protéine naturels. Ceci nous a incité à réaliser, pour ces 10 candidats ligands, des modèles moléculaires des complexes qu'ils sont susceptibles de former avec la cible MMP-14 à l'aide de méthodes de simulations utilisant le champ de force empirique CHARMM et qui seront présentés au paragraphe II.5.

II.3.3 Les motifs de liaison

Les motifs de liaison sur les protéines plates-formes identifiées ont une composition très différente d'une protéine à une autre, et différente également du motif de référence. En effet, la recherche effectuée à l'aide de STAMPS est indépendante de la nature des résidus du motif à greffer car elle repose uniquement sur la recherche de concordance topologique entre les atomes Cα et Cβ des résidus du motif sur le ligand naturel TIMP-2 et les atomes Cα et Cβ des résidus du motif présent dans la protéine plate-forme. Les chaines latérales ne sont donc pas prises en compte dans cette recherche. La composition des motifs de liaison est particulièrement riche en résidus Val et Pro. Sur un total de 10 plates-formes retenues, 6 possèdent un résidu Pro et/ou Val dans leur motif mime soit 60% des motifs. De manière plus générale, les motifs des 219 plates-formes identifiées par STAMPS ont été étudiés. Après la suppression des solutions redondantes, l'analyse des 113 plates-formes restantes a montré que 32% des plates-formes possèdent un résidu Pro dans leur motif de liaison et 24% des plates-formes possèdent un résidu Val.

Par ailleurs, malgré la grande diversité des structures 3D des plates-formes retenues, les motifs mimes au sein des plates-formes sont topologiquement similaires au motif présent dans la structure du TIMP-2 (Figure II.6), avec en particulier plusieurs plates-formes possédant des motifs mimes présentant un RMSD sur les positions des atomes pris en compte après superposition avec le motif de référence qui est très bas (compris entre 0,36 et 0,71 pour 2uuyB, 2f3cI, 1cz6A, 1edpA, 2it8A). Parmi les plates-formes identifiées *in silico*, on note la présence de l'endothelin 1 (1edpA) qui possède naturellement à son extrémité N-terminale la

séquence Cys_1-Ser_2-Cys_3-Ser_4 identique à celle présente dans le TIMP-2, avec les deux résidus Cys1 et Cys3 engagés dans des ponts disulfures.

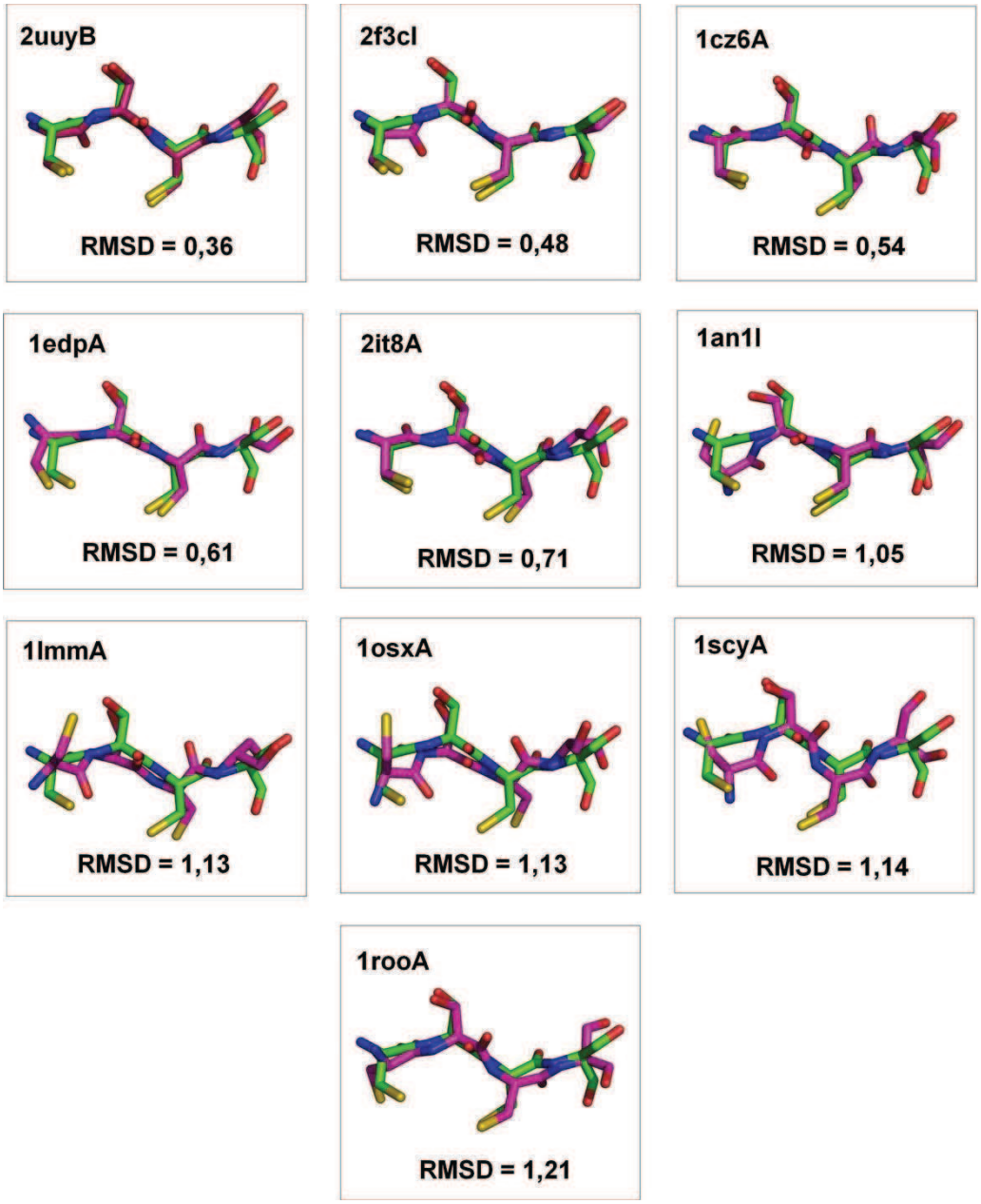


Figure II.6 : Superposition des motifs identifiés dans les différentes plates-formes avec le motif de référence. Les motifs sont représentés en mode bâtons, en rose pour le motif mime et en vert pour le motif du TIMP-2. Le code PDB au dessus des motifs indique la plate-forme à laquelle appartient le motif représenté en rose. Les valeurs de RMSD indiquées en dessous des motifs ont été calculées en superposant les atomes Cα et Cβ des résidus du motif mime natif (avant mutation des résidus) avec ceux des résidus du motif de référence. Pour plus de clarté sur la figure, les 4 résidus des motifs mimes ont été mutés en résidus Cys-Ser-Cys-Ser identiques à ceux du motif de référence.

II.4 Analyse des potentiels électrostatiques

Les potentiels électrostatiques des 10 plates-formes retenues ont été calculés à l'aide de l'équation de Poisson Boltzmann grâce au programme APBS (Méthodes). Comme nous l'avons vu précédemment, le complexe MMP-14/TIMP-2 présente une complémentarité électrostatique mise en évidence par l'examen des potentiels électrostatiques de surface des deux partenaires. Cette complémentarité électrostatique, impliquée dans la reconnaissance et l'association entre les deux partenaires, peut contribuer à l'affinité de liaison. Dans le cadre de cette approche mimétique de conception de ligands protéiques, nous avons choisi de rejeter les plates-formes qui présenteraient des potentiels électrostatiques significativement différents de celui du ligand naturel TIMP-2. Pour cette analyse, nous avons tout d'abord réalisé une analyse graphique des potentiels électrostatiques de surface des différents candidats ligands qui ont été comparés à celui du TIMP-2. Dans un deuxième temps, nous avons réalisé une comparaison quantitative de ces potentiels au voisinage du site d'interaction avec la cible.

II.4.1 Analyse graphique

La comparaison graphique des potentiels électrostatiques a été réalisée au niveau de la région des protéines contenant le site de liaison à la cible (Figure II.7). Pour ce faire, les différentes plates-formes ont été superposées au TIMP-2 via leurs motifs de liaison, puis tous les ligands ont été maintenus dans la même position et la même orientation les uns par rapport aux autres, afin de permettre la comparaison.

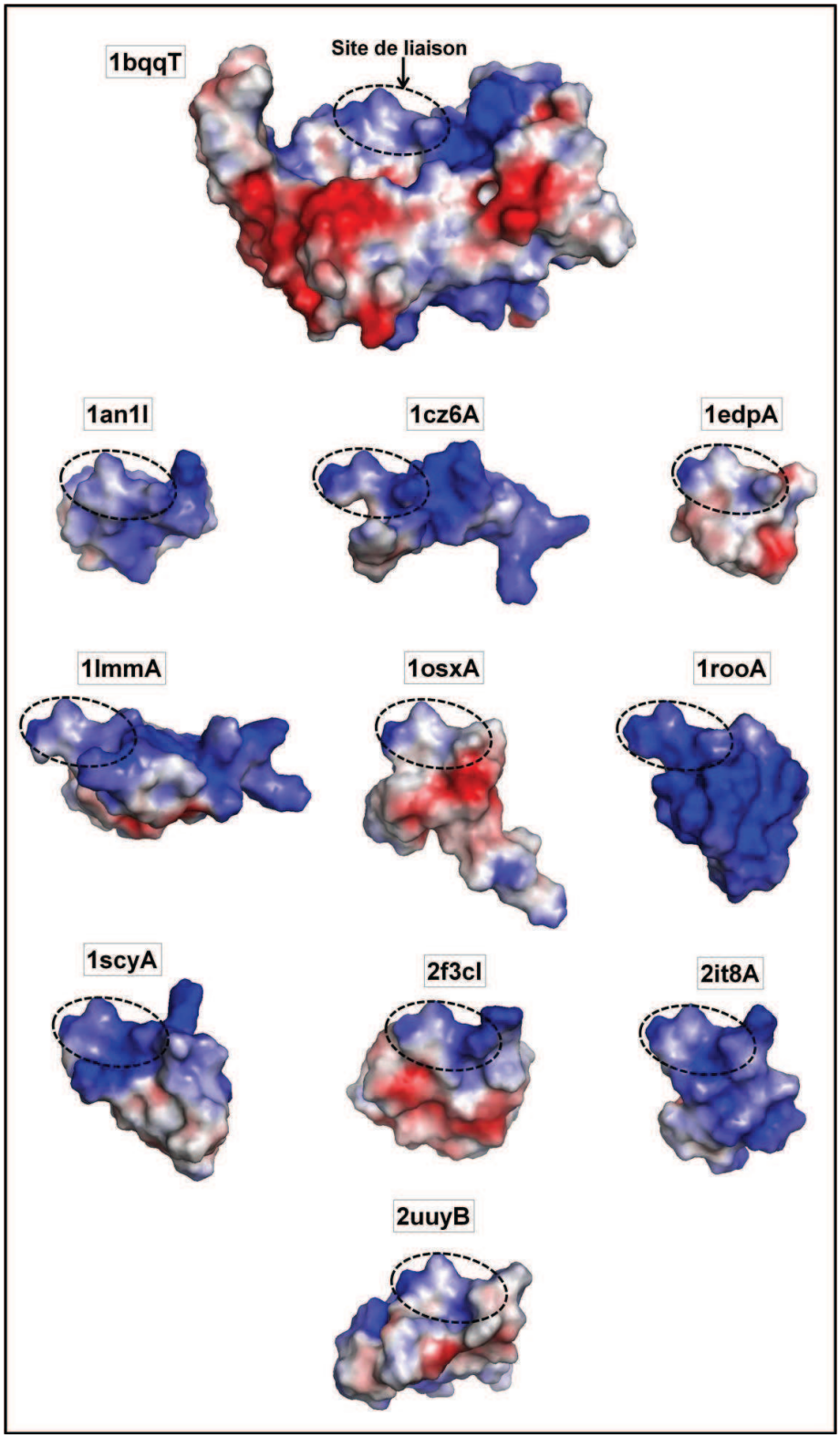


Figure II.7 : Comparaison des potentiels électrostatiques de surface des 10 plates-formes sélectionnées calculés par le logiciel APBS grâce à la résolution de l'équation de Poisson-Boltzmann, et visualisés sur la surface moléculaire des protéines accessible au solvant. Les surfaces bleues correspondent à un potentiel électrostatique positif, celles en rouge indiquent un potentiel négatif (échelle de potentiel : -2/+2). La région contenant le site de liaison des différents ligands pour les MMPs est entourée d'une ellipse noire. Dans ces calculs, les états de protonation des groupements acido-basiques ont été définis à pH 7 suivant la valeur de leur pK. Le groupement amino-terminal est sous forme NH_3^+.

L'analyse graphique des potentiels électrostatiques représentés sur la figure II.7 indique que les 10 plates-formes étudiées possèdent un potentiel électrostatique similaire à celui du TIMP-2 au niveau du site de liaison aux MMPs (zone encerclée en noir). Les sites de liaison pour les candidats ligands ainsi que pour le TIMP-2 sont caractérisés par un potentiel électrostatique positif. Cependant, des variations d'amplitude modérées de l'intensité de ces potentiels sont observées selon les ligands. Ainsi, les plates-formes 1edpA et 1osxA sont caractérisées par un potentiel électrostatique positif d'intensité plus faible que celui du TIMP-2 au niveau du site de liaison, tandis que les plates-formes 1rooA et 1cz6A possèdent des potentiels positifs d'intensité plus importante au niveau de ce site. Les régions des candidats-ligands distantes du motif de liaison greffé sont caractérisées par des potentiels électrostatiques différents selon les plates-formes. Cette observation importante pourrait être exploitée ultérieurement dans la phase d'optimisation des ligands protéiques de première génération obtenus à l'issue de la phase de conception *in silico*.

II.4.2 Analyse des indices de similarité de Hodgkin

La comparaison quantitative des potentiels électrostatiques des différents candidats ligands avec celui du ligand protéique naturel, le TIMP-2, a été réalisée grâce au programme PIPSA. Ce programme permet de comparer, deux-à-deux, les grilles de potentiels obtenues par la résolution de l'équation de Poisson-Boltzmann (Méthodes) avec le programme APBS. En particulier, la comparaison des grilles de potentiel électrostatique a été réalisée dans une région définie au niveau de la surface moléculaire correspondant au site de liaison avec la cible (Méthodes). Un indice de similarité électrostatique de Hodgkin (ESI_part) a été calculé pour chaque couple candidat ligand/TIMP-2 afin de comparer le potentiel électrostatique du candidat ligand dans la région de liaison à la cible à celui du ligand naturel TIMP-2 (Figure II.8).

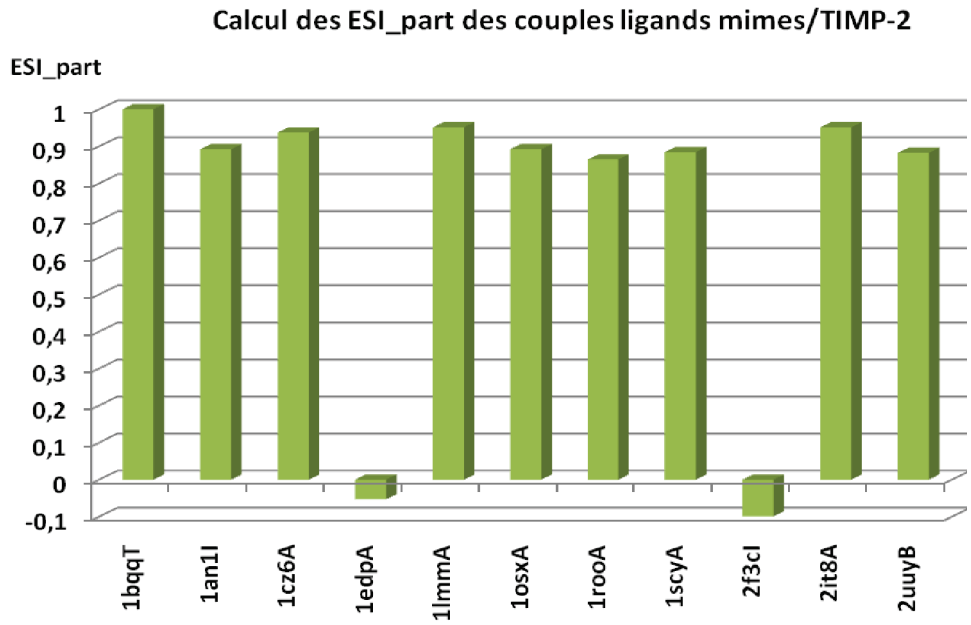


Figure II.8 : Indices de similarité électrostatique de Hodgkin (ESI) pour les couples candidats ligands /TIMP-2 dans la région de liaison à la cible.

L'analyse des indices de similarité électrostatique a montré que toutes les plates-formes, à l'exception de 1edpA et 2f3cI, possèdent des potentiels dans la région étudiée qui sont très similaires à celui du TIMP-2 (ESI_part compris entre 0,864 pour 1rooA et 0,951 pour 1lmmA et 2it8A). Les valeurs de ESI_part déterminées pour les candidats ligands 1edpA et 2f3cI sont proches de 0 (-0,053 et -0,099 respectivement) signifiant que les potentiels électrostatiques de ces deux plates-formes ne présentent pas de corrélation avec celui du TIMP-2. Comme nous l'avons montré au Chapitre II, des potentiels électrostatiques faiblement corrélés ou non corrélés peuvent être observés lors de la comparaison de ligands protéiques naturels capables de cibler un même site de liaison (Magis et al., soumis). Par exemple, la comparaison des potentiels électrostatiques de deux ligands naturels de la trypsine porcine : Ecballium elaterium trypsin inhibitor II (code PDB du complexe 1h9h) et trypsin-ovomucoid turkey egg white inhibitor (code PDB du complexe 1z7k), a conduit à une valeur de ESI_part de -0,01. De la même manière, une valeur de ESI_part de -0,30 a été obtenue lors de la comparaison de deux autres ligands naturels de la trypsine porcine : human tissue factor pathway inhibitor (code PDB 1tfx) et un peptide de 9-mer de la trypsine (code PDB 1uhb). Dans cette même étude, la comparaison deux à deux de 8 ligands naturels de la chymotrypsine a permis de démontrer que 6 paires de ligands sont caractérisées par des ESI_part négatifs compris entre -0,02 et -0,18.

Par ailleurs, la similarité de forme entre les différentes plates-formes et le ligand de référence a également été analysée. Le programme PIPSA a permis de calculer un indice de similarité de forme de Hodgkin (SSI) permettant de comparer les grilles de calcul de deux protéines au niveau de leurs régions « skin » en déterminant le nombre de points de ces grilles qui se trouvent à l'intersection des régions « skin » des deux protéines (Méthodes).

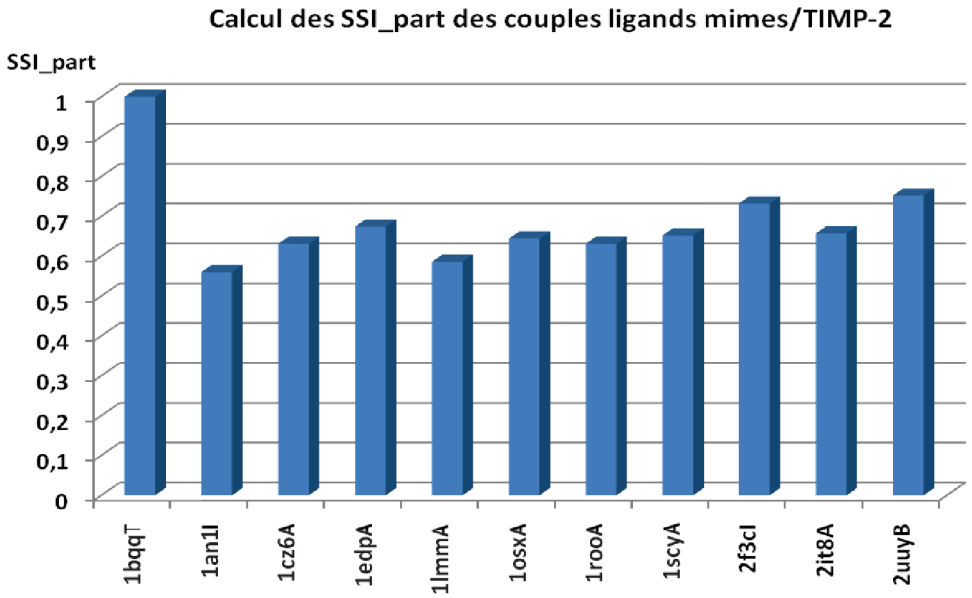


Figure II.9: Indices de similarité de forme de Hodgkin (SSI) pour les couples candidats ligands /TIMP-2 dans la région de liaison à la cible.

Cette analyse montre que toutes les plates-formes possèdent une similarité de forme satisfaisante avec le TIMP-2 dans la région de liaison avec la cible (Figure II.9). Les valeurs de SSI_part varient de 0,560 (1an1I) à 0,762 (2uuyB), avec une moyenne de 0,652.

A la lumière de ces indices de comparaison, les plates-formes sélectionnées selon les critères topologiques, stériques et structuraux précédemment décrits paraissent également capables de reproduire les critères électrostatiques et géométriques caractéristiques du ligand de référence. L'étude de ces 10 complexes a donc été poursuivie par la réalisation de modèles moléculaires des complexes candidats ligands / MMP-14 afin de pouvoir analyser en détails les interactions intermoléculaires susceptibles de s'établir entre les deux partenaires et de comparer les paramètres caractéristiques des interfaces ainsi formées aux valeurs observées dans des complexes protéine-protéine naturels.

II.5 Analyse des interfaces plateforme/cible

De façon similaire au protocole appliqué pour l'analyse de l'interface du complexe de référence TIMP-2/MMP-14, les complexes candidat ligand/MMP-14 ont été soumis à une procédure de relaxation par dynamique moléculaire et minimisation d'énergie sous contraintes positionnelles (Méthodes). Pour ces modèles moléculaires, les structures initiales des complexes candidat ligand/MMP-14 sont celles obtenues en sortie de STAMPS. L'utilisation de contraintes positionnelles et de contraintes de distances au niveau de la cible et du motif greffé sur le candidat ligand permet de préserver la structure 3D de la MMP-14 ainsi que la topologie du motif fonctionnel mime. Les distances de référence, imposées entre les 4 résidus du motif mime et les résidus de la cible avec lesquels ils vont interagir, sont déterminées à partir du complexe TIMP-2/MMP-14 (Structure 1bqq). Par ailleurs, des contraintes de distances intramoléculaires entre les Cα du candidat ligand ont été imposées pour préserver la structure 3D initiale du candidat ligand tout en autorisant des mouvements d'ensemble de celui-ci par rapport à la cible. Cette procédure permet la correction des problèmes stériques éventuels ou des interactions électrostatiques défavorables dans le modèle de départ. Elle permet également l'optimisation de la taille de l'interface au cours de la phase de relaxation. Pour chaque complexe, la structure obtenue à l'issue du dernier cycle de relaxation a été analysée et les caractéristiques des l'interfaces du complexe candidat ligand/MMP-14 ont été déterminées. Les principales caractéristiques relatives à l'interface du ligand dans sa conformation liée à la cible sont résumées dans le tableau II.3.

Tableau II.3 : Analyse des interfaces des plates-formes mimes en complexe avec la MMP-14 portant, dans l'ordre, sur le nombre de résidus du ligand, la taille de l'interface totale du complexe, la taille de l'interface du ligand avec la MMP, le nombre de résidus du ligand situés à l'interface avec la cible, la fraction de l'interface du ligand correspondant à des groupes non-polaires, le nombre de liaisons hydrogène/100 Å^2 de surface totale enfouie, le nombre d'interactions cation-π (Nb de CP), le nombre de ponts salins (Nb de PS) et le « gap volume index » rapporté à l'aire de l'interface du ligand.

Code PDB	Nb de résidus	Interface totale (Å^2)	Interface ligand (Å^2)	Nb de résidus à l'interface	% Fraction Non Polaire	Nb de HB/100 Å^2	Nb de CP	Nb de PS	GVI ligand (Å)
1edpA	17	1230	664	12	22	0,57 (5)	0	0	1,92
1an1I	40	2182	1073	20	27	0,59 (13)	1	1	2,15
1scyA	31	1445	771	15	15	0,73 (11)	0	2	1,91
2uuyB	52	2029	1047	22	30	0,82 (12)	1	1	1,87
1lmmA	40	1821	917	19	21	0,82 (15)	0	1	1,61
1rooA	35	1204	672	12	15	0,49 (6)	1	1	1,67
2f3cI	46	1514	826	15	8	0,66 (13)	0	0	1,84
2it8A	30	1370	786	13	18	0,65 (9)	1	0	1,70
1cz6A	25	1561	891	14	16	0,77 (12)	1	1	1,62
1osxA	26	1086	592	11	17	0,55 (6)	2	0	1,84
TIMP-2	184	3190	1643	36	42	1,03 (17)	0	0	1,95

L'analyse de ces modèles montre une absence de corrélation entre la taille de la plate-forme et la taille de l'interface. En effet, la plate-forme 1osxA de taille (26 résidus) supérieure à la plus petite plate-forme 1edpA (17 résidus) a cependant la plus petite interface (592 Å^2). De même, la plate-forme 1rooA de taille moyenne (35 résidus) possède également une interface très petite (672 Å^2) tandis que 1cz6A avec une petite taille (25 résidus) possède une grande interface (891Å^2).

Sur les 10 plates-formes analysées, la variation totale de surface accessible au solvant consécutive à l'association des deux partenaires (interface plate-forme + interface cible) est comprise entre 1086 Å^2 pour le candidat ligand 1osxA et 2182 Å^2 pour 1an1I. Ces observations montrent que les critères appliqués ont permis de sélectionner, dans la PDB,10 plates-formes de faible poids moléculaire (taille maximale de 60 résidus) pouvant conduire à des ligands protéiques susceptibles d'entrainer une variation totale de surface accessible au solvant située dans la gamme de celles observées dans la littérature pour des complexes protéine-protéine

naturels (1600 ± 400 $Å^2$; Lo Conte et al., 1999). Les tailles des interfaces des candidats ligands (Tableau II.3) confirment cette observation. Ces résultats montrent que le programme STAMPS a permis d'exploiter la diversité des structures de protéines de la PDB et d'identifier des plates-formes dont les caractéristiques structurales leur permettraient de satisfaire les critères topologiques requis pour la greffe du motif fonctionnel et offrant une interface de taille « standard » avec la cible.

Parmi les 10 complexes analysés, on observe que le nombre de résidus enfouis lors de l'association avec la cible est compris entre 11 et 22 selon la plate-forme. La composition chimique des interfaces est variable en fonction des plates-formes. Celle-ci peut être quantifiée en calculant la fraction de l'interface du candidat ligand correspondant à des groupements non polaires : parmi les 10 ligands, cette fraction varie entre 8 et 30%. On voit donc que la sélection réalisée en parcourant la PDB permet d'identifier des protéines conduisant à des interfaces de nature physico-chimique variée.

Par ailleurs, parmi les résidus présents à l'interface, un nombre variable d'entre eux participe à des interactions intermoléculaires de type liaison hydrogène, pont salin ou interaction cation-π. Parmi les 10 candidats ligands, ces résidus sont compris entre 5 (1edpA) et 15 (1lmmA). Cette importante variation doit être pondérée par la variation de l'aire de l'interface. Ainsi, parmi les différents ligands, la densité superficielle de liaisons hydrogène (calculée comme le rapport du nombre de liaisons hydrogène sur la variation totale de surface accessible au solvant, Tableau II.3) se situe dans la gamme de ce qui a été observé dans des complexes protéine-protéine naturels. Dans la plupart des cas, les 4 résidus du motif Cys-Ser-Cys-Ser contribuent de manière prépondérante au nombre d'interactions intermoléculaires observées à l'interface (6 interactions intermoléculaires avec la cible). Selon les modèles moléculaires réalisés, les plates-formes sélectionnées sont donc capables de reproduire la majorité des interactions intermoléculaires observées pour le motif 1-4 du TIMP-2 dans le complexe avec la MMP-14. On peut cependant noter que les modèles de complexes sont réalisés sous contraintes de manière à préserver la structure du complexe de départ et en particulier à imposer la conformation de la structure du motif 1-4 pour que celle-ci reste voisine de la structure adoptée par ces résidus dans le TIMP-2. Ces modèles ne permettent donc pas d'estimer la stabilité des interactions que pourraient établir les différents candidats ligands avec la MMP-14. Ce point sera étudié ultérieurement. Les résultats ci-dessus permettent cependant de valider la sélection

des plates-formes et de montrer que celles-ci sont stériquement compatibles pour l'association avec la cible considérée sans nécessiter de déformation importante de la structure 3D de la plate-forme, dans un mode de liaison qui conduit à reproduire les interactions intermoléculaires observées entre le motif 1-4 du TIMP-2 et la MMP14 et en offrant une interface de taille « standard ».

D'autres types d'interactions sont observés dans les modèles moléculaires des 10 complexes : interactions cation – π et ponts salins. Ces interactions impliquent des résidus du candidat ligand, présents à l'interface avec la MMP-14, autres que les résidus du motif greffé sur ces plates-formes. Elles sont donc fortuites car aucun critère concernant ces interactions n'a été appliqué lors de la recherche de plate-forme avec STAMPS. Dans le cas des ponts salins, ces interactions résultent de la présence d'un résidu chargé à la surface de la plate-forme, qui se trouve positionné à proximité d'un résidu de charge opposée à la surface de la cible. Le positionnement relatif des deux groupes autorisant leur interaction est une conséquence du mode de liaison du candidat ligand imposé par la présence du motif greffé. Il en va de même pour les interactions cation-π qui résultent du positionnement fortuit d'un groupement aromatique à l'interface d'un des partenaires à proximité d'un groupement portant une charge positive à la surface de l'autre partenaire. Ces interactions sont donc variables d'un candidat ligand à l'autre (Tableau II.3).

Les modèles moléculaires réalisés permettent également d'évaluer la compacité des interfaces. Celle-ci est mesurée par un indice appelé « gap volume index » (Méthodes), qui varie entre 1,61 (1lmmA) et 2,15 Å (1an1I) parmi les 10 interfaces analysées. Les valeurs de ces indices se situent dans la gamme observée pour des complexes protéine-protéine naturels et indiquent une bonne complémentarité géométrique des surfaces des partenaires au niveau du site de liaison.

II.6 Conclusion

La recherche de plates-formes *in silico* grâce au programme STAMPS a permis d'identifier 219 plates-formes protéiques satisfaisant les critères topologiques du motif de liaison ainsi que les critères stériques liés à la compatibilité géométrique avec la cible. Plusieurs étapes supplémentaires de sélection selon des critères structuraux ont permis de retenir 10

plates-formes possédant les caractéristiques structurales compatibles avec la liaison au site catalytique des MMPs. Ces plates-formes possèdent des séquences et des structures secondaires et tertiaires différentes entre elles et différentes du TIMP-2. Les 10 plates-formes mutées ont été validées grâce à une série d'analyses complémentaires. Deux critères globaux de similarité de forme et de potentiel électrostatique ont été utilisés. Ces analyses ont montré que les candidats ligands obtenus possèdent des formes et des potentiels électrostatiques suffisamment similaires à ceux du ligand naturel, le TIMP-2, dans la région ciblant le site catalytique des MMPs, pour qu'une seconde étape de validation soit entreprise. A cette fin, des modèles moléculaires affinés des complexes entre chacun des 10 candidats ligands et la MMP-14 ont été réalisés par dynamique moléculaire sous contrainte. L'analyse de ces modèles a permis de déterminer une série de paramètres caractéristiques des interactions protéine-protéine. Les résultats observés ont permis de valider *in silico* les 10 plates-formes comme candidats ligands. Nous avons donc entrepris la production de ces 10 plates-formes afin de réaliser leur évaluation expérimentale comme inhibiteurs de MMPs.

CHAPITRE III : Production des ligands conçus *in silico* et caractérisation comme inhibiteurs de MMPs

Les 10 candidats ligands protéiques conçus *in silico* ont été produits selon deux protocoles différents, en fonction des plates-formes. Certains candidats ligands ont été produits par synthèse peptidique sur support solide et d'autres ont été produits par voie recombinante chez la bactérie *E. coli*. Ces plates-formes doivent porter les résidus du motif fonctionnel d'intérêt, avoir l'extrémité N-terminale du motif libre et être correctement oxydées et repliées. Les plates-formes ainsi produites et purifiées sont soumises à plusieurs analyses et étapes de caractérisation afin de vérifier la masse, la pureté et le repliement correcte de la protéine finale. Ces plates-formes sont ensuite évaluées en tant qu'inhibiteurs des MMPs. Dans la suite, la plate-forme portant les résidus spécifiques du motif fonctionnel est désignée par plateforme_m et la plate-forme native (possédant des résidus topologiquement identiques à ceux du motif à greffer mais pour lesquels les substitutions n'ont pas été réalisées) est désignée par plate-forme. Par ailleurs, les mini-protéines natives 1edpA, 1scyA, 1lmmA, 1rooA ainsi que l'endothelin-3 (dont il n'existe pas de structure tridimensionnelle), étant disponibles dans le commerce, n'ont pas été synthétisées. Ces protéines commerciales ont également été testées.

III.1 Production des mini-protéines conçues *in silico* par synthèse chimique

La production des plates-formes 2f3cI, 2uuyB, 2it8A ainsi qu'une forme tronquée de 1edpA a été réalisée par synthèse chimique. Ces plates-formes ont été obtenues par synthèse peptidique sur support solide en utilisant la stratégie Fmoc/tBu. Après une étape d'oxydation en présence du couple oxydo-réducteur glutathion oxydé (GSSG)/glutathion réduit (GSH), les protéines repliées sont purifiées par RP-HPLC semi-préparative.

III.1.1 Synthèse sur support solide

Les séquences des protéines synthétisées 2f3cI, 2uuyB, 2it8A ainsi que la forme tronquée de 1edpA sont indiquées en Annexe II. Quelques modifications par rapport aux séquences natives ainsi que des modifications par rapport au protocole standard de synthèse peptidique

ont été apportées lors de la synthèse de certaines plates-formes ; ces modifications sont détaillées ci-dessous. Pour certaines protéines parmi celles citées plus haut, la synthèse par voie chimique n'a jamais été décrite dans la littérature (2uuyB et 2f3cI), ce qui a nécessité la mise au point d'un protocole adapté.

- Spécificités des séquences à synthétiser et rendements de l'étape de synthèse sur support solide

- **1edpA (Endothelin 1)** : l'endothelin 1 sous sa forme native possède 21 résidus. La structure 1edpA de la plate-forme identifiée lors de la recherche *in silico* par le programme STAMPS correspond à une forme tronquée de la protéine native possédant 17 résidus. Cette forme tronquée ne possède pas les 4 résidus C-terminaux $D_{18}I_{19}I_{20}W_{21}$. L'extrémité C-terminale est caractérisée par une très grande flexibilité illustrée par les disparités structurales des 20 conformations décrites dans la structure RMN correspondant au code PDB 1v6rA. Une analyse détaillée d'un modèle moléculaire de complexe réalisé sur la forme entière de l'endothelin 1 (structure cristallographique correspondant au code PDB 1ednA) avec le domaine catalytique de la MMP-14 montre que la partie C-terminale très flexible ne semble pas intervenir dans l'interaction avec la MMP, hormis le résidu D18 impliqué dans une liaison hydrogène avec un résidu de la cible. Ainsi, nous avons choisi de produire une forme tronquée de l'endothelin 1 sans les 3 derniers résidus $I_{19}I_{20}W_{21}$ (1edpA 1-18), en gardant donc le résidu D18.

- **2f3cI (Thrombin Inhibitor Infestin)** : la structure cristallographique de la plate-forme native 2f3cI possède un résidu Asp5 en position N-terminale de la protéine et du motif de liaison Cys_6-Ala_7-Cys_8-Pro_9. Le motif fonctionnel devant avoir son extrémité N-terminale libre de façon à ce que le 1er résidu Cys6 de ce motif puisse chélater le zinc, le résidu Asp5 a été éliminé de la séquence à synthétiser. Par ailleurs, la protéine 2f3cI contient une séquence de deux acides aminés $D_{19}G_{20}$ dont l'incorporation successive par synthèse chimique est susceptible de générer des difficultés lors de l'étape de couplage. Nous avons donc incorporé au cours de la synthèse, le bloc Fmoc-Asp-otbu-hmb-Gly-OH (1mmol) correspondant aux acides aminé D19 et G20 de la séquence primaire de la protéine.

- **2uuyB (Tick Derived Tryptase Inhibitor)** : constituée de 52 résidus, la séquence de cette protéine est la plus longue parmi les plates-formes retenues. Cette séquence est caractérisée par

une succession de certains résidus susceptibles d'entrainer des difficultés de synthèse par voie chimique. Contrairement aux résines utilisées dans les autres synthèses, on utilise dans le cas de 2uuyB une résine pré-fonctionnalisée avec le dernier résidu de la séquence (acide aminé C-terminal : Ala) déjà fixé sur la résine : Fmoc-Ala NovaSyn TGA. Cette résine est en outre caractérisée par un faible taux de chargement (0,26 mmol/g) et par la présence d'un bras espaceur de polyéthylène glycol (PEG) entre le linker et les billes de polystyrène, permettant d'augmenter l'accessibilité au peptide en élongation et d'éviter l'encombrement stérique. Une 1ère synthèse réalisée dans ces conditions a montré une chute brutale du signal de suivi de déprotection après l'incorporation de la séquence $G_{14}G_{15}$ conduisant à un signal résiduel de suivi de déprotection de 22% à la fin de la synthèse. Une analyse du produit final brut de synthèse par LC/MS a montré la présence de plusieurs pics correspondant à des peptides tronqués, essentiellement au niveau du dipeptide $G_{14}G_{15}$ tandis que le peptide d'intérêt n'est pas visible sur le spectre LC/MS. Une 2ème synthèse est alors réalisée, en prolongeant le temps de gonflage de la résine dans le NMP afin de permettre une meilleure accessibilité à tous les sites de la résine et éviter un effet d'encombrement stérique dû à la taille importante de la séquence à synthétiser. Dans ces conditions, le peptide d'intérêt a pu être détecté dans le produit brut de synthèse.

- **2it8A (Trypsin Inhibitor II)** : la séquence native de cette protéine est synthétisée, sans aucune mutation ou modification.

Par ailleurs, les résidus Met présents dans la séquence de certaines protéines sont susceptibles de subir une oxydation en solution au cours des différentes étapes de la synthèse chimique. Certaines synthèses réalisées au cours de ce travail dans lesquelles les résidus Met ont été maintenus ont montré que cette oxydation n'est pas systématique. Cependant, pour éviter cet éventuel problème, les résidus Met ont été substitués par des résidus Nle (Norleucine) lors des synthèses de 1edpA (1-18) et 2uuyB. Les deux résidus Met et Nle sont en effet analogues et de configuration identique à l'exception du soufre (-S-) dans la chaine latérale de Met remplacé par un méthylène ($-CH_2-$) dans le résidu Nle.

- <u>Rendements de l'étape de synthèse peptidique sur support solide</u>

L'efficacité de l'étape de synthèse peptidique sur support solide est mesurée par un suivi UV

lors de l'incorporation de chaque acide aminé de la séquence peptidique. En effet, lors de l'incorporation d'un acide aminé au peptide en élongation, un suivi de l'absorbance UV est réalisé au cours de l'étape de déprotection du groupement Fmoc porté par le dernier acide aminé. Le suivi UV permet donc de mesurer l'efficacité du couplage du dernier acide aminé incorporé et de détecter les positions de la séquence susceptibles d'entrainer des difficultés.

Tableau III.1 : Rendements de l'étape de synthèse peptidique déterminés par le signal de déprotection résiduel à la fin de la synthèse. ND : Non déterminé

Protéine	**Taille de la protéine**	**Rendement théorique**	**Rendement expérimental**
1edpA (1-18)	18	84%	ND
2it8A	29	76%	84%
2f3cI	50	60%	40%
2uuyB	52	59%	50%

L'efficacité de couplage théorique moyenne est de 99% à chaque cycle d'incorporation d'un acide aminé. De ce fait, le rendement de synthèse diminue en fonction du nombre de résidus de la séquence à synthétiser. Le rendement expérimental de l'étape de synthèse proprement dite des protéines est rapporté dans le tableau III.1. Dans le cas de 2it8A, le rendement expérimental est de 84%, suggérant que l'efficacité de couplage pourrait être supérieure à 99% (efficacité = 99,5%). Tandis que pour les plates-formes 2f3cI et 2uuyB, l'efficacité de couplage est inférieure à la valeur moyenne étant donné les faibles rendements obtenus par rapport aux rendements théoriques.

Plusieurs facteurs peuvent être limitants lors de la synthèse peptidique :

- La taille de la protéine : plus la taille est importante plus les rendements de synthèse diminuent du fait de l'encombrement stérique qui se produit lors de l'élongation de la chaine peptidique pouvant rendre inaccessibles certains sites de la résine.
- La composition en acides aminés de la séquence : comme dans le cas de la plate-forme 2uuyB, la succession des deux acides aminés $G_{14}G_{15}$ dans la séquence peptidique entraine une baisse importante du rendement de synthèse. En effet, après incorporation des acides aminés $G_{14}G_{15}$ le signal de suivi de déprotection chute brutalement conduisant à un signal résiduel à

50% en fin de synthèse. Il en est de même pour la plate-forme 2f3cI pour laquelle la synthèse se déroule normalement jusqu'à l'incorporation du bloc DG où le signal résiduel de suivi de déprotection passe à 40% et demeure constant jusqu'à la fin de la synthèse.

III.1.2 Clivage du produit de synthèse final

La durée de l'étape de clivage du produit de synthèse final de la résine résulte d'un compromis entre une efficacité de clivage qui augmente lorsque la durée de clivage augmente et une durée d'exposition du produit final aux réactifs de clivage qui doit être aussi brève que possible afin d'éviter des modifications chimiques au niveau de certains acides aminés. Dans un protocole standard, la durée optimale de clivage est de 2h. Cependant, dans le cas des protéines ayant des rendements de synthèse faibles tels que 2uuyB et 2f3cI, la durée de l'étape de clivage est prolongée (2h30 au lieu de 2h). Par ailleurs, un rinçage supplémentaire de la résine avec du TFA, après précipitation du produit de synthèse dans du TBME à froid, peut être effectué pour récupérer le maximum de produit de synthèse résiduel sur la résine. Dans ce cas, le TFA doit par la suite être évaporé. Cette étape supplémentaire est indispensable dans le cas de la plate-forme 2f3cI : le point isoélectrique de cette plate-forme étant très bas (pI = 5,26), une grande quantité de TFA risque de solubiliser une partie de la protéine présente dans le TBME. L'étape d'évaporation du TFA avant la centrifugation permet donc d'éviter une perte de rendement due à la solubilisation de la protéine par le TFA.

III.1.3 Purification du produit brut de synthèse par RP-HPLC

Avant la purification par RP-HPLC, le produit brut de synthèse est d'abord réduit dans un tampon dénaturant additionné de DTT. Cette étape est destinée à homogénéiser le mélange de produits contenus dans le brut de synthèse afin que le peptide d'intérêt soit élué en un seul pic correspondant à une seule forme entièrement réduite. Par la suite, la purification du peptide d'intérêt se fait par RP-HPLC sur des colonnes C18 en utilisant des gradients non linéaires de solvant organique dans un solvant aqueux. Lors de la purification, plusieurs pics apparaissent au niveau du chromatogramme, correspondants aux différents produits de synthèse incomplets. Ainsi, plus le rendement de l'étape de synthèse est faible, et plus on observe de pics sur le chromatogramme. De ce fait, une étape d'optimisation du gradient de séparation est nécessaire

pour chaque protéine afin de permettre une meilleure séparation du pic correspondant au produit d'intérêt des pics avoisinants plus ou moins nombreux et plus ou moins proches. Le rendement de cette étape de purification dépend également des caractéristiques physico-chimiques des protéines à purifier. En effet, la purification d'une protéine d'intérêt sur une colonne C18 en utilisant un mélange de solvants organiques et aqueux dépend essentiellement de la solubilité de cette protéine dans la phase mobile. L'indice GRAVY (Grand average of hydropathicity) caractéristique de chaque protéine donne une indication sur la solubilité de celle-ci (Kyte et Doolittle, 1982) : un indice GRAVY positif est caractéristique d'une protéine hydrophobe tandis qu'un indice négatif correspond à une protéine hydrophile. Les protéines 2f3cI, 2uuyB et 2it8A caractérisées par un GRAVY négatif sont correctement purifiées en suivant ce protocole. En revanche, l'endothelin-1 (1-18) est très hydrophobe (GRAVY $\geq$ 0,289) impliquant des phénomènes d'adsorption du peptide sur la phase hydrophobe de la colonne et une élution incomplète d'où une diminution importante du rendement de purification pour ce peptide.

Outre les limitations relatives aux rendements de purification, cette étape est également limitante du point de vue du temps requis pour purifier des quantités de l'ordre de 150 à 200 mg de produit brut de synthèse.

III.1.4 Formation des ponts disulfures

Le même protocole standard pour la formation des ponts disulfures est utilisé pour toutes les plates-formes. La formation des ponts disulfures a été réalisée en présence du couple oxydo-réducteur glutathion oxydé (GSSG) / glutathion réduit (GSH) à un ratio de 1/10. La durée moyenne de l'étape de formation de ponts disulfures, déterminée par un suivi de la disparition au cours du temps du pic correspondant à la forme réduite de la protéine, est de 48h. Le rendement de l'étape d'oxydation dépend du nombre de ponts disulfures présents dans la protéine. Par exemple, dans le cas de 2uuyB, cette étape est particulièrement limitante du fait de la présence de 4 ponts disulfures. En effet, lors de la purification du produit oxydé par RP-HPLC, plus le nombre de ponts disulfures d'une protéine est grand, plus on observe de pics au niveau du chromatogramme correspondant à plusieurs états d'oxydation d'une même protéine et plusieurs types d'appariements des ponts disulfures. De manière générale, les observations faites au cours de ce travail indiquent que le pic majoritaire correspond au peptide d'intérêt complètement oxydé, avec tous les ponts disulfures formés et correctement appariés.

Cependant, une caractérisation systématique est réalisée sur le produit final oxydé afin de vérifier son état d'oxydation ainsi que le bon appariement des ponts disulfures.

