Conception in silico d'inhibiteurs peptidiques d'interactions protéine-protéine

Maxence Delaunay

THESE DE DOCTORAT

NNT : 2022UPASQ069

université PARIS-SACLAY

Conception *in silico* d'inhibiteurs peptidiques d'interactions protéine-protéine

In silico design of peptide inhibitors of protein-protein interactions

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°569, innovation thérapeutique : du fondamental à l'appliqué (ITFA)
Spécialité de doctorat : Chimie thérapeutique
Graduate School : Santé et médicaments. Référent : Faculté de Pharmacie

Thèse préparée dans l'unité de recherche **BioCIS (Université Paris-Saclay, CNRS)**, sous la direction de **Tâp HA DUONG**, Professeur des Université.

Thèse soutenue à Paris-Saclay, le 15 Décembre 2022, par

Maxence DELAUNAY

Composition du Jury

Membres du jury avec voix délibérative

Luba TCHERTANOV Directrice de Recherche, ENS Paris-Saclay, Université Paris-Saclay	Président ou Présidente
Chantal PREVOST Chargée de Recherche HDR, Institut de Biologie Physico-Chimique, Université Paris Cité	Rapporteure & Examinatrice
Pierre TUFFERY Directeur de Recherche, Unité de Biologie Fonctionnelle et Adaptative, Université Paris Cité	Rapporteur & Examinateur
Esther KELLENBERGER Professeur des Universités, Laboratoire d'Innovation Thérapeutique, Université de Strasbourg	Examinatrice
Thomas SIMONSON Professeur des Universités, Laboratoire de Biochimie, Ecole Polytechnique	Examinateur

université PARIS-SACLAY

ÉCOLE DOCTORALE
Innovation thérapeutique
du fondamental à l'appliqué
(ITFA)

Titre : Conception *in silico* d'inhibiteurs peptidiques d'interactions protéine-protéine

Mots clés : Dynamique moléculaire, Peptide cyclique, conception de peptides, Amarrage moléculaire, Modélisation moléculaire, Conception par fragments

Résumé : Le travail de recherche consistera à développer une méthode computationnelle pour concevoir des molécules peptidomimétiques cycliques ayant une affinité optimale pour des interfaces protéine-protéine, dans le but d'inhiber leurs interactions. L'approche envisagée sera basée sur l'amarrage moléculaire puis l'assemblage d'acides aminés naturels ou non sur des surfaces protéiques visées. Elle sera notamment appliquée à la conception d'inhibiteurs d'agrégation des peptides Aβ et IAPP impliqués respectivement dans la maladie d'Alzheimer et le diabète de type 2.

Le travail de recherche consistera à développer une méthode computationnelle pour concevoir des molécules peptidomimétiques cycliques ayant une affinité optimale pour des interfaces protéine-protéine, dans le but d'inhiber leurs interactions. L'approche envisagée sera basée sur l'amarrage moléculaire puis l'assemblage d'acides aminés naturels ou non sur des surfaces protéiques visées. Elle sera notamment appliquée à la conception d'inhibiteurs d'agrégation des peptides Aβ et IAPP impliqués respectivement dans la maladie d'Alzheimer et le diabète de type 2.

Title : *In silico* design of peptidic inhibitors of protein-protein interactions

Keywords : cyclic peptides, peptide design, molecular dynamics, molecular modeling, fragment-based design, molecular docking

Abstract : The PhD research will consist in developing a computational method to design cyclic peptidomimetic molecules with optimal affinity for protein-protein interfaces, in order to inhibit their interactions. The approach will be based on molecular docking and then assembly of amino acids, natural or not on targeted protein surfaces. It will notably be applied to the design of inhibitors of the Aβ and IAPP peptide aggregations, involved in Alzheimer's disease and type 2 diabetes, respectively.

The PhD research will consist in developing a computational method to design cyclic peptidomimetic molecules with optimal affinity for protein-protein interfaces, in order to inhibit their interactions. The approach will be based on molecular docking and then assembly of amino acids, natural or not on targeted protein surfaces. It will notably be applied to the design of inhibitors of the Aβ and IAPP peptide aggregations, involved in Alzheimer's disease and type 2 diabetes, respectively.

Table des matières

Remerciements

Je tiens tout d'abord à remercier chaleureusement mon directeur de thèse, le Professeur Tâp Ha Duong, pour son encadrement tout au long de ces trois années de thèse.

Tâp, je souhaite non seulement te remercier mais également te témoigner toute ma reconnaissance pour tout ce que tu as fais durant cette thèse. Bien que ces trois années furent très difficile sur le plan de la mobilité à cause d'une certaine pandémie, tu m'as également permis de défendre mes résultats dans plusieurs congrès. Tu m'as suivi au quotidien pour me prodiguer de précieux conseils tout en me laissant explorer les idées que j'avais en tête. Ainsi, je te remercie de ta confiance qui m'est très précieuse. Enfin, je te remercie également pour toutes les discussions non scientifiques du quotidien qui ont grandement contribuées à rendre ces trois ans au sein du laboratoire très agréables.

Ce fut un honneur, un privilège mais aussi un plaisir de réaliser ce travail sous ta supervision. Encore merci.

Je tiens ensuite à remercier grandement les Docteurs Chantal Prévost et Pierre Tufféry d'avoir accepté d'être les rapporteurs de cette thèse. Je remercie également le Docteur Luba Tchertanov ainsi que les Professeurs Esther Kellenberger et Thomas Simonson d'avoir accepté d'évaluer mon travail et d'être examinateurs de mon jury de thèse.

Je remercie également très chaleureusement toute l'équipe FLUOPEPIT au sein de laquelle j'ai effectué cette thèse, Sandrine, Benoît, Julia, Nicolo, Thierry et Jean-Louis. Merci pour tous vos conseils et les moments passés tous ensembles. Je tiens à adresser un merci tout particulier à Julia pour avoir mené les tests expérimentaux de ma thèse. Merci pour la grande pédagogie dont tu as fais preuve pour m'avoir expliqué dans le détail ces travaux et m'avoir permis de pouvoir y assister. Cette expérience fut très enrichissante pour moi et cela a permis d'avoir un premier retour expérimental très important pour l'avenir de ce projet. Un grand merci également à tous les doctorants, Jacopo, José, Davide, Nicolo, Monika, Sondes, les post-doctorants, Léa et Lizeth ainsi que les nombreux stagiaires, Sujith, Ariane, Nicolas et Mélissa. Vous avez contribué à rendre la vie au laboratoire très agréable.

I want to end the lab acknowledgments with a special thanks to the three chinese PhD students with whom I shared the whole three years. First of all, Ping that was in molecular modeling with me. Thank you for your help and the everyday's discussions. I was really happy to share the office with you. I wish you the best for your PhD defence. Special thanks also for Tingting and Chenghui. We spend a lot of good time together and it was a real pleasure to meet you. I hope to see all of you in China one day!!

A Guillaume et Kendra pour leur amitié qui m'est indispensable au quotidien. Une pensée affectueuse à votre petit Elwan qui lira ces lignes d'ici quelques années, le temps d'apprendre à lire (avant 6 ans, il y a intérêt!).

Un grand merci à tous mes amis de Pharmacie Coralie, Coralie, Jeanne, Marjorie et Maxence avec qui j'ai passé et je passe encore des moments formidables.

Un grand merci à mes deux parents, Corinne et Laurent, pour m'avoir soutenu dans toutes les étapes de ma vie et de m'avoir permis d'en arriver à la soutenance d'un doctorat aujourd'hui.

Enfin, je souhaite remercier ma femme, Phuong Dung, qui m'a soutenu toutes ces 12 dernières années et m'a grandement aidé durant ces années de thèse.

Chapitre 1

Introduction

1.1 Les interactions protéine-protéine à la base du vivant

1.1.1 Rôles biologiques des interactions protéine-protéine

Les interactions protéine-protéine (IPP) occupent une place capitale au sein du vivant. Elles permettent la formation de grands ensembles protéiques stables pour constituer une structure cellulaire comme dans le cas de capsides virales. Lorsque ces IPP possèdent un caractère plus transitoire, elles sont à la base des mécanismes de régulation et de signalisation cellulaires. Au sein d'une seule cellule, il est estimé qu'au moins 130 000 interactions protéine-protéine ont lieu à tout moment et il existerait plusieurs centaines de milliers d'IPP en biologie humaine, dont seule une petite fraction est actuellement connue [287, 25, 194, 239]. On comprend donc qu'elles sont indispensables au bon fonctionnement cellulaire. Ainsi, on observe un intérêt croissant de la communauté scientifique pour celles-ci. En effet, une étude de 2010 utilisant la base de données BioGRID (https ://thebiogrid.org/) répertoriait 33 943 IPP [25]. Si l'on utilise cette base de données à l'heure actuelle, plus de 60 000 interactions protéine-protéine sont répertoriées. Cette base de données a ainsi doublé en l'espace de 10 ans, montrant un intérêt croissant pour les IPP. Ces interactions sont principalement détectées grâce à deux méthodes : la technique dite de *yeast two-hybrid* (Y2H) et la co-immunoprécipitation suivie par une spectrométrie de masse (SM) [25].

La méthode Y2H est basée sur la greffe d'un gène "rapporteur" sur une des deux protéines d'intéret. On couple la moitié du facteur de transcription sur chacune des deux protéines exprimées dans la levure. Si les deux protéines interagissent et donc se rapprochent, la transcription du gène rapporteur sera effectuée et donc son ARNm exprimé et détecté. Il s'agit d'une approche haut

débit ayant permis une explosion du nombre d'IPP découvertes [25, 51]. La co-immunoprécipitation suivie d'une SM est basée sur le fait que deux partenaires protéiques co-précipitent en Western Blot et sont détectables par spectrométrie de masse.

Compte tenu de l'immense diversité des IPP et de leur implication dans la plupart des fonctions biologiques, on ne pourra décrire l'ensemble des rôles biologiques dans lesquelles elles sont impliquées. Nous nous appuierons donc sur deux exemples de fonctions majeures dans lesquelles sont impliquées les IPP afin de comprendre l'importance de celles-ci.

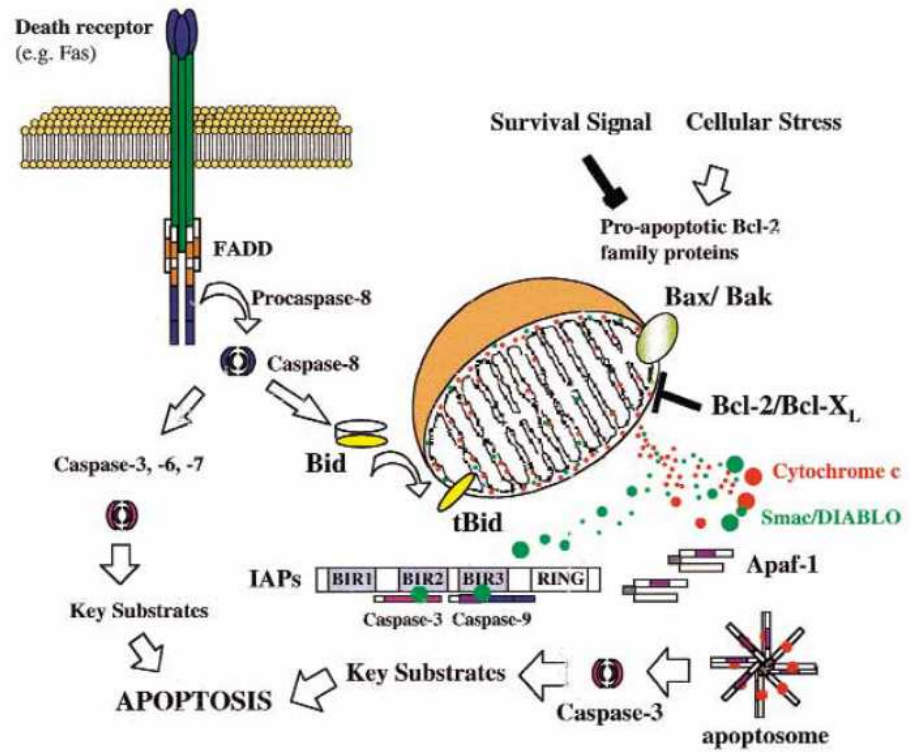

FIGURE 1.1 – Les différentes voies de l'apoptose [332]

L'apoptose ou mort cellulaire programmée (figure 1.1) est l'un des mécanismes cellulaires les plus finement régulés par une cascade de signalisations impliquant de nombreux complexes protéiques. Il s'agit d'un mécanisme pouvant avoir lieu selon deux voies principales : la voie des protéines de la famille Bcl-2 ou les récepteurs de mort [332].

Nous décrirons la voie impliquant les protéines de la famille Bcl-2 qui est particulièrement régulée par l'intervention de nombreuses IPP. En absence de stress cellulaire, l'apoptose peut en effet être inhibée par une protéine anti-

apoptotiques comme Bcl-XL en maintenant son interaction avec d'autres protéines proapoptotiques de la famille Bcl-2 telles que Bax ou Bak. Lors d'un stress cellulaire, les protéines proapoptotiques inhibées par Bcl-XL vont être translocuées du cytosol vers la mitochondrie où leurs partenaires antiapoptotique ne sont plus présentes. Une fois dans la mitochondrie, les protéines proapoptotiques permettent la libération du cytochrome c qui va lui-même entrainer l'oligomérisation des protéines Apaf-1. Lorsqu'Apaf-1 est oligomérisé. Conjointement les protéines Smac vont interagir avec les protéines IAPs afin de les inhiber. Les IAPS (XIAP, cIAP1 ...) étant inhibitrices de caspases, l'inhibition des IAPs entraine l'expression de caspases 3 et 9 qui vont réaliser l'apoptose par clivage protéique [332].

Cette voie de signalisation est essentielle dans le maintien d'un renouvellement cellulaire sain chez tout organisme. Ces différents mécanismes de régulation permettent d'assurer la survie des cellulaires saines et la mort de cellules dysfonctionnelles ou à risque tumorales.

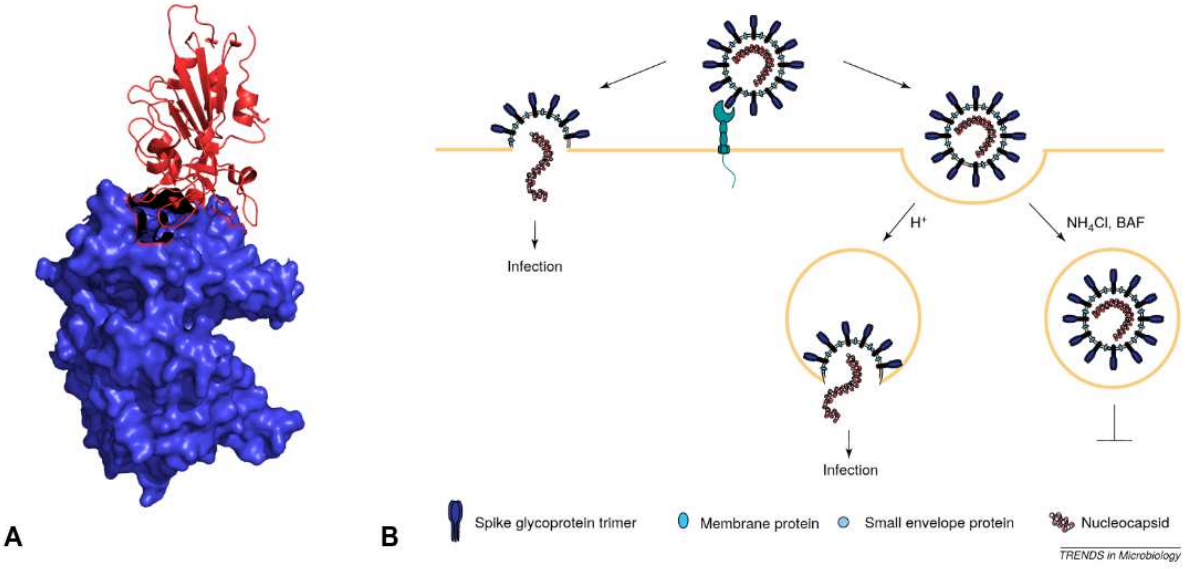

FIGURE 1.2 – A : Interaction entre la protéine Spike en rouge et ACE2 en bleu (PDB : 6M0J [155]). B : Mécanisme d'internalisation du SARS-CoV2 dans une cellule humaine [108].

Nous avons vu que les IPP pouvaient réguler des processus biologiques internes indispensables au bon fonctionnement d'un organisme. Cependant, elles peuvent aussi être utilisées comme point d'entrée cellulaire pour différentes bactéries ou éléments vivants au sein d'un autre organisme [237].

L'un des exemples actuel les plus connus étant l'interaction Spike/ACE2 (fi-

gure 1.2A) permettant l'internalisation du matériel génétique du SARS-CoV2 au sein des cellules humaines. La protéine Spike est exprimée à la surface du virus et va donc interagir avec le récepteur *angiotensin-converting enzyme 2* (ACE2) qui est habituellement impliquée dans la conversion et libération de l'angiotensine 1-7 favorisant la vasodilatation. Le virus est ensuite internalisé et son matériel génétique libéré dans certaines conditions (figure 1.2B) [108, 155].

1.1.2 Pathologie et dérégulation d'interactions protéine-protéine

Nous avons précédemment décrit l'importance des IPP dans la régulation fine de mécanismes cellulaires fins au sein d'un organisme vivant. On comprend donc qu'une simple dérégulation d'une seule de ces interactions peut conduire à de graves dysfonctionnements dans les cellules lorsqu'elle est impliquée dans une cascade de signalisation cellulaire. Ces dérèglements moléculaires sont à l'origine de très nombreuses pathologies. Les dérèglements des IPP sont principalement connus pour être impliqués dans de nombreux cancers, certaines maladies neurodégénératives ou encore diverses infections virales [237, 203, 108].

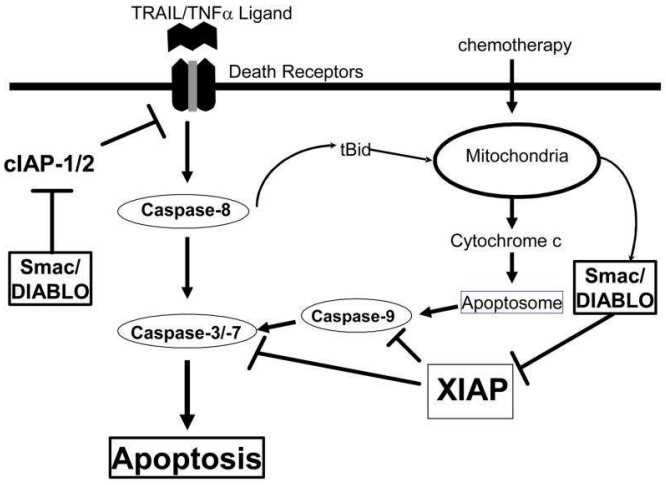

FIGURE 1.3 – Mécanisme de l'apoptose impliquant les protéines de la famille des IAPs et des Smac. Les IAPs (cIAP-1/2 et XIAP) inhibent l'expression des caspases proapoptotiques. Lorsqu'un signal de mort cellulaire apparaît, les protéines Smac favorisent l'apoptose par inhibition des protéines de la famille des IAPs [267].

Un premier type de dérégulation peut apparaître lorsqu'une interaction protéine-protéine est maintenue quelque soient les signaux cellulaires émis. Il s'agit

d'une dérégulation très souvent impliquée dans la prolifération tumorale. Cela peut premièrement résulter d'une surexpression d'un des partenaires protéiques. On sait en effet que la surexpression de la protéine antiapoptotique Bcl-XL (figure 1.1) est impliquée dans certains cancers tels que le carcinome du sein ou le cancer colorectal. On sait également que dans d'autres cancers tels que la leucémie lymphoïde chronique à cellules B ou certains cancers de la thyroïde, certaines protéines IAPs empêchant l'expression de caspases proapoptotiques sont surexprimées (figure 1.3) [277, 73].

Deuxièmement, au contraire, il arrive également qu'une interaction protéine-protéine ne soit pas suffisamment maintenue lorsque qu'un des partenaires protéiques, par exemple, est sous-exprimée. Ainsi, on sait que le déficit en antagonistes d'IAPs (antiapoptotique) tels que les Smac (favorisant l'apoptose par inhibition des IAPS, figure 1.3), est impliqué dans l'apparition de carcinomes pulmonaires ou de cancers de la vessie [73, 267].

Nous avons également vu un troisième type de dérégulation possible, par le biais de l'interaction Spike/ACE2, que ces voies pouvaient être utilisées par d'autres organismes vivants tels que le SARS-CoV2 mais également les Papillomavirus à but de réplication. Ainsi, un tel détournement conduit naturellement à l'apparition de pathologies chez l'organisme infecté. On peut ainsi expliquer l'apparition de nombreux symptômes pulmonaires graves dans la COVID par le fait que le récepteur ACE2 est particulièrement exprimé au niveau des bronches et des poumons chez l'être humain [25, 155].

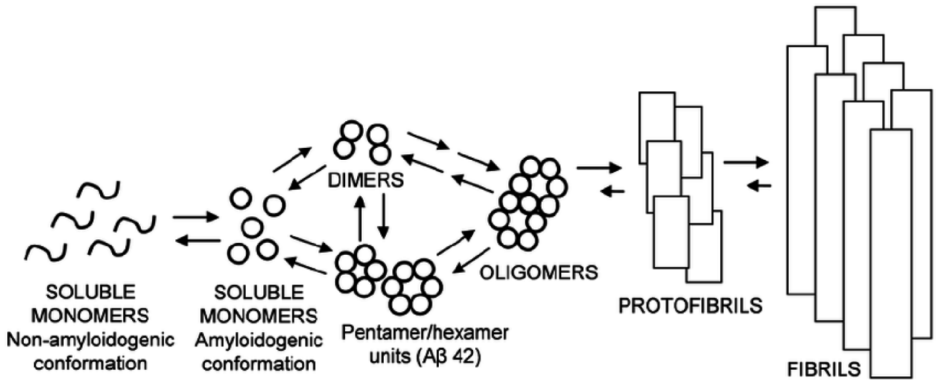

FIGURE 1.4 – Mécanisme moléculaire de la formation de la plaque amyloïde [9].

Enfin, Il est possible qu'une pathologie se déclare suite à l'apparition d'IPP non existantes chez un individu sain. C'est le cas de la formation de la plaque amyloïde résultante de l'agrégation du peptide Aβ responsable de la maladie d'Alzheimer. Il s'agit en effet du clivage de l'*amyloid precursor protein* (APP) par la γ-secretase, au lieu de l'α-secretase, qui aura pour conséquence la libéra-

tion du peptide Aβ dans l'espace extracellulaire. Aβ s'agrégera pour former des oligomères, des fibrilles puis la plaque amyloïde au niveau des neurofibrilles, responsable de la Maladie d'Alzheimer (figure 1.4) [31, 9].

1.1.3 Intérêt de moduler les interactions protéine-protéine en sciences pharmaceutiques par des peptides thérapeutiques

Nous avons précédemment vu l'impact que peuvent avoir les dérégulations d'IPP au sein d'une cellule et qu'elles peuvent mener à de sévères pathologies. Ainsi, la communauté scientifique s'est naturellement emparée de cette thématique afin de concevoir des principes actifs ciblant ces IPP afin de corriger une un possible dysfonctionnement de celles-ci. En effet, moduler une interaction protéine-protéine donnée permettrait d'agir de façon très efficace sur les causes d'une pathologie donnée [201].

Cependant, cibler les IPP s'est avéré plus difficile qu'escompté étant donné les caractéristiques structurales de leurs interfaces. En effet, celles-ci sont très souvent composée d'une cavité vaste et assez peu creusée dans laquelle va s'insérer un fragment peptidique du partenaire protéique [237]. Ces cavités rendent donc difficile l'utilisation de petites molécules en tant qu'inhibiteurs compétitifs compte tenu de leur faible affinité pour une cible si grande même s'il existe quelques succès comme certains inhibiteurs de Bcl-XL [323] ou encore quelques antiviraux ciblant les papillomavirus [237].

Les peptides à visée thérapeutiques semblent ainsi bien plus adaptés au ciblage d'interactions protéine-protéine. En effet, ils ont une taille plus adaptée aux caractéristiques structurales des IPP et ont donc une affinité et une spécificité beaucoup plus élevées envers leur cible [156].

Afin d'obtenir une vue d'ensemble au potentiel thérapeutique des peptides, nous avons réalisé une recherche systématique sur PubMed afin de regrouper les principaux peptides ou dérivés peptidiques modulateurs d'interactions protéine-protéine actuellement sur le marché ou en phase d'essais cliniques (tableau 1.1). Nous avons identifié ainsi une quinzaine de peptides ayant passé les premières phases d'essais cliniques, mais il faut aussi souligner que ces agents thérapeutiques sont de plus en plus nombreux dans les phases préliminaires de développement [156], montrant un intérêt récent mais croissant pour ce type de principes actifs.

Nous discuterons ainsi de plusieurs de ces peptides dans la partie concernant les méthodes expérimentales de développement de peptides modulateurs d'IPP afin d'illustrer la diversité d'approches possibles pour obtenir un médicament efficace avec des paramètres pharmacocinétiques optimaux.

Nom	Indications	Cible	Phase
eptifibatide	angor instable	intégrine $\alpha IIb\beta 3$	approuvé en 2016
cilengitide	cancer	intégrine $\alpha v\beta 3/\alpha v\beta 5$	phase III
balixafortide	cancer	CXCR4	phase II
solnatide	oedème pulmonaire	ENaC	phase II
ALRN-6924	cancer	MDM2/MDMX	phase II/III
ALRN-5281	maladies orphelines	GHRH	phase II/III
birinapant	cancer	XIAP/cIAP1	phase I/II
lcl-161	cancer	cIAP1/cIAP2	phase I/II
AT406	cancer	cIAP1/cIAP2	phase I
GDC-0917	cancer	cIAP1/cIAP2	phase I
delcasertib (KAI-9803)	infarctus du myocarde	δ-Protéine Kinase C	phase II (échec)
PPI-1019	Alzheimer	Aβ	phase I/II (échec en 2012)

TABLE 1.1 – Dérivés peptidiques sur le marché ou en essai clinique en 2022 [179, 55, 118, 295, 215, 88].

1.2 Stratégies expérimentales de conception de peptides modulateurs d'interactions protéine-protéine

Nous avons précédemment vu l'intérêt porté pour les peptides modulateurs d'IPP. Ainsi, de nombreuses stratégies expérimentales ont été mises en place afin d'entamer une démarche de conception rationnelle de peptides thérapeutiques (figure 1.5). L'une des techniques les plus importantes, ayant révolutionné la découverte de peptides pour une nouvelle cible donnée est le *phage display*, permettant un criblage haut débit de nombreuses séquences peptidiques dont les *hits* serviront de base à un futur principe actif. L'autre approche consiste à identifier un motif peptidique essentiel, adoptant souvent une structure secondaire particulière, à l'interaction avec une protéine cible [156].

Cependant, bien souvent, ces peptides ne possèdent pas de bons paramètres

pharmacocinétiques. En effet, ils sont rapidement reconnus et dégradés par les protéases sanguines. De plus, leur franchissement des barrières biologiques et des membranes est faible. Il convient donc de réduire leur nature peptidique afin de les rendre *druggable*. Ainsi, les séquences précédemment trouvées sont optimisées par l'ajout de nombreuses modifications chimiques avant d'en faire un candidat-médicament [156]. Parmi ces différentes techniques d'optimisation, nous détaillerons la cyclisation, la N-méthylation et l'ajout d'acides aminés modifiés à une séquence peptidique donnée.

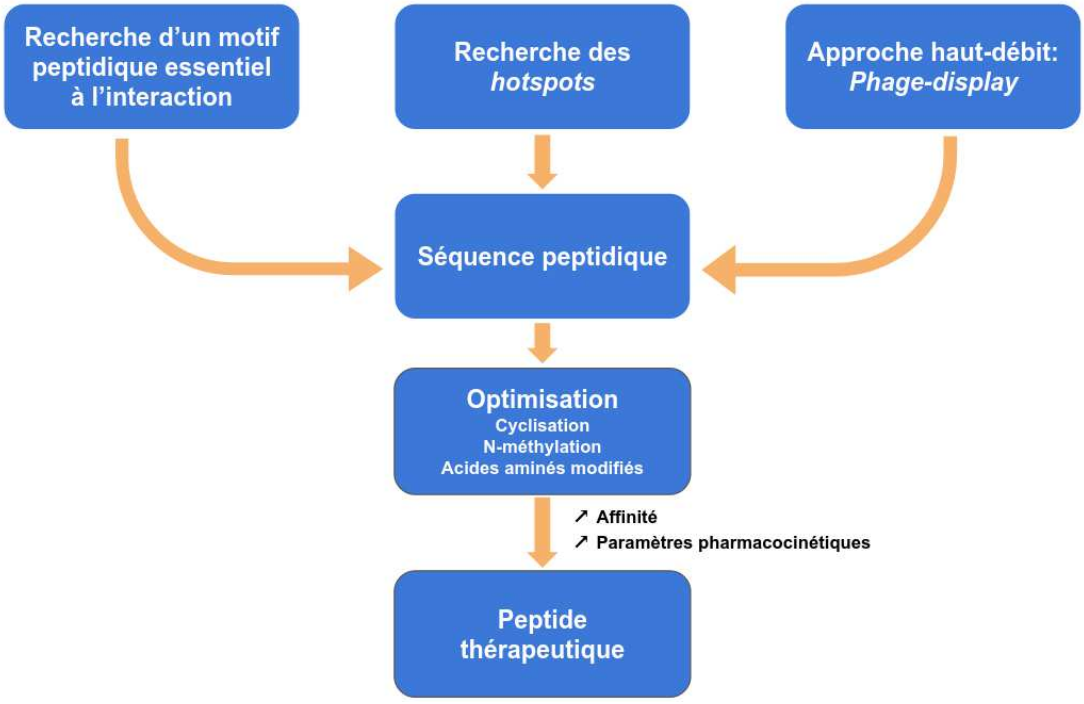

FIGURE 1.5 – Principales stratégies de conception expérimentale de peptides.

1.2.1 Identification de motifs peptidiques se liant à une protéine cible

De nombreuses interactions protéine-protéine se font par le biais d'un domaine globulaire d'une protéine et du fragment peptidique de son partenaire, adoptant bien souvent une structure secondaire particulière, fréquemment une hélice α ou un feuillet β. Lorsqu'il est identifié, ce fragment peptidique constitue le point de départ d'une séquence peptidique modulatrice qui sera optimisée ultérieurement [170].

L'identification de ce motif peptidique va se faire principalement par des mé-

thodes de caractérisations structurales classique telles que la cristallographie aux rayons X ou la RMN. Ces méthodes permettent d'élucider précisément la structure de l'interface entre deux protéines et va permettre le plus souvent d'identifier la structure secondaire qui servira de peptide modulateur. Cette séquence peut ensuite être affinée en mettant en oeuvre des tests d'interaction tels que la co-immunoprécipitation ou la *surface plasmon resonance* (SPR). Cependant, le caractère souvent transitoire des IPP rendent l'utilisation de ces méthodes difficile, ce qui a longtemps contribué à rendre très difficile le ciblage d'IPP en sciences pharmaceutiques [170].

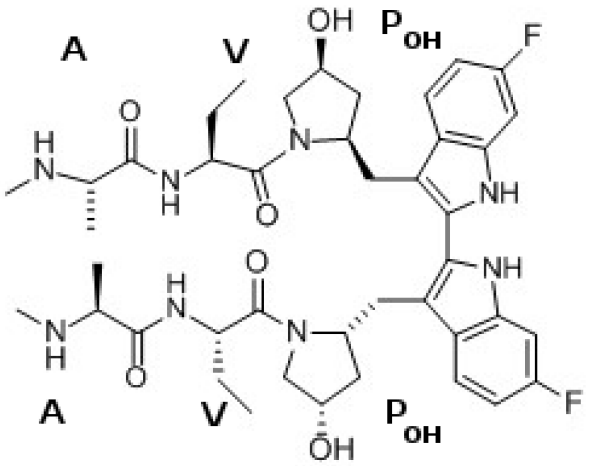

FIGURE 1.6 – Structure chimique du birinapant (P_OH est une hydroxyproline).

On peut toutefois citer un exemple de réussite de ces méthodes avec le birinapant (figure 1.6). Il s'agit d'un peptide inhibiteur de cIAP1 et XIAP qui va agir comme un mime des protéines Smac qui va permettre ainsi la libération des caspases et donc favoriser l'apoptose. Il est utilisé dans le traitement du cancer des ovaires avec résistance au traitement par le cisplatine, du cancer colorectal et du mélanome [206, 19]. Sa structure chimique possède deux motifs AVPI modifiés chimiquement. Cet inhibiteur a été développé selon la démarche précédemment détaillée. En effet, la résolution de la structure tridimensionnelle du complexe Smac/XIAP a mis en évidence l'importance du fragment AVPI de Smac dans l'interaction avec sa cible. Ce motif peptidique a ensuite été optimisé par diverses modifications chimiques, dont les deux plus importantes furent de remplacer l'isoleucine (I) par un cycle indole fluoré et de créer un motif dimérique d'AVPI. Ce dimère mime de façon plus complète l'interaction Smac/AVPI et a permis d'obtenir un K_D inférieur à 1 nM [19, 33, 281].

Enfin, il a également été mis en évidence que dans la plupart des interfaces protéine-protéine, un faible nombre de résidus contribue de façon majoritaire à l'énergie libre de liaison. Ces résidus sont appelés *hotspots*. La découverte de ces résidus essentiels à une IPP donnée a permis de faire émerger différentes stratégies de conception médicamenteuse. En effet, diverses petites molécules ont été conçues pour interagir avec les *hotspots* de façon spécifique pour réguler une IPP. En ce qui concerne la conception de peptides, cela permet de connaître les résidus d'un fragment peptidique que l'on doit absolument conserver lors de la phase d'optimisation. La détection des *hotspots* se fait à l'aide de la technique de l'*alanine scanning*. Il s'agit d'une méthode de substitution systématique des résidus d'une interface protéine-protéine par une alanine et d'analyser les conséquences sur l'interaction. Si l'interaction est diminuée ou perdue, cela signifique que le résidu muté est un *hotspot* [194].

1.2.2 Une approche haut-débit : le *phage display*

L'une des approches expérimentales les plus fréquemment utilisées dans la recherche de peptides inhibiteurs d'IPP est le *phage display*. Il s'agit d'une approche haut débit consistant à tester un très grand nombre de peptides sur une protéine cible donnée.

Cette méthode consiste à faire s'exprimer les peptides issus d'une librairie à la surface de micro-organismes (phages ou virions dans le cas de peptides macrocycliques [208]). Cette librairie peut être composée de peptides naturels linéaires, mais également de peptides cycliques, reliés à leurs extrémités par un pont disulfure [208]. Cela permet d'avoir une très grande diversité de structure à tester. Ensuite, la protéine cible est fixée sur une surface solide. Les micro-organismes exprimant les peptides à leur surface seront ensuite exposés à ces protéines. Après lavage, les phages restant accrochés à la surface solide sont ceux dont les peptides de surface interagissent avec une bonne affinité avec la protéine cible. Ces phages sont élués puis amplifiés pour récupérer leurs séquences peptidiques d'intérêt (figure 1.7) [122, 262, 210, 301].

Le *phage-display* est actuellement l'une des techniques les plus efficaces pour identifier de potentiels peptides modulateurs d'IPP. Cette méthode est à l'origine des nombreux composés prometteurs tels que des inhibiteurs de l'interaction entre la protéine p300 et HIF, qui est impliquée dans le mécanisme tumoral. Les études préliminaires de ces composés montrent un potentiel thérapeutique très intéressant [151].

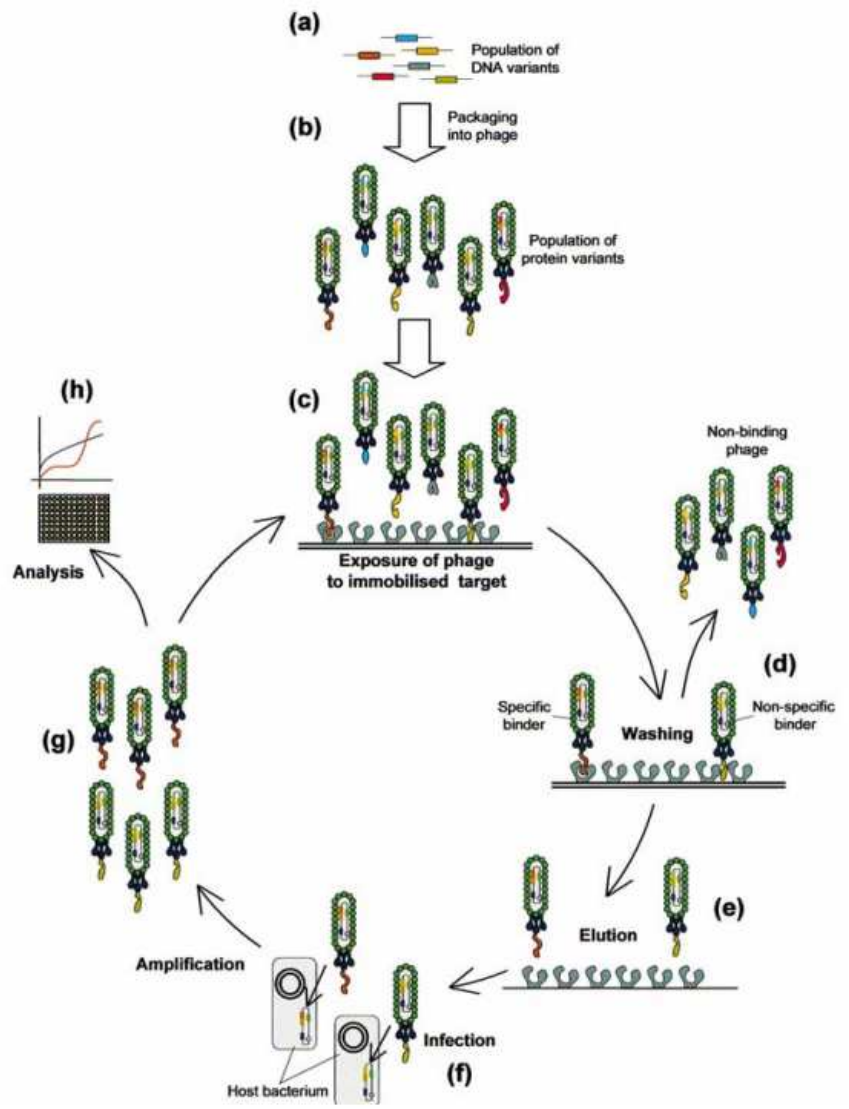

FIGURE 1.7 – Schéma de la méthode *phage-display*. (a) Synthèses des séquences d'ADN codant pour une librairie de peptides. (b) Intégration de la séquence aux phages. (c) Exposition des phages à la protéine cible. (d) Lavage des phages non en interaction. (e) Elution des phages liés afin d'infecter des bactéries hôtes (f) qui vont répliquer les séquences des peptides liés afin d'amplifier le signal (g) pour analyser les séquences sorties de l'expérimentation (h) [301].

1.2.3 Optimisation de peptides

Une fois trouvées les meilleures séquences modulatrices d'une IPP donnée, il est nécessaire d'optimiser le peptide afin d'en faire une molécule à visée thé-

rapeutique. En effet, les peptides sont connus pour avoir une faible biodisponibilité. Leur demi-vie est bien souvent très courte de par leur dégradation rapide par les protéases sériques. De plus, leur perméabilité membranaire est la plupart du temps très faible.

Ainsi, pour améliorer les paramètres pharmacocinétiques de ces composés, il convient de réduire leurs natures peptidiques par le biais de différentes méthodes, dont la cyclisation, la N-méthylation et l'ajout d'acides aminés modifiés, que nous décrirons ci-après.

1.2.3.1 La cyclisation

L'une des méthodes les plus courantes d'optimisation de peptide est la cyclisation. En effet, la nature moins flexible de ceux-ci leur permet tout d'abord d'avoir un coût entropique d'association plus bas que leur homologue linéaire et donc augmente leur affinité pour leur cible. La neutralisation des charges en N-terminal et C-terminal des peptides cycliques *head-to-tail* permet également non seulement de réduire leur reconnaissance par les protéases et donc d'augmenter leur biodisponibilité, mais également d'augmenter leur hydrophobicité, et leur perméabilité membranaire [299].

Il existe plusieurs méthodes de cyclisation peptidique qui peuvent être regroupées en plusieurs classes : les cyclisations *head-to-tail*, les cyclisations d'hélices-α en agrafe et les cyclisations de feuillets-β [299].

La cyclisation *head-to-tail*

La cyclisation *head-to-tail* est le fait de relier de façon covalente les extrémités N-terminal et C-terminal du peptide. Elle peut être effectuée par liaison peptidique simple en synthèse sur support solide en condition très diluée. La cyclisation par formation d'un pont disulfure via l'ajout de deux cystéines aux extrémités est également très fréquemment utilisée pour cycliser un peptide.

Cependant, compte tenu de la relative flexibilité d'un pont disulfure ou d'une liaison peptidique, ces méthodes peuvent parfois être insuffisantes pour maintenir un peptide dans une conformation voulue. Ainsi, il est également possible d'utiliser des groupements chimiques non peptidiques appelés *linkers* afin de relier les deux extrémités du peptides et former des monocycles (figure 1.8A) ou des bicycles (figure 1.8B) de conformation beaucoup plus stable. Ces *linkers* sont incorporés au peptide par un mécanisme de type substitution nucléophile d'ordre 2 (S_N2) [41, 15, 287].

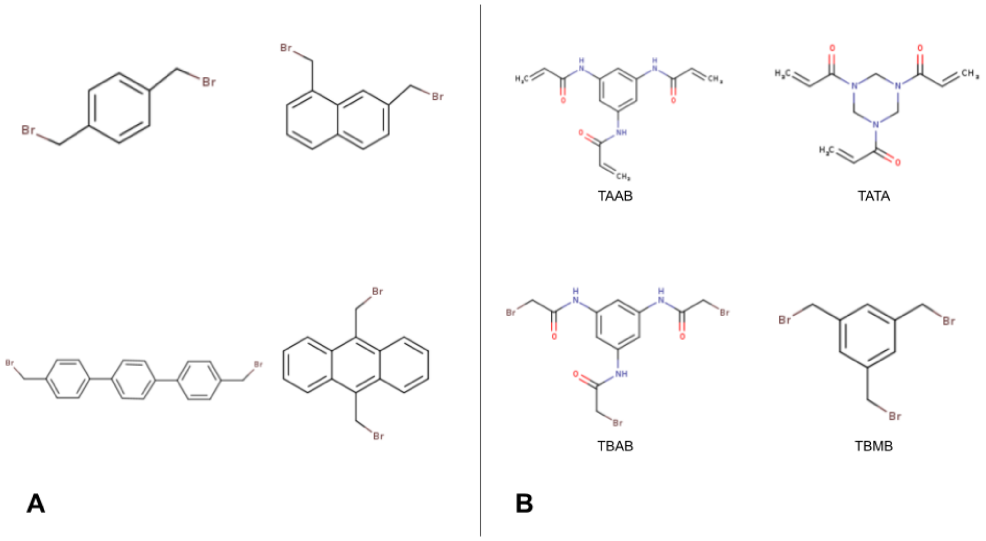

FIGURE 1.8 – Structures chimiques des *linkers* utilisés pour former des monocycles (A) ou des bicycles (B) peptidiques.

Plusieurs peptides cyclisés aux extrémités sont actuellement en cours d'essais cliniques. L'un des premiers approuvés par les autorités sanitaires est l'eptifibatide (figure 1.9), un peptide cyclique dérivé d'une toxine de serpent. Il est utilisé dans le cadre de symptômes coronariens en tant qu'antiagrégant plaquettaire [95] en empêchant l'interaction entre le facteur Willebrand et l'intégrine αIIbβ3. L'eptifibatide possède un motif Arginine-Glycine-Aspartate (RGD) identifié comme essentiel à l'interaction du peptide avec les protéines de la famille des intégrines [250]. Cependant, c'est bien la cyclisation par un pont disulfure stabilisant la conformation de ce motif qui a permis la spécificité de celui-ci pour l'intégrine αIIbβ3 [245, 244].

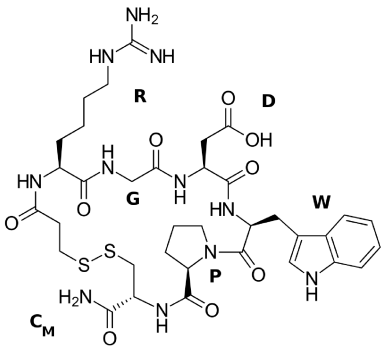

FIGURE 1.9 – Structure chimique de l'eptifibatide.

Les hélices-α agrafées

Lorsque le peptide d'intérêt thérapeutique adopte une structure secondaire en forme d'hélice-α, la cyclisation N-terminal-C-terminal n'est plus réalisable. Dans ce cas précis, la méthode de cyclisation la plus fréquente consiste en l'ajout d'une agrafe entre les chaînes latérales de deux résidus situés du même côté de l'hélice afin de stabiliser sa structure. Ces agrafes sont souvent des hydrocarbures, un pont disulfure ou des thioesters. Le plus fréquemment, la cyclisation est réalisée par métathèse d'alcènes (figure 1.10) entre un acide aminé modifié et son troisième, quatrième ou septième voisin [247].

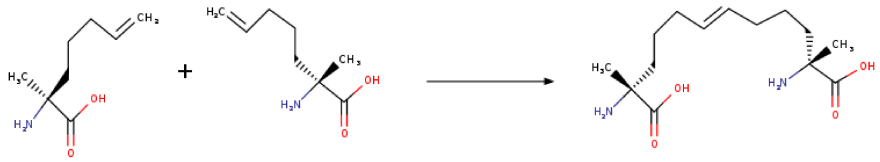

FIGURE 1.10 – Métathèse d'alcènes permettant la cyclisation d'hélices α.

Une autre possibilité pour réaliser cette agrafe est la cycloaddition entre un alcyne et un azoture (aussi appelée *"click chemistry"*) pour former un triazole qui connectera les deux acides aminés (figure 1.11) [201, 6].

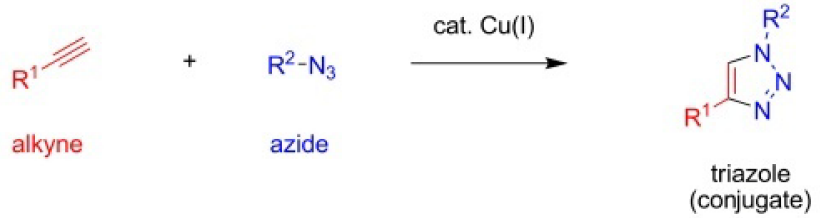

FIGURE 1.11 – Réaction d'un alcyne avec un azoture ayant pour produit un triazole [2].

Cette modification chimique permet principalement d'améliorer les paramètres pharmacocinétiques du peptide d'intérêt thérapeutique. En effet, en plus d'augmenter la stabilité conformationnelle de celui-ci, cette cyclisation est connue pour permettre au peptide d' échapper plus facilement à la dégradation par les protéases sanguines. Enfin, les hélices-α agrafées voient leur perméabilité cellulaire accrue de par une augmentation de leur hydrophobicité (les fonctions carboxyles et amines du squelette étant moins exposés au solvent) et un mode d'internalisation cellulaire qui leur serait propre [299, 50].

Plusieurs exemples de peptides contenant cette modification ont franchi l'étape des essais cliniques. L'un des exemples les plus connus étant l'ALRN-6924

(figure 1.12) utilisé dans le traitement de tumeurs solides ou hématologiques en empêchant l'interaction MDM2/MDMX, permettant ainsi au facteur p53 d'induire l'apoptose tumorale [1, 50].

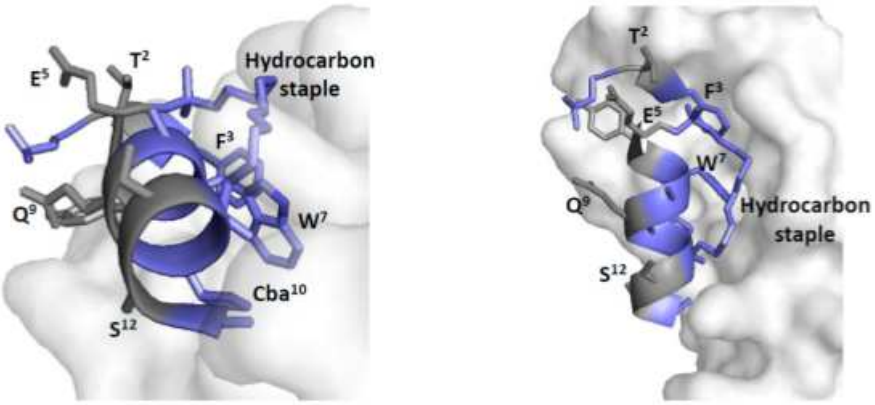

FIGURE 1.12 – Structure de l'ATSP (bleu), précurseur de l'ALRN-6924 (dont la structure est confidentielle), en interaction avec MDMX (blanc) [1].

Les feuillets β cyclisés

La cyclisation peut également permettre de stabiliser des peptides adoptant une conformation en feuillets β. Il est ainsi possible de favoriser une conformation en *beta-hairpin* en liant de façon covalente deux chaînes latérales (figure 1.13). Cependant, la méthode la plus fréquente consiste à ajouter un mime de coude β possédant un motif L-Pro-D-Pro. Ce motif est inséré aux extrémités libres de chacun des deux brins β [80].

Il existe actuellement un certain nombre de ces peptides en études précliniques. Un exemple prometteur est un inhibiteur compétitif de l'interaction HDM2/p53. Son interaction avec la protéine HDM2 permet la libération du facteur suppresseur de tumeur p53 et favorise donc l'apoptose tumorale. L'affinité de cet inhibiteur a été grandement accrue grâce à la stabilité du β*-hairpin* induite par la cyclisation[80, 299].

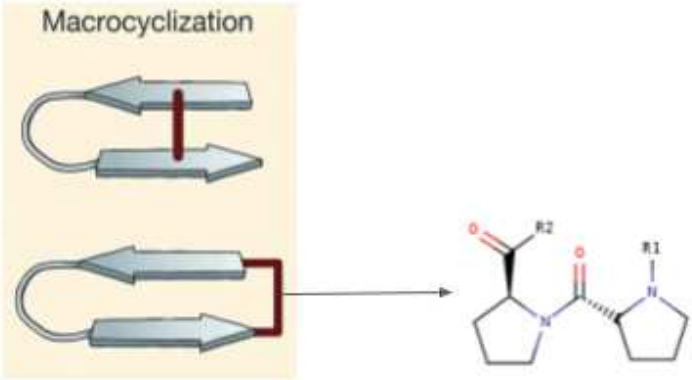

FIGURE 1.13 – β-*hairpins* stabilisés par liaison covalente entre deux chaînes latérales (haut) ou par mime de β-*turn* [299]

1.2.3.2 La N-méthylation

La N-méthylation est l'ajout d'un groupement méthyle sur un ou plusieurs azotes de la chaîne principale d'un peptide. Une telle modification chimique a pour conséquences un changement dans l'équilibre *cis*/*trans* des liaisons carbone/oxygène ou encore de réduire la flexibilité d'un peptide. De plus, un acide N-méthylé possède une analogie structurale avec la proline et peut donc être un stabilisateur de β-*hairpin*. C'est pourquoi la N-méthylation est retrouvée dans certains peptides inhibiteurs de l'agrégation du peptide amyloïde Aβ connu pour s'organiser en nombreux feuillets β. Néanmoins, ces composés sont encore à des stades relativement précoces de développement [142, 29].

On trouve néanmoins un exemple de peptide N-méthylé en phase II des essais cliniques : le cilengitide (figure 1.14). Tout comme l'eptifibatide décrit précédemment, il est un inhibiteur de protéines de la famille des intégrines. Il cible quant à lui les $\alpha_V\beta 3$ et $\alpha_V\beta 5$ et est utilisé dans le traitement du glioblastome de type II [186, 54].

En effet, ces intégrines sont impliquées dans l'angiogénèse tumorale et les inhiber freine donc le processus de vascularisation de la tumeur. Tout comme l'eptifibatide, le cilengitide possède le motif RGD nécessaire à l'interaction avec les intégrines mais acquiert sa spécificité d'action par la présence d'une phénylalanine et d'une valine. Dans le cas de ce peptide, la N-méthylation entre ces deux acides aminés a grandement augmenté l'affinité du peptide pour ses cibles grâce au maintien conformationnel de ces deux résidus [186, 35].

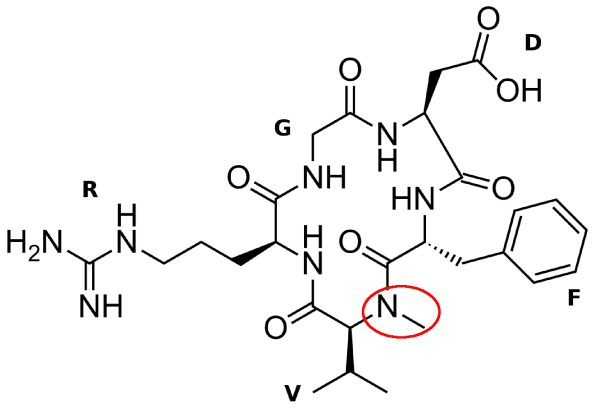

FIGURE 1.14 – Structure chimique du cilengitide

1.2.3.3 Ajout d'acides aminés modifiés

Enfin, la dernière stratégie d'optimisation peptidique que nous détaillerons consiste en l'ajout d'acides aminés modifiés à une séquence peptidique donnée. Le principal avantage d'insérer ces acides aminés modifiés va être de réduire la nature peptidique du composé. Ainsi, il échappera plus facilement au site de reconnaissance des protéases sanguines. De plus, de tels ajouts peuvent parfois fortement augmenter l'affinité et la spécificité d'un composé pour sa cible. Nous détaillerons d'abord deux modifications ciblant la chaîne principale d'un acide aminé. Nous insisterons ensuite sur les acides aminés fluorés puisqu'un chapitre entier des travaux de cette thèse leur est dédié.

Les modifications apportées sur la chaîne principale

Il existe deux principaux types d'acides aminés ayant une chaîne principale modifiée : Les acides-β-aminés et les peptoïdes (figure 1.15).

Un acide aminés β porte un carbone supplémentaire dans sa chaîne principale (figure 1.15B), ce qui place la fonction amine-terminale en β de l'acide-terminal. Les acides β aminés peuvent être incorporés dans des β-peptides (entièrement consistués d'acides β aminés) ou des α/β peptides (composés à la fois d'acides aminés α et d'acides aminés β). Ces acides aminés sont essentiellement conçus pour améliorer la stabilité protéolytique du peptide dans lequel ils sont incorporés. A titre d'exemple, de nombreux peptides comportant ces composés se sont avérés être de très bon inhibiteurs de l'interaction MDM2-p53 impliquée dans le cancer. En effet, les acides aminés β permettent aux peptides d'adopter une conformation en hélice α très stable et prédictible [201].

Le deuxième type de modification de la chaîne principale évoqué ici concerne

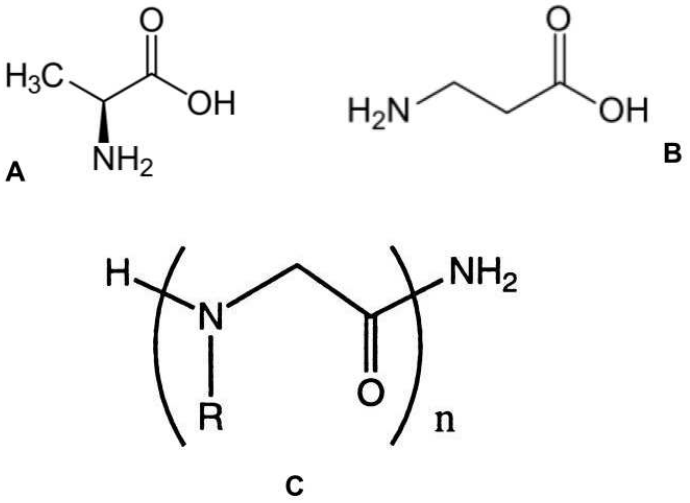

FIGURE 1.15 – (A) α-alanine. (B) β-alanine. (C) α peptoïde.

les peptoïdes (figure 1.15C). Ce sont des acides aminés dont la chaîne latérale est portée par la fonction amine. Ils ne sont que très peu dégradés par les protéases sériques compte tenu de leur nature peptidique très fortement altérée et auraient une perméabilité membranaire augmentée. Ces peptoïdes peuvent former un peptide entier, ou être incorporé dans une séquence peptidique naturelle. Ces composés ont déjà été testés en préclinique sur la protéine CXCR4 impliquée dans l'internalisation cellulaire du VIH. Un peptoïde hélicoïdal (figure 1.16) a également déjà été testé en tant qu'inhibiteur de l'interaction p53-HDM2 impliquée dans divers cancers [139, 268, 66].

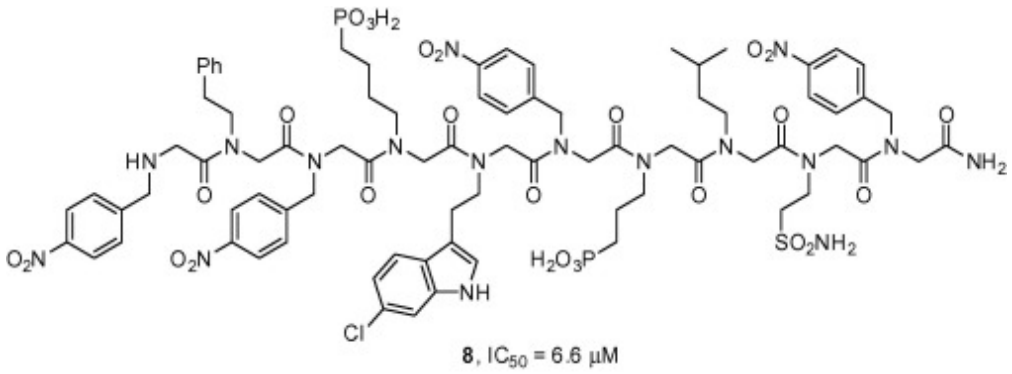

FIGURE 1.16 – Structure chimique d'un peptoïde inhibiteur de l'interaction HDM2/p53.

Les acides aminés fluorés

Le dernier type de modification dont nous parlerons dans cette introduction consiste en la fluoration d'acides aminés. Il sera détaillé également au chapitre 5 puisqu'il s'agit de tout un volet de ce travail de recherche.

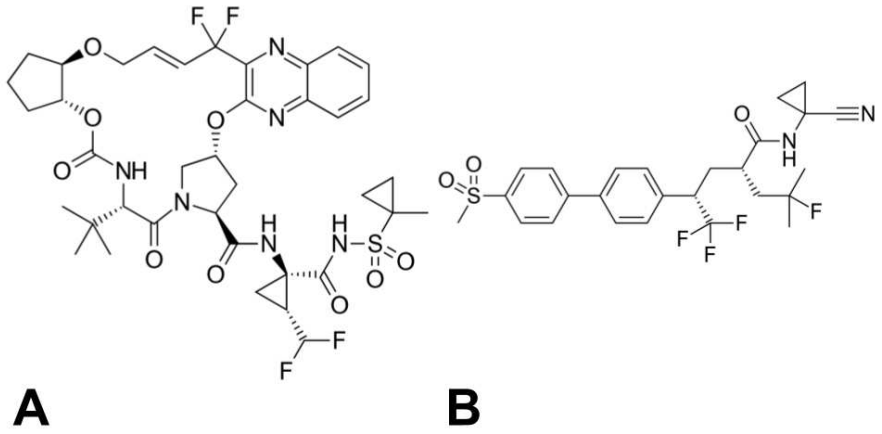

FIGURE 1.17 – Structure chimique du glecaprevir (A) et de l'odanacatib (B).

Les acides aminés fluorés ont pour principal avantage d'augmenter très fortement les stabilités thermique et chimique d'un composé et de le maintenir dans une structure secondaire donnée, que ce soit une hélice α ou un feuillet β [28]. Ils peuvent également augmenter l'affinité d'un peptide pour sa cible notamment par une augmentation de l'hydrophobicité du composé due au fluor. Cette hydrophobicité permet également d'augmenter la perméabilité cellulaire [28, 94].

Actuellement, on estime entre à 9 et 11 molécules fluorées approuvées par la FDA chaque année [191, 105]. Bien que cela ne concerne pas que les inhibiteurs d'IPP, il a été cité une quarantaine de peptides fluorés sur le marché et en phase clinique par *Mei et al* [190]. Ils sont utilisés dans de très nombreuses pathologies comme certaines maladies métaboliques, les infections virales et bactériennes, ou encore la maladie d'Alzheimer ou certains cancers [190, 3, 116, 184, 192, 57, 223].

On peut citer comme exemples le cas du glecaprevir (figure 1.17A), inhibiteur des protéases NS3/4A retrouvées chez le virus de l'hépatite C ou encore l'odanacatib, utilisé dans le traitement de certaines formes d'ostéoporose par inhibition de la cathépsine K (figure 1.17B) [190].

Chapitre 2

Approches computationnelles de conception de peptides inhibiteurs d'interactions protéine-protéine

Nous avons précédemment introduit les méthodes expérimentales de conception de peptides inhibiteurs d'interactions protéine-protéine ainsi que certaines de leurs limites. En effet, l'aspect souvent transitoire des interactions protéine-protéine rend parfois difficile la caractérisation structurale par les méthodes expérimentales classiques. Ainsi, il peut être difficile d'obtenir un motif peptidique minimal nécessaire à une IPP. De plus, les méthodes expérimentales dites de haut-débit telles que le *phage display* peuvent parfois être longues à réaliser et sont onéreuses. Le *phage display* est également difficile à mettre en place lorsque la cible est le résultat de l'oligomérisation de peptides comme dans le cas de protofibrilles d'Aβ. Enfin, même lorsque l'on obtient une séquence peptidique d'intérêt, les méthodes d'optimisation expérimentales nécessitent beaucoup de tests coûteux et longs à mettre en oeuvre afin de pouvoir évaluer l'influence d'une modification sur l'activité d'un peptide.

Ainsi, de nombreuses méthodes computationnelles ont vu le jour afin de pouvoir proposer plus rapidement et à moindre coût, diverses séquences peptidiques d'intérêt. Elles permettent également de proposer des inhibiteurs d'IPP même lorsque très peu d'informations sur la structure de celles-ci sont connues. Ainsi, nous présentons dans ce chapitre 1 les méthodes computationnelles de conception de peptides inhibiteurs d'interactions protéine-protéine. Nous avons présenté ces différentes approches selon qu'elles soient basées sur la recherche de séquences possédant des propriétés biologiques particulières ou des propriétés conformationnelles requises dans l'interaction avec une cible données (stratégies *peptide-based*) ou qu'elles soient basées sur l'utilisation de données structurales de l'interface protéine-protéine ciblée (stratégies *target-based*).

Ce chapitre est l'adaptation du chapitre de livre *Computational Tools and Strategies to Develop Peptide-Based Inhibitors of Protein-Protein Interactions, 2022* par *Delaunay and Ha-Duong* publié dans *Computational Peptide Science, Methods and Protocol, Methods in Molecular Biology, 2022* [64].

2.1 Introduction

Association and dissociation of proteins are molecular events at the basis of many crucial cellular processes. Therefore, perturbation of protein interaction networks generally leads to severe human diseases such as cancer or degenerative diseases. Infectious diseases also involve interactions between host and pathogen proteins [236]. Accordingly, protein-protein interactions (PPIs) have become the target of an increasing number of modulator molecules with therapeutic perspectives but also as chemical biology tools to study protein interactomes [194]. Notably, one advantage of targeting PPIs compared with single proteins is to reduce the probability of drug resistance. Indeed, protein-protein interfaces being highly complementary, a mutation in one protein would require a second complementary mutation in its partner to preserve their association, which is very unlikely [12].

Despite the generally large protein surfaces involved in PPIs, several small molecules were successfully developed for inhibiting protein-protein associations [254, 197, 300]. Nonetheless, similarly to small molecules which competitively inhibit enzymes by mimicking endogenous substrates, peptide derivatives which mimic peptide segments at protein-protein interfaces should be highly specific lead compounds. Moreover, many protein-protein associations, particularly in signaling pathways, are characterized by low affinities [318]. This offers a lot of room for developing peptidic binders with higher affinity. For these reasons, peptides are very promising starting points for deriving potent and selective PPI inhibitors [201, 55].

However, peptides exhibit well-known *in vivo* issues which have to be circumscribed in drug development projects : They are easily degraded by proteolytic enzymes, they generally have poor membrane permeability, and they potentially induce unwanted immune responses [85]. These drawbacks can be limited by reducing their peptidic nature with various approaches such as cyclisation, N-methylation, or incorporation of non-natural amino acids. These peptide derivatives are expected to keep a potent activity on targeted protein-protein interactions but with improved pharmacokinetic properties. Nonetheless, the daunting task in designing peptide derivatives to modulate protein-protein associations remains to find their optimal sequence of natural or non-natural amino acids.

In order to assist the medicinal chemists and chemical biologists in this task,

many computational approaches have been developed for the past decades. We attempt here to categorize the various *in silico* strategies that use these computational tools to design peptide-based molecules modulating PPIs. Similarly to the design of small molecules inhibiting enzymes or receptors, strategies to develop therapeutic peptide derivatives can be classified into ligand-based or structure-based approaches. In the first case, a peptide segment with interesting bioactivity has been identified but its targeted protein remains unknown. Peptide derivatives with improved properties will then be searched by similarity with the initial peptide segment. In the second case, the structure of the protein-protein complex is partially or entirely known and the strategies will consist in finding peptide derivatives having an optimal structural and chemical complementary with the targeted protein surfaces. Accordingly, the present chapter will be organized into two parts, the peptide-based and target-based strategies.

2.2 Peptide-based strategies

In ligand-based virtual screening, the search for small molecules similar to a query compound consists in comparing molecular descriptors which encode chemical and structural information into numbers. For example, molecular fingerprints encode in bit vectors the presence in molecules of particular substructures or fragments [32] and pharmacophore models capture in three-dimensional arrangements the chemical features of a ligand that are necessary for its bioactivity [134]. Regarding peptides, chemical information is encoded in their sequence and conformational information mainly lies in their secondary structures. Accordingly, we subdivided this section into sequence-based and conformation-based strategies.

2.2.1 Sequence-based approaches

Various properties of a peptide can be inferred by searching for homologous sequences in relevant databases. This can be routinely performed by using sequence alignment methods, such as the basic local alignment search tool (BLAST) [8]. For example, BLAST searches in database of essential genes (DEG) [322] or protein subcellular localization database (PSORTdb) [230] can rapidly provide useful information regarding the biological function of query protein sequences to identify novel targets [91]. Likewise, many databases dedicated to peptides, such as database of antimicrobial activity and structure of peptides (DBAAS) [219] or database of bioactive peptides (BIOPEP-UWMTM) [195], can be mined using BLAST to obtain information about biological and therapeutical acivities of query peptide sequences.

However, similarity searches based on purely sequential representation of proteins or peptides can fail in case of low coverage or low quality sequence alignments. Thus, alternative descriptors of protein and peptide chemical information can be used, such as the amino acid composition (AAC) which is simply the occurrence frequencies of the 20 native amino acids in their sequence, the dipeptide composition, or the pseudo amino acid composition (PseAAC) concept which additionally includes some sequence-order information via correlation factors of various chemical properties between (i,i+1), (i,i+2), and (i,i+3) pairs of amino acids [48]. By comparing these sequence descriptors with those of proteins and peptides in relevant databases, it is again possible to predict some of its functional, physical chemical, and structural features [225].

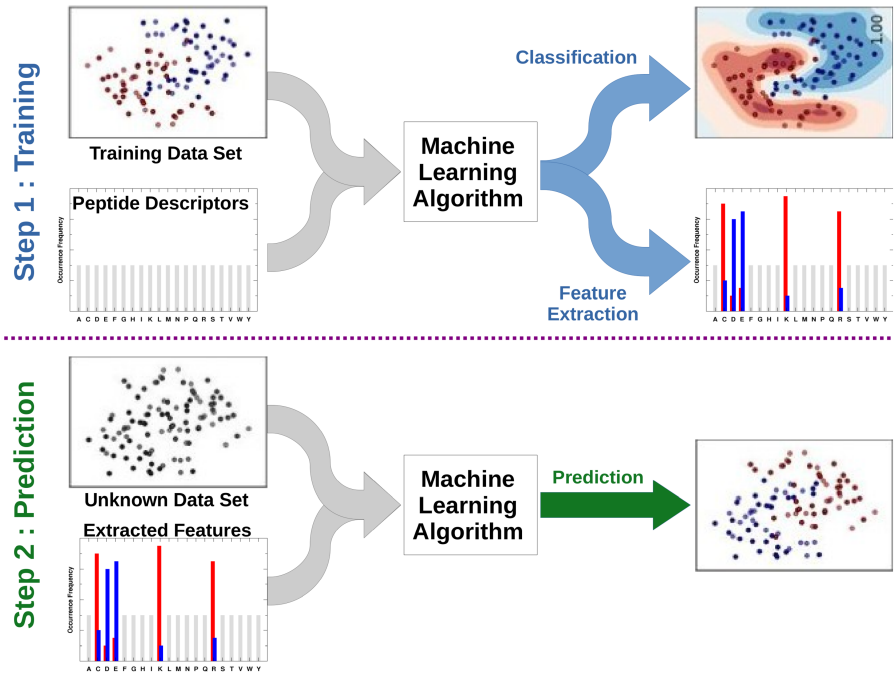

FIGURE 2.1 – Schematic description of sequence-based prediction of peptide therapeutic property through machine learning algorithms.

The common thread of sequence-based predictions of peptide bioactivity is first to build a training set of peptides with the experimentally validated desired and non-desired property, to choose some peptide chemical descriptors, and to train a machine learning algorithm for determining the relevant chemical features which separate the desired from non-desired peptides (Fig. 2.1). In a second stage, the trained machine learning algorithm is applied to unknown data sets of peptides to predict those with desired and non-desired properties according to the most discriminating features. In subsections below, we high-

light some studies which applied these sequence-based approaches to predict peptide bioactivity.

Prediction of peptide therapeutic property

Many studies applied this general framework to identify peptides against almost all main pathology classes. Among them, prediction of anticancer peptides has attracted great interest from several research groups [43, 307, 297]. By building specialised data sets of anti-angiogenic peptides, other predictors were developed for identifying more specific peptides inhibiting angiogenesis as promising cancer treatment [23, 152]. A second major class of diseases for which sequence-based searches for therapeutic peptides have been reported is infectious diseases. According to the property covered by the training data sets, several models were developed to predict antimicrobial [21], antibacterial [136], or antiviral [246] peptides.

To avoid peptide-induced unwanted immune responses, or, conversely, to design peptide-based vaccines, it can be interesting to anticipate the peptide immunogenicity property. To this end, several studies combining data sets of binders and non-binders to the major histocompatibility complex (MHC) proteins and various machine learning algorithms have been used to predict whether a peptide is likely to be a T-cell epitope presented on the cell surface [280, 130]. It is worthy to note that most recent peptide training data sets were built from the Immune Epitope Database (IEDB) [285] which also includes biological data about peptides involved in inflammatory disorders. Thus, by extracting data sets of peptides which trigger the secretion of inflammatory cytokines, it is possible to develop predictors of peptides with pro- or anti-inflammatory properties [101, 183].

Finally, it should be mentioned that these sequence-based methods can be virtually applied to investigate any peptide property, provided that high-quality training data sets can be built with validated peptides having and not having the desired property [298]. They have been widely used, for example, to predict peptide capability to cross membranes and penetrate into cells [273, 296, 211, 13]. Another peptide property which can be investigated by sequence-based approaches is their capability to bind proteins. The studies which used such methods to predict protein-peptide interactions are reviewed in the next subsection.

Prediction of protein-peptide interactions

Three different levels of detail about protein-peptide interactions can be obtained using sequence-based approaches [39]. The first one is to identify pro-

teins that bind a query sequence. In that case, it is first necessary to build training data sets of peptides which bind proteins and other ones which do not. Then, machine learning algorithms are trained to discriminate binders from non-binders by using relevant peptide chemical descriptors. It should be noted that many variants of this approach were originally developed for large scale predictions of protein-protein interactions in the perspective of deciphering interactomes of various organisms [103, 276, 234]. They were also applied to predict interactions between human and bacteria or virus proteins by using relevant data sets of host-pathogen protein-protein interactions [77, 169, 144].

Beside these genome scale predictions, several sequence-based studies have been conducted in order to specifically identify peptide segments that mediate protein-protein associations. These peptides can be classified into two types, the short linear motifs (SLiMs) which have generally less than ten residues, a few very conserved ones, and no particular secondary structures, and the molecular recognition features (MoRFs) which have a longer sequence and generally undergo a disorder-to-order transition upon binding. Accordingly, several sequence-based algorithms were specifically developed for mining SLiMs [270, 165, 114] and other ones for detecting MoRFs [70, 182, 104] in various data sets of protein-protein complexes. It is worthy to note that SLiMs generally bind to specific recognition modules such as SH3 or PDZ domains. Therefore, many specialized peptide motif predictors were trained and developed on specific data sets of these prevalent domains [270, 38, 165, 229, 242].

The second level of information that can be investigated with sequence-based approaches is the identification of amino acid residues at protein-peptide interfaces. Just like homology modeling which aims at predicting protein tertiary structures from sequences and data sets of experimentally resolved structures, comparative studies can also be used for determining protein-protein interfaces by searching homologs of query sequences in data sets of known protein-protein complexes [308, 89]. Nonetheless, instead of searching homologous proteins by using sequence alignment tools, such as BLAST, most of the sequence-based predictions of interface residues employ numerical vectors encoding key chemical features of protein sequences and various machine learning algorithms. The latter are generally trained on data sets of interface residues and non-interface ones which were built from known protein-peptide complexes. Commonly, interface residues are defined as those with a solvent accessible surface area (SASA) which decreases by more than 1 $Å^2$ upon binding or as those which are distant from a protein partner by less than a threshold parameter.

It is worthy to note that most sequence-based machine learning approaches were applied to predict residues of a protein which are likely to be involved

in binding any other protein partners [69, 126, 112]. Nonetheless, several other studies tackled the problem of predicting the interface residues on both proteins of specific complexes [5, 193, 240], providing precious detailed information on protein-protein binding modes, especially for transient complexes and those mediated by SLiMs and MoRFs. Regarding specifically protein-peptide interfaces, very few studies based on sequences only were reported in the literature. We found only two sequence-based predictors of protein-peptide binding sites, namely SPRINT [269] and SVMpep [327], both using support vector machines (SVM) for classification of binding and non-binding residues. It can be noted that, among the input physical chemical features of amino acid residues, SVMpep includes intrinsic disorder information predicted by the IUPred webserver [72], which seems to improve the prediction accuracy.

Finally, the third level of information that can be inferred from sequence-based approaches is the protein-peptide binding affinities. Although machine learning classifiers were developed to discriminate protein-protein interactions with low or high affinity [317, 265], quantitative predictions of binding free energies (ΔG) from sequences are generally performed with machine learning regression methods. Using training data sets of experimental binding free energies of known protein-protein complexes and various sequence descriptors, obtained correlations between predicted and experimental affinities are very diverse, ranging from 0.3 to 0.8 with an average value around 0.6, depending on the selected sequence features and external data sets used for testing [252, 196, 177, 265, 133]. However, these approaches seem to perform better in predicting changes in binding free energies ($\Delta\Delta G$) upon mutations on one of the two partner sequences [38, 252].

Indeed, thanks to data sets of experimental $\Delta\Delta G$ of mutations at the interface of protein-protein complexes [124, 121], different machine learning regression methods yielded correlations between predicted and experimental changes in binding free energies in the narrower range of 0.7-0.9 on various tested data sets [38, 252, 92, 233, 321]. It should be noted that most of these studies trained their machine leaning algorithms with descriptors extracted from three-dimensional structures of protein-protein complexes. Only two purely sequence-based predictors were so far reported in the literature [123, 168]. Moreover, it is important to mention that when predictors are blind tested on completely independent data sets of protein-protein complexes, then correlations between predicted and experimental $\Delta\Delta G$ upon mutations significantly drop to a range of 0.3-0.6 [92, 233, 123, 168, 321], indicating that there is still room for improving these predictors. Lastly, although these sequence-based approaches were mainly developed for protein-protein complexes, some of them have been applied to protein-peptide interactions, including PDZ-peptide associations [38,

252, 133] or complexes of MDM2 with p53 MoRF [92, 233]. These studies pave the way for the investigation and design of peptide sequences with optimal binding free energies for target proteins.

2.2.2 Conformation-based approaches

It is now recognized that protein sequence determines their three-dimensional conformational ensemble which, in turn, confers their biological activity. Therefore, many developments of peptide derivatives have been based on or oriented toward the structural properties of identified bioactive peptides. We highlight here two main *in silico* conformation-based approaches to discover or design new peptide derivatives, the peptide pharmacophore screening and the stabilization of secondary structure mimics (Fig. 2.2).

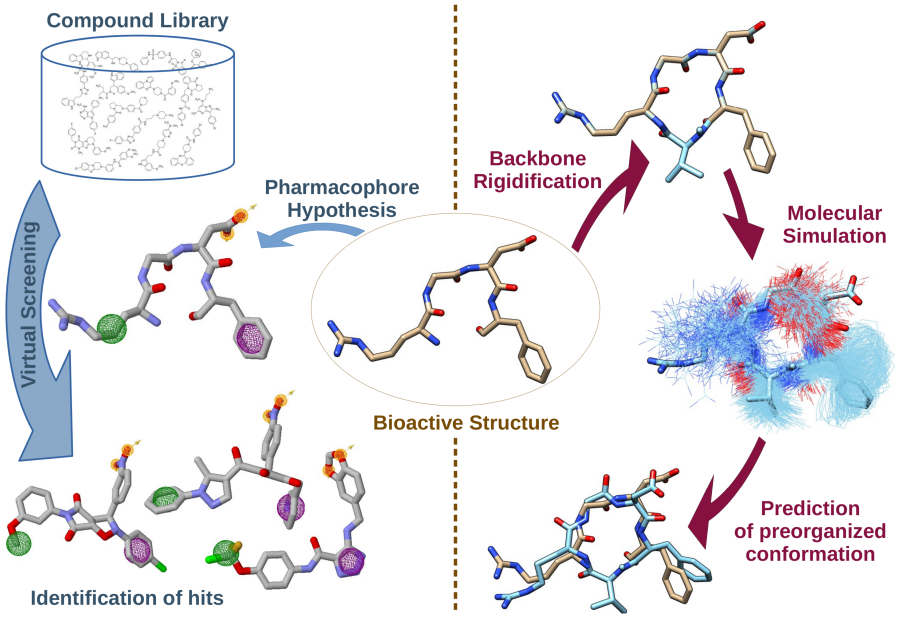

FIGURE 2.2 – Schematic description of a peptide pharmacophore screening method (left) and molecular simulation use for predicting preorganized conformations of a constrained peptide (right).

Peptide pharmacophore-based screening

Among the ligand-based approaches in drug discovery, the pharmacophore virtual screening is an efficient and popular computational tool which can harness the knowledge of peptide conformations. Indeed, when a peptide segment

is known to bind a target, then the residue side chains that are important for binding (hot spots) allow naturally to generate 3D-pharmacophore models. These, in turn, are used to screen compound libraries and identify new binders with a similar three-dimensional pharmacophoric arrangement. It should be mentioned that, although such drug developments are centered around a known peptide ligand, they often require the knowledge of its three-dimensional structure when bound to its target, conferring to these approaches a non purely ligand-based nature. This method was employed to discover inhibitors of several protein-protein complexes [187, 213, 81], notably involved in host-pathogen interactions [30, 102, 218]. Nonetheless, it should be noted that, after defining the peptide-based pharmacophore models, these studies often screened libraries of commercially available small compounds, which generally leads to hits being far from a peptide.

Constrained secondary structure mimics

Since many protein-protein interactions are mediated by peptide segments which are structured into α-helix, β-strand, or turns, a promising drug design strategy is to stabilize or constrain the peptide unbound state in these common secondary structures to minimize the entropy cost of binding and improve the affinity [125]. This can be achieved by using two main approaches, either by peptide cyclisation or by backbone stiffening. The first approach includes the α-helix stapling which consists in linking the side chains of two residues located on the same side of an α-helix, with hydrocarbon, lactam, or triazole staples for example [140, 98], and the β-sheet closure which consists in linking the two proximate residues at the extremities of a pair of β-strands, using hairpin loops or β-turn mimics [135, 158]. On the other hand, the backbone stiffening approach generally consists in inserting a chemical modification into the peptide backbone, such as disubstitution of the α-carbon [272] or substitution of the amide nitrogen [36, 243], in order to restrain its accessible conformational space.

In both previous strategies, a particularly helpful computational tool which can assist the design of these constrained peptides is molecular dynamics (MD) simulation. This technique numerically solves the Newton's equations of motion for a system of particles whose interactions are described by empirical potential functions usually referred to as force fields. When their timescales are sufficiently long, MD simulations can efficiently sample the peptide conformational ensembles and correctly predict their propensity to form secondary structures [188, 226, 34]. It could be noted that enhanced sampling techniques, such as replica exchange molecular dynamics [266] or metadynamics simulations [153] can also be used to generate more exhaustive conformational en-

sembles, especially for constrained cyclic peptides. Hence, more and more peptide derivative developments include MD studies to anticipate the impact of chemical modifications upon stabilization of secondary structures, as shortly presented below.

Regarding stapled helices, molecular simulations generally confirmed that they have more restricted conformational space than their non-stapled counterparts, but they still keep a high degree of conformational flexibility [131, 59, 53]. Importantly, these studies demonstrated that staples do not necessarily increase the helical propensity (or helicity) of stapled peptides, which seems to result from a fine balance between peptide sequences and position, length, and chemical nature of the staple [154, 53, 331]. Also, MD studies of stapled helices in free and bound states emphasized the point that high helicity of stapled peptides does not necessarily correlate with high binding affinity [131, 154, 53]. This could be due to the peptides' need for sufficient flexibility to adjust their structure in the partner binding site and/or to the fact that staples participate in and therefore modulate their binding [271].

Enhanced molecular simulations of cyclic peptides mimicking turns or β-structures also showed that cyclisation certainly reduced the heterogeneity of their conformations, but it still allows a significant amount of flexibility [264, 313, 314, 189]. Notably, cyclic backbones can still sample multiple conformational states, from compact to elongated structures, and a major question raised in these studies is whether, among them, there is a pre-organized one close to a bioactive conformation [224, 228, 313, 286, 79]. If such a bioactive conformation can be identified within the peptide conformational ensemble, then additional chemical modifications of the peptide backbone, such as α-carbon disubstitution or N-methylation, can be introduced to shift the conformational equilibrium in favor of it. Here again, enhanced MD simulations can help to rationalize and optimize the impact of these modifications on peptide derivative conformational space [286, 209, 259, 243]. All together, these peptide conformation-based studies can guide chemists away from less interesting modulators in order to limit costs of long synthesis campaigns.

2.3 Target-based strategies

Target-based approaches for designing peptides modulating protein-protein interactions require to gather structural information about the studied complexes. In many cases, these data are difficult to obtain experimentally due to technical limitations but also due to the low affinity and/or the transient character of many protein-protein associations [164]. In that context, several computational tools have been developed these last decades to gain a better in-

sight into the structural determinants of these interactions. In this section, we will first describe different *in silico* approaches to investigate the interface of protein-peptide complexes by collecting data about cavities and hot spots or by performing protein-peptide docking. Next, we will see how these tools and information can help the rational design of regulatory peptides by finding minimal recognition motifs or by peptide library virtual screening. We will also discuss the optimisation methods to enhance affinity and specificity of these compounds.

2.3.1 Structural characterisation of protein-peptide interfaces

When a protein-peptide binding mode is unknown but the tertiary structure of the unbound protein is resolved, it is possible to anticipate the ligand binding sites on the protein by predicting its cavities and/or the few amino acids that predominantly contribute to the binding free energy (hot spots). It is also possible to model the complex three-dimension structures with protein-peptide docking techniques (Fig. 2.3).

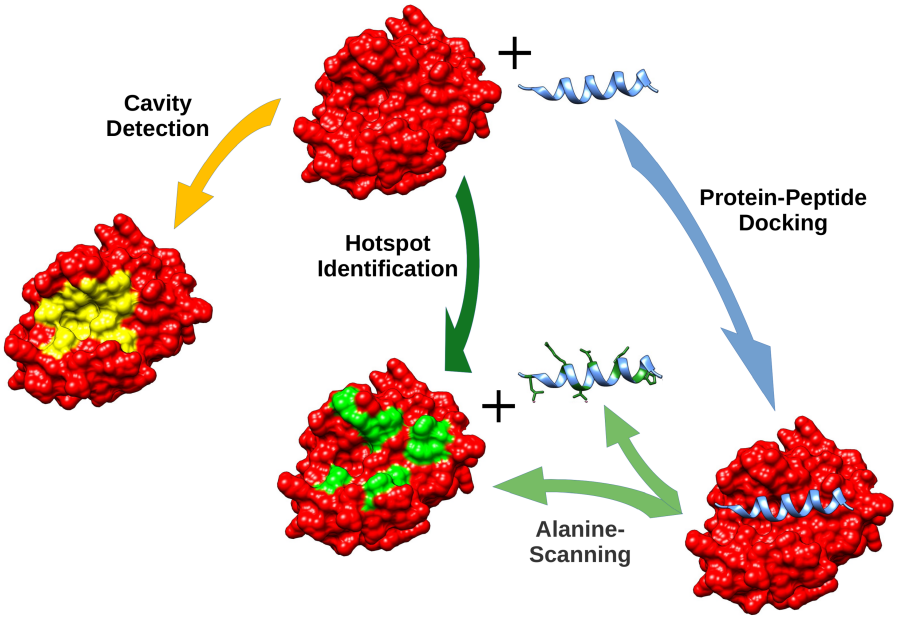

FIGURE 2.3 – Computational approaches that can provide structural information about a protein-peptide interface. Computational alanine-scanning is one method to identify hot spots from the three-dimensional structure of protein-peptide complexes.

Cavity detection

In classical structure-based drug design, an exploration of druggable cavities on a protein surface is generally performed prior to chemical library virtual screening or fragment-based design approaches [96]. This can be done with the web-servers CASTp [22] or FPOCKET [160], for example. However, as far as we know, these algorithms were mostly applied to detect protein cavities for small ligands but not for peptide binding sites which are wider and more difficult to identify.

We found in the literature only one study which investigates the binding pocket on a protein involved in protein-peptide interactions [100]. Using accelerated molecular dynamics simulations and a pocket identification method called VISM-CFA [99], the authors characterized the dynamic behavior of the Bad peptide binding site on Bcl-xL protein. They showed that the binding pocket of the unbound protein is often in a non-druggable closed state with a volume below 100 $Å^3$. Nevertheless, they also could identify minor conformations of apo Bcl-xL ($\sim$10%) with a more open binding pocket which could accommodate Bad peptide or small ligands [100]. This study reminds us that detection of druggable pockets on an unbound protein should preferentially be performed on its conformational ensemble rather than on a single structure.

Hot spot identification

As mentioned in the sequence-based section, hot spots of protein-protein complexes can be determined with machine-learning algorithms trained on data sets of interface and non-interface residues. The sequence descriptors used in those cases are generally intrinsic properties (polarity, hydrophilicity, hydrophobicity...) of protein amino acids. However, the accuracy of these predictors can be greatly improved by including structural properties such as residue solvent accessible surface areas or inter-residue distances in known tertiary and quaternary protein structures [172]. Thus, several high-performance hot spots predictors using protein three-dimensional structures have been developed and successfully applied to many protein-protein complexes [279, 305, 291, 67, 222].

Alternative physics-based or energy-based methods were also developed to identify hot spots on protein surfaces. Most of them use fragment-based approaches which aim at determining the preferential binding sites of small organic probes on known structures of proteins. This can be achieved by running docking calculations of small compounds into target cavities with classical protein-ligand docking programs, such as Gold [129] or Autodock Vina [278], as demonstrated in the study by Wang *et al.* of human activin receptor hot spots [292]. Another possibility to explore fragment binding sites on proteins is to use mo-

lecular simulations, such as the grand canonical ensemble Monte Carlo simulations employed by Kulp III *et al.* to identify hot spots on various proteins, including lysozyme, RecA, HIV protease, dihydrofolate reductase, elastase, MDM2, and peptide deformylase [147, 146]. Importantly, these studies indicate that hot spots are more correctly predicted by locating high affinity binding sites for organic fragments which are also low affinity binding sites for water molecules.

Experimentally, protein hot spots can be identified by using the alanine-scanning mutagenesis method [56]. In the same spirit, they can be predicted by using the computational alanine-scanning (CAS). From a known quaternary structure of a protein-peptide complex, the technique consists in estimating the binding free energy change ($\Delta\Delta G$) upon mutation of residues at the interface into alanine. Mutations that significantly impair the protein-peptide binding energy identify the hot spots. This general scheme was implemented into several molecular modeling software packages, such as Rosetta (Flex_ddG) [16] or BUDE (BudeAlaScan) [117]. The main difference between these programs lies in the methods used to compute binding free energies which can be fast empirical energy functions, MM/PBSA calculations, or thermodynamic integrations [185]. Thanks to its rapidity and low-cost, computational alanine-scanning was applied to identify hot spots of many protein-protein interactions [312, 60, 106, 157, 326], including the recent SARS-CoV-2 spike glycoprotein binding to host ACE2 receptors.

Protein-peptide docking

The knowledge of the three-dimensional structure of a targeted protein-protein complex is an invaluable information for structure-based design of PPI inhibitors. When only the tertiary structures of two unbound partners are known, protein-protein or protein-peptide docking are the main computational tools to generate structural models of their binding mode. The first protein-protein docking programs commonly consider proteins as rigid bodies. They generally consist in two or three steps : First, the shape complementary between the two protein structures is optimized [40, 17]. Then the obtained quaternary structures are re-scored by taking into account physical criteria such as electrostatic, van der Waals interactions, or desolvation energies [45, 63]. Frequently, these two steps are performed simultaneously. Generally, they are followed by a third step consisting in molecular dynamics simulations to allow local relaxation of the protein-protein interface.