III.2 Production des plates-formes protéiques par voie recombinante

La plupart des candidats ligands conçus au cours de ce travail ont été produits à l'aide d'une stratégie de production de protéines par voie recombinante. Cette stratégie s'inscrit dans la perspective ultérieure d'introduire de la variabilité au niveau des séquences d'intérêt en vue de les optimiser et d'améliorer leur affinité et/ou leur sélectivité vis à vis des cibles d'intérêt. Dans le contexte de la production des protéines par voie recombinante, l'enjeu est de pouvoir produire chez la bactérie *Escherichia coli* des petites protéines (masse moléculaire < 6 kDa) à ponts disulfures (entre 2 et 4 ponts disulfures) sous forme oxydée et correctement repliées. Par ailleurs, les plates-formes identifiées *in silico* doivent également être produites avec leur extrémité N-terminale libre, portant les résidus constituant le motif fonctionnel d'intérêt. Toutes les plates-formes identifiées *in silico* ont été sélectionnées pour une production par voie recombinante avec le motif fonctionnel de référence, à l'exception de l'endothelin 1 qui possède naturellement le même motif de liaison que le ligand de référence (CSCS) et qui a été produite par synthèse chimique.
Les travaux relatifs à la production recombinante des mini-protéines d'intérêt ont été réalisés dans le cadre de la mise en place au sein du service d'une plate-forme dédiée à la production recombinante de protéines à ponts disulfures chez *E. coli*. Ces travaux ont donc nécessité plusieurs phases d'optimisation avant d'atteindre des résultats satisfaisants.

III.2.1 Construction des plasmides d'expression

Des travaux de criblage à grande échelle ont été réalisés afin de déterminer des conditions d'expression permettant la production recombinante des mini-protéines chez *E. coli*.

III.2.1.1 Essais et mise au point des différentes stratégies d'expression

Les premiers essais d'expression des protéines plates-formes identifiées *in silico* ont été

réalisés avec une stratégie d'expression permettant l'exportation dans le périplasme d'*E. coli*. En effet, ce compartiment constitue un environnement oxydant qui possède une machinerie enzymatique catalysant la formation des ponts disulfures. Cependant, parmi toutes les protéines dont l'expression périplasmique a été testée, seule 1an1I_m a pu être produite sous forme soluble avec des rendements satisfaisants. Par conséquent, la production de toutes les autres plates-formes (à l'exception de 1edpA) a nécessité l'emploi de stratégies d'expression recombinante alternatives chez *E. coli,* notamment cytoplasmiques. Dans cette optique, un criblage systématique à grande échelle de différents paramètres (souche d'expression, partenaire de fusion solubilisant, expression cytoplasmique ou periplasmique) a été réalisé afin d'identifier les conditions optimales d'expression des mini-protéines à ponts disulfures chez *E. coli*.

- Production dans le cytoplasme d'*E. coli*

De manière générale, le cytoplasme d'*E. coli* constitue un environnement réducteur défavorable à la formation de ponts disulfures dans les protéines. De ce fait, les protéines cytoplasmiques ne contiennent pas de ponts disulfures structuraux. Par ailleurs, les petites protéines ayant une taille inférieure à 10 kDa sont souvent difficiles à produire dans *E. coli*, car elles ont tendance à être rapidement dégradées ou à former des agrégats insolubles appelés corps d'inclusion. Ainsi, les protéines d'intérêt sont souvent produites en fusion avec un partenaire solubilisant qui peut éventuellement posséder un effet redox favorable à la formation des ponts disulfures. Parmi les partenaires classiquement utilisés, on trouve la thiorédoxine, DsbC, DsbA, GST et MBP.

Notre stratégie de production des mini-protéines à ponts disulfures dans le cytoplasme d'*E. coli* repose sur l'utilisation d'un plasmide d'expression portant une construction contenant :
- un partenaire de fusion solubilisant possédant le cas échéant un rôle rédox ;
- une étiquette 6(His) pour la purification de la protéine d'intérêt présente dans le lysat cellulaire
- un site de clivage spécifique d'une protéase permettant d'éliminer la protéine de fusion et l'étiquette d'affinité afin de libérer la séquence d'intérêt. Les enzymes de clivage les plus couramment utilisées sont la thrombine, le facteur Xa, l'entérokinase et la TEV protéase. Cette dernière a été utilisée dans toute cette étude.

La meilleure condition d'expression obtenue pour chacune des plates-formes est illustrée dans la figure III.1.

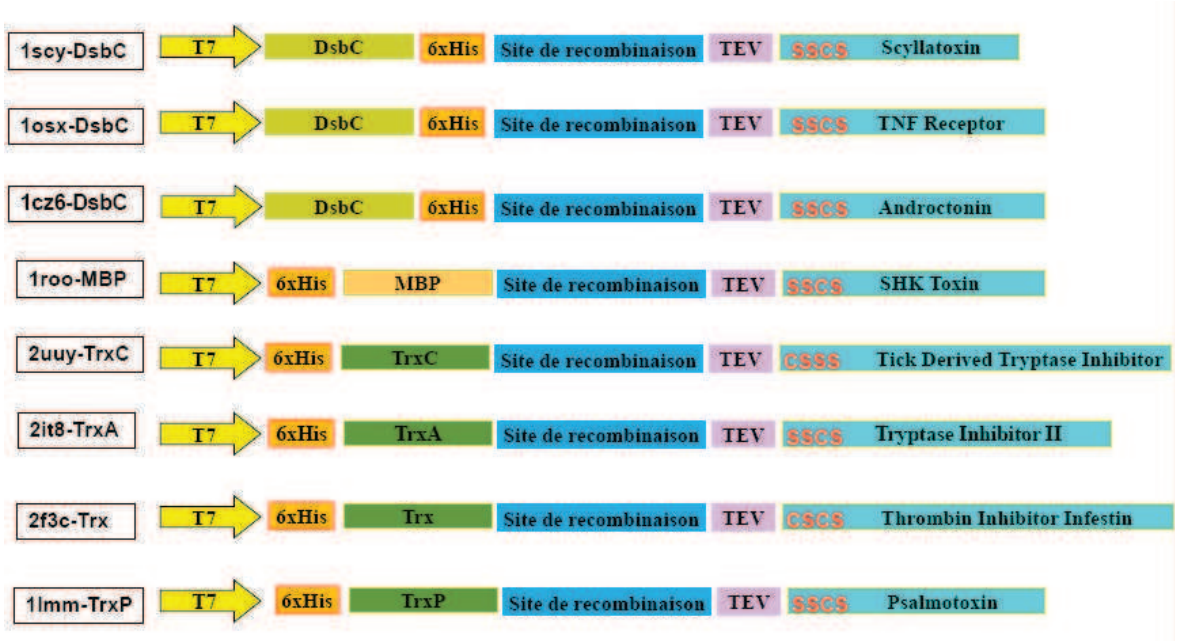


Figure III.1 : Construction des plasmides d'expression permettant la production des plates-formes dans le cytoplasme d'*E. coli*, et contenant une séquence codant pour une protéine de fusion (DsbC, MBP ou Trx), une séquence codant pour une étiquette 6(His), une séquence de clivage à la protéase TEV et la séquence de la protéine d'intérêt.

- Production dans le périplasme d'*E. coli*

Dans le cas de la plate-forme 1an1I_m, une stratégie de production dans le périplasme d'*E. coli* a été adoptée. Le plasmide d'expression porte une construction illustrée dans la figure III.2 et contenant :

- le peptide signal pelB permettant l'adressage de la protéine vers le périplasme d'*E. coli* : contrairement au cytoplasme, le périplasme d'*E. coli* est un environnement oxydant permettant la formation des ponts disulfures des protéines. Le peptide signal pelB permet donc le passage de la protéine d'intérêt à travers la membrane entre le cytoplasme et le périplasme, puis il subit un clivage spécifique par la Leader Peptidase présente au sein de cette membrane, permettant ainsi à la protéine de se replier correctement dans le périplasme.
- une étiquette 6(His)
- un site de coupure par la protéase TEV

Figure III.2 : Construction du plasmide d'expression permettant la production de la plate-forme 1an1I_m dans le périplasme d'*E. coli*, et contenant une séquence codant le peptide signal pelB, la séquence de la protéine d'intérêt, une séquence de clivage à la protéase TEV et une séquence codant pour une étiquette 6(His).

Cette stratégie, contrairement à la précédente, permet d'obtenir la mini-protéine d'intérêt avec son extrémité N-terminale libre, après clivage du peptide signal pelB. En revanche, après coupure à la TEV protéase, une séquence de 6 acides aminés résiduels reste attachée à l'extrémité C-terminale.

Toutes les plates-formes ont pu être produites à grande échelle dans les conditions d'expression décrites ci-dessus, en utilisant la stratégie de production dans le cytoplasme ou dans le périplasme. Toutefois, ces stratégies n'ont pas conduit à des résultats optimaux notamment en termes de rendements finaux et ne seront pas présentés dans ce rapport. En effet, les étapes de production chez *E. coli* ainsi que l'extraction du cytoplasme ou du périplasme ont pu être réalisées pour toutes les protéines d'intérêt. En revanche, les étapes de purification, de caractérisation et d'évaluation de l'affinité de ces protéines pour les cibles d'intérêt ont révélé un certain nombre de difficultés. Plusieurs autres phases d'optimisation des conditions d'expression ont finalement permis d'identifier une stratégie unique et optimale pour toutes les plates-formes.

III.2.1.2 Stratégie optimale adoptée pour l'expression des mini-protéines d'intérêt

Suite à la mise au point des protocoles et des conditions d'expression, la construction finale utilisée correspond à des gènes synthétiques contenant :

- une séquence non canonique du site de clivage par la TEV protéase : ENLYFQ ;
- la séquence correspondant à la mini-protéine d'intérêt possédant le motif fonctionnel de liaison à l'extrémité N-terminale ;
- les sites de recombinaison Gateway de part et d'autre du gène synthétique.

La TEV protéase est spécifique d'une séquence ENLYFQG qu'elle clive entre les acides aminés Q et G, laissant ainsi un résidu G au niveau de l'extrémité N-terminale de la protéine d'intérêt. Notre objectif étant de produire les mini-protéines d'intérêt avec leur extrémité N-terminale portant le motif de liaison libre afin de chélater le zinc catalytique, le résidu Gly a été

éliminé de la séquence correspondant au site de clivage par la TEV protéase (Figure III.3).

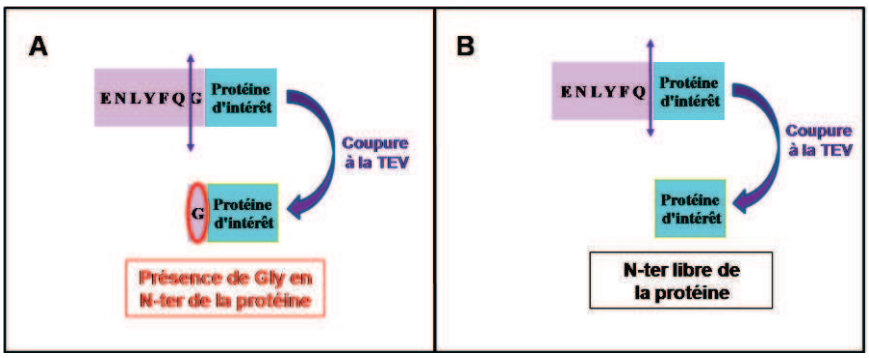


Figure III.3 : Sites de clivage à la TEV protéase. A : Site canonique ENLYFQG impliquant la présence d'un résidu G résiduel au niveau de l'extrémité N-terminale de la protéine après clivage à la TEV protéase. B : Site de clivage sans le résidu Gly permettant l'obtention d'une protéine dont l'extrémité N-terminale est libre.

Par ailleurs, les séquences des protéines d'intérêt correspondent aux séquences natives des protéines, à l'exception des 4 résidus formant le motif de liaison qui ont été mutés afin d'introduire les résidus spécifiques du motif fonctionnel de référence Cys-Ser-Cys-Ser (Annexe II et Méthodes). Pour ce faire, les résidus Cys natifs en positions 1 et/ou 3 du motif des plates-formes natives sont préservés. Lorsque le résidu en position 1 ou 3 du motif de liaison de la plate-forme native n'est pas une cystéine, ce résidu est muté en serine. Cette stratégie permet d'éviter les problèmes d'oxydation durant les différentes étapes de production, liés à la présence de résidus Cys non natifs libres. Tous les autres résidus du motif de liaison des plates-formes mimes sont mutés sérine. Ainsi, pour les plates-formes possédant un résidu Cys natif en position 3 du motif, le motif transféré est SSCS (1scyA, 1an1I, 1lmmA, 1rooA, 1cz6A, 1osxA et 2it8A) tandis que pour les plates-formes possédant un résidu Cys natif en position 1 du motif, le motif transféré est CSSS (2uuyB). La plate-forme 2f3cI possède deux résidus Cys natifs en positions 1 et 3 du motif, le motif transféré CSCS est donc strictement identique au motif de référence. Par ailleurs, les protéines natives 2f3cI et 1cz6A possèdent un résidu supplémentaire en position N-terminale du motif (Asp5 et Arg1 respectivement). Ces résidus sont supprimés de la séquence afin de garder l'extrémité N-terminale du motif de liaison (résidu en position 1 du motif) libre.

Les gènes ainsi construits sont clonés dans le vecteur de destination pETG-82A portant le gène codant pour la protéine DsbC et l'étiquette 6(His) par un clonage LR Gateway™ grâce aux sites de recombinaison portés par le gène. Les résultats obtenus lors du criblage des différentes

conditions d'expression montrent que la protéine de fusion DsbC est la plus adéquate pour la solubilisation et l'oxydation des protéines d'intérêt. En particulier, des phénomènes de précipitation ont été observés pour les protéines en fusion avec les thiorédoxines lors de la purification par affinité sur des colonnes de nickel. Les plasmides d'expression construits codent donc pour une protéine de fusion constituée de la protéine DsbC – Etiquette 6(His)- Site de clivage TEV protéase - Mini-protéine d'intérêt (Figure III.4).

Figure III.4 : Séquence génétique contenue dans le plasmide d'expression permettant la production de la protéine d'intérêt comme protéine de fusion.

III.2.2 Production et purification des protéines de fusion

La production des protéines recombinantes a été réalisée en suivant un protocole décrit récemment dans les données de la littérature (Nguyen et al., 2011). Ce protocole repose sur la co-expression de la protéine d'intérêt avec la sulfhydrile oxydase Erv1p et la protéine disulfide isomérase humaine PDI. La sulfhydrile oxydase Erv1p est une enzyme qui catalyse naturellement la formation *de novo* des ponts disulfures dans une protéine à partir de l'oxygène moléculaire. La présence simultanée dans le cytoplasme d'*E. coli* de la sulfhydrile oxydase Erv1p avec une protéine catalysant l'isomérisation des ponts disulfures telle que DsbC ou PDI permet l'obtention de protéines dont les ponts disulfures sont formés et correctement appariés. L'utilisation de ce protocole devrait résulter en une nette amélioration des rendements de production par rapport à un protocole standard sans co-expression de la protéine d'intérêt avec la sulfhydrile oxydase Erv1p et la protéine PDI.

Ainsi, le protocole de production consiste à transformer, par un protocole de choc thermique, une souche d'expression *E. coli* BL21-Star (DE3) avec le plasmide d'expression portant le gène de la mini-protéine d'intérêt et un gène de résistance à l'ampicilline ainsi que le plasmide pMJS9 portant les gènes codant pour la sulfhydrile oxydase Erv1p et la protéine PDI ainsi qu'un gène de résistance au chloramphénicol. La mise en culture des souches bactériennes se fait dans un milieu additionné d'ampicilline et de chloramphénicol afin de sélectionner uniquement les cellules transformées par les deux plasmides. L'expression de la sulfhydrile

oxydase et de la PDI est induite dans un 1er temps par ajout d'arabinose, puis l'expression de la protéine recombinante d'intérêt est induite par ajout d'IPTG

Toutes les protéines ont été produites dans 1 litre de milieu de culture à l'exception de 2uuyB_m, produite dans 2 litres de culture du fait des faibles rendements de production finaux. Toutes les protéines sont produites selon le même protocole : mise en culture des cellules bactériennes, récupération des cellules bactériennes par centrifugation, lyse des culots bactériens et éclatement des cellules par un système de broyage cellulaire puis purification du surnageant sur colonne de nickel. A cette étape, les quantités de protéines de fusion pures obtenues après purification sur la colonne de nickel et déterminées par dosage UV à 280 nm sont très variables : de 60 mg/L de culture (2uuyB_m) à 336 mg/L de culture (1osxA_m).

III.2.3 Clivage des protéines de fusion et purification des mini-protéines d'intérêt

La mini-protéine d'intérêt est libérée de la fusion présente au niveau de son extrémité N-terminale par clivage à la TEV protéase. Pour ce faire, la solution contenant la protéine de fusion est soumise à une dialyse afin d'éliminer l'imidazole contenu dans le tampon d'élution de la protéine purifiée sur colonne de nickel. Plusieurs tampons de dialyse de compositions différentes ont été testés afin de trouver un tampon compatible avec l'activité de la TEV protéase et avec les caractéristiques physico-chimiques des protéines d'intérêt. Pour toutes les protéines produites, le tampon de dialyse utilisé est composé de Tris-HCl (50 mM) à pH 8, à l'exception de 1osxA_m. En effet, dans le cas de cette protéine, à la suite du clivage à la TEV protéase, une grande partie de la protéine devient insoluble et forme un précipité dense en solution. Ce phénomène peut s'expliquer par le point isoélectrique de cette protéine (pI = 7,66) très proche du pH du tampon de clivage (pH = 8). Pour améliorer la solubilité de la protéine d'intérêt, un autre tampon de dialyse a donc été utilisé : 50 mM MES, pH 6,5. Par ailleurs, le pH du tampon de clivage par la TEV protéase n'étant pas optimal dans le cas de 1osxA_m, le clivage est réalisé à 25°C au lieu de 4°C afin d'augmenter l'efficacité de clivage. L'efficacité de cette étape de clivage a été déterminée par une analyse du produit de clivage par LC/MS mettant en évidence la disparition plus ou moins totale du pic correspondant à la protéine de fusion au profit des pics correspondants à la protéine DsbC seule et à la protéine d'intérêt respectivement.

Suite au clivage par la TEV protéase, le mélange de clivage contient la protéine d'intérêt libre, la protéine DsbC en fusion avec l'étiquette 6(His) ainsi que la TEV protéase portant également

une étiquette 6(His). Un 1er protocole a été utilisé pour la purification du mélange de clivage sur des billes de nickel : la protéine DsbC portant l'étiquette 6(His) ainsi que la TEV protéase restent fixées sur les billes de nickel tandis que la protéine d'intérêt est récupérée dans l'éluat après lavage des billes. Ce protocole était destiné à éliminer les protéines de grandes tailles telles que DsbC et la TEV protéase afin de limiter les problèmes de surcharge des colonnes de purification par RP-HPLC. Cependant, les rendements obtenus avec ce protocole étaient médiocres du fait de la perte en protéine d'intérêt par adhésion non spécifique sur les billes et du fait de la présence d'une quantité résiduelle de la protéine DsbC qui ne s'est pas fixée sur les billes. Par conséquent, la purification sur les billes de nickel était systématiquement suivie d'une purification par RP-HPLC. Afin d'améliorer les rendements de purification, l'étape de purification sur les billes de nickel a été éliminée et le mélange de clivage a directement été soumis à une purification par RP-HPLC. Les quantités de protéines de fusion contenues dans le mélange de clivage à purifier variant entre 60 et 220 mg, plusieurs cycles de purification par RP-HPLC ont été nécessaires pour purifier la protéine d'intérêt. Par ailleurs, une colonne C4 semi-préparative a été utilisée afin de permettre la séparation et la purification des protéines de grandes tailles telles que DsbC.
Après les étapes de purification, les mini-protéines d'intérêt pures ont été lyophilisées puis solubilisées dans le tampon approprié pour les études ultérieures.

III.2.4 Caractérisation des mini-protéines produites par voie recombinante

La protéine d'intérêt finale est caractérisée à l'aide de plusieurs méthodes analytiques. Une analyse par RP-HPLC couplée à la spectrométrie de masse (LC/MS) a permis de vérifier la pureté du produit final par la présence d'un pic unique et fin au niveau du chromatogramme. Par ailleurs, l'analyse de masse a indiqué que le produit final a la masse correcte attendue correspondant à la protéine d'intérêt sous forme oxydée, avec tous les ponts disulfures formés et possédant les résidus spécifiques du motif de liaison de référence au niveau de son extrémité N-terminale libre. La concentration de la solution finale du produit pur et oxydé dans le tampon adéquat pour la réalisation des tests d'inhibition est déterminée par une analyse de la composition en acides aminés ainsi qu'un dosage UV à 280 nm.

L'appariement correct des ponts disulfures a été vérifié grâce à des tests de liaison des protéines recombinantes avec leurs cibles naturelles. La plate-forme Thrombin Inhibitor Infestin

recombinante (2f3cI_m) a été testée pour sa capacité à inhiber une de ses cibles naturelles : la trypsine. La valeur de Ki obtenue (Ki = 0,53±0,15 nM) est proche de celle décrite dans la littérature (Ki = 2 nM) (Campos et al., 2002). Cette observation suggère fortement que la plate-forme recombinante est correctement oxydée et repliée de sorte que son activité inhibitrice envers sa cible naturelle est maintenue. De manière similaire, la plate-forme Leech Derived Tryptase Inhibitor recombinante (1an1I_m) a été testée pour sa capacité à inhiber une de ses cibles naturelles : la trypsine. Cette protéine a montré un Ki expérimental (Ki = 1,77±0,24 nM) identique à celui rapporté dans les données de la littérature (Ki = 1,8 nM) (Stubbs et al., 1997). La plate-forme Tick Derived Tryptase Inhibitor recombinante (2uuyB_m) a montré une affinité nM pour la trypsine (Ki = 3,1±0,37 nM) du même ordre de grandeur que celle rapportée dans les données de la littérature (Ki < 10 nM) (Valdez-Cruz et al., 2011). Un test d'activité a également été réalisé afin d'évaluer l'activité inhibitrice de la plate-forme Trypsin Inhibitor II recombinante (2it8A_m) envers la trypsine. Les résultats ont permis de déterminer une valeur de Ki = 73±27 nM. Cette valeur est 20 fois supérieure à la valeur trouvée dans la littérature pour l'inhibition de la trypsine par la protéine Trypsin Inhibitor II native 2it8A (Ki = 0,3 nM) qui diffère de la forme recombinante produite par les 3 résidus modifiés dans le motif de liaison (GVCP et SSCS respectivement) (Sommerhoff et al., 2010). Cependant, la constante d'affinité mesurée de l'ordre du nM est très basse et suggère fortement que la protéine recombinante a gardé une activité inhibitrice correcte envers la trypsine, suggérant un appariement correct des ponts disulfures.

III.2.5 Rendement de production des mini-protéines d'intérêt

Toutes les plates-formes portant le motif spécifique du ligand de référence ont pu être produites par voie recombinante, à l'exception de l'analogue de l'endothéline 1 qui a été produit par synthèse chimique. Parmi les 9 plates-formes produites par voie recombinante, 3 plates-formes ont été produites selon un protocole standard sans co-expression avec la sulfhydrile oxydase (2f3cI_m, 1scyA_m et 2it8A_m), 4 ont été produites en co-expression avec la sulfhydrile oxydase (1osxA_m, 1cz6A_m, 1lmmA_m et 1an1I_m) et 2 ont été produites selon les deux protocoles (2uuyB_m et 1rooA_m). Le tableau III.2 récapitule les rendements de production des plates-formes obtenues par voie recombinante, selon les protocoles avec ou sans co-expression avec la sulfhydrile oxydase. La protéine de fusion correspond à la protéine d'intérêt fusionnée à la protéine DsbC portant l'étiquette 6(His). Ainsi, la quantité initiale

théorique de protéine d'intérêt contenue dans la protéine de fusion est calculée à partir des % massiques des deux protéines de la fusion. Le rendement de production correspond donc au rapport entre la quantité initiale de protéine d'intérêt contenue dans la fusion et la quantité finale de protéine d'intérêt pure et oxydée.

Les quantités de protéines de fusion obtenues selon les deux protocoles sont très variables en fonction des protéines : elles varient de 73 à 336 mg/L de culture. Cette différence importante peut être en relation avec la toxicité plus ou moins élevée de la protéine produite pour les cellules bactériennes. En effet, on observe une relation entre la DO finale de la culture à 600 nm avant la lyse des cellules et la quantité finale de protéine de fusion produite (Données non fournies dans ce rapport). La plate-forme 1rooA_m ayant la DO la plus basse en fin de culture (DO_{600nm} = 3,9) est associée au plus faible rendement en protéine de fusion (73 mg/L) et à l'inverse, la plate-forme 1osxA_m ayant la DO la plus élevée en fin de culture (DO_{600nm} = 6,1) est associée au rendement le plus important en protéine de fusion (336 mg/L).

La comparaison des deux protocoles montre que les rendements de production observés pour deux séries de mini-protéines sont en moyenne plus élevés en présence de sulfhydrile oxydase. En effet, en l'absence de sulfhydrile oxydase, les rendements observés pour la première série de mini-protéines varient entre 1,6 et 8,8% (Tableau III.2). En présence de sulfhydrile oxydase, les rendements obtenus pour une seconde série de mini-protéines sont compris entre 1,9 à 34,7%. Lorsque l'on compare les rendements obtenus en présence et en absence de sulfhydrile oxydase pour une même protéine (1rooA_m), la différence est de l'ordre d'un facteur 10. Cette différence s'explique très vraisemblablement par la capacité de l'enzyme sulfhydrile oxydase à catalyser la formation des ponts disulfures dans la protéine d'intérêt. Associée aux isomérases PDI ou DsbC, la sulfhydrile oxydase favorise donc l'obtention de protéines solubles, oxydées et adoptant le bon repliement.

Par ailleurs, les quantités finales en protéines d'intérêt pures et correctement oxydées et repliées sont de l'ordre du mg (comprises entre 0,178 et 13 mg). Pour certaines protéines (1lmmA_m et 1an1I_m), les quantités finales obtenues en protéines d'intérêt sont très importantes, (respectivement 13 et 12 mg). A l'inverse, deux plates-formes sont produites en quantités inférieures au milligramme : 2uuyB_m et 1osxA_m avec 0,319 mg et 0,565 mg respectivement. Ces quantités sont néanmoins suffisantes pour l'évaluation de la capacité des protéines

correspondantes à inhiber l'activité des MMPs. La faible quantité de protéine finale obtenue pour la plate-forme 1osxA_m (0,565 mg), en dépit de la quantité importante de protéine de fusion (336 mg), est probablement due aux caractéristiques physico-chimiques de la protéine d'intérêt entrainant une perte de rendement considérable lors des étapes de dialyse, de clivage et de purification par RP-HPLC. En effet, malgré une taille très petite (26 résidus), cette protéine est très hydrophobe (GRAVY = +0,369 ; Indice Aliphatique = 90) et donc très peu soluble dans les différents solvants et tampons utilisés, impliquant une perte de produit au niveau des étapes finales de production. Concernant la plate-forme 2uuyB_m, la très faible quantité de protéine finale obtenue est probablement due au nombre élevé de ponts disulfures contenus dans la protéine d'intérêt (4 ponts disulfures) impliquant des pertes importantes correspondant à une accumulation de formes non oxydées ou partiellement oxydées, et celles dont les ponts disulfures ne sont pas correctement appariés.

Tableau III.2: Rendements de production des plates-formes obtenues par voie recombinante, avec ou sans co-expression de la sulfhydrile oxydase.

	Production **sans co-expression** avec la sulfhydrile oxydase			
Code PDB	Quantité de protéine de fusion (mg/L de culture)	Quantité théorique de protéine d'intérêt (mg)	Quantité finale de protéine d'intérêt (mg)	Rendement total de production (%)
2f3cI_m	120	20,4	0,700	3,5
2uuyB_m	87	14,8	0,250	1,6
1scyA_m	117	12,9	1,15	8,8
1rooA_m	73	9,3	0,178	1,9
2it8A_m	291	29,1	0,890	3,0
	Production **avec co-expression** avec la sulfhydrile oxydase			
2uuyB_m	75	12,7	0,319	2,5
1osxA_m	336	30,2	0,565	1,9
1cz6A_m	145	13	1,75	13,5
1lmmA_m	277	38,8	13	33,5
1rooA_m	104	13,2	2,3	17,4
1an1I_m	240	34,5	12	34,7

En conclusion, toutes les plates-formes identifiées *in silico* (à l'exception de l'endothelin 1 produite par synthèse chimique) ont pu être produites par voie recombinante en utilisant une stratégie reposant sur : une production en protéine de fusion avec DsbC comme partenaire solubilisant et oxydant, une co-expression de la protéine de fusion avec la sulfhydrile oxydase pour l'amélioration des rendements de production et d'oxydation, une purification par affinité grâce à une étiquette 6(His) et un clivage à la protéase TEV pour libérer la protéine d'intérêt. Ce protocole optimisé a permis de produire des protéines très différentes du point de vue de la taille, du nombre de ponts disulfures et des propriétés physico-chimiques. Les quantités finales de peptide pur oxydé sont de l'ordre du mg et sont suffisantes pour la réalisation des tests de caractérisation et d'inhibition sur les MMPs. Les analyses par chromatographie liquide couplée à la spectrométrie de masse montrent que le peptide d'intérêt est correctement oxydé, que son extrémité N-terminale est libre, et qu'il possède la séquence du motif d'intérêt.

III.3 Evaluation des mini-protéines comme inhibiteurs de MMPs

Toutes les protéines natives produites par voie chimique, les protéines commerciales ainsi que les protéines produites par voie recombinante et portant le motif fonctionnel d'intérêt ont été testées pour leur capacité à inhiber les MMPs. L'inhibition des MMPs par les mini-protéines produites a été évaluée par des tests enzymatiques dont le principe repose sur l'effet de la présence de la mini-protéine étudiée sur le clivage d'un substrat à fluorescence quenchée (le substrat subit une exaltation de sa fluorescence lorsqu'il est clivé) par les MMPs.

Les candidats ligands ont été conçus via le greffage d'un motif fonctionnel identifié sur un inhibiteur à large spectre des MMPs, le TIMP-2. Par ailleurs, aucun critère de sélection spécifique d'un membre particulier de la famille des MMPs n'a été utilisé pour la sélection des protéines plates-formes capables de reproduire la topologie fonctionnelle du motif greffé. De ce fait, aucune sélectivité particulière n'est attendue pour ces différents candidats ligands. Leur activité inhibitrice a donc été déterminée sur plusieurs MMPs. Au total, 6 membres de la famille des MMPs ont été testés : MMP-2, -8, -9, -12, -13 et -14. Les résultats obtenus sont présentés dans le tableau III.3.

Tableau III.3 : Constantes d'inhibition (µM) de l'activité enzymatique des MMPs par les plates-formes natives portant le motif de liaison natif (vert) et les plates-formes portant les résidus spécifiques du motif de liaison du ligand de référence (rouge). (ND : Non Déterminé).

Code PDB	Motif	MMP-2	MMP-8	MMP-9	MMP-12	MMP-13	MMP-14
1edpA	CSCS	7	2	7	1	0,45	ND
1edpA (1-18)	CSCS	ND	3	10	7	3	ND
Endothelin 3	CTCF	10	5	4	ND	0,80	ND
1scyA	AFCN	10	12	7	79	ND	ND
1scyA_m	SSCS	150	75	130	160	440	ND
1rooA	RSCI	ND	70	ND	ND	ND	ND
1rooA _m	SSCS	182	54	102	457	381	591
1lmmA	EDCI	ND	65	ND	ND	ND	ND
1lmmA _m	SSCS	18	11	32	34	68	24
2f3cI	CACP	56	56	51	77	143	217
2f3ci _m	CSCS	32	6	36	31	73	100
2uuyB	CTVP	10	4	20	8	9	129
2uuyB _m	CSSS	69	24	39	160	158	ND
2it8A	GVCP	155	84	142	> 300	> 400	ND
2it8A_m	SSCS	173	70	157	> 600	> 600	> 600
1an1I_m	SSCS	22	5	30	136	38	132
1cz6A_m	SSCS	92	40	45	582	183	213
1osxA_m	SSCS	97	54	58	197	206	582

Les résultats obtenus montrent que tous les candidats ligands conçus *in silico* et produits expérimentalement sont capables d'inhiber les MMPs testées avec des constantes d'inhibition mesurables comprises entre 450 nM et 590 µM, à l'exception de la plate-forme 2it8A. En effet, les MMP-12, -13 et -14 ne montrent pas de variation d'activité catalytique en présence de 600 µM de 2it8A_m.

L'inhibiteur le plus affin envers les MMPs testées est l'endothelin 1 qui possède naturellement le motif fonctionnel du TIMP-2 (CSCS), avec des valeurs de Ki comprises entre 1 et 7 µM et un Ki sub-micromolaire pour la MMP-13 (Ki = 450 nM). A titre d'exemple, la courbe d'inhibition correspondant au test réalisé avec la plate-forme native 1edpA sur la MMP-13 est représentée dans la figure III.5.

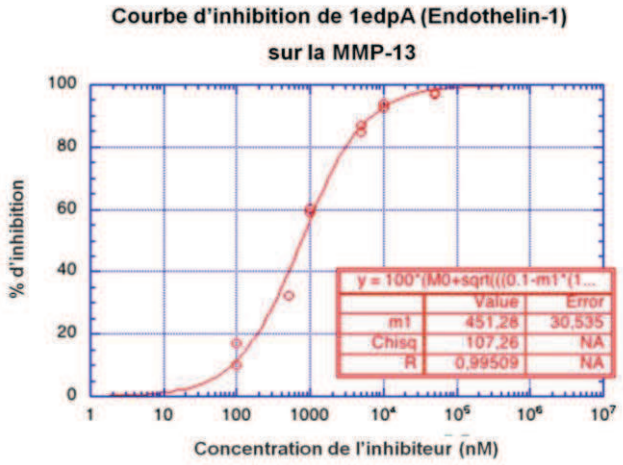


Figure III.5: Courbe d'inhibition de la plate-forme native 1edpA sur la MMP-13 représentant le pourcentage d'inhibition obtenu sur l'enzyme testée en fonction de la concentration de la plate-forme. La valeur de la constante d'inhibition (m1) est de 451 nM.

La forme native de la plate-forme 2uuyB a également montré une bonne affinité envers les MMPs testées : les constantes d'inhibition sont proches du micromolaire pour les MMP-2, -8, -9, -12 et -13 (entre 4 et 20 µM). Sur la MMP-8 en particulier, plusieurs plates-formes ont montré des affinités élevées : 2uuyB, 1an1I_m et 2f3cI_m avec des Ki compris entre 4 et 6 µM.

Par ailleurs, la comparaison entre les constantes d'affinité obtenues pour les plates-formes portant un motif natif et celles obtenues pour les plates-formes mutées portant le motif à greffer permet d'évaluer l'importance de la nature des résidus spécifiques du motif de référence. De manière générale, on note que les plates-formes portant le motif natif (1scyA, 1rooA, 1lmmA, 2f3cI, 2uuyB et 2it8A) possèdent également des affinités mesurables pour les différentes MMPs testées, comprises entre 7 et 217 µM. Ce résultat était attendu car les résidus du motif fonctionnel établissent des interactions intermoléculaires avec différents résidus de la cible via leur chaine principale et non par le biais de leurs chaines latérales. Dans le cas de la plate-forme 2it8A, ni la forme mutée portant le motif de référence (SSCS) ni la forme native (GVCP) ne sont de bons inhibiteurs de MMPs ; les constantes d'inhibition obtenues dans les deux cas pour les différentes MMPs testées sont similaires. En revanche, dans le cas de 1scyA, 2f3cI et 2uuyB, les constantes d'inhibition varient entre les formes portant le motif natif et celles portant le motif de référence. Avec la plate-forme 2f3cI_m portant le motif CSCS, on note un léger accroissement de l'affinité pour les MMPs testées d'un facteur 1,5 à 9,5 par rapport à la plate-forme native portant le motif CACP. Cependant, pour 2uuyB, la plate-forme portant un motif natif (CTVP) possède une affinité plus élevée pour les MMPs testées que la version de

cette plate-forme 2uuyB_m portant le motif CSSS ; la perte d'affinité est d'un facteur 20 sur la MMP-12 à 2 sur la MMP-9. De manière similaire, en greffant le motif de référence SSCS sur la plate-forme 1scyA native portant le motif AFCN, une perte d'affinité est observée sur les MMPs testées avec en particulier une différence d'un facteur 18,5 sur les constantes d'inhibition déterminées sur la MMP-9 et un facteur 2 sur celles déterminées sur la MMP-12. Ces résultats indiquent donc que l'importance de la nature des résidus du motif dépend de la plate-forme considérée. Les effets constatés peuvent avoir de multiples origines enthalpiques ou entropiques dont l'identification demanderait une étude thermodynamique et structurale détaillée.

Les 10 candidats ligands conçus *in silico* grâce au programme STAMPS et produits expérimentalement ont montré des affinités mesurables pour les 6 MMPs testées. L'inhibition de l'activité catalytique de ces enzymes montre également que les mini-protéines ciblent le site catalytique de ces MMPs. Les constantes d'association observées sont comprises entre 450 nM et quelques centaines de micromolaires. Ces données sont résumées sur la figure III.6.

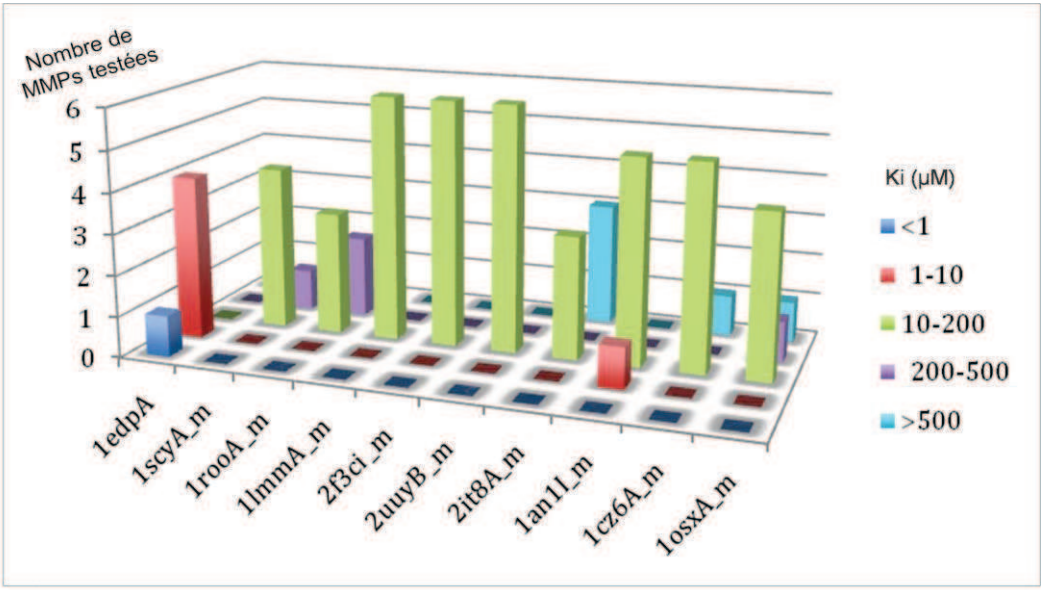


Figure III.6: Constantes d'inhibition mesurées pour les 10 ligands protéiques conçus *in silico* pour une série de MMPs (MMP-2, -8, -9, -12, -13 et -14). Les données sont classées par gamme de constante d'inhibition exprimée en µM. Le nombre de valeurs figure en ordonnées.

Cette figure montre que, à l'exception de 1edpA, la majorité des constantes d'association obtenues pour les 9 ligands protéiques se situe entre 10 et 200 µM. Cette observation montre que l'approche utilisée permet d'atteindre l'objectif assigné : concevoir une série de ligands

protéiques par greffage d'un motif fonctionnel sur une série de plates-formes protéiques capables de cibler un site précis à la surface d'une enzyme. Toutefois, nous observons que les constantes d'association obtenues sont le plus souvent modestes. Cette observation soulève des questions importantes concernant les limites de la méthode et la nécessité de prendre en compte d'autres facteurs permettant de caractériser la reproduction des interactions observées dans le complexe naturel. Cette question sera abordée dans la dernière partie des résultats. Par ailleurs, une phase d'optimisation expérimentale des candidats ligands obtenus est nécessaire afin d'améliorer l'affinité et/ou la sélectivité de ces ligands protéiques vis-à-vis des membres de la famille des MMPs.

III.4 Optimisation de certaines mini-protéines identifiées *in silico*

Des modifications ponctuelles des ligands protéiques conçus peuvent être envisagées pour accroître leur affinité initiale. Ceci a été envisagé pour les composés 1edpA (1-18) et 1an1I_m en exploitant le mode de liaison décrit dans les modèles moléculaires réalisés. La plate-forme 1edpA (1-18) a été sélectionnée pour exploiter la structure particulière du sous site S_1' des MMPs afin d'accroître l'affinité du composé initial. Des analogues de la position 2 de 1edpA (1-18) (position dite P_1' dans les substrats interagissant dans le sous site S_1' de l'enzyme) ont été réalisés. La plate-forme 1an1I_m a été retenue pour étudier l'effet de mutations à sa surface en utilisant la structure du complexe issue de la phase de conception *in silico* et une méthode d'évaluation empirique de la variation d'énergie libre de liaison (FoldX).

III.4.1 Variants de la plate-forme 1edpA pour la position 2 du motif de liaison

Sur la base d'un modèle moléculaire réalisé avec la MMP-3 en interaction avec la plate-forme de l'endothelin 1, des substitutions ponctuelles d'acides aminés ont été proposées. En particulier, pour la conception de variants de l'endothelin 1, nous avons choisi d'exploiter une caractéristique structurale des MMPs, qui est la présence d'une cavité profonde et hydrophobe désignée par le sous-site S_1' (Chapitre IV). Le résidu en position 2 du motif de liaison du ligand naturel TIMP-2 ainsi que de 1edpA se positionne dans la cavité S_1' du site actif des MMPs. Des études antérieures ont montré une préférence de ce sous site pour des résidus hydrophobes. Sur

la base de ces observations, nous avons incorporé dans la séquence de la plate-forme 1edpA (1-18) en position 2 du motif, des résidus non-naturels possédant de longues chaînes hydrophobes susceptibles de s'insérer dans le sous site S_1' des différentes MMPs (Figure III.7).

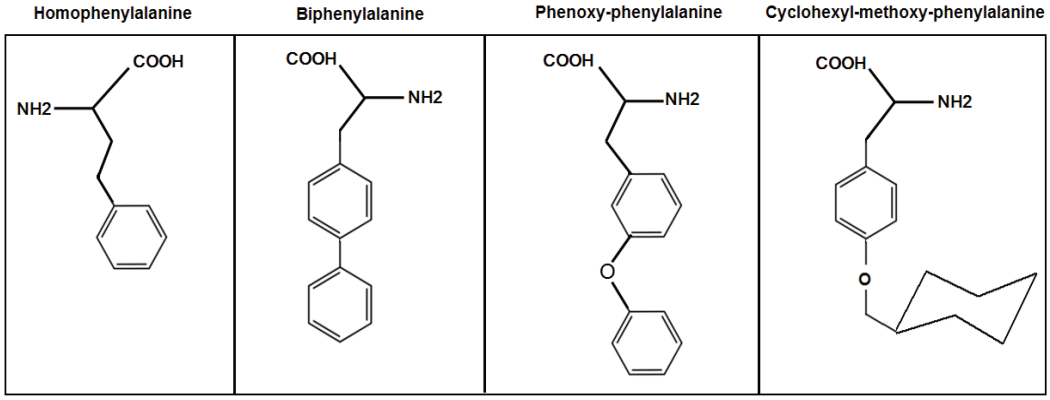


Figure III.7: Différents résidus non naturels introduits en position 2 du motif de liaison de l'endothelin 1. De gauche à droite : Homophenylalanine, Biphenylalanine, Phenoxy-phenylalanine et Cyclohexyl-methoxy-phenylalanine.

Quatre variants pour la position 2 du motif de 1edpA (1-18) ont été produits : 1edpA (1-18) S2hΦ avec un résidu homophenylalanine en position 2, 1edpA (1-18) S2ΦΦ avec un résidu biphenylalanine, 1edpA (1-18) S2ΦoΦ où la sérine est mutée en un résidu phenoxy-phenylalanine et 1edpA (1-18) S2ΦoM où la sérine est mutée en un résidu cyclohexyl-methoxy-phenylalanine. Ces 4 variants ont été produits par synthèse chimique selon le protocole de production mis au point et optimisé pour l'endothelin 1 (1-18). Ces analogues ont été purifiés, caractérisés et testés sur les 6 MMPs considérées. Les résultats de ces tests d'inhibition sont rapportés dans le tableau III.4.

Tableau III.4 : Constantes d'inhibition (µM) de l'activité enzymatique des MMPs par les variants de l'endothéline native 1edpA. Les acides aminés non naturels sont notés en rouge. S : Sérine ; hΦ : Homophenylalanine ; M: Méthionine ; nL: norLeucine ; ΦoΦ : Phenoxy-phenylalanine ; ΦoM : Cyclohexyl-methoxy-phenylalanine ; ΦΦ : Biphenylalanine.

Protéine	Motif	MMP-2	MMP-8	MMP-9	MMP-12	MMP-13	MMP-14
1edpA (1-18)	CSCS	nd	3	10	7	3	nd
1edpA (1-18) S2hΦ	ChΦCS	**0,62**	**0,76**	**0,40**	**0,29**	**0,34**	nd
1edpA (1-18) M7nL	CSCS	20	4	8	6	4	108
1edpA (1-18) S2ΦoΦ, M7nL	CΦoΦCS	23	8	4	5	9	55
1edpA (1-18) S2ΦoM, M7nL	CΦoMCS	15	11	13	1	10	107
1edpA (1-18) S2ΦΦ, M7nL	CΦΦCS	1	1	**0,67**	**0,63**	**0,41**	nd

Ces résultats montrent que la substitution de la sérine par le résidu non naturel homophenylalanine entraine une augmentation de l'affinité du composé d'un facteur de l'ordre de 20 pour les 5 MMPs testées (composés 1edpA (1-18) S2hΦ / 1edpA (1-18)). Cette substitution donne donc un variant plus affin que le ligand initial pour les MMPs testées avec des constantes d'inhibition comprises entre 290 et 760 nM. Une observation similaire est réalisée avec la substitution de la sérine en position 2 dans le composé1edpA (1-18) M7nL par le résidu biphenylalanine (composé 1edpA (1-18) S2ΦΦ, M7nL) qui entraine également une augmentation de l'affinité du composé jusqu'à un facteur 20 par rapport au composé initial. En revanche les substitutions de la sérine en position 2 par les acides aminés phenoxy-phenylalanine et cyclohexyl-methoxy-phenylalanine ne conduisent pas à des augmentations d'affinité importantes. Les différences d'effet notées pour les 4 variants peuvent s'expliquer par la géométrie particulière des 4 acides aminés non naturels incorporés en position 2. En effet, un modèle de complexe entre la MMP-3 et le variant 1edpA (1-18) S2hΦ, illustré dans la figure III.8, a montré que la présence du groupement méthylène supplémentaire entre le Cα et le groupement aromatique de la chaine latérale du résidu homophenylalanine permet à ce résidu un ajustement de sa conformation pour pénétrer dans la cavité S_1'.