Naturally, rigid docking methods are not appropriate for highly flexible proteins, especially for those which bind their partner through peptide segments such as SLiMs or MorFs. In these cases, protein-peptide docking programs should be preferred since they take into account the peptide flexibility at an early stage.

As for hot spot predictions, one can distinguish knowledge-based from physics-based protein-peptide docking. In knowledge-based approaches, also called template-based docking, the protein structure and peptide sequences are first used to search for homologous protein-peptide complexes in databases of experimentally resolved quaternary structures. Then, similarly to homology modeling, protein structure alignment and peptide sequence alignment are used to generate models of the protein-peptide binding mode. In this type of docking, the peptide backbone flexibility is taken into account by the different homologous peptide structures found in the database. Most often, model building is followed by an energy-based optimisation to allow further structural flexibility, such as in GalaxyPepDock [162], HDOCK [310], or InterPep2 [127].

The physics-based methods for flexible peptide docking can be subdivided into three different approaches : ensemble docking, *ab initio* docking, and fragment-based docking. In ensemble docking, the unbound peptide conformations are pre-sampled and the representative structures are rigidly docked into the protein. PepATTRACT [248], MdockPep [309], or PIPER-FlexPepDock [7], or HPEPDOCK [330] can be classified as ensemble docking methods. In *ab initio* approaches, the peptide conformations are sampled on-the-fly during the docking process, using mainly molecular simulations as in FlexPepDock [227], AnchorDock [18], or CABS-dock[149]. In fragment-based methods, the peptide is cut into shorter compounds and the fragments are docked onto protein. Then, the best modes of binding of each fragment are linked to generate the binding mode of the initial peptide. DINC [10] and IDP-LZerD [216] belong to this type of protein-peptide docking.

2.3.2 Identification and optimisation of peptide hits

In drug design, hit identification is the process consisting in finding compounds which bind a target and modify its activity. In this subsection, we describe computational methods to identify peptide hits modulating protein-protein interactions. Peptide hits can be derived mainly from minimal recognition motifs at structurally known protein-protein interface or with (structure-based) virtual screening of peptide libraries (Fig. 2.4).

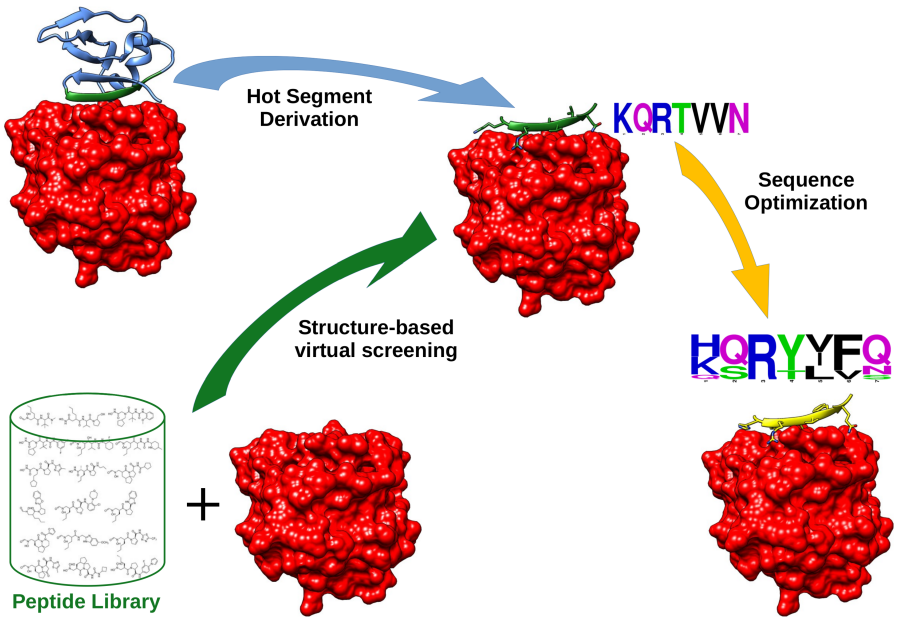

FIGURE 2.4 – Computational approaches used to identify a peptide hit and to optimize its sequence for higher affinity and selectivity.

Derivation of minimal recognition motifs

At many protein-protein interfaces, even those involving globular proteins, a short peptide segment predominantly contributes to the binding energy and is required to stabilize the complex [173]. Finding this hot segment, also called minimal recognition motif or self-inhibitory peptide, is often a good starting point for developing potent protein-protein inhibitors [174]. When the quaternary structure of a complex is known, several computational tools can assist the researchers to derive these minimal recognition motifs.

A first approach consists in identifying the hot spots of a complex and then in extracting the shortest peptide segment which contains as many hot spots as possible [205, 204]. Generally, the binding energies of these hot segments with their targets are subsequently estimated by using docking calculations or molecular simulations and compared to the initial protein-protein interactions to support the minimal recognition motif design [173, 205, 204, 119]. Another example of this approach was reported in two different studies of the same target, the Hsp90 dimer. The identification by computational alanine-scanning of four hot spots on the protein C-terminal α-helix served as a basis for designing several peptide-based inhibitors of Hsp90 [90, 26].

In the previous approach, the first and last residues of the minimal recog-

nition motif still have to be chosen by the researchers, and validation of these choices by computing binding energies can be quite tedious. Thus, a systematic method called Rosetta Peptiderive has been developed to automatically identify hot segments from the three-dimensional structure of a given protein-protein complex [251]. In this algorithm, a sliding window of user-defined size runs along one protein sequence and, at each position, isolates a peptide segment whose binding energy with the protein partner is computed using the Rosetta energy function [173]. Peptides which contribute the most to the protein-protein interaction are selected as hot segments. Peptiderive was made available to the scientific community through a web server [251] and allowed several groups to rapidly design from identified self-inhibitory peptides several inhibitors of various protein-protein interactions [110, 167, 274, 119]. It should be noted that, if the hot segments found have the appropriate geometry, then Peptiderive can automatically derive cyclic peptides by mutating their terminal residues into cysteine and linking them by a disulfide-bond [251].

Structure-based virtual screening

When the tertiary structure of a protein is known, one major strategy for drug discovery consists in docking millions of compounds from various chemical libraries into identified target cavities. Structure-based virtual screening has been applied to search for inhibitors of various protein-protein interactions, but mainly within libraries of small organic molecules [93, 128, 141]. Probably because docking peptides requires more computational resources than for small compounds, few papers reported the discovery of protein-protein inhibitors by using structure-based screening of peptide libraries.

Nevertheless, with the continuous increase of computing power, recent studies using peptide screening have been reported in the literature. In these studies, libraries of natural peptides extracted from food were docked into angiotensin-conversion enzymes [303, 315] or xanthine oxidase [316] to identify potent peptide-based inhibitors of these proteins. Nonetheless, it should be mention that the used libraries were mainly composed of very short tri- or tetrapeptides, limiting the possibility to discover peptides long enough to competitively inhibit large protein-protein interfaces.

In this respect, it is worthy to mention that several computational tools can boost virtual screening of peptides by facilitating the generation of libraries of various peptides. The Robetta server, for example, can be used to easily generate libraries of helical, loop, or extended peptides [137]. Another example is the program CycloPs which can simply generate large and diverse libraries of cyclic peptides from natural and commercially available non-natural amino acids [75]. However, as far as we know, no structure-based virtual screening of CycloPs li-

braries has been reported in the literature so far. This could be due again to the computationally demanding calculations required for reliably docking several thousands of peptides with more than five residues.

Improving peptide affinity and selectivity by sequence optimisation

After having found a peptide hit, it is generally worthwhile to increase its affinity for its target to improve its inhibition potency. Moreover, selectivity of therapeutic compounds is an important requirement in drug development to lower the risk of off-targeting. In the case of peptide-based inhibitors, computational tools can help to improve the affinity and selectivity of identified peptide hits by optimising their sequence. The guiding principle of this hit-to-lead process is similar to that used in protein redesign to improve their stability [113], since the physical forces that drive protein folding also drive protein-protein and protein-peptide binding.

In favorable cases where the protein-peptide quaternary structure is known, redesign techniques generally consist in exploring the sequence space of the fixed-backbone peptide and finding those which minimize an energy score. This can be the binding free energy variation ($\Delta\Delta G$) relative to the initial peptide sequence for affinity improvement, or the difference between binding free energies of the same sequence but for two different protein partners for selectivity enhancement. In essence, these approaches are similar to the computational alanine-scanning technique, except that each residue of the redesigned peptide can be mutated into all possible amino acids.

Rosetta [143] is probably the most used software to design or redesign proteins, peptides and their associations, but several other programs can be used to perform these tasks, including K* [232], ORBIT [253], Proteus [258], or dTERMen [86]. These programs exploit different algorithms to explore the protein and peptide sequence space, such as minimization methods, genetic algorithm, or Monte Carlo sampling. They also differ in their energy functions which combine to varying degrees ingredients of physics-based all-atom force fields, implicit solvation models, and knowledge-based potentials derived from protein complex structures [221, 24]. Interestingly, dTERMen uses a scoring function derived from statistical potentials between tertiary structural motifs (TERMs) frequently observed in protein three-dimensional structures [180]. Since these TERMs have characteristic sequence preferences [329], the structure-based interactions are converted into sequence-based scoring functions which are extremely fast to evaluate, allowing to exhaustively explore sequence spaces of long peptides and proteins [86].

Many applications of computational protein-peptide interface redesign have been reported in the literature and subsequent experimental validations of their

predictions highlight the reliability of these approaches to improve the affinity and selectivity of peptides for their target proteins. Among the success stories, highly selective peptides were computationally designed against bZIP proteins [97, 42], PDZ domains [260, 232, 328], amyloid fibrils [257], the cytokine TNFα [320, 311], and several anti-apoptotic proteins of the Bcl-2 family [84, 20, 86]. Interestingly, two studies among the previously cited redesigned peptide inhibitors with D-amino acids [257, 311], paving the way for the development of peptide-based drugs with high affinity, selectivity, and metabolic stability.

2.4 Conclusions

In this review, we classified the computational tools and strategies for designing peptide-based inhibitors of PPIs into the two conventional ligand-based and structure-based categories. Nevertheless, the border between the two classes becomes more and more porous and several peptide developments combined both approaches. For example, sequence-based predictions of hot spots at protein-peptide interfaces by machine learning algorithms are more accurate when molecular descriptors include structural information, such as solvent accessible surface areas or inter-residue distances. Hybrid approaches will probably become more frequent in the near future.

In both categories, the main challenge remains to determine the optimal peptide sequences which bind a protein target with the best affinity and selectivity. This requires to be able to compute as accurately as possible binding free energies and their relation to sequences, structures, and dynamics. Notably, regarding peptides which have generally more degrees of freedom than small organic compounds, this objective calls for correctly characterising their conformational ensemble to quantitatively estimate the entropy cost of association, especially for peptides which undergo a disorder-to-order transition upon binding.

Lastly, in the perspective of drug design, it remains crucial to reduce the peptidic nature of the identified peptide hits for increasing their stability against proteolytic enzymes (without decreasing their affinity and selectivity). This can be achieved by introducing non-natural amino acids, such as D-amino acids, peptoids, or chemically modified side chains, in the early stages of the development of PPI peptide-based inhibitors. Their membrane permeability is also an important property which is worth investigating as early as possible in order to maximize the chances of success in clinical trials.

Chapitre 3

Des3PI : Une approche computationnelle *fragment-based* de conception de peptides cycliques modulateurs d'interacions protéine-protéine

Par le biais du chapitre précédent, nous avons exploré les diverses méthodes computationnelles pour la conception de peptides inhibiteurs d'IPP. Nous avions souligné l'existence de deux grands types de méthodes : les approches *peptide-based*, basées principalement sur des méthodes par apprentissage, la structure 3D des cibles n'étant pas résolue, ou les méthodes *target-based*, utilisant la structure tridimensionnelle des cibles pour concevoir des inhibiteurs peptidiques.

Parmi toutes ces diverses méthodes, la conception *fragment-based* de peptides n'est que très peu représentée dans le *design* de peptides thérapeutiques ciblant les IPP. Pourtant, nous avons décrit précédemment que la plupart des interfaces protéine-protéine sont constituées de résidus réalisant l'essentiel de l'énergie libre de liaison, les *hotspots*. Or, une approche par fragments permettrait de cibler de façon précise et optimale ces *hotspots* pour créer un peptide hautement spécifique à sa cible.

Ainsi, l'objectif de cette thèse était de développer une approche *fragment-based* basée sur le *docking* d'une librairie d'acide aminés : Des3PI (***Des****ign* ***P****eptides targeting* ***P****rotein-****P****rotein* ***I****nteractions*). Nous pensons en effet que développer une telle méthode pourrait permettre de proposer de façon rapide plusieurs séquences ayant potentiellement une bonne énergie de liaison sur une protéine cible donnée. Cela permettrait ainsi de réduire le nombre de séquences

peptidiques à tester pour un expérimentateur, faisant ainsi bénéficier d'un gain de temps mais également une économie de moyens.

Nous décrirons ici de façon détaillée la méthode de Des3PI à travers trois exemples d'applications qui nous ont permis de calibrer l'approche. Les systèmes étudiés ont été choisis pour la diversité de leur type d'interface. Le système Mcl-1/PUMA implique majoritairement une hélice α, l'interface Ras/Raf est formée à partir d'une hélice α et d'un feuillet β et l'oligomère d'Aβ est le résultat de l'interaction de multiples feuillets β entre eux.

Ce chapitre est l'adaptation de l'article, *Des3PI : a fragment-based approach to design cyclic peptides targeting protein-protein interactions, Delaunay and Ha-Duong, Journal of Computer-Aided Molecular Design, 2022* [65].

3.1 Introduction

Many basic activities of cells, such as metabolic pathways or signal transduction, are carried out by series of associations and dissociations of biomolecules, especially proteins. Thus, deregulations of the network of protein-protein interactions (PPIs) often lead to cellular dysfunctions and to severe diseases, such as cancers or degenerative diseases [236]. Accordingly, targeting abnormal PPIs with modulator molecules offers an attractive opportunity to discover new therapeutic compounds [194]. It is noteworthy that, compared to strategies that target isolated enzymes or receptors, targeting PPIs reduces the probability of drug resistances, since a mutation in one protein should need a second mutation in its partner to maintain the PPI functional [12]. However, in contrast to substrate binding sites in enzymes or receptors, protein-protein interfaces are generally broad and flat, and often need large molecules to be disrupted [254, 197, 300]. For this reason, peptide derivatives are particularly well appropriate for efficiently modulating PPIs, given their intermediate size between small organic compounds and antibodies. Also, peptide derivatives have generally higher specificity for their target than small compounds, reducing the probability of undesirable side effects [201, 55]. Consequently, peptide derivatives have emerged as promising therapeutic avenues [156, 283], and, at the present time, a dozen of peptides targeting PPIs are currently in clinical trials [14, 179].

Nonetheless, the peptide approach remains quite under-exploited, mainly due to non-optimal pharmacokinetic properties inherent in peptides : they are easily degraded by the proteases, they have difficulties to pass physiological barriers, and they can induce undesirable immune responses [85, 283]. To overcome these limitations, it is recommended to reduce the peptidic nature of these molecules, for example by using non-natural amino acids or by cyclizing them. Particularly, in addition to a lower sensitivity to proteases and a higher

membrane permeability, cyclic peptides have a more constrained conformation than their linear counterparts, which reduces the entropy cost of binding and improves the affinity to their target. It could be noted that cyclic peptide conformations generally have preferential orientations of the amino acid side chains. To design peptides with side chain orientations optimal for binding a targeted protein, various computational techniques can be used to predict the conformations of cyclic peptides depending on their cyclization method (stapling, head-to-tail, disulfide, side chain to side chain...) and on the insertion of non-natural amino acids (N-methylated, α,α-disubstituted...) [189, 64].

That being said, the main challenge in the drug discovery process remains to find the sequence of the peptide derivatives that will bind a target with high affinity and high specificity. Among the experimental approaches, the display-based technologies [27, 202], especially the phage display screening [261, 319], are probably the most widely used for this purpose. Basically, the method consists in generating huge libraries of peptides displayed on phage capsids, in incubating these phages with proteins of interest attached to a surface, and in identifying those that can bind with high affinity the immobilized proteins. It is worthy to note that phage-displayed cyclic peptides can be obtained by forming a covalent bond between two cysteine residues or between a cysteine and an inserted non-canonical amino acid bearing an electrophilic reactive group [27, 294]. Despite the undeniable efficiency of phage display techniques to discover peptide binders of a protein, there is nonetheless no guarantee that the identified peptides bind the protein surface involved in the targeted PPI, especially for large proteins or those which interact with multiple partners. Furthermore, applying this technique to identify binders of amyloid protein aggregates seems to be tricky and no such application was reported in the literature, as far as we know. Thus, it still remains worthwhile to develop new methods for designing cyclic peptides targeting not only a protein but a specific protein surface.

When the three-dimensional structure of a targeted protein-protein complex is known, a traditional approach to identify a peptide hit is to isolate from the binding interface the short peptide segments that mostly contribute to the complex binding energy. Then, these so-called minimal recognition motifs, hot segments, or self-inhibitory elements are modified to optimize their affinity, specificity, and pharmacokinetic properties. In this regard, it is worthy to mention that several computational tools were developed to find the optimal linkers to perform the cyclization of minimal recognition motifs [251, 256, 241]. If pharmacophores of a minimal recognition motif can be identified, then virtual screening of various peptide libraries can be performed on these pharmacophores to discover peptide binders of the targeted protein [74]. To help in this task, several computational tools were developed to facilitate the generation of libra-

ries of diverse peptides. For example, the Robetta server can easily generate libraries of helical, loop, or extended peptides [137]. Of particular interest is the program CycloPs which can simply generate large and diverse libraries of cyclic and constrained peptides from natural and commercially available non-natural amino acids [75].

When the three-dimensional structure of only one partner of a protein-protein complex is known, and as long as the binding interface can be inferred, *de novo* design methods of cyclic peptides could be advantageous. Among the recent research in this direction, one can mention the stochastic evolutionary algorithm proposed by Soler *et al.* [263] or the anchor extension strategy developed by Hosseinzadeh *et al.* [111]. Alternative approaches that have emerged these last years to design PPI modulators are the fragment-based methods. They consist in screening a library of molecular fragments according to their affinity for a targeted protein and selecting those with the best binding energies. The selected fragments are then linked to form one modulator molecule. These approaches are however tricky to be implemented experimentally because small molecular fragments generally have low affinities for the target which are difficult to be measured by standard experiments. To overcome the experimental limits, *in silico* fragment-based approaches have been developed since computational methods, such as molecular docking, can quantify protein-ligand interactions even if the affinity is very low.

Nevertheless, when targeting large protein surfaces, the molecular fragments should not be too small to have an appreciable binding specificity. Thus, only two types of libraries were successfully used in fragment-based design of PPI modulators, those composed of FDA-approved compounds or those constituted of natural substances [194]. However, the molecules yielded by these fragment libraries are far from being peptides and the chemical linking of already elaborated fragments can be tricky. In this context, we developed a novel *in silico* fragment-based approach to design peptidic PPI inhibitors. This method, called Des3PI (Design of Peptides targeting Protein-Protein Interactions), performs docking calculations of a library of amino acids on a targeted protein surface and then links those with good binding energy in order to generate the sequence and structure of cyclic peptides which will likely bind the protein target with high affinity and specificity.

3.2 Methods

3.2.1 Building the fragment library

In this study, the fragment library is simply composed of the twenty natural proteinogenic amino acids. An initial three-dimensional structure for each of them was generated by using the 2D to 3D structure conversion program MarvinSketch 6.2.1 from ChemAxon [37]. Then, each amino acid structure was charged using the AM1-BCC model [120] and shortly minimized using 5 000 steepest descent steps and the Generalized AMBER Force Field (GAFF) [290]. It should be noted that all fragment amine and carboxyl groups were modeled in their neutral form, except the aspartate, glutamate, lysine, and arginine side chains which were considered in their ionic state.

3.2.2 Finding the fragment preferential binding positions

First, the targeted protein surface was delineated by centering and sizing an Autodock Vina search box [217] around the area of interest (Figure 3.1A). Then, each fragment was docked 50 times onto the defined protein surface using Autodock Vina [278]. For each docking, the 9 best scores were retained, yielding $9 \times 50 \times 20 = 9\,000$ binding modes of all the 20 amino acids on the protein surface.

Then, the positions of the α-carbons of all the binding modes were clustered using a hierarchical algorithm based on the method of centroids [284]. A criterion of 3.5 Å for the minimal distance between two centroids was chosen, this threshold being slightly lower than the mean distance between two successive α-carbons in proteins and peptides (3.8 Å) [214]. To avoid considering the sparsely populated clusters, those with less than 0.1% of the 9 000 binding modes were not further taken into account, yielding the most significantly populated clusters as the hotspots of the future modulator peptide (Figure 3.1B). By default, this cutoff parameter was fixed to 0.1%, but it can be adjusted in order to have a manageable number of hotspots. For example, in the case of Mcl-1 protein, cutoff values of 0.1%, 0.5%, and 1% lead to 9, 8, and 6 hotspots, respectively (Figure 3.3). Finally, the most frequently found amino acid in each cluster is selected to generate the sequence of the peptide that Des3PI considers as a good binder of the targeted protein surface (Figure 3.1C).

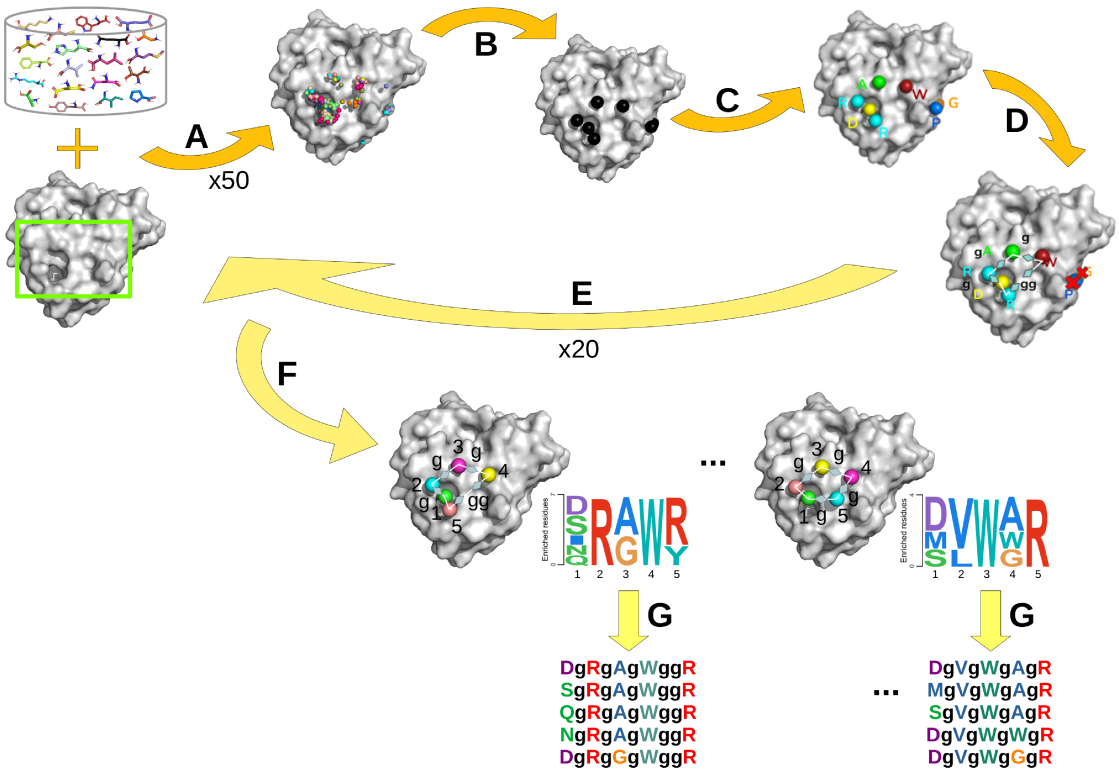

FIGURE 3.1 – Des3PI workflow : (A) The library of amino acids is docked 50 times onto the targeted protein surface delineated by a green rectangular box. (B) All the binding modes are clustered to determine the hotspot locations. (C) The most recurrent amino acids are identified for each hotspot. (D) The hotspots that are close from each other are linked with glycine residues. (E) Steps A, B, C, and D are repeated 20 times. (F) For each class of peptides, the amino acid occurrence in the generated sequences is calculated, and (G) the five most promising peptide sequences are output. Lowercase **g** indicates glycine linkers.

3.2.3 Linking the hotspots into a cyclic peptide

Once the hotspots and their relative positions were determined, it is possible to visually choose those that are close enough to form one peptide and manually link them to generate a cyclic sequence (Figure 3.1D). Alternatively, it is possible to use an algorithm that we developed to automatically perform these two tasks. This independent module that we integrated into Des3PI has been satisfactorily tested in several cases but may give some inaccurate or unexpected cyclic sequences when the hotspots do not have a clear cyclic geometry. This automated approach is presented below.

First, to identify the hotspots that are not too far from each other for forming

one cyclic peptide, the positions of all previously found hotspots were clustered using the same method as above but with a criterion of 12.5 Å, which allows to separate groups of hotspots distant by more than 3 times the mean distance between two successive α-carbons in proteins and peptides. Moreover, we considered that a group of hotspots can form a promising cyclic peptide if it is composed of at least 4 hotspots.

Then, for a given group of hotspots, we determined the cyclic peptide sequence as follows : (i) The average plane of the group of hotspots was determined by using a principal component analysis of their positions, and the projections of the hotspots on this plane were calculated. (ii) The most populated hotspot (which is arbitrarily defined as the first residue of the sequence) is chosen as the origin of a reference frame of the plane, whose the axes are the two first principal components previously found. (iii) In this reference frame, the other hotspots are ordered according to their polar angle, from the lowest to the largest one, in the interval [-180° ; 180°]. It should be noted that, in the case where two hotspots have close polar angle values (differing by less than 15°), the hotspot with the lowest radial distance is prioritized, except for the two last hotspots for which the highest radial distance is prioritized in order to close the peptide cycle.

Finally, having the sequence of hotspots in the designed peptide, Des3PI determines how many linkers are required to link two consecutive hotspots. In this study, we chose the glycine as a linker and we added between two consecutive hotspots a number of linkers equal to the integer part of their separating distance divided by 4.5 Å (Figure 3.1D). This parameter was determined by using a trial and error approach on Mcl-1 protein : a too small value (below 4.0 Å) yielded too many glycine residues between hotspots and too large peptides. Conversely, a too large value (above 5.0 Å) led to too few linkers and too compact peptides. Subsequently, both too large and too compact peptides could not be correctly docked onto the targeted protein surface close to the hotspot positions determined by Des3PI (see the validation subsection).

3.2.4 Generating the most promising peptide sequences

At this point, it should be noted that, when repeating steps A to D, different peptides can be obtained because of the stochastic search algorithm implemented in Autodock Vina [278]. To provide sequence diversity, the whole protocol described above was repeated 20 times to generate 20 peptides (Figure 3.1E). The latter were categorized into different classes according to the number of hotspots and their geometry (Figure 3.1F). We considered here that two peptides have a similar geometry when the RMSD between their hotspots is below 1.75 Å. This parameter was fixed by using a trial and error approach on Ras protein : we

tested different values from 1.0 to 2.0 Å and visually inspected whether similar hotspot geometries were effectively in the same class, and conversely, whether different geometries could be separated in different classes (Figure 3.2). Once the peptide classes were defined, we output for each of them the amino acid occurrence at each position of the peptide sequences, which can be visualized using the PSSMSearch server [145]. Finally, a score is attributed to each amino acid proportional to its occurrence and the peptide sequences with the highest sum of these scores are considered as the most promising cyclic peptides for binding the targeted protein surface (Figure 3.1G).

To complete the description of Des3PI, the numbers of runs (20) and of docking calculations (50) are shortly discussed hereinafter. Their impact on the sequences generated by Des3PI was assessed in the case of Mcl-1 protein (Figure 3.15). This benchmark shows that, for 10 runs, the amino acid occurrences in the generated sequences slightly differ when the number of docking varies from 25 to 75. For 20 runs, the occurrences seem to converge for a number of docking larger than 50, and, for 30 runs, they are similar for all tested numbers of docking. From these tests, we decided to fix the default numbers of runs and docking calculations to 20 and 50, respectively.

3.2.5 Validation using blind docking

To validate the method, the peptides proposed by Des3PI have to be synthesized and their affinity and/or PPI inhibition activity have to be experimentally quantified. However, these experimental validations can be difficult and long to implement. Thus, we propose here a two-step procedure to computationally support whether or not the generated peptides are likely to succeed.

The first step consists in a blind docking of the best cyclic peptides designed by Des3PI on the targeted protein, and to verify whether they preferentially bind the targeted surface. It should be stressed here that, unlike the previous dockings of the single amino acids which were restricted to protein surfaces involved in protein-protein interfaces, the blind dockings of the designed peptides were performed on the entire proteins without specifying any targeted surface. Among the protein-peptide docking programs that could deal with cyclic flexible peptides, we chose AutoDock CrankPep (ADCP) [324, 325] which just requires to input the protein PDB file and the peptide sequence string. It is noteworthy that ADCP could yield different results depending on the first and last residues of the cyclic sequence given as input. Thus, for each peptide composed of n residues, we performed n docking calculations with different inputs of the first and last residues of the cyclic sequence. Each docking run consisted in 50 independent searches of 2 500 000 Monte Carlo steps, and generated 100 best binding modes. Finally, the $100 \times n$ preferential binding modes of each peptide were analyzed by

computing the root-mean-square deviation (RMSD) of the α-carbons relative to the hotspots generated by Des3PI. Then we checked for each peptide whether one or several binding modes among the 5% best scores were found close to the targeted protein surface with a Cα RMSD relative to the Des3PI hotspots lower than 10 Å.

3.2.6 Checking complex stability using MD simulations

In a second step, we selected the protein-peptide complex with the lowest peptide RMSD (among the 5% best scores), and verify its stability by using molecular dynamics simulations performed with the GROMACS 2019.1 package [4]. The AMBER99SB-ILDN [275] and GAFF [290] force fields were used for the protein and peptide, respectively. Each complex was placed in a cubic box, so that the minimal distance between the solute and the cube faces was equal to 1 nm. Then the complex was solvated with TIP3P water molecules and neutralized with 0.15 mol/L of sodium chloride. The Lennard-Jones potentials were cut off at 1.2 nm and the Coulomb interactions were treated using the smooth PME method [78]. Each system was first minimized using 10 000 steps of the steepest descent method, then submitted to two short equilibration runs of 1 ns each, the first one to heat the system to 310 K using a Berendsen thermostat and the second one to equilibrate the pressure around 1 bar using the Parinello-Rahman method. After that, a 200 ns production run was performed in the isothermal–isobaric (NPT) ensemble using the Nose-Hoover and Parrinello-Rahman coupling algorithms [207, 109, 212] with the time constants $\tau_T = 0.5$ ps and $\tau_P = 2.5$ ps. The Newton's equations of motion were integrated using the leap-frog algorithm with a time step of 2 fs, while keeping constant the length of all covalent bonds using the LINCS procedure [107]. MD trajectory frames were saved every 20 ps for subsequent analysis. Notably, contact residues were computed using the GROMACS *gmx mindist* tool and a cutoff value of 0.5 nm.

3.3 Results and discussion

3.3.1 Peptides generated by Des3PI

Des3PI was first applied to identify cyclic peptides targeting three proteins which are involved in three different types of protein-protein interfaces : the protein Ras which binds Raf via an α-helix and a β-strand, Mcl-1 which interacts with the α-helical BH3 motif of PUMA, and a protofibril of Aβ which mainly involves β-strand/β-strand interactions.

The three-dimensional structure of Ras was taken from a crystallographic

structure of Ras-Raf complex (PDB ID : 3KUD [83]). The 20 runs of Des3PI on Ras generated peptides with either 4, 5, or 6 hotspots. The 20 peptides could be categorized in four classes whose amino acid occurrences are displayed in Figure 3.2. From the occurrences, we output the five best peptide sequences in each class. When comparing the positions and amino acid compositions of class IV hotspots with the Raf residues in contact with Ras, hotspots 2 and 6 which are mainly populated with Arg are located at the same positions as two Lys residues. Furthermore, hotspot 3 which is essentially a Val is retrieved at the same location as a Raf Val residue. The three other hotspots are not clearly related to the RAF residues observed around their positions. Overall, half the six hotspots are composed of amino acids with similar properties as the Raf residues at the same locations.

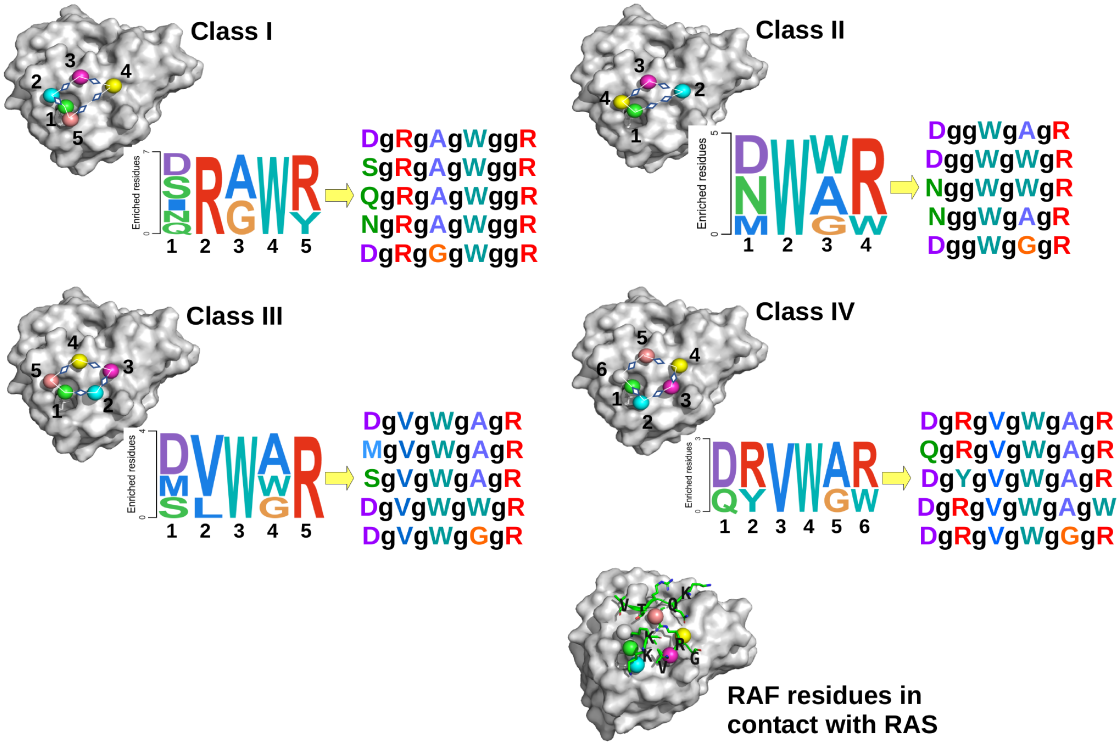

FIGURE 3.2 – Des3PI generated 4 classes of peptides potentially binding Ras protein. The five best peptide sequences of each class were generated according to the amino acid occurrences in the peptide hotspots.

In experimental efforts to discover inhibitors of the oncogenic K-Ras proteins, Wu *et al.* performed a screening of about 3 millions cyclic peptides against the K-Ras G12V mutant and identified 20 sequences that can bind K-Ras with submicromolar affinity and disrupt its interactions with Raf [304]. Interestingly,

these identified sequences are rich in aromatic residues and Arg, similarly to the sequences output by Des3PI. Notably, their most promising cyclic peptide (compound **12**) has the sequence dNle-Fpa-Arg-dNal-Arg-Arg, where dNle is a D-norleucine, Fpa a fluorophenylalanine, and dNal a D-2-naphthylalanine [304]. The similarity in composition of their sequence with those of Des3PI Class IV peptides suggests that our computational approach generated relevant peptide binders of Ras.

For Mcl-1, we applied Des3PI on a three-dimensional structure extracted from the first model of the NMR structure of Mcl-1 in complex with PUMA (PDB ID : 2ROC [62]). In one run, Des3PI only found 3 hotspots, and in the other 19, the algorithm generated 6 hotspots (representing at least 1% of the 9 000 binding modes of the 20 amino acids, instead of the default threshold of 0.1%). In the latter case, the hotspots have a very similar geometry and can be grouped into one class of peptides. However, although the sixth hotspot was close enough to the other five to form a peptide (according to the clustering criterion of 12.5 Å), it led to a non obvious cyclic geometry (Figure 3.3). Therefore, we decided to manually remove this hotspot and only keep the remaining five that were able to form a cyclic peptide. The amino acid occurrences in the 19 peptides and the derived five best peptide sequences are reported in Figure 3.3. It could be noted that these five hotspots only partially occupy the Mcl-1 binding groove which normally accommodate the PUMA α-helix. More specifically, referring to the pocket nomenclature of Mcl-1 binding groove by Denis *et al.* [68], the Des3PI hotspots are located in the hydrophobic pockets P2 and P3. It is therefore not surprising that the amino acids most frequently found at hotspots 2, 3, 4, and 5 are mainly hydrophobic ones. An Asn residue is always found at hotspot 1, close to the position of a Puma Arg residue (which makes an intramolecular salt bridge with an Asp). Overall, the amino acids frequently found at these five hotspots are consistent with those of Puma involved in binding Mcl-1.

FIGURE 3.3 – Des3PI generated one class of peptides potentially binding Mcl-1 protein. The five best peptide sequences were generated according to the amino acid occurrences in the peptide hotspots.

Among the known inhibitors of Mcl-1, many are small organic compounds with a central indolic, heterocyclic, or aromatic scaffold which occupies the P2 pocket, another hydrophobic group connected to the central scaffold which occupies the P3 pocket, and a carboxylic acid group which interacts with Mcl-1 Arg263 [68]. Alternatively, peptide inhibitors of Mcl-1 have been identified by screening BH3-based libraries [161, 76]. It was observed that these helical peptide binders had very similar sequences to natural BH3 helix ones, with four hydrophobic residues on one side of the α-helix which occupy the four Mcl-1 binding pockets, and one Asp residue between the third and fourth ones which makes a salt bridge with Mcl-1 Arg263. In comparison, Des3PI also found four hydrophobic hotspots but they are not aligned as those in BH3-like helices and only occupy two over the four Mcl-1 binding pockets (P2 and P3). Moreover, we did not retrieved an Asp residue close to Mcl-1 Arg263. Instead, Des3PI output and Asn at hotspot 1 which could easily make a hydrogen bond with it. Overall, the peptides designed by Des3PI have different topology from BH3-based α-helices, but might tightly occupy half of the Mcl-1 binding groove.

Regarding the design of peptides targeting Aβ protofibril, the results provided by Des3PI are more diverse than for Ras and Mcl-1, due to a larger area of the targeted surface. The latter is the surface perpendicular to the principal axis of the dimeric S-shaped protofibril of Aβ resolved by solid-state NMR (PDB ID : 5KK3 [52]). We extracted from the PDB structure the inner 2×5 Aβ molecules of the protofibril and applied Des3PI to the axial surface composed of the chains C and L. In several runs, Des3PI was able to find 2 or 3 groups of hotspots close enough to form cyclic peptides. Over the 20 runs, Des3PI identified three different areas that could be potential peptide binding sites (Figure 3.4). The first one was systematically retrieved in the 20 runs, and Des3PI provided here one class of cyclic peptides with 4 hotspots. The second area was identified 17 times over 20 runs, and, here also, Des3PI found only one class of cyclic peptides with 4 hotspots. The third area was retrieved 11 times over 20 runs, but our algorithm generated here 4 different classes of cyclic peptides, all of them having 4 hotspots except the last one which has 5 (Figure 3.4). All together, we could provide 12 different peptide sequences that potentially bind 3 different areas of the surface perpendicular to the Aβ protofibril axis.

The nature of the amino acids which most frequently occur in Des3PI hotspots are generally consistent with the Aβ residues at the protein-protein interface. In class I peptides, an Ile is mainly encountered at hotspot 2 which is located at the same position as an Aβ Ile residue. At hotspots 3 and 4, two Ser were found close to two Aβ Gly and one His. Lastly, an Arg is always found at hotspot 1 close to Aβ Glu and His residues. In class II peptides, hydrophobic amino acids are always found at the hotspots 1, 2, and 4 which are located in the area of

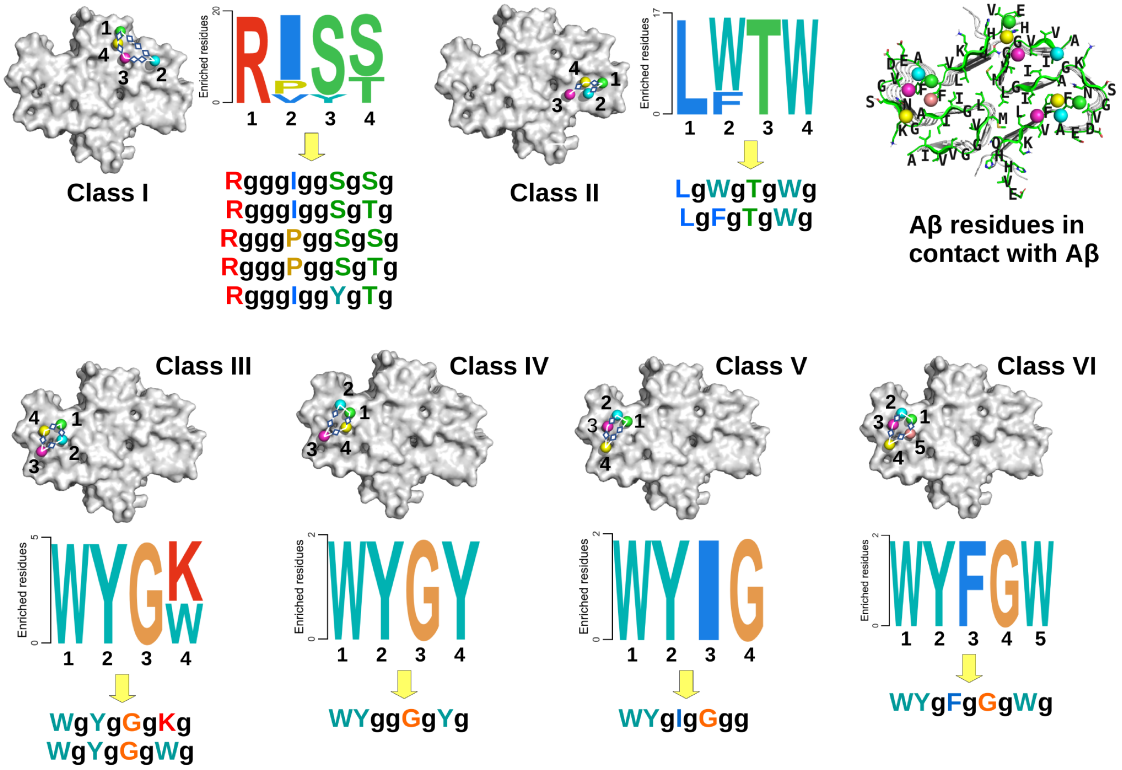

FIGURE 3.4 – Des3PI generated six classes of peptides potentially binding Aβ protofibril. For each class, at most five best peptide sequences were generated according to the amino acid occurrences in the peptide hotspots.

four Aβ hydrophobic residues (two Phe, one Ala, and one Val). The Thr found in hotspot 3 is situated at an Aβ Val residue position. Similarly, the class VI peptides have four hotspots composed of hydrophobic amino acids and positioned in the vicinity of two Phe, one Ala, and one Val. A Gly is always retrieved in the hotspot 4 which is situated at the place of an Aβ Lys residue (Figure 3.4). Overall, except for class I peptides which rather bind the Aβ C-terminal segment, Des3PI mainly generated hydrophobic peptides which target the ^{18}VFFA^{21} central region.

This observation is an encouraging outcome of our computational approach since the Aβ self-recognition element ^{16}KLVFFA^{21} is a major target of Aβ aggregation inhibitors. Naturally, many of them are peptides or peptidomimetics designed from this sequence [238], including cyclic peptides [44, 11, 178]. Nonetheless, high throughput screening approaches allows to identify peptide inhibitors with more diverse sequences than the self-recognition one. For instance, Richman *et al.* synthesized a library of head-to-tail cyclic D,L-α-hexapeptides and identified among them the two sequences lLwHsK and sHwHsK (where lower

and upper case letters denote D- and L-amino acids, respectively) which can inhibit Aβ aggregation [231]. In another study, Wang *et al.* performed an *in silico* screening of amyloidogenic hexapeptide databases to find those which are prone to dimerize into a β-sheet. Among 11 identified ones, 6 hexapeptides exhibited strong binding affinity to Aβ in SPR experiments, and among them, the two sequences CTRIYWG and GTVWWG could strongly inhibit Aβ aggregation in ThT fluorescence assays [293]. It could be noted that these two experimental studies revealed peptides with amino acids (Ile, Leu, Thr, Trp, and Tyr) similar to those which frequently appear in Des3PI sequences. This suggests that our approach could design potential good peptide binders and inhibitors of Aβ oligomers.