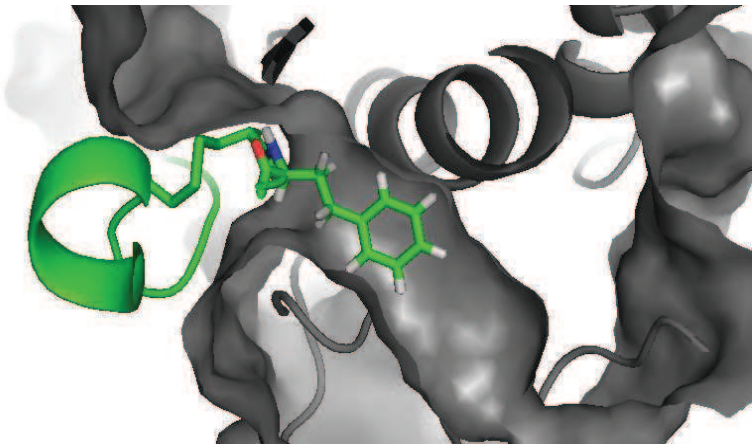


Figure III.8: Positionnement du résidu en position 2 du motif de liaison du ligand dans la cavité S_1' de la MMP. Modèle de complexe entre la MMP-3 représentée en mode surface et cartoon en gris et le variant 1edpA (1-18) S2hΦ représenté en mode cartoon en vert. Le résidu en position 2 homophenylalanine est représenté en mode bâtons dans la cavité S_1'.

L'absence de ce groupement méthylène entre le Cα et le groupement aromatique dans les résidus phenoxy-phenylalanine et cyclohexyl-methoxy-phenylalanine pourrait entrainer une gêne stérique à l'origine d'une pénalité énergétique compensant l'effet favorable attendu lié à la présence des groupements hydrophobes dans la cavité S_1'.

En conclusion, l'insertion d'un acide aminé non naturel à la place de la sérine en position 2 du motif a permis d'accroître, dans le cas des variants 1edpA (1-18) S2hΦ et 1edpA (1-18) S2ΦΦ, l'affinité de la plate-forme de 1ère génération 1edpA. Les effets de ces substitutions confirment le mode de liaison prévu par l'approche de conception de ligand protéique par greffage de motif fonctionnel. Ces composés constituent des molécules d'intérêt possédant des constantes d'inhibition de l'ordre de quelques centaines de nanomolaire qui seraient utilisables pour un travail systématique d'introduction de variabilité à l'interface avec les cibles afin de rechercher des composés sélectifs d'un membre particulier de la famille des MMPs.

III.4.2 Variants de la plate-forme 1an1I_m par mutagénèse dirigée

La plate-forme 1an1I_m recombinante a été choisie comme candidat pour la phase d'optimisation de séquence du fait du rendement élevé de production par voie recombinante de ce ligand (quantité de peptide final pur et oxydé de l'ordre de 12 mg). Ainsi, des substitutions d'acides aminés réalisées sur le ligand protéique artificiel 1an1I_m ont été recherchées afin

d'obtenir des variants plus affins de ce ligand protéique initial pour les MMPs testées. Cette étude s'est fondée sur l'exploitation d'un modèle moléculaire d'interaction de ce ligand protéique avec sa cible.

Dans le but d'identifier les positions dans la séquence de 1an1I_m, situées à l'interface avec la cible et sur lesquelles des substitutions de résidus peuvent être introduites, un modèle moléculaire de complexe MMP-8/1an1I_m a été réalisé. Le choix de la MMP-8 pour la réalisation du modèle est justifié par l'affinité observée de la plate-forme 1an1I_m envers la MMP-8 (Ki = 4 µM). Le modèle moléculaire du complexe MMP-8/1an1I_m est réalisé en utilisant la structure MMP-14/1an1I_m issue de la conception *in silico* de ce ligand protéique. Le domaine catalytique de la MMP-8 est superposé au domaine catalytique de la MMP-14 dans le modèle MMP-14/1an1I_m. Les coordonnées résultantes de la MMP-8 et de 1an1I_m sont utilisées comme coordonnées initiales. Le protocole de relaxation par minimisation d'énergie et dynamique moléculaire, utilisé lors de la conception des ligands protéiques, est ensuite appliqué à la structure initiale MMP-8/1an1I_m. L'interface de la structure du complexe issu du dernier cycle de relaxation est analysée. Cette analyse a permis d'identifier 3 positions clé au niveau de la structure du ligand 1an1I_m pour l'introduction de mutations ponctuelles. Ces positions identifiées correspondent aux résidus Lys8, Leu10 et Ala22. Par ailleurs, de manière similaire au travail d'optimisation réalisé sur la plate-forme 1edpA(1-18), la position 2 du motif de liaison de la plate-forme 1an1I_m recombinante, correspondant au résidu Ser3, a également été retenue pour l'introduction de la variabilité. Les quatre positions étudiées sont représentées sur la figure III.9.

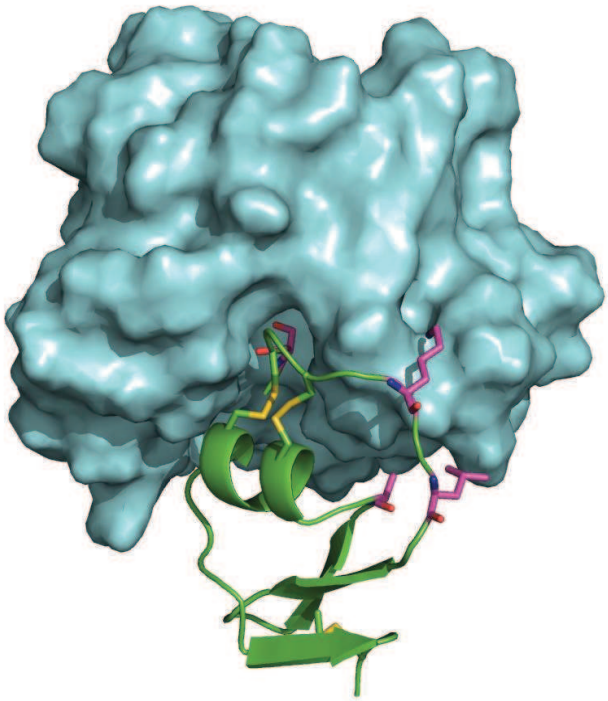


Figure III.9: Structure 3D du complexe MMP-8/1an1I_m. La MMP-8 est représentée en mode surface en cyan et la plate-forme 1an1I_m est représentée en mode cartoon vert. Les résidus de 1an1I_m en bâtons roses situés à l'interface avec la cible représentent les positions privilégiées pour introduire des modifications ponctuelles.

Le logiciel FoldX (Guerois et al., 2002) a été utilisé pour déterminer les substitutions les plus favorables du point de vue énergétique au niveau de ces 4 positions. Ce logiciel calcule la variation d'énergie libre ΔΔG liée à la mutation d'un résidu d'une séquence protéique. La variation d'énergie libre ΔΔG est donc calculée pour chacune des 20 substitutions possibles au niveau des 4 positions identifiées *in silico*. Les résultats reportés sur la figure III.10 montrent que les résidus les plus favorables énergétiquement sont Ile ou Met en position 3, Trp, Tyr et Met en positions 8, 10 et 22 respectivement. Afin de vérifier ce résultat obtenu sur la 15ème structure issue de la relaxation du complexe MMP-8/1an1I_m, les mêmes calculs ont été réalisés sur 2 autres structures issues du protocole de relaxation. Les substitutions les plus favorables pour les positions 3, 8, 10 et 22 dans ces deux structures sont identiques à celles obtenues avec la structure 15.

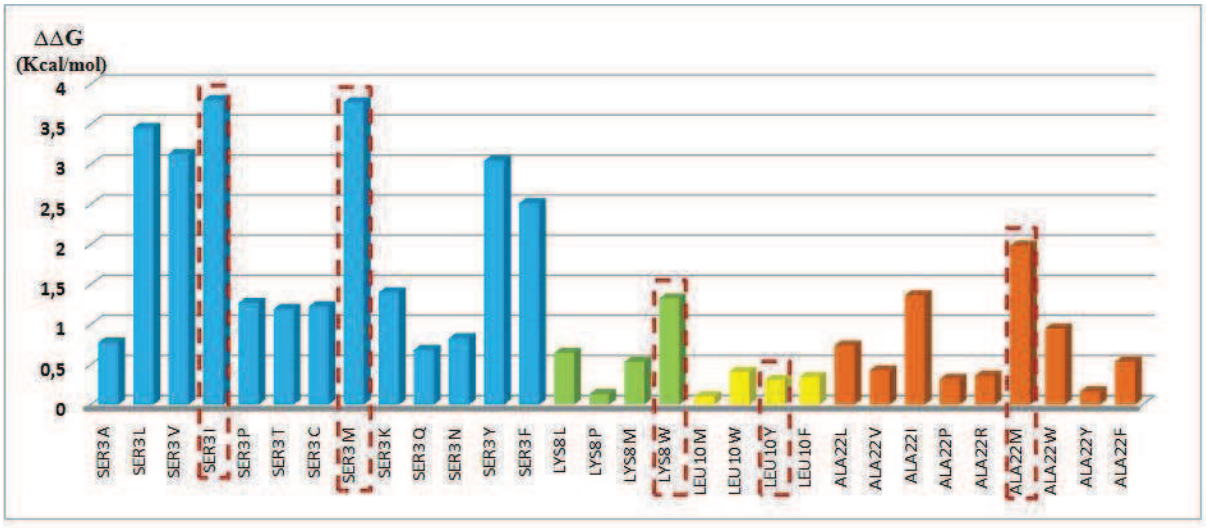


Figure III.10: Variation d'énergie libre ΔΔG (Kcal/mol), calculée avec le logiciel FoldX, et résultant de la mutation des résidus natifs de la plate-forme 1an1I_m en positions 3, 8, 10 et 22 par les 20 acides aminés naturels. Seules les substitutions énergétiquement favorables sont représentées dans le graphique (ΔΔG > 0).

Sur la base de ces résultats, nous avons choisi de produire des variants basés sur des doubles mutations en gardant systématiquement la substitution Ser3Ile : Ser3Ile/Lys8Trp ; Ser3Ile/Leu10Tyr et Ser3Ile/Ala22Met, ainsi que deux variants basés sur des quadruples mutations :

- Ser3Ile/Lys8Trp/Leu10Tyr/Ala22Met ;
- Ser3Met/Lys8Trp/Leu10Tyr/Ala22Met.

Les résidus Ile et Met semblent énergétiquement équivalents pour la position 3, des quadruples variants avec l'un ou l'autre de ces deux résidus ont donc été produits.

Sur le plan expérimental, les variants de la plate-forme 1an1I_m ainsi conçus ont été produits par biologie moléculaire en système bactérien selon le protocole utilisé pour la production de la plate-forme recombinante 1an1I_m. Les différents plasmides d'expression pour les 5 variants sont obtenus à partir du plasmide d'expression portant le gène codant pour la protéine 1an1I_m en réalisant de la mutagénèse dirigée par PCR. Après production, purification et caractérisation des différents variants conçus *in silico*, l'affinité de ces variants pour les MMPs ciblées a été déterminée. Les résultats sont rapportés dans le tableau III.5.

Tableau III.5: Constantes d'inhibition (μM) de l'activité enzymatique des MMPs pour la plate-forme 1an1I_m portant le motif d'intérêt (en rouge) et ses variants (en noir).

Code PDB	Motif	MMP-2	MMP-8	MMP-9	MMP-12	MMP-13	MMP-14
1an1I_m	SSCS	22	5	30	136	38	132
1an1I_m (S3M, K8W, L10Y, A22M)	SMCS	13	2	20	36	11	59
1an1I_m (S3I, K8W, L10Y, A22M)	SICS	16	3	20	71	23	99
1an1I_m (S3I, A22M)	SICS	19	18	33	122	169	70
1an1I_m (S3I, K8W)	SICS	17	7	31	85	42	93
1an1I_m (S3I, L10Y)	SICS	26	16	24	85	64	51

Les résultats des tests d'inhibition montrent que les deux variants portant les quadruples mutations S3M ou S3I/K8W/L10Y/A22M possèdent une meilleure affinité pour les MMPs par rapport à la plate-forme de 1ère génération 1an1I_m portant le motif d'intérêt. Toutefois les gains d'affinité observés sont très faibles (un facteur 1,5 à 3,7 entre le ligand initial 1an1I_m et le variant 1an1I_m S3M/K8W/L10Y/A22M sur les différentes MMPs testées).

Ces résultats appellent plusieurs remarques. En dépit de l'utilisation du modèle moléculaire de l'interaction du composé initial 1an1I_m avec la cible MMP-8, le choix des substitutions d'acides aminés sur la base de l'estimation empirique de la variation d'énergie libre à l'aide d'une méthode comme FoldX (Guerois et al., 2002) ou Rosetta (Kortemme et Baker, 2002) comporte une incertitude importante. Le paramétrage de ces méthodes empiriques sur des bases de données de quelques centaines de mutations expérimentales dans quelques dizaines de complexes conduit à une situation où la déviation moyenne est de l'ordre de 1 kcal/mol. Cependant, l'examen détaillé de ces données montre que les écarts entre les données calculées et les données expérimentales peuvent être supérieurs à 5 kcal/mol dans certains cas. Par ailleurs, la plupart des mutations réalisées par des méthodes expérimentales aux interfaces protéine-protéine conduisent à des variations d'énergie libre d'association faibles (< 1 kcal/mol). Il serait donc nécessaire de pouvoir disposer de méthodes de calculs possédant une justesse de l'ordre de cette valeur. Il faut également garder en mémoire que la variation d'énergie libre consécutive à la substitution de plusieurs acides aminés ne correspond pas à la somme des contributions de chacune des substitutions. Ces observations suggèrent que l'optimisation d'interface protéine-protéine est à ce jour difficilement réalisable *in silico* par des méthodes de calculs reposant sur des estimateurs empiriques de la variation d'énergie libre

tels que FoldX ou Rosetta. Cette optimisation doit être abordée à l'aide d'une approche expérimentale permettant de produire et de tester un grand nombre de ces variants impliquant de multiples substitutions simultanées d'acides aminés à l'interface considérée.

III.5 Caractérisation structurale de l'interaction mini-protéine/MMP

Les données de liaison des candidats ligands et de certains de leurs analogues, obtenues au cours de ce travail suggèrent fortement que le mode de liaison de ces composés correspond aux modèles d'interactions établis au cours de la conception *in silico*. Ce mode est guidé par la présence du motif fonctionnel « greffé » sur les différentes plates-formes. Nous avons néanmoins cherché à caractériser expérimentalement ce mode de liaison. Ainsi, des études de cristallogenèse de l'interaction entre certains ligands protéiques conçus *in silico* et le domaine catalytique de la MMP-12 ont été entreprises. Pour ce faire, nous avons sélectionné certains des variants synthétiques de 1edpA possédant les affinités les plus élevées pour les domaines catalytiques testés tels 1edpA (1-18) S2ΦΦ. Des essais de cristallogenèse ont également été obtenus avec le composé 1edpA possédant une sérine en position 2 (1edpA (1-18)). Le choix de mener ces essais de cristallogenèse sur ces deux composés en interaction avec la MMP-12 était motivé par leur affinité élevée pour ce domaine catalytique mais également par l'expérience acquise au laboratoire concernant la cristallisation de complexes de la MMP-12.

La méthodologie utilisée dans les essais de cristallogenèse est celle de la goutte assise en diffusion de vapeur permettant l'obtention de cristaux de MMP en complexe avec des inhibiteurs possédant des affinités μM pour leur cible. Les cristaux de MMP sont formés à partir d'un mutant inactif F67D du domaine catalytique de la MMP-12 humaine, en présence d'un inhibiteur à large spectre des MMPs, le AHA, avant d'être transférés dans un bain contenant le ligand protéique d'intérêt. Des cristaux ont été obtenus pour les deux analogues 1edpA (1-18) et 1edpA (1-18) S2ΦΦ. Les cristaux obtenus avec le composé 1edpA (1-18) S2ΦΦ ont diffracté à une résolution de 2,2 Å et ceux obtenus avec 1edpA (1-18) à 2,8 Å. L'analyse des données obtenues pour 1edpA (1-18) S2ΦΦ a permis de caractériser la structure du motif 1-4 porté par 1edpA (1-18) S2ΦΦ dans le site catalytique de la MMP-12. La densité électronique correspondant à ce motif est représentée sur la figure III.11.

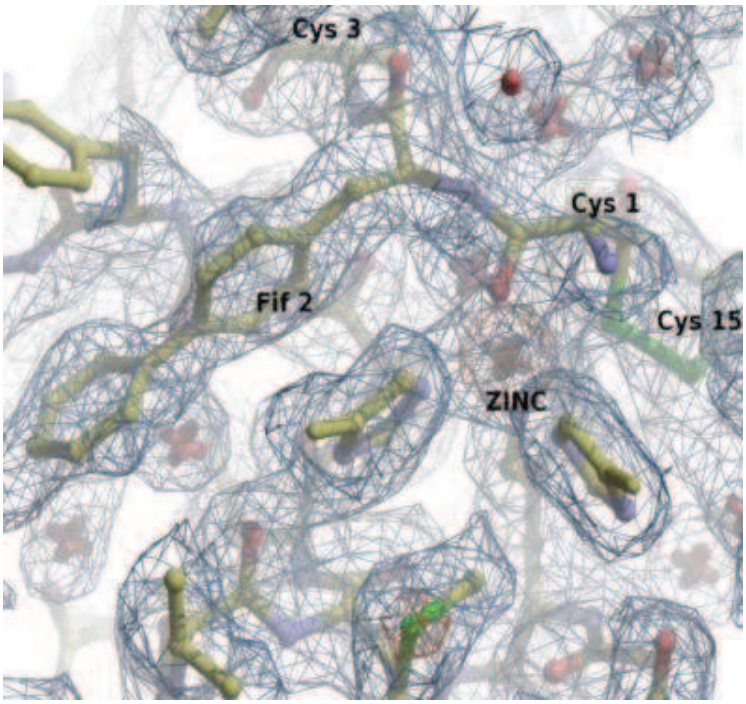


Figure III.11: Densité électronique correspondant à la structure cristallographique du variant 1edpA (1-18) S2ΦΦ en complexe avec la MMP-12. La vue agrandie montre la densité électronique des atomes des 4 résidus formant le motif de liaison du ligand dans le site actif de la MMP.

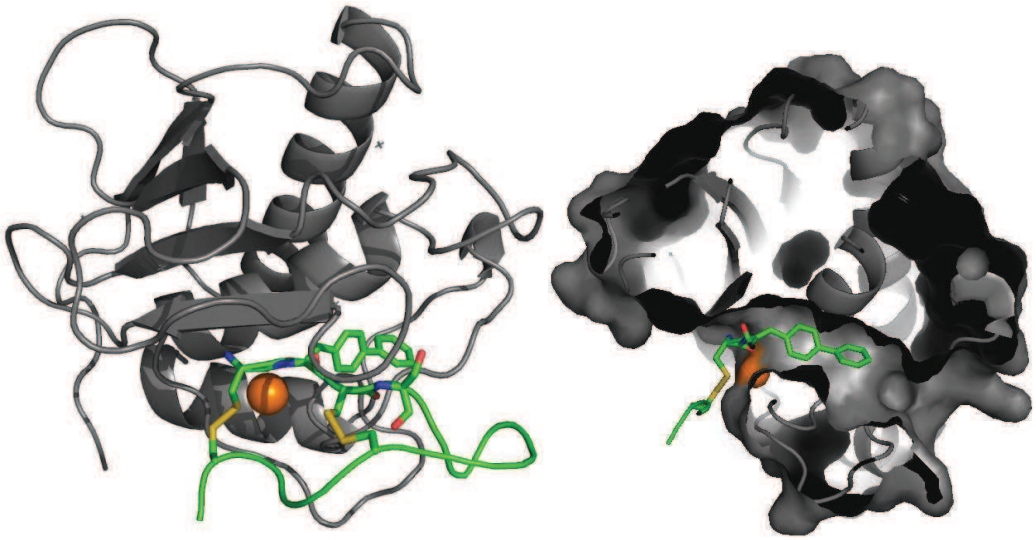


Figure III.12 : Structure cristallographique aux rayons X du variant 1edpA (1-18) S2ΦΦ en complexe avec le domaine catalytique de la MMP-12h. A gauche : Vue de face du complexe. Les protéines sont représentées en mode cartoon en gris pour la MMP-12h et en vert pour le ligand ; le motif de liaison du ligand est représenté en mode bâtons et le zinc catalytique est représenté par une sphère orange. A droite : Vue en coupe de la MMP-12h représentée en mode surface, mettant à nu la cavité S_1' dans laquelle plonge l'acide aminé biphenylalanine.

La structure expérimentale du complexe MMP-12/1edpA (1-18) S2ΦΦ montre que le ligand interagit dans le site actif de la MMP-12 de manière identique au mode d'interaction décrit par les modèles moléculaires établis au cours de la phase de conception *in silico*, et de manière identique aux inhibiteurs naturels de MMPs : les TIMPs. En particulier, les résultats indiquent que le mode d'interaction du motif 1-4 présent au niveau de l'extrémité N-terminale de l'analogue 1edpA (1-18) S2ΦΦ est similaire à celui observé dans le ligand naturel TIMP-2. La figure III.10 montre le résidu Cys 1 du motif de liaison en interaction avec le zinc catalytique de la MMP-12 qui est entouré des résidus His217 et His222. Dans le cas particulier du complexe MMP-12h/1edpA (1-18) S2ΦΦ, la structure cristallographique met en évidence l'acide aminé biphenylalanine plongeant dans la cavité S_1' (Figure III.12). En confirmant le bon mode de liaison de la plate-forme avec la cible, cet important résultat valide les hypothèses sur lesquelles se fonde la conception de mini-protéines inhibitrices de MMPs dans ce projet.

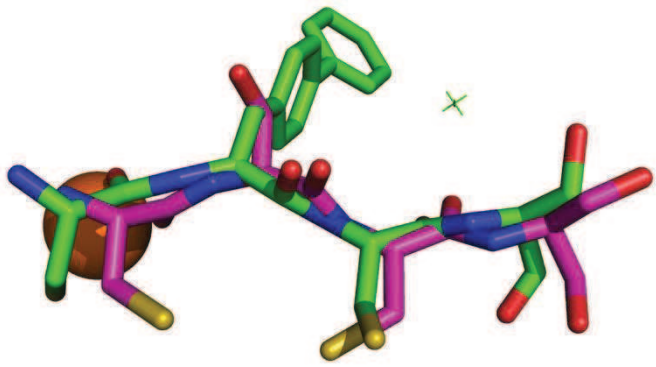


Figure III.13: Superposition des motifs 1-4 du variant 1edpA (1-18) S2ΦΦ et du ligand de référence TIMP-2. Les motifs sont représentés en mode bâtons et correspondent à la structure cristallographique du variant 1edpA (1-18) S2ΦΦ (vert) issue du complexe avec la MMP-12, et à la structure cristallographique du TIMP-2 (rose) issue du complexe 1bqq. Le zinc catalytique est représenté par une sphère orange.

La superposition des résidus du motif 1-4 du composé 1edpA (1-18) S2ΦΦ dans le cristal obtenu en interaction avec la MMP-12 montre un excellent accord avec la structure du segment 1-4 de l'inhibiteur naturel TIMP-2 en complexe avec la MMP-14 (Figure III.13) ; le calcul de RMSD sur les positions des 8 atomes Cα et Cβ du motif mime (C-ΦΦ-C-S) du variant 1edpA (1-18) S2ΦΦ superposés aux Cα et Cβ du motif de référence (C-S-C-S) du TIMP-2 (structure 1bqq) donne une valeur égale à 0,566 Å.

Des observations très similaires ont été faites en analysant la structure 3D obtenue par diffraction X pour le complexe du composé 1edpA (1-18) en interaction avec la MMP-12. Le

segment 1-4 (CSCS) de ligand protéique interagit au cœur du site catalytique de l'enzyme et le résidu sérine en position 2 occupe le sous-site S_1'.

Les distances interatomiques entre le motif CSCS de 1edpA (1-18) et la MMP-12 ainsi que celles caractérisant l'interaction du motif C-ΦΦ-C-S de 1edpA (1-18) S2ΦΦ avec le domaine catalytique de la MMP-12 ont été calculées à partir de la structure cristallographique des complexes obtenus. Ces distances ont été comparées avec celles observées entre le motif fonctionnel CSCS du ligand naturel TIMP-2 et les résidus du domaine catalytique de la MMP-14, calculées à partir de la structure cristallographique du complexe 1bqq (Tableau III.6).

Tableau III.6 : Distances interatomiques (Å) entre le motif CSCS du ligand de référence TIMP-2 et le domaine catalytique de la MMP-14, calculées à partir de la structure cristallographique du complexe 1bqq. Ces distances interatomiques sont comparées avec celles observées entre le motif CSCS de 1edpA (1-18) ou le motif C-ΦΦ-CS du variant 1edpA (1-18) S2ΦΦ et le domaine catalytique de la MMP-12, à partir des structures cristallographiques des complexes correspondants. Pour les résidus des MMPs, la numérotation correspond à celle de la MMP-14 à gauche et à celle de la MMP-12 à droite.

	Résidu	N°	Atome		Résidu	N°	Atome	1bqq	1edpA (1-18)	1edpA (1-18)S2ΦΦ
Ligand	CYS	1	N	MMP	ZN	289/264	ZN	**2,9**	**2,0**	**4,0**
Ligand	CYS	1	O	MMP	ZN	289/264	ZN	**2,0**	**2,1**	**2,0**
Ligand	CYS	1	N	MMP	GLU	240/219	OE2	**2,7**	**2,7**	**2,9**
Ligand	SER/ΦΦ	2	N	MMP	ALA	200/182	O	**3,1**	**3,5**	**3,0**
Ligand	SER/ΦΦ	2	O	MMP	ALA	200/182	N	**3,8**	**3,7**	**3,5**
Ligand	SER/ΦΦ	2	OG	MMP	GLU	240/219	OE1	**2,9**	**4,5**	**-**
Ligand	SER/ΦΦ	2	O	MMP	LEU	199/181	N	**2,7**	**2,7**	**2,8**
Ligand	CYS	3	N	MMP	PRO	259/238	O	**3,1**	**3.7**	**3,4**
Ligand	CYS	3	O	MMP	TYR	261/240	N	**3,0**	**3,6**	**2,8**
Ligand	SER	4	N	MMP	GLY	197/179	O	**3,1**	**3,0**	**2,9**

Pour les deux complexes obtenus avec les analogues de 1edpA, la majorité des distances calculées pour les deux ligands artificiels sont proches de celles observées pour le ligand naturel. Pour ce qui concerne le ligand 1edpA (1-18) S2ΦΦ, 9 distances sur les 10 mesurées diffèrent de moins de 0,3 Å de celles observées dans le complexe TIMP-2/MMP-14. Cette importante similarité des distances caractéristiques des interactions intermoléculaires est illustrée sur la figure III.14. Dans le cas du composé 1edpA (1-18), 6 distances sur les 10 mesurées diffèrent de moins de 0,1 Å de celles observées dans le complexe TIMP-2/MMP-14

et 3 distances montrent des écarts compris entre 0,1 et 0,6 Å (Tableau III.6). Pour les deux ligands artificiels, l'écart le plus important observé avec la structure TIMP-2/MMP-14 concerne l'interaction de l'atome N terminal du motif 1-4 avec le Zinc (0,9 et 1,1 Å pour 1edpA (1-18) et 1edpA (1-18) S2ΦΦ respectivement). On voit que pour les deux analogues de 1edpA, la plupart des distances interatomiques observées avec des atomes du domaine catalytique de la MMP-12 sont compatibles avec l'établissement des différentes interactions identifiées dans le complexe de référence. Toutefois les différences observées peuvent avoir des conséquences sur les énergies d'interaction de ces composés avec les MMPs.

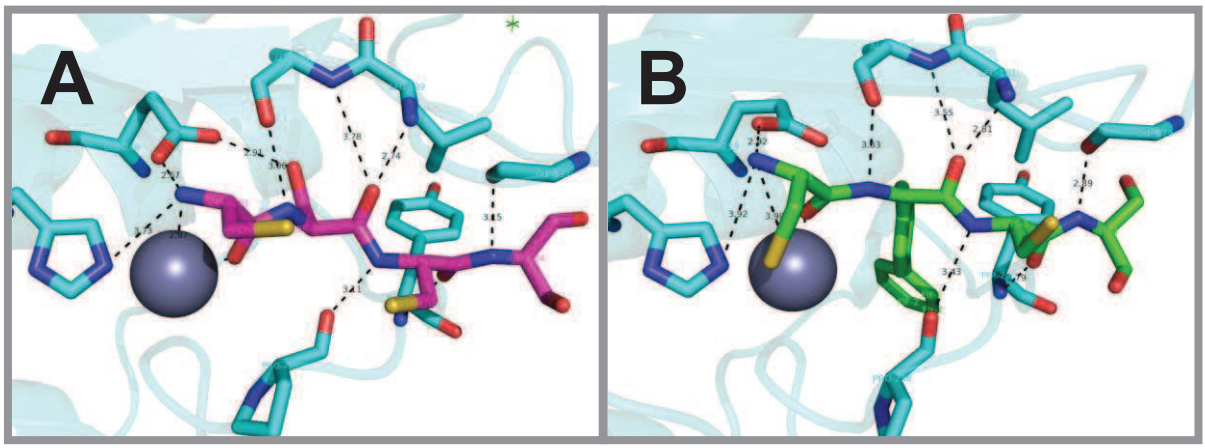


Figure III.14 : Distances interatomiques entre le motif de liaison du ligand et les résidus avoisinants du domaine catalytique de la MMP. La MMP est représentée en mode cartoon en cyan et le zinc catalytique en sphère grise. Les résidus impliqués dans les interactions interatomiques sont représentés en bâtons et les distances interatomiques en pointillés noirs. A : Motif fonctionnel CSCS du TIMP-2 en interaction avec les résidus de la MMP-14 (Code PDB 1bqq). B : Motif de liaison C-ΦΦ-C-S du variant 1edpA (1-18) S2ΦΦ en interaction avec les résidus de la MMP-12.

L'obtention de la structure tridimensionnelle des variants 1edpA (1-18) et 1edpA (1-18) S2ΦΦ par cristallographie avec une bonne résolution (2,8 et 2,2 Å respectivement) a permis de déterminer sans ambigüité le positionnement du motif de liaison dans le site actif de la MMP-12. Ces résultats montrent que les résidus du motif 1-4 se positionnent de manière similaire dans le contexte des ligands artificiels et dans celui du ligand naturel. Cette observation valide l'approche de conception *in silico* qui a permis de concevoir ces ligands protéiques artificiels en greffant le motif fonctionnel d'intérêt sur des plates-formes protéiques. La majorité des distances interatomiques caractéristiques des interactions intermoléculaires établies entre le motif 1-4 des ligands artificiels et le domaine catalytique de la MMP sont proches de celles mesurées entre les résidus du motif fonctionnel du ligand naturel TIMP-2 et les résidus

avoisinants de la MMP-14. Cependant, quelques distances présentent des différences qui pourraient contribuer à la plus faible affinité observée pour les ligands protéiques artificiels comparés aux ligands naturels. Ceci nous a incités à étudier en détails les caractéristiques géométriques et la stabilité des interactions intermoléculaires établies par ce motif 1-4 du TIMP-2 dans les différents contextes structuraux des ligands protéiques artificiels.

III.6 Etude de la stabilité des interactions par dynamique moléculaire

Les observations réalisées dans les complexes des ligands artificiels 1edpA (1-18) et 1edpA (1-18) S2ΦΦ en interaction avec la MMP-12 ont montré que le principe de la méthode de conception de ligand protéique mise en œuvre dans ces travaux pouvait permettre d'obtenir des ligands artificiels en contrôlant la position ciblée par ces ligands. Ceci est réalisé via le greffage d'un motif fonctionnel reproduisant un ensemble d'interactions intermoléculaires observées dans le complexe d'un ligand de référence en interaction avec la cible considérée. Dans le cas présent, les observations structurales faites avec les analogues de 1edpA montrent que, si une très grande majorité des interactions peut être reproduite, les caractéristiques géométriques de celles-ci peuvent présenter des différences avec celles observées dans le ligand naturel. Ces différences peuvent être à l'origine de la moindre affinité observée dans la série des 10 ligands protéiques artificiels comparée à celle des ligands naturels.

Malgré les efforts importants réalisés, nous n'avons pas pu obtenir de cristaux d'autres ligands artificiels en interaction avec des domaines catalytiques de MMPs. Ces données structurales auraient pu permettre une meilleure compréhension des raisons pour lesquelles les affinités mesurées pour ces autres ligands protéiques se situent majoritairement entre le micromolaire et quelques centaines de micromolaires. Ces faibles affinités peuvent avoir de multiples origines, en particulier des interfaces « non idéales » avec la cible. Cependant, il est possible que, dans le contexte des plates-formes identifiées, les interactions réalisées entre le motif 1-4 et la cible ne soient pas reproduites de manière satisfaisante. Les observations faites dans les complexes des analogues de 1edpA ont montré que, bien que le positionnement relatif des groupements des résidus 1-4 vis-à-vis de la MMP-12 soit similaire à ce qui est observé dans le ligand naturel, des différences dans la géométrie précise de ces interactions peuvent exister. Nous avons donc

cherché à mieux caractériser ces interactions et à évaluer leur stabilité relative en réalisant pour les 10 ligands protéiques artificiels en complexe avec la MMP-14, des simulations de dynamique moléculaire dans l'eau.

Les trajectoires obtenues pour les 10 ligands protéiques en complexe avec la MMP-14 ont été analysées et comparées à la trajectoire réalisée pour le complexe naturel TIMP-2/MMP-14. Deux paramètres ont été calculés pour les interactions caractéristiques établies entre les atomes du segment 1-4 et les atomes correspondants de la MMP-14 : les distances moyennes au long de la trajectoire et les taux d'occupation de ces interactions définis comme la fraction de temps pendant laquelle l'interaction est établie sur la durée de la trajectoire analysée. Cette durée a été déterminée en analysant les fluctuations des positions des atomes Cα des deux protéines par rapport à la structure de départ. A titre d'exemple, cette évolution est illustrée pour le complexe MMP-14/1edpA (1-18) sur la figure III.15. On peut voir que, au cours de la première partie de la simulation, la structure évolue sensiblement par rapport à la structure de départ. Sur cette figure on constate que le RMSD sur les positions des deux partenaires n'évolue plus après environ 200 ps. Une constatation similaire est faite concernant l'aire de l'interface du ligand 1edpA (1-18) dont la valeur évolue dans les premières 200 ps pour se stabiliser dans la suite de la trajectoire.

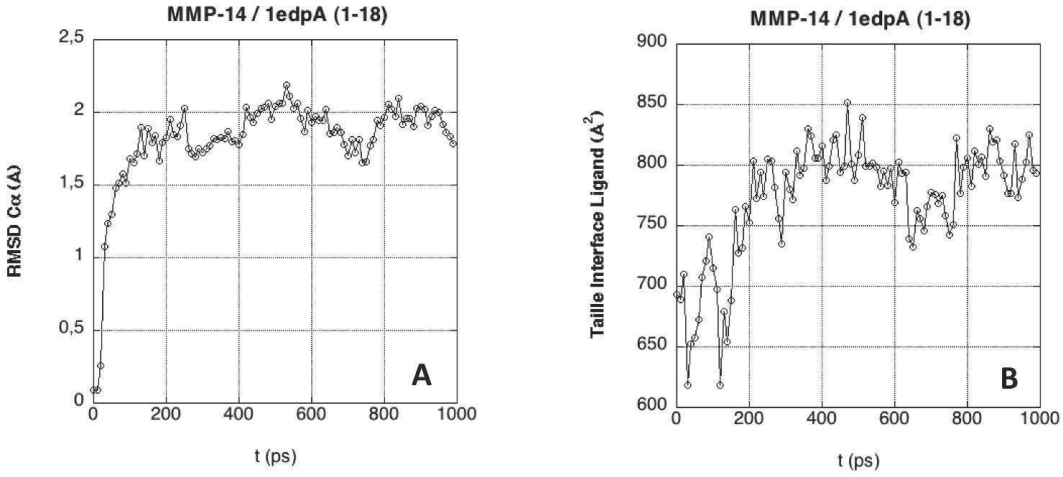


Figure III.15**:** Analyse de la simulation réalisée pour le complexe MMP-14/1edpA (1-18) dans l'eau. Le RMSD sur les positions des Cα des deux partenaires par rapport à la structure initiale (A) et l'évolution de la taille de l'interface du ligand (B) au cours de la trajectoire sont indiqués.

Il est à noter que la valeur moyenne de cette aire est supérieure de 10 à 15 % à la valeur calculée dans la structure initiale suggérant qu'une optimisation de l'interface a lieu dans les conditions de cette simulation. Toutefois, les différences structurales observées entre la structure initiale et des structures extraites de la seconde moitié de la trajectoire sont très faibles. Ces observations nous ont conduit à réaliser les calculs de moyenne de distance sur la deuxième moitié des trajectoires obtenues pour ces 10 ligands protéiques (t compris entre 500 ps et 1 ns).

La figure III.16 illustre les fluctuations observées pour un jeu des distances caractéristiques de l'interaction du motif 1-4 dans le ligand protéique 1edpA (1-18) avec la MMP-14. Après la période d'établissement de l'équilibre, les distances fluctuent autour d'une valeur moyenne. Les amplitudes de ces fluctuations sont variables et pour certaines d'entre elles importantes. Cette observation nous a conduit à calculer, pour les complexes étudiés, les taux d'occupation caractéristiques des interactions entre le motif 1-4 et la MMP-14 que nous cherchons à mimer. Ces occupations sont définies pour chaque interaction comme la fraction du temps pendant laquelle l'interaction est considérée comme établie. Ces valeurs ont été comparées à celles observées dans la simulation du complexe naturel TIMP-2/MMP-14. Les données sont reportées dans le tableau III.7. Dans le cas du complexe naturel, à l'exception de la liaison hydrogène MMP-14 A200(O) – TIMP-2 S2(N), présente 62 % du temps, les interactions analysées sont stables sur la durée de la trajectoire analysée (500-1000 ps) (taux d'occupation 1). Une situation similaire est observée pour 1edpA (1-18) où les interactions analysées, à l'exception d'une seule d'entre elles, présentent des taux d'occupation supérieurs à 80 %. Parmi les 10 ligands protéiques, 3 montrent une absence de l'interaction entre l'amine terminale du motif 1-4 et le zinc catalytique (1an1I_m, 1rooA_m et 1scyA_m). L'examen des trajectoires montre que ceci est lié à l'établissement d'interactions intramoléculaires des ligands entre cette amine terminale et des groupes présents à la surface du ligand. L'absence de contrôle structural lié à la nature du résidu 1 (sérine) autorise les modifications de conformation nécessaires à l'établissement de ces interactions intramoléculaires. Toutefois, on voit également dans d'autres plates-formes, que la présence d'une sérine en position 1 peut conduire à la reproduction correcte de l'interaction de l'amine terminale avec le zinc (1cz6A_m, 1lmmA_m, 1osxA_m, 2it8A_m et 2uuyB_m).

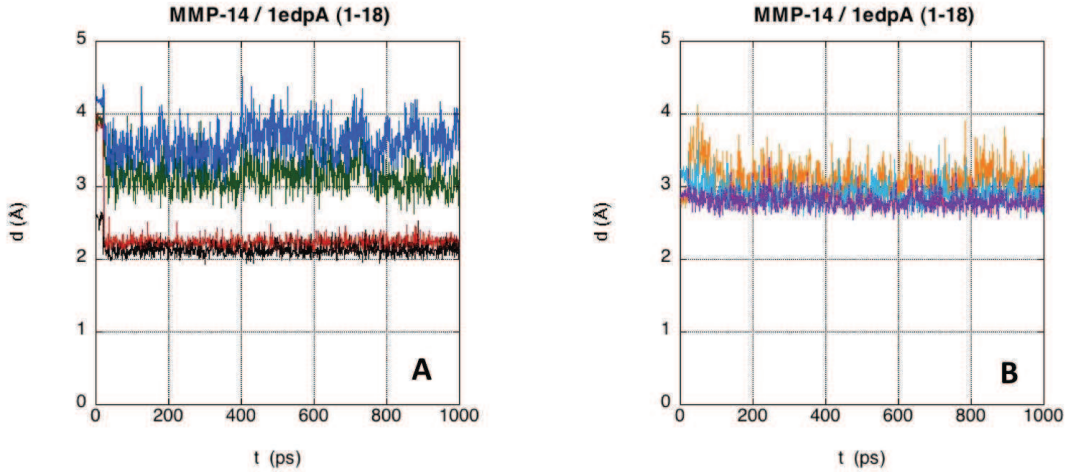


Figure III.16 **:** Analyse de la simulation réalisée pour le complexe MMP-14/1edpA (1-18) dans l'eau. Les distances reportées en A sont : MMP-14(ZN) – 1edpA (1-18) Cys1-N (rouge), MMP-14(ZN) – 1edpA (1-18) Cys1(O) (noir), MMP-14 A200(O) – 1edpA(1-18) Ser2(N) (vert), MMP-14 A200(N) – 1edpA(1-18) Ser2(O) (bleu) et en (B) : MMP-14 P259(O) – 1edpA(1-18) Cys3(N) (orange), MMP-14 P261(N) – 1edpA(1-18) Cys3(O) (turquoise) et MMP-14 G197(O) – 1edpA(1-18) Ser4(N) (violet).

Tableau III.7: Occupations calculées sur la trajectoire de dynamique moléculaire dans d'eau (500-1000 ps) pour un ensemble d'interactions entre les résidus du motif 1-4 dans les ligands naturels et artificiels et la MMP-14. Les valeurs inférieures à 0,8 sont indiquées en gras. Pour les interactions impliquant le zinc catalytique, le critère retenu est que la distance entre les deux atomes soit inférieure à 2,9 Å. Pour les liaisons hydrogène, un double critère est appliqué : que la distance ente le proton lié au donneur et l'accepteur soit inférieure à 2,8 Å et que l'angle DHA soit inférieur à 120 °.

Ligand Protéique	1_N	1_O	2_N	2_O	3_N	3_O	4_N
MMP-14	ZN1	ZN1	ALA200(O)	ALA200(N)	PRO259(O)	TYR261(N)	GLY197(O)
1bqqT	1	1	**0,62**	1	0,99	1	1
1an1I_m	**0**	1	**0,02**	0,98	0,97	**0,62**	**0**
1cz6A_m	1	1	**0,79**	1	0,98	0,97	**0**
1edpA	1	1	0,96	1	**0,72**	1	1
1lmmA_m	1	1	0,96	0,99	0,99	0,96	0,89
1osxA_m	1	1	**0,57**	1	0,99	**0,66**	1
1rooA_m	**0**	1	**0**	**0,67**	0,95	1	0,96
1scyA_m	**0**	1	**0,02**	0,91	0,99	1	**0,08**
2f3cI_m	1	1	0,81	1	0,95	**0,27**	0,92
2it8A_m	1	1	**0,35**	**0,57**	0,94	0,99	0,99
2uuyB_m	1	1	**0,52**	1	**0,64**	**0,47**	1

L'analyse de ces trajectoires montre que la plupart des interactions sont reproduites dans les 10 ligands étudiés. Toutefois, il existe quatre plates-formes pour lesquelles on observe une ou plusieurs interactions possédant un taux d'occupation voisin ou égal à 0 (1cz6A_m : 1,

1rooA_m : 2, 1an1I_m et 1scyA_m : 3). Les distances interatomiques moyennes correspondant à ces interactions calculées sur la seconde moitié de la trajectoire (Tableau III.8) confirment les observations faites avec les taux d'occupation. On note également l'excellent accord entre les distances moyennes calculées sur la trajectoire réalisée pour le complexe naturel MMP-14/TIMP-2 et les valeurs obtenues dans la structure cristallographique 1bqq (Tableau III.6).

Tableau III.8**:** Distances interatomiques moyennes calculées au cours de la simulation de dynamique moléculaire dans l'eau, caractéristiques des interactions établies entre le motif 1-4 dans les ligands naturel/artificiels et la MMP-14.

Ligand		Cible			1bqqT	1an1I_m	1cz6A_m	1edpA	1immA_m	1osxA_m	1rooA_m	1scyA_m	2f3cI_m	2it8A_m	2uuyB_m
Resid	Atom	Resn	Resid	Atom											
1	N	ZN2	289	ZN	2,3	5,14	2,33	2,29	2,31	2,31	4,01	3,98	2,33	2,29	2,28
1	O	ZN2	289	ZN	2,13	2,21	2,13	2,14	2,15	2,15	2,12	2,13	2,11	2,11	2,14
1	N	GLU	240	OE1	3,92	5,86	4	2,96	4,06	4	4,13	4,15	3,83	4	2,93
1	N	GLU	240	OE2	2,88	5,81	2,9	2,77	2,91	2,89	4,28	4,1	2,84	2,92	2,83
2	N	ALA	200	O	3,52	3,92	3,46	3,19	3,18	3,58	4,48	3,94	3,26	3,57	3,6
2	O	ALA	200	N	4,05	4,09	3,99	3,62	3,85	4,08	5,27	4,6	3,82	4,62	3,94
2	OG	GLU	240	OE1	2,71	4,16	2,7	5,14	2,68	2,69	3,28	2,7	2,71	2,77	4,66
2	OG	GLU	240	OE2	4,6	3,94	4,56	6,3	4,54	4,55	4,81	4,45	4,54	4,6	5,31
2	O	LEU	199	N	2,89	2,98	2,87	2,87	2,9	2,9	3,67	3,2	2,88	3,42	2,93
3	N	PRO	259	O	2,91	2,9	3,11	3,1	2,96	2,93	2,94	2,89	3,02	3,04	3,94
3	O	TYR	261	N	2,93	3,31	3,12	2,92	3,09	3,38	2,98	2,91	3,64	3,02	3,24
4	N	GLY	197	O	2,9	4,14	2,86	2,82	3	2,86	2,91	4,32	2,98	2,93	2,85

III.7 Conclusion

Cette étude montre que la majorité des 10 plates-formes sélectionnées pour le greffage du motif fonctionnel 1-4 du TIMP-2 peut conduire à une reproduction satisfaisante des interactions intermoléculaires établies par ce motif dans le contexte du ligand protéique naturel, le TIMP-2. Toutefois, les simulations de dynamique moléculaire réalisées avec les 10 complexes montrent également que plusieurs de ces plates-formes sont susceptibles de ne pas reproduire certaines de ces interactions. D'autre part, la stabilité de certaines interactions reproduites dans le contexte du ligand artificiel pourrait être réduite par rapport à celle observée dans le complexe de la MMP-14 avec le ligand naturel.