3.3.2 Validation of Des3PI peptides by blind docking

To validate the method, we set up a two-step computational procedure to check whether or not the generated peptides are likely to succeed. First, the selected peptides were blindly docked on the protein target by using the ADCP program [324, 325]. We considered that a peptide passes this test if at least one binding mode among the 5% lowest scores is retrieved close to the Des3PI hotspots with a RMSD of the peptide Cα atoms lower than 10 Å. Secondly, we selected the protein-peptide complex with the lowest peptide RMSD (among the 5% best scores), and verify its stability by molecular dynamics simulations.

For Ras protein, the graphs displaying the ADCP scores *versus* RMSD of the docked peptides (Figures 3.5 and 3.6) show that all 20 peptides generated by Des3PI have at least one low energy binding mode close to the targeted surface. It should be noted that the peptides with only 4 hotspots (class II) have overall moderate binding energies, as might be expected. However, although the peptides with 6 hotspots (class IV) have binding energies among the lowest ones, they did not outperform those of class I which have only 5. This indicates that the presence of specific amino acids in the peptides is more important for the binding than their size. For the second step of the validation procedure, we could have checked the stability of all the 20 peptides in complex with Ras, but, because of our limited computational resources, we chose to submit only one representative peptide of each class to the MD simulation step (QRAWR, NWAR, DVWGR, and DRVWAW).

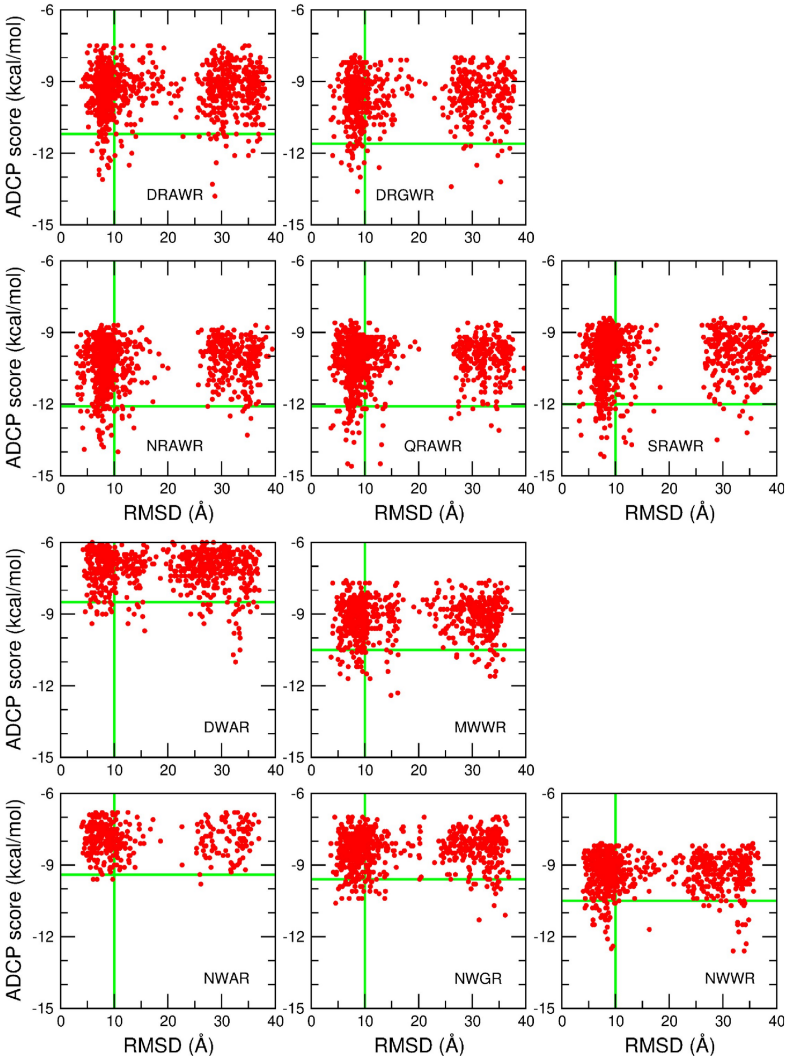

FIGURE 3.5 – ADCP score of the binding modes on Ras protein of the 20 best peptides (class I and II) generated by Des3PI as a function of their RMSD relative to the hotspot positions.

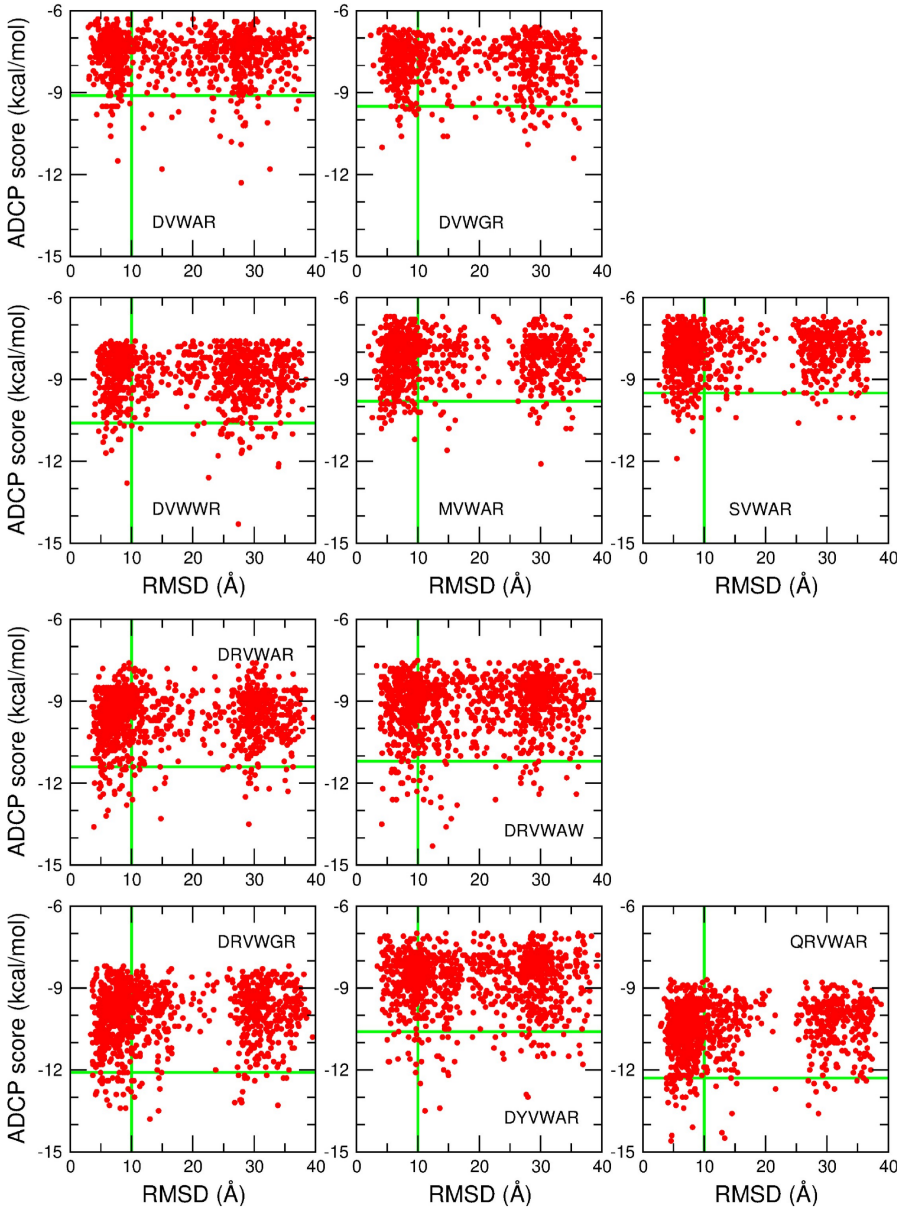

FIGURE 3.6 – ADCP score of the binding modes on Ras protein of the 20 best peptides (class III and IV) generated by Des3PI as a function of their RMSD relative to the hotspot positions.

In contrast, the blind docking of the peptides generated by Des3PI for Mcl-1 is more ambivalent than for Ras (Figure 5.5). Among the five selected peptides, only NFWIW clearly has low energy binding modes in the targeted surface of Mcl-1. Nevertheless, for each of the two peptides NFFKW and NWFIW, one binding mode was found at the boundary of the criteria for validating the test.

Compared to Ras protein, this mitigated success for Mcl-1 might be due to the shape of its binding interface with PUMA BH3 α-helix which is longer and narrower than the rather flat targeted surface of Ras. This might explain not only the lower number of peptide classes found by Des3PI for Mcl-1 (Figure 3.3) when compared to Ras (Figure 3.2) but also the medium success rate of the peptide blind docking on Mcl-1, given the fact that the designed peptides are cyclic, rather plane and not helical. Despite this, we continued the validation procedure of the 3 mentioned peptides (NFWIW, NFFKW, and NWFIW) by checking the stability of their complex with Mcl-1 using MD simulations.

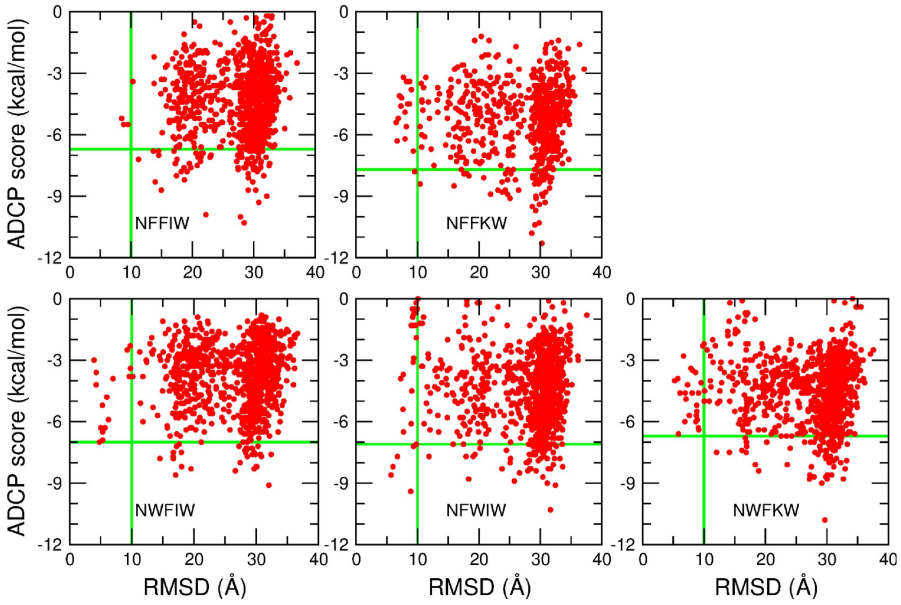

FIGURE 3.7 – ADCP score of the binding modes on Mcl-1 protein of the 5 best peptides generated by Des3PI as a function of their RMSD relative to the hotspot positions.

Regarding the 12 peptides designed by Des3PI for targeting Aβ, the results of their blind docking (Figure 3.8) show that (i) none of the 5 peptides of class I could be successfully docked close to the Des3PI hotspots. Nevertheless, the peptide RISS had one binding mode at the boundary of the test criteria and was further considered in the second validation step; (ii) each of the two sequences LFTW and LWTW of class II has one binding mode satisfying the criteria for validating the test; and (iii) 3 over the 5 last peptides generated by Des3PI (WYGK, WYGW, and WYIG) were successfully docked to their target. Given the very large area of the Aβ protofibril targeted surface, we estimate that the success rate of 6 peptides over 12 is rather encouraging. These 6 peptides were further evaluated by submitting their complexes with Aβ protofibril to MD simulations.

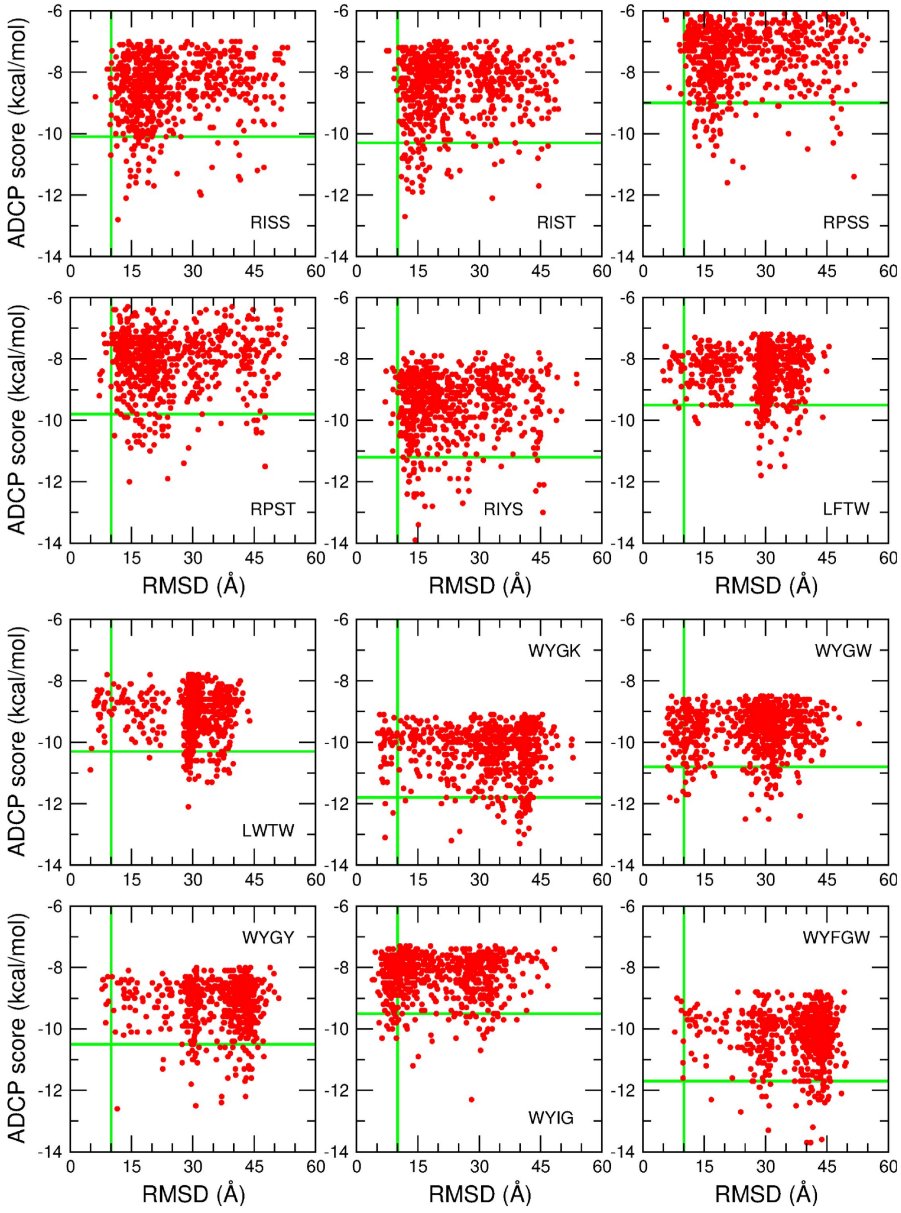

FIGURE 3.8 – ADCP score of the binding modes on Aβ protofibril of the 12 best peptides generated by Des3PI as a function of their RMSD relative to the hotspot positions.

All together, over the 37 peptides targeting Ras, Mcl-1, or Aβ, 29 of them could be successfully docked onto the protein targeted surfaces. This satisfactory success rate (78%) attests that Des3PI can generate cyclic peptide sequences with high probabilities to bind a targeted protein interface.

3.3.3 Checking protein-peptide complex stability by MD simulations

The best binding mode (i.e. that one with the lowest RMSD with respect to the Des3PI hotspots among the 5% best scores) of each of the 4 peptides NWAR, DVWGR, QRAWR, and DRVWAW on Ras protein was used as the starting conformation for two independent MD simulations. In all 8 simulations, the Ras protein RMSD relative to its initial conformation are stabilized between 1 and 2 Å (Figure 3.16). Overall, the 4 peptides remain on the targeted surface of Ras (Figure 3.9), except in one simulation of peptide NWAR (class II) and, in a lesser extent, in one simulation of DVWGR (class III), corroborating the moderate binding energies output from their docking calculations (Figures 3.5 and 3.6). The 2 peptides QRAWR (class I) and DRVWAW (class IV) have the most stable positions on the protein surface. Notably, the 6 hotspots DRVWAW peptide largely occupies the Ras surface that is involved in the binding to Raf and appears to be the most promising potent inhibitor of Ras-Raf interactions.

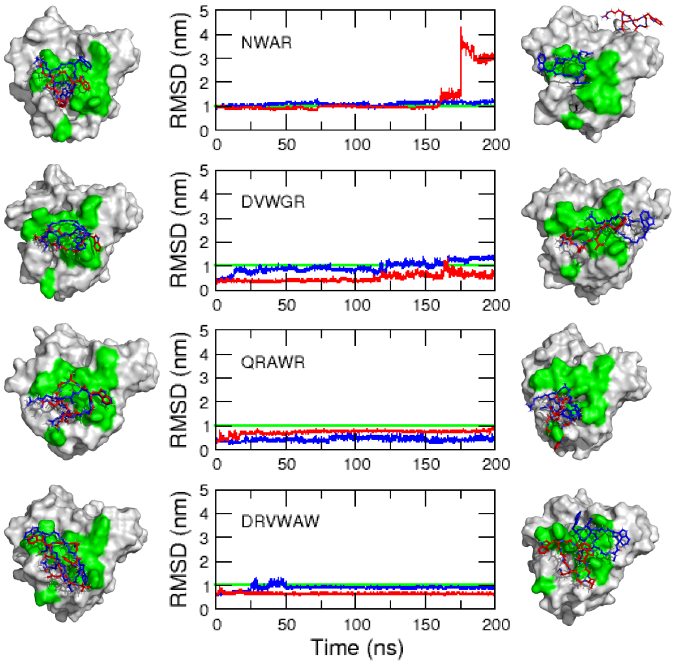

FIGURE 3.9 – Time evolution of the peptide Cα RMSD relative to the Des3PI hotspots in the Ras-peptide complex MD simulations. Snapshots in the left and right columns represent the complex initial and final structures, respectively. Green patches on protein surface indicate Ras residues in contact with Raf (PDB ID : 3KUD [83]).

Regarding Mcl-1, all MD simulations of its complexes with the 3 peptides that were successfully docked show that the ligand positions in the protein binding cavity are stable during 200 ns (Figure 3.10). It is interesting to note that, given the rather plane shape of these cyclic peptides which does not fit well the rather long and narrow binding site of Mcl-1, we expected that their binding to the protein would not be very stable. However, our MD simulations indicate the opposite tendency which can be accounted for by the fact that Mcl-1 can distort to well accommodate the cyclic peptides [181]. Indeed, as shown in Figure 3.17, the protein RMSD increased to larger values (between 3 and 4 Å) than those of Ras which does not need to deform to bind the cyclic peptides. At the end, the 3 peptides NFWIW, NFFKW, and NWFIW successfully passed the two-step assessment.

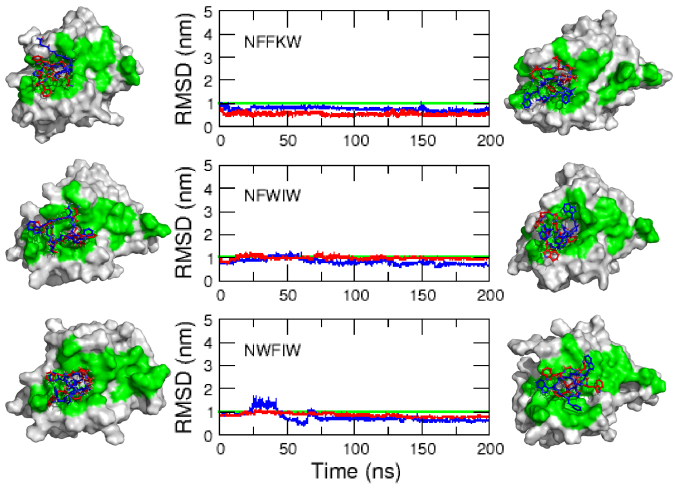

FIGURE 3.10 – Time evolution of the peptide Cα RMSD relative to the Des3PI hotspots in the MD simulations of their complexes with Mcl-1. Snapshots in the left and right columns represent the complex initial and final structures, respectively. Green patches on protein surface indicate Mcl-1 residues in contact with PUMA (PDB ID : 2ROC [62]).

In the MD simulations of the 6 peptides bound to Aβ, the protofibril RMSD stabilized at higher values (between 3 and 7 Å) than those of the globular proteins Ras and Mcl-1 (Figure 3.18). Regarding the peptides, the following observations can be made (Figure 3.11) : (i) the RISS peptide unbound the Aβ protofibril in one of its simulations, crossed the simulation box, and bound the opposite surface of the protofibril (figure not shown). In the second simulation, the peptide remained overall attached to the targeted surface but transiently unbound the protofibril, indicating that this peptide is not tightly held in place ; (ii) the peptides LFTW and LWTW of class II remained attached and quite close to the protofibril targeted surface, even if translations away from their initial position could be observed in half of their simulations. It should be noted that these translations may led these peptides to interact with the Aβ residues symmetrical to those initially contacted ; (iii) the 3 last assessed peptides also remained bound to the targeted surface of Aβ protofibril. Nevertheless, the peptides WYGW and WYIG also moved away from their putative binding site in half of their simulations, contrary to WYGK which firmly stayed around its targeted surface in both its simulations.

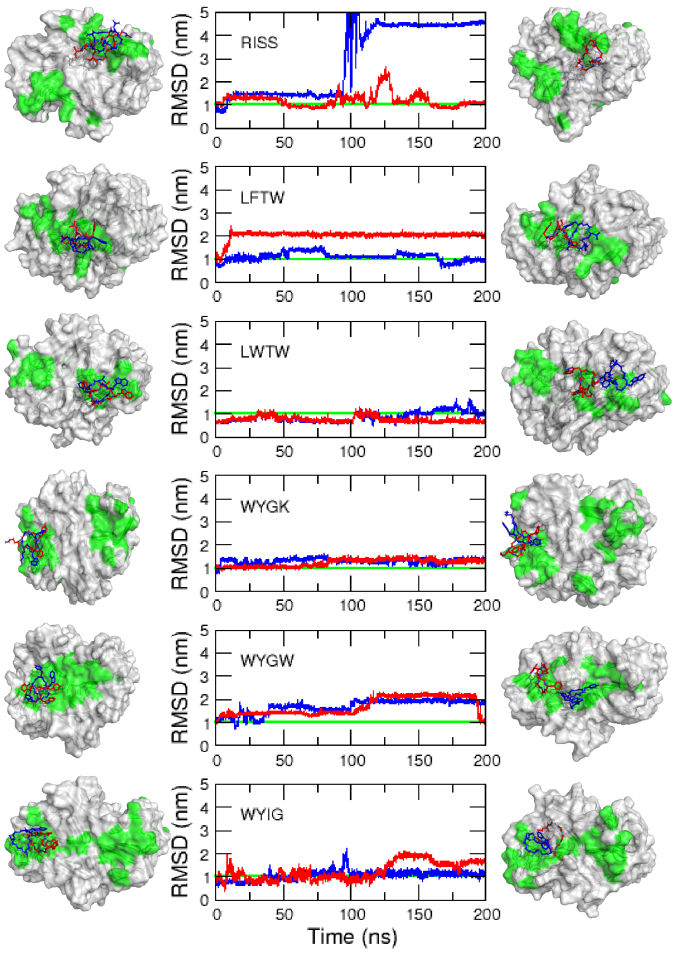

FIGURE 3.11 – Time evolution of the peptide Cα RMSD relative to the Des3PI hotspots in the MD simulations of their complexes with Aβ protofibril. Snapshots in the left and right columns represent the complex initial and final structures, respectively. Green patches on protofibril surface indicate Aβ residues initially in contact with docked peptides or residues symmetrical to the former ones.

To sum up, three quarters of the peptides targeting Ras, three over three targeting Mcl-1, and two thirds of those targeting Aβ were shown to form steady dynamic complexes with their targeted protein. This success rate of 77% suggests again that the peptides designed by Des3PI have good chances to bind in a stable way a targeted protein surface.

3.3.4 Analyzing the targeted protein contact residues

For each of the three targeted proteins, the most promising inhibitory peptides exhibit, after blind docking and MD simulations, RMSD relative to Des3PI hotspots between 0.4 and 1.1 nm. These values which can be considered as significant indicate that the peptide binding modes observed in simulations are not exactly those expected by Des3PI and reflect some conformational changes, global rotations, and/or translations of the peptides on the protein surface. Nevertheless, as illustrated in Figures 3.9 and 3.10, in most of the simulated protein-peptide complexes, the ligand steadily occupied a large part of the interface area and might therefore competitively inhibit the protein partner binding. To support this assumption, we compared the protein residues that are contacted by the peptides during simulations with those that are in contact with the partners in experimental complex structures.

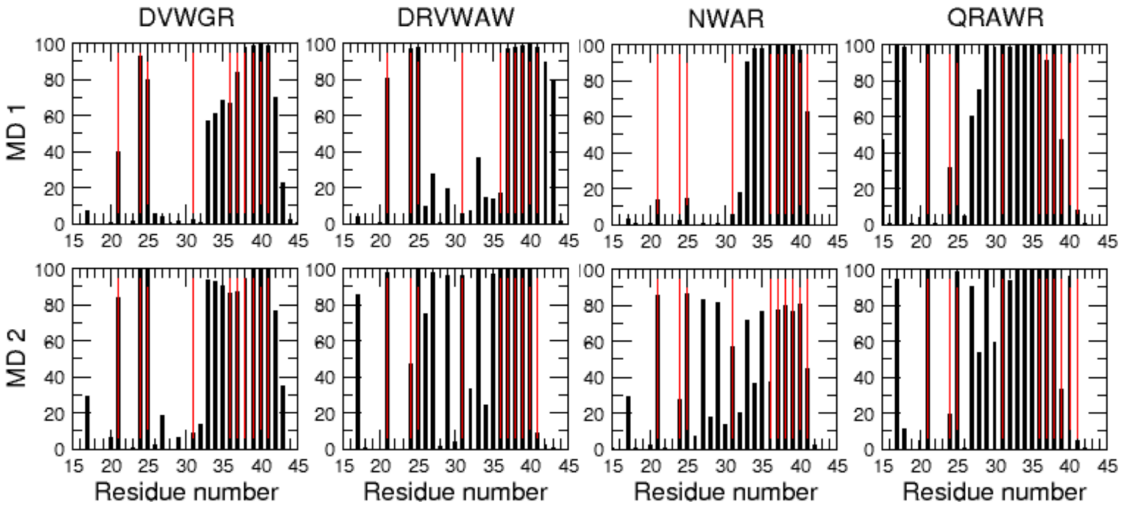

FIGURE 3.12 – Percentage of the MD trajectory times for which Ras residues are contacted by the inhibitory peptides (black bars). Red bars indicate Ras residues in contact with Raf in the 3KUD structure [83]).

Regarding Ras protein, over the 10 residues of Ras involved in the interface with Raf, 7 residues (Ile21, Gln25, Ile36, Glu37, Asp38, Ser39, and Tyr40) are contacted by the Des3PI peptides during an extensive part of all the MD trajectories, except in the first MD of Ras-NWAR complex (Figure 3.12). This indicates that the three peptides DVWGR, DRVWAW, and QRAWR steady occupy a large portion of the Ras-Raf interface and thus might be potent inhibitors of this complex.

In the six MD simulations of Mcl-1 complexes, the peptides could not occupy the whole binding groove which accommodates the PUMA α-helix in the

2ROC structure [62]). Indeed, the four residues Phe299, Phe300, Val302, and Gln303, which are located at one extremity of the groove (right side of the binding site colored in green in Figure 3.10), are never contacted by the three simulated peptides (data not shown). Nevertheless, over the 18 residues that constitute the rest of the Mcl-1 interface with PUMA, at least 10 residues (His205, Ala208, Phe209, Met212, Lys215, Val230, His233, Val234, Thr247, and Phe251) are in contact with the peptides during a large part of the MD trajectories (Figure 3.13). By stably occupying about half of the Mcl-1 binding groove, the three designed peptides might be considered as promising inhibitors of Mcl-1 interactions with PUMA.

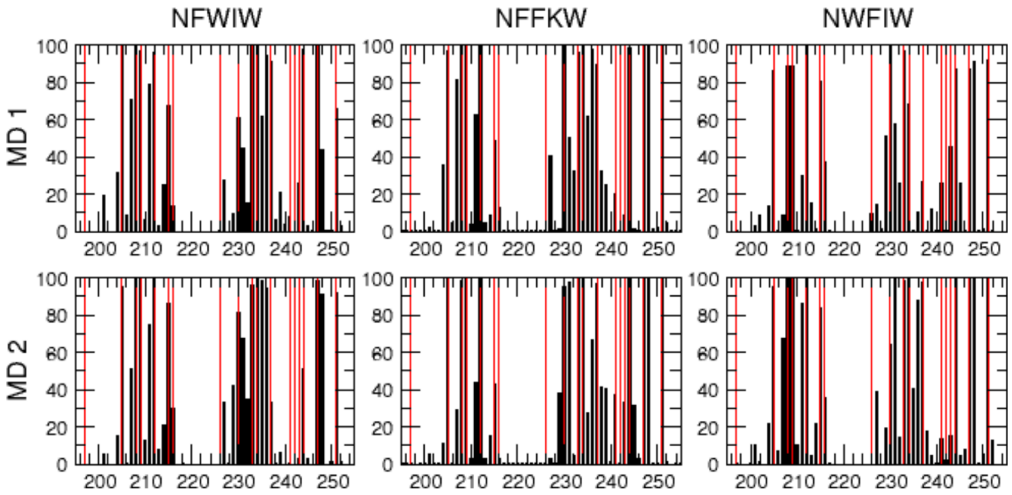

FIGURE 3.13 – Percentage of the MD trajectory times for which Mcl-1 residues are contacted by the inhibitory peptides (black bars). Red bars indicate Mcl-1 residues in contact with PUMA in the 2ROC structure [62]).

About Aβ protofibrils, the targeted binding surface is composed of the residues 11-42 of two chains and is clearly too large to be entirely occupied by the six peptides designed with Des3PI. That is why we compared here the Aβ most frequently contacted residues in simulations with those from the docking calculations to check the stability of the binding modes of the docked peptides. As displayed in Figure 3.14, the first simulated peptide, RISS, could not remain in the same location as the one predicted by docking, confirming that this binding mode is not stable. About the two peptides LFTW and LWTW, Figure 3.14 shows that, overall, they both kept their docking position on the Aβ protofibril surface during the simulations. It should be noted that, in the first MD simulation, LFTW slightly shifted to region 25-34 of the Aβ first chain, while it significantly moved toward the symmetrical residues 14-20 and 30-34 of the Aβ second chain, in the second trajectory. Regarding the peptides of the last class, we can observe a

similar dynamic behavior for the two peptides WYGK and WYIG, which, overall, remained in the same area as their docking position, but slightly moved toward the Aβ region 36-41 during their simulations (Figure 3.14). In contrast, WYGW is more mobile and can span across the binding surface to reach the symmetrical Aβ chain. Altogether, these analyses suggest that the four peptides LFTW, LWTW, WYGK, and WYIG are good binders of the Aβ protofibril targeted surface and might be potential good inhibitors of Aβ aggregation.

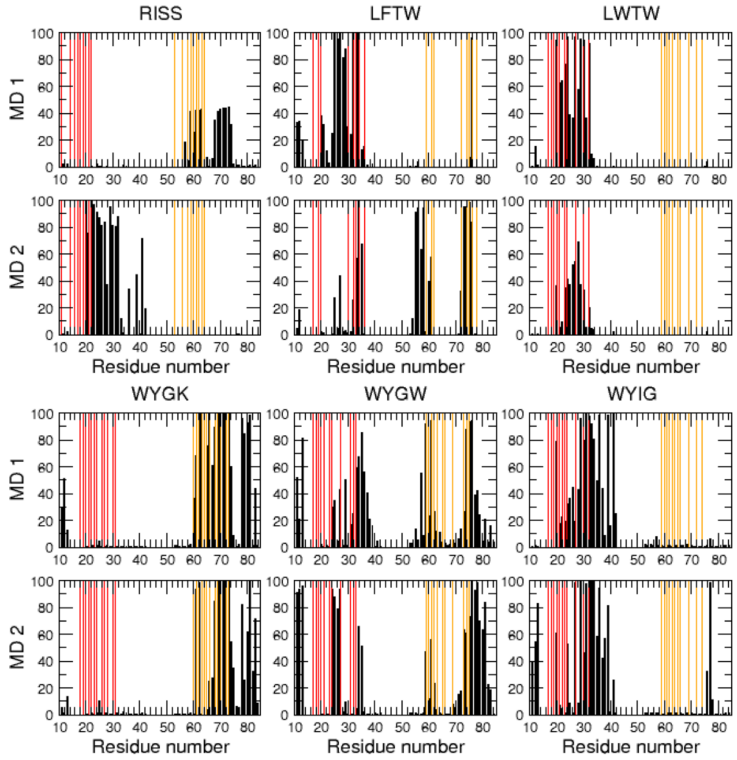

FIGURE 3.14 – Percentage of the MD trajectory times for which Aβ residues are contacted by the inhibitory peptides (black bars). Red bars indicate here Aβ residues in contact with each peptide after the blind docking. Orange bars indicate the Aβ residues symmetrical to the previous ones.

3.3.5 Peptide pharmacological properties

As highlighted by Vinogradov *et al.*, a major challenge in cyclic peptide design remains to optimize their pharmacological properties [283]. We think that this task is beyond the scope of this study which mainly aims at finding peptide sequences with potential high affinity for a targeted protein surface. Neverthe-

less, we would like to briefly discuss here some pharmacological properties of the most promising peptides found by DesPI (Table 3.1).

Peptide	Mass Weight	Polar Surface Area	LogP	LogS
DVWGR	841.87	381.46	-4.46	-4.92
QRAWR	985.06	477.16	-6.07	-4.88
DRVWAW	1099.16	455.45	-4.16	-8.03
NFFIW	993.07	349.88	-1.21	-8,77
NFFKW	1009.10	377.52	-2.99	-8.10
NWFIW	1032.11	365.67	-1.49	-9.32
LFTW	775.85	268.82	-0.62	-7.11
LWTW	814.89	284.61	-0.75	-7.67
WYGK	763.82	296.46	-2.87	-5.96
WYIG	747.80	268.82	-1.04	-6.64

TABLE 3.1 – Some pharmacological properties of the most promising Des3PI peptides. All these quantities were estimated using the SwissADME webserver [58]. Mass Weight and Polar Surface Area are in g/mol and $Å^2$, respectively. The reported logP and logS values are the average values of five and three predictions, respectively [58].

Unsurprisingly, the cyclic peptide molecular weights exceed the Lipinski's threshold of 500 Da [171], and, due to the large number of backbone hydrogen bond donors and acceptors, their polar surface areas are much larger than the 140 $Å^2$ criteria of Veber *et al.* [282]. It could also be noted that, despite the presence of many hydrophobic side chains in Des3PI peptides, they are globally rather hydrophilic as indicated by their negative octanol-water partition coefficient LogP. All together, these data suggest that our peptides could hardly diffuse through biological membranes and would have a poor oral bioavailability. Nevertheless, beyond the fact that many drugs do not satisfy the rule of 5 [71], it is possible to improve the peptide passive diffusion through biological membranes by reducing the number of backbone hydrogen bond donors. This could be achieved by using N-methylated amino acids, such as in the orally bioavailable immunosuppressant cyclosporine, a cyclic 11-residue peptide containing seven N-methylated amino acids [288].

Finally, it should be noted that SwissADME [58] predicts that our peptides are between poorly and moderately soluble in water, as indicated by their LogS values between -10 and -4. This is probably due to the presence of many hydrophobic and aromatic side chains in Des3PI peptides. Unfortunately, improving the peptide solubility by modifying these side chains would probably lead to a deterioration of both their affinity for the protein and their membrane permea-

bility. As for about 40% of the approved drugs, strategies to administer poorly soluble compounds remain to use drug delivery systems, such as emulsion, liposome, or polymer encapsulation [132].

3.3.6 Computation times

Before concluding, we would like to give an idea about the time needed for finding new peptide binders of a targeted protein with Des3PI (Table 3.3.6). The generation of the peptide hotspots and sequences per se is a process that takes less than a day, at the end of which the user is provided with several peptide candidates which can be synthesized and tested experimentally. An increased confidence in Des3PI results can be achieved by running protein-peptide blind docking calculations, at the cost of a few extra computation hours. Naturally, the protein-peptide complex stability assessment by MD simulations is by far the most time consuming process, but we consider this step as optional if experimental studies can be envisaged in a near future.

	Peptide generation (50 x 20 x 20 docking)	Peptide blind docking (n docking*)	Complex stability (1 simulation of 200 ns)
Ras	16 h	2 h	252 h
Mcl-1	20 h	3 h	353 h
Aβ	19 h	3 h	505 h

Computation time of each step of Des3PI approach for the three proteins studied herein. These times were obtained on a local Linux workstation with 12 CPU of 3.7 Ghz. *Each peptide blind docking consists in n docking calculations where n is the peptide number of residues.

3.4 Conclusion

Given a protein surface involved in a protein-protein interaction that we want to perturb with a cyclic peptide, Des3PI aims at identifying sequences that potentially bind the targeted surface with high affinity. In this report, the principle and algorithm of the fragment-based approach implemented in Des3PI were described in detail in the Methods section. Des3PI was applied to three different protein interfaces, one composed of α-helices (Mcl-1), a second one of β-strands (Aβ protofibrils), and a third one comprising both of them (Ras). For each of these targets, Des3PI was able to provide at least five different peptide sequences with potential high affinity for the proteins. For large and flat protein surfaces, such as Ras or Aβ protofibril, Des3PI can even generate a dozen or more peptide hits.

The ability of the peptides designed by Des3PI to correctly and stably bind their targeted protein surfaces were tested by a two-step validation protocol consisting in a blind docking of the peptides onto the targeted proteins, followed by stability tests of the complexes found among the 5% best scores and with the lowest peptide RMSD with respect to Des3PI hotspots. For Ras, all the 20 peptides provided by Des3PI were successfully docked, and 4 representative ones in complex with Ras were submitted to MD simulations. Three of them exhibited a good stability of their binding to Ras targeted surface. Among the 5 peptides that Des3PI yielded for Mcl-1, three successfully passed both the blind docking and stability test. Finally, when targeting Aβ protofibril, Des3PI generated 12 different peptides, half of them were successfully docked, and 4 over 6 exhibited steady binding to the targeted surface perpendicular to the protofibril axis.

At the end, the overall success rate of the two-step validation procedure is about 60% for the three protein targets studied herein. This encouraging result suggests that our peptide design program Des3PI is a reliable tool to identify in the immense space of peptide sequences those which are most likely to bind a protein surface target. Of course, these identified peptides require to be synthesized and tested *in vitro* to fully validate this approach. These experimental studies have been initiated for the peptides targeting Mcl-1 and Aβ protofibril, and are in progress.

Supplementary informations

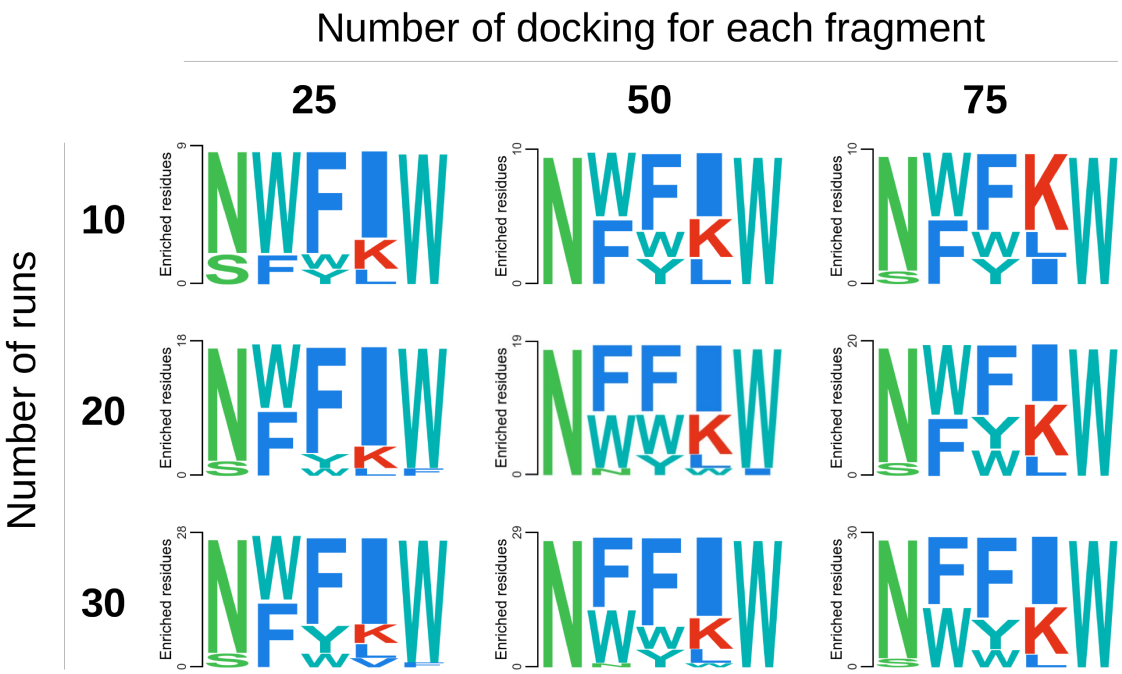

FIGURE 3.15 – Amino acid occurrences in the peptides generated by Des3PI for targeting Mcl-1 protein as a function of the number of docking calculations of each amino acid of the library (columns) and as a function of the number of runs (rows).

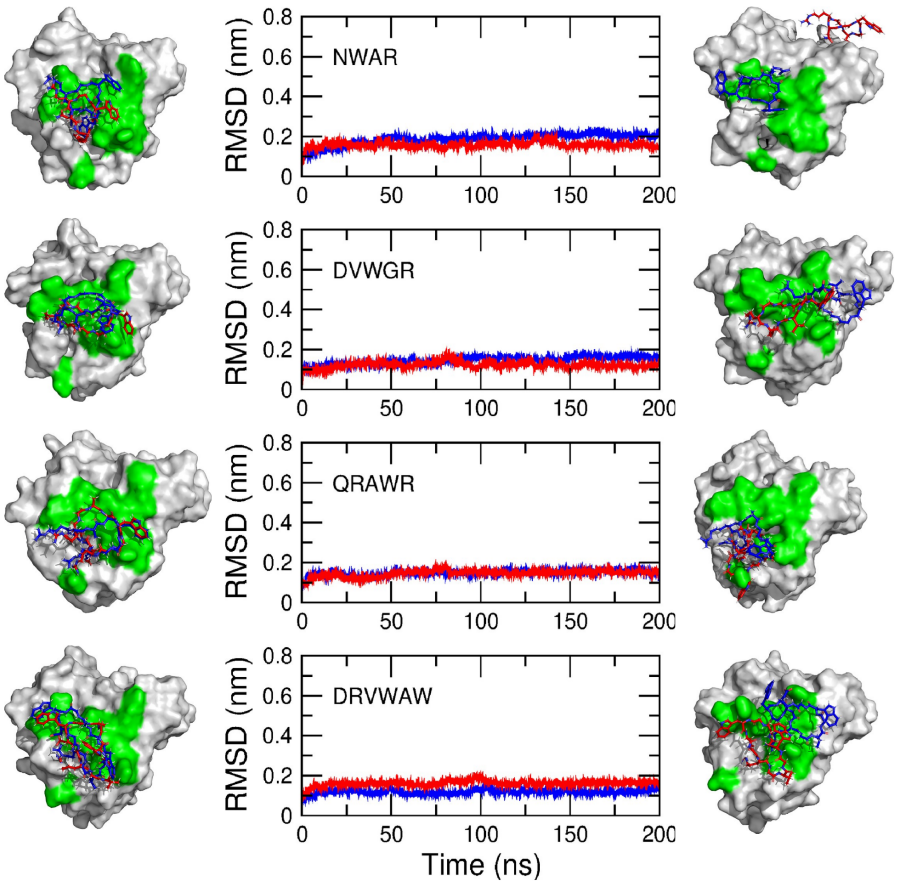

FIGURE 3.16 – Time evolution of Ras protein Cα RMSD relative to its initial conformation in the Ras-peptide complex MD simulations. Snapshots in the left and right columns represent the complex initial and final structures, respectively. Green patches on protein surface indicate Ras residues in contact with Raf (PDB ID : 3KUD [83]).

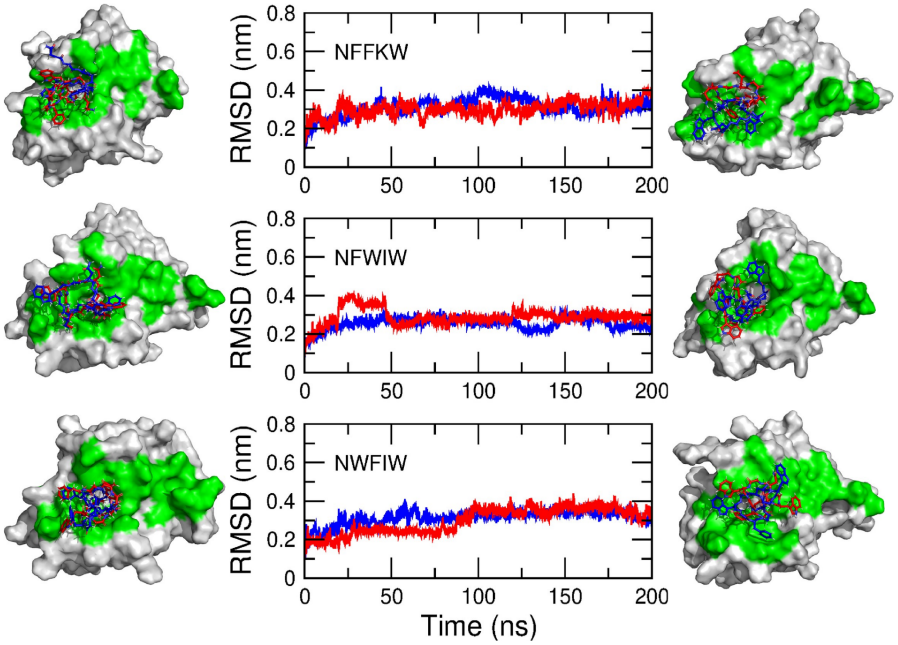

FIGURE 3.17 – Time evolution of Mcl-1 protein Cα RMSD relative to its initial conformation in the protein-peptide complex MD simulations. Snapshots in the left and right columns represent the complex initial and final structures, respectively. Green patches on protein surface indicate Mcl-1 residues in contact with PUMA (PDB ID : 2ROC [62]).

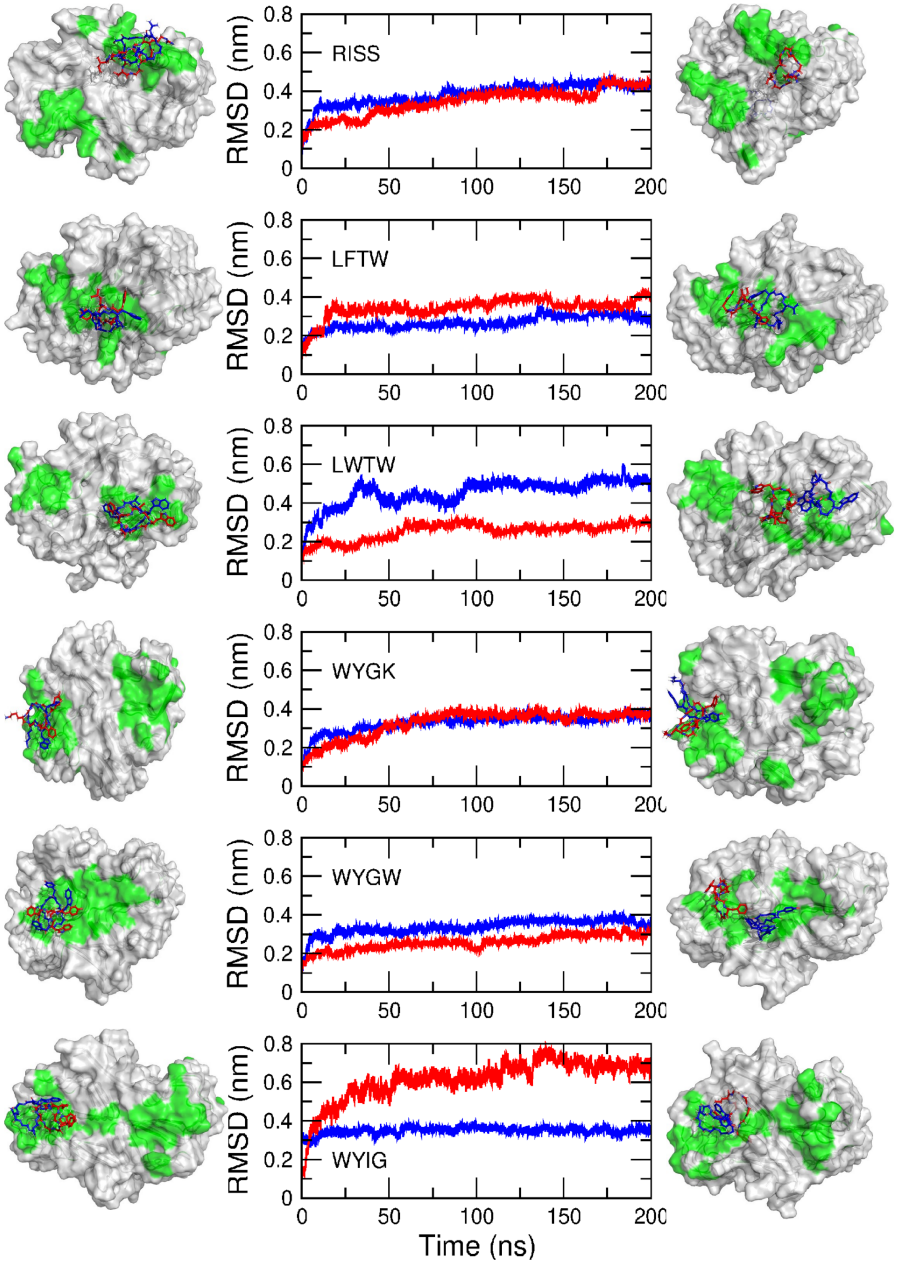

FIGURE 3.18 – Time evolution of the Aβ protofibril Cα RMSD relative to its initial conformation in the protofibril-peptide complex MD simulations. Snapshots in the left and right columns represent the complex initial and final structures, respectively. Green patches on protofibril surface indicate Aβ residues initially in contact with redocked peptides or residues symmetrical to the former ones.

Chapitre 4

Conception *in silico* de peptides cycliques inhibiteurs d'interactions protéine-peptide

Nous avons précédemment présenté une approche *in silico* produisant des résultats prometteurs. En effet, le protocole de validation computationnelle a permis de mettre en lumière plusieurs séquences pour chacun des trois systèmes visés qui semblent avoir une bonne énergie libre de liaison. Bien que nous devions attendre les résultats expérimentaux pour confirmer ou non l'efficacité de notre approche, nous voulions tout de même chercher à savoir s'il était possible d'élargir le champs d'action de Des3PI.

En effet, bien que les trois systèmes précédents Ras, Mcl-1 et Aβ possédaient une diversité de structure intéressante, il s'agissait dans les trois cas d'interfaces possédant une zone d'interaction très vaste et peu creusée, comme dans de nombreuses IPP. Néanmoins, au sein des IPP, il existe un type d'interface faisant intervenir une cavité en sillon beaucoup plus étroite avec comme partenaire un petit fragment peptidique linéaire (*Short linear Motif*, SLiM). Ce type d'interface est notamment retrouvé dans les domaines PDZ ou SH3.

Nous voulions donc savoir si notre approche, initialement pensée pour une interface plane et vaste, était capable de cibler ce type d'interface protéine-peptide. Nous avons donc appliqué Des3PI sur quatre cibles protéiques d'interfaces différentes. Les domaines PDZ et SH3, avec une interface en forme de sillon, mais également le récepteur CXCR4 et l'intégrine αVβ3. L'interface de ces protéines est une poche profonde proche de ce que l'on peut observer dans des interfaces enzyme-substrat, mais bien plus grandes. De plus, ces deux dernières protéines ont pour avantage d'avoir chacune été cocristallisée en interaction avec un peptide cyclique inhibiteur. Ainsi nous pouvions comparer directement les peptides cycliques générés par Des3PI avec ceux connus expéri-

mentalement. Les domaines PDZ et SH3 étaient quant à eux cocristallisés avec un petit peptide linéaire chacun, que nous comparerons également à nos séquences peptidiques.

Le but de ce travail a été donc de savoir si l'utilisation de Des3PI est adaptée à différentes formes d'interfaces protéine-peptide mais également de savoir si cette approche nous permet de retrouver des séquences proches de celles connues pour être affines expérimentalement. Le cas échéant, nous voulions également savoir si les peptides générés par Des3PI étaient plus affins pour leur cible que les peptides cocristallisés.

Ce chapitre est une adaptation du manuscrit *Computational Design of Cyclic Peptides to Inhibit Protein-Peptide Interactions, Delaunay et Ha-Duong* soumis au journal *Proteins : Structure, Function, and Bioinformatics, 2022*.

4.1 Introduction

Modulating protein-protein interactions (PPIs) is a promising strategy to interfere with normal or pathological cellular activities. Thus, developing molecules able to selectively inhibit or stabilize PPIs has become a very active research field, for either medicinal or chemical biology purposes [175]. Many PPIs tightly occur between folded globular proteins to form elaborate supramolecular architectures which fulfill specific biological functions, such as DNA replication, protein synthesis, or molecular transport. In this type of complexes, the protein-protein interfaces are generally large and their modulation by small molecules is considered difficult [255].

In this context, peptide derivatives, due to their larger size and higher specificity than those of small compounds, have attracted great interest in developments of efficient PPI inhibitors [156]. Among these, head-to-tail cyclic oligopeptides (with around 10 amino acid residues) have a rather flat geometry and are thus appropriate to strongly bind flat protein surfaces involved in PPIs [333]. Moreover, cyclic peptides generally have a better stability against proteases and an increased membrane permeability, which improves their pharmacokinetic properties, relative to their linear counterparts [289]. Nevertheless, since large PPI interfaces are generally composed of intricate patterns of various hydrophobic, polar, and charged patches [159], the identification of cyclic peptide sequences with high affinity for a given targeted surface remains a major challenge.

To help in designing such cyclic peptides, we have recently developed a computational fragment-based approach, called Des3PI (Design of Peptides targeting Protein-Protein Interactions). Briefly, the algorithm performs docking calculations of a library of amino acids on a targeted protein surface and then links those with good binding energy to generate the cyclic peptide sequences

and structures. This *in silico* peptide design was efficiently applied to three large protein surfaces involved in PPIs : the Ras and Mcl-1 proteins, as well as an Aβ protofibril [65].

However, many other protein-protein complexes do not result from the association of two globular proteins, but are rather formed through the binding of one folded domain with one short peptide segment, also called short linear motif (SLiM). Well known examples of such protein-peptide complexes are formed by SH3 [150] or PDZ [163] domains, which play crucial roles in signal transduction pathways and regulations of many cellular activities. These domains specifically bind peptide segments of around 6 amino acids in length [61]. This limited number of residues at these protein-peptide interfaces accounts for their relative low binding affinities, often resulting in reversible and transient associations.

Given the importance of these protein-peptide interactions in controlling various biological processes, targeting these PPIs has become a promising and attractive strategy to develop drugs against several diseases, including cancer, neurological disorders, and viral infections [176, 49]. It could be noted that the low affinity of these PPIs can facilitate the discovery of good competitive binders, but the large number of proteins possessing those folded domains challenges the design of inhibitors with high specificity.

In this study, we assessed the ability of our program Des3PI [65] to generate head-to-tail cyclic peptides that can competitively bind protein surfaces involved in interactions with SLiMS or peptide segments. For this purpose, we applied Des3Pi on 4 folded proteins, the αVβ3 integrin [306], the CXCR4 chemokine receptor [302], the Src tyrosine kinase SH3 domain [82], and the PDZ domain of Shank3 protein [220]. Then we compared the Des3PI peptide sequences and binding modes to those of the peptide co-crystallized with each of the 4 previous proteins, namely the cyclic RGD ligand (PDB ID : 1L5G), the cyclic CVX15 antagonist (PDB ID : 3OE0), the proline-rich RLP2 peptide (PDB ID : 1RLQ), and the C-terminal GKAP peptide (PDB ID : 5OVC), respectively.

4.2 Methods

4.2.1 Des3PI algorithm

The details of Des3PI method were reported in a previous article [65], thus only the principal elements of the algorithm are underlined here (Figure 4.1). Des3PI is a fragment-based approach which first performs docking of the 20 natural amino acids onto a targeted protein surface using Autodock Vina [278]. Then, the amino acid Cα of all the binding modes are clustered and the most abundant amino acids in each cluster define the hotspots of the designed pep-

tide. The hotspots closed enough to compose a cyclic peptide are finally linked to each other with one or several glycine residues according to their relative distance.

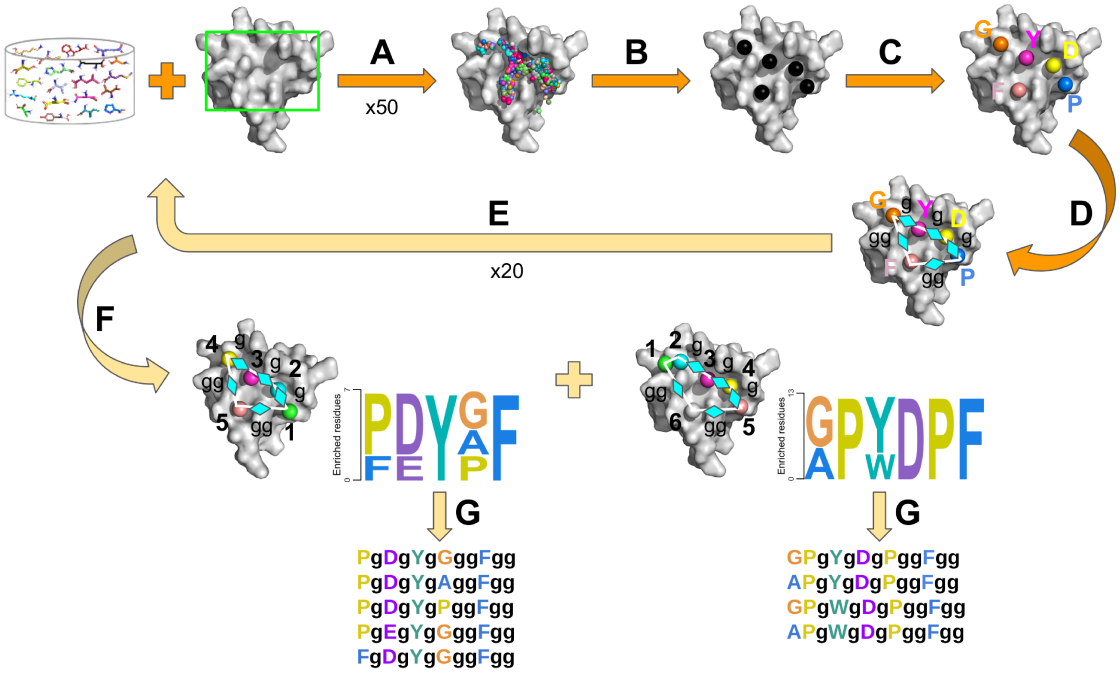

FIGURE 4.1 – Des3PI workflow. Each amino acid is docked 50 times (A). All the binding modes are clustered (B). The most abundant amino acid is determined in each cluster to obtain the hotspots (C). The hotspots are linked using one or several glycines to provide a cyclic peptide (D). Steps A,B,C, and D are repeated 20 times to generate multiple peptides (E). Peptides with equal number of hotspots and similar geometry are grouped into classes of peptides (F). For each class, the amino acid occurrences and the five best peptide sequences are determined (G). Lowercase **g** indicates glycine linkers.

4.2.2 Peptide docking

As a first validation step, the cyclic peptides generated by Des3PI were blindly docked on their target using Autodock CrankPep (ADCP) [324]. This software is dedicated to docking peptides onto proteins and is notably able to dock head-to-tail cyclic peptides. As inputs, ADCP requires the three-dimensional structure of the targeted protein and the peptide string sequence. For cyclic peptides, we noticed that ADCP could generate slightly different results depending on the input first and last residues. Thus, for a peptide composed of N residues,

we docked the cyclic sequence N times, using N different starting residues. Each ADCP calculation consisted in 50 independent searches of 2 500 000 Monte Carlo steps, and output 100 best binding modes.

We analyzed the ADCP results by computing the root mean square deviation (RMSD) of the docked peptide Cα with respect to the hotspots generated by Des3PI. We considered that a cyclic peptide successfully passes this validation step if at least one binding mode among the top 5% of ADCP scores was found close to the Des3PI hotspots with a RMSD lower than 10 Å.

4.2.3 Molecular dynamics simulations

In a second validation step, the protein-peptide complex with the lowest peptide RMSD among the top 5% of the best binding energies was selected and submitted to two independent MD simulations to check its stability. The GROMACS 2019.1 package [4] was used to perform the simulations. The AMBER-99SB-ILDN [275] and GAFF [290] force fields were employed for the protein and peptide, respectively. Each complex was initially placed in a cubic box, so that the minimal distance between the solute and the cube faces was equal to 1 nm. Then the complex was solvated with TIP3P water molecules and neutralized with 0.15 mol/L of sodium chloride. The Lennard-Jones potentials were cut off at 1.2 nm and the Coulomb interactions were treated using the smooth PME method [78]. Each system was first minimized using 10 000 steps of the steepest descent method, then submitted to two short equilibration runs of 1 ns each, the first one to heat the system to 310 K using a Berendsen thermostat and the second one to equilibrate the pressure around 1 bar using the Parinello-Rahman method. After that, a 200 ns production run was performed in the isothermal–isobaric (NPT) ensemble using the Nose-Hoover and Parrinello-Rahman coupling algorithms with the time constants $\tau_T = 0.5$ ps and $\tau_P = 2.5$ ps, respectively [207, 109, 212]. The equations of motion were integrated using the leap-frog algorithm with a 2 fs time step, while keeping constant the length of all covalent bonds using the LINCS procedure [107]. MD trajectory frames were saved every 20 ps for subsequent analysis.

4.2.4 Free energy calculations

Finally, a third validation step consisted in estimating the binding free energy of cyclic peptides that can form stable complexes with their targeted protein, and in comparing it with the binding free energy of the reference co-crystallized peptide segments. These free energy calculations were based on the molecular

mechanics Poisson–Boltzmann surface area (MM-PBSA) approach :

$$\Delta G_{Bind} = \Delta E_{MM} + \Delta G_{Solv} - T\Delta S_{Conf} \quad (4.1)$$

where ΔE_{MM} is the change in the mean potential energy of the protein-ligand association in vacuum. $\Delta G_{Solv} = \Delta G_{PB} + \Delta G_{SA}$ is the variation of the solvation free energy which polar and apolar contributions can be calculated by using the Poisson-Boltzmann (PB) equation and the molecular solvent-accessible surface areas (SA), respectively. Lastly, $T\Delta S_{Conf}$ is the configurational entropy change upon the complex formation in gas phase.

The first two terms of Equation 5.1 were computed by applying the *g_mmpbsa* tool [148] on the last 100 ns of the stable complex simulations. We kept the default values for all *g_mmpbsa* parameters, except for the solute dielectric constant used in ΔG_{PB} calculation. Indeed, preliminary benchmarks with different values of ϵ_{solute} (2, 4, or 8) showed that $\epsilon_{solute} = 8$ yielded the closest results to available experimental data for SH3 and PDZ complexes [220, 82]. The configurational entropy term were estimated using the quasi-harmonic method in the dihedral angle space [166]. For that, the covariance matrix of the backbone (ϕ, ψ) angles was calculated over the last 100 ns of each trajectory [200] and diagonalized using the *gmx covar* GROMACS tool. Then the eigenvectors are analyzed with *gmx anaeig* which computes the configurational entropy using the Schlitter's approximation [249].

At this point, it is important to mention that *g_mmpbsa* calculations are based on the assumption that the protein and ligand structures do not change upon binding [148]. Indeed, free energies of the two molecules in their unbound state are computed on the conformations artificially extracted from the MD simulations of their complex. This assumption is reasonable for folded globular proteins and small compounds with few degrees of freedom, but not for highly flexible molecules, such as peptides. Thus, to better take into account the contribution of the peptide conformational changes in both the solvation energy and configurational entropy terms, we employed the procedure described below (Figure 4.2).

First, from a protein-peptide complex MD simulation, we extracted the molecular mechanics term ΔE_{MM} which is the average of the sum of the intermolecular Lennard-Jones ($E_{LJ}^{Prot-Pept}$) and Coulomb ($E_{Coul}^{Prot-Pept}$) energies. The application of *g_mmpbsa* on the complex trajectory was also used to computed the solvation free energy of the complex ($G_{Solvation}^{Cplx}$) and of the unbound protein ($G_{Solvation}^{UnboundProt}$) which is assumed to undergo negligible conformational changes. Lastly, the complex simulation allowed to compute the configurational entropy of the peptide in the bound state ($S_{Conf}^{Pept \in Cplx}$).

Secondly, we ran an independent 200 ns MD simulation of the unbound peptide in water using the same parameters as described in the previous subsec-

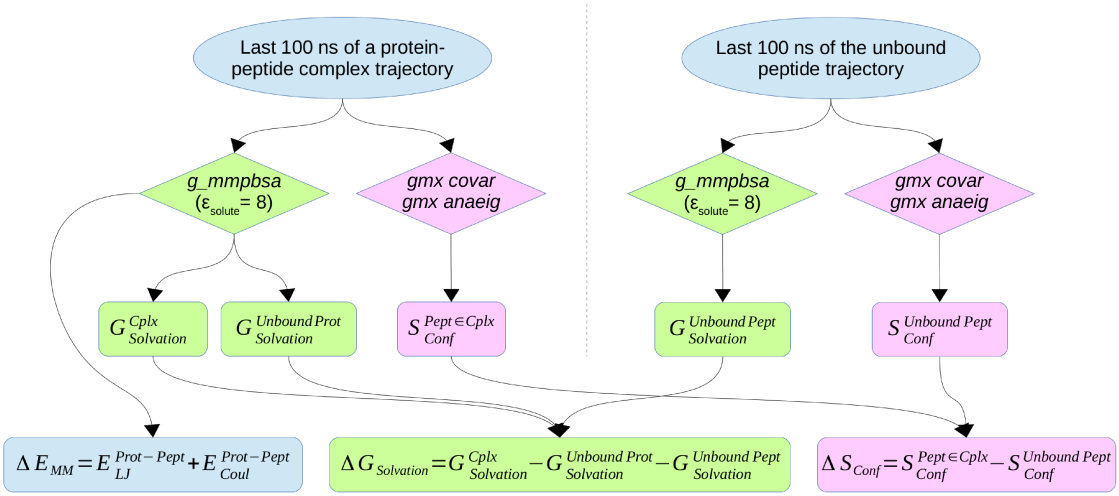

FIGURE 4.2 – Schematic description of the approach used to compute the three terms of the binding free energy of a protein-peptide complex in Equation 5.1.

tion. The last 100 ns of the trajectory were then used to compute both the solvation free energy ($G^{UnboundPept}_{Solvation}$) and the configurational entropy ($S^{UnboundPept}_{Conf}$) of the unbound peptide. These quantities were finally combined with those extracted from the complex simulations to yield the protein-peptide binding free energy (Figure 4.2). It should be noted here that *g_mmpbsa* yields estimations of the different contributions of the binding free energy with their associated standard deviations. But for the configurational entropy, *gmx covar* and *gmx anaeig* tools do not provide standard deviations of the estimations. To compensate this lack, we performed bootstrap analyses as follows : The last 100 ns of each MD trajectories were first divided into 10 equal parts. Secondly, 1 000 bootstrap trajectories of 100 ns each were generated by randomly taking with replacement 10 parts of the original trajectories. Then, the *gmx covar* and *gmx anaeig* tools were applied on each of the 1 000 bootstrap trajectories to yield the configurational entropy mean values and standard deviations.

4.3 Results and Discussion

4.3.1 Peptide sequences generated by Des3PI

As previously mentioned, we would like to compare the peptide sequences generated by Des3PI with those of peptides which are experimentally known to be good binders of four protein examples. We first considered the possibility to design peptides which can compete with the cyclic pentapeptide Arg-Gly-

Asp-dPhe-mVal (where dPhe and mVal denote a D-Phe and a N-methylated Val) bound to the extracellular segment of the $\alpha V\beta 3$ integrin (PDB ID : 1L5G) [306].

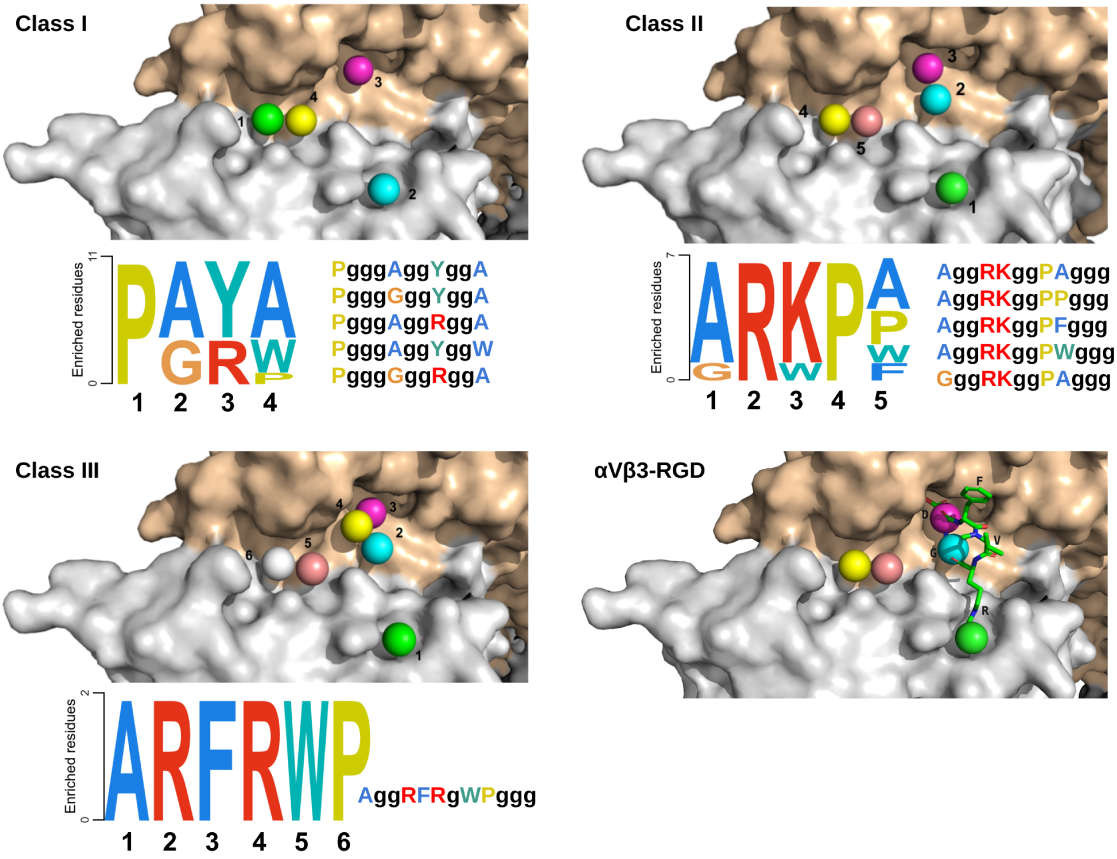

FIGURE 4.3 – Des3PI generated 3 classes of peptides which can bind the $\alpha V\beta 3$ integrin at the RGD cyclic peptide binding site. Bottom right panel displays RGD binding mode superimposed on class II hotspots.

The application of Des3PI on the integrin binding area of the co-crystallized pentapeptide yielded 3 classes of peptides which amino acid occurences in the hotspots are displayed in Figure 4.3. These peptides have 4, 5, or 6 hotspots which lie at the bottom of the cleft between the propeller and βA domains of the integrin head. It should be noted that these hotspots occupy a larger area on the protein surface than that of the pentapeptide which mainly makes contacts with the integrin through its RGD motif [306]. More specifically, the locations of the 3 residues Arg, Gly, and Asp match with the hotspot positions 1, 2, and 3 of peptide classes II, respectively (Figure 4.3).

However, the amino acids most frequently found by Des3PI at these three hotspot positions are very different from the RGD sequence. At position 1, an

Ala or a Gly is found, while an Arg is predominant at position 2. But the most striking difference is that Des3PI cannot retrieve an Asp residue at position 3. It should be noted that this inability to generate an Asp residue at this position is observed in the presence of the divalent ions Mn^{2+} co-crystallized with the protein, suggesting that this might be due to a poor accuracy of the ligand-ion interaction treatment in the Autodock Vina scoring function [278].

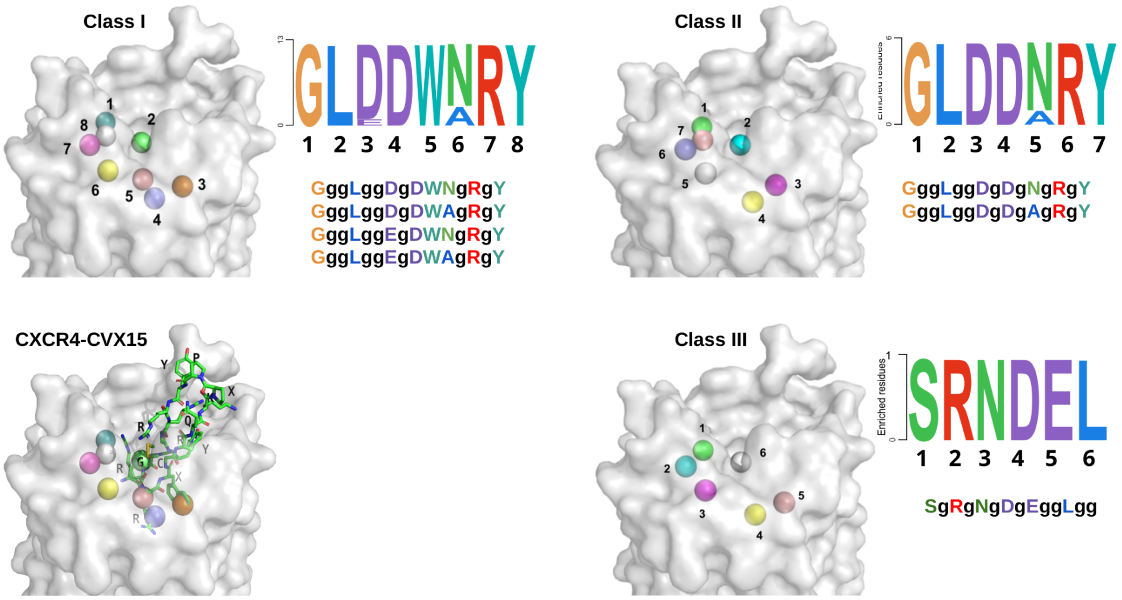

FIGURE 4.4 – Des3PI generated 3 classes of peptides which can bind the CXCR4 receptor at the CVX15 cyclic peptide binding site. Bottom left panel displays CVX15 binding mode superimposed on class I hotspots.

On CXCR4 receptor (PDB ID : 3OE0), Des3PI generated 3 classes of cyclic peptides, with 6 to 8 hotspots (Figure 4.4). Referring to the subdivision of the receptor transmembrane cavity into a deep major sub-pocket and a shallow minor one [235], the class I hotpsots 1, 7, and 8 are located in the minor sub-pocket, the hotspots 3, 4, and 5 in the major one, and the 2 and 6 at the edge of the two sub-pockets. In contrast, the co-crystallized 16-residues peptide CVX15, which is structured into a β-haiprin cyclized by a disulfide bond between Cys4 and Cys13, mainly occupies the major sub-pocket. Notably, its residues Arg2 and Nal3 are localized at the bottom of the major sub-pocket close to class I hotpspots 3, 4, and 5, while both ends Arg1 and dPro16 lie at the edge of the two sub-pocket, near hotspots 6 and 2, respectively (PDB ID : 3OE0) [302].

But, comparing the compositions in amino acids of these hotspots, they differ completely from the CVX15 residues in interaction with CXCR4 : Although a Trp is found at position 5 instead of position 3 where the naphtalene ring of

Nal3 is localized, a negatively charged Asp is mostly yielded at hotspots 3 and 4 where a positively charged Arg and a hydrophobic residue are expected. For hotspots 2, a hydrophobic residue Leu is always found which is rather consistent with the location of the dPro16, but Des3Pi mostly generated an Asn at position 6, where the CVX15 Arg1 lies. Altogether, it seems that Des3PI has difficulties to generate correct charged residues at expected location of the targeted surface.

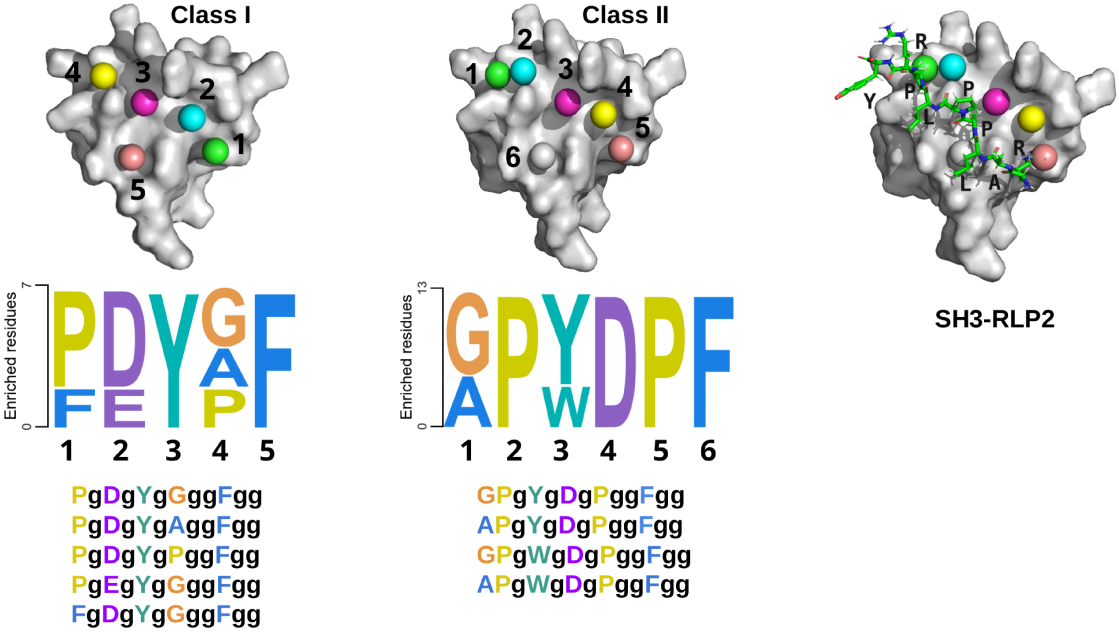

FIGURE 4.5 – Des3PI generated 2 classes of peptides which can bind the Src tyrosine kinase SH3 domain at the proline-rich RLP2 peptide binding site. Right panel displays RLP2 binding mode superimposed on class II hotspots.

When applied to the SH3 domain (PDB ID : 1RLQ), Des3PI generated 2 classes of cyclic peptides, with 5 or 6 hotspots. The hotspots are approximately aligned (except one of them in each class) nearly parallel to the axis of the RLP2 peptide (Figure 4.5). Interestingly, in both classes, several prolines are retrieved among the most frequently encountered amino acids in Des3PI sequences, like in the proline-rich RLP2 peptide. In class II peptides, the proline found at hotspot 2 is similarly located as RLP2 Pro77, but the one at hotspot 5 is retrieved at a location close to RLP2 Arg71. Close to RLP2 residues Pro74 and Pro75, Des3PI generated the hotspot 3 that is rather an aromatic and hydrophobic residue (Tyr of Trp), and a Phe hotspot is always found where is located RLP2 Leu73. Finally, it could be noted that an Asp residue is retrieved at hotspot 4 near Arg19 of SH3 domain. Overall, the peptides designed by Des3PI have residues quite similar in physical chemical properties to those of the RLP2 ligand (except its

Arg71) and their sequences appear rational to bind the SH3 domain with good affinities.

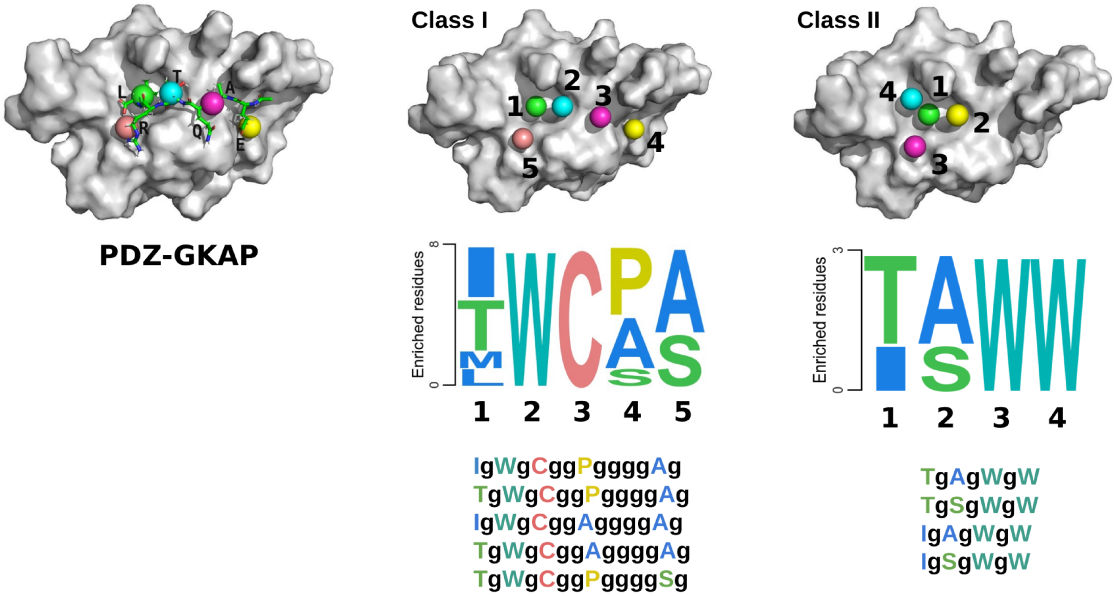

FIGURE 4.6 – Des3PI generated 2 classes of peptides which can competitively bind the PDZ domain of Shank3 protein at the C-terminal GKAP peptide binding site. Left panel displays GKAP binding mode superimposed on class I hotspots.

On the PDZ domain (PDB ID : 5OVC), Des3PI generated 2 classes of cyclic peptides (Figure 4.6). The first one has 5 nearly aligned hotspots which occupy almost the entire binding surface of GKAP peptide on the protein. The second one has only 4 hotspots which are grouped in a star geometry and occupy a bit more than half of the protein-peptide interface. Furthermore, the most frequently encountered amino acids in the class II hotspots are mostly hydrophobic, whereas GKAP peptide has several charged and polar residues. In contrast, the class I peptides has more polar amino acids, notably a Cys residue at hotspot 3 which occupies the same position as GKAP Gln989 residue. At hotspots 1 and 2, hydrophobic amino acids (Ile, Trp) are frequently retrieved at the locations of the peptide Leu992 and Thr990 residues. Nevertheless, at hotspots 4 and 5, a Pro and Ala amino acids frequently appear where GKAP Glu987 and Arg991 residues are lying, respectively.

In summary, on $\alpha V\beta 3$ integrin and CXCR4 receptor, Des3PI generated cyclic peptides with sequences and configurations weakly related to those of known peptide inhibitors, indicating that our algorithm is poorly effective to find better peptide binders for these 2 proteins. In contrast, on SH3 and PDZ domains which bind short linear motifs, Des3PI was able to design cyclic peptides with

appropriate configurations and consistent sequences for binding these 2 proteins with good affinities. To further assess the quality of Des3PI peptides, we performed their blind docking onto SH3 and PDZ domains, carried out MD simulations of the most promising complexes, and estimated their binding free energies.

4.3.2 Docking of Des3PI peptides on SH3 and PDZ domains

The best peptide sequences output by Des3PI (Figures 4.5 and 4.6) were blindly docked on the entire domains (without specifying any restricted surface) to check whether they favorably bind the targeted surfaces. Regarding the peptides designed against SH3, they can generally reach the targeted binding site, since one or several binding modes among the top 5% ADCP scores have Cα RMSD lower than 10 Å with respect to Des3PI hotspots (Figure 4.7). Among them, the four class I peptides PDYGF, PDYAF, FDYGF, and PEYGF have interesting low energy binding modes close to the 5 Des3PI hotspots. We also notice that the class II peptides, which have one additional specific hotspot, overall do not do better than the class I ones, both in terms of binding energies and RMSD, probably because the supplementary amino acids (an Ala or Gly) are too small to significantly improve the binding energy.

In contrast, the proline-rich RLP2 peptide could be blindly docked into SH3 domain with much better binding energies and RMSD (relative to its Cα except those of the 2 last residues (Arg78 and Tyr79) which are not in contact with the protein) than those of the 9 tested Des3PI peptides (Figure 4.7). Aside from indicating that ADCP is rather effective to predict the binding modes of peptides onto proteins, these results suggest that the peptides designed by Des3PI might be weak competitors of the RLP2 peptide to bind the SH3 domain.

Regarding the docking of the best Des3PI peptides on PDZ domain, it can be observed that the two classes of peptides yielded different RMSD-ADCP score graphs (Figure 4.8). The class II peptides, which have compact configurations, are more frequently docked close to the Des3PI hotspots than the class I peptides which have more extended conformations. Nevertheless, the lowest ADCP binding energies of the latter are generally lower than those of the class II peptides. However, similarly to the SH3 case, none of the 9 tested Des3PI peptides could be docked with better binding energy and RMSD than the GKAP peptide co-crystallized with PDZ (Figure 4.8). Again, this suggests that Des3PI peptides might not efficiently compete with GKAP for the binding of PDZ domain. Nevertheless, it is known that energy scores of protein-ligand docking programs are only rough estimations of binding free energies, and in particular, do not well account for the solvation and entropy contributions. To deeper investigate the relative affinities of Des3PI peptides versus those of RLP2 and GKAP for SH3

and PDZ domains, respectively, we performed and analyzed MD simulations in explicit solvent of the most promising protein-peptide complexes.

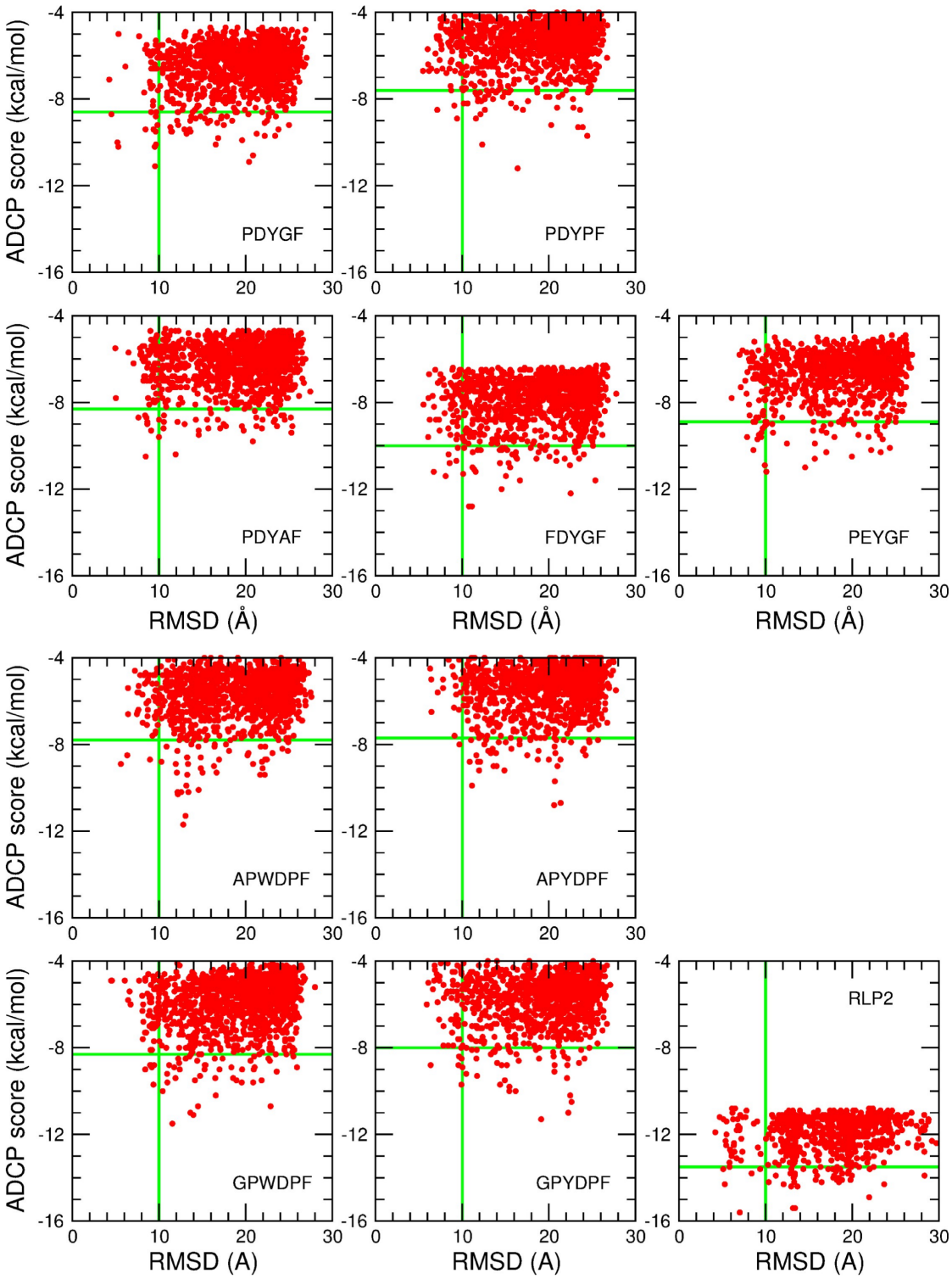

FIGURE 4.7 – ADCP score of the binding modes on SH3 domain of the 9 best Des3PI peptides (class I and II) and of RLP2 peptide as a function of their Cα RMSD relative to the hotspot positions or its conformation in the complex, respectively.