On peut enfin noter qu'il n'est pas possible de corréler directement les effets dynamiques observés concernant les interactions entre le motif greffé et la MMP-14 avec les affinités mesurées des composés correspondants. En effet, d'autres facteurs varient d'une plate-forme à l'autre, en particulier la composition spécifique de chacune des interfaces. Il est néanmoins probable qu'une reproduction imparfaite des interactions du motif greffé, tant sur le plan géométrique que dynamique, puisse contribuer à la plus faible affinité observée pour les 10 candidats ligands conçus dans le cadre de ce travail. Ceci suggère également que la sélection des plates-formes pourrait tirer bénéfice de la prise en compte des aspects dynamiques révélés par les simulations présentées ci-dessus.

CONCLUSION ET PERSPECTIVES

Affinités mesurables et diversité de ligands : des atouts pour le développement des méthodes hybrides :

La conception *in silico* de ligands protéiques constitue aujourd'hui une voie de grand intérêt pour le développement de molécules d'intérêt thérapeutique ou diagnostique. Les progrès très importants réalisés dans ce domaine sont liés à la conjonction de plusieurs facteurs. Tout d'abord, les données structurales concernant les structures tridimensionnelles de protéines et de complexes protéiques naturels disponibles ne cessent de s'accroitre. La PDB contient à ce jour 84223 structures dont 77958 structures de Protéine (RCSB 08/12), la très grande majorité d'entre elles sont des structures obtenues par diffraction des rayons X. D'autre part, pour exploiter cette ressource structurale, des programmes de plus en plus rapides et efficaces ont été développés permettant l'identification de motifs de résidus dans une topologie définie au sein des structures de protéines. Parmi ceux-ci, le programme STAMPS développé au laboratoire a déjà permis de réaliser plusieurs applications (Amrein et al., 2012; Magis et al., 2006). Les moyens de calculs aujourd'hui accessibles constituent également un atout important pour la conception *in silico* de protéines artificielles. Enfin, on doit noter que les derniers succès spectaculaires reportés en matière de conception *in silico* de ligands protéiques (Azoitei et al., 2012; Azoitei et al., 2011; Fleishman et al., 2011) et de conception d'enzymes artificielles (Jiang et al., 2008; Rothlisberger et al., 2008; Siegel et al., 2010) sont liés au couplage de méthodes de conception *in silico* avec des méthodes d'optimisation expérimentales extrêmement puissantes des structures protéiques initiales obtenues *in silico* (Peisajovich and Tawfik, 2007). Ces observations suggèrent que l'avenir proche pourrait voir se généraliser des approches hybrides de conception de ligands protéiques dans lesquelles la phase de conception *in silico* joue un rôle central (Figure IV.1). Le succès de telles approches dépendra de l'efficacité de la phase initiale de conception *in silico*. A l'issue de cette phase de conception, il est essentiel que les ligands protéiques conçus possèdent une affinité mesurable et suffisamment élevée pour la cible, permettant de faciliter leur évaluation expérimentale. D'autre part, pouvoir disposer d'une série de ligands protéiques présentant des caractéristiques physico-chimiques différentes et des modes de liaison différents à la surface de la cible pourrait accroitre les chances finales de succès. En effet, chaque candidat ligand protéique est susceptible de ne pas satisfaire l'une des nombreuses étapes séparant sa conception *in silico* de son utilisation finale dans un contexte diagnostique ou thérapeutique. En effet, la facilité de production de la protéine plate-forme avec un rendement suffisant, sa sélectivité vis à vis de la

cible, sa pharmaco-cinétique, sa pharmaco-dynamique et sa stabilité *in vivo*, sont autant de paramètres qui ne peuvent pas être pris en compte dans la conception initiale. Pouvoir disposer de plusieurs plates-formes pour le développement d'un ligand peut donc constituer un atout important pour atteindre l'objectif.

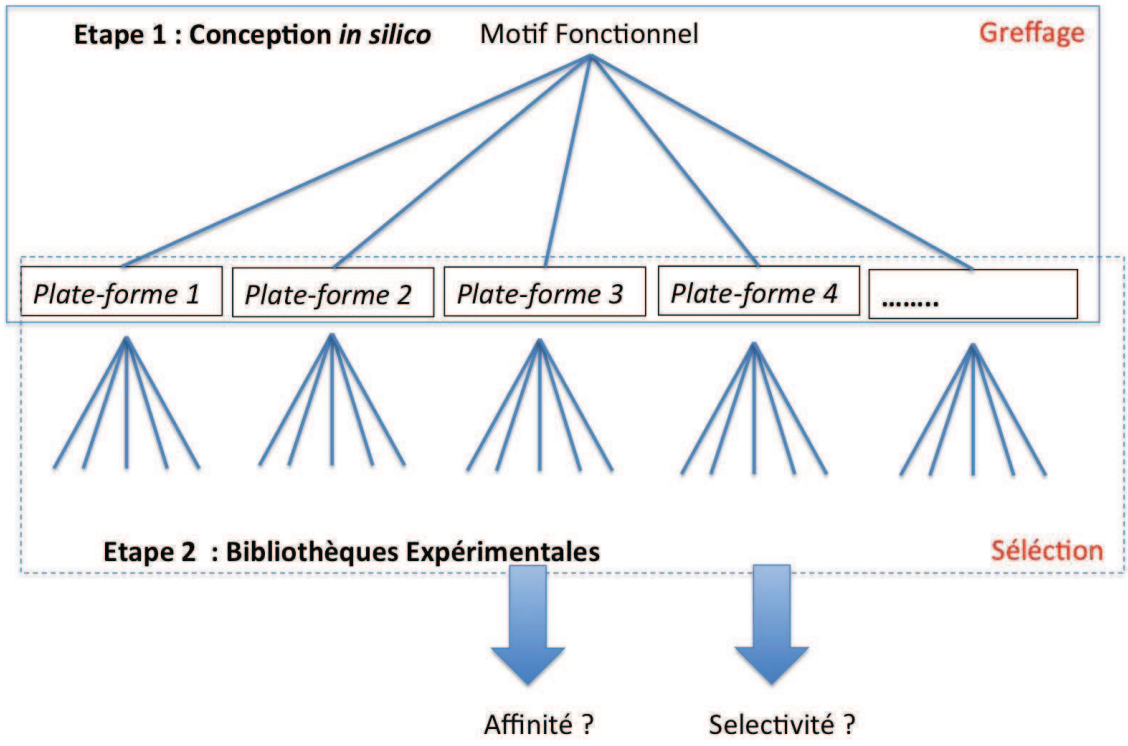


Figure IV.1 : Schéma de principe des approches hybrides de conception de ligands protéiques par greffage de motifs fonctionnels. Des candidats ligands sont conçus *in silico* grâce au greffage de motifs de résidus à la surface de plates-formes protéiques de petites tailles. Après mise en évidence d'une affinité initiale, des bibliothèques de variants conçus en exploitant le mode de liaison imposé par le greffage du motif fonctionnel sont réalisées. Les variants possédant les propriétés recherchées sont sélectionnés dans ces bibliothèques.

Rendement des méthodes de conception in silico

En dépit des succès reportés dans la littérature, l'analyse des données accessibles dans les études citées plus haut montre que le taux de succès des méthodes *in silico* appliquées (fraction de composés conçus *in silico* possédant effectivement une affinité/activité mesurable) dans ces travaux est très faible. Ainsi, dans le cas de la conception d'un ligand capable d'interagir dans une région de la tige de la protéine hémagglutinine (HA) du virus pandémique H1N1 1918, seuls 2 des 73 candidats ligands issus de la conception *in silico* et qui ont été exprimés à la surface de phages ont montré une affinité mesurable (respectivement 200 nM et 5 µM (Fleishman et al., 2011). Dans ce cas, l'approche *in silico* était fondée sur le greffage d'un motif

d'acides aminés, dont les positions optimales ont été identifiées à l'aide d'une méthode d'ancrage *in silico* à la surface de la cible. Ces motifs ont été greffés sur des plates-formes protéiques sélectionnées pour reproduire la topologie de ce motif. Une seconde étude, portant sur la conception d'une protéine capable de présenter un épitope de la protéine d'enveloppe du virus HIV, la gp120, a également montré un taux de succès très faible. Dans ce travail, les auteurs ont greffé deux segments d'acides aminés présents dans la gp120 sur des plates-formes sélectionnées pour leur capacité à recevoir ces greffes de squelette (Azoitei et al., 2011). Sur les 62 candidats testés, seuls 25 ont pu être exprimés et produits sous forme soluble. Cette observation peut s'expliquer en partie par le travail de remodelage important réalisé à l'aide de méthodes empiriques présentant une incertitude considérable. De plus, un seul de ces 25 candidats ligands testés a montré une affinité mesurable mais très faible (300 μM). Les travaux du même groupe portant sur le complexe 2F5/gp41 ont montré que seuls 35 variants sur les 98 conçus pour interagir avec l'anticorps 2F5 ont pu être produits *in vitro* (Azoitei et al., 2012). Dans le domaine du développement d'enzymes artificielles, fondée sur la recherche de plates-formes protéiques pour le greffage d'un motif fonctionnel, des observations similaires ont été réalisées. Les travaux de conception *de novo* d'enzymes catalysant la réaction d'élimination de Kemp ont permis d'obtenir 8 enzymes montrant une activité mesurable sur 59 plates-formes testées donnant un rendement de 13% (Röthlisberger et al., 2008). Un rendement de 4% a été obtenu pour les travaux relatifs à la conception *de novo* d'enzymes DielsAlderase avec seulement 2 variants montrant une activité DielsAlderase sur 50 plates-formes testées (Siegel et al., 2010). De plus, il faut noter que dans d'autres études, le taux de succès de la phase *in silico* n'est pas connu soit parce que les données ne sont pas reportées soit parce que la sélection des plates-formes aboutit à un candidat ligand unique dont l'étude expérimentale est poursuivie (Liu et al., 2007).

Nécessité d'évaluer l'efficacité de la méthode de conception *in silico*

La méthode de conception *in silico* utilisée au cours de ce travail a été précédemment appliquée à la conception de ligands des canaux Kv1.2 (Magis et al., 2006). Cette méthode est fondée sur l'exploitation systématique des structures tridimensionnelles de protéines naturelles décrites dans la PDB pour le greffage de chaines latérales. Ce greffage peut être indépendant de la structure secondaire portant les résidus du motif dans le ligand protéique naturel à mimer. Un objectif important de cette méthode est de sélectionner des structures et de les utiliser en leur

apportant la perturbation la plus faible possible lors du greffage du motif fonctionnel. Notre approche se distingue donc de celle reportée par Azoitei et al. (Azoitei et al., 2011) qui utilisent la sélection et le remodelage massifs *in silico* de la structure de la plate-forme. En principe, cette dernière approche possède l'avantage de pouvoir permettre le greffage d'épitopes de plus grande taille. Cependant, les incertitudes inhérentes aux méthodes de remodelage fondées sur des fonctions d'énergies empiriques contribuent certainement à la faiblesse du taux de succès observé. La première application de notre approche avait abouti à la conception de quatre ligands dont la production et l'évaluation comme ligands des canaux Kv1.2 avaient été réalisées. Les résultats de cette étude avaient conduit à un taux de succès de 75% et trois de ces ligands possédaient des affinités proches de 1 μM. Les études de conception *in silico* de ligands protéiques sont cependant encore rares et il est nécessaire de poursuivre leur évaluation. Par ailleurs, comme nous l'avons déjà souligné, une caractéristique générale des approches *in silico* de conception de ligands protéiques dans les différentes études reportées est l'obtention de ligands protéiques initiaux de faible affinité, en général plusieurs ordres de grandeur plus faible que l'affinité des composés naturels vis-à-vis des mêmes cibles. Cette dernière observation montre que, quels que soient les méthodes mises en œuvre, certains éléments importants pour la liaison des ligands protéiques de première génération ne sont pas pris en compte dans la conception *in silico* de ces composés.

Complémentarité de surface et aire d'interface

L'objectif de ce travail de thèse était de concevoir une série de ligands protéiques capables de se lier au site catalytique des membres de la famille des MMPs. Le choix du motif à greffer repose sur les connaissances acquises concernant l'interaction des inhibiteurs protéiques naturels des MMPs, les TIMPs, tant sur le plan structural (Fernandez-Catalan et al., 1998; Gomis-Rüth et al., 1997) que sur le plan fonctionnel (Butler et al., 1999 ; Huang et al., 1997 ; Meng et al., 1999). Ces travaux nous ont conduit à sélectionner le motif 1-4 du TIMP-2 comme motif fonctionnel. Le programme STAMPS a permis d'identifier, dans la PDB, des plates-formes capables de reproduire la topologie du motif fonctionnel. Les critères de sélection appliqués au cours de cette recherche concernent la topologie du motif de résidus équivalents au motif 1-4 du TIMP-2 tel que décrit dans la structure du complexe de cet inhibiteur naturel avec la MMP-14 et la compatibilité stérique de la plate-forme pour l'interaction avec la cible. Cette compatibilité est évaluée par le programme STAMPS à l'aide d'un estimateur appelé

facteur de recouvrement stérique. Ce critère très simple a permis la sélection d'une série de 10 plates-formes. L'évaluation de la similarité de forme, à l'aide de l'indice de Hodgkin, entre chacun des 10 ligands protéiques conçus *in silico* et le ligand naturel n'a également pas révélé de différence significative dans la région d'interaction avec la cible. En effet, les valeurs de ces indices et leur fluctuation dans la série des 10 ligands se situent dans la gamme observée dans des complexes protéine-protéine naturels (Magis et al., soumis ; Chapitre II, II.6). Des modèles moléculaires d'interaction des 10 partenaires protéiques ont été réalisés dans le champs de force de CHARMM par minimisation d'énergie et dynamique moléculaire. Dans ces modèles, nous avons observé que la compatibilité stérique était satisfaisante car, pour les 10 plates-formes, les indices de « gap volume » des interfaces se situent dans la gamme obtenue pour ce paramètre dans des complexes protéine-protéine naturels (Jones et Thornton, 1996). On peut également noter que les modèles moléculaires n'ont pas révélé d'incompatibilité des partenaires sur le plan structural ou énergétique. Nous avons donc montré qu'il était possible de reconstituer des interfaces présentant des caractéristiques de forme adaptées à la cible considérée avec les 10 plates-formes possédant des structures secondaire/tertiaire très différentes entre elles et différentes de celle du ligand naturel TIMP-2. Cette observation est conforme aux observations reportées pour des ligands protéiques naturels (DeLano et al., 2000 ; Keskin et Nussinov, 2005). Une observation similaire a été réalisée dans le cas de séries de ligands protéiques naturels de trois sérine-protéases (Magis et al, soumis ; Chapitre II, II.6).

Par ailleurs, les analyses des modèles moléculaires ont montré que les 10 plates-formes protéiques sélectionnées possédaient des aires d'interface situées dans la gamme de celles observées dans les complexes protéine-protéine naturels (Jones et Thornton, 1996 ; Lo Conte et al., 1999). Ces interfaces sont représentées sur la figure IV.2 pour les 10 ligands protéiques artificiels et le ligand naturel dans les modèles d'interaction avec le domaine catalytique de la MMP-14. Cette observation est intéressante car aucun critère particulier de sélection concernant l'aire de l'interface n'a été utilisé lors de la recherche de plates-formes. Il se peut donc que ce résultat soit fortuit mais cela montre également qu'il est possible, à l'aide du criblage systématique de la PDB réalisé par STAMPS, d'obtenir des plates-formes possédant les caractéristiques topologiques permettant le greffage du motif tout en offrant une interface de compacité et d'aire suffisantes pour la conception d'un ligand protéique artificiel. Par ailleurs, l'inclusion future, dans STAMPS, d'un calcul d'aire d'interface entre la plate-forme et la cible permettrait d'étudier systématiquement ce paramètre. Ainsi, s'il s'avérait dans d'autres cas

qu'une fraction significative des plates-formes identifiées présentait des aires d'interface trop faibles, il serait possible d'introduire un nouveau critère de sélection sur ce paramètre.

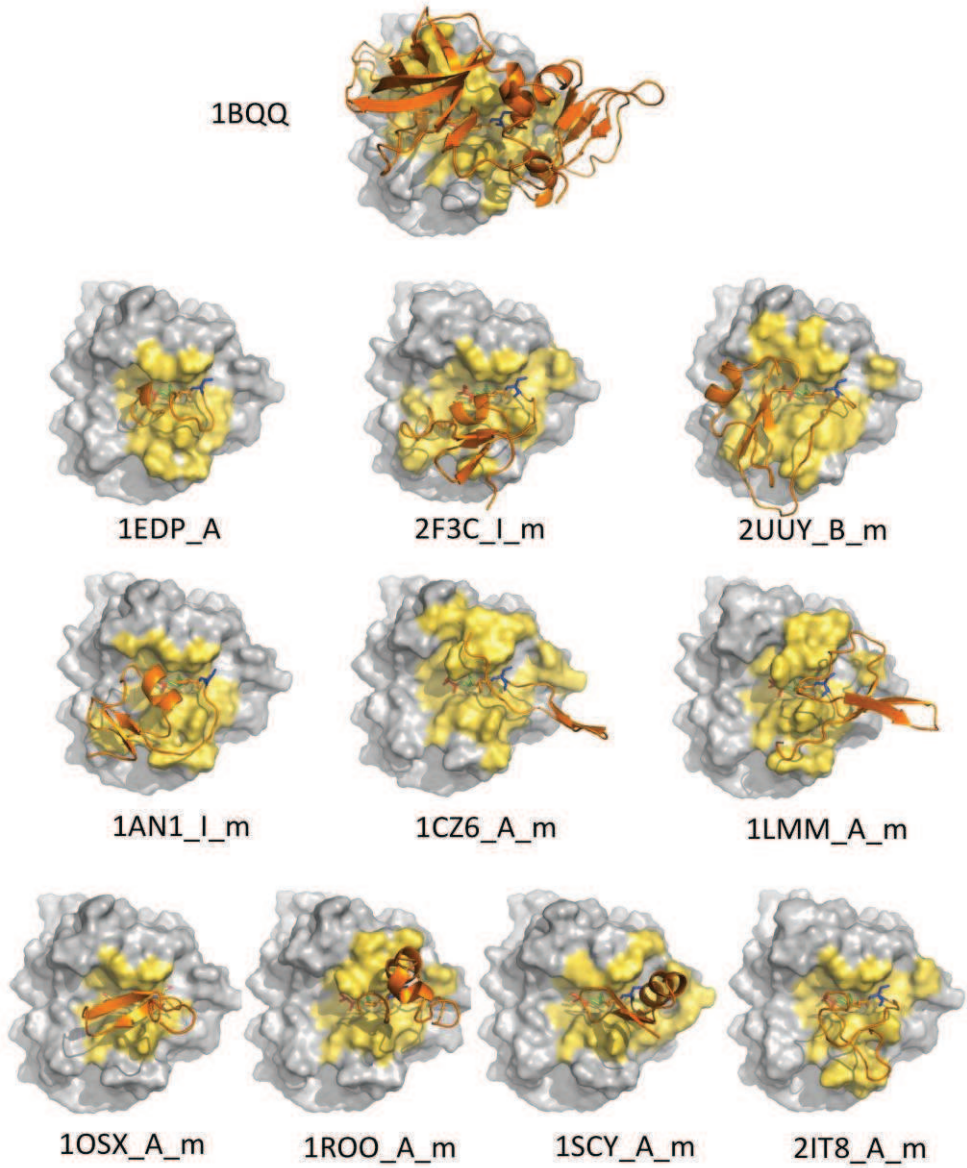


Figure IV.2: Représentation des aires des interfaces de la MMP-14 avec les 10 ligands protéiques artificiels et le ligand protéique naturel, le TIMP-2. Les structures correspondent à celles obtenues lors de la conception *in silico* des ligands protéiques artificiels. Pour chaque complexe, le domaine catalytique de la MMP-14 est représenté en mode surface de couleur grise, excepté la partie de l'enzyme située à une distance inférieur à 3,5 Å du ligand protéique, qui est représentée en jaune. Les ligands protéiques sont représentés en mode cartoon (orange).

On peut également noter que la taille des interfaces est un critère important pour la phase d'optimisation des plates-formes qui peut être réalisée en modifiant, sur le ligand protéique de

première génération, la nature des résidus présents à l'interface avec la cible. La figure IV.2 montre que les régions de la MMP-14 impliquées dans la liaison diffèrent selon les plates-formes. Ce dernier point pourrait se révéler capital pour le développement futur de composés affins et sélectifs d'un membre particulier de la famille des MMPs. En effet, des bibliothèques construites sur la base de l'une ou l'autre des plates-formes pourraient permettre d'exploiter des régions particulières de la surface du domaine catalytique. Des stratégies d'optimisation spécifiques de chaque plate-forme initiale pourraient donc être définies, à partir des composés initiaux pour obtenir des ligands protéiques de meilleure affinité et/ou de meilleure sélectivité.

Propriétés électrostatiques

Il est communément admis que les propriétés électrostatiques des protéines jouent un rôle central dans les interactions établies avec leurs partenaires (Honig et Nicholls, 1995). Le champ électrostatique est susceptible de jouer un rôle dans l'association des partenaires (Wade et al., 1998a) mais également dans l'affinité et la sélectivité des interactions protéine-protéine. L'indice de similarité de Hodgkin a été utilisé pour identifier des propriétés communes de protéines possédant de grandes divergences de structure primaire (Botti et al., 1998 ; De Rienzo et al., 2000 ; Demchuk et al., 1994 ; Wade et al., 1998b). Cet indice a également été utilisé pour aligner des protéines de fonctions similaires mais présentant des structures primaires, secondaires et tertiaires différentes (Ullmann et al., 2000 ; Ullmann et al., 1997). Dans leur étude, De Rienzo et ses collaborateurs ont étudié la sensibilité de l'indice de similarité de Hodgkin aux fluctuations structurales et ont pu montrer que la distribution des indices de similarité des potentiels électrostatiques entre des conformères issus de l'analyse par résonance magnétique nucléaire d'une même protéine était centrée sur 0,85. Dans le cadre de cette thèse, la comparaison quantitative des potentiels électrostatiques des 10 ligands protéiques artificiels avec celui du ligand protéique naturel a permis de montrer que, pour la plupart d'entre eux (8/10), les potentiels des ligands artificiels étaient similaires dans la région d'interaction avec la MMP-14 à celui du ligand naturel TIMP-2. Deux exceptions ont néanmoins été notées puisque les potentiels électrostatiques de deux ligands protéiques artificiels ont montré une absence de corrélation avec celui du ligand naturel. Toutefois, ce phénomène a également été observé dans des familles de complexes protéine-protéine naturels (Magis et al., soumis, Partie B, Chapitre II, II.6). De plus, un exemple a montré que deux protéines NC10 et NC41, capables de lier une cible commune la neuraminidase, ne possédaient pas de similarité de leurs potentiels

électrostatiques dans la région d'interaction avec la cible commune (McCoy et al., 1997).

En résumé, nous avons sélectionné 10 plates-formes permettant le greffage du motif fonctionnel 1-4 du TIMP-2, autorisant un mode de liaison reproduisant les interactions observées avec la cible dans le ligand naturel. Plusieurs paramètres permettant de caractériser la complémentarité de surface, l'aire des interfaces et les caractéristiques électrostatiques dans la région impliquée dans l'interaction ont été déterminés. Des modèles moléculaires décrivant l'interaction présumée de ces 10 ligands artificiels avec la MMP-14 ont été établis et les interactions intermoléculaires analysées. Au total, aucun des critères utilisés n'a conduit à exclure l'un des 10 candidats sélectionnés à l'issue de la recherche de plates-formes dans la PDB grâce au programme STAMPS.

Mimétisme

Les 10 ligands protéiques conçus *in silico* ont donc été produits par voie chimique ou par voie recombinante et évalués expérimentalement en tant qu'inhibiteurs des MMPs. Les résultats obtenus ont montré que les 10 plates-formes sélectionnées conduisaient à des ligands protéiques artificiels présentant des affinités mesurables pour les MMPs testées, comprises entre 450 nM et 590 µM sans optimisation des composés initiaux conçus *in silico*. La plupart des constantes d'inhibition mesurées pour les 6 MMPs se situent entre 1 et 200 µM. Ce résultat montre que la méthode utilisée pour la conception de ces composés conduit bien à des molécules capables de cibler spécifiquement la zone du site catalytique des MMPs. Cependant, les affinités mesurées pour ces 10 ligands artificiels sont significativement plus faibles que celles mesurées pour les inhibiteurs protéiques naturels, les TIMPs dont les Ki sont généralement compris entre 0,1 nM et 200 nM, avec plusieurs valeurs situées autour de 1 nM (Hamze et al., 2007 ; Melendez-Zajgla et al., 2008). Il existe donc un facteur 10^4 à 10^5 entre les constantes d'inhibition des ligands protéiques artificiels et les constantes d'inhibition mesurées pour les TIMPs. Cette observation suggère que, bien que les molécules conçues soient capables de cibler le site catalytique des MMPs, certains paramètres importants pour l'interaction avec la cible sont mal reproduits sur ces plates-formes. Les investigations réalisées par cristallographie ont permis de confirmer, dans le cas de deux analogues de la plate-forme 1edpA en interaction avec le domaine catalytique de la MMP-12, la position du motif fonctionnel dans le site actif de l'enzyme et de déterminer les interactions intermoléculaires reproduites par ces deux composés.

Ces structures ont montré que le motif 1-4 de ces deux variants était capable de reproduire toutes les interactions observées pour le motif 1-4 du TIMP-2 en interaction avec la MMP-14. Toutefois, ces structures ont également montré que des différences dans les caractéristiques géométriques de ces interactions pouvaient exister. Ces observations nous ont incité à réaliser des simulations de dynamique moléculaire dans l'eau afin d'estimer la capacité des différentes plates-formes à reproduire l'ensemble des interactions intermoléculaires. Les résultats de ces simulations indiquent que parmi les 10 ligands protéiques obtenus, plusieurs d'entre eux sont capables de reproduire les interactions observées entre les résidus du motif 1-4 et la cible dans le complexe TIMP-2/MMP-14 (6 ligands protéiques sur 10). Cependant, cette étude a également montré que plusieurs plates-formes protéiques ne permettaient de reproduire qu'une fraction de ces interactions (entre 4 et 6 interactions sur 7 pour 4 des ligands protéiques). De plus, des différences concernant la stabilité de ces interactions sont également observées entre le ligand naturel et certains ligands artificiels. Ces observations pourraient expliquer en partie les affinités modestes observées pour les différents ligands vis-à-vis des 6 MMPs testées. Il faut toutefois noter qu'il n'est pas possible d'établir une corrélation directe entre le nombre d'interactions reproduites par un ligand protéique et la constante d'association mesurée pour celui-ci. En effet, les différents ligands possèdent de nombreuses autres caractéristiques structurales, dynamiques et de composition des interfaces avec la cible qui contribuent aux différences d'affinité observées. Une conclusion importante de ces travaux est que les différences observées concernant la géométrie des interactions intermoléculaires et leur stabilité entre les ligands protéiques artificiels et le ligand naturel devraient être prises en compte dans la sélection de protéines plates-formes. En particulier, des simulations de dynamique moléculaire pourraient être systématiquement réalisées pour évaluer la stabilité des interactions à mimer. Enfin, ces résultats soulignent l'importance du contrôle structural apporté par la plate-forme. En effet, la conformation du motif 1-4 de TIMP-2 est contrôlée par deux ponts disulfures. Les données cristallographiques et de simulation concernant les ligands protéiques 1edpA (1-18) et 2f3cI_m, qui possèdent également deux ponts disulfures impliquant les cystéines 1 et 3, montrent que ces deux plates-formes reproduisent de manière satisfaisante les interactions observées pour ce motif dans le complexe TIMP-2/MMP-14. La reproduction de ces interactions est moins satisfaisante pour les plates-formes possédant une seule des deux cystéines du motif 1-4 et en particulier lorsque celle-ci se trouve en position 3 du motif. Ces observations suggèrent que, conformément au principe de l'approche mise en œuvre pour le greffage de motif, la plate-forme doit apporter le meilleur contrôle structural possible. Dans le

cas présent, seules deux protéines satisfaisant l'ensemble des critères de sélection et possédant des cystéines impliquées dans des ponts disulfures en position 1 et 3 ont été identifiées.

Rendement de la méthode

Comme nous l'avons vu plus haut, les structures secondaires/tertiaires des plates-formes des 10 ligands protéiques conçus sont différentes les unes des autres. La plate-forme 1edpA correspond à l'endothéline 1. Une étude a été reportée montrant que la sarafotoxin 6b, une petite protéine de la famille des endothélines était capable de se lier à plusieurs MMPs (Lauer-Fields et al., 2007). Cette protéine avait été identifiée de manière empirique grâce à la similarité de structure tertiaire de son extrémité N-terminale avec celle des TIMPs. Cependant, dans ces travaux, l'affinité obtenue avec la sarafotoxine native était très faible (> 100 μM pour la MMP-3 et la MMP-14). Des modifications successives de la structure native ont permis d'atteindre des affinités comprises entre 1 et 20 μM pour les MMP-1, MMP-2 et MMP-9. Dans notre étude, l'endothéline native présente une affinité submicromolaire pour la MMP-13 (450 nM) et des affinités comprises entre 1 et 7 μM pour toutes les autres MMPs testées. Des variants de l'endothéline produits dans le cadre de ce travail de thèse ont montré des affinités jusqu'à 290 nM pour la MMP-12. La méthode de conception *in silico* mise en œuvre au cours de ces travaux a permis de concevoir 9 autres protéines dont certaines possèdent des affinités proches de la dizaine de μM et qui présentent des interfaces avec la cible de taille plus importante que l'endothéline (Figure IV.2). Ces nouveaux points de départ pourraient permettre d'obtenir des ligands protéiques affins et sélectifs de certains membres de la famille des MMPs.

Ainsi, l'efficacité de l'approche de conception *in silico* de ligands protéiques développée au sein de notre équipe, mise en évidence dans des travaux antérieurs, a été confirmée par le succès des travaux présentés dans ce mémoire. Ainsi, les 10 ligands protéiques conçus *in silico* et sélectionnés pour une production et une évaluation expérimentales ont montré des affinités mesurables pour les MMPs. Dans un précédent travail de thèse réalisé au sein de notre équipe, l'approche de conception *in silico* de ligands protéiques a été appliquée au développement de bloqueurs du canal potassique Kv1.2 par transfert d'un motif fonctionnel de 4 résidus présents à la surface de la protéine BgK, un bloqueur naturel du canal Kv1.2 (Magis et al., 2006). Cette approche avait permis l'obtention de 3 ligands protéiques artificiels ayant des affinités comprises entre 500 nM et 1,6 μM pour ce canal, sans étapes d'optimisation supplémentaires.

Avec une méthode très similaire à celle utilisée au sein de notre équipe, basée sur des critères topologiques relatifs aux atomes Cα et Cβ des résidus du motif fonctionnel ainsi que des critères stériques, Luhua Lai et ses collaborateurs ont conçu un mime de l'érythropoïétine humaine EPO par transfert des 3 résidus clés essentiels à la liaison à son récepteur EPOR. Dans ce travail, une seule plate-forme (PLCδ1-PH) parmi les 15 sélectionnées, ainsi que des variants de cette plate-forme, ont été produits expérimentalement. Le triple mutant de cette plate-forme portant les 3 résidus clés de la liaison (ERPH1) a montré l'affinité de liaison la plus élevée pour EPOR avec un Kd de 24 nM (Liu et al., 2007). Les résultats issus de ces différents travaux démontrent donc l'intérêt de l'approche *in silico* de conception de ligands protéiques artificiels.

Cependant, les résultats obtenus au cours de ces différentes études montrent qu'il est nécessaire de coupler les approches de conception *in silico* de ligands protéiques avec des méthodes expérimentales de « maturation d'affinité » pour améliorer l'affinité des ligands de 1ère génération. L'objectif de cette optimisation est d'introduire des modifications de certains résidus présents à la surface de la plate-forme permettant la formation de nouvelles interactions intermoléculaires. Il faut noter que, contrairement aux inhibiteurs synthétiques de faible poids moléculaire, les mini-protéines sélectionnées dans ce travail sont caractérisées par une interface importante avec la cible. Ces plates-formes offrent la possibilité d'introduire de la variabilité au niveau de leur surface d'interaction avec la cible afin d'améliorer leur affinité et/ou leur sélectivité vis-à-vis des MMPs. Toutefois, les quelques analogues des composés 1edpA (1-18) et 1an1I_m réalisés au cours de ce travail, sur un nombre restreint de positions, n'ont pas permis d'obtenir de gain d'affinité important. Les constantes d'inhibition ont été améliorées d'un facteur 25 pour la plate-forme 1edpA (1-18) et d'un facteur 4 au mieux pour la plate-forme 1an1I_m. Ces résultats montrent que l'optimisation doit être envisagée de manière plus large à l'aide de méthodes capables de produire une importante diversité de variants dans un temps court. La création de bibliothèques de ligands protéiques et la mise en place d'une stratégie de criblage de ces bibliothèques par « phage display » ou « yeast display » peut permettre d'atteindre cet objectif. Des telles stratégies de criblage pourraient également être appliquées en vue d'améliorer la sélectivité des ligands envers un membre particulier de la famille des MMPs. Cette approche a permis de développer des variants de TIMP-2 capables d'inhiber sélectivement et à haute affinité la MMP-1 avec un facteur de sélectivité de 10^4 par rapport à la MMP-3 et MMP-14 (Bahudhanapati et al., 2011). L'exploitation des 10 ligands protéiques conçus au cours de cette thèse à l'aide d'une stratégie de criblage par « phage

display » ou « yeast display » pourrait donc permettre d'obtenir des mini-protéines sélectives de certains membres de la famille des MMPs.

ANNEXES

Annexe I : Résultats de la recherche par STAMPS

Tableau I : Résultats des plates-formes retournées par STAMPS pour les deux recherches S_Cys 1 et S_Cys 3 réalisées sur la version entière de la PDB. Ce tableau indique, pour chaque plate-forme, son code PDB et l'identité de la chaine (Pdbtitle), sa taille (size), le RMSD calculé sur les atomes Cα et Cβ du motif, le facteur stérique intermoléculaire (S_inter) et intramoléculaire (S_intra) ainsi que le nom et le numéro des résidus de la plate-forme équivalents au motif recherché.

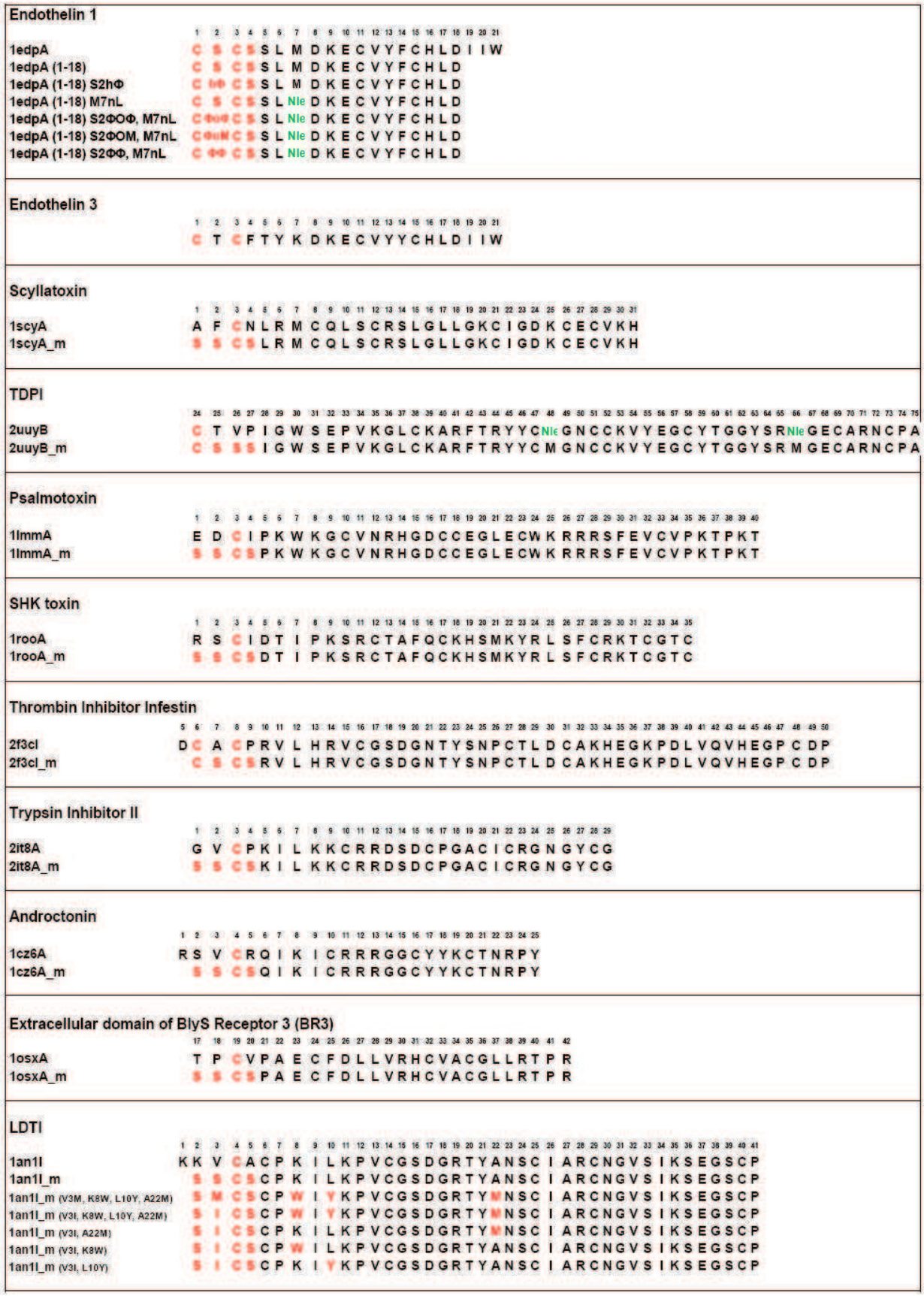
pdbtitle	Size	RMSD	orientation	S_inter	S_intra	Res 1		Res 2		Res 3		Res 4	
S_Cys1													
pdb2obqD_00	16	0.28	14.63	1.06	0.44	CYS	22	VAL	23	VAL	24	ILE	25
pdb2uuyB_00	52	0.36	9.71	1.05	0.25	CYS	24	THR	25	VAL	26	PRO	27
pdb2oslP_00	20	0.47	31.41	2.12	0.34	CYS	167	GLU	168	PRO	169	ALA	170
pdb2f3cI_00	46	0.48	16.76	1.18	0.30	CYS	6	ALA	7	CYS	8	PRO	9
pdb2rquB_00	19	0.56	39.94	2.42	0.50	CYS	1578	ILE	1579	ILE	1580	SER	1581
pdb1edpA_00	17	0.61	21.39	2.12	0.11	CYS	1	SER	2	CYS	3	SER	4
pdb1b9pA_00	34	0.68	39.27	1.43	0.15	CYS	1	ALA	2	VAL	3	GLU	4
pdb1fgeA_00	20	0.69	21.86	0.84	0.06	CYS	407	GLU	408	ALA	409	PRO	410
pdb1choI_00	53	0.74	43.42	2.10	0.22	CYS	8	ASP	7	VAL	6	SER	5
pdb3o6qB_00	43	0.80	37.94	0.31	0.27	CYS	9	GLU	10	PRO	11	ARG	12
pdb1paoA_00	17	0.82	33.33	1.18	0.38	CYS	129	LYS	130	SER	131	THR	132
pdb1kgmA_00	35	0.83	54.65	2.37	0.26	CYS	4	THR	3	VAL	2	GLU	1
pdb1r48A_00	33	0.83	37.42	2.26	0.13	CYS	1	GLY	2	ASN	5	ASP	4
pdb1srbA_00	21	0.83	35.81	2.22	0.06	CYS	1	SER	2	CYS	3	LYS	4
pdb1m4eA_00	20	0.84	47.97	2.45	0.21	CYS	5	ILE	3	CYS	2	ILE	1
pdb3o6qD_00	44	0.88	47.01	0.28	0.24	CYS	9	GLU	10	PRO	11	ARG	12
pdb1y1bA_00	48	1.02	53.13	1.35	0.23	CYS	4	ASP	3	PRO	2	LYS	1
pdb1hx2A_00	60	1.06	54.08	2.33	0.16	CYS	4	VAL	3	PHE	2	ASN	1
pdb1v6rA_00	21	1.08	37.79	1.59	0.11	CYS	1	SER	2	CYS	3	SER	4
pdb1axhA_00	37	1.10	74.93	2.34	0.19	CYS	4	THR	3	PRO	2	SER	1
pdb1kioA_00	35	1.21	51.98	1.87	0.21	CYS	4	THR	3	VAL	2	GLU	1
S_Cys3													
pdb2kggA_00	52	0.27	17.80	2.44	0.13	SER	1	VAL	2	CYS	3	ALA	4
pdb1ewsA_00	32	0.29	15.48	2.32	0.18	MET	1	PRO	2	CYS	3	SER	4
pdb2l61A_00	26	0.35	21.35	2.11	0.29	SER	2	GLY	3	CYS	4	ASP	5
pdb2v1vA_00	29	0.37	22.14	1.20	0.16	ARG	1	VAL	2	CYS	3	PRO	4
pdb2l3iA_00	30	0.38	25.82	1.77	0.02	ILE	2	ARG	3	CYS	4	PRO	5
pdb1nqlB_00	48	0.42	26.03	2.07	0.19	SER	4	GLU	5	CYS	6	PRO	7
pdb1y2pA_00	34	0.43	15.21	1.94	0.27	ALA	1	SER	2	CYS	3	ARG	4
pdb2fk3A_00	59	0.43	26.61	0.95	0.16	GLU	131	ALA	132	CYS	133	LYS	134
pdb1g1pA_00	31	0.45	22.11	2.30	0.41	ASP	1	ASP	2	CYS	3	ILE	4
pdb1hx2A_00	60	0.46	21.88	1.48	0.15	PHE	2	VAL	3	CYS	4	PRO	5
pdb2ikeA_00	54	0.47	18.79	1.97	0.27	LEU	71	SER	72	CYS	73	LEU	74
pdb2f3cI_01	46	0.48	16.76	1.18	0.30	CYS	6	ALA	7	CYS	8	PRO	9
pdb2juyA_00	27	0.48	28.76	1.83	0.36	PHE	1	PHE	2	CYS	3	PRO	4
pdb1dkcA_00	38	0.49	29.59	0.98	0.17	ALA	1	GLY	2	CYS	3	ILE	4
pdb1ie6A_00	33	0.50	20.40	1.80	0.14	GLY	1	ASP	2	CYS	3	LEU	4
pdb2w9uA_00	46	0.52	24.45	2.26	0.23	MET	2	ASP	3	CYS	4	THR	5
pdb1zntA_00	28	0.53	31.47	2.42	0.33	GLY	2	GLU	3	CYS	4	VAL	5
pdb1cz6A_00	25	0.54	17.80	1.76	0.17	SER	2	VAL	3	CYS	4	ARG	5
pdb2g81I_00	55	0.54	13.63	2.15	0.34	PRO	17	CYS	18	CYS	19	ASP	20
pdb1tleA_00	58	0.55	13.47	1.24	0.18	ARG	1	PRO	2	CYS	3	GLN	4
pdb1cixA_00	44	0.61	30.47	2.21	0.22	SER	2	ARG	3	CYS	4	GLN	5
pdb1edpA_00	17	0.61	21.39	2.12	0.11	CYS	1	SER	2	CYS	3	SER	4
pdb1gpsA_00	47	0.62	28.24	2.45	0.38	LYS	1	ILE	2	CYS	3	ARG	4
pdb2gwpA_00	32	0.62	38.35	2.48	0.21	LEU	2	LEU	3	CYS	4	TYR	5
pdb1ldtL_00	46	0.64	20.24	2.49	0.80	LYS	2I	VAL	3I	CYS	4I	ALA	5I
pdb2dcvA_00	42	0.65	36.87	1.80	0.14	VAL	2	SER	3	CYS	4	LEU	5
pdb1hymA_01	44	0.69	42.31	1.60	0.28	GLY	5	PRO	4	CYS	3	SER	2
pdb1q2kA_00	31	0.69	28.46	1.47	0.15	ALA	1	ALA	2	CYS	3	TYR	4