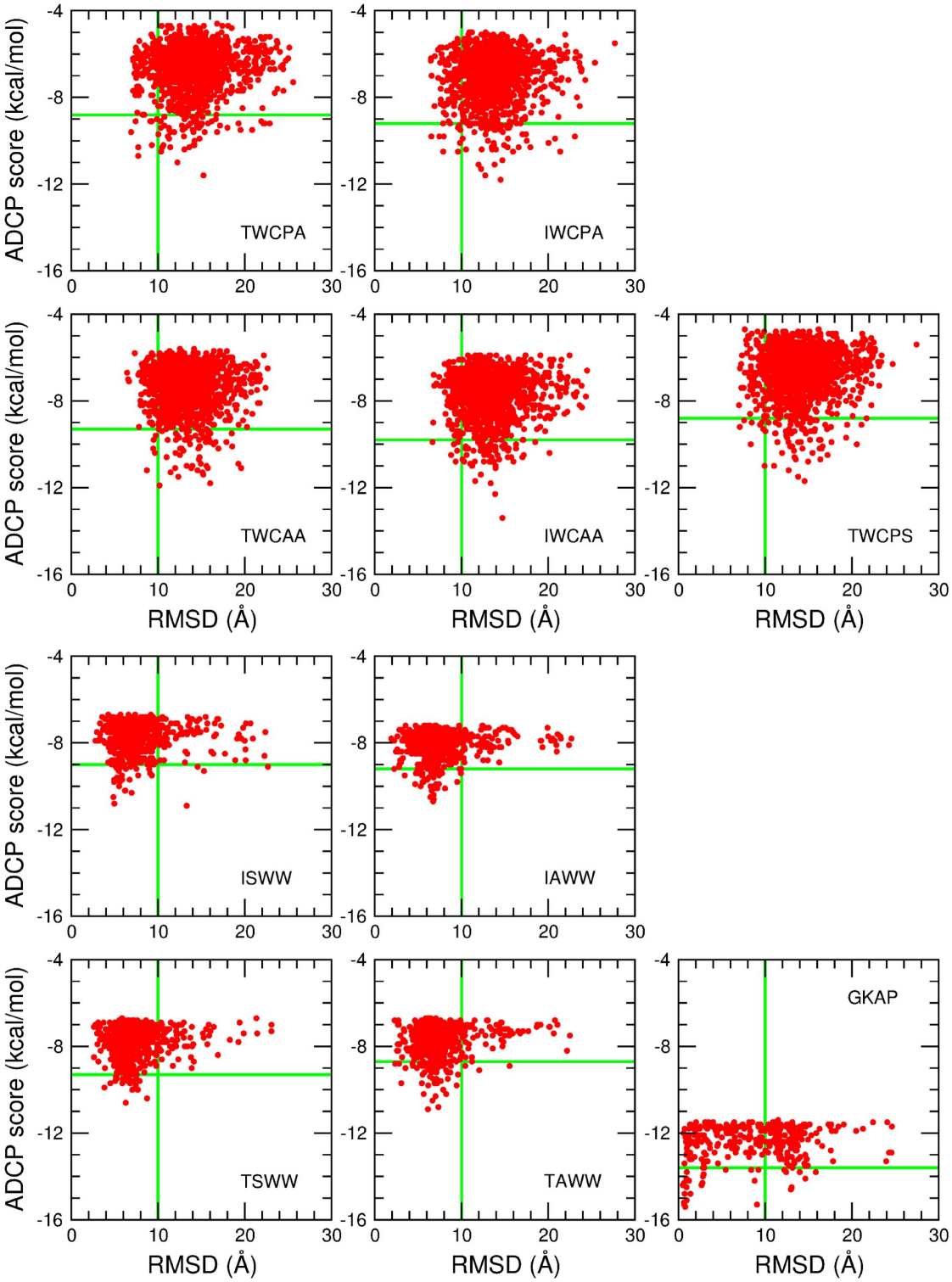

FIGURE 4.8 – ADCP score of the binding modes on PDZ domain of the 9 best Des3PI peptides (class I and II) and of GKAP peptide as a function of their Cα RMSD relative to the hotspot positions or its conformation in the complex, respectively.

4.3.3 Stability of protein-peptide complexes

For each of the cyclic peptides designed with Des3PI and docked onto either SH3 or PDZ domain, the binding mode among the 5% best scores with the lowest RMSD relative to Des3PI hotspots was selected as the initial structure for a duplicate MD simulation of 200 ns. Figures 4.9 and 4.10 depict the motions on SH3 surface of class I and class II peptides, respectively, in terms of their Cα RMSD relative to Des3PI hotspots. Over the 5 class I peptides, only FDYGF (and PDYAF to a lesser extent) remained stable in SH3 binding site during both its 200 ns simulations. This dynamic stability could be also observed for the two cyclic peptides APWDPF and GPWDPF of class II. When comparing these sequences to those of peptides which are less stable (in terms of RMSD), these results suggest that a Phe is prefereed to a Pro at Class I hotspot 1 and a Trp is preferred to a Tyr at class II hotspot 3 for peptide good stability. As a control of these numerical experiments, the linear RLP2 peptide also remained tightly attached to the SH3 binding site during its 200 ns MD simulation, pointing out that Des3PI peptides that moved away from their binding area are probably worse binders of SH3 domain than RLP2. Nevertheless, it could be noted that, during its MD simulation, RLP2 conformations deviated from its initial left-handed polyproline II helical structure observed experimentally (Figures 4.10). This stresses possible imperfections of the used force field (GAFF) for peptides which probably limit the accurate modeling of protein-peptide associations in our MD simulations.

Regarding the cyclic peptides designed to target the PDZ domain, their motions on the protein surface are illustrated in Figures 4.11 and 4.12. Among the 5 class I peptides, TWCPA and IWCPA exhibited steady RMSD in their duplicate simulations. Visual inspections of their MD trajectories confirmed that they kept their position in the PDZ binding site, contrary to the 3 other peptides of this class. The two peptides ISWW and IAWW of class II also remained attached to the targeted PDZ binding surface during their duplicate 200 ns trajectories. The two other peptides TSWW and TAWW of this class (as well as TWCAA and TWCPS of class I) are less tightly bound to the PDZ binding site, indicating that the hydrophobic residue Ile at hotspot 1 should be more appropriate than a Thr for good affinities. Comparatively, the linear GKAP peptide which has notably a Leu and a Thr located at hotspot 1 and 2 positions, respectively, has a marked stability in the dynamics simulation of its complex with PDZ (Figure 4.12).

To sum up, MD simulations of Des3PI peptides in complex with their protein target showed mixed results. While all the assessed cyclic peptides could be satisfactorily docked onto SH3 or PDZ domain, less than half of them could remain attached to the binding site in duplicate 200 ns MD simulations. Without ruling out the known limitations of atomic MD simulations (force field accuracy

and sampling exhaustiveness), we are rather seeing here some possible shortcomings of Des3PI algorithm : It generates cyclic peptides with 4 to 6 hotspot residues linked by as much or more glycines which weakly contribute to the binding energy of the protein-peptide complexes. We believe that there is room for improvement of our approach in future works.

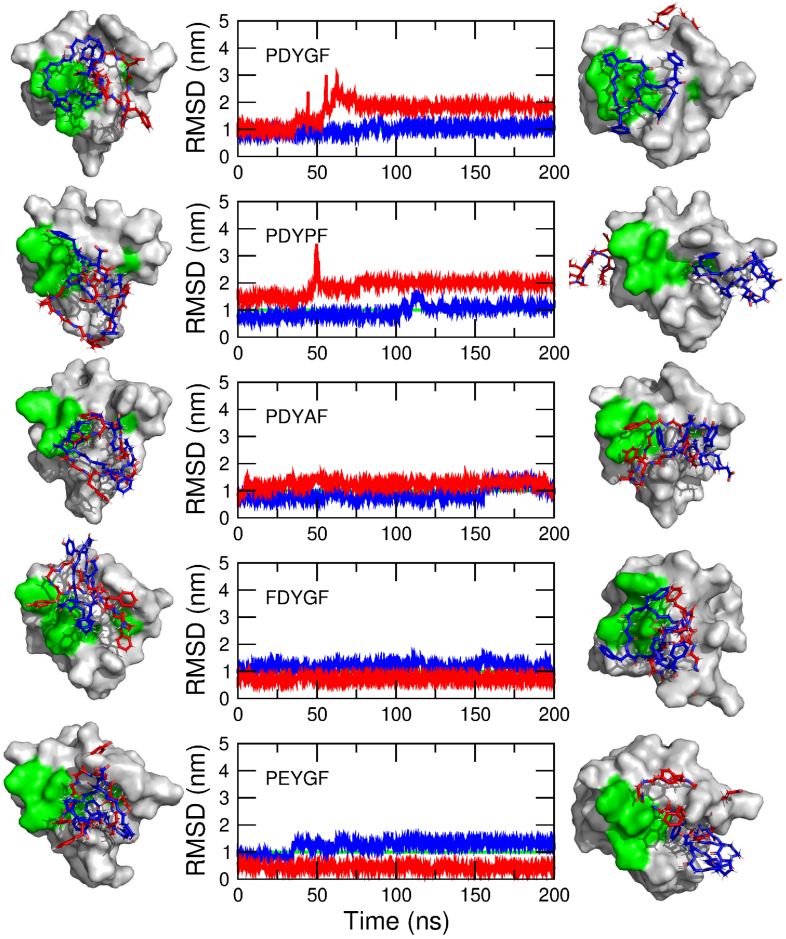

FIGURE 4.9 – Time evolution of class I peptide Cα RMSD relative to Des3PI hotspots in MD simulations of their complexes with SH3 domain. Snapshots in the left and right columns represent the complex initial and final structures, respectively. Green patches on protein surface indicate residues initially in contact with docked peptides.

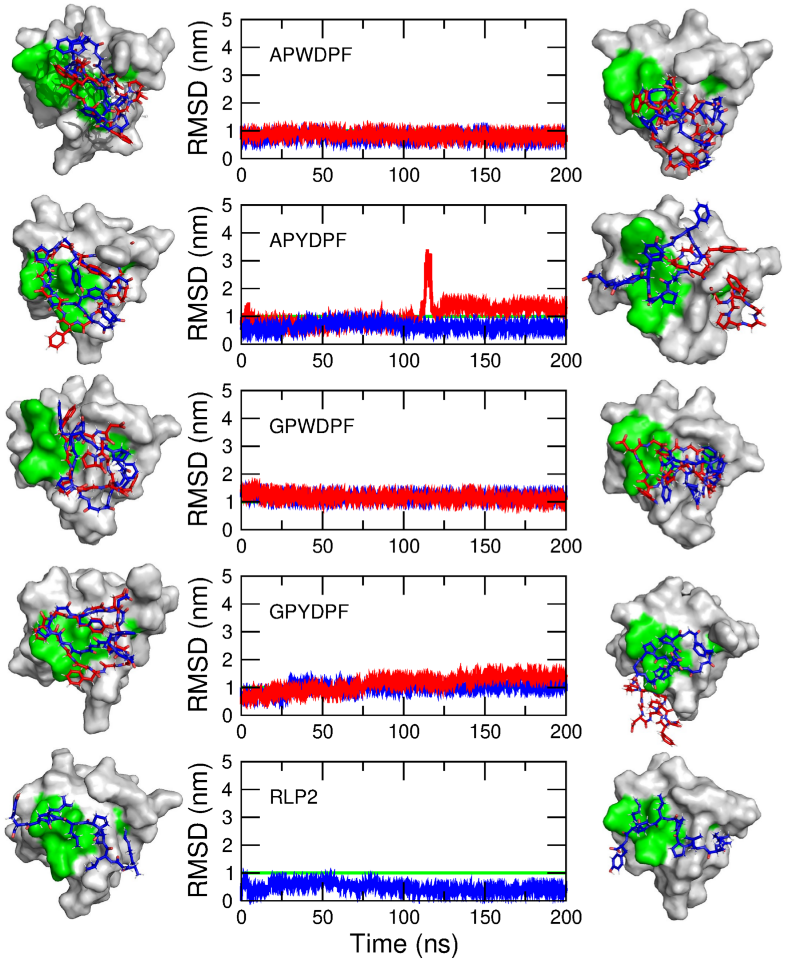

FIGURE 4.10 – Time evolution of class II peptide Cα RMSD relative to Des3PI hotspots in MD simulations of their complexes with SH3 domain. Snapshots in the left and right columns represent the complex initial and final structures, respectively. Green patches on protein surface indicate residues initially in contact with docked peptides. Bottom graph and snapshots relate to RLP2 peptide MD simulation.

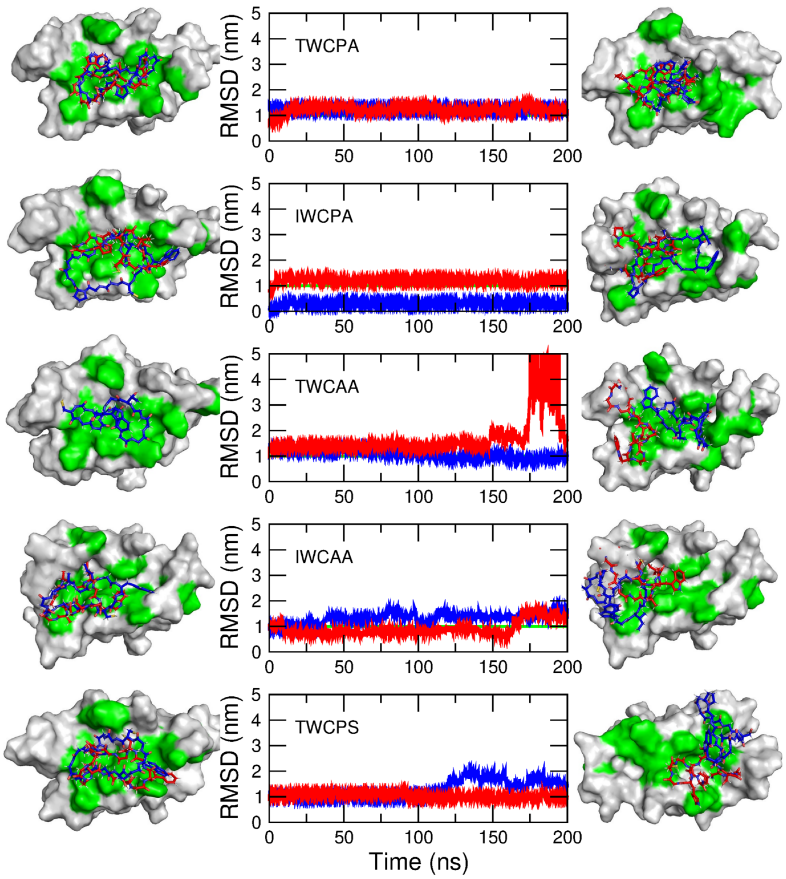

FIGURE 4.11 – Time evolution of class I peptide Cα RMSD relative to Des3PI hotspots in MD simulations of their complexes with PDZ domain. Snapshots in the left and right columns represent the complex initial and final structures, respectively. Green patches on protein surface indicate residues initially in contact with docked peptides.

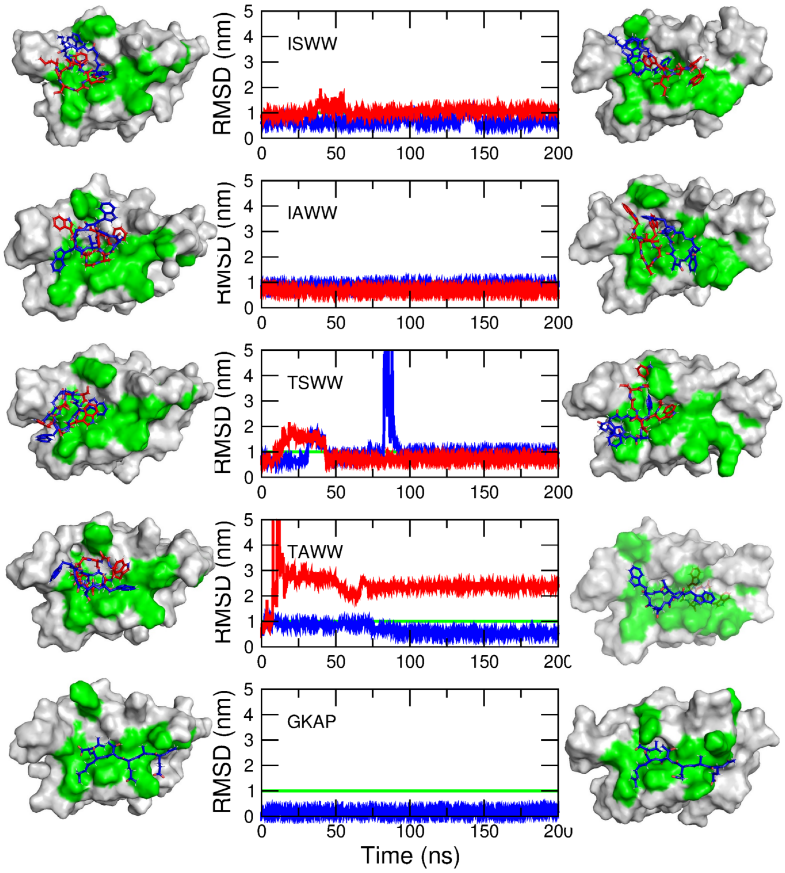

FIGURE 4.12 – Time evolution of class II peptide Cα RMSD relative to Des3PI hotspots in MD simulations of their complexes with PDZ domain. Snapshots in the left and right columns represent the complex initial and final structures, respectively. Green patches on protein surface indicate residues initially in contact with docked peptides. Bottom graph and snapshots relate to GKAP peptide MD simulation.

4.3.4 Binding free energy of protein-peptide complexes

As mentioned in the docking subsection, we intended to analyze the protein-peptide complex stable trajectories to estimate more accurately the binding free energies of Des3PI peptides versus those of RLP2 and GKAP for SH3 and PDZ domains, respectively. Table 4.1 shows the different contributions of the binding free energies computed with the MM-PBSA approach for the SH3 peptide binders FDYGF, APWDDPF, GPWDPF, and RLP2. Compared to the three Des3PI peptides which have only one negatively charged residue, the latter has two Arg residues and both its N-terminal and C-terminal ends are charged. Therefore,

the striking difference between Des3PI peptides and RLP2 lies in the electrostatic terms E_{Coul} and ΔG_{PB} which are much larger (in absolute value) for RLP2 than for the other three. However, the sum of theses two terms has close values to that one of Des3PI peptides, except of GPWDPF. For the latter, the change in electrostatic terms is lower than for RLP2, and at the end, its total binding free energy is lower than that one of RLP2.

Peptide	E_{LJ}	E_{Coul}	ΔG_{PB}	ΔG_{SA}	$-T\Delta S_{Conf}$	ΔG_{Bind}
FDYGF 1	-35 $\pm$ 6	6 $\pm$ 2	31 $\pm$ 6	-4 $\pm$ 1	3 $\pm$ 3	1 $\pm$ 9
FDYGF 2	-28 $\pm$ 3	3 $\pm$ 1	36 $\pm$ 5	-3 $\pm$ 1	14 $\pm$ 3	22 $\pm$ 9
APWDPF 1	-45 $\pm$ 6	-1 $\pm$ 3	37 $\pm$ 11	-4 $\pm$ 1	15 $\pm$ 2	2 $\pm$ 8
APWDPF 2	-43 $\pm$ 5	-3 $\pm$ 3	43 $\pm$ 9	-5 $\pm$ 1	7 $\pm$ 3	-1 $\pm$ 14
GPWDPF 1	-46 $\pm$ 5	1 $\pm$ 3	25 $\pm$ 9	-5 $\pm$ 1	7 $\pm$ 2	-18 $\pm$ 9
GPWDPF 2	-41 $\pm$ 5	6 $\pm$ 1	20 $\pm$ 6	-4 $\pm$ 1	10 $\pm$ 3	-9 $\pm$ 14
RLP2	-41 $\pm$ 6	-32 $\pm$ 3	68 $\pm$ 24	-4 $\pm$ 1	5 $\pm$ 2	-4 $\pm$ 24

TABLE 4.1 – Binding free energies of Des3PI peptides that formed stable complexes with SH3 domain. Data for RLP2 peptide are shown in the last row. All energy terms are in kcal/mol.

It is interesting to note that, unexpectedly, the entropy cost of the linear peptide RLP2 is generally lower that than that one of Des3PI cyclic ones (Table 4.1). We think that this might be related to the number of backbone dihedral angles considered in configuration entropy calculations which is lower in RLP2 (9 residues) than in the Des3PI peptides (12 to 13 residues). We also would like to draw attention to the large value of the standard deviation for RPL2 polar solvation term (ΔG_{PB}) which greatly impacts the standard deviation of its total binding free energy. We attribute this large deviation to the high flexibility of RLP2 peptide which conformational changes during its MD simulation in unbound state can significantly modify the ionic interactions between its charged chemical groups. Overall, these results indicate that Des3PI could design one cyclic peptide which would be a good competitive binders for SH3 domain.

Regarding the PDZ-peptide complexes, the energy decomposition of their binding free energies is shown in Table 4.2. Here also, the main difference between Des3PI peptides and GKAP is in the two electrostatic terms E_{Coul} and ΔG_{PB}, due to the fact that the latter is rather polar and charged (with a Glu, an Arg, and a deprotonated C-terminus) unlike Des3PI peptides which are mainly apolar. Strikingly, the variation of the electrostatic solvation free energy is about three times higher for GKAP peptide than for Des3PI ones. It is also much larger than ΔG_{PB} of RLP2 bound to SH3 domain. These data can be explained by the fact that the three GKAP charged chemical groups are involved in ionic and/or

Peptide	E_{LJ}	E_{Coul}	ΔG_{PB}	ΔG_{SA}	$-T\Delta S_{Conf}$	ΔG_{Bind}
TWCPA 1	-47 $\pm$ 4	-3 $\pm$ 1	25 $\pm$ 6	-5 $\pm$ 1	10 $\pm$ 5	-20 $\pm$ 10.0
TWCPA 2	-47 $\pm$ 6	-5 $\pm$ 1	30 $\pm$ 7	-5 $\pm$ 1	6 $\pm$ 4	-21 $\pm$ 10
IWCPA 1	-51 $\pm$ 7	6 $\pm$ 2	32 $\pm$ 6	-5 $\pm$ 1	9 $\pm$ 4	-9 $\pm$ 10.0
IWCPA 2	-60 $\pm$ 5	-8 $\pm$ 2	45 $\pm$ 6	-6 $\pm$ 1	7 $\pm$ 4	-22 $\pm$ 10.0
ISWW 1	-41 $\pm$ 7	-1 $\pm$ 1	28 $\pm$ 5	-4 $\pm$ 1	7 $\pm$ 1	-11 $\pm$ 7
ISWW 2	-37 $\pm$ 5	-4 $\pm$ 1	33 $\pm$ 9	-4 $\pm$ 1	0 $\pm$ 1	-12 $\pm$ 11
IAWW 1	-40 $\pm$ 3	-2 $\pm$ 1	18 $\pm$ 4	-4 $\pm$ 1	15 $\pm$ 1	-13 $\pm$ 5
IAWW 2	-50 $\pm$ 4	-5 $\pm$ 1	29 $\pm$ 4	-5 $\pm$ 1	14 $\pm$ 1	-17 $\pm$ 7
GKAP	-54 $\pm$ 5	-31 $\pm$ 4	96 $\pm$ 17	-5 $\pm$ 1	9 $\pm$ 1	15 $\pm$ 16

TABLE 4.2 – Binding free energies of Des3PI peptides that formed stable complexes with PDZ domain. Data for GKAP peptide are shown in the last row. All energy terms are in kcal/mol.

hydrogen bonds with PDZ (PDB ID : 5OVC) whereas only one Arg of RLP2 is making a non-covalent interactions with SH3, its other charged groups remaining in the solvent (PDB ID : 1RLQ). Therefore, the desolvation free energy of GKAP is much larger than those of RPL2 and Des3PI peptides, and accounts for its less favorable total binding free energy compared to the other peptide binders. These results suggest that Des3PI cyclic peptides might be good binders and potentially good inhibitors of PDZ domain.

4.4 Conclusions

Many protein-protein complexes are mediated by the association of a folded domain with a peptide segment. Generally, these protein-peptide interactions are highly specific but have low affinity, giving the opportunity to design competitive inhibitors of these PPIs. In this report, we evaluated the ability of our program Des3PI to design cyclic peptides which could potentially inhibit 4 known protein-peptide complexes : the αVβ3 integrin-RGD ligand (PDB ID : 1L5G) [306], the CXCR4 receptor-CVX15 antagonist (PDB ID : 3OE0) [302], the SH3 domain-RLP2 proline-rich peptide (PDB ID : 1RLQ) [82], and the PDZ-GKAP C-terminal peptide (PDB ID : 5OVC) [220].

Regarding the two first systems, Des3PI generated peptides with hotspots that are very different, in nature and in position, from the cyclic peptides RGD and CVX15 co-crystallized with αVβ3 and CXCR4, respectively. These discrepancies might result from the treatment of ionic interactions by Autodock Vina, the threshold for clustering the amino acid binding modes into hotspots, and/or the

particular shape of the two binding sites. Deeper investigations of these two cases are certainly needed, but, given the obtained peptide sequences which were not particularly promising, we did not further study their interactions with their targeted proteins.

More encouraging results were obtained on SH3 and PDZ systems. First, Des3PI could find amino acid hotspots which were roughly aligned like in the short linear peptides RLP2 and GKAP that bind the SH3 and PDZ domains, respectively. This result is probably due to the linear groove shape of the two binding sites. Secondly, Des3PI found amino acids with similar nature at similar positions to those of RLP2 and GKAP peptides, suggesting that Des3PI peptides could compete with them for binding the targeted proteins. The blind docking of the most promising cyclic peptides on SH3 or PDZ domains allowed to retrieve among the 5% best scores at least one binding mode close to the hotspots identified by Des3PI. Nevertheless, over the 18 assessed cyclic peptides, 7 of them could steadily remain in the binding site of their target in duplicate 200 ns MD simulations. When computing their binding free energies with the MM-PBSA method, we found that one peptide (GPWDPF) had a better affinity for SH3 domain than RLP2, and that the four simulated ones (TWCPA, IWCPA, ISWW, and IAWW) had lower binding free energies for PDZ than GKAP.

Overall, despite some failings, Des3PI appears as a promising tool for "fishing in the ocean" of peptide sequences those with potential good affinities for a targeted protein surface. A future study on the treatment of ionic interactions in protein-ligand docking program would be probably needed to rectify the observed under-representation of charged amino acids in the output sequences. Another path to improvement would be to refine the clustering of amino acid binding modes (step B of Figure 4.1) to generate cyclic peptides with more hotspots and less glycine linkers and enhance their affinity and specificity for their targets.

Chapitre 5

Enrichissement de la librairie d'acides aminés de Des3PI par des acides aminés fluorés.

L'étude précédente sur l'influence de la forme de la cavité d'une protéine cible sur les résultats de Des3PI a montré que le programme était capable de proposer des séquences peptidiques quelque soit le système étudié. Cependant, nous ne retrouvons aucun résidu en commun par rapport aux peptides cocristallisés avec CXCR4 et l'intégrine αVβ3. Pour les domaines SH3 et PDZ, les peptides générés par Des3PI possédaient plusieurs résidus de nature similaire à ceux des SLiMs cocristallisés. De plus, des calculs d'énergie libre de liaison par MM-PBSA a montré qu'un peptides générés par Des3PI ciblant SH3 semblait légèrement plus affins que le peptide cocristallisé. Dans le cas du domaine PDZ, les peptides cycliques conçus *in silico* avaient des affinités significativement meilleures que le peptide cocristallisé.

Ainsi, bien que les résultats soient encourageants, nous souhaitions améliorer les performances de notre logiciel afin de proposer des peptides possédant des énergies libres de liaison encore plus favorables. Pour cela, nous proposons donc d'enrichir la librairie de Des3PI avec divers acides aminés fluorés. Ceux-ci ont été choisis pour leur capacité à augmenter l'affinité d'un peptide pour sa cible et leurs propriétés pharmacocinétiques. De plus, l'équipe FLUOPEPIT à laquelle nous appartenons est spécialisée dans la synthèse de peptides fluorés. Ainsi, ajouter ces acides aminés fluorés ouvrerait la voie à une possible collaboration expérimentale avec les chimistes médicinaux de notre équipe.

L'objectif de ce chapitre était donc de savoir, si une fois incorporé à la librairie, des acides aminés fluorés allaient ressortir dans les différentes séquences proposées, et si ces séquences avaient de meilleures affinités que leurs homologues naturels non fluorés. Ce chapitre est l'adaptation d'un manuscrit en cours

de rédaction qui sera soumis très prochainement.

5.1 Introduction

Although fluorine has been used since the end of the 19th century in classical organic chemistry, it only appeared in the field of medicinal chemistry in the middle of the last century. Since then, it has become an essential element in modern strategies of drug design development. It is estimated that 20 to 30$\%$ of the drugs currently on the market or in clinical trials are fluorinated compounds [115, 199]. Currently, we seem to be on a stable trend of 9 to 11 fluorinated molecules approved by the FDA each year [191, 105].

The reason why fluorine is so important in drug design is that it has many chemical properties that imrpove the pharmacodynamics and pharmacokinetics properties of drug candidate. Indeed, fluorine can greatly increase the thermal and chemical stability of a molecule. It is also known to increase the hydrophobicity of these molecules and thus, to enhance their cellular permeability [47, 28, 192, 94, 138, 198].

In parallel, pharmaceutical sciences have seen a growing interest in therapeutic peptides since the second half of the 20th century. These compounds are indeed very much studied for their high affinity and specificity for their protein targets. Moreover, the emergence of high throughput discovery methods such as phage display has made the search for therapeutic peptides more accessible [201].

Naturally, multiple studies have been conducted to combine the advantages of these two branches of medicinal chemistry by incorporating fluorinated amino acids into peptide sequences. These amino acids bring chemical stability and improve the peptide hydrophobicity that is necessary to cross biological barriers. They also strongly stabilize the fluorinated peptide secondary structures, especially in the case of α-helix and β-sheet [28]. Finally, the fluorination of peptides significantly increases their half-life by preventing their recognition by blood proteases [190, 199, 46] and their metabolic stability [115] (with the exception of aromatic fluorinated amino acids).

Fluorinated peptides are nowadays increasingly used in therapeutics. A review by Mei *et al.* details about forty of these compounds currently on the market or in clinical trials, including, as an example, the glecaprevir used in the treatment of hepatitis C [190]. These peptides are used in many therapeutic areas such as bacteriology, virology, the treatment of certain metabolic diseases or cancers. Interestingly, these different pathologies are very of them caused by a deregulation of protein-protein interactions which are the molecular basis of cell signaling and regulation processes [190, 3, 116, 184, 192, 57, 223]. Indeed, fluo-

rinated peptides appear optimal to target protein-protein interfaces which are generally composed of a large, shallow and hydrophobic cavity. For that reason, the development of small chemical compounds targeting PPIs has been a relative failure due to their lack of high affinity for such interfaces [201]. In contrast, peptides have a more suitable size which allows to bind PPIs with high affinity and specificity. And the incorporation of fluorinated amino acids additionally allow to overcome some pharmacokinetic limits of peptides due to the properties mentioned above [201].

In this context, We propose here an upgrading of Des3PI , our computational fragment-based approach for the development of cyclic peptides inhibiting protein-protein interactions [65], by incorporing int its library some fluorinated amino-acids. Designing cyclic peptides was a first answer to increase their cell permeability and metabolic stability [201]. By incorporating fluorinated amino acids into Des3PI library, we intend to propose a method that would generate fluorinated cyclic peptides with improved affinities for a given protein-protein interface when compared to their non-fluorinated counterparts.

We present in this report the results of using Des3PI with a library enriched with 16 fluorinated amino acids on two previously studied systems, Mcl-1 and Ras and compare the results of the non-fluorinated version of Des3PI with the new one.

5.2 Methods

5.2.1 Library enrichment

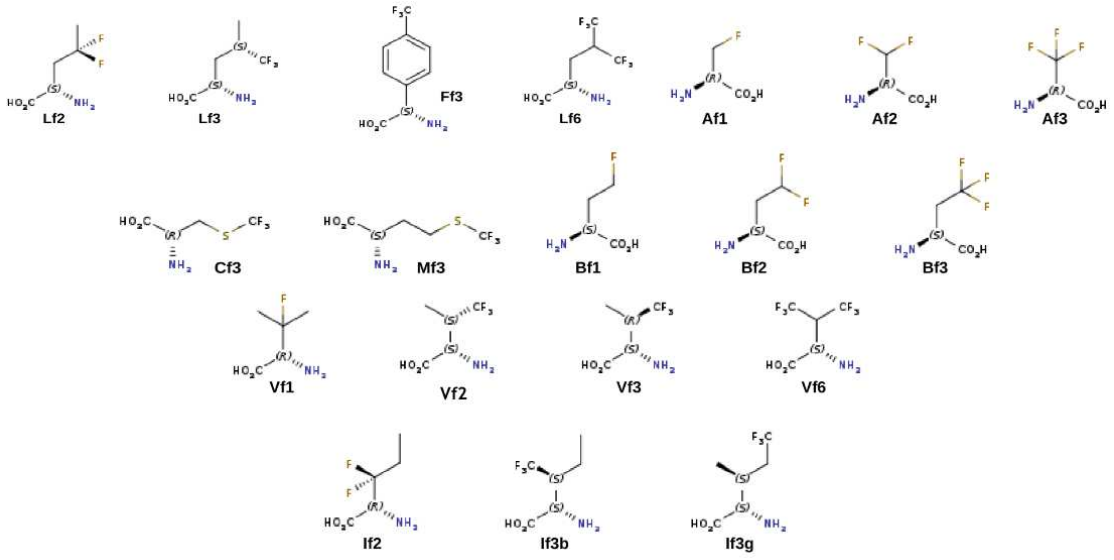

FIGURE 5.1 – Fluorinated amino acids added in the library of amino acides of Des3PI. The first letter of their names refers to the natural amino acid they are based on (except B_{f1}, B_{f2} and B_{f3} which denotes).

In this study, the Des3PI library of 20 amino acids has been enriched with 16 fluorinated amino acids (figure 5.1). These residues were chosen for their ease of synthesis, so it would be easier to initiate further experimental validations [199]. A three dimensional structure of these amino acids was built using MarvinSketch 6.2.1 [37]. Their AM1-BCC charges were computed [120] and their conformation were shortly minimized using 5000 steepest descent steps and the Generalized Amber Force Field (GAFF) [290].

5.2.2 Des3PI's method

As Des3PI algorithm was fully described in a previous work, we only provide here a summary of the method (figure 5.2). Des3PI docks 50 times each amino acid of the library on a targeted protein surface. The binding modes are then clustered and the most represented amino acid in each cluster defines the peptide hotspots. Glycine linkers are used to connect the hotspots and generated

the full cyclic peptide sequence. These steps are repeated 20 times and peptides with identical geometry are regrouped in classes. The five best sequences of each class are determined according to the amino acid occurrence at each hotspot of each class.

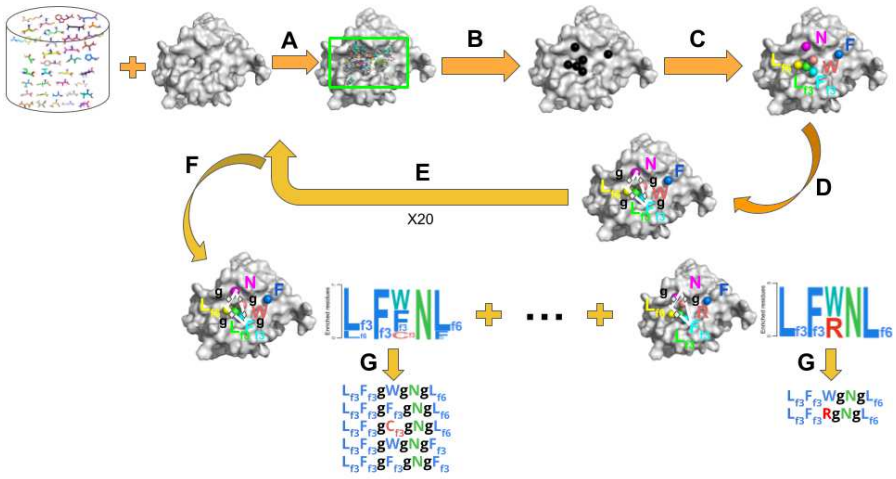

FIGURE 5.2 – Des3PI workflow. (A) Each amino acid (natural and fluorinated) is docked 50 times. (B) All the binding modes are clustered. (C) The most represented amino acid is determined in each cluster to compute the hotspots. (D) The hotspots are linked using one or several glycines. (E) Steps A,B,C, and D are repeated 20 times to generate multiple sequences. (F) Peptides with the same number of hotspots and a similar geometry are grouped into classes of peptides. (G) For each class, the amino acid occurrences and the five best sequences are determined. Lowercase **g** indicates glycine linkers.

5.2.3 Peptide docking on protein target

Unlike natural peptides, the fluorinated peptides designed by our new version of Des3PI cannot be docked by using protein-peptide docking methods such as Autodock CrankPep (ADCP) because fluorinated amino acids parameters are not implemented [324]. Thus, we decided to use Autodock VINA for the blind docking step although it is not able to fully take into account the whole flexibility of a peptide. While ADCP could generate 3D structure of peptide from sequence, this time we had to generate a three-dimensional structure of our fluorinated peptides before docking. For that, the peptides sequences structures were written in SMILES and then converted into PDB using OpenBabel. The peptides were then charged using the mmff94 method and a minimization

was performed with a 10 000 steps Steepest Descent algorithm.

Then, we performed the blind docking of each peptide 10 times. Their binding modes were analyzed by computing the root-mean-square deviation (RMSD) of the α-carbons relative to the hotspots previously generated by Des3PI. Finally, we checked if there is at least one binding mode among the 5% best VINA scores on the targeted protein surface with a RMSD relative to Des3PI hotspots lower to 10 Å.

5.2.4 MD simulations of protein-peptide complexes

For each selected protein-peptide complex, we performed two independant 200 ns MD simulations in explicit water. The AMBER99SB-ILDN [275] and GAFF [290] force fields were used for the protein and peptide, respectively. Each complex was placed in a cubic box, so that the minimal distance between the solute and the cube faces was equal to 1 nm. Then the complex was solvated with TIP3P water molecules and neutralized with 0.15 mol/L of sodium chloride. The Lennard-Jones potentials were cut off at 1.2 nm and the Coulomb interactions were treated using the smooth PME method [78]. Each system was first minimized using 10 000 steps of the steepest descent method, then submitted to two short equilibration runs of 1 ns each, the first one to heat the system to 310 K using a Berendsen thermostat and the second one to equilibrate the pressure around 1 bar using the Parinello-Rahman method. After that, a 200 ns production run was performed in the isothermal–isobaric (NPT) ensemble using the Nose-Hoover and Parrinello-Rahman coupling algorithms [207, 109, 212] with the time constants $\tau_T = 0.5$ ps and $\tau_P = 2.5$ ps. The Newton's equations of motion were integrated using the leap-frog algorithm with a time step of 2 fs, while keeping constant the length of all covalent bonds using the LINCS procedure [107]. MD trajectory frames were saved every 20 ps for subsequent analysis.

5.2.5 Calculation of the protein-peptide binding free energy

Finally, we estimated the binding free energy of cyclic peptides that can form stable complexes with their targeted protein, and in comparing it with the binding free energy of the reference co-crystallized peptide segments. These free energy calculations were based on the molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) approach :

$$\Delta G_{Bind} = \Delta E_{MM} + \Delta G_{Solv} - T\Delta S_{Conf} \tag{5.1}$$

where ΔE_{MM} is the change in the mean potential energy of the protein-ligand association in vacuum. $\Delta G_{Solv} = \Delta G_{PB} + \Delta G_{SA}$ is the variation of the solvation

free energy which polar and apolar contributions can be calculated by using the Poisson-Boltzmann (PB) equation and the molecular solvent-accessible surface areas (SA), respectively. Lastly, $T\Delta S_{Conf}$ is the configurational entropy change upon the complex formation in gas phase.

The details of the method used to compute the 3 terms of Equation S1 are described in our previous study (Chapter 4). We would just like to mention that to take into account the contribution of peptide conformational changes in δG_{solv} and δS_{conf} terms, we additionally ran an independant 200ns MD simulation of the unbound peptide, and extracted from the last 100ns trajectory its solvation free energy and conformational entropy (Chapter 4).

5.3 Results and Discussion

5.3.1 Peptides generated with the new library of Des3PI

In this section, we will describe the results generated by the new Des3PI library and will compare them to the results published in a previous study (Chapter 3).

Thus, we applied Des3PI on the cavity of Mcl-1 interacting with PUMA (PDB : 2ROC [62]) in the same way as in our previous work. This generated 5 different classes of peptides (figure 5.3), two of them have 4 hotspots and 3 of them have 5 hotspots, compared to only one class with 5 hotspots obtained with the previous version of Des3PI (Chapter 3). Clearly, the addition of new amino acids has greatly increased the diversity in the number of sequences and conformations proposed by Des3PI. In terms of occurrences, the new classes I, IV and V appeared 5,7 and 5 times over 20 while the class II and III were only found 2 and 1 times, respectively.

Regarding the general position of the hotspots, we can first note that Des3PI places them mostly at the center-left of the Mcl-1 cavity (figure 5.3 and still do not cover the entire cavity. Thus, this version is able to propose 11 sequences instead of 5 even though there is not enough diversity in 4 classes to compute 5 sequences for them. To describe the hotspots positions, we use class IV numbering. The positions of the hotspots 1, 2 and 4 are found in all the new peptide classes. These particularly conserved hotspots thus appear to be areas of high docked fragment density. In contrast, the presence of hotspot 3, 5 and 6 is then relatively variable. Moreover, we can observe that the newly generated class IV is identical in its configuration to the class I generated by the previous version (figure 3.3, Chapter 3). This class is still obtained by excluding cluster 6 which prevented the generation of a stable cyclic conformation. In contrast, hotspot 6 can be easily inserted in a cyclic peptide and thus, was conserved in peptide of

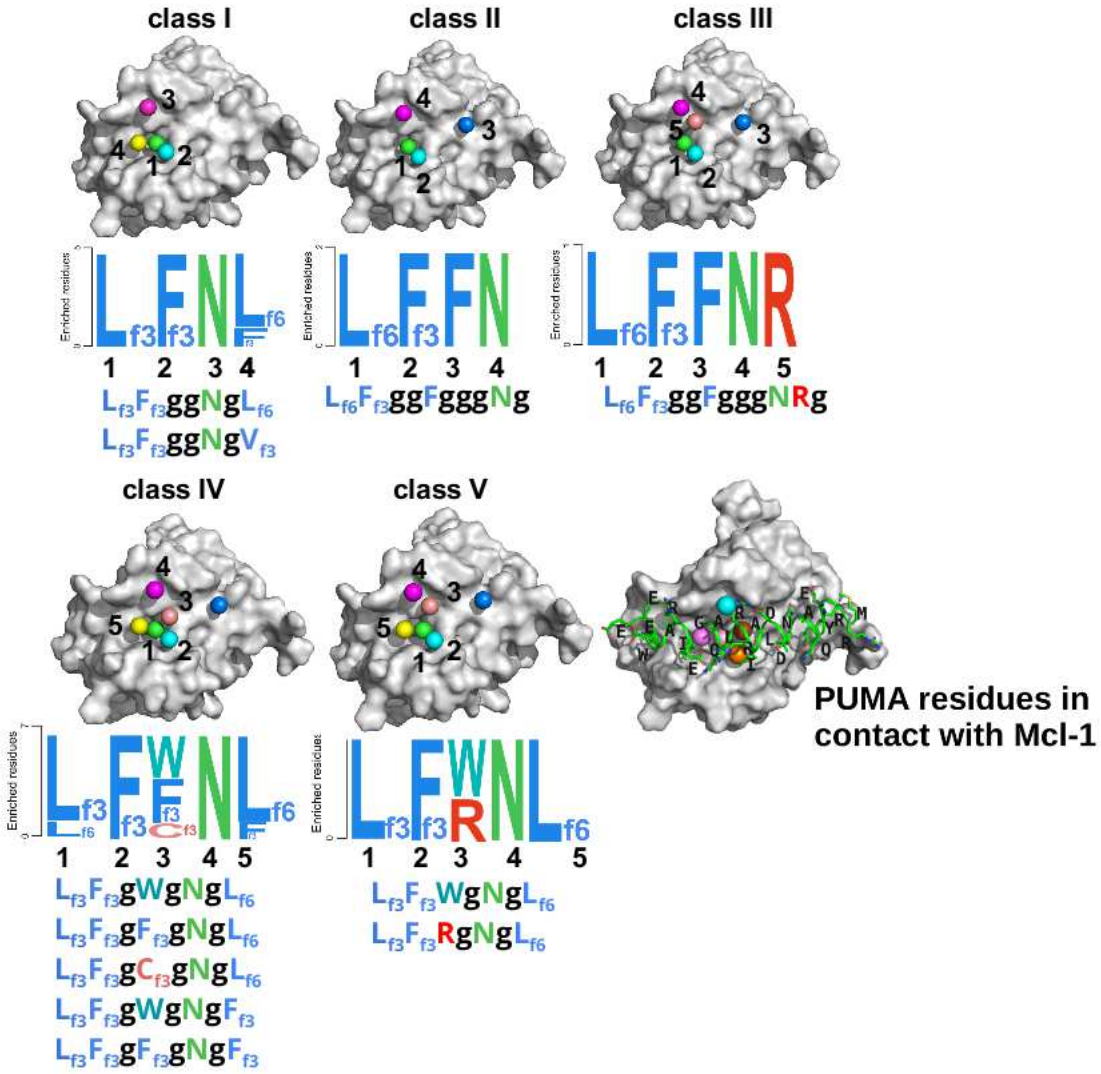

FIGURE 5.3 – Results generated by Des3PI targeting Mcl-1 with the enriched library.

class II and III. Note that the difference in geometry between class IV and V is due to a slight shift in hotspot 3 toward hotspot 4.

Then, concerning the nature of the amino acids found by Des3PI, we observe at first sight that fluorinated amino acids are widely represented in most of the hotspots of all the calculated classes. However, the asparagine of the hotspot 4 at the top of the cavity is found without any alternative in all classes. It is positioned in place of the Asp of PUMA interacting with Arg263 of Mcl-1. Although we still do not find any ionic interaction at this location, we still have the possibility

to form a hydrogen bond at this location through this Asparagine. Most of the other hotspots, are located in Mcl-1 hydrophobic pockets and are rationally hydrophobic amino acids. If we compare with the old Des3PI results, the isoleucine strongly represented is mostly substituted by a fluorinated Leucine (L_{f3} or L_{f6}) in cluster 1 of the new version. The cluster 2 of the new version is entirely represented by a triply fluorinated phenylalanine (F_{f3}). This cluster being previously represented by a Phenylalanine or a Tryptophan (cluster 3), we keep an aromatic amino acid at this position. On the contrary, the current cluster 5 is mostly represented by the residue L_{f6} whereas it was previously occupied by a tryptophan. Although the aromatic character of the amino acid is lost, the L_{f6} residue is still very hydrophobic so the main nature of the residue is kept. Finally, it is interesting to note that the cluster 3 in class V (and 5 in class III), not existing in the previous application, generated an Arginine. This may echo the lysine found in the previous version, not far from it. Finally, we noticed that Des3PI gives two identical sequences in terms of hotspots ($L_{f3}F_{f3}WNL_{f6}$ of class IV and V) but they have different conformation and linkers so we will treat them as different.

Thus, we can conclude that the use of the fluorinated library on Mcl-1 gives results more diverse than the version in terms of conformation. Moreover, a large number of hydrophobic residues are substituted by more hydrophobic fluorinated analogues when it exists.

The targeting of Ras with the fluorinated library (figure 5.4) also led to five classes of peptides instead of four with the previous version. Nevertheless, the current version computed only 14 different peptides sequences instead of 20 because of the low occurrences of class I, II and V.

In contrast, classes III and IV are particularly well represented since they appear in 7 and 9 of the 20 runs, respectively. Note that two of the four geometries of the previous version are conserved with this new calculation (class III and class V, figure 3.4) and that many hotspots are common. Indeed, the position of hotspot 1, 3, 4 and 6 (class IV is taken as reference) generated by the application of the fluorinated library are particularly well preserved through the various classes. Moreover, they have also the same position as in the various classes of the old version, which makes it possible to compare their nature.

Concerning this nature, we notice that the nature of the hotspots already present in our previous work are retrieved with new version. For example, we find the aspartate of the previous classes in the hotspot 1 of class III IV of fluorinated Des3PI. Arginine from the westernmost cluster of the cavity is also found as well as alanine and glycine from the northernmost cluster, although fluorinated analogues of alanine exist in the library. Similarly, the valine found in the southeast cluster is found also unfluorinated in the new application. Surprisingly, cluster 1 (green) found in all classes is often a hydrophobic fluorinated amino

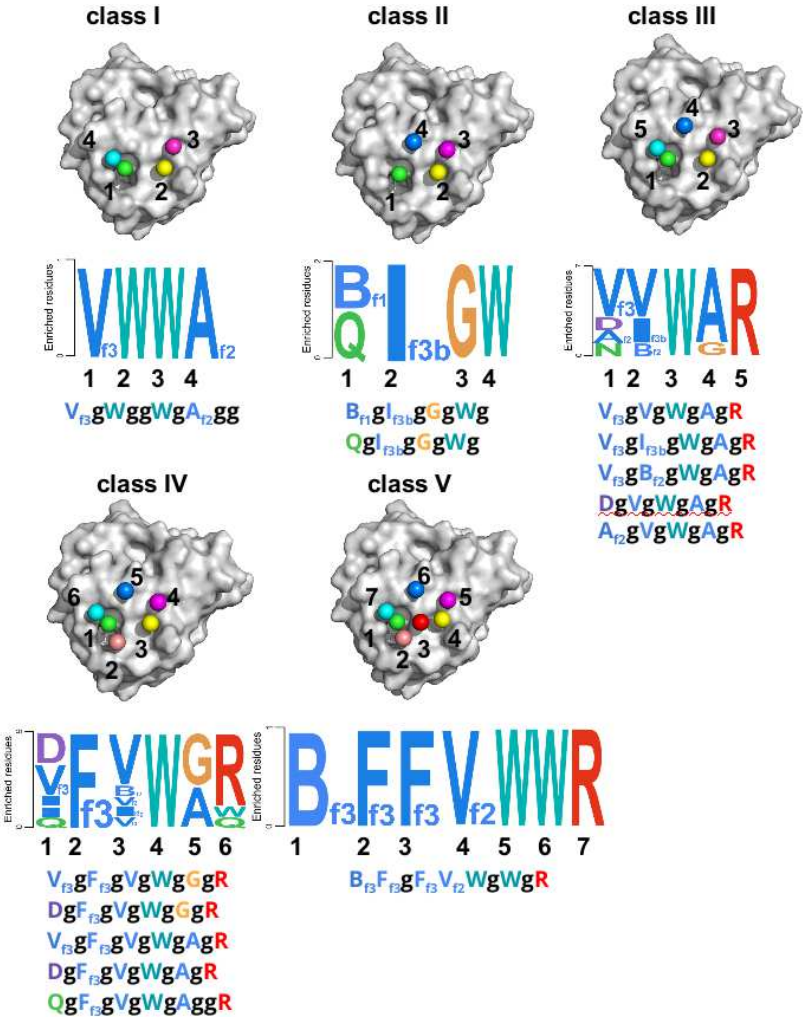

FIGURE 5.4 – Results generated by Des3PI targeting Ras with the enriched library.

acid (V_{f3}, B_{f1} or B_{f3}) whereas it was generally polar or even charged previously. This might be due to the fact that hotspot 1 is located in a small polar cavity which can accomodate a thin polar side chain but not big fluorinated side chain which prefer to be located on the edge of this cavity.

Thus, in the case of Ras, the amino acids found seem to be less substituted by fluorinated analogues than in the case of Mcl-1, with even the presence of a completely non-fluorinated sequence still proposed (DVWAR, class IV).

5.3.2 Docking of the peptides generated by Des3PI

As in our previous work, we wanted to validate *in silico* our peptide sequences. However, as mentioned in the method section, we were confronted with the fact that the step of blind redocking was not feasible with Autodock CrankPep due to the addition of fluorinated amino acids. To our knowledge, there is currently no peptide-protein docking program that can handle our fluorinated peptides. We therefore decided to try to redock our peptides using Autodock VINA, keeping in mind that its algorithm is not dedicated to the docking of flexible peptides. However, we wanted to know if it was possible to generate different conformations close to those of our hotspots to perform a molecular dynamics simulation that would give more accurate results on the stability of our peptides.

In addition, as one of the objectives of this work was to find out if the fluorinated peptides generated by Des3PI formed more stable complexes than their non-fluorinated analogues, we also performed blind docking of the non-fluorinated equivalent of the peptides that passed this step for further explorations. These non-fluorinated analogues were generated using the same method as their fluorinated counterparts.

For peptides targeting Mcl-1 (figure 5.5), we observed 4 peptides out of 11 that successfully pass the blind docking validation : L$_{f7}$F$_{f3}$FN, L$_{f6}$F$_{f3}$FNR L$_{f3}$F$_{f3}$WNL$_{f6}$ and L$_{f3}$F$_{f3}$NL$_{f6}$. The non-fluorinated analogs of the two last one also success to reach Des3PI hotspots position. Among these four fluorinated peptides, those from the class IV show a lower binding energy. Moreover, this energy appears to be in a same range as those of their non-fluorinated analogs. With these four peptides with a binding mode among the top 5% of the scores with an RMSD lower than 10$\mathrm{\AA}$, the success rate of 4 over 11 (roughly 36%) appears to be lower than the results of our previous work with the natural peptides and ADCP (3 over 5, about 60%).

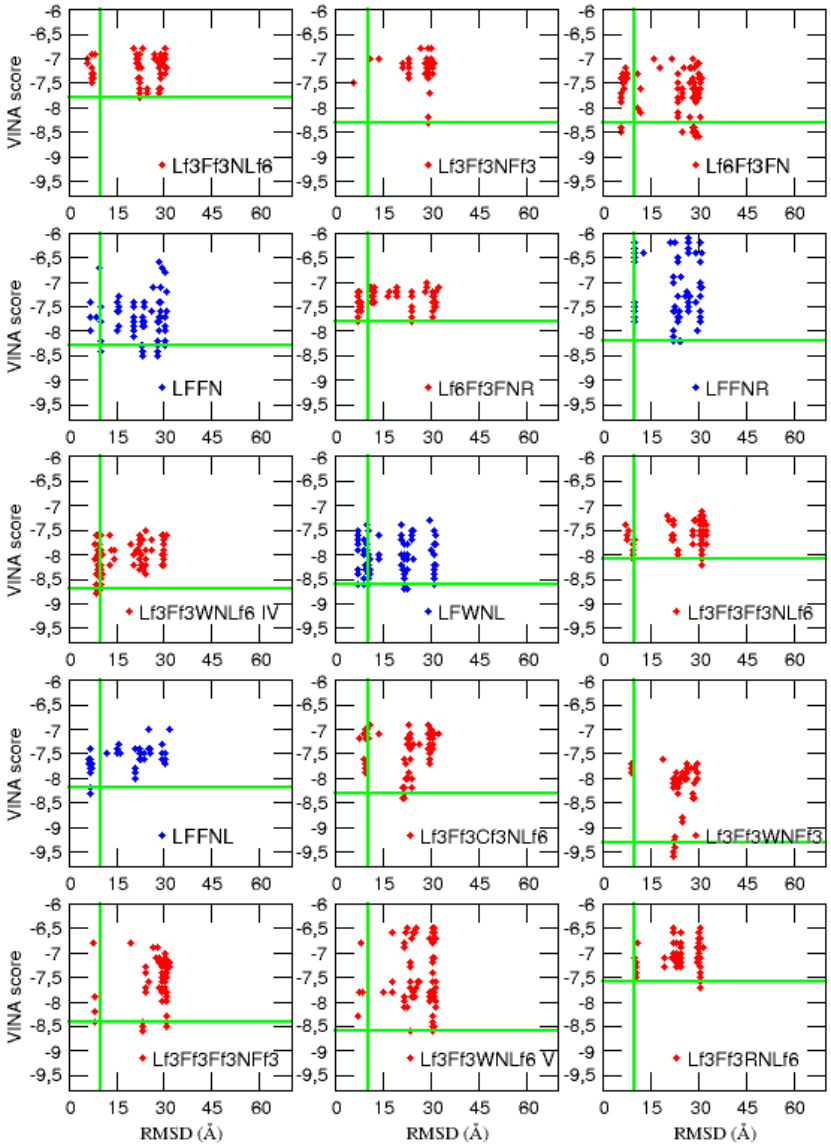

FIGURE 5.5 – VINA score (in kcal/mol) of the binding modes on Mcl-1 of the Des3PI fluorinated peptides (red) and the non fluorinated analogues (blue) of those successfully pass the validation step (LFFN, LFFNR, LFWNL and LFFNL) as a function of their Cα RMSD relative to the hotspot positions.

Nevertheless, we have to keep in mind that VINA is probably less efficient than ADCP to deal with flexible peptides. However, these results allow us to test

the stability of different complexes.

We docked the 13 fluorinated peptides and the non fluorinated peptide designed by the new Des3PI's version on Ras protein and found that 4 of them successfully passed this step (figure 5.6). To these 4 peptides, we also docked their non fluorinated analogues (QIGW, AVWAR, LFFVWWR) or fluorinated analog for DVWAR (DVf6WAR). Interestingly, we observed that the 4 peptides designed by Des3PI have lower binding free energy than their counterparts. Although the success rate is lower than in our previous study, this still allows us to perform stability tests later.

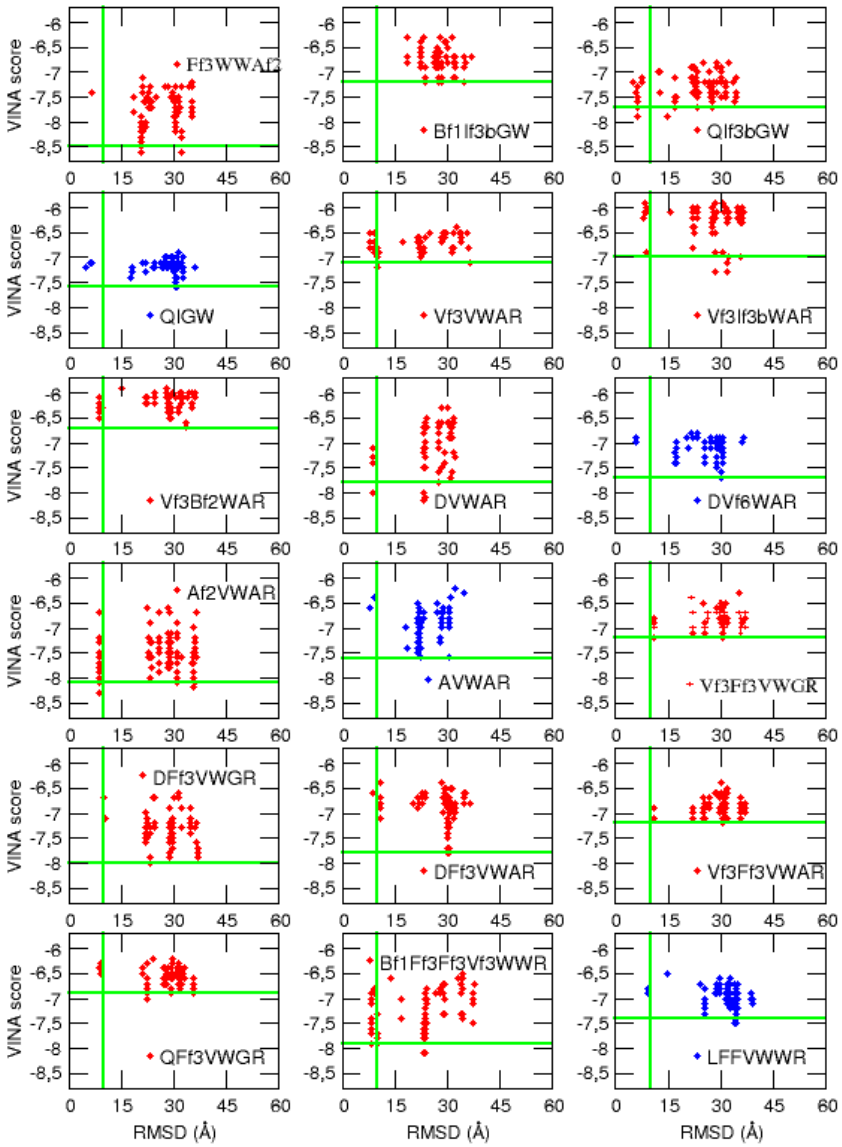

FIGURE 5.6 – VINA score (in kcal/mol) of the binding modes on Ras of the Des3PI peptides (red) and the analogues (blue) of those successfully pass the validation step (QIGW, DVf6WAR, AVWAR and LFFVWWR) as a function of their Cα RMSD relative to the hotspot positions.

5.3.3 Evaluating stability of the generated complexes by MD simulations

We have seen previously that for each of the 2 systems, 4 peptides and their analogues had passed the blind docking step. Due to limited computational resources for this project, we launched two independant simulations for only two peptides and their two analogues for each of the targets.

In the case of the analogues of the peptides generated by Des3PI, we chose, when possible, to respect the validation criteria of the previous step to choose the complex to simulate. When it was not possible, we chose the pose with the lowest RMSD, and thus the conformation closest to the one calculated by Des3PI.

For Mcl-1, we observe that all simulated peptides remain relatively stable within the protein cavity, with RMSD varying around 1 nm relative to Des3PI hotspots. Most of the peptides appear to remain strongly interacting with the larger hydrophobic pocket of Mcl-1 where the Des3PI hotspots were located. Moreover, we notice that the peptide LFFN (non-fluorinated analogue of Lf6Ff3FN), initially located at the center-right of the cavity, translates to a central site, much closer to the conformation obtained by Des3PI. This indicates us that the binding mode proposed by Autodock VINA did not seems to be the most optimal. Given the stability of all complexes, we therefore retained the full set of simulations for MM-PBSA analysis of the last 100 nanoseconds of each simulation to see if the fluorinated peptides had a better association free energy than their fluorinated counterparts.

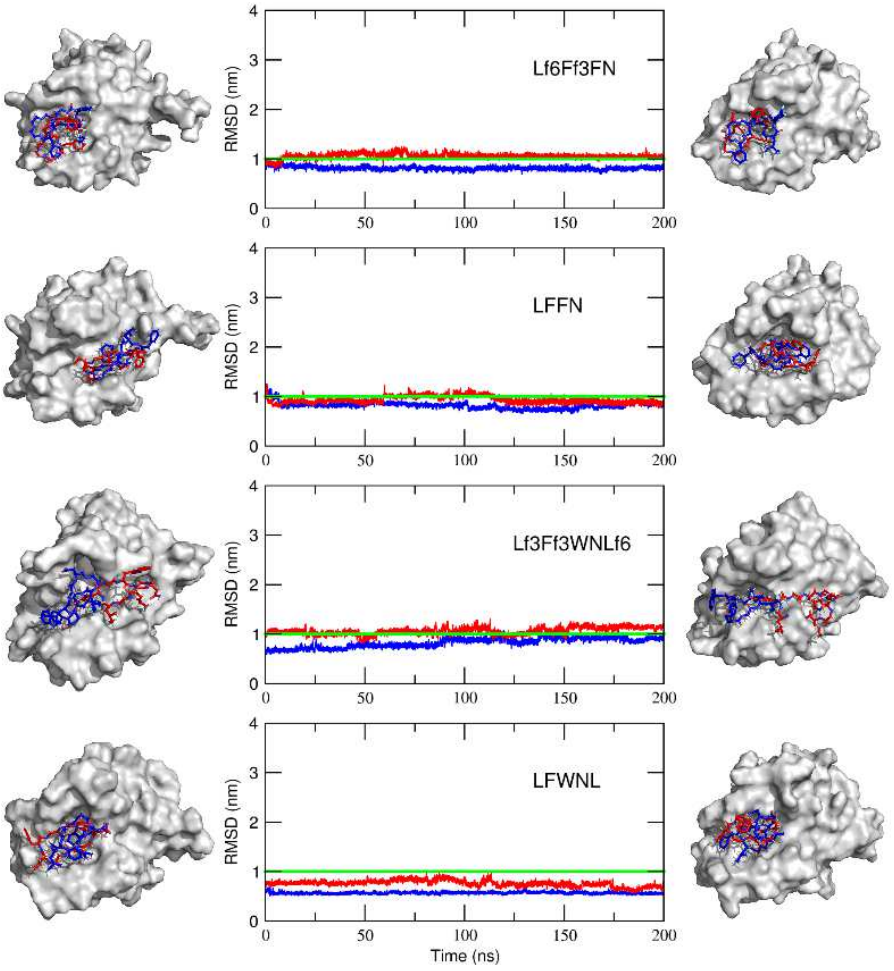

FIGURE 5.7 – Time evolution of the peptide Cα RMSD relative to the Des3PI hotspots in the Mcl1-fluorinated peptide complex MD simulations. Pictures in the left and right columns represent the complex initial and final structures, respectively.

For Ras, we chose to run a peptide from two different classes (figure 5.8). The B$_{f1}$F$_{f3}$F$_{f3}$V$_{f3}$WWR peptide from class V which had good associated VINA binding energy and a large number of hotspots, and DVWAR which was the only fully non-fluorinated peptide generated by the new version of Des3PI, and thus could serve as a reverse control comparing it with a valine-fluorinated analog not generated by Des3PI.

We observe that B$_{f1}$F$_{f3}$F$_{f3}$V$_{f3}$WWR remains relatively stable within its cavity along the 2 replicates, with an RMSD approaching 1nm relative to the Des3PI hotspots (figure 5.8). However, the results of its analogue (LFFVWWWR) show

that the peptide drops out of the cavity to lodge in another binding site. Similarly, the DVWAR peptide generated by Des3PI is stable in the cavity while the chosen analog exits the cavity and translates slightly under the protein. Thus, in both cases, the choice of the Des3PI algorithm to select or not a fluorinated amino acid seems to be optimal as supported by our validation protocol since only the peptides generated by Des3PI passed both steps. Although we would have liked to obtain a set of stable complexes to be able to perform MM-PBSA calculations in order to estimate the binding free energy, these studies provide a first piece of evidence in favor of the Des3PI results.

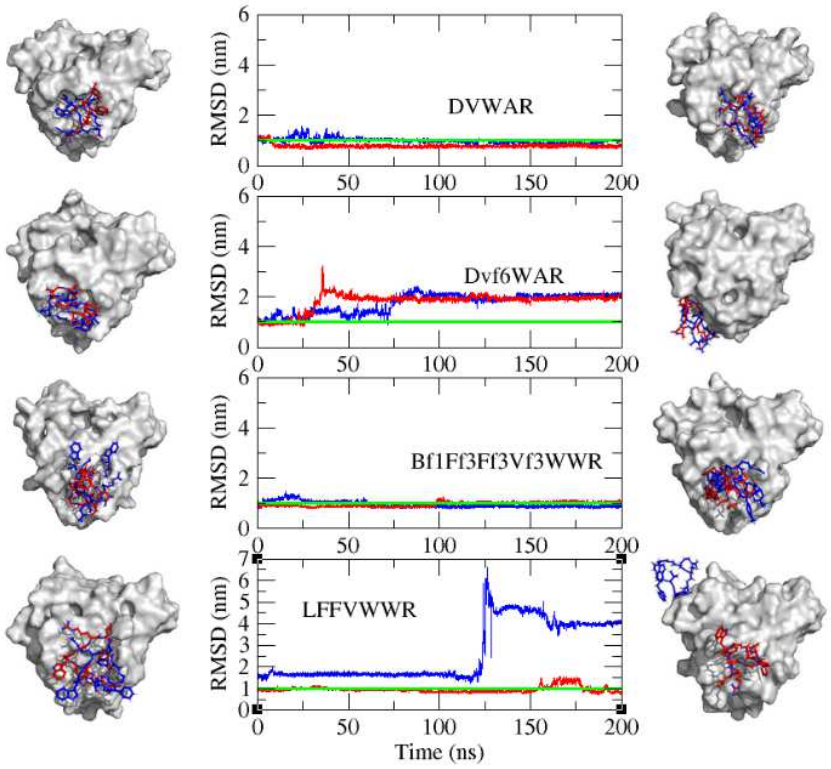

FIGURE 5.8 – Time evolution of the peptide Cα RMSD relative to the Des3PI hotspots in the Ras-fluorinated peptide complex MD simulations. Pictures in the left and right columns represent the complex initial and final structures, respectively.

5.3.4 Binding free energy of protein-fluorinated peptide complexes of Mcl-1

To assess whether the fluorinated peptides generated by Des3PI have a higher affinity for their target than their non-fluorinated analogues, we performed MM-PBSA binding free energy estimation from simulations of all complexes involving Mcl-1 (table 5.1).

Peptide	E_{LJ}	E_{Coul}	ΔG_{PB}	ΔG_{SA}	$-T\Delta S_{Conf}$	ΔG_{Bind}
$L_{f6}F_{f3}FN$ 1	-67 ± 5	-1 ± 1	18 ± 3	-6 ± 0.4	12 ± 2	-44 ± 7
$L_{f6}F_{f3}FN$ 2	-63 ± 3	-1 ± 0.4	15 ± 3	-6 ± 0.3	15 ± 2	- 40 ± 5
LFFN 1	-62 ± 6	-1 ± 1	16 ± 4	-5 ± 1	-9 ± 2	-61 ± 8
LFFN 2	-58 ± 4	-1 ± 1	15 ± 3	-5 ± 0.3	0 ± 3	-49 ± 7
$L_{f3}F_{f3}WNL_{f6}$ 1	-62 ± 6	0 ± 0.4	15 ± 3	-6 ± 0.4	-6 ± 2	-59 ± 8
$L_{f3}F_{f3}WNL_{f6}$ 2	-66 ± 10	0 ± 1	18 ± 5	-7 ± 1	1 ± 1	-54 ± 8
LFWNL 1	-62 ± 4	0 ± 0.3	11 ± 2	-5 ± 0.3	1 ± 1	-55 ± 5
LFWNL 2	-61 ± 6	-1 ± 1	16 ± 3	-6 ± 0.4	-9 ± 2	-61 ± 8

TABLE 5.1 – Binding free energies of all the Des3PI peptides and their non-fluorinated analogs in complex with Mcl-1. All energy terms are in kcal/mol.

The first general observation is that fluorinated and non-fluorinated peptides seem to have overall similar binding free energy. We notice that the different terms of the equation are also all of the same order, whatever the nature of the peptide. The predominant term is the LJ contribution which confirms that the protein-peptide interactions are mainly hydrophobic. Surprisingly, some value of $-T\Delta S_{Conf}$ are even negative. It is this term that explains the significant difference between the simulations of the peptide $L_{f6}F_{f3}FN$ and the first simulation of its non-fluorinated analogue LFFN. Nevertheless, this negative term $-T\Delta S_{Conf}$ might be due to a limited supply of the cyclic peptides by the simulations.

Overall, results indicate that there is a significant difference in the binding free energy between a fluorinated peptide generated by Des3PI and its non-fluorinated analog.

5.4 Conclusion and perspectives

It is known that the addition of fluorinated amino acids allows in many cases to improve not only the pharmacokinetic properties of a peptide, but also sometimes to enhance the affinity for its target, in particular thanks to an increase their hydrophobicity.

We therefore investigated whether the addition of such amino acids to the Des3PI library would make it possible to propose peptide sequences with fluorinated amino acids and to find out whether these are more affine for their targets than their natural analogues.

Thus, by applying this new version of Des3PI on the first proteins that we had studied, we saw that the new algorithm generated classes of peptides with relatively similar geometries allowing us to compare easily with the old results.

We also noticed that the algorithm did propose quite similar sequences as previously but with some specific substitutions of amino acids by their fluorinated equivalent.

We observed that between 30 to 40 % of these peptides for the two systems could be docked close to the binding mode proposed by Des3PI in their best scores. Stability testing on Ras by MD simulation unfortunately did not allow for further evaluation of affinity by MM-PBSA given the lack of stability of the chosen analogues. However, the fact that the peptides proposed by Des3PI remained stable is still encouraging. On the contrary, the MD studies on peptides in complex with Mcl-1 were stable for both the fluorinated peptides and their non-fluorinated analogues. This allowed us to perform binding free energy calculations. However, these calculations did not lead to a significant difference in the affinity between the fluorinated and non-fluorinated peptides.

In conclusion, the incorporation of fluorinated amino acids gave encouraging results despite of the difficulties encountered to correctly estimate their affinity with targeted proteins. It will be interesting to find a more accurate computational evaluation of such compounds while waiting for eventual experimental evaluations that would really allow to assess on the efficiency of such peptides.

Chapitre 6

Essais expérimentaux de peptides cycliques générés par Des3PI sur Aβ

6.1 Introduction

En parallèle du travail *in silico* fourni lors de cette thèse, nous avons cherché à initier diverses collaborations avec des équipes de biophysique expérimentale afin de valider de notre approche en testant *in vitro* divers peptides proposés par Des3PI. Ainsi, 4 collaborations expérimentales ont été initiées à ce jour. Deux d'entre elles concernent des tests d'inhibition de l'agrégation d'Aβ et IAPP en présence de nos peptides, réalisés par le Dr Julia Kaffy au sein de notre équipe FLUOPEPIT. Une autre collaboration concerne 3 peptides générés par Des3PI ciblant Mcl-1 pour lesquels des études RMN sont en cours par le Dr Ludovic Carlier à l'Université Paris-Sorbonne. Enfin, nous avions également appliqué Des3PI sur l'helicase bactérienne DnaB et généré des peptides avec lesquels des tests d'interactions sont en train d'être réalisés.

Grâce au travail du Dr Julia Kaffy, nous avons pu obtenir des résultats concernant les 4 peptides générés par Des3PI ciblant l'agrégation d'Aβ et que nous présenterons brièvement dans ce chapitre.

Ainsi, les peptides LWTW, LFTW, WYGK et WYIG ont été mis en présence de peptides Aβ pour un test de fluorescence à la Thioflavine T [87]. Cette molécule a pour particularité de fluorescer lorsqu'elle interagit avec des feuillets β. Ainsi, si les peptides Aβ s'agrègent en protofibrilles riches en feuillets β, la fluorescence augmentera. Si notre peptide inhibe ou retarde cette agrégation, nous observerons un retard et/ou une moindre hausse de la fluorescence.

6.2 Méthode

Les quatres peptides LWTW, LFTW, WYGK et WYIG ont été synthétisé sous leur forme cyclique par la plateforme PeptLab de Cergy-Paris Université. Nous disposons également des versions linéaires de LFTW et WYGK. Aβ1-42 a été acheté chez Bachem et la ThT a été obtenu chez Sigma. Le peptide Aβ1-42 été dissous dans de l'hexafluoro-isopropanol (HFIP) pur à une concentration de 1 mg/mL et incubé pendant 5 mn à température ambiante pour dissoudre les agrégats préformés. Ensuite, l'HFIP a été évaporé avec de l'azote gazeux sec suivi d'une dessiccation sous vide pendant au moins 2 heures, solubilisé dans l'eau (1 mg/mL) et le peptide a été lyophilisé. Le film peptidique résultant a ensuite été dissous dans une solution aqueuse d'ammoniac à 1 % jusqu'à une concentration de 1 mM puis, juste avant l'utilisation, a été dilué à 0,2 mM avec un tampon Tris-HCl 10 mM et NaCl 100 mM (pH 7,4). Les solutions mères des composés à tester ont été dissoutes dans du DMSO (20 mM), puis dans de l'eau à 2 mM et 0,2 mM. La concentration finale de DMSO a été maintenue constante à 0,5 % (v/v). La fluorescence du ThT a été mesurée pour évaluer le développement des fibrilles d'Aβ1-42 au cours du temps à l'aide d'un lecteur de plaques de fluorescence (Fluostar Optima, BMG labtech) avec des plaques de microtitration noires standard de 96 puits (volume final dans les puits de 200 μL). Les expériences d'inhibition ont été lancées en ajoutant le peptide (concentration finale de Aβ1-42 égale à 10 μM) dans un mélange contenant 40 μM de ThT dans un tampon Tris-HCl 10 mM et NaCl 100 mM (pH 7,4) avec et sans les composés à différentes concentrations (100 et 10 μM) à température ambiante. L'intensité de fluorescence du ThT de chaque échantillon (réalisé en triplicat) a été enregistrée avec des filtres d'excitation/émission de 440/480 nm réglés pendant 42 h en effectuant une double agitation orbitale de 10 s avant le premier cycle. Les essais de fluorescence ont été réalisés deux fois à des jours différents, avec le même lot de peptide. La capacité des peptides à inhiber l'agrégation d'Aβ1-42 a été évaluée en considérant le temps de demi-agrégation (t1/2) et l'intensité du plateau de fluorescence expérimental (F), les deux valeurs ont été obtenues en ajustant les données cinétiques obtenues avec une courbe sigmoïde de Boltzmann en utilisant Matplotlib. L'extension/réduction relative de t1/2 est définie comme le t1/2 expérimental en présence du composé testé par rapport à celui obtenu sans le composé et est évaluée comme le pourcentage suivant : [t1/2 (Aβ + composé) - t1/2 (Aβ)] / t1/2 (Aβ) × 100. L'extension/réduction relative du plateau de fluorescence expérimental est définie de la même manière comme le pourcentage suivant : (F(Aβ + composé) - F(Aβ)) / F(Aβ) × 100. Les résultats des triplicats ont été moyennés et les courbes ajustées en pourcentage par rapport à la valeur de fluorescence maximale d'Aβ contrôle.

6.3 Résultats et discussion

On observe tout d'abord que la courbe de fluorescence d'Aβ contrôle adopte bien une forme sigmoïde (figure 6.1) avec existence d'une phase de latence en début de cinétique, puis apparition d'un plateau de fluorescence. L'allongement de la phase de latence (extension/réduction de t1/2 > 100% dans le tableau 6.1) et l'abaissement du plateau de fluorescence (extension/réduction du plateau expérimental < 0% dans le tableau 6.1) en présence d'un de nos peptides traduira une activité sur l'agrégation d'Aβ.

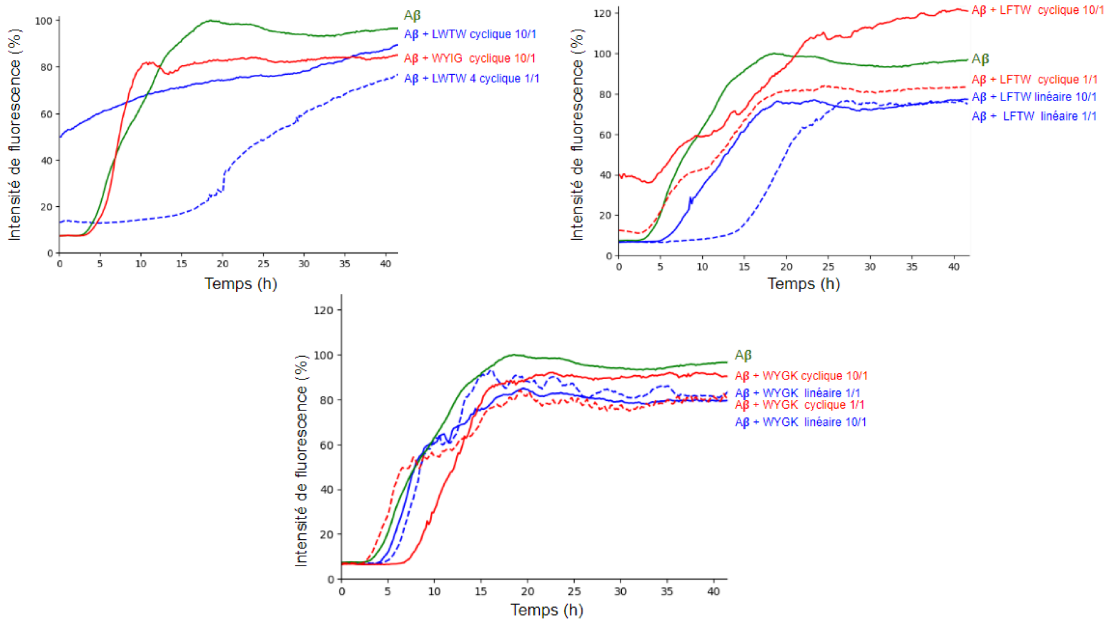

FIGURE 6.1 – Courbes de fluorescence d'Aβ en fonction du temps et de la présence d'un inhibiteur. Le ratio 10/1 signifie que le peptide inhibiteur est en concentration 10 fois plus importante qu'Aβ. Le ratio 1/1 signifie que le peptide inhibiteur est à la même concentration qu'Aβ.

De façon générale, tous les peptides générés par Des3PI semblent avoir une meilleure activité à concentration égale avec Aβ plutôt qu'au ratio 10/1. On observe en effet bien souvent une phase de latence plus longue et un abaissement du plateau plus conséquent. L'une des hypothèses les plus plausibles est qu'à trop grande concentration, nos peptides ont tendance à s'auto-agréger et ne sont plus disponibles pour interagir avec Aβ.

Avec un ratio 1/1, LWTW cyclique et LFTW linéaire semblent avoir une activité importante sur l'agrégation d'Aβ avec une extension de la phase de latence supérieure à 120% et un abaissement du plateau de fluorescence de 20%. Ainsi,

Peptides	Ratio de concentrations	Extension/réduction du t1/2	Extension/réduction du plateau
LWTW cyclique	10/1	Absence de phase de latence	-7%
LWTW cyclique	1/1	**172%**	**-24%**
LWTW linéaire	10/1	non testé	non testé
LWTW linéaire	1/1	non testé	non testé
WYIG cyclique	10/1	-13%	-12%
WYIG cyclique	1/1	non testé	non testé
LFTW linéaire	10/1	36%	-20%
LFTW linéaire	1/1	**129%**	**-20%**
LFTW cyclique	10/1	37%	27%
LFTW cyclique	1/1	19%	-13%
WYGK linéaire	10/1	-23%	-12%
WYGK linéaire	1/1	8%	-3%
WYGK cyclique	10/1	44%	-4%
WYGK cyclique	1/1	-16%	-12%

TABLE 6.1 – Tableau récapitulatif des extension réductions des valeurs de t1/2 et du plateau après ajustement à une sigmoïde de Boltzmann

bien que LFTW soit normalement proposé avec une conformation cyclique par Des3PI, nous avons tout de même deux peptides sur les 4 testés qui ont une activité significative sur l'agrégation d'Aβ, qui est freinée en présence de ces inhibiteurs.

Le peptide WYGK semble avoir une faible activité sur Aβ quelque soit sa concentration et sa forme cyclique ou linéaire. Il ne provoque qu'un léger retard de l'agrégation et un léger abaissement du plateau qui ne semblent pas suffisamment significatif, avec des extensions du t1/2 de l'ordre de 10 à 20% et un abaissement du plateau de l'ordre de 5 à 10%.

Enfin, les résultats de WYIG ne montrent pas non plus d'influence significative sur l'agrégation d'Aβ. Il est à noter que le ratio 1/1 n'a pu être testé par manque de produits sur la fin de la manipulation.

6.4 Conclusion

Les résultats des tests expérimentaux de l'activité des peptides générés par Des3PI ont montré un ralentissement de l'agrégation d'Aβ clair pour deux d'entre eux. Les résultats ainsi présentés sont donc très encourageants. Bien que nous n'ayons à l'heure actuelle les résultats que pour un seul système, trois autres collaborations expérimentales ont été engagées. Un autre test de fluorescence à la Thioflavine T est envisagé afin de tester des peptides générés par Des3PI sur des oligomères de la protéine IAPP impliquée dans le diabète de type II au laboratoire BioCIS, également supervisés par le Dr Julia Kaffy. Ensuite, des expériences de caractérisation structurale par RMN sont actuellement en cours pour trois de nos peptides ciblant Mcl-1, supervisés par le Dr Ludovic Carlier

(Sorbonne-Université). Enfin, des tests d'interactions en présence de peptides générés par Des3PI ciblant l'helicase bactérienne DnaB sont en cours de réalisation dans l'équile de Sophie Quevillon Cheruel à l'I2BC.

Conclusion générale

Cette thèse avait pour but de développer une méthode computationnelle de conception de dérivés peptidiques cycliques inhibiteurs d'interactions protéine-protéine.

Ces IPP sont en effet à la base de la plupart des mécanismes de régulation et de signalisation cellulaires. Elles sont ainsi impliquées dans des processus biologiques majeurs, tels que l'apoptose médiée par une chaîne d'interactions entre protéines pro et anti-apoptotiques servant de verrous à la libération des caspases provoquant la mort cellulaire. On comprend ainsi qu'une simple dérégulation d'une de ces interactions peut conduire à de nombreuses pathologies graves. Par exemple, la surexpression de la protéine antiapoptotique Bcl-xL déséquilibre les voies de l'apoptose et induit certains cancers.

Ainsi, bien que considérées comme difficile à moduler, l'intérêt thérapeutique pour la conception de principes actifs ciblant ces IPP reste majeur. Le relatif échec des petits composés chimiques a mis en lumière qu'il était nécessaire d'opter pour des molécules plus grandes et plus spécifiques afin d'espérer moduler de telles interactions. Le développement de peptides inhibiteurs d'IPP s'est donc fortement accéléré ces dernières décennies. Aujourd'hui, plusieurs dizaines d'inhibiteurs sont sur le marché ou en essais cliniques et de très nombreux composés en études préliminaires dans des aires thérapeutiques très variées.

Au chapitre 1, nous avons exposé comment chimistes et biologistes concevaient ces peptides thérapeutiques. Le point de départ de ces composés est obtenu soit par une identification d'un motif peptidique minimum se liant à une des protéines-cibles ou, de plus en plus fréquemment et efficacement par une technique de criblage haut-débit, le *phage-display*. Ces peptides sont ensuite optimisés afin d'améliorer l'affinité pour leur cible et optimiser leurs faibles propriétés pharmacocinétiques qu'ils possèdent. Pour cela, il est possible de les cycliser, les N-méthyler ou encore d'y incorporer des acides aminés modifiés.

Compte tenu du coût important et du temps conséquent des développements expérimentaux, nous avons ensuite vu que les approches computationnelles avaient toute leur place dans la conception de peptides inhibiteurs d'IPP

(Chapitre 2). Nous avons vu qu'il était même possible de proposer des séquences peptidiques d'intérêt lorsque les informations sur les cibles protéiques étaient très minces. Les méthodes dites *sequence-based* permettent ainsi de prédire les propriétés thérapeutiques d'une séquence peptidique ou encore de savoir si celle ci a le potentiel d'interagir avec une cible principalement par des approches dites *sequence-based*. Il est également possible de déterminer les éléments clés d'une interface protéine-protéine à l'aide d'outils computationnels, tels que les cavités ou de *hotspots* permettant de guider la conception de peptides. Lorsque la structure de la cible est connue et grâce à l'augmentation de la puissance des ordinateurs, il est de plus en plus possible de réaliser des criblages virtuels de librairies de peptides. Enfin, il existe également des outils d'optimisation de séquences peptidiques consistant en de nombreuses estimations d'énergies libres de liaison pour tester rapidement l'influence d'une modification de séquence.

Nous avons tiré comme conséquence de cet état de l'art que le défi principal restait de trouver la ou les meilleures séquences peptidiques capables d'inhiber une IPP donnée. Or, il s'avérait que les approches dites *fragment-based*, pourtant efficaces dans la conception de petites molécules, étaient sous représentées dans ce domaine d'étude. C'est dans ce contexte que nous avons conçu et développé Des3PI, notre outil de conception de peptides par fragemnts. Nous avons choisi de calibrer notre approche sur trois systèmes sélectionnés à l'interface protéine-protéine. Mcl-1 étant en interaction avec une hélice α, Ras avec une hélice α et un feuillet β et les oligomères d'Aβ constitués de feuillets β (Chapitre 3).

L'application de Des3PI a ainsi permis de générer 5 séquences ciblant Mcl-1, 20 ciblant Ras et 12 ciblant Aβ. Ces séquences sont globalement très hydrophobes, avec notamment le tryptophane qui est régulièrement représenté au sein des séquences trouvées. Afin de pouvoir convaincre les expérimentateurs d'initier une collaboration, nous avons mis en place un protocole de validation *in silico* consistant en un étape de *docking* aveugle suivie d'un test de stabilité par dynamique moléculaire pour les complexes les plus prometteurs. Pour chacun des trois systèmes, au moins trois peptides ont validé l'ensemble du protocole de validation. Ces résultats encourageants ont permis de déboucher sur une collaboration expérimentale concernant Aβ avec le Dr. Julia Kaffy de notre équipe ainsi que pour Mcl-1 par le Dr. Ludovic Carlier de Sorbonne-Université.

Dans l'attente de résultats expérimentaux, nous avons poursuivis les tests de robustesse de notre approche. Dans une deuxième étude, nous avons recherché à estimer l'influence de la forme de la cavité ciblée sur les résultats de Des3PI, mais aussi à comparer les séquences générées par notre algorithme à des séquences peptidiques connues pour interagir avec une cible donnée (Chapitre 4). Ainsi, nous avons appliqué notre approche sur les domaines PDZ et

SH3 cocristallisés avec un petit motif peptidique linéaire, dont la cavité en sillon est plus étroite qu'une interface protéine-protéine classique. Nous avons également utilisé Des3PI sur CXCR4 et l'intégrine αVβ3 qui ont pour intérêt d'être cocristallisées avec un peptide cyclique et dont les cavités ont plutôt la forme d'un entonnoir profond.