pdb2dcwA_00	42	0.69	25.76	1.85	0.12	ILE	2	THR	3	CYS	4	LEU	5
pdb1wo0A_00	17	0.71	31.87	1.63	0.16	LYS	1	TRP	2	CYS	3	PHE	4
pdb2it8A_00	29	0.71	16.40	1.17	0.30	GLY	1	VAL	2	CYS	3	PRO	4
pdb1oqeO_00	31	0.72	31.44	1.04	0.26	THR	2	PRO	3	CYS	4	VAL	5
pdb1mctI_00	28	0.73	28.41	2.47	0.26	ARG	1	ILE	2	CYS	3	PRO	4
pdb2b5kA_00	19	0.74	51.52	1.53	0.22	ARG	2	TRP	3	CYS	4	PHE	5
pdb1wxnA_00	42	0.75	25.63	2.34	0.30	THR	2	ALA	3	CYS	4	SER	5
pdb1c9tI_00	53	0.76	46.15	1.67	0.22	THR	8	PRO	9	CYS	10	GLY	11
pdb1c9tK_00	53	0.77	45.88	1.64	0.26	THR	8	PRO	9	CYS	10	GLY	11
pdb1c9tL_00	53	0.79	47.19	1.57	0.27	THR	8	PRO	9	CYS	10	GLY	11
pdb1hymA_00	44	0.79	23.38	1.44	0.28	SER	1	SER	2	CYS	3	PRO	4
pdb1ibiA_00	59	0.79	28.88	2.42	0.13	GLU	118	LYS	119	CYS	120	SER	121
pdb1znfA_01	25	0.79	46.35	2.25	0.05	LEU	5	GLY	4	CYS	3	LYS	2
pdb3ccsU_00	53	0.79	30.24	2.49	0.08	ARG	4	GLU	5	CYS	6	ASP	7
pdb1c9tH_00	53	0.80	46.91	1.69	0.24	THR	8	PRO	9	CYS	10	GLY	11
pdb1oqeP_00	31	0.80	30.76	0.87	0.27	THR	2	PRO	3	CYS	4	VAL	5
pdb1pb5A_00	35	0.80	36.18	1.57	0.17	GLU	2	ALA	3	CYS	4	GLU	5
pdb1bj6A_00	42	0.81	26.93	2.40	0.11	VAL	13	LYS	14	CYS	15	PHE	16
pdb1jy4B_00	32	0.82	33.93	2.45	0.03	GLY	2	GLU	3	CYS	4	LYS	5
pdb2ctiA_00	29	0.82	61.30	2.40	0.18	ARG	5	PRO	4	CYS	3	VAL	2
pdb1p0t0_00	25	0.83	35.49	2.16	0.31	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0t5_00	25	0.83	35.57	2.12	0.29	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0t7_00	25	0.83	35.76	2.00	0.28	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tb_00	25	0.83	35.79	1.99	0.30	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tB_00	25	0.83	35.89	1.99	0.31	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tE_00	25	0.83	35.78	2.00	0.28	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tF_00	25	0.83	35.79	2.03	0.30	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tH_00	25	0.83	35.74	2.01	0.28	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0th_00	25	0.83	35.75	1.99	0.34	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tL_00	25	0.83	35.46	2.10	0.30	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tl_00	25	0.83	35.75	1.99	0.28	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tP_00	25	0.83	35.88	1.97	0.28	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tT_00	25	0.83	35.94	1.99	0.30	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tt_00	25	0.83	35.87	1.98	0.32	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tv_00	25	0.83	35.78	1.97	0.28	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tx_00	25	0.83	35.56	2.14	0.31	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tZ_00	25	0.83	35.88	1.99	0.28	THR	17	PRO	18	CYS	19	VAL	20
pdb1p8bA_00	37	0.83	41.22	2.41	0.34	ALA	1	SER	2	CYS	3	ASN	4
pdb1srbA_00	21	0.83	35.81	2.22	0.06	CYS	1	SER	2	CYS	3	LYS	4
pdb2o8mD_00	18	0.83	39.17	0.95	0.41	LYS	220	GLY	221	CYS	222	VAL	223
pdb1ejaB_00	53	0.84	48.53	2.20	0.23	THR	8	PRO	9	CYS	10	GLY	11
pdb1m4eA_00	20	0.84	47.97	2.45	0.21	CYS	5	ILE	3	CYS	2	ILE	1
pdb1p0t1_00	25	0.84	35.82	2.01	0.30	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0t2_00	25	0.84	35.84	2.01	0.31	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0t3_00	25	0.84	35.69	2.02	0.28	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0t4_00	25	0.84	35.69	2.02	0.31	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0t6_00	25	0.84	35.60	2.00	0.28	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0t8_00	25	0.84	35.74	2.01	0.30	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0t9_00	25	0.84	35.97	1.96	0.29	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tA_00	25	0.84	35.80	2.01	0.29	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0ta_00	25	0.84	35.70	2.01	0.28	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tc_00	25	0.84	35.89	2.01	0.30	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tC_00	25	0.84	35.91	1.95	0.30	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0td_00	25	0.84	35.79	1.99	0.30	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0te_00	25	0.84	35.80	2.01	0.30	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tf_00	25	0.84	35.95	2.01	0.31	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tG_00	25	0.84	35.92	2.01	0.30	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0ti_00	25	0.84	35.70	2.01	0.28	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tI_00	25	0.84	36.00	2.02	0.31	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tj_00	25	0.84	35.73	2.01	0.28	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tJ_00	25	0.84	35.91	2.00	0.30	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tK_00	25	0.84	35.91	2.01	0.31	THR	17	PRO	18	CYS	19	VAL	20

pdb1p0tk_00	25	0.84	35.80	2.01	0.31	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tM_00	25	0.84	35.89	2.00	0.28	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tm_00	25	0.84	35.88	2.01	0.28	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tN_00	25	0.84	35.83	1.99	0.28	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tn_00	25	0.84	35.88	2.01	0.32	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tO_00	25	0.84	35.81	2.01	0.30	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0to_00	25	0.84	35.81	2.03	0.31	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tp_00	25	0.84	35.76	1.99	0.31	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tQ_00	25	0.84	35.75	1.99	0.30	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tq_00	25	0.84	35.69	1.99	0.26	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tR_00	25	0.84	35.83	2.01	0.28	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tr_00	25	0.84	35.85	2.01	0.28	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tS_00	25	0.84	35.86	2.01	0.31	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0ts_00	25	0.84	35.66	2.17	0.28	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tU_00	25	0.84	35.81	2.01	0.30	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tu_00	25	0.84	35.89	2.01	0.30	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tV_00	25	0.84	36.00	1.99	0.30	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tW_00	25	0.84	35.80	2.00	0.31	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tw_00	25	0.84	35.81	1.99	0.27	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tX_00	25	0.84	35.90	1.99	0.31	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tD_00	25	0.85	35.85	2.00	0.30	THR	17	PRO	18	CYS	19	VAL	20
pdb1p0tg_00	25	0.85	35.82	1.87	0.29	THR	17	PRO	18	CYS	19	VAL	20
pdb2w9uA_01	46	0.85	37.05	2.01	0.25	ALA	1	ASP	3	CYS	4	THR	5
pdb1c9tG_00	53	0.86	47.24	1.62	0.27	THR	8	PRO	9	CYS	10	GLY	11
pdb1c9tJ_00	53	0.86	46.99	1.94	0.26	THR	8	PRO	9	CYS	10	GLY	11
pdb1oqeQ_00	31	0.86	30.64	0.95	0.24	THR	2	PRO	3	CYS	4	VAL	5
pdb1wt7A_00	41	0.86	51.84	2.24	0.33	THR	4	SER	3	CYS	2	TRP	1
pdb2fk3F_00	59	0.86	43.90	2.18	0.16	GLU	131	ALA	132	CYS	133	LYS	134
pdb2nz3A_00	40	0.87	33.00	1.97	0.17	ALA	1	THR	2	CYS	3	ASP	4
pdb1oqeR_00	31	0.88	31.36	0.93	0.24	THR	2	PRO	3	CYS	4	VAL	5
pdb1p0tY_00	25	0.88	34.85	1.91	0.29	THR	17	PRO	18	CYS	19	VAL	20
pdb1g1zA_00	31	0.89	30.41	1.21	0.34	ASP	1	ASP	2	CYS	3	ILE	4
pdb1oqeK_00	31	0.89	29.22	0.90	0.24	THR	2	PRO	3	CYS	4	VAL	5
pdb1oqeL_00	31	0.89	30.02	2.40	0.26	THR	2	PRO	3	CYS	4	VAL	5
pdb1wpdA_00	34	0.89	35.66	2.02	0.30	VAL	1	SER	2	CYS	3	GLY	5
pdb1zfuA_00	40	0.89	39.39	1.85	0.31	PHE	2	GLY	1	CYS	4	ASN	5
pdb1kweA_00	16	0.90	59.42	2.18	0.32	ILE	175	SER	174	CYS	173	PRO	172
pdb1ncpN_00	17	0.90	55.87	2.25	0.07	ASN	5	PHE	4	CYS	3	LYS	2
pdb1txmA_00	34	0.90	68.93	2.33	0.22	GLY	5	THR	4	CYS	3	SER	2
pdb1qjlA_00	28	0.91	40.98	1.17	0.09	ASN	40	THR	39	CYS	38	ILE	37
pdb1beiA_00	34	0.92	57.69	1.48	0.13	ASP	5	ILE	4	CYS	3	SER	2
pdb1znfA_00	25	0.93	29.49	2.36	0.05	TYR	1	LYS	2	CYS	3	GLY	4
pdb2a05A_01	57	0.93	47.01	0.91	0.18	GLY	187	SER	188	CYS	189	ALA	190
pdb2jniA_00	21	0.93	28.85	1.84	0.09	ARG	1	TRP	2	CYS	3	VAL	4
pdb2qa4U_00	53	0.93	29.58	2.46	0.08	ARG	4	GLU	5	CYS	6	ASP	7
pdb1jlzA_00	23	0.94	76.24	2.28	0.22	SER	4	GLY	3	CYS	2	ALA	1
pdb1f8zA_00	39	0.96	44.88	1.04	0.14	ALA	1	THR	2	CYS	3	ARG	4
pdb1gl1I_00	34	0.97	26.90	1.87	0.34	ILE	2	SER	3	CYS	4	GLU	5
pdb2w9vA_00	44	0.97	43.69	1.56	0.34	MET	2	ASP	3	CYS	4	THR	5
pdb1aggA_00	48	0.97	67.65	2.28	0.34	ASP	2	GLU	1	CYS	4	ILE	5
pdb1beiA_01	34	0.99	36.46	1.31	0.13	ARG	1	SER	2	CYS	3	ILE	4
pdb1n4nA_00	47	0.99	47.49	2.35	0.19	ALA	1	THR	2	CYS	3	LYS	4
pdb2kmrA_00	44	0.99	25.83	1.85	0.31	LYS	2	VAL	3	CYS	4	ALA	5
pdb1oqeM_00	31	1.00	32.26	1.06	0.25	THR	2	PRO	3	CYS	4	VAL	5
pdb1ji9A_00	37	1.01	40.31	2.36	0.21	LYS	32	SER	33	CYS	35	SER	36
pdb1emxA_00	30	1.02	50.37	2.04	0.17	ASP	1	ASP	2	CYS	3	GLY	4
pdb2kb8A_00	37	1.02	66.12	2.35	0.12	THR	4	ASN	3	CYS	2	LYS	1
pdb1v6rA_01	21	1.03	62.57	1.27	0.15	SER	5	SER	4	CYS	3	SER	2
pdb1wj0A_00	58	1.03	32.85	1.84	0.04	ILE	125	CYS	126	CYS	127	GLN	128
pdb1zpxA_00	21	1.03	70.67	0.83	0.38	SER	16	GLY	15	CYS	14	PRO	13
pdb3g4sU_00	53	1.03	35.94	2.39	0.25	ARG	4	GLU	5	CYS	6	ASP	7
pdb1k81A_00	36	1.04	56.33	1.86	0.17	VAL	108	ILE	109	CYS	110	ARG	111

pdb1w7zG_00	30	1.04	65.79	1.92	0.27	ARG	4	PRO	3	CYS	2	GLY	1
pdb2e3gA_00	40	1.04	51.71	1.94	0.16	LEU	5	ASP	4	CYS	3	THR	2
pdb2fk3C_00	59	1.04	26.86	0.98	0.17	GLU	131	ALA	132	CYS	133	LYS	134
pdb2ldfA_00	24	1.04	79.32	2.04	0.13	ASP	5	ASN	4	CYS	3	CYS	1
pdb1an1I_00	40	1.05	47.38	2.27	0.28	LYS	2	VAL	3	CYS	4	ALA	5
pdb2a05A_00	57	1.05	51.29	1.62	0.17	SER	191	ALA	190	CYS	189	GLY	187
pdb2bayD_00	56	1.05	30.99	2.48	0.28	MET	1	LEU	2	CYS	3	ALA	4
pdb1p1pA_00	24	1.06	70.02	1.96	0.34	SER	5	GLY	4	CYS	3	CYS	2
pdb2ldeA_00	25	1.06	68.80	1.43	0.10	ASP	5	THR	4	CYS	3	SER	2
pdb1mxpA_00	15	1.07	59.27	1.39	0.20	TYR	5	SER	4	CYS	3	CYS	2
pdb1v6rA_00	21	1.08	37.79	1.59	0.11	CYS	1	SER	2	CYS	3	SER	4
pdb2f3cI_00	46	1.09	35.18	0.77	0.31	ASP	5	ALA	7	CYS	8	PRO	9
pdb2uuxA_00	55	1.09	32.13	1.72	0.27	ALA	22	GLU	23	CYS	24	THR	25
pdb1uybA_00	16	1.10	48.41	1.89	0.17	ASP	2	ASP	3	CYS	4	GLU	5
pdb1y1bA_00	48	1.11	42.94	1.40	0.23	PRO	2	ASP	3	CYS	4	PRO	5
pdb2l62A_01	26	1.11	39.21	2.44	0.71	SER	2	GLY	1	CYS	4	ASP	5
pdb1mxpA_01	15	1.12	70.43	1.18	0.32	TYR	5	SER	4	CYS	3	GLY	1
pdb1znmA_00	26	1.12	53.91	1.21	0.14	PHE	6	THR	5	CYS	4	GLN	3
pdb3njpC_00	47	1.12	64.94	2.44	0.15	LEU	8	PRO	7	CYS	6	GLU	5
pdb1c49A_00	35	1.13	43.71	2.29	0.36	ILE	5	SER	6	CYS	7	THR	8
pdb1lmmA_00	40	1.13	53.52	1.97	0.24	GLU	1	ASP	2	CYS	3	ILE	4
pdb1osxA_00	26	1.13	67.41	1.77	0.18	PRO	21	VAL	20	CYS	19	PRO	18
pdb1osxA_01	26	1.13	51.76	1.21	0.23	THR	17	PRO	18	CYS	19	VAL	20
pdb2fk3G_00	59	1.13	31.38	0.93	0.17	GLU	131	ALA	132	CYS	133	LYS	134
pdb2o8mC_00	23	1.13	49.19	1.14	0.38	LYS	220	GLY	221	CYS	222	VAL	223
pdb1scyA_00	31	1.14	48.18	0.49	0.16	ALA	1	PHE	2	CYS	3	ASN	4
pdb2g9jC_00	37	1.14	74.23	1.96	0.01	SER	252	GLY	250	CYS	249	GLY	248
pdb2l62A_00	26	1.14	38.94	2.39	0.46	SER	2	GLY	3	CYS	4	ASP	5
pdb1tcjA_00	19	1.15	56.29	2.48	0.21	ARG	1	ASP	2	CYS	3	CYS	4
pdb2oqjC_00	18	1.15	67.81	1.76	0.04	PRO	4	PRO	3	CYS	2	ALA	1
pdb3mywI_00	55	1.15	66.98	2.34	0.33	SER	15	ASP	14	CYS	13	CYS	12
pdb2bayB_00	56	1.16	38.57	2.27	0.25	MET	1	LEU	2	CYS	3	ALA	4
pdb2bayC_00	55	1.16	35.07	2.35	0.30	MET	1	LEU	2	CYS	3	ALA	4
pdb1tckA_00	19	1.17	62.76	2.25	0.18	ARG	1	ASP	2	CYS	3	CYS	4
pdb2oqjI_00	17	1.17	70.98	1.59	0.05	PRO	4	PRO	3	CYS	2	ALA	1
pdb1xu2S_00	37	1.18	68.08	0.90	0.33	GLN	10	SER	9	CYS	8	GLN	7
pdb2v1vA_01	29	1.18	81.76	1.92	0.16	ARG	5	PRO	4	CYS	3	VAL	2
pdb1ajjA_00	37	1.19	71.74	2.33	0.24	ALA	7	SER	6	CYS	5	PRO	4
pdb1pmcA_00	36	1.19	30.65	2.27	0.20	ILE	2	SER	3	CYS	4	GLU	5
pdb1terA_00	21	1.19	53.46	0.93	0.14	ALA	1	LEU	2	CYS	3	ASN	4
pdb2it7A_00	28	1.19	68.54	2.34	0.38	ARG	4	PRO	3	CYS	2	GLY	1
pdb1wm8A_00	28	1.20	62.10	1.16	0.31	VAL	1	GLY	2	CYS	3	GLU	4
pdb1rooA_00	35	1.21	51.94	0.80	0.13	ARG	1	SER	2	CYS	3	ILE	4
pdb2kmrA_01	44	1.21	55.02	1.36	0.41	LYS	1	VAL	3	CYS	4	ALA	5
pdb1panA_00	17	1.22	73.99	1.03	0.34	SER	131	LYS	130	CYS	129	ALA	128
pdb2oqjL_00	19	1.25	68.77	1.91	0.05	PRO	4	PRO	3	CYS	2	ALA	1

Annexe II : Séquences des plates-formes testées



Annexe III : Synthèse peptidique

Principe :

Les peptides sont synthétisés en phase solide sur des résines. Le protocole utilisé est celui de la stratégie Fmoc, groupement protecteur de la fonction α-amine des acides aminés. Les synthèses sont effectuées de manière automatisée sur un synthétiseur de peptides (Applied Biosystems 433A). La synthèse se fait de l'acide aminé C-terminal vers l'acide aminé N-terminal, l'acide aminé C-terminal étant attaché au support solide. Le support solide dans les résines est constitué par des billes de polystyrène réticulé par du divinylbenzène. Deux types de résines ont été utilisées : des résines de type Rink amide, non préchargées, et une résine de type NovaSyn TGA préchargée.

Protocole :

La résine est préalablement immergée dans le NMP pendant 45min afin de la faire gonfler et rendre ainsi tous ses sites accessibles, la séquence est ensuite construite selon le cycle suivant :

1) Déprotection du groupement Fmoc : étape de déprotection du groupement Fmoc porté par le dernier acide aminé fixé au peptide lié à la résine; cette déprotection est réalisée par une amine secondaire, la pipéridine.

Lors de la déprotection du groupement Fmoc, un suivi de l'absorbance à 305 nm est réalisé. En effet, l'adduit fulvène-pipéridine libéré lors de cette déprotection absorbe à cette longueur d'onde et sa quantification permet d'estimer l'efficacité du couplage du dernier acide aminé incorporé mais surtout de vérifier si la réaction de déprotection est difficile, ce qui pourrait entrainer un couplage difficile de l'acide aminé suivant. Les informations fournies par ce monitorage permettent d'adapter les temps de couplage et de déprotection aux difficultés rencontrées.

2) Lavage : l'excès de pipéridine est enlevé par lavage à la NMP une fois la déprotection terminée (après environ 5 à 10 min). Le peptide lié à la résine est alors prêt pour l'étape suivante qui correspond au couplage.

3) Pré-activation : l'acide aminé suivant est couplé par formation d'une liaison peptidique au niveau de la fonction N-terminale déprotégée du peptide en voie de synthèse. Afin de permettre le couplage, le groupe carboxylique de l'acide aminé à incorporer doit être activé.

Lors de cette étape, 1 mmol de l'acide aminé protégé au niveau de sa fonction α-aminée est dissoute dans 1 ml de NMP et 1 ml d'une solution de HOBt 1 M dans la NMP. Puis 1 ml d'une solution de DCCI 1 M dans la NMP est ajoutée.

4) Couplage : Après 40 à 50 min d'activation, l'ester actif est transféré dans le réacteur qui contient la résine, tandis que le précipité de dicyclohexylurée (DCU) reste dans la chambre d'activation. C'est à ce moment que l'étape de couplage proprement dite a lieu, c'est à dire que l'acide aminé Fmoc activé réagit avec la fonction N-terminale du peptide en cours de synthèse par formation d'une liaison peptidique. Durant cette étape il est crucial d'obtenir l'efficacité de couplage la plus importante possible afin d'aboutir à un rendement final intéressant.

5) Capping : l'efficacité de couplage n'atteignant jamais les 100%, il est nécessaire de faire suivre cette étape de couplage par une étape de capping qui va permettre d'acétyler les fonctions amines libres (fonctions N-terminales des chaînes qui n'ont pas réagi lors de l'étape de couplage précédente). Cette étape permet d'éviter l'apparition de peptides comportant des délétions qui seraient difficiles à séparer du peptide entier lors des étapes de purification ultérieures. Cette réaction s'effectue dans solution de capping contenant un mélange Ac_2O/ Collidine / NMP (50 / 60 / 890, v/v/v).

6) Lavage : après lavage à la NMP, la résine est prête pour de nouveaux cycles de déprotection/couplage.

Une fois le peptide synthétisé, il doit être clivé de la résine et les groupements protecteurs présents sur les chaines latérales doivent être enlevés. Ces deux opérations sont effectuées simultanément par traitement du peptide lié à la résine par un acide faible : le TFA. Le clivage se fait également en présence de « scavengers » tels que l'eau et le TIS, qui sont des molécules capables de piéger les cations réactifs formés par certains groupements protecteurs des chaines latérales. Le mélange de clivage utilisé est donc composé de 10 mL de TFA/TIS/H2O dans les proportions 95/2,5/2,5 (v/v/v).

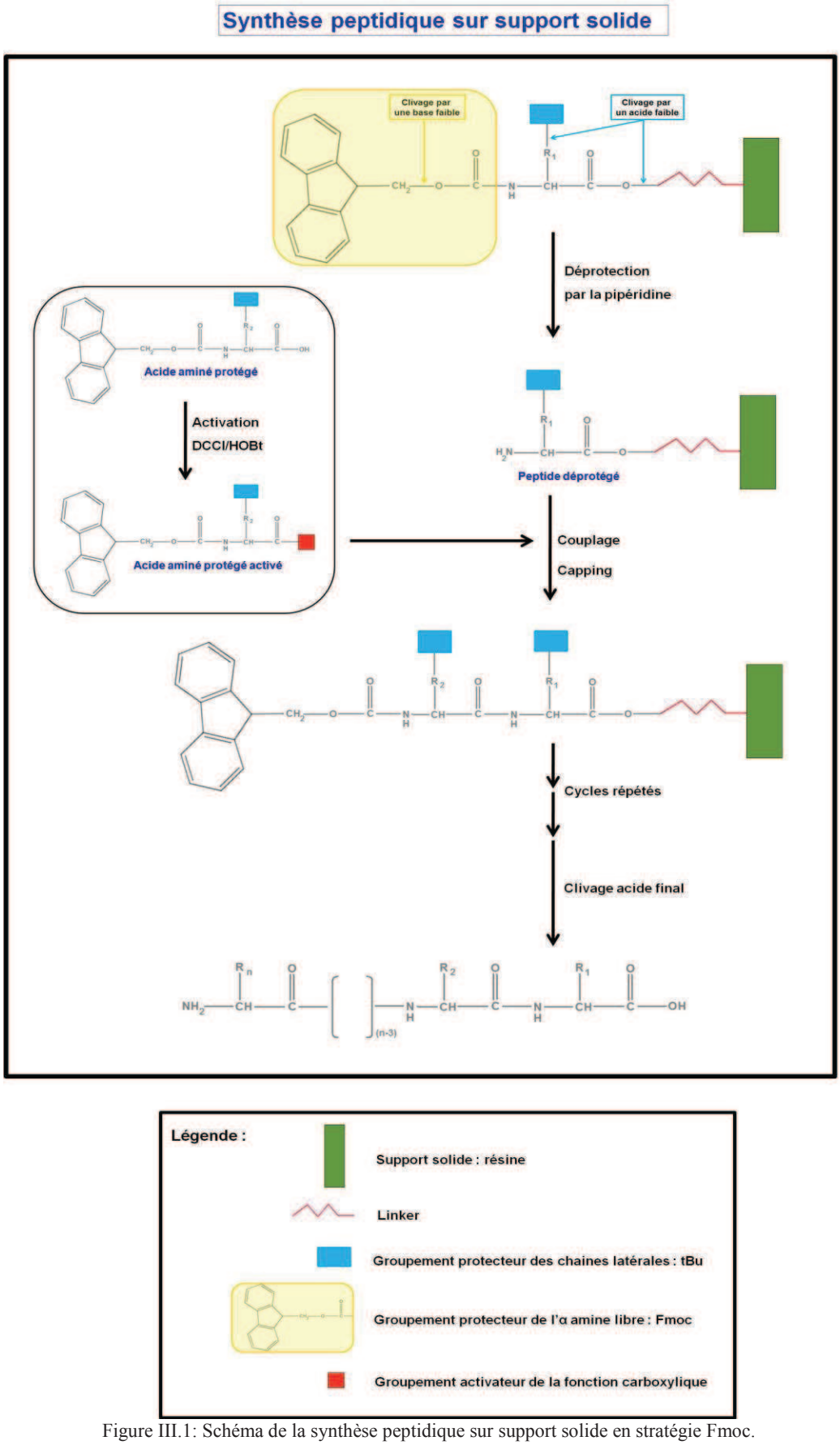


Figure III.1: Schéma de la synthèse peptidique sur support solide en stratégie Fmoc.

ARTICLES PUBLIÉS

the FEBS Journal

Grafting of functional motifs onto protein scaffolds identified by PDB screening – an efficient route to design optimizable protein binders

Rym Tlatli, Hervé Nozach, Guillaume Collet, Fabrice Beau, Laura Vera, Enrico Stura, Vincent Dive and Philippe Cuniasse

Service d'Ingénierie Moléculaire des Protéines, Institut de Biologie et Technologies de Saclay (IBITEC-S), Commissariat à l'Energie Atomique, Gif-sur-Yvette, France

Keywords
functional motif; grafting; MMP inhibitor; protein binder design; protein–protein interactions

Correspondence
P. Cuniasse, Service d'Ingénierie Moléculaire des Protéines, Institut de Biologie et Technologies de Saclay (IBITEC-S), Commissariat à l'Energie Atomique, F-91191 Gif-sur-Yvette, France
Fax: +33 1 69 08 90 71
Tel: +33 1 69 08 56 35
E-mail: philippe.cuniasse@cea.fr

(Received 21 August 2012, revised 19 October 2012, accepted 30 October 2012)

doi:10.1111/febs.12056

Artificial miniproteins that are able to target catalytic sites of matrix metalloproteinases (MMPs) were designed using a functional motif-grafting approach. The motif corresponded to the four N-terminal residues of TIMP-2, a broad-spectrum protein inhibitor of MMPs. Scaffolds that are able to reproduce the functional topology of this motif were obtained by exhaustive screening of the Protein Data Bank (PDB) using STAMPS software (search for three-dimensional atom motifs in protein structures). Ten artificial protein binders were produced. The designed proteins bind catalytic sites of MMPs with affinities ranging from 450 nM to 450 μM prior to optimization. The crystal structure of one artificial binder in complex with the catalytic domain of MMP-12 showed that the inter-molecular interactions established by the functional motif in the artificial binder corresponded to those found in the MMP-14–TIMP-2 complex, albeit with some differences in geometry. Molecular dynamics simulations of the ten binders in complex with MMP-14 suggested that these scaffolds may allow partial reproduction of native inter-molecular interactions, but differences in geometry and stability may contribute to the lower affinity of the artificial protein binders compared to the natural protein binder. Nevertheless, these results show that the *in silico* design method used provides sets of protein binders that target a specific binding site with a good rate of success. This approach may constitute the first step of an efficient hybrid computational/experimental approach to protein binder design.

Introduction

In recent years, considerable efforts have been made to design alternatives to antibodies for development of therapeutic or diagnostic compounds. Approaches based on production and screening of large libraries involving randomization of restricted regions in moderate size protein scaffolds with a stable core were proposed. Impressive successes were obtained by these approaches [1], and some of these proteins are currently being investigated for therapeutic applications [2], demonstrating the potential of moderate-size proteins in this context. However, these approaches do not offer precise control of the binding location of the protein binder at the surface of the target or its binding orientation. This may be an important drawback in the context of drug development, where specific interactions are mandatory.

Abbreviations
MMP, matrix metalloproteinase; rmsd, root mean square deviation; TIMP, tissue inhibitor of metalloproteinases.

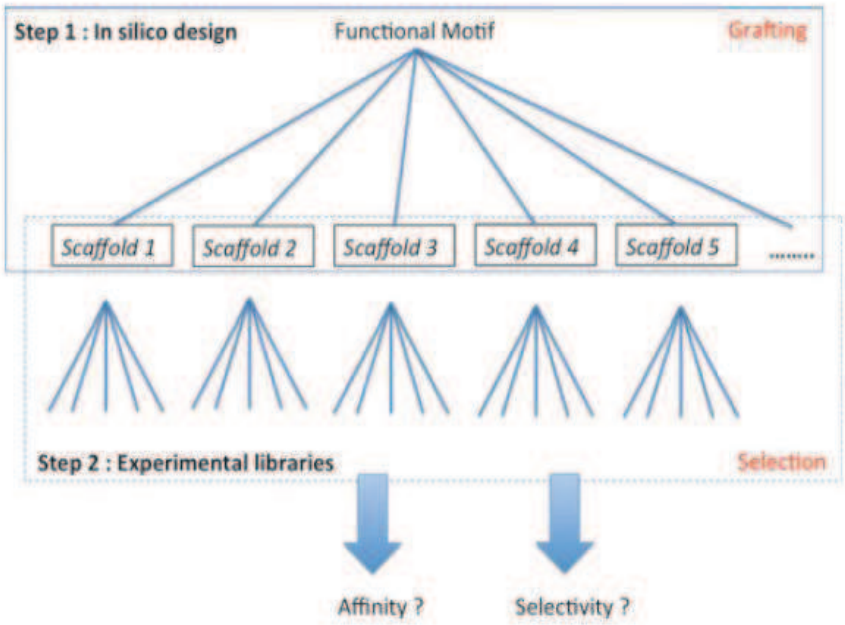
 © 2012 The Authors Journal compilation © 2012 FEBS

In silico functional motif grafting methods have been proposed as an alternative to generic scaffolds for structure-based design of artificial protein binders. These approaches are based on the observation that a limited set of residues, referred to as hot spots, located at the surface of a protein binder, contribute predominantly to the binding free energy [3]. Artificial protein binders may be obtained by grafting such hot spots on protein scaffolds that are able to reproduce the functional topology of these residues. In contrast to library-based approaches, grafting methods allow the design of protein binders that interact at a specific location because they are based on reproduction of precise interactions. In early studies, the grafted motifs were restricted to residues belonging to a single structural element that were reproduced on a different protein [4–9]. Although important successes were obtained using this principle, its potential is clearly limited by the availability of scaffolds. In addition, functional epitopes are often discontinuous, and the residues that constitute the motif may be located in separated structural elements. To overcome these limitations, the backbone grafting approach was recently proposed [10,11]. In this method, linear epitopes are transplanted onto a protein scaffold by replacing portions of native backbone by the linear segments to be grafted. This rather aggressive method may result in an important proportion of non-viable proteins. In addition, due to limited scaffold availability, this approach will also be restricted to simultaneous transfer of a very small number of linear epitopes.

A secondary structure-independent approach for side chain grafting may offer an interesting alternative to the above methods [12,13]. It is a generalization of secondary structure-driven side-chain grafting, based on the observation that proteins that possess radically different folds may bind structurally similar protein interfaces [3,14,15]. Furthermore, it was shown that structurally dissimilar proteins that interact at one binding site used the same contact residues at the surface of the target [3]. These observations led to the design of artificial protein binders by grafting a functional motif onto a scaffold possessing a 3D structure unrelated to that of the natural protein binder but capable of reproducing the functional topology of the motif. In this case, scaffold identification is achieved by detecting a set of residues in a protein structure that adopt a topology similar to that adopted by the selected motif in the natural protein binder, but not necessarily in the same structural context. This may be achieved if only Ca and Cb coordinates of the residues are taken into account for the scaffold search. This principle also renders the screening independent of the type of amino acids that are natively present on the scaffolds examined. To fully exploit the potential of the method, systematic examination of the complete Protein Data Bank (PDB; more than 75 000 protein structures in April 2012) rather than a limited subset of pre-selected structures is required. For this purpose, software capable of analysing protein structures at the atomic scale and of detecting sets of Ca, Cb atoms in a defined topology were developed [16,17]. It should be noted that this search principle, implemented in STAMPS software (search for three-dimensional atom motifs in protein structures) [16], allows identification of scaffolds for continuous and discontinuous epitopes. We previously used this approach and STAMPS to design three miniproteins that are capable of binding the Kv1.2 channel with affinity between 500 nM and 1.6 μM [13], without further optimization. The grafting of a discontinuous epitope of erythropoietin (EPO) at the surface of the rat pleckstrin homology domain of phospholipase Cδ1 led to an artificial erythropoietin receptor (EPOr) binder with binding affinity of 24 nM after optimization [18]. However, applications of this approach are still scarce, and the available data do not yet permit evaluation of its robustness.

The secondary structure-independent *in silico* design approach may constitute the first step of a general hybrid computational/experimental strategy (Scheme 1), in which initial protein binders showing measurable albeit weak affinities for the target are obtained (step 1). In a second step, beyond the scope of the present paper, focused libraries of these protein binders may be designed to select miniproteins with the desired affinity/selectivity (step 2). The coupling of *in silico* approaches with powerful experimental methods for testing and optimizing the initial protein binder candidates has already resulted in important successes in the field of protein binder design [10,19]. However, these studies also suggested that the limiting step of this hybrid approach remains discovery of the initial binders. The global success rate of the design approach may be significantly improved if the *in silico* design step were able to produce a large set of initial protein binders rather than a single one.

The goal of the present paper is to evaluate the capacity of the secondary structure-independent grafting approach to respond to the challenge of designing a set of initial protein binders that target a specific binding site (step 1). As a test case, we applied it to the design of miniproteins targeting the catalytic site of matrix metalloproteinases (MMPs). MMPs are involved in many pathological conditions in humans, such as cancer, arthritis and cardiovascular diseases [20,21]. This family contains more than 20 members characterized by high structural similarity of their



Scheme 1. General strategy for protein binder design.

catalytic domains. Artificial miniprotein binders may offer an interesting alternative to synthetic inhibitors [22–26] for obtaining compounds with high affinity/selectivity. Catalytic domains of MMPs are characterized by a high degree of similarity of their sequence and 3D structures [27,28]. Their action is regulated by TIMPs (tissue inhibitors of metalloproteinases), endogenous potent inhibitors proteins with sub-nanomolar affinities for MMPs [29]. Mutagenesis [30–33] and structural studies [34–37] revealed the particular role of the first four residues of TIMPs that interact in the S1–S3′ region of the catalytic domain of MMPs. The structural studies also confirmed interaction of the α-amino and carbonyl groups of residue in position 1 of TIMPs with the catalytic Zn of MMPs, as previously evidenced via chemical modification of the N-terminal residue in TIMP-2 [38] and TIMP-4 [39]. Thus, these four residues were used as the motif for grafting onto miniprotein scaffolds to design miniprotein MMP inhibitors. The scaffolds were identified by systematic screening of the PDB using STAMPS [16]. Ten protein binder candidates were produced and evaluated as MMP inhibitor. We also sought to evaluate whether these scaffolds permitted reproduction of the interactions established by the natural protein binder with the targets.

Results

Motif to be grafted

The atom coordinates of residues C1, S2, C3 and S4 of TIMP-2 selected as the functional motif for grafting were taken from the 3D structure of the catalytic domain of MMP-14 in interaction with TIMP-2 (PDB code 1BQQ) [35]. In this structure, residues 2–4 exhibit a series of hydrogen bonds between main-chain donor and acceptor groups and residues of the catalytic domain of the enzyme (Fig. 1). In the present work, this structure was considered as the prototype of MMP–TIMP interaction.

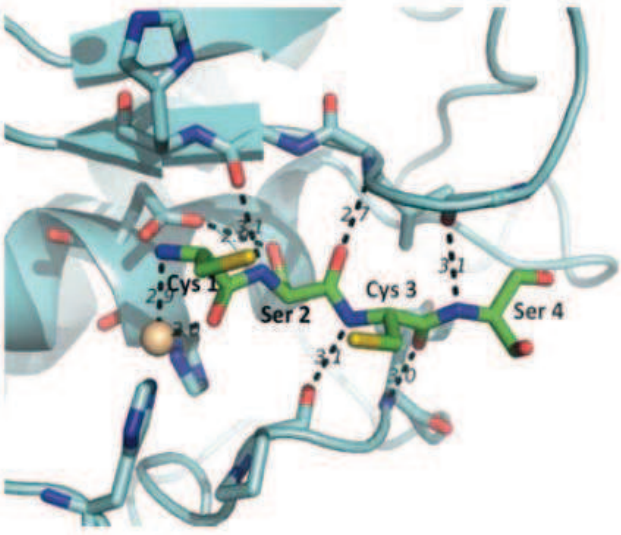


Fig. 1. Inter-molecular interactions between the four N-terminal residues of TIMP-2 (shown as sticks) and residues of the catalytic domain of MMP-14 (also shown as sticks) (PDB code 1BQQ). The catalytic zinc atom of the enzyme is shown in orange (sphere representation). Residues are coloured by atom type (N, blue; O, red; S, yellow; C, green for TIMP-2 residues; cyan for the catalytic domain of MMP-14).

Scaffold search with STAMPS

Scaffold searches were performed using the developmental version of STAMPS version 3.1.1. Several constraints were imposed during these searches. The residue equivalent to C1 in the scaffold must be located in the N-terminal region of the scaffolds to be able to reproduce a similar interaction as observed with the catalytic Zn atom of MMPs in TIMP complexes. Taking into account the limited size of the motif (four residues) and its widespread structure, the number of solutions may be very large. This was confirmed by preliminary STAMPS searches performed without restriction on the type of the equivalent residues (data not shown). To limit the number of solutions, two searches were performed using restraints on the native residues present in the scaffolds. In search 1 (S_Cys1), the native residue in the position equivalent to C1 in the scaffold must be a cysteine, the other residues (2–4) being one of the 20 amino acids. In search 2 (S_Cys3), we restricted the selected scaffolds to those naturally possessing a cysteine in the position equivalent to C3 of the motif, with residues 1, 2 and 4 being one of the 20 amino acids. Other parameters were included to limit the number of scaffolds identified by STAMPS. We searched for the equivalent motif in the first five residues of the proteins, with the possibility of deleting one residue in the N-terminal position of the native sequence of the scaffold. The size of the scaffolds was limited to a range of 15–60 residues, and the maximum root mean square deviation (rmsd) of the Ca–Cb motifs between the topology observed in the complex 1BQQ and that of the selected scaffold was set to 1.25 Å. Searches S_Cys1 and S_Cys3 resulted in 21 and 198 solutions, respectively. As STAMPS exhaustively examines all PDB files, the solutions identified include some redundancies (Table S1). Figure 2A shows that the sizes of the scaffolds identified are distributed between 15 and 60 residues. The topological similarity between the searched motif and those identified in these scaffolds is high, as shown by the rmsd distribution for the 219 solutions (Fig. 2B). A similar conclusion may be drawn from the distribution of the mean Ca–Cb angle after superimposition on the TIMP-2 motif (Fig. 2C). Furthermore, the distribution of steric factors for the 219 solutions in the binding mode required to reproduce the position of the functional motif in TIMP-2 relative to the cat-

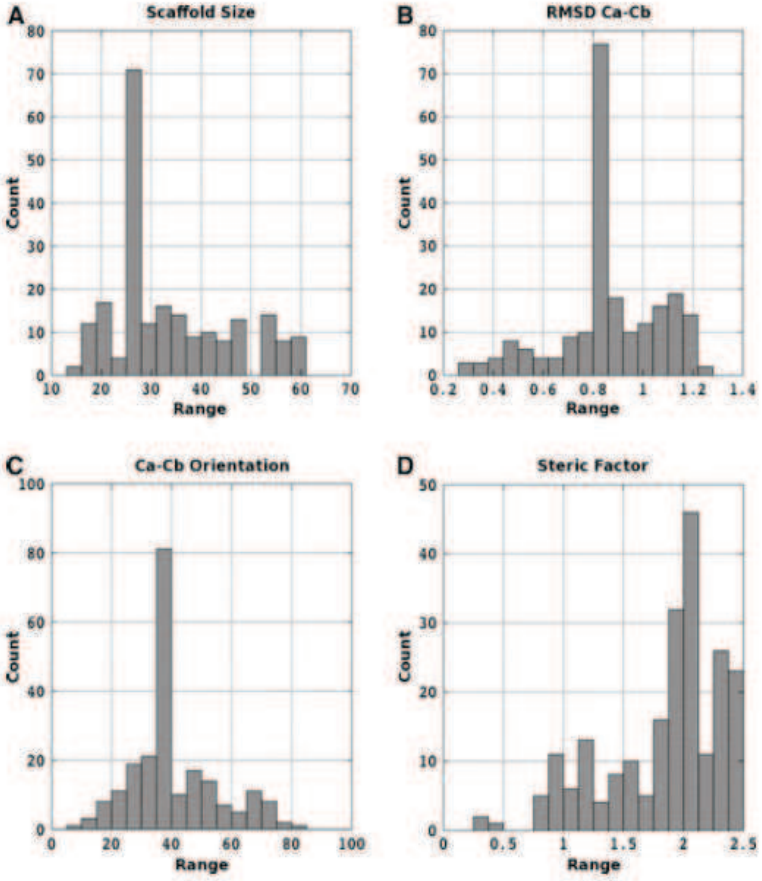


Fig. 2. Global analysis of the scaffolds identified by STAMPS. (A) Distribution of the protein scaffold sizes. (B) rmsd of the positions of the Ca,Cb atoms after superimposition on the corresponding atoms in the functional motif. (C) Geometric deviation measured on the basis of the mean value of the Ca–Cb angles between the functional motif and the corresponding vectors in the considered scaffold. (D) Geometric deviation measured on the basis of the inter-molecular steric factor.

alytic domain of MMP-14 shows satisfactory steric compatibility with the target (Fig. 2D). Visual examination of the 219 complexes identified by STAMPS was performed to check for possible minor steric hindrance not detected by the crude approach implemented in STAMPS and to eliminate scaffolds in which the topologically equivalent motif was located in an unstructured region. We finally selected a set of ten scaffolds that were further investigated (Table 1). Their 1D structures and those of the mutated forms are shown in Fig. 3. The 3D structures of the ten selected scaffolds are shown in Fig. 4. This illustrates well the diversity of the secondary/tertiary structures of these scaffolds and the structural similarity of the functional motifs in the various contexts.

Electrostatic and shape similarity between the protein binder candidates and the natural binder TIMP-2

We compared the shapes of the ten protein binder candidates with that of the natural inhibitor TIMP-2 in the region involved in interaction with the MMP using PIPSA software [40]. Shape similarity is quantified by calculation of the Hodgkin shape similarity index (SSI_H), which varies from 0 in the absence of shape similarity to 1 when the shapes are identical. Figure 5A shows the values obtained for the ten protein binder candidates. The shape similarity in the binding region between the natural ligand TIMP-2 and the ten scaffolds selected is satisfactory, as SSI_H_part ranges from 0.56 to 0.75.

Electrostatic potential plays an important role in binding affinity. This long-range interaction is determined by the full charge distribution of the protein scaffold. As no restriction on the electrostatic potential was applied during the scaffold search, it is important to check for electrostatic similarity between the selected scaffolds and the natural binder. We calculated the electrostatic potential for the ten scaffolds using the Poisson–Boltzmann equation in APBS software [41]. The structures used were those derived from STAMPS that included mutations of the motif residues when necessary (see below). The resulting electrostatic potential grids were compared with that calculated for the natural inhibitor TIMP-2 in a region restricted to the binding region with the target (see Experimental procedures). The results indicate that most scaffolds exhibit a similar electrostatic potential in their binding region, with the exception of 1EDP_A and 2F3C_I_m, which showed uncorrelated potentials compared to the natural binder (Fig. 5B). However, such a situation also occurs in some cases in natural binders (P. Cuniasse, unpublished data). As none of the ten scaffolds

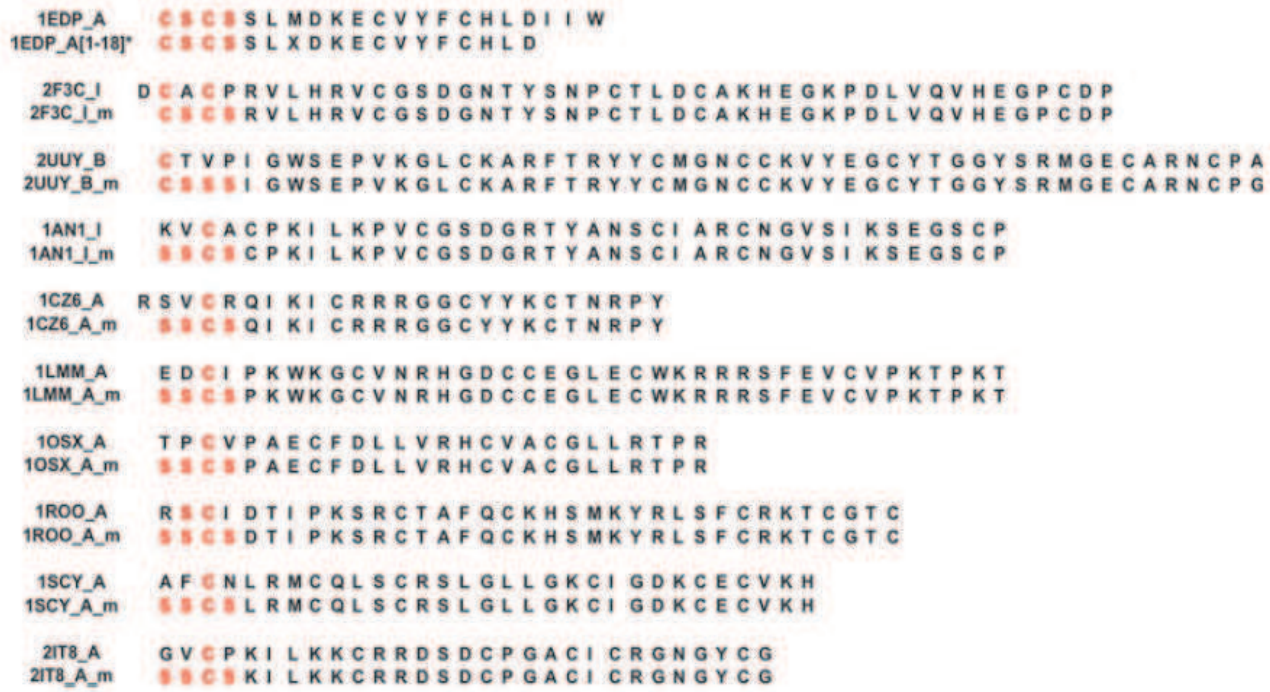


Fig. 3. Amino acid sequences of the ten protein binders selected for molecular modelling and experimental evaluation. The residues of the functional motif are shown in red. For each scaffold, the first line indicates the amino acid sequence found in the PDB file and the second line indicates the sequence of the protein produced. X, norleucine.

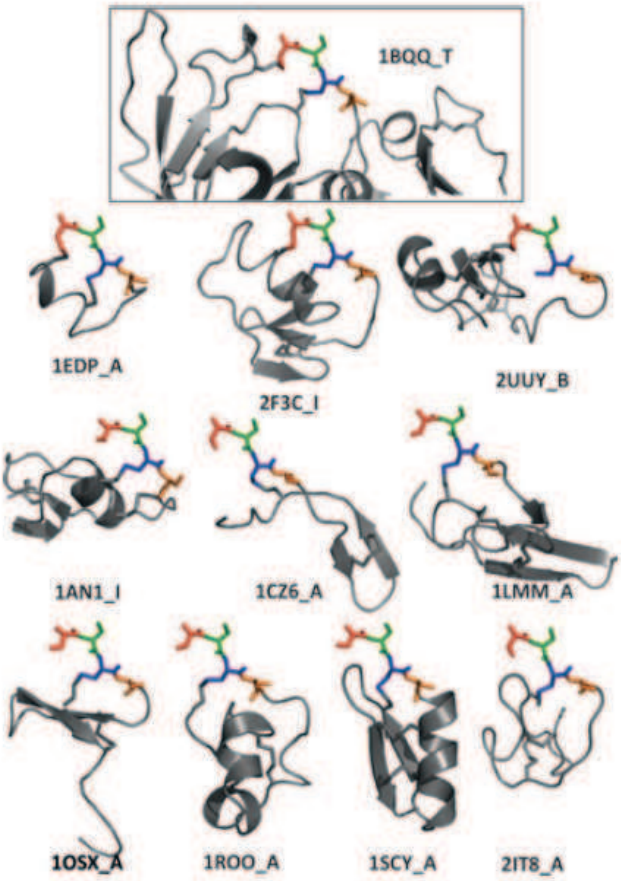


Fig. 4. 3D structures of the scaffolds identified by STAMPS via screening of the PDB and selected for further evaluation. The inset at the top of the figure shows the structure of TIMP-2 in its complex with MMP-14 (PDB code 1BQQ). The structures are shown in cartoon representation in grey. Native disulfides and the residues constituting the grafted motif are indicated by stick representation. Hydrogen atoms are omitted.

displayed anti-correlated electrostatic potentials in the binding region with respect to TIMP-2, they were all retained for further evaluation.

Molecular dynamics refinement of the selected candidates

The ten initial complexes identified by STAMPS were refined by energy minimization and restrained molecular dynamics *in vacuo*. The resulting refined structures were analysed, and a series of parameters characterizing protein–protein interactions were determined. For each complex, positional restraints were applied to all non-hydrogen atoms of the catalytic domain of MMP-14 to preserve its 3D structure and to the four residues of the protein binder designs to maintain their position with respect to the catalytic domain of MMP-14. A second type of restraint was applied to the scaffolds during refinement to preserve their overall 3D structure while allowing global motion of the scaffold with respect to the target. This allowed optimization of surface complementarity of the two partners. These restraints correspond to a set of distances joining the Ca of each residue to the Ca atoms of all other residues in the scaffold. This distance matrix was automatically generated at the start of the refinement protocol, and the reference distances were taken from the coordinates of the X-ray/NMR structure of the scaffold obtained from the PDB (see Experimental procedures). The initial step of the refinement protocol consisted of an energy minimization to correct for possible unfavourable local steric or electrostatic interactions in the starting complex. This step is followed by 15 cycles of molecular dynamics (25 000 steps each). At the end of each cycle, the structure was subjected to energy minimization under positional restraints. The resulting structure was saved for further analysis. Parameters characterizing these protein complexes were determined (Table 2). The total variation in accessible surface area upon binding of the complex (ΔASA_bind) ranges from 1086 to 2182 $Å^2$ for 1OSX_A_m and 1AN1_I_m, respectively. With the

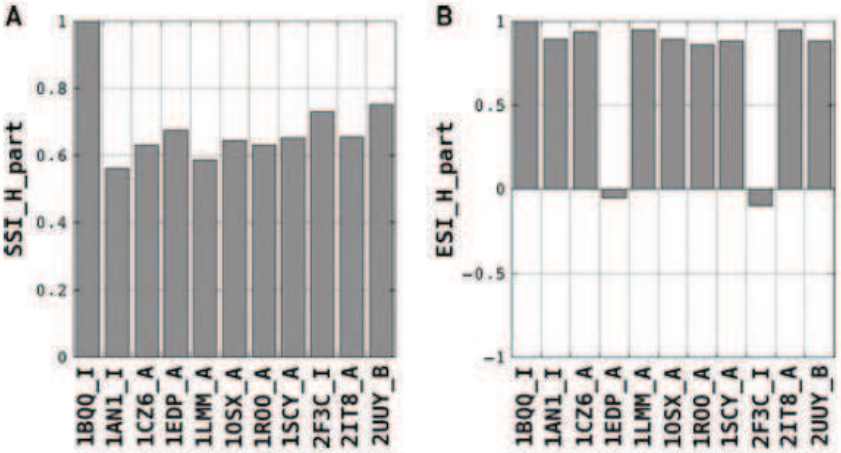


Fig. 5. Shape (A) and electrostatic (B) Hodgkin similarity indices calculated between the reference ligand TIMP-2 and the ten selected protein binder candidates in the region involved in interaction with the target MMP-14. SSI_H_part, Hodgkin shape similarity index; ESI_H_part, Hodgkin electrostatic similarity index.

Table 1. Scaffolds identified by STAMPS after exhaustive screening of the PDB, and selected for further *in silico* and experimental evaluation.

Search	Name	Protein size	Motif size	rmsd[a]	Orientation	Steric (inter/intra)	Residue 1		Residue 2		Residue 3		Residue 4	
S_Cys1	1EDP_A	17	4	0.61	21.39	2.12/0.11	Cys	1	Ser	2	CYS	3	SER	4
	2F3C_I_m	46	4	0.48	16.76	1.18/0.30	Cys	6	Ala	7	CYS	8	PRO	9
	2UUY_B_m	52	4	0.36	9.71	1.05/0.25	Cys	24	Thr	25	VAL	26	PRO	27
S_Cys3	1AN1_I_m	40	4	1.05	47.38	2.27/0.28	Lys	2	Val	3	CYS	4	ALA	5
	1CZ6_A_m	25	4	0.54	17.80	1.76/0.17	Ser	2	Val	3	CYS	4	ARG	5
	1EDP_A_m	17	4	0.61	21.39	2.12/0.11	Cys	1	Ser	2	CYS	3	SER	4
	1LMM_A_m	40	4	1.13	53.52	1.97/0.24	Glu	1	Asp	2	CYS	3	ILE	4
	1OSX_A_m	26	4	1.13	51.76	1.21/0.23	Thr	17	Pro	18	CYS	19	VAL	20
	1ROO_A_m	35	4	1.21	51.94	0.80/0.13	Arg	1	Ser	2	CYS	3	ILE	4
	1SCY_A_m	31	4	1.14	48.18	0.49/0.16	Ala	1	Phe	2	CYS	3	ASN	4
	2F3C_I_m	46	4	0.48	16.76	1.18/0.30	Cys	6	Ala	7	CYS	8	PRO	9
	2IT8_A_m	29	4	0.71	16.40	1.17/0.30	Gly	1	Val	2	CYS	3	PRO	4

[a] Calculated based on the position of Ca,Cb atoms.

exception of scaffold 1OSX_A_m, the values observed for these scaffolds are in the range reported for natural complexes (1600 ± 400 Å^2) [42]. A similar conclusion may be drawn from the variation in accessible surface area of the protein scaffolds upon binding (ΔASA_interf_lig), which ranges from 592 Å^2 for 1OSX_A_m to 1073 Å^2 for 1AN1_I_m. The number of residues of the protein binder located at the interface with the target varies from 11 to 22. This variability is also illustrated in Fig. 6, in which the binding surfaces of the protein target in the modelled complexes are shown in orange. In addition, the protein binder candidates interact with slightly different regions on the surface of the target. The ten protein binder candidates presented a predominantly polar interface, with the non-polar fraction varying between 8% and 30% (Table 2). The gap volume indices characterizing surface complementarity between the protein binders and the target varied between 1.61 and 2.15 Å, which is in the range observed for natural protein–protein complexes [43]. We also analysed the inter-molecular hydrogen bonds established between the two partners. This number ranges from 5 in the case of 1EDP_A to 13 in the case of 1LMM_A_m (Table 2). When the number of hydrogen bonds were normalized to the interface area, the resulting hydrogen bond surface densities varied between 4.0 and 8.2×10^{-3} Å^{-2}, which is also in the range observed for natural hetero-dimeric protein–protein complexes [43,44]. In addition, salt bridges and cation–π interactions were observed in the models of the MMP-14/designs complexes (Table 2). The above observations show that the selected scaffolds are able to establish a set of inter-molecular interactions, some of which involve residues of the grafted motif and others corresponding to residues located at the interface with the target. In addition, this modelling study showed that the parameters characterizing the ten artificial interfaces are within the ranges observed for natural protein–protein complexes.

Production and evaluation of the ten protein binder candidates

The above results prompted us to produce the ten protein binder candidates in order to evaluate their ability to bind a series of MMPs (MMP-2, MMP-8, MMP-9, MMP-12, MMP-13 and MMP-14). Some of the designed artificial protein binders were produced by solid-phase peptide chemical synthesis, and others were produced by recombinant expression in *Escherichia coli*. The mutations depended on the native composition of each particular scaffold (Fig. 3). In the

Table 2. Characterization of refined MMP-14-protein binder candidate complexes. ΔASA_bind, total variation of accessible surface area upon binding; ΔASA_interf_lig, Protein–binder interface area; Nb_Res_Inter_lig, number of residue on the protein-binder side of the interface; %Res_NP_lig, proportion of apolar residues on the protein-binder side of the interface; % Surf_NP_lig, proportion of atoms belonging to apolar groups on the protein binder side of the interface; Gap_Vol_In_ligand, gap volume index; Nb_interprot_HB, number of inter-molecular hydrogen bonds; Nb_Salt_Bridges, number of inter-molecular salt bridges; Nb_CP, number of inter-molecular cation–π interactions; Interprot_HB_SD, inter-molecular hydrogen bond surface density (number of hydrogen bonds/total buried accessible surface area upon binding).

Scaffold	1BQQ_T	1EDP_A	2F3C_I_m	2UUY_B_m	1AN1_I_m	1CZ6_A_m	1LMM_A_m	1OSX_A_m	1ROO_A_m	1SCY_A_m	2IT8_A_m
ΔASA_bind (Å²)	2966	1230	1514	2029	2182	1561	1821	1086	1204	1445	1370
ΔASA_interf_lig (Å²)	1552	664	826	1043	1073	891	917	592	672	771	786
Nb_Res_Inter_lig	36	12	15	22	20	14	19	11	12	15	13
%Surf_NP_lig	42	22	8	30	27	16	21	17	15	15	18
Gap_Vol_In_ligand (Å)	1.92	1.72	1.84	1.87	2.15	1.62	1.61	1.84	1.67	1.91	1.70
Nb_interprot_HB	17	5	13	12	13	12	15	6	6	11	9
Nb_Salt_Bridges	0	0	0	1	1	1	1	0	1	2	0
Nb_CP	2	0	0	1	1	1	0	2	1	0	1
Interprot_HB_SD (10^{-3} Å^{-2})	5.7	4.0	6.6	5.9	5.9	7.7	8.2	5.5	4.9	7.3	6.5

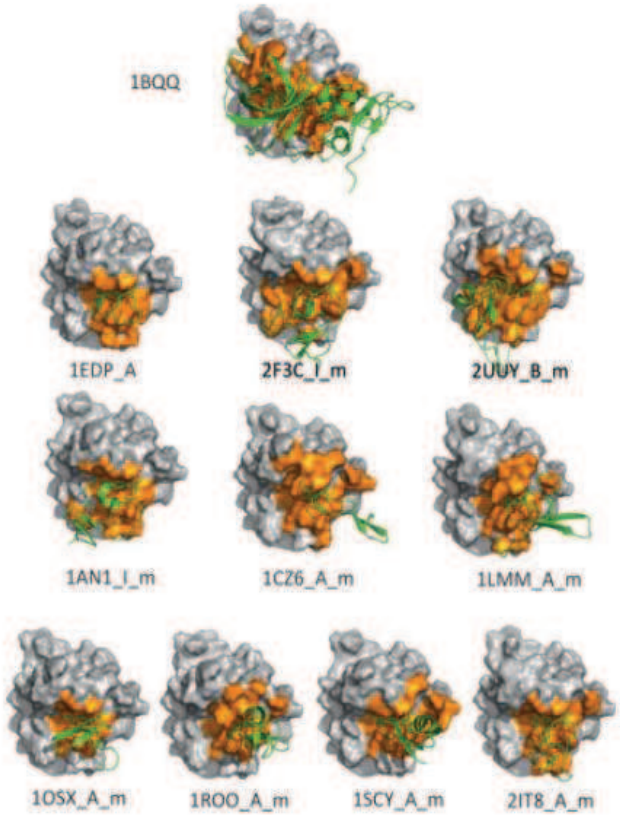


Fig. 6. Graphical representation of the complex between the catalytic domain of MMP-14 and the natural inhibitor TIMP-2 or the designed protein binders. The catalytic domain of MMP-14 is shown in surface represented in grey. Protein ligands are shown in cartoon representation in green. The MMP-14 interface (residues at a distance < 3.5 Å) is coloured in orange.

case of endothelin-1 (1EDP_A), no mutation was necessary as the four native residues identified by STAMPS in this scaffold had the CSCS sequence. In the case of scaffold 2F3C_I_m, the residue in position 1 of the native scaffold was deleted. This left the native cysteine in position 2 of the scaffold as the N-terminal residue in 2F3C_I_m. The residues in positions 3 and 5 of the native sequence (corresponding to positions 2 and 4 of the motif) were mutated to serine. Similarly to 2F3C_I_m, residue R1 present in the native structure of 1CZ6_A was deleted, making native residue C4 residue C3 of the motif. In the other scaffolds, cysteines present in positions 1 and 3 in the native structure were preserved. When the residue in position 1 or 3 in the native scaffold was not a cysteine, it was replaced by a serine, a residue that is sterically similar to cysteine. This strategy avoids unwanted effects of the presence of a free cysteine during the protein production step. In all cases, residues corresponding to positions 2 and 4 of the motif were replaced by serine.

Table 3. Binding affinity for MMPs (K_i values in μM; mean ± standard error) of the ten artificial protein binders. ND, not determined.

Protein binders	Scaffold name	MMP-2	MMP-8	MMP-9	MMP-12	MMP-13	MMP-14
1EDP_A	Endothelin-1	7 ± 1	2 ± 0.2	7 ± 0.9	1 ± 0.1	0.45 ± 0.03	ND
2F3C_I_m	Thrombin inhibitor infestin	32 ± 3	6 ± 1	36 ± 5	31 ± 2	73 ± 7	100 ± 15
2F3C_I[a]	Thrombin inhibitor infestin	56 ± 6	56 ± 19	51 ± 5	77 ± 7	143 ± 17	217 ± 39
2UUY_B_m	Tick-derived tryptase inhibitor	69 ± 8	24 ± 3	39 ± 6	160 ± 12	158 ± 20	ND
2UUY_B[a]	Tick-derived tryptase inhibitor	10 ± 1	4 ± 0.2	20 ± 2	8 ± 0.9	9 ± 0.7	129 ± 19
1AN1_I_m	Leech-derived tryptase inhibitor	22 ± 2	5 ± 0.7	30 ± 4	136 ± 10	38 ± 7	132 ± 11
1CZ6_A_m	Androctonin	92 ± 17	40 ± 6	45 ± 6	>500	183 ± 18	213 ± 15
1LMM_A_m	Psalmotoxin-1	18 ± 1	11 ± 3	32 ± 3	34 ± 1	68 ± 5	24 ± 3
1OSX_A_m	Extracellular domain of BLyS R3	97 ± 13	54 ± 9	58 ± 10	197 ± 12	206 ± 29	>500
1ROO_A_m	SHK toxin	182 ± 21	54 ± 8	102 ± 14	457 ± 35	381 ± 44	>500
1SCY_A_m	Scyllatoxin	150 ± 15	75 ± 5	130 ± 13	160 ± 47	440 ± 46	ND
2IT8_A_m	Trypsin inhibitor 2	173 ± 15	70 ± 8	157 ± 20	>500	>500	>500

[a] Native N-terminal sequence.

Table 4. Binding affinity for MMPs (K_i values in μM; mean ± standard error) of the analogues of 1EDP_A and 1AN1_I. ND, not determined; hF, homo-phenyl alanine; FF, a pseudo residue with a methyl-biphenyl moiety.

Protein binder	Scaffold name/mutation	MMP-2	MMP-8	MMP-9	MMP-12	MMP-13	MMP-14
1EDP_A	Endothelin-1	7 ± 1	2.0 ± 0.2	7.0 ± 0.9	1.0 ± 0.1	0.45 ± 0.03	ND
1EDP_[1-18][a]		20 ± 3	4.0 ± 0.3	8.0 ± 1.0	6.0 ± 0.3	4.0 ± 0.3	ND
1EDP_[1-18]S2hF[a]	S2hF	0.6 ± 0.4	0.7 ± 0.6	0.4 ± 0.3	0.2 ± 0.2	0.3 ± 0.2	100 ± 12
1EDP_[1-18]S2FF[a]	S2FF	1.0 ± 0.4	1.0 ± 0.5	0.7 ± 0.3	0.6 ± 0.4	0.4 ± 0.3	ND
1AN1_I	Leech-derived tryptase inhibitor	22 ± 2	5 ± 1	30 ± 4	136 ± 10	38 ± 7	132 ± 11
1AN1_I_m1	(K2S, V3I, A5S, L10Y)	26 ± 4	16 ± 2	24 ± 3	85 ± 6	64 ± 7	51 ± 6
1AN1_I_m2	(K2S, V3I, A5S, A22M)	19 ± 2	18 ± 2	33 ± 5	122 ± 100	169 ± 43	70 ± 3
1AN1_I_m3	(K2S, V3I, A5S, K8W)	17 ± 1	7 ± 0.2	31 ± 4	85 ± 3	42 ± 4	93 ± 6
1AN1_I_m4	(K2S, V3I, A5S, K8W, L10Y, A22M)	16 ± 2	3.0 ± 0.3	20 ± 3	71 ± 4	23 ± 2	99 ± 8
1AN1_I_m5	(K2S, V3M, A5S, K8W, L10Y, A22M)	13 ± 1	2.0 ± 0.1	20 ± 3	36 ± 4	11 ± 1	59 ± 2

[a] Residue M7 was replaced by a norleucine.

As the functional motif is taken from the natural broad-spectrum inhibitor of MMPs, TIMP-2, the protein binders are expected to interact with all MMPs. The binding of the ten protein binder candidates to a series of six MMPs was thus evaluated by determining their ability to inhibit the cleavage of a synthetic fluorogenic substrate (see Experimental procedures). The corresponding inhibition constants (K_i) are reported in Table 3. In the case of MMP-2, MMP-8 and MMP-9, all the designed binders were able to inhibit the catalytic activity of these metalloproteinases, with K_i values ranging from 2 to 182 μM. In the case of MMP-12 and MMP-13, 1EDP_A, 2F3C_I_m, 2UUY_B_m, 1AN1_I_m and 1LMM_A_m displayed K_i values ranging from 450 nM to 158 μM. 1CZ6_A_m showed a K_i value of 183 μM for MMP-13, and no inhibition of catalytic activity was observed for MMP-12 at 500 μM. 1ROO_A_m displayed weak inhibition of MMP-12 and MMP-13 (K_i > 300 μM), and 2IT8_A_m displayed no inhibition at 500 μM for these two enzymes. Only 2F3C_I_m, 1AN1_I_m, 1CZ6_A_m and 1LMM_A_m were able to inhibit MMP-14, with K_i values ranging from 24 to 213 μM. In the case of 2F3C_I and 2UUY_B, we also evaluated binding of the native scaffolds (Table 3). 2F3C_I_m displayed binding affinities that were slightly improved by factors of 2–10 for the MMPs tested, compared to the native sequence. In contrast, 2UUY_B_m showed lower binding affinity for these MMPs compared to 2UUY_B, by factors ranging from 5 to 17. In this study, the interactions between the grafted motif and the target involved main-chain chemical functions. It is therefore expected that native scaffolds reproduce these interactions. However, the contrasting results obtained with 2UUY_B_m and 2F3C_I_m as compared to their native analogues show that the consequences of mutating the native residues of the scaffold to graft the functional motif may depend on the particular context of each scaffold.

Analogues of some protein binders

1EDP_A and 1AN1_I_m were selected for optimization. We produced four analogues of 1EDP_A. Exami-

Table 5. Inter-atomic distances (Å) between the CSCS motif in the artificial binder 1EDP_A and the catalytic domain of MMP-12 deduced from the X-ray structure of this complex. The distances (Å) observed for the CSCS motif in the X-ray structure of the natural binder TIMP-2 in complex with MMP-14 (PDB code 1BQQ) are also shown.

	Residue name	Residue number	Atom		Residue name	Residue number	Atom	1BQQ	1EDP_A
Binder	Cys	1	N	MMP	Zn2	289	ZN	2.9	2.0
Binder	Cys	1	O	MMP	Zn2	289	ZN	2.0	2.1
Binder	Cys	1	N	MMP	Glu	240/127	OE1	2.7	3.7
Binder	Ser	2	N	MMP	Ala	200/87	O	3.1	3.5
Binder	Ser	2	O	MMP	Ala	200/87	N	3.8	3.7
Binder	Ser	2	OG	MMP	Glu	240/127	OE1	2.9	4.5
Binder	Ser	2	O	MMP	Leu	199/86	N	2.7	2.7
Binder	Cys	3	N	MMP	Pro	259/146	O	3.1	3.7
Binder	Cys	3	O	MMP	Tyr	261/148	N	3.0	3.6
Binder	Ser	4	N	MMP	Gly	197/84	O	3.1	3.0

nation of the model of this compound in interaction with the catalytic domain of MMP-14 suggested that the last three residues may be removed. We thus synthesized 1EDP_[1-18]* and two analogues in which residue S2 was replaced by non-natural amino acids possessing bulky aromatic side chains. This strategy is based on the specific structure of the S1′ sub-site of most MMPs, which is able to accommodate long non-natural side chains. This has influenced the development of synthetic inhibitors of MMPs and generally led to improved binding affinity compared to natural side chains [23]. The 1EDP analogues were produced by solid-phase peptide synthesis. Table 4 shows their binding affinity for six MMPs. As expected, the presence of homo-phenyl alanine in position 2 led to improved binding affinities by factors ranging from 5 for MMP-8 to 32 for MMP-2 compared to a serine in this position (1EDP_[1-18]*). When the side chain of S2 was replaced by a methyl-biphenyl moiety (compound 1EDP_[1-18]S2FF), we also observed improved affinities for the targets by factors up to 20 compared to 1EDP_[1-18]*. These observations are consistent with a binding mode whereby the residue in position 2 in the 1EDP_A analogue occupies the S1′ sub-site of the MMPs tested. We observed binding constants of up to 290 nM for MMP-12 (compound 1EDP_[1-18] S2hF*, Table 4). We also produced a series of analogues of 1AN1_I_m in which the residues in positions 2, 3, 5, 8, 10 and 22 (Table 4) were substituted. These mutations were suggested by the FOLDX software [45] based on a modelled structure of 1AN1_I_m in interaction with MMP-8. This complex was selected as 1AN1_I_m displayed highest affinity for the catalytic domain of MMP-8. However, these substitutions only led to modest improvements in affinity for some MMPs by factors up to ~ 4, and the best binder displayed a K_i value of 2 μM for MMP-8 (Table 4).

Structural characterization of 1EDP_A in complex with MMPs

The above results strongly suggested that the designed binders interact with the MMPs in the expected binding mode. In order to characterize the binding mode of these artificial binders and determine to what extent they are able to reproduce the interactions established by the natural binder TIMP-2 with MMPs (PDB code 1BQQ), crystallization assays were performed. For this study, we selected 1EDP_A. Soaking experiments performed with 1EDP_A on MMP-12 crystals caused lattice disorder, and data were collected for the one crystal that was still viable. Despite the difficulties with the experiment, the surviving crystal diffracted to a resolution of 2.8 Å, and the data were used to calculate an electron density map in which it is possible to discern unambiguously a peptide fragment interacting in the S1–S3′ region of the enzyme. The electron density is strong for residues 1–4 of 1EDP_A, so the position of the atoms and characterization of their interaction with the catalytic domain of the enzyme may be considered reliable. The electron density corresponding to regions of the scaffold not directly involved in the interaction with MMP-12 is less well defined. The reasons for the weaker electron density may relate to the short length of the soak, partial cleavage of the scaffold and conformational variability of the scaffold within the crystal lattice. Conformational variability is also observable in the various structures of endothelin-1 deposited in the PDB, which display important conformational variability from each other. The sequence that matches the electron density for the four-residue stretch corresponds well to that for 1EDP_A. The presence of stronger electron density for the sulfur atoms that correspond to the cysteines in positions 1 and 3 of 1EDP_A is also reassuring. Residues C1 to S4 adopt a conformation within the

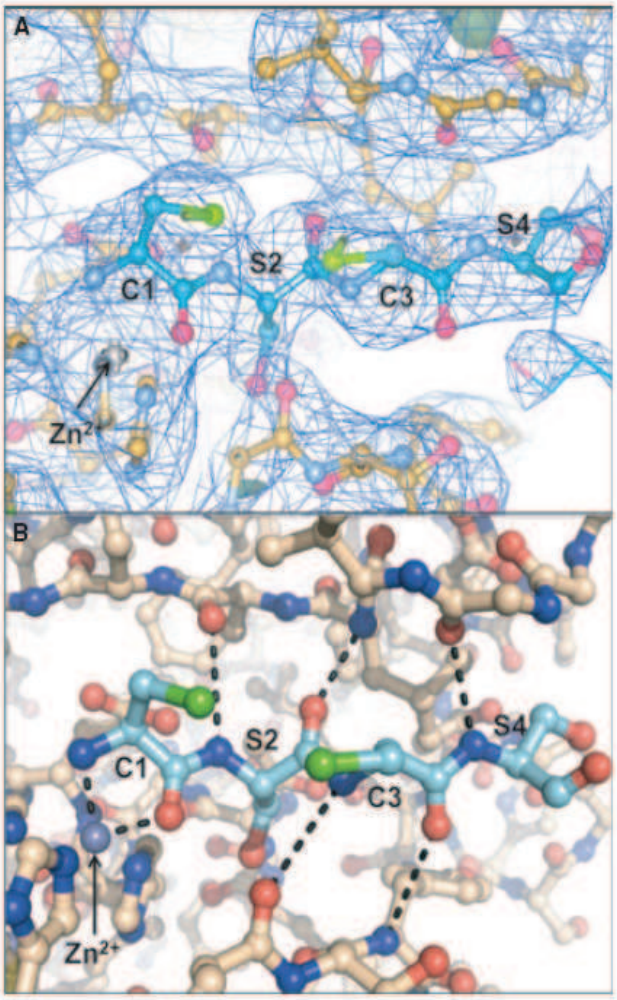


Fig. 7. X-ray structure of 1EDP_A in complex with MMP-12. The structures are shown in ball and stick representation and coloured by element type (N in dark blue, O in red and S in green). The carbon atoms are coloured in orange for the catalytic domain of MMP-12 and in blue for 1EDP_A. (A) The electron density is shown in mesh representation coloured in blue at 0.66 σ. (B) Only residues 1–4 of 1EDP_A are shown. The electron density is omitted and the inter-molecular interactions between residues 1–4 of 1EDP_A and the catalytic domain of MMP-12 are indicated by dashed lines.

active site of MMP-12 that is similar to that observed for the equivalent residues from the natural binder, TIMP-2, in complex with MMP-14 (Fig. 7). The N-terminal nitrogen and carbonyl oxygen of residue C1 interact with the catalytic Zn^{2+} atom, and the corresponding distances are close to that observed in the TIMP-2–MMP-14 complex (1BQQ, Table 4). Several distances between hydrogen bond donors and acceptors belonging to residues 1–4 of 1EDP_A and the catalytic domain of MMP-12 are consistent with the presence of inter-molecular hydrogen bonds (Table 5). Comparison of these distances also showed that some differences exist between the artificial binder 1EDP_A and the natural one. Four distances between the two partners involving hydrogen bond donors or acceptors were longer by 0.6–1.6 Å in the MMP-12–1EDP_A complex compared to the MMP-14–TIMP-2 complex (1BQQ). The distance separating the hydroxyl oxygen of S2 in 1EDP_A and the carboxylate oxygen of residue E240 in the enzyme showe an increase of 1.6 Å. This is due to the different rotamers adopted by residue S2 in the two structures. Finally, four distances in the MMP-12–1EDP_A complex were similar within 0.4 Å to those observed in the MMP-14–TIMP-2 complex.

Determination of the stability of the inter-molecular interactions in the artificial binders by molecular dynamic simulation

To further evaluate the possible differences in geometry and dynamic stability of the interactions consequent to grafting of the functional motif on the various scaffolds, we performed molecular dynamics simulations of the ten binders in complex with MMP-14 in water solution. For each complex, a 1 ns molecular dynamics trajectory was calculated using the complex identified by STAMPS as the starting structure, and further refined by energy minimization and molecular dynamics under positional restraints (see Experimental procedures). Twelve mean distances characterizing interaction of the selected motif with the target were calculated from these simulations (Table S2). We observed excellent agreement between the mean distances determined in simulation of the natural complex TIMP-2–MMP-14 and the corresponding distances observed in the X-ray structure of this complex (1BQQ). Binders 1CZ6_A_m, 1LMM_A_m, 1OSX_A_m, 2F3C_I_m and 2IT8_A_m reproduce these interatomic distances well. According to the simulations, other binders reproduce only partially the distances observed in the natural complex. 1EDP_A displayed nine distances that were similar within 0.5 Å to those found in the natural complex (Table S2). We also noted a good agreement between the distances determined in this simulation and that obtained from the X-ray structure of 1EDP_A in complex with MMP-12. In the case of 1SCY_A_m and 2UUY_B_m, eight distances are well reproduced. Designs 1ROO_A_m and 1AN1_I_m showed seven and six distances, respectively, that were similar within 0.5 Å to the corresponding ones in the TIMP-2–MMP-14 complex. We further examined, along the molecular dynamics trajectories of the ten complexes, the stability of seven interactions that we wished to reproduce. Two of them involve the N-terminal residue of the binder interacting with the catalytic Zn^{2+} ion. Five hydrogen bonds

Table 6. Occupancies of a set of inter-molecular interactions determined in the molecular dynamics simulation of the complexes between the catalytic domain of MMP-14 and the natural ligand TIMP-2 (1BQQ) or the ten artificial protein binders. The atoms involved in these interactions are indicated at the top of each column for the binder and the target MMP-14. Occupancies below 0.8 are shown in bold.

Protein binder atom Protein target atom	N_1 ZN_1	O_1 ZN_1	N_2 ALA_200_O	O_2 ALA_200_N	3_N PRO_259_O	3_O TYR_261_N	4_N GLY_197_O
1BQQ_T	1.00	1.00	**0.62**	1.00	0.99	1.00	1.00
1AN1_I_m	**0.00**	1.00	**0.02**	0.98	0.97	**0.62**	**0.00**
1CZ6_A_m	1.00	1.00	**0.79**	1.00	0.98	0.97	**0.00**
1EDP_A	1.00	1.00	0.96	1.00	**0.72**	1.00	1.00
1LMM_A_m	1.00	1.00	0.96	0.99	0.99	0.96	0.89
1OSX_A_m	1.00	1.00	**0.57**	1.00	0.99	**0.66**	1.00
1ROO_A_m	**0.00**	1.00	**0.00**	**0.67**	0.95	1.00	0.96
1SCY_A_m	**0.00**	1.00	**0.02**	0.91	0.99	1.00	**0.08**
2F3C_I_m	1.00	1.00	0.81	1.00	0.95	**0.27**	0.92
2IT8_A_m	1.00	1.00	**0.35**	**0.57**	0.94	0.99	0.99
2UUY_B_m	1.00	1.00	**0.52**	1.00	**0.64**	**0.47**	1.00

involving donors and acceptors of main-chain residues of the motif were also analysed (Table 6). rmsds for the positions of the protein atoms and the interface area between the partners indicated that equilibrium was reached in all cases in < 500 ps. We thus determined the occupancies (fraction of the time where the interaction is observed on the total time analyzed) of the selected inter-molecular interactions in the last 500 ps of the trajectories. To this end, we used a maximum distance between the acceptor atom and the hydrogen atom bound to the donor of 2.8 Å, and a maximum donor/hydrogen acceptor angle of 120°. In the case of the natural complex TIMP-2–MMP-14, six interactions were stable in the time frame analysed (occupancy of ~ 1). One hydrogen bond between the amide nitrogen of residue S2 and the carbonyl oxygen of residue A200 is formed only about 60% of the time. When considering the stabilities of these interactions in the ten protein binders, we found only three cases in which the interaction between the nitrogen of the residue in position 1 and the Zn^{2+} ion was not observed in the time frame (1AN1_I, 1ROO_A and 1SCY_A). In these structures, the hydrogen bond between the amide of the residue in position 2 of the binders and the carbonyl oxygen of residue Ala200 of the enzyme was also absent. Examination of these trajectories indicated rotation of the ψ angle of the residue in position 1 and the ϕ angle of the residue in position 2 of the three binders compared to the starting structure, and that the nitrogen of residue 2 establishes alternative intra-molecular hydrogen bonds (data not shown). It should be noted that these three scaffolds do not possess a cysteine involved in a disulfide bond in position 1 that would have precluded adoption of this conformation. However, this movement is not only due to the presence of a serine in position 1, as this conformation was not observed in scaffolds 1CZ6_A_m, 1LMM_A_m, 1OSX_A_m and 2IT8_A_m. Overall, Table 6 shows that, according to the simulations, some protein binders reproduce only partially the set of hydrogen bonds observed in the natural complex. In addition, several scaffolds compatible with reproduction of the hydrogen bonds observed in the natural complex displayed differences in occupancy compared to this complex. These differences may play a role in the lower affinity observed for the designed protein binders compared to the natural one, but other factors, specific for each scaffold, also contribute to this observation.

Discussion

Recent studies have shown that high-affinity protein binders may be obtained using *in silico* design approaches based on grafting of a functional motif onto protein scaffolds [10,18,19]. These studies clearly demonstrated the role of computational design in defining the framework within which mutagenesis or molecular evolution will enable achievement of the desired affinity [46]. However, these authors point out that, despite the power of the method, directed evolution/selection is unlikely to be successful unless the starting structure already possesses, even at a very low level, the activity that one wishes to select. These studies also suggested that computational design remains the limiting step of these approaches. Obtaining initial protein binders is clearly not an easy task. For instance, in a study reporting the design of protein binders targeting the b12 antibody, only one of 62 proteins designed was shown to possess measurable affinity, albeit at a low level [10]. This low success rate may be due, at least in part, to the rather aggressive

backbone grafting approach used in that study that resulted in a large number of deletions and mutations in the protein scaffolds. This may also explain why only 25 of the 62 protein designed could be expressed in soluble form and purified. In a study describing the design of a protein binder targeting the conserved stem region of the influenza hemagglutinin, the authors grafted hot spots derived from computational docking onto protein scaffolds identified using a protocol that has similarity with the secondary structure-independent approach used in the present paper [19]. Seventy-three of the 88 proteins designed were evaluated by yeast display. However, only two of these proteins displayed measurable but weak affinity for the target. This shows that computational protein binder design is not a solved problem. We still lack a protein design method allowing a high success rate.

Here, we used an *in silico* secondary structure-independent approach based on grafting of a functional motif onto miniprotein scaffolds to design binders targeting the catalytic sites of MMPs. This approach has already shown its ability to create binders that bind in the micromolar to nanomolar range prior to optimization [13,18]. The scaffold search was performed by systematic screening of the PDB using STAMPS software [16]. The search principle implemented in STAMPS allows identification of both continuous and discontinuous functional motifs. The grafted motif corresponded to the continuous N-terminal tetrapeptide of TIMP-2, a natural broad-spectrum protein inhibitor of MMPs. The scaffold search relied on identification of a set of side chains with a topology similar to that of the functional motif. This principle allows us to take advantage of the existing conformational state of the scaffold. It also avoids aggressive protein structure manipulation, with grafting of the functional site being achieved via a limited number of mutations at the surface of the selected scaffold. Here we sought to evaluate the yield of the method in terms of generating a set of active protein binders. To this end, we produced ten of the protein binder candidates obtained by the *in silico* approach. We showed that the ten proteins were able to target most of the MMP catalytic domains tested with affinities of up to 450 nM prior to optimization. This observation is consistent with the fact that the interactions established by the selected motif are expected to be similar with all MMPs because their catalytic domains possess a high structural similarity in the targeted region. Furthermore, none of the additional criteria used for scaffold selection (structural complementarity with the target and electrostatic similarity with the natural broad-spectrum protein inhibitor TIMP-2) permit discrimination between different MMP catalytic domains.

The ten selected scaffolds possess various 3D structures. Endothelin-1 was identified as a suitable scaffold. This protein naturally contains the CSCS N-terminal motif. A previous study showed that a truncated analogue of the related miniprotein sarafotoxin 6b possessing a similar fold to that of endothelin-1 and a CSCS N-terminal motif was capable of inhibiting MMP-1, MMP-2 and MMP-9 with affinities between 1 and 20 μM [47]. In this work, the authors identified sarafotoxin 6b as a suitable scaffold based on its N-terminal 1D sequence homology with TIMPs. It should be noted that we also identified sarafotoxin 6b as a suitable scaffold (pdb1srbA_00, Table S2), but it was not selected for experimental evaluation due to its fold similarity with endothelin-1. In the present work, nine other scaffolds whose 3D structures were unrelated to that of with endothelin-1 were identified by systematic screening of the PDB and selected for experimental evaluation. Scaffold 2F3C_I possess cysteine residues in positions 1 and 3. It displayed a fold similarity with endothelin-1 in the region of the two disulfides. The corresponding protein displayed binding affinities for the MMPs tested ranging from 6 to 100 μM. One scaffold naturally possessing a cysteine residue in the N-terminal position (2UUY_B) gave a protein binder with affinities between 4 and 129 μM for the MMPs tested. All other scaffolds possessed a cysteine residue in position 3 and interact with most MMPs tested, albeit with moderate affinities. Overall, as compared to previously reported studies, the yield of the approach may be considered as good as the ten proteins designed displayed measurable binding affinities for the targets. In addition, all these binders were produced by solid-phase peptide synthesis or recombinant techniques.

In the present work, we also wished to evaluate to which extent the selected scaffold permitted reproduction of the interactions established by the functional motif in the natural complex. This question has been addressed in some studies via structural characterization of the designed proteins in complex with their protein target in the case of secondary structure-driven motif grafting [48] and backbone grafting [10]. These studies revealed a high degree of structural mimicry for compounds displaying high affinity and specificity similar to that of the natural binder. In the present study, we showed high structural mimicry of the functional motif in the context of 1EDP_A, but also found some differences in the precise geometry of the interactions. We could not solve the 3D structure of other binders in interaction with one of the MMPs. Molecular dynamics simulations of the ten designed binders in interaction with the catalytic domain of MMP-14

suggested that some scaffolds allow reproduction of the interactions observed in the natural ligand. However, depending on the scaffolds, some differences in the dynamics of these interactions were observed. In this regard, the conformational control due to residue C1 may be important in the present case. These observations suggest that, in addition to topological considerations guiding scaffold selection, it may be important to take their dynamics in the bound state into account. This will enable selection of scaffolds offering the best conformational stability of the grafted residues.

The present results also illustrate one limit of grafting-based design methods. Any mutation performed to graft the motif may alter the conformational ensemble of the protein scaffold in the free state. This may prevent reproduction of the inter-molecular interactions with the target. This could also lead to an important free energy penalty for returning to the correct conformation at the bound state. In addition, increasing the conformational ensemble of the protein binder in the free state may lead to an increased unfavourable entropy variation upon binding. These phenomena may contribute to the moderate affinity obtained for most of the initial protein binders designed by grafting methods. However, several remarks may be made about our method. First, our approach entails a limited number of mutations (maximum four in this case) to graft the functional motif on the protein scaffold. The present approach is thus far less aggressive for protein structures than backbone grafting approaches, thereby increasing the possibility that the designed proteins actually bind the target. Second, the presence of disulfides in some scaffolds may contribute to preserving both the structure and the native conformational ensemble after mutation. Third, our method was designed to produce a series of protein binder candidates rather than a single one, thus increasing the possibility of obtaining at least one binder with the desired property. This is the reason why we are continuing to improve the STAMPS software to take advantage of the complete structural diversity of the PDB. Further efforts to correctly sample the conformational ensemble accessible to the molecule in the free state and the bound state by simulation approach are also required. This may permit the introduction of new scaffold selection criteria.

One advantage of the grafting approach is that it permits control of the location where the binder interacts at the target surface. Furthermore, knowledge of the binding mode of the initial moderate-affinity binder may be used to guide its optimization. This permitted production of an analogue of endothelin-1 in which the residue at position 2 was replaced by a bulky non-natural amino acid. This led to an increased binding affinity compared to the parent analogue for all MMPs tested (up to 290 nM for MMP-12). This is consistent with previous reports showing that bulky hydrophobic non-natural side chains interacting at the S1′ sub-site may lead to improved affinity of pseudo-peptide inhibitors. However, optimization of initial binders by single mutations often gives disappointing results, as observed in the case of scaffold 1AN1_I. Alternatively, coupling of efficient computational design methods with rapid experimental techniques for testing and evolving the proteins designed *in silico* could become a standard to develop artificial protein binders. The present work permitted the design of ten artificial protein binders targeting the catalytic domain of six MMPs, with affinities between 450 nM and 450 μM. Molecular models of these ten artificial protein binders in interaction with MMP-14 indicated that they possess interface areas in the range 700-1000 $Å^2$. These initial protein binder candidates may be used to create libraries focused on specific positions at these interfaces. This would permit selection of variants with improved affinity. Furthermore, differential selection may also be used to identify variants of these artificial protein binders displaying specificity for any particular MMP. Phage display and selection based on the scaffold of the broad-spectrum natural MMP inhibitor TIMP-2 [49] permitted us to obtain variants of this natural protein binder selective for MMP-1. The ten artificial miniprotein binders designed in the present study offer new starting points to develop selective artificial miniprotein inhibitors for various members of the matrixin family.

Experimental procedures

Scaffold search in the PDB

The scaffold search was performed using developmental version 3.1.1 of the STAMPS software (search for three-dimensional atom motifs in protein structures) [16]. This software permits identification of proteins possessing a set of residues in a topology similar to that of the functional motif used as input. STAMPS is able to screen all PDB files (> 75 000 protein structures, April 2012) very efficiently. Some flexibility was introduced via the STAMPS parameters used for the searches. The maximum rmsd for the position of the Ca,Cb atoms of a selected scaffold compared to the corresponding atoms in the selected motif was 1.25 Å (MAXRMSD). The difference between one of the Ca,Cb distances in the selected motif and the

corresponding distance in the motif examined must not exceed 1.85 Å (MAXDDIST). In addition, the maximum deviation between one of the Ca–Cb vectors in the selected motif and the corresponding vector in the examined scaffold must be lower than 130° (MAXANGLE), and the mean deviation for all Ca–Cb angles was limited to 90° (DELTANGLE). The size of the protein scaffolds was limited to a minimum of 15 residues (MINSIZE) and a maximum of 60 residues (MAXSIZE). The intermolecular steric hindrance criterion (STERICERR) allowing evaluation of steric compatibility of the scaffold with the target MMP-14 was set to 2.5. The parameter permitting evaluation of the intra-molecular steric hindrance after automated grafting of the motif residues on the scaffold (STERICINTRA) was set to 1.0. The 3D structures of scaffolds satisfying the above criteria in complex with the target MMP-14 were saved by STAMPS in PDB format. The search was performed using the full PDB.

Molecular modelling refinement of the selected scaffolds

The structures of the MMP-14–scaffold complexes identified by STAMPS were subjected to relaxation by molecular dynamics and energy minimization under positional restraints using CHARMM software version 35b1 [50] and CHARMM force field version 27 [51]. For each complex, after reconstruction of the hydrogen atoms absent in the original PDB structures, the initial structure was minimized under restraints. Standard harmonic restraints were applied to the Cα atoms of the target MMP-14 with a force constant of 50 kcal·mol^{-1}·A^{-2} except for residues of the target located at a distance less than 7 Å from the protein scaffold, for which the harmonic constant was set to 0.5 kcal·mol^{-1}·A^{-2}. The secondary structure of the protein scaffold was preserved using all intra-molecular Ca,Ca distance restraints, with the harmonic constant set to 50 kcal·mol^{-1}·A^{-2}. These distance restraints were defined via the Nuclear Overhauser Effect (NOE) facility of CHARMM. This allowed preservation of the 3D structure of the scaffold without preventing global movement of the scaffold with respect to the target. A set of positional restraints was also applied to the side chain of the grafted functional motif to preserve its position relative to the target. In all cases, we observed that the initial minimization allowed correction of some local steric hindrance between the two partners that resulted from the crude steric evaluation in STAMPS. After 300 initial energy minimization steps, 375 000 molecular dynamics steps (duration 0.0005 ps) at 300 K were performed. These calculations were performed *in vacuo*. After each 25 000 steps, the structure of the complex was quenched using a 3000 steps energy minimization, and the minimized structure was saved for further analysis.

Electrostatic analysis of the protein binder candidates

The electrostatic potentials of the various protein binder-target complexes were calculated using the Poisson–Boltzmann model with APBS software version 1.3 [41]. The coordinates of the structures used in these calculations are those of the X-ray structures after reconstruction of hydrogen atoms and, if necessary, of the missing coordinates of non-hydrogen atoms. Residues were considered to be in a protonation state similar to that of free amino acids without a pK shift at pH 7. After superimposing the structures of the common protein–target complex in each set, the resulting coordinates of the protein binder-target complexes were saved. Then the electrostatic potentials of the protein binder complexes were calculated in the absence of the target. The charge and radius used were those of the PARSE parameter set [52]. The grid dimensions and their spacing were defined in order to encompass all the scaffolds studied. This allowed further comparison of the electrostatic potential using PIPSA software [40]. The electrostatic potential was calculated using the focusing method. The non-linear version of the Poisson–Boltzmann equation was used with an ionic strength of 150 mM, and the radius of these ions was set to 2.0 Å. The dimensions of a coarse grid in the three directions were set to 97 points, with spacing of 2.0 Å. The solvent radius was set to 1.4 Å. The solute and solvent dielectric constants were set to 1 and 80, respectively. During the first calculation, we used the multiple Debye–Hückel boundary conditions. During the focusing step, the grid dimensions were set to 97 points in each direction, the spacing was set to 0.8 Å, and the boundary condition was obtained from the coarse grid calculation. The electrostatic potentials calculated were written in UHBD format [53] to allow further analysis using PIPSA software. The electrostatic potential was compared using version 2 of the PIPSA software. This program calculates the Hodgkin similarity indices to compare the molecular potential M of two proteins a and b [54,55]. The molecular potentials are calculated at point (i, j, k) on a three-dimensional grid superimposed on the molecule analysed, and the Hodgkin similarity index $ESI_H_{a,b}$, is defined as:

$$ESI_H_{a,b} = \frac{2(M_a, M_b)}{(M_a^2 + M_b^2)}$$

In the case of comparison of the electrostatic potentials Φ_a and Φ_b:

$$(M_a, M_b) = \sum_{i,j,k} \Phi_a(i,j,k)\Phi_b(i,j,k)$$

Therefore, the value of ESI_H varies from −1, when the electrostatic potentials are anti-correlated, to 1 when they are identical, with a value of 0 corresponding to uncorrelated potentials.

For greater physical relevance, the points (i, j, k) are selected in a region called 'skin' around each protein. In the present case, the skin is defined as a layer located between a distance of 3 Å (probe radius) and a distance of 7 Å from the van der Waals surface. This corresponded to a layer of thickness 4 Å. The Hodgkin similarity indices for comparison of the electrostatic potential of proteins a and b are calculated using the grid points within the intersection of their skins. The ESI_H calculation is performed either on the complete molecular skins (ESI_H_all) or is restricted to the binding site region (ESI_H_part). This region is defined as the intersection of the skin with a sphere of radius 10 Å centred on the coordinates of the N atom of residue C1 in TIMP-2 (structure 1BQQ).

The Hodgkin similarity index may be used to compare the shape of two proteins a and b. In this case, the shape similarity index $SSI_H_{a,b}$, is defined as:

$$SSI_H_{a,b} = \frac{2(n_{inter})}{(n_a + n_b)}$$

where n_a and n_b are the number of grid points in the skin of proteins a and b, respectively, and n_{inter} is the number of grid points in the intersection of skins of proteins a and b. Therefore, the value of SSI_H varies from 0 when the shapes of the proteins are dissimilar, to 1 when they are identical. This shape similarity index may be calculated in the complete intersection of proteins a and b (SSI_H_all) or may be restricted to the binding region. This region was defined as the intersection of the skins in the binding region (SSI_H_part). This intersection was defined as a sphere of 10 Å centred on the coordinates of the N atom of residue C1 in TIMP-2 (structure 1BQQ).

DNA constructs and protein purification

Synthetic genes optimized for recombinant expression of miniproteins in *E. coli* were obtained from Geneart AG (Regensburg, Germany). These genes contain the sequence of a tobacco etch virus (TEV) protease cleavage site (ENLYFQ), followed by sequence corresponding to miniproteins containing the functional motif, with Gateway recombination sites at each extremity of the gene. These synthetic constructs were cloned by Gateway™ LR (Life Technologies, Grand Island, NY, USA) cloning technology using pETG-82A (EMBL, Heidelberg, Germany) as the destination vector. The resulting expression plasmids encode the fusion protein of interest. Variants of miniproteins were obtained by performing site-directed mutagenesis using the PCR technique. Nucleotide sequences of all expression plasmids were verified by sequencing (Eurofins MWG, Ebersberg, Germany).

For production of recombinant proteins, 1 L of LB medium containing 1 mL ampicillin (100 mg·mL^{-1}) was inoculated with 10 mL of an overnight culture of the BL21 Star™ (DE3), (Life technologies) *Escherichia coli* strain transformed with expression plasmids, and incubated at 37 °C. When the attenuance at 600 nm was 0.8, the culture was incubated at 20 °C for 30 min, and expression of recombinant proteins was induced by adding 1 mL of 1 M isopropyl thio-β-D-galactoside, and incubation was continued overnight at 20 °C. Cells were then harvested by centrifugation (4500 ***g***, 30 min, 4 °C), and the pellet was resuspended in 50 mL lysis buffer (100 mM Tris/HCl, pH 8, 150 mM NaCl, 5% glycerol); cells were then disrupted using a cell disrupter (Constant System Ltd, Daventry, UK). The lysate was cleared by centrifugation (39 000 ***g***, 30 min, 4 °C) and the supernatant was loaded onto a 5 mL HisTrap FF column (GE Healthcare, Freiburg, Germany). The His6-tagged fusion proteins were then eluted using a linear gradient of buffer B (100 mM Tris/HCl, pH 8, 150 mM NaCl, 5% glycerol, 500 mM imidazole) in buffer A (100 mM Tris/HCl, pH 8, 150 mM NaCl, 5% glycerol, 40 mM imidazole) (0–100% B over 30 min at a flow rate of 2 mL·min^{-1}). The fractions containing the His6-tagged fusion protein were pooled and dialysed for 3 h against buffer C (50 mM Tris/HCl, pH 8) using a Spectra/Por® (Spectrum Europe BV, Breda, the Netherlands) dialysis membrane (molecular weight cut-off 3500). The protein of interest was then cleaved using 10% w/w TEV protease overnight at 4 °C. Then the protein of interest was purified by reversed-phase HPLC on a C4 column (Vydac 214TP1010 Grace Discovery Science, Columbia, MD, USA; particle diameter 10 μm; pore size 300 Å; column width and length 10 × 250 mm) using a non-linear gradient of solvent B [90% acetonitrile, 10% H_2O, 0.09% trifluoroacetic acid (TFA)] in solvent A (95% H_2O, 5% acetonitrile 0.1% TFA) at a flow rate of 4 mL·min^{-1}.

Solid-phase peptide synthesis

Proteins 1EDP_A, 2UUY_B, 2F3C_I, 2IT8_A and mutants of 1EDP_A were obtained by solid-phase synthesis on an automated peptide synthesizer (model 433A; Applied Biosystems, Life Technologies, Carlsbad, CA, USA) using the fluorenylmethyloxycarbonyl/tert-butyl ether (Fmoc/tBu) strategy and a modified program. As a standard procedure, 0.1 mmol of Fmoc–C-terminal rink amide methylbenzhydrylamine (MBHA) resin (loading rate 0.66 mmol·g^{-1}), a tenfold excess of each Fmoc-amino acid (Novabiochem, MERCK BIOSCIENCES AG, Läufelfingen, Switzerland), DCCI (dicyclohexylcarbodiimide)/HOBt (1-hydroxybenzotriazole) activation and HBTU (O-(Benzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate)/NMP (N-Methyl-2-pyrrolidone) coupling conditions were used. After drying under vacuum, the miniprotein was liberated from the resin, and protecting groups were removed by acidic cleavage with 10 mL TFA/triisopropylsilane (TIS)/H_2O (95/2.5/2.5 v/v/v) for 2 h at room temperature. The resin was then filtered, and the free miniprotein was precipitated in 100 mL ice-cold methyl *tert*-butyl ether (TBME). After centrifugation (3000 ***g***, 30 min,

4 °C), the pellet was dissolved in 30 mL 10% acetic acid and freeze-dried.

Purification by reversed-phase HPLC

The crude extracts were reduced in denaturing buffer (0.5 M Tris, pH 8, 6 M guanidine) with 10 mM dithiothreitol for 90 min at 37 °C, and the solution was then acidified to pH 2 using TFA/H_2O (50/50 v/v) and subjected to reversed-phase HPLC purification on a Discovery® BIO Wide Pore C18 semi-preparative column (Supelco, Sigma-Aldrich, St Louis, MO, USA; particle diameter 5 μm; pore size 300 Å; column width 10 mm and length 250 mm) using non-linear gradients of solvent B (90% acetonitrile, 10% H_2O, 0.09% TFA) in solvent A (90% H_2O, 10% acetonitrile, 0.1% TFA) at a flow rate of 5 mL·min^{-1}. Detection was performed at 214 nm using a Merck (Whitehouse Station, NJ, USA) L4000 UV detector equipped with a 5 mm cell.

Disulfide pairing

Disulfide pairing was performed in denaturing buffer containing 5 mM reduced and 0.5 mM oxidized glutathione and the protein at a final concentration of 1 mg·mL^{-1}, at 4 °C for 24–48 h. Monitoring of disulfide pairing was performed by following the disappearance of the fully reduced protein by analytical reversed-phase HPLC. The solution was purified as described above.

Characterization of proteins

After HPLC purification, chemically and biologically produced proteins were lyophilized and then solubilized in the appropriate buffer for further studies. Purity of proteins was assessed by analytical HPLC coupled to a high-resolution ion trap mass spectrometer (Esquire HCT; Bruker Daltonics, Billerica, MA, USA). The correct mass after disulfide bond formation was confirmed by MALDI-TOF/TOF 4800 analysis (Applied Biosystems), and correct disulfide pairing was checked by binding assays using natural targets of miniproteins, circular dichroism or crystallography. Concentrations of miniproteins were determined using a JLC-500/V AminoTac amino acid analyser (JEOL, Tokyo, Japan) and absorbance measurement at 280 nm on a Beckmann (Fullerton, CA, USA) spectrometer using a molar extinction coefficient calculated on the basis of the known amino acid content. Disulfide pairing was tested indirectly by checking the native function of the scaffold when possible and when the grafting did not entail mutation of residues involved in the native function. Scaffold 2F3C_I is a trypsin inhibitor. We observed that 2F3C_I_m inhibits trypsin with a K_i value of 0.5 ± 0.1 nM, which is close to the value reported in the literature (2 nM) [56]. Similarly, scaffold 1AN1_I is a trypsin inhibitor. We found that 1AN1_I_m was able to inhibit trypsin with a K_i value of 1.7 ± 0.2 nM, consistent with the value reported in the literature (1.8 nM [57]). The K_i value determined for inhibition of trypsin activity by 2UUY_B_m was 3.0 ± 0.2 nM, again consistent with the value reported in the literature (< 10 nM [58]). Finally, the inhibition constant for 2IT8_A_m was 73.0 ± 0.2 nM, a factor that is ~20 times higher than the value reported in the literature [59]. However, in this case, three residues involved in the native function of the scaffold were mutated for grafting the functional motif. Nevertheless, the high inhibition constant for trypsin suggests that the mutated protein is correctly folded.

MMP inhibition assays

MMP inhibition by the miniproteins was tested by competition experiments using 1.8 μM of the fluorogenic substrate McaMat (Mca-Pro-Leu-Gly-Leu-Dpa-Ala-Arg-NH_2, Calcobiochem, Merck) and human MMPs in a nanomolar range concentration (from R & D Systems, Minneapolis, MN, USA, or produced in the laboratory as described previously [60]), in 50 mM Tris/HCl buffer pH 6.8, 10 mM $CaCl_2$. The substrate and enzyme concentrations for the experiments were chosen so as the substrate utilization remains well below 10% of the total amount and to observe the initial rates. For each inhibitor, the percentage of inhibition was determined in duplicate experiments. Four inhibitor concentrations were chosen to observe a 20–80% range of inhibition. After 45 min incubation under shaking at 25 °C, inhibition assays were performed by recording the fluorescence increase induced by cleavage of the fluorogenic substrate, using 96-well non-binding surface plates (Corning-Costar, Lowell, MA, USA). Fluorescence signals were monitored using a Fluoroscan Ascent photon counter spectrophotometer (Thermolab) equipped with a temperature control device and a plate shaker. K_i values were determined as described previously [61].

Crystallization

The protein solution for crystallization consisted of 643 μM of the catalytic domain of the F67D mutant of human MMP-12 (residues 106–263) with 1 mM acetohydroxamic acid to prevent self-degradation of the proteinase during crystallization in 0.1 M glycine pH 8.5 and 27% polyethylene glycol 10 000. Crystals of MMP-12 were grown in the presence of acetohydroxamic acid by sitting-drop vapor diffusion experiments at 20 °C in a cooled incubator. The crystals were harvested and transferred to a cryo-solution (90 mM Tris/HCl pH 8, 27% polyethylene glycol 8000, 15% monomethyl ether/polyethylene glycol 550, 10% glycerol) for 5 min to remove acetohydroxamic acid, then transferred to the same cryo-solution with 1 mM 1EDP_A for 5 min before the crystals were picked up with a loop and plunged into liquid nitrogen.

Structure determination and refinement

The data for crystals were collected at the European Synchrotron Radiation Facility (Grenoble, France) on beam line ID14-2 at 100 K from a single crystal. Most MMP-12 crystals soaked with 1EDP_A were badly disordered, and only one crystal gave usable data to only 2.8 Å resolution. Similar crystals that had followed the same soaking procedure but with small nanomolar inhibitors diffracted to 1.6–2.3 Å. Data reduction was performed using MOSFLM [62]. The crystals belong to tetragonal space group $P2_12_12$, with cell parameters $a = 69.1$ Å, $b = 63.2$ Å, $c = 37.0$ Å, with one molecule in the asymmetric unit. The structure was solved by rigid-body refinement using REFMAC [63], starting with PBD entry 3LIR [64]. The CCP4 suite of programs [65] was used for manipulation of structure factors and coordinates. Final fitting and stereochemical analysis of the refined model were performed using COOT [66]. The figures were created using PyMOL (Schrödinger Inc., New York, NY, USA).

Molecular dynamics simulation in water solution

The calculations were carried using CHARMM version 35b1 [50] and version 27 of the CHARMM parameter set for proteins [51]. For each artificial binder, the initial MMP-14 complex was obtained after relaxation by energy minimization and molecular dynamics under positional restraints *in vacuo* of the complexes identified by STAMPS (see above). This structure was immersed in a cubic box of pre-equilibrated TIP3P water model. After removal of the water molecules overlapping the protein complex, a short minimization was performed to optimize the water–protein interface. During this step, the coordinates of the proteins were fixed. This was followed by a second energy minimization under positional restraints applied to the protein atoms. Then, molecular dynamics simulation was performed. The system was heated from 0 to 298 K in 100 ps in steps of 50 K. During this equilibration step, harmonic restraints (5 kcal·mol^{-1}·Å^{-2}) were applied to the positions of zinc, calcium and non-hydrogen backbone atoms. After equilibration, all positional restraints were removed. Molecular dynamics simulation was performed using the leapfrog Verlet algorithm at constant pressure (1 atm) and temperature (298 K). Electrostatic interactions were calculated using the particle mesh Ewald summation method. The time step was set to 1 fs, and the coordinates were saved every 0.1 ps.

Acknowledgements

Gael Debret is gratefully acknowledged for her work in development of the STAMPS software.

References

1 Gebauer M & Skerra A (2009) Engineered protein scaffolds as next-generation antibody therapeutics. *Curr Opin Cheml Biol* **13**, 245–255.
2 Beck A, Wurch T, Bailly C & Corvaia N (2010) Strategies and challenges for the next generation of therapeutic antibodies. *Nat Rev Immunol* **10**, 345–352.
3 DeLano WL, Ultsch MH, de Vos AM & Wells JA (2000) Convergent solutions to binding at a protein–protein interface. *Science* **287**, 1279–1283.
4 Vita C, Drakopoulou E, Vizzavona J, Rochette S, Martin L, Menez A, Roumestand C, Yang YS, Ylisastigui L, Benjouad A *et al.* (1999) Rational engineering of a miniprotein that reproduces the core of the CD4 site interacting with HIV-1 envelope glycoprotein. *Proc Natl Acad Sci USA* **96**, 13091–13096.
5 Domingues H, Cregut D, Sebald W, Oschkinat H & Serrano L (1999) Rational design of a GCN4-derived mimetic of interleukin-4. *Nat Struct Biol* **6**, 652–656.
6 Martin L, Stricher F, Misse D, Sironi F, Pugniere M, Barthe P, Prado-Gotor R, Freulon I, Magne X, Roumestand C *et al.* (2003) Rational design of a CD4 mimic that inhibits HIV-1 entry and exposes cryptic neutralization epitopes. *Nat Biotechnol* **21**, 71–76.
7 Sia SK & Kim PS (2003) Protein grafting of an HIV-1-inhibiting epitope. *Proc Natl Acad Sci USA* **100**, 9756–9761.
8 Cobos ES, Pisabarro MT, Vega MC, Lacroix E, Serrano L, Ruiz-Sanz J & Martinez JC (2004) A miniprotein scaffold used to assemble the polyproline II binding epitope recognized by SH3 domains. *J Mol Biol* **342**, 355–365.
9 Li C, Liu M, Monbo J, Zou GZ, Li CQ, Yuan WR, Zella D, Lu WY & Lu WY (2008) Turning a scorpion toxin into an antitumor miniprotein. *J Am Chem Soc* **130**, 13546–13548.
10 Azoitei ML, Correia BE, Ban YEA, Carrico C, Kalyuzhniy O, Chen L, Schroeter A, Huang PS, McLellan JS, Kwong PD *et al.* (2011) Computation-guided backbone grafting of a discontinuous motif onto a protein scaffold. *Science* **334**, 373–376.
11 Azoitei ML, Ban YEA, Julien JP, Bryson S, Schroeter A, Kalyuzhniy O, Porter JR, Adachi Y, Baker D, Pai EF *et al.* (2012) Computational design of high-affinity epitope scaffolds by backbone grafting of a linear epitope. *J Mol Biol* **415**, 175–192.
12 Liang SD, Li WZ, Xiao L, Wang JS & Lai LH (2000) Grafting of protein–protein interaction epitope. *J Biomol Struct Dynamics* **17**, 821–828.
13 Magis C, Gasparini D, Lecoq A, Le Du MH, Stura E, Charbonnier JB, Mourier G, Boulain JC, Pardo L, Caruana A *et al.* (2006) Structure-based secondary structure-independent approach to design

protein ligands: application to the design of Kv1.2 potassium channel blockers. *J Am Chem Soc* **128**, 16190–16205.
14 Keskin O & Nussinov R (2005) Favorable scaffolds: proteins with different sequence, structure and function may associate in similar ways. *Protein Eng Des Sel* **18**, 11–24.
15 Keskin O & Nussinov R (2007) Similar binding sites and different partners: implications to shared proteins in cellular pathways. *Structure* **15**, 341–354.
16 Debret G, Martel A & Cuniasse P (2009) Rasmot-3d Pro: a 3D motif search webserver. *Nucleic Acids Res* **37**, W459–W464.
17 Zhang CS & Lai LH (2012) AutoMatch: target-binding protein design and enzyme design by automatic pinpointing potential active sites in available protein scaffolds. *Proteins* **80**, 1078–1094.
18 Liu S, Liu S, Zhu X, Liang H, Cao A, Chang Z & Lai L (2007) Nonnatural protein–protein interaction-pair design by key residues grafting. *Proc Natl Acad Sci USA* **104**, 5330–5335.
19 Fleishman SJ, Whitehead TA, Ekiert DC, Dreyfus C, Corn JE, Strauch EM, Wilson IA & Baker D (2011) Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* **332**, 816–821.
20 McCawley LJ & Matrisian LM (2001) Matrix metalloproteinases: they're not just for matrix anymore! *Curr Opin Cell Biol* **13**, 534–540.
21 Rydlova M, Holubec L, Ludvikova M, Kalfert D, Franekova J, Povysil C & Ludvikova M (2008) Biological activity and clinical implications of the matrix metalloproteinases. *Anticancer Res* **28**, 1389–1397.
22 Whittaker M, Floyd CD, Brown P & Gearing AJH (1999) Design and therapeutic application of matrix metalloproteinase inhibitors. *Chem Rev* **99**, 2735–2776.
23 Cuniasse P, Devel L, Makaritis A, Beau F, Georgiadis D, Matziari A, Yiotakis A & Dive V (2005) Future challenges facing the development of specific active-site-directed synthetic inhibitors of MMPs. *Biochimie* **87**, 393–402.
24 Fisher JF & Mobashery S (2006) Recent advances in MMP inhibitor design. *Cancer Metast Rev* **25**, 115–136.
25 Tu GG, Xu WF, Huang HM & Li SH (2008) Progress in the development of matrix metalloproteinase inhibitors. *Curr Med Chem* **15**, 1388–1395.
26 Yiotakis A & Dive V (2008) Synthetic active site-directed inhibitors of metzincins: achievement and perspectives. *Mol Aspects Med* **29**, 329–338.
27 Bode W (2003) Structural basis of matrix metalloproteinase function. *Biochem Soc Symp* **70**, 1–14.
28 Tallant C, Marrero A & Gomis-Ruth FX (2010) Matrix metalloproteinases: fold and function of their catalytic domains. *Biochim Biophys Acta* **1803**, 20–28.
29 Brew K, Dinakarpandian D & Nagase H (2000) Tissue inhibitors of metalloproteinases: evolution, structure and function. *Biochim Biophys Acta* **1477**, 267–283.
30 Butler GS, Hutton M, Wattam BA, Williamson RA, Knauper V, Willenbrock F & Murphy G (1999) The specificity of TIMP-2 for matrix metalloproteinases can be modified by single amino acid mutations. *J Biol Chem* **274**, 20391–20396.
31 Meng Q, Malinovskii V, Huang W, Hu YJ, Chung L, Nagase H, Bode W, Maskos K & Brew K (1999) Residue 2 of TIMP-1 is a major determinant of affinity and specificity for matrix metalloproteinases but effects of substitutions do not correlate with those of the corresponding P1' residue of substrate. *J Biol Chem* **274**, 10184–10189.
32 Wei S, Chen Y, Chung L, Nagase H & Brew K (2003) Protein engineering of the tissue inhibitor of metalloproteinase 1 (TIMP-1) inhibitory domain – in search of selective matrix metalloproteinase inhibitors. *J Biol Chem* **278**, 9831–9834.
33 Hamze AB, Wei S, Bahudhanapati H, Kota S, Acharya KR & Brew K (2007) Constraining specificity in the N-domain of tissue inhibitor of metalloproteinases-1; gelatinase-selective inhibitors. *Protein Sci* **16**, 1905–1913.
34 GomisRuth FX, Maskos K, Betz M, Bergner A, Huber R, Suzuki K, Yoshida N, Nagase H, Brew K, Bourenkov GP *et al.* (1997) Mechanism of inhibition of the human matrix metalloproteinase stromelysin-1 by TIMP-1. *Nature* **389**, 77–81.
35 Fernandez-Catalan C, Bode W, Huber R, Turk D, Calvete JJ, Lichte A, Tschesche H & Maskos K (1998) Crystal structure of the complex formed by the membrane type 1-matrix metalloproteinase with the tissue inhibitor of metalloproteinases-2, the soluble progelatinase A receptor. *EMBO J* **17**, 5238–5248.
36 Morgunova E, Tuuttila A, Bergmann U & Tryggvason K (2002) Structural insight into the complex formation of latent matrix metalloproteinase 2 with tissue inhibitor of metalloproteinase 2. *Proc Natl Acad Sci USA* **99**, 7414–7419.
37 Maskos K, Lang R, Tschesche H & Bode W (2007) Flexibility and variability of TIMP binding: X-ray structure of the complex between collagenase-3/MMP-13 and TIMP-2. *J Mol Biol* **366**, 1222–1231.
38 Higashi S & Miyazaki K (1999) Reactive site-modified tissue inhibitor of metalloproteinases-2 inhibits the cell-mediated activation of progelatinase A. *J Biol Chem* **274**, 10497–10504.
39 Troeberg L, Tanaka M, Wait R, Shi YE, Brew K & Nagase H (2002) *E. coli* expression of TIMP-4 and comparative kinetic studies with TIMP-1 and TIMP-2: insights into the interactions of TIMPs and matrix

metalloproteinase 2 (gelatinase A) . *Biochemistry* **41**, 15025–15035.
40 Blomberg N, Gabdoulline RR, Nilges M & Wade RC (1999) Classification of protein sequences by homology modeling and quantitative analysis of electrostatic similarity. *Proteins* **37**, 379–387.
41 Baker NA, Sept D, Joseph S, Holst MJ & McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA* **98**, 10037–10041.
42 Lo Conte L, Chothia C & Janin J (1999) The atomic structure of protein–protein recognition sites. *J Mol Biol* **285**, 2177–2198.
43 Jones S & Thornton JM (1996) Principles of protein–protein interactions. *Proc Natl Acad Sci USA* **93**, 13–20.
44 Xu D, Tsai CJ & Nussinov R (1997) Hydrogen bonds and salt bridges across protein–protein interfaces. *Protein Eng* **10**, 999–1012.
45 Guerois R, Nielsen JE & Serrano L (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol* **320**, 369–387.
46 Peisajovich SG & Tawfik DS (2007) Protein engineers turned evolutionists. *Nat Methods* **4**, 991–994.
47 Lauer-Fields JL, Cudic M, Wei S, Mari F, Fields GB & Brew K (2007) Engineered sarafotoxins as tissue inhibitor of metalloproteinases-like matrix metalloproteinase inhibitors. *J Biol Chem* **282**, 26948–26955.
48 Huang CC, Stricher F, Martin L, Decker JM, Majeed S, Barthe P, Hendrickson WA, Robinson J, Roumestand C, Sodroski J *et al.* (2005) Scorpion-toxin mimics of CD4 in complex with human immunodeficiency virus gp120: crystal structures, molecular mimicry, and neutralization breadth. *Structure* **13**, 755–768.
49 Bahudhanapati H, Zhang YN, Sidhu SS & Brew K (2011) Phage display of tissue inhibitor of metalloproteinases-2 (TIMP-2): identification of selective inhibitors of collagenase-1 (metalloproteinase 1 (MMP-1)). *J Biol Chem* **286**, 31761–31770.
50 Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S & Karplus M (1983) CHARMM – a program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* **4**, 187–217.
51 MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S *et al.* (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* **102**, 3586–3616.
52 Sitkoff D, Sharp KA & Honig B (1994) Accurate calculation of hydration free-energies using macroscopic solvent models. *J Phys Chem* **98**, 1978–1988.
53 Madura JD, Briggs JM, Wade RC, Davis ME, Luty BA, Ilin A, Antosiewicz J, Gilson MK, Bagheri B, Scott LR *et al.* (1995) Electrostatics and diffusion of molecules in solution – simulations with the University of Houston Brownian dynamics program. *Comput Phys Commun* **91**, 57–95.
54 Wade RC, Gabdoulline RR & De Rienzo F (2001) Protein interaction property similarity analysis. *Int J Quantum Chem* **83**, 122–127.
55 Hodgkin EE & Richards WG (1987) Molecular similarity based on electrostatic potential and electric field. *Int J Quantum Chem* **32**, 105–110.
56 Campos ITN, Amino R, Sampaio CAM, Auerswald EA, Friedrich T, Lemaire HG, Schenkman S & Tanaka AS (2002) Infestin, a thrombin inhibitor presents in *Triatoma infestans* midgut, a Chagas' disease vector: gene cloning, expression and characterization of the inhibitor. *Insect Biochem Mol* **32**, 991–997.
57 Stubbs MT, Morenweiser R, Sturzebecher J, Bauer M, Bode W, Huber R, Piechottka GP, Matschiner G, Sommerhoff CP, Fritz H *et al.* (1997) The three-dimensional structure of recombinant leech-derived tryptase inhibitor in complex with trypsin – implications for the structure of human mast cell tryptase and its inhibition. *J Biol Chem* **272**, 19931–19937.
58 Valdez-Cruz NA, Conde R, Zamudio F & Possani LD (2011) Anticoagulants from scorpion venoms. In Toxins and Homeostasis: From Bench to Bedside (Marklan FS, ed.), pp. 255–266, Springer, NY, USA.
59 Sommerhoff CP, Avrutina O, Schmoldt HU, Gabrijelcic-Geiger D, Diederichsen U & Kolmar H (2010) Engineered cystine knot miniproteins as potent inhibitors of human mast cell tryptase β. *J Mol Biol* **395**, 167–175.
60 Devel L, Rogakos V, David A, Makaritis A, Beau F, Cuniasse P, Yiotakis A & Dive V (2006) Development of selective inhibitors and substrate of matrix metalloproteinase-12. *J Biol Chem* **281**, 11152–11160.
61 Horovitz A & Levitzki A (1987) An accurate method for determination of receptor ligand and enzyme-inhibitor dissociation constants from displacement curves. *Proc Natl Acad Sci USA* **84**, 6654–6658.
62 Leslie AGW (2006) The integration of macromolecular diffraction data. *Acta Crystallogr D* **62**, 48–57.
63 Murshudov GN, Vagin AA & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D* **53**, 240–255.
64 Devel L, Garcia S, Czarny B, Beau F, Lajeunesse E, Vera L, Georgiadis D, Stura E & Dive V (2010) Insights from selective non-phosphinic inhibitors of MMP-12 tailored to fit with an S-1' loop canonical conformation. *J Biol Chem* **285**, 35900–35909.
65 Bailey S (1994) The Ccp4 suite – programs for protein crystallography. *Acta Crystallogr D* **50**, 760–763.

66 Emsley P, Lohkamp B, Scott WG & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D* **66**, 486–501.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site:

Table S1. STAMPS output for the two searches of the PDB.

Table S2. Mean inter-atom distances determined in molecular dynamics simulations of the complexes between the protein binder candidates and the catalytic domain of MMP-14.

REFERENCES BIBLIOGRAPHIQUES

Adams GP, Schier R, McCall AM, Simmons HH, Horak EM, Alpaugh RK et al. (2001) High affinity restricts the localization and tumor penetration of single-chain fv antibody molecules. *Cancer Res*; 61: 4750-4755.

Albrand, J.-P., Blackledge, M. J., Pascaud, F., Hollecker, M. and Marion, D. (1995) NMR and restrained molecular dynamics study of the three-dimensional solution structure of toxin FS2, a specific blocker of the L-type calcium channel, isolated from black mamba venom. *Biochemistry*; 34: 5923-5937.

Almagro J.C. and Fransson J. (2008) Humanization of antibodies. *Front. Biosci*.; 13: 1619-1633.

Amour A., Slocombe PM., Webster A., Butler M., Knight CG., Smith BJ., Stephens PE., Shelley C., Hutton M., Knäuper V., Docherty AJP. and Murphy G. (1998) TNF-Alpha converting enzyme (TACE) is inhibited by TIMP-3. *FEBS Lett*; 435: 39-44.

Amrein, B., Schmid, M., Collet, G., Cuniasse, P., Gilardoni, F., Seebeck, F.P., and Ward, T.R. (2012) Identification of two-histidines one-carboxylate binding motifs in proteins amenable to facial coordination to metals. *Metallomics*; 4: 379-388.

Andarawewa KL, Boulay A, Masson R, Mathelin C, Stoll I, Tomasetto C, Chenard MP, Gintz M, Bellocq JP, Rio MC. (2003) Dual stromelysin-3 function during natural mouse mammary tumor virus-ras tumor progression. *Cancer Res*; 63: 5844-5849.

Antil-Delbeke, S., Gaillard C., Tamiya T., Corringer P.J., Changeux J.P., Servent D. and Ménez A. (2001) Molecular determinants by which long chain toxins from snake venom interacts with the neuronal α7-nicotinic acetylcholine receptor. *J. Biol. Chem.;* 275: 29594-29601.

Apte S.S., Olsen B.R. and Murphy G. (1995) The gene structure of tissue inhibitor of metalloproteinases (TIMP)-3 and its inhibitory activities define the distinct TIMP gene family. *J. Biol. Chem.*; 270: 14313-14318.

Artymiuk, P. J., Poirrette, A.R., Grindley, H.M., Rice, D.W. and Willett, P. (1994) A graph-theoretic approach to the identification of three-dimensional patterns of amino acid side-chains in protein structures. *J. Mol. Biol*.; 243: 327-344.

Arumugam, S. and Van Doren, S.R. (2003) Global orientation of bound MMP-3 and N-TIMP-1 in solution via residual dipolar couplings. *Biochemistry*; 42: 7950-7958.

Azoitei ML, Ban YE, Julien JP, Bryson S, Schroeter A, Kalyuzhniy O, Porter JR, Adachi Y,

Baker D, Pai EF, Schief WR. (2012) Computational design of high-affinity epitope scaffolds by backbone grafting of a linear epitope. *J Mol Biol*.; 415: 175-192.

Azoitei ML, Correia BE, Ban YE, Carrico C, Kalyuzhniy O, Chen L et al. (2011) Computation-guided backbone grafting of a discontinuous motif onto a protein scaffold. *Science*; 334: 373-376.

Bahudhanapati, H., Zhang, Y.N., Sidhu, S.S. and Brew, K. (2011) Phage display of tissue inhibitor of metalloproteinases-2 (TIMP-2): identification of selective inhibitors of collagenase-1 (metalloproteinase 1 (MMP-1)). *Journal of Biological Chemistry*; 286(36): 31761-31770.

Bailey S. (1994) The Ccp4 Suite - Programs for Protein Crystallography. *Acta Crystallogr D*; 50: 760-763.

Baker M, and Jones TD. (2007) The identification and removal of immunogenicity in therapeutic proteins. *Curr. Opin. Drug Discovery Dev*.; 10(2): 219-227.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA*; 98: 10037-10041.

Batra, J., Robinson, J., Soares, A.S., Fields, A.P., Radisky, D.C., Radisky, E.S. (2012) Matrix metalloproteinase-10 (MMP-10) interaction with tissue inhibitors of metalloproteinases TIMP-1 and TIMP-2: binding studies and crystal structure. *J. Biol. Chem*.; 287(19): 15935-15946.

Beckman RA., Weiner LM. and Davis HM. (2007) Antibody constructs in cancer therapy: protein engineering strategies to improve exposure in solid tumors. *Cancer*; 109: 170-179.

Blomberg, N., Gabdoulline, R.R., Nilges, M., and Wade, R.C. (1999) Classification of protein sequences by homology modeling and quantitative analysis of electrostatic similarity. *Proteins*; 37: 379-387.

Bloom, L. and Calabro, V. (2009) FN3: a new protein scaffold reaches the clinic. *Drug Discov. Today*; 14: 949-955.

Bode W., Gomis-Rüth FX. and Stocker W. (1993) Astacins, serralysins, snake venom and matrix metalloproteinases exhibit identical zinc-binding environments (HEXXHXXGXXH and Met-turn) and topologies and should be grouped into a common family, the 'metzincins'. *FEBS Lett*; 331: 134-40.

Bode W. (1995) A helping hand for collagenases: the haemopexin-like domain. *Structure*; 3(6): 527-530.

Bode, W., Schwager, P. and Huber, R. (1978) The transition of bovine trypsinogen to a trypsin-like state upon strong ligand binding. *J. Mol. Biol*.; 118: 99-112.

Bogan AA and Thorn KS (1998) Anatomy of hot spots in protein interfaces. *J Mol Biol*; 280: 1-9.

Bolon, D.N., Grant, R.A., Baker, T.A. and Sauer, R.T. (2005) Specificity versus stability in computational protein design. *Proc. Natl Acad. Sci. USA*; 102: 12724-12729.

Botti, S.A., Felder, C.E., Sussman, J.L. and Silman, I. (1998) Electrotactins: a class of adhesion proteins with conserved electrostatic and structural motifs. *Protein Eng*; 11: 415-420.

Branson B.M. (2004) FDA approves OraQuick for use in saliva. On March 25, the FDA approved the first rapid test for HIV in oral fluids. *AIDS Clin. Care*; 16(5): 39.

Brew K., Dinakarpandian D. and Nagase H. (2000) Tissue inhibitors of metalloproteinases: evolution, structure and function. *Biochim. Biophys. Acta*; 1477: 267-283.

Brillard-Bourdet, M., Nguyên, V., Ferrer-di Martino, M., Gauthier, F. and Moreau, T. (1998) Purification and characterization of a new cystatin inhibitor from Taiwan cobra (*Naja naja atra*) venom. *Biochem J*.; 331(1): 239-244.

Brinckerhoff C.E. and Matrisian L.M. (2002) Matrix Metalloproteinases: a tail of a frog that became a prince. Nature Reviews: *Molecular Cell Biology*; 3: 207-214.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM - A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem*.; 4: 187-217.

Bryson, S., Julien, J.P., Hynes, R.C. and Pai, E.F. (2009) Crystallographic definition of the epitope promiscuity of the broadly neutralizing anti-human immunodeficiency virus type 1 antibody 2F5: vaccine design implications. *J. Virol*.; 83: 11862-11875.

Burrage PS., Mix KS. and Brinckerhoff CE. (2006) Matrix metalloproteinases: role in arthritis. *Front Biosci*; 11*:* 529-543.

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.

Butler, G.S., Hutton, M., Wattam, B.A., Williamson, R.A., Knauper, V., Willenbrock, F., and Murphy, G. (1999) The specificity of TIMP-2 for matrix metalloproteinases can be modified by single amino acid mutations. *Journal of Biological Chemistry*; 274: 20391-20396.

Cabrera H.A. (1969) A comprehensive evaluation of pregnancy tests. *Am. J. Obstet. Gynecol*.; 103: 32-38.

Campos IT, Amino R, Sampaio CA, Auerswald EA, Friedrich T, Lemaire HG, Schenkman S, Tanaka AS. (2002) Infestin, a thrombin inhibitor presents in Triatoma infestans midgut, a Chagas' disease vector: gene cloning, expression and characterization of the inhibitor. *Insect Biochem Mol Biol*.; 32(9): 991-997.

Capon D.J., S.M. Chamow, J. Mordenti, S.A. Marsters, T. Gregory, H. Mitsuya, et al. (1989) Designing CD4 immunoadhesins for AIDS therapy. *Nature*; 337: 525-531.

Caravella J. and Lugovskoy A. (2010) Design of next-generation protein therapeutics. *Current Opinion in Chemical Biology*; 14: 520-528.

Carter P, Kelley RF, Rodrigues ML, Snedecor B, Covarrubias M, Velligan MD, Wong WL, Rowland AM, Kotts CE, Carver ME, et al. (1992) High level Escherichia coli expression and production of a bivalent humanized antibody fragment. *Biotechnology*; 10(2): 163-167.

Carter, P.J. (2006) Potent antibody therapeutics by design. *Nat. Rev. Immunol*.; 6: 343-357.

Carter PJ. (2011) Introduction to current and future protein therapeutics: a protein engineering perspective. *Exp Cell Res*; 317(9): 1261-1269.

Chakrabarti P. and Janin J. (2002) Dissecting protein–protein recognition sites. *Proteins: Struct. Funct. Genet*.; 47: 334-343.

Chames P., Van Regenmortel M., Weiss E. and Baty D. (2009) Therapeutic antibodies: successes, limitations and hopes for the future. *Br J Pharmacol*; 157: 220-233.

Cheetham JC, Smith DM, Aoki KH, Stevenson JL, Hoeffel TJ, Syed RS, Egrie J, Harvey TS. (1998) NMR structure of human erythropoietin and a comparison with its receptor bound conformation. *Nat Struct Biol*.; (10): 861-866.

Chothia, C. and Janin, J. (1975) Principles of protein-protein recognition. *Nature*; 256: 705-708.

Cines, D.B., Yasothan, U. and Kirkpatrick, P. (2008) Romiplostim. *Nat. Rev. Drug Discov*.; 7:

887-888.

Cizkova D., Marsala J., Lukacova N., Marsala M., Jergova S., Orendacova J. and Yaksh TL. (2002) Localization of N-type Ca2+ channels in the rat spinal cord following chronic constrictive nerve injury. *Exp. Brain. Res*.; 147: 456-463.

Clackson T. and Wells JA. (1995) A hot spot of binding energy in a hormone-receptor interface. *Science*; 267: 383-386.

Cortez-Retamozo, V., Backmann, N., Senter, P.D., Wernery, U., De Baetselier, P., Muyldermans, S. and Revets, H. (2004) Efficient cancer therapy with a nanobody-based conjugate. *Cancer Res*.; 64: 2853-2857.

Coussens LM, Tinkle CL, Hanahan D, et al. (2000) MMP-9 supplied by bone marrow-derived cells contributes to skin carcinogenesis. *Cell*; 103: 481-490.

Crowley P.B. and Golovin A. (2005) Cation-pi interactions in protein-protein interfaces. *Proteins*; 59(2): 231-239.

Dahiyat B.I. and Mayo S.L. (1997) De novo protein design: fully automated sequence selection. *Science*; 278: 82-87.

Davis M.E. and McCammon J.A. (1990) Electrostatics in biomolecular structure and dynamics. *Chem. Rev.*; 90(3): 509-521.

De Rienzo F., Gabdoulline RR., Menziani MC. and Wade RC. (2000) Blue copper proteins: a comparative analysis of their molecular interaction properties. *Protein Sci*.; 9(8): 1439-1454.

Debret G, Martel A, and Cuniasse P. (2009) RASMOT-3D PRO: a 3D motif search webserver. *Nucleic Acids Res*.; 37: 459-464.

DeLano W.L. (2002) The PyMOL Molecular Graphics System, DeLano Scientific, San Carlos, CA.

DeLano W.L., Ultsch M.H., de Vos A.M. and Wells J.A. (2000) Convergent solutions to binding at a protein–protein interface. *Science*; 287: 1279-1283.

Demchuk, E., Mueller, T., Oschkinat, H., Sebald, W. and Wade, R.C. (1994) Receptor-Binding Properties of 4-Helix-Bundle Growth-Factors Deduced from Electrostatic Analysis. *Protein Sci*; 3: 920-935.

Devel L., Rogakos V., David A., Makaritis A., Beau F., Cuniasse P., Yiotakis A. and Dive V.

(2006) Development of selective inhibitors and substrate of matrix metalloproteinase-12. *The Journal of biological chemistry*; 281: 11152-11160.

Devel, L., Garcia, S., Czarny, B., Beau, F., Lajeunesse, E., Vera, L., Georgiadis, D., Stura, E., and Dive, V. (2010) Insights from Selective Non-phosphinic Inhibitors of MMP-12 Tailored to Fit with an S-1' Loop Canonical Conformation. *Journal of Biological Chemistry*; 285: 35900-35909.

De Weille JR., Schweitz H., Maes P., Tartar A., Lazdunski M. (1991) Calciseptine, a peptide isolated from black mamba venom, is a specific blocker of the L-type calcium channel. *Proc. Natl. Acad. Sci. USA*; 88: 2437-2440.

Dickinson, C.D., Veerapandian, B., Dai, X.P., Hamlin, R.C., Xuong, N.H., Ruoslahti, E. and Ely, K.R. (1994) Crystal structure of the tenth type III cell adhesion module of human fibronectin. *J. Mol. Biol*.; 236: 1079-1092.

Dineen SP., Sullivan LA., Beck AW., Miller AF., Carbon JG., Mamluk R., Wong H. and Brekken RA. (2008) The Adnectin CT-322 is a novel VEGF receptor 2 inhibitor that decreases tumor burden in an orthotopic mouse model of pancreatic cancer. *BMC Cancer*; 8: 352.

Dollery CM. and Libby P. (2006*)* Atherosclerosis and proteinase activation. *Cardiovasc Res*; 69: 625-635.

Drakopoulou, E., Vizzavona, J. and Vita, C. (1998) Engineering a CD4 mimetic inhibiting the binding of the human immunodeficiency virus-1 (HIV-1) envelope glycoprotein gp120 to human lymphocyte CD4 by the transfer of a CD4 functional site to a small natural scaffold. *Lett. Pep. Sci*.; 5: 241-245.

Earl ST., Masci PP., de Jersey J., Lavin MF. and Dixon J. (2012) Drug development from Australian elapid snake venoms and the Venomics pipeline of candidates for haemostasis: Textilinin-1 (Q8008), Haempatch™ (Q8009) and CoVase™ (V0801). *Toxicon*.; 59(4): 456-463.

Egeblad M. and Werb Z. (2002) New functions for the matrix metalloproteinases in cancer progression. *Nat*. *Rev*. *Cancer*; 2*:* 161-174.

Emanuel, S.L., Engle, L.J., Cao, C., et al. (2010) Adnectins as a platform for multi-specific targeted biologics: a novel bispecific inhibitor of EGFR and IGF-IR growth factor receptors. 101st Annual Meeting of American Association of Cancer Research (17-21 April, Philadelphia), Abst 2586.

Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of

Coot. *Acta Crystallogr D*; 66: 486-501.

Eng, J., Kleinman, W.A., Singh, L., Singh, G. and Raufman, J.P. (1992) Isolation and characterization of exendin-4, an exendin-3 analogue, from *Heloderma suspectum* venom. Further evidence for an exendin receptor on dispersed acini from guinea pig pancreas. *J. Biol. Chem.;* 267: 7402-7405.

Engel CK., Pirard B., Schimanski S., Kirsch R., Habermann J., Klingler O., Schlotte V., Weithmann KU. and Wendt KU. (2005) Structural basis for the highly selective inhibition of MMP-13. *Chemistry & biology*; 12: 181-189.

Escalante, T., Franceschi, A., Rucavado, A. and Gutierrez, J.M. (2000) Effectiveness of batimastat, a synthetic inhibitor of matrix metalloproteinases, in neutralizing local tissue damage induced by BaP1, a hemorrhagic metalloproteinase from the venom of the snake *Bothrops asper*. *Biochem. Pharmacol*.; 60: 269-274.

Escoubas, P., De Weille J.R., Lecoq A., Diochot S., Waldmann R., Champigny G., Moinier D., Ménez A. and Lazdunski M. (2000) Isolation of a tarantula toxin specific for a class of proton-gated Na+ channels. *J. Biol. Chem.;* 275: 25116-25121.

Fernandez-Catalan C., W. Bode, R. Huber, D. Turk, J.J. Calvete, A. Lichte, et al., (1998) Crystal structure of the complex formed by the membrane type 1-matrix metalloproteinase with the tissue inhibitor of metalloproteinases-2, the soluble progelatinase A receptor. *EMBO J*.; 17: 5238-5248.

Filippovich I., Sorokina N., Masci PP., de Jersey J., Whitaker AN., Winzor DJ., Gaffney PJ. and Lavin MF. (2002) A family of textilinin genes, two of which encode proteins with antihaemorrhagic properties. *Br. J. Haematol.*; 119: 376-384.

Fingleton B. (2003) Matrix metalloproteinase inhibitors for cancer therapy: the current situation and future prospects. *Expert Opin Ther Targets*; 7: 385-397.

Fleishman SJ., Whitehead TA., Ekiert DC., Dreyfus C., Corn JE., Strauch EM., Wilson IA. and Baker D. (2011) Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science*; 332: 816-821.

Flight SM., Johnson LA., Du QS., Warner RL., Trabi M., et al. (2009) Textilinin-1, an alternative anti-bleeding agent to aprotinin: Importance of plasmin inhibition in controlling blood loss. *British J Hematol*; 145: 207-211.

Flower D.R. (1995) Multiple molecular recognition properties of the lipocalin protein family. *J. Mol. Recognit*.; 8: 185-195.

Folgueras, A.R., Pendas, A.M., Sanchez, L.M., and Lopez-Otin, C. (2004) Matrix metalloproteinases in cancer: from new functions to improved inhibition strategies. *Int. J. Dev. Biol*.; 48: 411-424.

Food and Drug Administration (2010) FDA: Pfizer Voluntarily Withdraws Cancer Treatment Mylotarg from U.S. Market, http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm216448.htm.

Francis B. and Kaiser II. (1993) Inhibition of metalloproteinases in *Bothrops asper* venom by endogenous peptides. *Toxicon*.; 31(7): 889-899.

Fujimori K., Covell D.G., Fletcher J.E. and Weinstein J.N.A (1990) Modeling analysis of monoclonal antibody percolation through tumors: a binding site barrier. *J. Nucl. Med*.; 31: 1191-1198.

Fujinaga M., Sielecki AR., Read RJ., Ardelt W., Laskowski M. Jr, and James MN. (1987) Crystal and molecular structures of the complex of alpha-chymotrypsin with its inhibitor turkey ovomucoid third domain at 1.8 Å resolution. *J Mol Biol*; 195(2): 397-418.

Garcia-Calvo, M., Leonard, R.J., Novick, J., Stevens, S.P., Schmalhofer, W., Kaczorowski, G.J. and Garcia, M.L. (1993) Purification, characterization, and biosynthesis of margatoxin, a component of *Centruroides margaritatus* venom that selectively inhibits voltage-dependent potassium channels. *The Journal of Biological Chemistry;* 268(25): 18866-18874.

Gebauer M. and Skerra A. (2009) Engineered protein scaffolds as next-generation antibody therapeutics. *Curr Opin Chem Biol*.; 13: 245-255.

Goldstein, G., Schindler, J. and Tsai, H. (1985) A randomized clinical trial of OKT3 monoclonal antibody for acute rejection of cadaveric renal transplants. *N. Engl. J. Med*.; 313: 337-342.

Gomis-Rüth FX., Maskos K., Betz M., Bergner A., Huber R., Suzuki K., Yoshida N., Nagase H., Brew K., Bourenkov GP., Bartunik H. and Bode W. (1997) Mechanism of inhibition of the human matrix metalloproteinase stromelysin-1 by TIMP-1. *Nature;* 389: 77-81.

Greenberg AS, Avila D, Hughes M, et al. (1995) A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. *Nature*; 374: 168-173.

Greene J., Wang M., Liu YE., Raymond LA., Rosen C. and Shi YE. (1996) Molecular cloning and characterization of human tissue inhibitor of metalloproteinase 4. *J. Biol. Chem*.;

271: 30375-30380.

Gross, M.L. (2002) Ethics, policy, and rare genetic disorders: the case of Gaucher disease in Israel. *Theor. Med. Bioeth*.; 23: 151-170.

Grossman, M., Tworowski, D., Dym, O., Lee, M.H., Levy, Y., Murphy, G. and Sagi, I. (2010) The intrinsic protein flexibility of endogenous protease inhibitor TIMP-1 controls its binding interface and affects its function. *Biochemistry*; 49: 6184-6192.

Guerois R., Nielsen J. E. and Serrano L. (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol*; 320: 369-387.

Guharoy M. and Chakrabarti P. (2005) Conservation and relative importance of residues across protein–protein interfaces. *PNAS*; 102(43): 15447-15452.

Hamers-Casterman C, Atarhouch T, Muyldermans S, et al. (1993) Naturally occurring antibodies devoid of light chains. *Nature*; 363: 446-448.

Hamze AB, Wei S, Bahudhanapati H, Kota S, Acharya KR and Brew K. (2007) Constraining specificity in the N-domain of tissue inhibitor of metalloproteinases-1; gelatinase-selective inhibitors. *Protein Sci.*; 16(9): 1905-1913.

Hartman, G.D., Egbertson, M.S., Halczenko, W., Laswell, W.L., Duggan, M.E., Smith, R.L. et al. (1992) Non-Peptide Fibrinogen Receptor Antagonists. 1. Discovery and Design of Exosite Inhibitors. *Journal of Medicinal Chemistry*; 35(24): 4640-4642.

Harvey AL. and Robertson B. (2004) Dendrotoxins: structure-activity relationships and effects on potassium ion channels. *Curr Med Chem*.; 11(23): 3065-3072.

Havranek J.J. and Harbury P.B. (2003) Automated design of specificity in molecular recognition. *Nat. Struct. Biol*.; 10: 45-52.

Hearst DP. and Cohen FE. (1994) GRAFTER: a computational aid for the design of novel proteins. *Protein Eng*; 7: 1411-1421.

Hegyi, H. and Gerstein, M. (1999) The relationship between protein structure and function: a comprehensive survey with application to the yeast genome. *J. Mol. Biol*.; 288: 147-164.

Hellinga HW. and Richards FM. (1991) Construction of new ligand binding sites in proteins of known structure. I. Computer-aided modeling of sites with pre-defined geometry. *J Mol Biol*.; 222(3): 763-785.

Hernandez-Barrantes, S., Bernardo, M., Toth, M., and Fridman, R. (2002) Regulation of membrane type-matrix metalloproteinases. *Semin. Cancer Biol.;* 12: 131-138.

Hey, T., Fiedler, E., Rudolph, R. and Fiedler, M. (2005) Artificial, non-antibody binding proteins for pharmaceutical and industrial applications. *Trends in Biotechnology***;** 23 (10): 514-522.

Higashi S. and Miyazaki K. (1999) Reactive site-modified tissue inhibitor of metalloproteinases-2 inhibits the cell-mediated activation of progelatinase A. *J Biol Chem*.; 274(15): 10497-10504.

Hohlbaum A.M. and Skerra A. (2007) Anticalins a novel class of therapeutic binding proteins. *Innovations in Pharmaceutical Technology*; 23: 32-38.

Hokama Y., Iwanaga S., Tatsuki T. and Suzuki T. (1976) Snake venom proteinase inhibitors. III. Isolation of five polypeptide inhibitors from the venoms of *Hemachatus haemachatus* (Ringhal's corbra) and *Naja nivea* (Cape cobra) and the complete amino acid sequences of two of them. *J Biochem*.; 79(3): 559-578.

Holliger P. and Hudson P.J. (2005) Engineered antibody fragments and the rise of single domains. *Nat. Biotechnol*.; 23: 1126-1136.

Honig B. and Nicholls A. (1995) Classical Electrostatics in Biology and Chemistry. *Science;* 268: 1144-1149.

Hornischer K. and Blocker H. (1996) Grafting of discontinuous sites: a protein modeling strategy. *Protein Eng*; 9: 931-939.

Horovitz A., and Levitzki A. (1987) An Accurate Method for Determination of Receptor Ligand and Enzyme-Inhibitor Dissociation-Constants from Displacement Curves. *PNAS USA;* 84: 6654-6658.

Fazelinia H., Cirino P.C., and Maranas C.D. (2009) OptGraft: A computational procedure for transferring a binding site onto an existing protein scaffold. *Protein Sci*.; 18(1): 180-195.

Huang KF, Hung CC, Wu SH, Chiou SH. (1998) Characterization of three endogenous peptide inhibitors for multiple metalloproteinases with fibrinogenolytic activity from the venom of Taiwan habu (*Trimeresurus mucrosquamatus*). *Biochem Biophys Res Commun*.; 248(3): 562-568.

Huang, W., Meng, Q., Suzuki, K., Nagase, H., and Brew, K. (1997) Mutational Study of the Amino-terminal Domain of Human Tissue Inhibitor of Metalloproteinases 1 (TIMP-1)

Locates an Inhibitory Region for Matrix Metalloproteinases. *Journal of Biological Chemistry* 272 (35): 22086–22091.

Huang, W., Suzuki, K., Nagase, H., Arumugam, S., Van Doren, S.R., and Brew, K. (1996) Folding and characterization of the amino-terminal domain of human tissue inhibitor of metalloproteinases- 1 (TIMP- 1) expressed at high yield in *E. coli*. *FEBS Lett*. 384: 155–161.

Hutton M., Willenbrock F., Brocklehurst K. and Murphy G. (1998) Kinetic analysis of the mechanism of interaction of full-length TIMP-2 and gelatinase A: evidence for the existence of a low-affinity intermediate. *Biochemistry* 37: 10094–10098.

Iyer, S., Wei, S., Brew, K. and Acharya, K.R. (2007) Crystal structure of the catalytic domain of matrix metalloproteinase-1 in complex with the inhibitory domain of tissue inhibitor of metalloproteinase-1. *J.Biol.Chem*. 282(1): 364-371.

Janin J. and Chothia C. (1990) The structure of protein-protein recognition sites. *J. Biol.Chem.* 265: 16027-16030.

Jarrin, A., Andrieux, A., Chapel, A., Buchou, T. and Marguerie, G. (1994) A synthetic peptide with anti-platelet activity derived from a CDR of an anti-GPIIb-IIIa antibody. *FEBS Lett* 354: 169-172.

Jiang L, Althoff EA, Clemente FR, Doyle L, Rothlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas CF III, Hilvert D, Houk KN, Stoddard BL and Baker D. (2008) De novo computational design of retro-aldol enzymes. *Science* 319: 1387-1391.

Joachimiak, L.A., Kortemme, T., Stoddard, B.L. and Baker, D. (2006) Computational design of a new hydrogen bond network and at least a 300-fold specificity switch at a protein-protein interface. *J. Mol. Biol*. 361: 195-208.

Johnson JL., George SJ., Newby AC. and Jackson CL. (2005) Divergent effects of matrix metallo-proteinases 3, 7, 9, and 12 on atherosclerotic plaque stability in mouse brachiocephalic arteries. *Proc Natl Acad Sci U S A*; 102: 15575-15580.

Jones S. and Thornton J.M. (1995) Protein-protein interactions: a review of protein dimer structures. *Prog Biophys Mol Biol*.; 63(1): 31-65.

Jones S. and Thornton J.M. (1996) Principles of protein-protein interactions. *Proc. Natl Acad. Sci. USA*, 93: 13-20.

Julien, J.P., Bryson, S., Nieva, J.L. and Pai, E.F. (2008) Structural details of HIV-1 recognition by the broadly neutralizing monoclonal antibody 2F5: epitope conformation, antigen-

recognition loop mobility, and anion-binding site. *J. Mol. Biol*. 384: 377-392.

Kashiwagi M., Tortorella M., Nagase H. and Brew K. (2001) TIMP-3 is a potent inhibitor of aggrecanase 1 (ADAM-TS4) and aggrecanase 2 (ADAM-TS5). *J Biol Chem*; 276: 12501-12504.

Kawashima, T., Berthet-Colominas, C., Wulff, M., Cusack, S. and Leberman, R. (1996) The structure of the E. coli EF-Tu. EF-Ts complex at 2.5 Å resolution. *Nature*, 379: 511-518.

Keskin O. and Nussinov R. (2005) Favorable scaffolds: proteins with different sequence, structure and function may associate in similar ways. *Protein Eng Des Sel*.; 18(1): 11-24.

Keskin O. and Nussinov R. (2007) Similar binding sites and different partners: implications to shared proteins in cellular pathways. *Structure* (3): 341-354.

Kleywegt G.T. (1999) Recognition of spatial motifs in protein structures. *J. Mol. Biol*., 285: 1887-1897.

Koide, A., Bailey, C.W., Huang, X. and Koide, S. (1998) The fibronectin type III domain as a scaffold for novel binding proteins. *J. Mol. Biol*. 284: 1141-1151.

Kolosov A., Goodchild CS. and Cooke I. (2010) CNSB004 (Leconotide) causes antihyperalgesia without side effects when given intravenously: a comparison with ziconotide in a rat model of diabetic neuropathic pain. *Pain Med*.; 11(2): 262-273.

Kortemme T. and Baker D. (2002) A simple physical model for binding energy hot spots in protein-protein complexes. *PNAS*, 99(22): 14116–14121.

Kyte J. and Doolittle RF. (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol*.; 157(1): 105-132.

Lang R., M. Braun, N.E. Sounni, A. Noel, F. Frankenne, J.M. Foidart, et al. (2004) Crystal structure of the catalytic domain of MMP-16/MT3-MMP: characterization of MT-MMP specific features. *J. Mol. Biol*. 336: 213-225.

Laskowski R.A. (1991) SURFNET computer program (Department of Biochemistry and Molecular Biology, University College, London, England).

Lauer-Fields, J.L., Cudic, M., Wei, S., Mari, F., Fields, G.B. and Brew, K. (2007) Engineered sarafotoxins as tissue inhibitor of metalloproteinases-like matrix metalloproteinase inhibitors. *Journal of Biological Chemistry*; 282(37): 26948-26955.

Lauwereys M., Arbabi Ghahroudi M., Desmyter A., et al. (1998) Potent enzyme inhibitors derived from dromedary heavy-chain antibodies. *EMBO J*; 17: 3512-20.

Lawrence M.C. and Colman P.M. (1993) Shape complementarity at protein/protein interfaces. *J. Mol. Biol.*; 234: 946-950.

Leader, B., Baca, Q.J. and Golan, D.E. (2008) Protein therapeutics: a summary and pharmacological classification. *Nature Reviews Drug Discovery*; 7: 21-39.

Lee M.H., Rapti M. and Murphy G. (2003) Unveiling the surface epitopes that render tissue inhibitor of metalloproteinases-1 inactive against membrane type 1-matrix metalloproteinase, *J. Biol. Chem*.; 278: 40224-40230.

Lee SC, Park K, Han J, Lee JJ, Kim HJ, Hong S, Heu W, Kim YJ, Ha JS, Lee SG, Cheong HK, Jeon YH, Kim D and Kim HS. (2012) Design of a binding scaffold based on variable lymphocyte receptors of jawless vertebrates by module engineering. *PNAS*; 109(9): 3299-3304.

Leslie A.G.W. (2006) The integration of macromolecular diffraction data. *Acta Crystallogr D* 62: 48-57.

Li C., Liu M., Monbo J., et al. (2008) Turning a scorpion toxin into an antitumor miniprotein. J *Am Chem Soc*; 130: 13546-13548.

Li J., Yang C., Xia Y., Bertino A., Glaspy J., Roberts M. and Kuter D.J., (2001) Thrombocytopenia caused by the development of antibodies to thrombopoietin. *Blood* 98: 3241-3248.

Liang J., Liu E. and Yu Y. (2006) Macrophage metalloelastase accelerates the progression of atherosclerosis in transgenic rabbits. *Circulation*; 113: 1993-2001.

Liang S., Liu Z., Li W., Ni L. and Lai L. (2000) Construction of protein binding sites in scaffold structures. *Biopolymers*; 54: 515-523.

Liu S., Liu S., Zhu X., Liang H., Cao A., Chang Z. and Lai L. (2007) Nonnatural protein–protein interaction-pair design by key residues grafting. *Proc Natl Acad Sci USA*; 104: 5330-5335.

Livett BG., Sandall DW., Keays D., Down J., Gayler KR., Satkunanathan N. and Khalil Z. (2006) Therapeutic applications of conotoxins that target the neuronal nicotinic acetylcholine receptor. *Toxicon*; 48: 810-829.

Lo Conte L., Chothia C. and Janin J. (1999) The atomic structure of protein–protein recognition

sites. *J Mol Biol*; 285: 2177-2198.

Löfblom J., Feldwisch J., Tolmachev V., Carlsson J., Ståhl S. and Frejd FY. (2010) Affibody molecules: Engineered proteins for therapeutic, diagnostic and biotechnological applications. *FEBS Letters*, 584(12): 2670-2680.

Lonberg N. (2008a) Human monoclonal antibodies from transgenic mice. *Handb. Exp. Pharmacol*.; 181: 69-97.

Lonberg N. (2008b) Fully human antibodies from transgenic mouse and phage display platforms. *Curr. Opin. Immunol*.; 20: 450-459.

Lu J., Yang H., Yu H., Gao W., Lai R., Liu J. and Liang X. (2008) A novel serine protease inhibitor from *Bungarus fasciatus* venom. *Peptides*; 29(3): 369-374.

Lutterotti A. and Martin R. (2008) Getting specific: monoclonal antibodies in multiple sclerosis. *Lancet Neurol*.; 7: 538-547.

Luttun A, Lutgens E, Manderveld A, et al. (2004) Loss of matrix metalloproteinase-9 or matrix metalloproteinase-12 protects apolipoprotein E-deficient mice against atherosclerotic media destruction but differentially affects plaque growth. *Circulation*; 109: 1408-1414.

MacKerell, A.D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J. et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B*; 102: 3586-3616.

Magis C., D. Gasparini, A. Lecoq, M. H. Le Du, E. Stura, J. B. Charbonnier, G. Mourier, J.-C. Boulain, L. Pardo, A. Caruana, A. Joly, M. Lefranc, M. Masella, A. Menez, and P. Cuniasse (2006) Structure-Based Secondary Structure-Independent Approach To Design Protein Ligands: Application to the Design of Kv1.2 Potassium Channel Blockers. *J. AM. CHEM. SOC*.; 128: 16190-16205.

Main, A.L., Harvey, T.S., Baron, M., Boyd, J. and Campbell, I.D. (1992) The three-dimensional structure of the tenth type III module of fibronectin: an insight into RGD-mediated interactions. *Cell;* 71: 671-678.

Malisi C., Kohlbacher O. and Höcker B. (2009) Automated scaffold selection for enzyme design. *Proteins*; 77: 74-83.

Malmberg A.B. and Yaksh T.L. (1995) Effect of continuous intrathecal infusion of ω-conopeptides, N-type calciumchannel blockers, on behaviour and antinociception in the formalin and hot-plate tests in rats. *Pain*; 60: 83-90.

Mamluk R., Carvajal IM., Morse BA., Wong H., Abramowitz J., Aslanian S., Lim AC. et al. (2010) Anti-tumor effect of CT-322 as an adnectin inhibitor of vascular endothelial growth factor receptor-2. *MAbs*; 2(2): 199-208.

Martin L., Stricher F., Missé D., Sironi F., Pugnière M. et al. (2003) Rational design of a CD4 mimic that inhibits HIV-1 entry and exposes cryptic neutralization epitopes. *Nat Biotechnol*.; 21: 71-76.

Marvin J.S. and Hellinga H.W. (2001) Conversion of a maltose receptor into a zinc biosensor by computational design. *Proc. Natl. Acad. Sci*.; 98: 4955-4960.

Maskos K. (2005) Crystal structures of MMPs in complex with physiological and pharmacological inhibitors. *Biochimie*; 87(3-4): 249-263.

Maskos, K., Lang, R., Tschesche, H. and Bode, W. (2007) Flexibility and variability of TIMP binding: X-ray structure of the complex between collagenase-3/MMP-13 and TIMP-2. *J. Mol. Biol*.; 366: 1222-1231.

McCoy, A.J., Chandana Epa, V. and Colman, P.M. (1997) Electrostatic complementarity at protein-protein interfaces. *J. Mol. Biol*.; 268: 570-584.

McGivern J.G. (2007) Ziconotide: a review of its pharmacology and use in the treatment of pain. *Neuropsychiatr Dis Treat*.; 3(1): 69-85.

McIntosh, J.M., Santos, A.D. and Olivera, B.M. (1999) *Conus* peptides targeted to specific nicotinic acetylcholine receptor subtypes. *Annu. Rev. Biochem.*; 68: 59-88.

Melendez-Zajgla, J., Del Pozo, L., Ceballos, G., and Maldonado, V. (2008) Tissue Inhibitor of Metalloproteinases-4. The road less travelled. *Molecular Cancer;* 7: 85.

Meng, Q., Malinovskii, V., Huang, W., Hu, Y., Chung, L., Nagase, H., Bode, W., Maskos, K. and Brew, K. (1999) Residue 2 of TIMP-1 is a major determinant of affinity and specificity for matrix metalloproteinases but effects of substitutions do not correlate with those of the corresponding P1' residue of substrate. *J. Biol. Chem*.; 274: 10184-10189.

Mer, G., Kellenberger, E. and Lefèvre, J.-F. (1998) α-Helix mimicry of a β-turn. *J. Mol. Biol*. 281: 235-240.

Miller C., Moczydlowski E., Latorre R. and Phillips M. (1985) Charybdotoxin, a protein inhibitor of single Ca2+-activated K+ channels from mammalian skeletal muscle. *Nature*; 313(6000): 316-318.

Miller, S., Janin, J., Lesk, A. M. and Chothia, C. (1987) Interior and surface of monomeric proteins. *J. Mol. Biol*.; 196: 641-656.

Moczydlowski, E., Olivera, B.M., Gray, W.R. and Strichartz, G.R. (1986) Discrimination of muscle and neuronal Na-channel subtypes by binding competition between [3H]saxitoxin and μ-conotoxins. *Proc. Natl Acad. Sci. USA;* 83: 5321-5325.

Moks, T., Abrahmsén, L., Nilsson, B., Hellman, U., Sjöquist, J. and Uhlén, M. (1986) Staphylococcal protein A consists of five IgG-binding domains. *Eur. J. Biochem.;* 156: 637-643.

Morgunova E., Tuuttila A., Bergmann U., Isupov M., Lindqvist Y., Schneider G. and Tryggvason K. (1999) Structure of human pro-matrix metalloproteinase-2: activation mechanism revealed. *Science*; 284: 1667-1670.

Morgunova, E., Tuuttila, A., Bergmann, U. and Tryggvason, K. (2002) Structural insight into the complex formation of latent matrix metalloproteinase 2 with tissue inhibitor of metalloproteinase 2. *Proc. Natl. Acad. Sci. USA*; 99(11): 7414-7419.

Mott J.D. and Werb Z. (2004) Regulation of matrix biology by matrix metalloproteinases. *Curr. Opin. Cell Biol*.; 16: 558-564.

Müller H.N. and Skerra A. (1993) Functional expression of the uncomplexed serum retinolbinding protein in *Escherichia coli*. Ligand binding and reversible unfolding characteristics. *J. Mol. Biol*.; 230: 725-732.

Murphy, G., Houbrechts, A., Cockett MI., Williamson RA., O'Shea M. and Docherty AJP. (1991) The N-terminal domain of human tissue inhibitor of metalloproteinases retains metalloproteinase inhibitory activity. *Biochemistry;* 30: 8097-8102.

Murphy G. and Willenbrock F. (1995) Tissue inhibitors of matrix metalloendopeptidases. *Methods Enzymol*.; 248: 496-510.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D*; 53: 240-255.

Nadzirin N., Gardiner EJ., Willett P., Artymiuk PJ. and Firdaus-Raih M. (2012) SPRITE and ASSAM: web servers for side chain 3D-motif searching in protein structures. *Nucleic Acids Research*.; 40: 380-386.

Neuberger MS., Williams GT., Mitchell EB., Jouhal SS., Flanagan JG. and Rabbitts TH. (1985) A hapten-specific chimaeric IgE antibody with human physiological effector function. *Nature*; 314: 268-270.

Nguyen VD., Hatahet F., Salo KE., Enlund E., Zhang C. and Ruddock LW. (2011) Pre-expression of a sulfhydryl oxidase significantly increases the yields of eukaryotic disulfide bond containing proteins expressed in the cytoplasm of *E. coli*. *Microb Cell Fact.*; 10: 1.

Nilsson B., Moks T., Jansson B., Abrahmsén L., Elmblad A., Holmgren E., Henrichson C., Jones TA. and Uhlén M. (1987) A synthetic IgG-binding domain based on staphylococcal protein A. *Protein Eng.*; 1: 107-113.

Nobeli I., Favia AD. and Thornton JM. (2009) Protein promiscuity and its implications for biotechnology. *Nat Biotechnol.*; (2): 157-167.

Nooren I.M.A. and Thornton J.M. (2003) Structural characterisation and functional significance of transient protein-protein interactions. *J Mol Biol;* 325 (5): 991-1018.

Norton RS., Pennington MW. and Wulff H. (2004) Potassium channel blockade by the sea anemone toxin ShK for the treatment of multiple sclerosis and other autoimmune diseases. *Curr Med Chem.*; 11(23): 3041-3052.

Nuttall S.D. and Walsh R.B. (2008) Display scaffolds: protein engineering for novel therapeutics. *Current Opinion in Pharmacology*; 8: 609-615.

Nygren P.Å. (2008) Alternative binding proteins: affibody binding proteins developed from a small three-helix bundle scaffold. *FEBS J.*; 275: 2668-2676.

Nygren P.A. and Uhlen M. (1997) Scaffolds for engineering novel binding sites in proteins. *Curr Opin Struct Biol*; 7(4): 463-469.

Ofek, G., Tang, M., Sambor, A., Katinger, H., Mascola, J. R., Wyatt, R. and Kwong, P.D. (2004) Structure and mechanistic analysis of the anti-human immunodeficiency virus type 1 antibody 2F5 in complex with its gp41 epitope. *J. Virol.*; 78: 10724-10737.

Ofran Y. and Rost B. (2007) Protein–protein interaction hotspots carved into sequences. *PLoS Comput Biol;* 3(7): 1169-1176.

Ondetti MA., Williams NJ., Sabo EF., Pluscec J., Weaver ER. and Kocy O. (1971) Angiotensin-converting enzyme inhibitors from the venom of *Bothrops jararaca*. Isolation, elucidation of structure and synthesis. *Biochemistry*; 19: 4033-4039.

Overall CM. and Kleifeld O. (2006) Tumour microenvironment - opinion: Validating matrix metalloproteinases as drug targets and anti-targets for cancer therapy. *Nat Rev Cancer*; 6: 227-239.

Padlan E.A. (1990) On the nature of antibody combining sites: unusual structural features that may confer on these sites an enhanced capacity for binding ligands. *Proteins*; **7**: 112-124.

Peisajovich S.G. and Tawfik D.S. (2007) Protein engineers turned evolutionists. *Nat Methods;* 4: 991-994.

Perkins, J.R., Parker, C.E. and Tomer, K.B. (1993) The characterization of snake venoms using capillary electrophoresis in conjunction with electrospray mass spectrometry: Black mambas. *Electrophoresis*; 14: 458-468.

Peterson JT. (2004) Matrix metalloproteinase inhibitor development and the remodeling of drug discovery. *Heart Fail Rev*; 9: 63-79.

Philo JS., Aoki KH., Arakawa T., Narhi LO. and Wen J. (1996) Dimerization of the Extracellular Domain of the Erythropoietin (EPO) Receptor by EPO: One High-Affinity and One Low-Affinity Interaction. *Biochemistry;* 35: 1681-1691.

Plaxco, K.W., Spitzfaden, C., Campbell, I.D. and Dobson, C.M. (1997) Comparison of the folding kinetics and thermodynamics of two homologous fibronectin type III modules. *J. Mol. Biol*.; 270: 763-770.

PRISM-PLUS (The Platelet Receptor Inhibition in Ischemic Syndrome Management in Patients Limited by Unstable Signs and Symptoms Study Investigators) (1998) Inhibition of the Platelet Glycoprotein IIb/IIIa Receptor with Tirofiban in Unstable Angina and Non–Q-Wave Myocardial Infarction. *N Engl J Med*; 338: 1488-1497.

Reichert J.M. (2003) Trends in development and approval times for new therapeutics in the United States. *Nature Rev. Drug Discov*.; 2: 695-702.

Reichert J.M. (2010) Antibodies to watch in 2010. *MAbs*; 2: 28-45.

Reichert J.M. (2011) Antibody-based therapeutics to watch in 2011. *MAbs*.; 3(1): 76-99.

Reina, J., Lacroix, E., Hobson, S.D., Fernandez-Ballester, G., Rybin, V., Schwab, M.S., Serrano, L. and Gonzalez, C. (2002) Computer-aided design of a PDZ domain to recognize new target sequences. *Nat. Struct. Biol*.; 9: 621-627.

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.

Robeva A, Politi V, Shannon JD, Bjarnason JB, Fox JW. (1991) Synthetic and endogenous inhibitors of snake venom metalloproteinases. *Biomed Biochim Acta*.; 50(4-6): 769-73.

Rosenberg GA. (2005) Matrix metalloproteinases biomarkers in multiple sclerosis. *Lancet;* 365*:* 1291-1293.

Rothlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, et al. (2008) Kemp elimination catalysts by computational enzyme design. *Nature*; 453: 190-195.

Samsa, G.P., Matchar, D.B., Williams, G.R. and Levy, D.E. (2002) Cost-effectiveness of ancrod treatment of acute ischaemic stroke: results from the Stroke Treatment with Ancrod Trial (STAT). *J. Eval. Clin. Pract.*; 8: 61-70.

Sandall DW., Satkunanathan N., Keays DA., Polidano MA., Liping X., Pham V., Down JG., Khalil Z., Livett BG. and Gayler KR. (2003) A novel α-conotoxin identified by gene sequencing is active in suppressing the vascular response to selective stimulation of sensory nerves *in vivo*. *Biochemistry*; 42: 6904-6911.

Scarborough RM., Naughton MA., Teng W., Rose JW., Phillips DR., Nannizzi L., Arfsten A., Campbell AM. and Charo IF. (1993) Design of potent and specific integrin antagonists. Peptide antagonists with high specificity for glycoprotein IIb-IIIa. *J. Biol. Chem*.; 268: 1066-1073.

Schlehuber S. and Skerra A. (2002) Tuning ligand affinity, specificity, and folding stability of an engineered lipocalin variant – a so-called 'anticalin' – using a molecular random approach. *Biophys. Chem*.; 96: 213-228.

Schlehuber S. and Skerra A. (2005) Lipocalins in drug discovery: from natural ligand-binding proteins to "anticalins". *Drug Discov Today*; 10(1): 23-33.

Schulz H., Bohlius JF., Trelle S. et Al. (2007) Immunochemotherapy with rituximab and overall survival in patients with indolent or mantle cell lymphoma: a systematic review and meta-analysis [archive], *J Natl Cancer Inst*; 99: 706-714.

Schweitz, H., Heurteaux, C., Bois, P., Moinier, D., Romey, G. and Lazdunski, M. (1994) Calcicludine, a venom peptide of the Kunitz-type protease inhibitor family, is a potent blocker of high threshold Ca2+ channels with a high affinity for L-type channels in cerebellar granule neurons. *Proc. Natl. Acad. Sci. USA*; 91: 878-882.

Scott, D.A., Wright, C.E. and Angus, J.A. (2002) Actions of intrathecal ω-conotoxins CVID, GVIA, MVIIA, and morphine in acute and neuropathic pain in the rat. *Eur. J. Pharmacol.;* 451: 279-286.

Sheinerman FB., Norel R. and Honig B. (2000) Electrostatic aspects of protein-protein interactions. *Curr Opin Struct Biol*; 10(2): 153-159.

Shifman JM. and Mayo SL. (2002) Modulating calmodulin binding specificity through computational protein design. *J Mol Biol*; 323: 417-423.

Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, St Clair JL, Gallaher JL, Hilvert D, Gelb MH, Stoddard BL, Houk KN, Michael FE, Baker D (2010) Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science*; 329: 309-313.

Silverman J, Lu Q., Bakker A., To W., Duguay A., et al. (2005) Multivalent avimer proteins evolved by exon shuffling of a family of human receptor domains. *Nature Biotechnology*; 23: 1556-1561.

Sitkoff, D., Sharp, K.A. and Honig, B. (1994) Accurate calculation of hydration free-energies using macroscopic solvent models. *Journal of Physical Chemistry*; 98: 1978-1988.

Skerra A. and Schlehuber S. (2003) PIERIS Proteolab AG. Muteins of human neutrophil gelatinase-associated lipocalin and related proteins, WO 03/029463 A2.

Slagle, S.P., Kozack, R.E. and Subramaniam, S. (1994) Role of electrostatics in antibody-antigen association: anti-hen egg lysozyme/lysozyme complex (HyHEL-5/HEL). *J. Biomol. Struct. Dynam*.; 12: 439-456.

Smith J.W., Hu D., Satterthwait A., Pinz-Sweeney S., and Barbas C.F., III (1994) Building synthetic antibodies as adhesive ligands for integrins. *J. Biol. Chem.;* 269: 32788-32795.

Smith JW., Tachias K. and Madison EL. (1995) Protein loop grafting to construct a variant of tissue-type plasminogen activator that binds platelet integrin alpha IIb beta 3. *J Biol Chem.*; 270(51): 30486-30490.

Smith, M., Cabot, P.J., Ross, F.B., Robertson, A.D. and Lewis, R.J. (2002) The novel N-type calcium channel blocker, AM336, produces potent dose-dependent antinociception after intrathecal dosing in rats and inhibits substance P release in rat spinal cord slices. *Pain*; 96: 119-127.

Sommerhoff, CP., Avrutina, O., Schmoldt, H.U., Gabrijelcic-Geiger, D., Diederichsen, U. and Kolmar, H. (2010) Engineered cystine knot minoproteins as potent inhibitors of human mast cell tryptase beta. *J Mol Biol*.; 395(1): 167-175.

Steed P.M., Tansey M.G., Zalevsky J., Zhukovsky E.A., Desjarlais J.R., Szymkowski D.E., Abbott C., Carmichael D., et al. (2003) Inactivation of TNF signaling by rationally designed dominant-negative TNF variants. *Science*; 301: 1895-1898.

Stocker, W., Grams, F., Baumann, U., Reinemer, P., Gomis-Ruth, F.X., McKay, D.B. and Bode, W. (1995) The metzincins - topological and sequential relations between the astacins, adamalysins, serralysins, and matrixins (collagenases) define a superfamily of zinc-peptidases. *Protein Sci*.; 4: 823-840.

Stracke JO., Hutton M., Stewart M., Pendas AM., Smith B., Lopez-Otin C., Murphy G. and Knauper V. (2000) Biochemical characterization of the catalytic domain of human matrix metalloproteinase 19. Evidence for a role as a potent basement membrane degrading enzyme. *J Biol Chem.;* 275(20): 14809-14816.

Stricher F., Martin L. and Vita C. (2006) Design of Miniproteins by the Transfer of Active Sites Onto Small-Size Scaffolds. *Methods in Molecular Biology*; 340: 113-149.

Strydom DJ. (1977) Snake venom toxins. The amino acid sequence of toxin Vi2, a homologue of pancreatic trypsin inhibitor, from *Dendroaspis polylepis polylepis* (black mamba) venom. *Biochim Biophys Acta*.; 491(2): 361-369.

Stubbs MT., Morenweiser R., Stürzebecher J., Bauer M., Bode W., Huber R., Piechottka GP., Matschiner G., Sommerhoff CP., Fritz H. and Auerswald EA. (1997) The three-dimensional structure of recombinant leech-derived tryptase inhibitor in complex with trypsin. Implications for the structure of human mast cell tryptase and its inhibition. *J. Biol. Chem.*; 272 (32): 19931-19937.

Szayna M., Doyle ME., Betkey JA., Holloway HW., Spencer RG., Greig NH. and Egan JM. (2000) Exendin-4 decelerates food intake, weight gain, and fat deposition in Zucker rats. *Endocrinology*; 141: 1936-1941.

Takahashi H., Iwanaga S., Kitagawa T., Hokama Y. and Suzuki T. (1974) Snake venom proteinase inhibitors. II. Chemical structure of inhibitor II isolated from the venom of Russell's viper (*Vipera russelli*). *J Biochem*.; 76(4): 721-733.

Tam R. and Saier M.H.Jr. (1993) Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiol. Rev*.; 57: 320-346.

Tasumi S., Velikovsky CA., Xu G., Gai SA., Wittrup KD., Flajnik MF., Mariuzza RA. and Pancer Z. (2009) High-affinity lamprey VLRA and VLRB monoclonal antibodies. *Proc Natl Acad Sci USA*; 106: 12891-12896.

Tolmachev V. (2008) Imaging of HER-2 overexpression in tumors for guiding therapy. *Curr. Pharm. Des*.; 14: 2999-3019.

Troeberg, L., Tanaka, M., Wait, R., Shi, Y. E., Brew, K. and Nagase, H. (2002) *E. coli*

expression of TIMP-4 and comparative kinetic studies with TIMP-1 and TIMP-2: insights into the interactions of TIMPs and matrix metalloproteinase 2 (gelatinase A). *Biochemistry*; 41: 15025-15035.

Tuba, Z., Maho, S. and Vizi, E.S. (2002) Synthesis and structure activity relationships of neuromuscular blocking agents. *Curr. Med. Chem.*; 9: 1507-1536.

Tyagi M., Shoemaker BA., Bryant SH. and Panchenko AR. (2009) Exploring functional roles of multibinding protein interfaces. *Protein Sci.*; 8: 1674-1683.

Ullmann, G.M., Hauswald, M., Jensen, A. and Knapp, E.W. (2000) Structural alignment of ferredoxin and flavodoxin based on electrostatic potentials: Implications for their interactions with photosystem I and ferredoxin-NADP reductase. *Proteins-Structure Function and Bioinformatics*; 38: 301-309.

Ullmann, G.M., Hauswald, M., Jensen, A., Kostic, N.M. and 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.

Urvoas A., Guellouz A., Valerio-Lepiniec M., Graille M., Durand D., Desravines DC., van Tilbeurgh H., Desmadril M. and Minard P. (2010) Design, production and molecular structure of a new family of artificial alpha-helicoidal repeat proteins (αRep) based on thermostable HEAT-like repeats. *J Mol Biol*; 404(2): 307-327.

Valdez-Cruz N.A., Conde R., Zamudio F. and Possani L.D. (2011) Anticoagulants from scorpion venoms en: Marklan, F.S. Toxins and Hemostasis: From Bench to Bedside. Springer; 255-266.

Valente RH., Dragulev B., Perales J., Fox JW. and Domont GB. (2001) BJ46a, a snake venom metalloproteinase inhibitor. Isolation, characterization, cloning and insights into its mechanism of action. *Eur J Biochem.*; 268(10): 3042-3052.

Visse R. and Nagase H. (2003) Matrix metalloproteinases and tissue inhibitors of metalloproteinases: structure, function, and biochemistry. *Circ Res*; 92(8): 827-839.

Vita C, Drakopoulou E, Vizzavona J, Rochette S, Martin L, Ménez A et al. (1999) Rational engineering of a miniprotein that reproduces the core of the CD4 site interacting with HIV-1 envelope glycoprotein. *Proc. Natl. Acad. Sci. USA*; 96: 13091-13096.

Vita, C., Vizzavona, J., Drakopoulou, E., Zinn-Justin, S., Gilquin, B. and Ménez, A. (1998) Novel miniproteins engineered by the transfer of active sites to small natural scaffolds. *Biopolymers*; 47: 93-100.

Wade, R.C., Gabdoulline, R.R., Ludemann, S.K. and Lounnas, V. (1998a) Electrostatic steering and ionic tethering in enzyme-ligand binding: Insights from simulations. *Proc. Natl. Acad. Sci. USA*; 95: 5942-5949.

Wade, R.C., Gabdoulline, R.R. and Luty, B.A. (1998b) Species dependence of enzyme-substrate encounter rates for triose phosphate isomerases. *Proteins*; 31: 406-416.

Waldmann T.A. (2003) Immunotherapy: past, present and future. *Nat. Med*.; 9: 269-277.

Walsh G. (2010) Biopharmaceuticals benchmarks 2010. *Nature Biotechnology;* 28: 917-924.

Wang WM., Ge G., Lim NH., Nagase H. and Greenspan DS. (2006) TIMP-3 inhibits the procollagen N-proteinase ADAMTS-2. *Biochem J;* 398: 515-519.

Wang, Y.X., Gao, D., Pettus, M., Phillips, C. and Bowersox, S.S. (2000) Interactions of intrathecally administered Ziconotide, a selective blocker of neuronal N-type voltage sensitive calcium channels, with morphine on nociception in rats. *Pain*; 84: 271-281.

Welsh J.H. (1967) Acetylcholine in snake venoms. In *Animal Toxins* (Russell, F.E. and Saunders, P.R., Eds.) 363-368, Pergamon, New York.

Will H., Atkinson S.J., Butler G.S., Smith B. and Murphy G. (1996) The soluble catalytic domain of membrane type 1 matrix metalloproteinase cleaves the propeptide of progelatinaseA and initiates autoproteolytic activation. Regulation by TIMP-2 and TIMP-3. *J. Biol. Chem*.; 271: 17119-17123.

Willenbrock F. and Murphy G. (1994) Structure–function relationships in the tissue inhibitors of metalloproteinases. *Am. J. Resp. Crit. Care Med.;* 150: 5165-5170.

Williamson RA., Hutton M., Vogt G., Rapti M., Knäuper V., Carr MD., et al. (2001) Tyrosine 36 plays a critical role in the interaction of the AB loop of tissue inhibitor of metalloproteinases-2 with matrix metalloproteinase-14. *J Biol Chem*; 276: 32966-32970.

Wingfield, P.T., Sax, J.K., Stahl, S.J., Kaufman, J. et al. (1999) Biophysical and functional characterization of full-length, recombinant human tissue inhibitor of metalloproteinases-2 (TIMP-2) produced in *Escherichia coli*: comparison of wild type and amino-terminal alanine appended variant with implications for the mechanism of TIMP functions. *J. Biol. Chem*.; 274: 21362-21368.

Witty JP., Lempka T., Coffey RJ. Jr, et al. (1995) Decreased tumor formation in 7,12-dimethylbenzanthracene treated stromelysin-1 transgenic mice is associated with

alterations in mammary epithelial cell apoptosis. *Cancer Res*; 55: 1401-1406.

Wu A.M. and Senter P.D. (2005) Arming antibodies: prospects and challenges for immunoconjugates. *Nat. Biotechnol*.; 23: 1137-1146.

Wurch, T., Lowe, P., Caussanel, V., Bes, C., Beck, A., and Corvaia, N. (2008) Development of novel protein scaffolds as alternatives to whole antibodies for imaging and therapy: status on discovery research and clinical validation. *Curr. Pharm. Biotechnol*.; 9: 502-509.

Xu, D., Tsai, C-J. and Nussinov, R. (1997) Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein Engineering*; 10(9): 999-1012.

Xu L., Aha P., Gu K., Kuimelis RG., Kurz M., et al. (2002) Directed evolution of high-affinity antibody mimics using mRNA display. *Chem Biol*; 9: 933-942.

Yang W., Arii S., Gorrin-Rivas MJ., et al. (2001) Human macrophage metalloelastase gene expression in colorectal carcinoma and its clinicopathologic significance. *Cancer*; 91: 1277-1283.

Zahnd C., Kawe M., Stumpp MT., de Pasquale C., Tamaskovic R., Nagy-Davidescu G., et al. (2010) Efficient tumor targeting with high-affinity designed ankyrin repeat proteins: effects of affinity and molecular size. *Cancer Res*; 70: 1595-1605.

Zahnd C., Wyler E., Schwenk JM., Steiner D., Lawrence MC., McKern NM., Pecorari F., Ward CW., Joos TO. and Plückthun A. (2007) A designed ankyrin repeat protein evolved to picomolar affinity to Her2. *J Mol Biol*; 369: 1015-1028.

Zanghellini A., Jiang L., Wollacott AM., Cheng G., Meiler J., Althoff EA., Röthlisberger D. and Baker D. (2006) New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci*; 15: 2785-2794.

Zhang C. and Lai L. (2012) AutoMatch: target-binding protein design and enzyme design by automatic pinpointing potential active sites in available protein scaffolds. *Proteins*; 80(4): 1078-1094.

Résumé : Les travaux réalisés portent sur la conception de ligands protéiques capables de cibler le site catalytique des métalloprotéases matricielles (MMPs) grâce à une méthode d'ingénierie développée au laboratoire qui repose sur le greffage de motifs fonctionnels. Le motif fonctionnel choisi correspond aux 4 résidus N-terminaux du TIMP-2, un inhibiteur naturel des MMPs. Des plates-formes protéiques possédant des motifs d'acides aminés dans une topologie similaire à celle du motif de référence dans le complexe TIMP-2/MMP-14 ont été identifiées par criblage systématique de la PDB à l'aide du logiciel STAMPS (Search for Three-dimensional Atom Motif in Protein Structure). Dix candidats ligands satisfaisant les contraintes topologiques, stériques et de similarité électrostatique avec le ligand naturel TIMP-2 ont été sélectionnés. Ces ligands ont été produits par synthèse chimique ou par voie recombinante puis leur capacité à inhiber une série de 6 MMPs a été évaluée. Les résultats indiquent que tous les ligands protéiques conçus *in silico* sont capables de lier les sites catalytiques des MMPs avec des constantes d'association allant de 450 nM à 590 μM, sans optimisation supplémentaire. La caractérisation structurale par diffraction X de 2 variants d'un de ces ligands protéiques a permis de montrer que les interactions établies par le motif 1-4 dans ces ligands étaient similaires à celles observées dans le complexe TIMP-2/MMP-14, avec cependant des différences dans la géométrie de certaines d'entre elles. Des études de simulation par dynamique moléculaire ont également permis de mettre en évidence de possibles différences dans la géométrie et la stabilité de certaines des interactions reproduites dans les 10 plates-formes, pouvant contribuer aux affinités modestes observées pour ces ligands. Cependant, les résultats obtenus montrent que la méthode de conception *in silico* utilisée est capable de fournir une série de ligands protéiques de 1ère génération ciblant de manière spécifique un site catalytique d'intérêt avec un bon rendement. Cette méthode pourrait constituer la 1ère étape d'une approche hybride de conception *in silico* de ligands combinée à des techniques de sélection expérimentales.

Abstract : Artificial mini-proteins able to target catalytic sites of matrix metalloproteinases (MMPs) were designed using a functional motif grafting approach. The motif corresponded to the 4 N-terminal residues of TIMP-2, a broad-spectrum natural protein inhibitor of MMPs. Scaffolds able to reproduce the functional topology of this motif as described in the TIMP-2/MMP-14 complex were obtained by exhaustive screening of the Protein Data Bank (PDB) using the STAMPS software (Search for Three-dimensional Atom Motif in Protein Structure). Ten artificial protein binders satisfying all topologic, steric and electrostatic criteria applied for selection were produced for experimental evaluation. These binders targeted catalytic sites of MMPs with affinities ranging from 450 nM and 590 μM prior to optimization. The crystal structures of two artificial binders in complex with the catalytic domain of MMP-12 showed that the intermolecular interactions established by the functional motif in these artificial binders corresponded to those found in the TIMP-2/MMP-14 complex, albeit with some differences in their geometry. Molecular dynamics simulations of the 10 binders in complex with MMP-14 suggested that these scaffolds could allow reproducing in part the native intermolecular interactions, but some differences in geometry and stability could contribute to the lower affinity of the artificial protein binders as compared to the natural one. Nevertheless, these results show that the *in silico* design method used can provide sets of starting protein binders targeting a specific binding site with a good rate of success. This approach could constitute the first step of an efficient hybrid computational-experimental protein binder design approach.