Les résultats sur CXCR4 et l'intégrine furent relativement décevants puisque nous n'avons pas retrouvé de point commun entre les séquences générées par Des3PI et celles des différents peptides cycliques cocristallisés. Il s'avère même que dans le cas de l'intégrine αVβ3, Des3PI propose un résidu chargé positivement alors qu'il est sensé interagir avec un cation Mn^{2+} présent au sein de la protéine. Cela peut être dû au fait que VINA prend mal en compte les interaction avec les ions des partenaires protéiques. Cependant, les résultats sur les domaines PDZ et SH3 furent plus encourageants. En effet, notre algorithme est capable de retrouver des résidus de nature très similaires à ceux des peptides cocristallisés. Parmi ceux-ci, 1 sur 3 peptides ciblant SH3 et 4sur 4 peptides ciblant PDZ ont une énergie libre d'association (estimée par MM-PBSA) plus favorable que les peptides cocristallisés avec ces 2 domaines. Cette étude a permis de renforcer l'idée que bien que perfectible, notre approche semblait toujours encourageante.

Le dernier travail computationnel de cette thèse a consisté à enrichir la librairie d'acides aminés de Des3PI afin d'augmenter la diversité des peptides générés. Nous avons ainsi ajouté 16 acides aminés fluorés, connus pour améliorer l'affinité des peptides et leurs paramètres pharmacocinétiques (Chapitre 5). Nous avons testé cette nouvelle librairie sur des protéines déjà ciblées, Ras et Mcl-1 afin de pouvoir comparer nos résultats. Cette approche a permis de retrouver des peptides de conformations relativement similaires à ceux obtenus avec la librairie non enrichie mais a considérablement augmenté la diversité de séquences. Dans le cas de Mcl-1, 5 classes de peptides ont été générées contre une précédemment. Certains acides aminés se sont montré particulièrement conservés et d'autres ont bel et bien été substitués par leur analogues fluorés.Dans le cas de Ras, le nombre de nouvelles séquences (14) reste comparable aux 20 obtenues précédemment, et de même que pour Mcl-1, certains acides aminés ont été strictement conservés et d'autres ont été substitués par des acides aminés fluorés. Les simulations de dynamique moléculaire des peptides fluorés ciblant Mcl-1, ainsi que celles de leurs analogues non fluorés se sont avérées stables. Les calculs de MM-PBSA n'ont pas mis en lumière de différences significatives en terme d'affinité entre les peptides fluorés et non fluorés. Dans le cas de Ras, nous n'avons malheureusement pas pu réaliser de tels tests puisque bien que les peptides générés par Des3PI se soient avérés stables, leurs analogues ne restaient pas en interaction dans la cavité. Cependant, ces résul-

tats nous ont apporté l'information que le choix d'un acide aminé fluoré ou non par Des3PI permettait d'apporter une certaine stabilité au peptide au sein de la cavité. Mais il reste à initier des collaborations expérimentales afin de valider la pertinence de nos composés par rapport aux non-fluorés.

Enfin, cette thèse se termine sur des résultats expérimentaux préliminaires sur l'effet de nos peptides sur l'aggrégation d'Aβ. Les travaux de Julia Kaffy ont montré que deux de nos quatre composés ont une activité intéressante sur l'aggrégation d'Aβ, ce qui constitue un résultat très encourageant. En parallèle, Des3PI a été appliqué sur des oligomères d'IAPP, et les tests des peptides obtenus seront réalisés dans les prochains mois. Les peptides ciblant Mcl-1 ont, quant à eux, été synthétisés et les tests expérimentaux sur la protéine Mcl-1 sont en cours. Enfin, les résultats expérimentaux concernant des peptides inhibiteurs de l'helicase bactérienne DnaB sont en voie de finalisation.

Les enjeux concernant la suite de ce projet sont multiples. Tout d'abord, les résultats prometteurs de notre approche doivent être confirmés expérimentalement sur un plus grand nombre de systèmes. D'un point de vue computationnel, plusieurs pistes d'améliorations peuvent être envisagées. L'une des limites de notre approche est qu'il est difficile de déterminer avec précision l'orientation des chaînes latérales des *hotspots* compte tenu du fait que seuls les Cα sont clusterisés. Ainsi, une alternative à développer serait de réaliser le clustering des modes de liaison des acides aminés à la fois sur le Cα mais également sur le centre de masse de la chaîne latérale afind'obtenir des séquences plus précises et peut être plus diverses. Nous avons également vu que la présence d'un ion à l'interface d'une protéine cible pouvait provoquer une baisse de l'efficacité de Des3PI. Ainsi, il serait intéressant d'implémenter dans Des3PI une possibilité de choisir parmi plusieurs programmes de docking celui qui serait le plus adapté à une cible donnée. Dans cette optique, nous avons initié une collaboration avec le Pr Sanner pour développer une version d'ADCP qui tient compte des acides aminés non naturels. Cependant, l'absence de données expérimentales sur la structure de complexes de protéines avec des peptides fluorés retarde la validation de ces développements.

Pour finir, nous souhaiterions mettre Des3PI à disposition des scientifiques travaillant autour du peptide. Dans cette optique, nous souhaitons développer une version "web-serveur" de Des3PI pour faciliter son utilisation et diffuser largement cet outil au sein des communautés de chimistes médicinaux mais aussi de chémobiologie.

Table des figures

Liste des tableaux

Bibliographie

[1] Aileron Therapeutics Initiates Phase 1 Cancer Study Of ALRN-6924 In Advanced Hematologic And Solid Malignancies With Wild Type P53.

[2] Click chemistry in Applications.

[3] Therapeutic peptides containing fluorinated amino acids - ProQuest.

[4] M.J. Abraham, T. Murtola, R. Schulz, S. Pall, J.C. Smith, B. Hess, and E. Lindahl. GROMACS : High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2 :19–25, 2015.

[5] Fayyaz ul Amir Afsar Minhas, Brian J. Geiss, and Asa Ben-Hur. PAIRpred : Partner-specific prediction of interacting residues from sequence and structure : Interface Prediction Using PAIRpred. *Proteins : Structure, Function, and Bioinformatics*, 82 :1142–1155, 2014.

[6] Sandip G. Agalave, Suleman R. Maujan, and Vandana S. Pore. Click chemistry : 1,2,3-triazoles as pharmacophores. *Chemistry, an Asian Journal*, 6 :2696–2718, 2011.

[7] Nawsad Alam, Oriel Goldstein, Bing Xia, Kathryn A. Porter, Dima Kozakov, and Ora Schueler-Furman. High-resolution global peptide-protein docking using fragments-based PIPER-FlexPepDock. *PLOS Computational Biology*, 13 :e1005905, 2017.

[8] Stephen F. Altschul, Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman. Basic local alignment search tool. *Journal of Molecular Biology*, 215 :403–410, 1990.

[9] Damien Amouyel. In *Dosage des peptides amyloïdes Aβ1-42 et Aβ1-40 dans le liquide cérébro-spinal et intérêt de leur rapport dans le diagnostic de la maladie d'Alzheimer*. 2018.

[10] Dinler A. Antunes, Mark Moll, Didier Devaurs, Kyle R. Jackson, Gregory Lizée, and Lydia E. Kavraki. DINC 2.0 : A New Protein-Peptide Docking Webserver Using an Incremental Approach. *Cancer Research*, 77 :e55–e57, 2017.

[11] Tadamasa Arai, Takushi Araya, Daisuke Sasaki, Atsuhiko Taniguchi, Takeshi Sato, Youhei Sohma, and Motomu Kanai. Rational Design and Identification of a Non-Peptidic Aggregation Inhibitor of Amyloid-β Based on a Pharmacophore Motif Obtained from *cyclo* [-Lys-Leu-Val-Phe-Phe-]. *Angew. Chem.*, 126 :8375–8378, 2014.

[12] Alexander I. Archakov, Vadim M. Govorun, Alexander V. Dubanov, Yuri D. Ivanov, Alexander V. Veselovsky, Paul Lewi, and Paul Janssen. Protein-protein interactions as a target for drugs in proteomics. *Proteomics*, 3 :380–391, 2003.

[13] Muhammad Arif, Saeed Ahmad, Farman Ali, Ge Fang, Min Li, and Dong-Jun Yu. TargetCPP : accurate prediction of cell-penetrating peptides from optimized multi-scale features using gradient boost decision tree. *Journal of Computer-Aided Molecular Design*, 34 :841–856, 2020.

[14] Michelle R. Arkin, Yinyan Tang, and James A. Wells. Small-Molecule Inhibitors of Protein-Protein Interactions : Progressing toward the Reality. *Chemistry & Biology*, 21 :1102–1114, 2014.

[15] Vanessa Baeriswyl, Sara Calzavarini, Shiyu Chen, Alessandro Zorzi, Luca Bologna, Anne Angelillo-Scherrer, and Christian Heinis. A Synthetic Factor XIIa Inhibitor Blocks Selectively Intrinsic Coagulation Initiation. *ACS Chemical Biology*, 10(8) :1861–1870, 2015.

[16] Kyle A. Barlow, Shane Ó Conchúir, Samuel Thompson, Pooja Suresh, James E. Lucas, Markus Heinonen, and Tanja Kortemme. Flex ddG : Rosetta Ensemble-Based Estimation of Changes in Protein-Protein Binding Affinity upon Mutation. *Journal of Physical Chemistry. B*, 122 :5389–5399, 2018.

[17] Alper Baspinar, Engin Cukuroglu, Ruth Nussinov, Ozlem Keskin, and Attila Gursoy. PRISM : a web server and repository for prediction of protein–protein interactions and modeling their 3D complexes. *Nucleic Acids Research*, 42 :W285–W289, 2014.

[18] Avraham Ben-Shimon and Masha Y. Niv. AnchorDock : Blind and Flexible Anchor-Driven Peptide Docking. *Structure*, 23 :929–940, 2015.

[19] Christopher A. Benetatos, Yasuhiro Mitsuuchi, Jennifer M. Burns, Eric M. Neiman, Stephen M. Condon, Guangyao Yu, Martin E. Seipel, Gurpreet S. Kapoor, Matthew G. LaPotre, Susan R. Rippin, Yijun Deng, Mukta S. Hendi, Pavan K. Tirunahari, Yu-Hua Lee, Thomas Haimowitz, Matthew D. Alexander, Martin A. Graham, David Weng, Yigong Shi, Mark A. McKinlay, and Srinivas K. Chunduru. Birinapant(TL32711), a Bivalent Smac Mimetic, Targets TRAF2-associated cIAPs, Abrogates TNF-induced NF-κB Activation and

is Active in Patient-Derived Xenograft Models. *Molecular Cancer Therapeutics*, page molcanther.0798.2013, 2014.

[20] Stephanie Berger, Erik Procko, Daciana Margineantu, Erinna F Lee, Betty W Shen, Alex Zelter, Daniel-Adriano Silva, Kusum Chawla, Marco J Herold, Jean-Marc Garnier, Richard Johnson, Michael J MacCoss, Guillaume Lessene, Trisha N Davis, Patrick S Stayton, Barry L Stoddard, W Douglas Fairlie, David M Hockenbery, and David Baker. Computationally designed high specificity inhibitors delineate the roles of BCL2 family proteins in cancer. *eLife*, 5 :e20352, 2016.

[21] Pratiti Bhadra, Jielu Yan, Jinyan Li, Simon Fong, and Shirley W. I. Siu. AmPEP : Sequence-based prediction of antimicrobial peptides using distribution patterns of amino acid properties and random forest. *Scientific Reports*, 8 :1697, 2018.

[22] T. Andrew Binkowski, Shapor Naghibzadeh, and Jie Liang. CASTp : Computed Atlas of Surface Topography of proteins. *Nucleic Acids Research*, 31 :3352–3355, 2003.

[23] Jose Liñares Blanco, Ana B. Porto-Pazos, Alejandro Pazos, and Carlos Fernandez-Lozano. Prediction of high anti-angiogenic activity peptides in silico using a generalized linear model and feature selection. *Scientific Reports*, 8 :15688, 2018.

[24] F Edward Boas and Pehr B Harbury. Potential energy functions for protein design. *Current Opinion in Structural Biology*, 17 :199–204, 2007.

[25] Laura Bonetta. Interactome under construction. *Nature*, 468(7325) :851–852, 2010. Number : 7325 Publisher : Nature Publishing Group.

[26] Bertan Bopp, Emanuele Ciglia, Anissa Ouald-Chaib, Georg Groth, Holger Gohlke, and Joachim Jose. Design and biological testing of peptidic dimerization inhibitors of human Hsp90 that target the C-terminal domain. *Biochimica et Biophysica Acta*, 1860 :1043–1055, 2016.

[27] Tjibbe Bosma, Rick Rink, Markus A. Moosmeier, and Gert N. Moll. Genetically Encoded Libraries of Constrained Peptides. *ChemBioChem*, 20 :1754–1758, 2019.

[28] Benjamin C. Buer and E. Neil G. Marsh. Fluorine : A new element in protein design. *Protein Science*, 21(4) :453–462, 2012.

[29] Patricia A. Burke, Sally J. DeNardo, Laird A. Miers, Kathleen R. Lamborn, Siegfried Matzku, and Gerald L. DeNardo. Cilengitide targeting of alpha(v)beta(3) integrin receptor synergizes with radioimmunotherapy to increase efficacy and apoptosis in breast cancer xenografts. *Cancer Research*, 62 :4263–4272, 2002.

[30] Fabiana Caporuscio, Andrea Tafi, Emmanuel González, Fabrizio Manetti, José A. Esté, and Maurizio Botta. A dynamic target-based pharmacophoric model mapping the CD4 binding site on HIV-1 gp120 to identify new inhibitors of gp120–CD4 protein–protein interactions. *Bioorganic & Medicinal Chemistry Letters*, 19 :6087–6091, 2009.

[31] M D Carter, G A Simms, and D F Weaver. The Development of New Therapeutics for Alzheimer's Disease. *Clinical Pharmacology & Therapeutics*, 88(4) :475–486, 2010.

[32] Adrià Cereto-Massagué, Maria José Ojeda, Cristina Valls, Miquel Mulero, Santiago Garcia-Vallvé, and Gerard Pujadas. Molecular fingerprint similarity search in virtual screening. *Methods*, 71 :58–63, 2015.

[33] J. Chai, C. Du, J. W. Wu, S. Kyin, X. Wang, and Y. Shi. Structural and biochemical basis of apoptotic activation by Smac/DIABLO. *Nature*, 406 :855–862, 2000.

[34] Maud Chan-Yao-Chong, Célia Deville, Louise Pinet, Carine van Heijenoort, Dominique Durand, and Tâp Ha-Duong. Structural Characterization of N-WASP Domain V Using MD Simulations with NMR and SAXS Data. *Biophysical Journal*, 116 :1216–1227, 2019.

[35] Jayanta Chatterjee, Chaim Gilon, Amnon Hoffman, and Horst Kessler. N-methylation of peptides : a new perspective in medicinal chemistry. *Accounts of Chemical Research*, 41 :1331–1342, 2008.

[36] Jayanta Chatterjee, Florian Rechenmacher, and Horst Kessler. N-Methylation of Peptides and Proteins : An Important Element for Modulating Biological Functions. *Angewandte Chemie International Edition*, 52 :254–269, 2013.

[37] ChemAxon. Software solutions and services for chemistry & biology. https ://www.chemaxon.com, 2020.

[38] Jiunn R Chen, Bryan H Chang, John E Allen, Michael A Stiffler, and Gavin MacBeath. Predicting PDZ domain–peptide interactions from primary sequences. *Nature Biotechnology*, 26 :1041–1045, 2008.

[39] Muhao Chen, Chelsea J T Ju, Guangyu Zhou, Xuelu Chen, Tianran Zhang, Kai-Wei Chang, Carlo Zaniolo, and Wei Wang. Multifaceted protein–protein interaction prediction based on Siamese residual RCNN. *Bioinformatics*, 35 :i305–i314, 2019.

[40] Rong Chen, Li Li, and Zhiping Weng. ZDOCK : an initial-stage protein-docking algorithm. *Proteins*, 52 :80–87, 2003.

[41] Shiyu Chen, Julia Morales-Sanfrutos, Alessandro Angelini, Brian Cutting, and Christian Heinis. Structurally Diverse Cyclisation Linkers Impose

Different Backbone Conformations in Bicyclic Peptides. *ChemBioChem*, 13(7) :1032–1038, 2012.

[42] T. Scott Chen, Aaron W. Reinke, and Amy E. Keating. Design of Peptide Inhibitors That Bind the bZIP Domain of Epstein–Barr Virus Protein BZLF1. *Journal of Molecular Biology*, 408 :304–320, 2011.

[43] Wei Chen, Hui Ding, Pengmian Feng, Hao Lin, and Kuo-Chen Chou. iACP : a sequence-based tool for identifying anticancer peptides. *Oncotarget*, 7 :16895–16909, 2016.

[44] Pin-Nan Cheng, Cong Liu, Minglei Zhao, David Eisenberg, and James S. Nowick. Amyloid β-sheet mimics that antagonize protein aggregation and reduce amyloid toxicity. *Nat. Chem.*, 4 :927–933, 2012.

[45] Tammy Man-Kuang Cheng, Tom L. Blundell, and Juan Fernandez-Recio. pyDock : electrostatics and desolvation for effective scoring of rigid-body protein-protein docking. *Proteins*, 68 :503–515, 2007.

[46] Hsien-Po Chiu, Bashkim Kokona, Robert Fairman, and Richard P. Cheng. Effect of Highly Fluorinated Amino Acids on Protein Stability at a Solvent-Exposed Position on an Internal Strand of Protein G B1 Domain. *Journal of the American Chemical Society*, 131 :13192–13193, 2009.

[47] Hsien-Po Chiu, Yuta Suzuki, Donald Gullickson, Raheel Ahmad, Bashkim Kokona, Robert Fairman, and Richard P. Cheng. Helix Propensity of Highly Fluorinated Amino Acids. *Journal of the American Chemical Society*, 128(49) :15556–15557, 2006.

[48] Kuo-Chen Chou. Prediction of protein cellular attributes using pseudo-amino acid composition. *Proteins : Structure, Function, and Genetics*, 43 :246–255, 2001.

[49] Nikolaj R. Christensen, Jelena Čalyševa, Eduardo F. A. Fernandes, Susanne Lüchow, Louise S. Clemmensen, Linda M. Haugaard-Kedström, and Kristian Strømgaard. PDZ Domains as Drug Targets. *Adv. Therap.*, 2 :1800143, 2019.

[50] Qian Chu, Raymond E. Moellering, Gerard J. Hilinski, Young-Woo Kim, Tom N. Grossmann, Johannes T.-H. Yeh, and Gregory L. Verdine. Towards understanding cell penetration by stapled peptides. *MedChemComm*, 6 :111–119, 2015.

[51] Pj Coates and Pa Hall. The yeast two-hybrid system for identifying protein–protein interactions. *The Journal of Pathology*, 199(1) :4–7, 2003.

[52] Michael T. Colvin, Robert Silvers, Qing Zhe Ni, Thach V. Can, Ivan Sergeyev, Melanie Rosay, Kevin J. Donovan, Brian Michael, Joseph Wall, Sara Linse,

and Robert G. Griffin. Atomic Resolution Structure of Monomorphic Aβ42 Amyloid Fibrils. *J. Am. Chem. Soc.*, 138 :9663–9674, 2016.

[53] Sean P. Cornillie, Benjamin J. Bruno, Carol S. Lim, and Thomas E. Cheatham. Computational Modeling of Stapled Peptides toward a Treatment Strategy for CML and Broader Implications in the Design of Lengthy Peptide Therapeutics. *Journal of Physical Chemistry B*, 122 :3864–3875, 2018.

[54] Dermot Cox, Marian Brennan, and Niamh Moran. Integrins as therapeutic targets : lessons and opportunities. *Nature Reviews. Drug Discovery*, 9 :804–820, 2010.

[55] Anna D Cunningham, Nir Qvit, and Daria Mochly-Rosen. Peptides and peptidomimetics as regulators of protein–protein interactions. *Current Opinion in Structural Biology*, 44 :59–66, 2017.

[56] B. C. Cunningham and J. A. Wells. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science*, 244 :1081–1085, 1989.

[57] Meghna Dabur, Joana A. Loureiro, and Maria Carmo Pereira. The current state of amyloidosis therapeutics and the potential role of fluorine in their treatment. *Biochimie*, 2022.

[58] Antoine Daina, Olivier Michielin, and Vincent Zoete. SwissADME : a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.*, 7 :42717, 2017.

[59] João M. Damas, Luís C.S. Filipe, Sara R.R. Campos, Diana Lousa, Bruno L. Victor, António M. Baptista, and Cláudio M. Soares. Predicting the Thermodynamics and Kinetics of Helix Formation in a Cyclic Peptide Model. *Journal of Chemical Theory and Computation*, 9 :5148–5157, 2013.

[60] F. Dapiaggi, S. Pieraccini, and M. Sironi. In silico study of VP35 inhibitors : from computational alanine scanning to essential dynamics. *Molecular BioSystems*, 11 :2152–2157, 2015.

[61] Norman E. Davey, Kim Van Roey, Robert J. Weatheritt, Grischa Toedt, Bora Uyar, Brigitte Altenberg, Aidan Budd, Francesca Diella, Holger Dinkel, and Toby J. Gibson. Attributes of short linear motifs. *Mol. BioSyst.*, 8 :268–281, 2012.

[62] Catherine L. Day, Callum Smits, F. Cindy Fan, Erinna F. Lee, W. Douglas Fairlie, and Mark G. Hinds. Structure of the BH3 Domains from the p53-Inducible BH3-Only Proteins Noxa and Puma in Complex with Mcl-1. *J. Mol. Biol.*, 380 :958–971, 2008.

[63] Bernard Degryse, Juan Fernandez-Recio, Valentina Citro, Francescol Blasi, and Maria Vittoria Cubellis. In silico docking of urokinase plasminogen activator and integrins. *BMC Bioinformatics*, 9 :S8, 2008.

[64] Maxence Delaunay and Tâp Ha-Duong. Computational Tools and Strategies to Develop Peptide-Based Inhibitors of Protein-Protein Interactions. In Thomas Simonson, editor, *Computational Peptide Science : Methods and Protocols*, Methods in Molecular Biology, pages 205–230. Springer US, New York, NY, 2022.

[65] Maxence Delaunay and Tâp Ha-Duong. Des3PI : a fragment-based approach to design cyclic peptides targeting protein-protein interactions. *Journal of Computer-Aided Molecular Design*, 36(8) :605–621, 2022.

[66] Oliver Demmer, Andreas O. Frank, Franz Hagn, Margret Schottelius, Luciana Marinelli, Sandro Cosconati, Ruth Brack-Werner, Stephan Kremb, Hans-Jürgen Wester, and Horst Kessler. A conformationally frozen peptoid boosts CXCR4 affinity and anti-HIV activity. *Angewandte Chemie (International Ed. in English)*, 51 :8110–8113, 2012.

[67] Lei Deng, Jihong Guan, Xiaoming Wei, Yuan Yi, Qiangfeng Cliff Zhang, and Shuigeng Zhou. Boosting prediction performance of protein-protein interaction hot spots by using structural neighborhood properties. *Journal of Computational Biology*, 20 :878–891, 2013.

[68] Camille Denis, Jana Sopková-de Oliveira Santos, Ronan Bureau, and Anne Sophie Voisin-Chiret. Hot-Spots of Mcl-1 Protein : Miniperspective. *J. Med. Chem.*, 63 :928–943, 2020.

[69] Kaustubh Dhole, Gurdeep Singh, Priyadarshini P. Pai, and Sukanta Mondal. Sequence-based prediction of protein–protein interaction sites with L1-logreg classifier. *Journal of Theoretical Biology*, 348 :47–54, 2014.

[70] Fatemeh Miri Disfani, Wei-Lun Hsu, Marcin J. Mizianty, Christopher J. Oldfield, Bin Xue, A. Keith Dunker, Vladimir N. Uversky, and Lukasz Kurgan. MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. *Bioinformatics*, 28 :i75–i83, 2012.

[71] Bradley Croy Doak, Bjorn Over, Fabrizio Giordanetto, and Jan Kihlberg. Oral Druggable Space beyond the Rule of 5 : Insights from Drugs and Clinical Candidates. *Chem. Biol.*, 21 :1115–1142, 2014.

[72] Zsuzsanna Dosztányi, Veronika Csizmok, Peter Tompa, and István Simon. IUPred : web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*, 21 :3433–3434, 2005.

[73] Laurence Dubrez, Jean Berthelet, and Valérie Glorian. IAP proteins as targets for drug development in oncology. *OncoTargets and therapy*, 6 :1285–1304, 2013.

[74] Fergal J. Duffy, Darragh O'Donovan, Marc Devocelle, Niamh Moran, David J. O'Connell, and Denis C. Shields. Virtual Screening Using Combinatorial Cyclic Peptide Libraries Reveals Protein Interfaces Readily Targetable by Cyclic Peptides. *J. Chem. Inf. Model.*, 55 :600–613, 2015.

[75] Fergal J. Duffy, Mélanie Verniere, Marc Devocelle, Elise Bernard, Denis C. Shields, and Anthony J. Chubb. CycloPs : Generating Virtual Libraries of Cyclized and Constrained Peptides Including Nonnatural Amino Acids. *Journal of Chemical Information and Modeling*, 51 :829–836, 2011.

[76] Sanjib Dutta, Stefano Gullá, T. Scott Chen, Emiko Fire, Robert A. Grant, and Amy E. Keating. Determinants of BH3 Binding Specificity for Mcl-1 versus Bcl-xL. *J. Mol. Biol.*, 398 :747–762, 2010.

[77] Fatma-Elzahraa Eid, Mahmoud ElHefnawi, and Lenwood S. Heath. DeNovo : virus-host sequence-based protein–protein interaction prediction. *Bioinformatics*, 32 :1144–1150, 2016.

[78] Ulrich Essmann, Lalith Perera, Max L. Berkowitz, Tom Darden, Hsing Lee, and Lee G. Pedersen. A smooth particle mesh Ewald method. *J. Chem. Phys.*, 103 :8577–8593, 1995.

[79] Chandler B Est, Parth Mangrolia, and Regina M Murphy. ROSETTA-informed design of structurally stabilized cyclic anti-amyloid peptides. *Protein Engineering, Design and Selection*, 32 :47–57, 2019.

[80] Rudi Fasan, Ricardo L. A. Dias, Kerstin Moehle, Oliver Zerbe, Jan W. Vrijbloed, Daniel Obrecht, and John A. Robinson. Using a β-Hairpin To Mimic an α-Helix : Cyclic Peptidomimetic Inhibitors of the p53–HDM2 Protein–Protein Interaction. *Angewandte Chemie International Edition*, 43 :2109–2112, 2004.

[81] S. M. Fayaz and G. K. Rajanikant. Modelling the molecular mechanism of protein–protein interactions and their inhibition : CypD–p53 case study. *Molecular Diversity*, 19 :931–943, 2015.

[82] Sibo Feng, James K. Chen, Hongtao Yu, Julian A. Simon, and Stuart L. Schreiber. Two Binding Orientations for Peptides to the Src SH3 Domain : Development of a General Model for SH3-Ligand Interactions. *Science*, 266 :1241–1247, 1994.

[83] Daniel Filchtinski, Oz Sharabi, Alma Rüppel, Ingrid R. Vetter, Christian Herrmann, and Julia M. Shifman. What Makes Ras an Efficient Molecular Switch : A Computational, Biophysical, and Structural Study of Ras-GDP Interactions with Mutants of Raf. *J. Mol. Biol.*, 399 :422–435, 2010.

[84] Glenna Wink Foight, Jeremy A. Ryan, Stefano V. Gullá, Anthony Letai, and Amy E. Keating. Designed BH3 peptides with high affinity and specificity for targeting Mcl-1 in cells. *ACS chemical biology*, 9 :1962–1968, 2014.

[85] Keld Fosgerau and Torsten Hoffmann. Peptide therapeutics : current status and future directions. *Drug Discovery Today*, 20 :122–128, 2015.

[86] Vincent Frappier, Justin M. Jenson, Jianfu Zhou, Gevorg Grigoryan, and Amy E. Keating. Tertiary Structural Motif Sequence Statistics Enable Facile Prediction and Design of Peptides that Bind Anti-apoptotic Bfl-1 and Mcl-1. *Structure*, 27 :606–617.e5, 2019.

[87] Kirsten Gade Malmos, Luis M. Blancas-Mejia, Benedikt Weber, Johannes Buchner, Marina Ramirez-Alvarado, Hironobu Naiki, and Daniel Otzen. ThT 101 : a primer on the use of thioflavin T to investigate amyloid formation. *Amyloid : The International Journal of Experimental and Clinical Investigation : The Official Journal of the International Society of Amyloidosis*, 24(1) :1–16, 2017.

[88] Ankit Ganeshpurkar, Rayala Swetha, Devendra Kumar, Gore P. Gangaram, Ravi Singh, Gopichand Gutti, Srabanti Jana, Dileep Kumar, Ashok Kumar, and Sushil K. Singh. Protein-Protein Interactions and Aggregation Inhibitors in Alzheimer's Disease. *Current Topics in Medicinal Chemistry*, 19 :501–533, 2019.

[89] Javier Garcia-Garcia, Victòria Valls-Comamala, Emre Guney, David Andreu, Francisco J. Muñoz, Narcis Fernandez-Fuentes, and Baldo Oliva. iFrag : A Protein–Protein Interface Prediction Server Based on Sequence Fragments. *Journal of Molecular Biology*, 429 :382–389, 2017.

[90] Jason Gavenonis, Nicholas E. Jonas, and Joshua A. Kritzer. Potential C-terminal-domain inhibitors of heat shock protein 90 derived from a C-terminal peptide helix. *Bioorganic & Medicinal Chemistry*, 22 :3989–3993, 2014.

[91] Priyanka Gawade and Payel Ghosh. Genomics driven approach for identification of novel therapeutic targets in Salmonella enterica. *Gene*, 668 :211–220, 2018.

[92] Cunliang Geng, Anna Vangone, Gert E. Folkers, Li C. Xue, and Alexandre M. J. J. Bonvin. iSEE : Interface structure, evolution, and energy-based machine learning predictor of binding affinity changes upon mutations. *Proteins : Structure, Function, and Bioinformatics*, 87 :110–119, 2019.

[93] Tim Geppert, Stefanie Bauer, Jan A. Hiss, Elea Conrad, Michael Reutlinger, Petra Schneider, Martin Weisel, Bernhard Pfeiffer, Karl-Heinz Altmann, Zoe Waibler, and Gisbert Schneider. Immunosuppressive Small Molecule Discovered by Structure-Based Virtual Screening for Inhibitors of Protein–Protein Interactions. *Angewandte Chemie International Edition*, 51 :258–261, 2012.

[94] Christoph Giese, Sandra Lepthien, Linda Metzner, Matthias Brandsch, Nediljko Budisa, and Hauke Lilie. Intracellular uptake and inhibitory activity of aromatic fluorinated amino acids in human breast cancer cells. *ChemMedChem*, 3(9) :1449–1456, 2008.

[95] Karen L. Goa and Stuart Noble. Eptifibatide. *Drugs*, 57 :439–462, 1999.

[96] Ragul Gowthaman, Sven A. Miller, Steven Rogers, Jittasak Khowsathit, Lan Lan, Nan Bai, David K. Johnson, Chunjing Liu, Liang Xu, Asokan Anbanandam, Jeffrey Aubé, Anuradha Roy, and John Karanicolas. DARC : Mapping Surface Topography by Ray-Casting for Effective Virtual Screening at Protein Interaction Sites. *Journal of Medicinal Chemistry*, 59 :4152–4170, 2016.

[97] Gevorg Grigoryan, Aaron W. Reinke, and Amy E. Keating. Design of protein-interaction specificity gives selective bZIP-binding peptides. *Nature*, 458 :859–864, 2009.

[98] Danielle A. Guarracino, Jacob A. Riordan, Gianna M. Barreto, Alexis L. Oldfield, Christopher M. Kouba, and Desiree Agrinsoni. Macrocyclic Control in Helix Mimetics. *Chem. Rev.*, 119 :9915–9949, 2019.

[99] Zuojun Guo, Bo Li, Joachim Dzubiella, Li-Tien Cheng, J. Andrew McCammon, and Jianwei Che. Evaluation of Hydration Free Energy by Level-Set Variational Implicit-Solvent Model with Coulomb-Field Approximation. *Journal of Chemical Theory and Computation*, 9 :1778–1787, 2013.

[100] Zuojun Guo, Atli Thorarensen, Jianwei Che, and Li Xing. Target the More Druggable Protein States in a Highly Dynamic Protein–Protein Interaction System. *Journal of Chemical Information and Modeling*, 56 :35–45, 2016.

[101] Sudheer Gupta, Parul Mittal, Midhun K. Madhu, and Vineet K. Sharma. IL17eScan : A Tool for the Identification of Peptides Inducing IL-17 Response. *Frontiers in Immunology*, 8 :1430, 2017.

[102] Pamela R. Hall, Andrei Leitão, Chunyan Ye, Kathleen Kilpatrick, Brian Hjelle, Tudor I. Oprea, and Richard S. Larson. Small molecule inhibitors of hantavirus infection. *Bioorganic & Medicinal Chemistry Letters*, 20 :7085–7091, 2010.

[103] Somaye Hashemifar, Behnam Neyshabur, Aly A Khan, and Jinbo Xu. Predicting protein-protein interactions through sequence-based deep learning. *Bioinformatics*, 34 :i802–i810, 2018.

[104] Hao He, Jiaxiang Zhao, and Guiling Sun. Computational prediction of MoRFs based on protein sequences and minimax probability machine. *BMC Bioinformatics*, 20 :529, 2019.

[105] Jingrui He, Ziyi Li, Gagan Dhawan, Wei Zhang, Alexander E. Sorochinsky, Greg Butler, Vadim A. Soloshonok, and Jianlin Han. Fluorine-containing drugs approved by the FDA in 2021. *Chinese Chemical Letters*, 2022.

[106] Liping He, Jingxiao Bao, Yunpeng Yang, Suzhen Dong, Lujia Zhang, Yifei Qi, and John Z. H. Zhang. Study of SHMT2 Inhibitors and Their Binding Mechanism by Computational Alanine Scanning. *Journal of Chemical Information and Modeling*, 59 :3871–3878, 2019.

[107] Berk Hess. P-LINCS : A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.*, 4 :116–122, 2008.

[108] Heike Hofmann and Stefan Pöhlmann. Cellular entry of the SARS coronavirus. *Trends in Microbiology*, 12(10) :466–472, 2004.

[109] William G. Hoover. Canonical dynamics : Equilibrium phase-space distributions. *Phys. Rev. A*, 31 :1695–1697, 1985.

[110] Shoichiro Horita, Yayoi Nomura, Yumi Sato, Tatsuro Shimamura, So Iwata, and Norimichi Nomura. High-resolution crystal structure of the therapeutic antibody pembrolizumab bound to the human PD-1. *Scientific Reports*, 6 :35297, 2016.

[111] Parisa Hosseinzadeh, Paris R. Watson, Timothy W. Craven, Xinting Li, Stephen Rettie, Fátima Pardo-Avila, Asim K. Bera, Vikram Khipple Mulligan, Peilong Lu, Alexander S. Ford, Brian D. Weitzner, Lance J. Stewart, Adam P. Moyer, Maddalena Di Piazza, Joshua G. Whalen, Per Greisen, David W. Christianson, and David Baker. Anchor extension : a structure-guided approach to design cyclic peptides targeting enzyme active sites. *Nat. Commun.*, 12 :3384, 2021.

[112] Qingzhen Hou, Paul F G De Geest, Christian J Griffioen, Sanne Abeln, Jaap Heringa, and K Anton Feenstra. SeRenDIP : SEquential REmasteriNg to DerIve profiles for fast and accurate predictions of PPI interface positions. *Bioinformatics*, 35 :4794–4796, 2019.

[113] Po-Ssu Huang, Scott E. Boyken, and David Baker. The coming of age of de novo protein design. *Nature*, 537 :320–327, 2016.

[114] Willy Hugo, See-Kiong Ng, and Wing-Kin Sung. D-SLIMMER : Domain-SLiM Interaction Motifs Miner for Sequence Based Protein-Protein Interaction Data. *Journal of Proteome Research*, 10 :5285–5295, 2011.

[115] Susanne Huhmann and Beate Koksch. Fine-Tuning the Proteolytic Stability of Peptides with Fluorinated Amino Acids. *European Journal of Organic Chemistry*, 2018(27-28) :3667–3679, 2018.

[116] Susanne Huhmann, Elisabeth K. Nyakatura, Anette Rohrhofer, Johann Moschner, Barbara Schmidt, Jutta Eichler, Christian Roth, and Beate Koksch. Systematic Evaluation of Fluorination as Modification for Peptide-Based Fusion Inhibitors against HIV-1 Infection. *ChemBioChem*, 22(24) :3443–3451, 2021.

[117] Amaurys A. Ibarra, Gail J. Bartlett, Zsöfia Hegedüs, Som Dutt, Fruzsina Hobor, Katherine A. Horner, Kristina Hetherington, Kirstin Spence, Adam Nelson, Thomas A. Edwards, Derek N. Woolfson, Richard B. Sessions, and Andrew J. Wilson. Predicting and Experimentally Validating Hot-Spot Residues at Protein–Protein Interfaces. *ACS Chemical Biology*, 14 :2252–2263, 2019.

[118] Vidhya V. Iyer. A Review of Stapled Peptides and Small Molecules to Inhibit Protein-Protein Interactions in Cancer. *Current Medicinal Chemistry*, 23 :3025–3043, 2016.

[119] Farzaneh Jafary, Mohamad Reza Ganjalikhany, Ali Moradi, Mahdie Hemati, and Sepideh Jafari. Novel Peptide Inhibitors for Lactate Dehydrogenase A (LDHA) : A Survey to Inhibit LDHA Activity via Disruption of Protein-Protein Interaction. *Scientific Reports*, 9 :4686, 2019.

[120] Araz Jakalian, David B. Jack, and Christopher I. Bayly. Fast, efficient generation of high-quality atomic charges. AM1-BCC model : II. Parameterization and validation. *Journal of Computational Chemistry*, 23 :1623–1641, 2002.

[121] Justina Jankauskaitė, Brian Jiménez-García, Justas Dapkūnas, Juan Fernández-Recio, and Iain H. Moal. SKEMPI 2.0 : an updated benchmark of changes in protein–protein binding energy, kinetics and thermodynamics upon mutation. *Bioinformatics*, 35 :462–469, 2019.

[122] Weronika Jaroszewicz, Joanna Morcinek-Orłowska, Karolina Pierzynowska, Lidia Gaffke, and Grzegorz Węgrzyn. Phage display and other peptide display technologies. *FEMS Microbiology Reviews*, 46(2) :fuab052, 2022.

[123] Sherlyn Jemimah, Masakazu Sekijima, and M Michael Gromiha. ProAffiMuSeq : sequence-based method to predict the binding free energy change of protein–protein complexes upon mutation using functional classification. *Bioinformatics*, 36 :1725–1730, 2019.

[124] Sherlyn Jemimah, K. Yugandhar, and M. Michael Gromiha. PROXiMATE : a database of mutant protein–protein complex thermodynamics and kinetics. *Bioinformatics*, 33 :2787–2788, 2017.

[125] M. Jesus Perez de Vega, Mercedes Martin-Martinez, and Rosario Gonzalez-Muniz. Modulation of Protein-Protein Interactions by Stabilizing/Mimicking Protein Secondary Structure Elements. *Current Topics in Medicinal Chemistry*, 7 :33–62, 2007.

[126] Jianhua Jia, Zi Liu, Xuan Xiao, Bingxiang Liu, and Kuo-Chen Chou. iPPBS-Opt : A Sequence-Based Ensemble Classifier for Identifying Protein-Protein Binding Sites by Optimizing Imbalanced Training Datasets. *Molecules*, 21 :95, 2016.

[127] Isak Johansson-Åkhe, Claudio Mirabello, and Bjorn Wallner. InterPep2 : global peptide–protein docking using interaction surface templates. *Bioinformatics*, 36 :2458–2465, 2020.

[128] David K. Johnson and John Karanicolas. Ultra-High-Throughput Structure-Based Virtual Screening for Small-Molecule Inhibitors of Protein–Protein Interactions. *Journal of Chemical Information and Modeling*, 56 :399–411, 2016.

[129] Gareth Jones, Peter Willett, Robert C Glen, Andrew R Leach, and Robin Taylor. Development and validation of a genetic algorithm for flexible docking11Edited by F. E. Cohen. *Journal of Molecular Biology*, 267 :727–748, 1997.

[130] Kasper W. Jorgensen, Michael Rasmussen, Soren Buus, and Morten Nielsen. NetMHCstab - predicting stability of peptide-MHC-I complexes; impacts for cytotoxic T lymphocyte epitope discovery. *Immunology*, 141 :18–26, 2014.

[131] Thomas L. Joseph, David P. Lane, and Chandra S. Verma. Stapled BH3 Peptides against MCL-1 : Mechanism and Design Using Atomistic Simulations. *PLOS One*, 7 :e43985, 2012.

[132] Sandeep Kalepu and Vijaykumar Nekkanti. Insoluble drug delivery strategies : review of recent advances and business prospects. *Acta Pharm. Sin. B*, 5 :442–453, 2015.

[133] Hetunandan Kamisetty, Bornika Ghosh, Christopher James Langmead, and Chris Bailey-Kellogg. Learning Sequence Determinants of Protein :Protein Interaction Specificity with Sparse Graphical Models. *Journal of Computational Biology*, 22 :474–486, 2015.

[134] Teresa Kaserer, Katharina Beck, Muhammad Akram, Alex Odermatt, and Daniela Schuster. Pharmacophore Models and Pharmacophore-Based Virtual Screening : Concepts and Applications Exemplified on Hydroxysteroid Dehydrogenases. *Molecules*, 20 :22799–22832, 2015.

[135] Omid Khakshoor and James S Nowick. Artificial β-sheets : chemical models of β-sheets. *Current Opinion in Chemical Biology*, 12 :722–729, 2008.

[136] Maede Khosravian, Fateme Kazemi Faramarzi, Majid Mohammad Beigi, Mandana Behbahani, and Hassan Mohabatkar. Predicting Antibacterial Peptides by the Concept of Chou's Pseudo-amino Acid Composition and Machine Learning Methods. *Protein and Peptide Letters*, 20 :180–186, 2013.

[137] David E. Kim, Dylan Chivian, and David Baker. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Research*, 32 :W526–W531, 2004.

[138] Kenneth L. Kirk. Fluorine in medicinal chemistry : Recent therapeutic applications of fluorinated small molecules. *Journal of Fluorine Chemistry*, 127(8) :1013–1029, 2006.

[139] Kent Kirshenbaum, Annelise E. Barron, Richard A. Goldsmith, Philippe Armand, Erin K. Bradley, Kiet T. V. Truong, Ken A. Dill, Fred E. Cohen, and Ronald N. Zuckermann. Sequence-specific polypeptoids : A diverse family of heteropolymers with stable secondary structure. *Proceedings of the National Academy of Sciences of the United States of America*, 95 :4303–4308, 1998.

[140] Mark Klein. Stabilized helical peptides : overview of the technologies and its impact on drug discovery. *Expert Opinion on Drug Discovery*, 12 :1117–1125, 2017.

[141] David R. Koes, Alexander Dömling, and Carlos J. Camacho. AnchorQuery : Rapid online virtual screening for small-molecule protein–protein interaction inhibitors. *Protein Science*, 27 :229–232, 2018.

[142] Nicoleta Kokkoni, Kelvin Stott, Hozefa Amijee, Jody M. Mason, and Andrew J. Doig. N-Methylated peptide inhibitors of beta-amyloid aggregation and toxicity. Optimization of the inhibitor structure. *Biochemistry*, 45 :9906–9918, 2006.

[143] Tanja Kortemme, Lukasz A. Joachimiak, Alex N. Bullock, Aaron D. Schuler, Barry L. Stoddard, and David Baker. Computational redesign of protein-protein interaction specificity. *Nature Structural & Molecular Biology*, 11 :371–379, 2004.

[144] Irfan Kösesoy, Murat Gök, and Cemil Öz. A new sequence based encoding for prediction of host–pathogen protein interactions. *Computational Biology and Chemistry*, 78 :170–177, 2019.

[145] Izabella Krystkowiak, Jean Manguy, and Norman E Davey. PSSMSearch : a server for modeling, visualization, proteome-wide discovery and annotation of protein motif specificity determinants. *Nucleic Acids Research*, 46 :W235–W241, 2018.

[146] John L. Kulp, Ian S. Cloudsdale, John L. Kulp, and Frank Guarnieri. Hotspot identification on a broad class of proteins and RNA suggest unifying principles of molecular recognition. *PLOS One*, 12 :e0183327, 2017.

[147] John L. Kulp, John L. Kulp, David L. Pompliano, and Frank Guarnieri. Diverse Fragment Clustering and Water Exclusion Identify Protein Hot Spots. *Journal of the American Chemical Society*, 133 :10740–10743, 2011.

[148] Rashmi Kumari, Rajendra Kumar, and Andrew Lynn. g_mmpbsa—A GROMACS Tool for High-Throughput MM-PBSA Calculations. *J. Chem. Inf. Model.*, 54(7) :1951–1962, 2014.

[149] Mateusz Kurcinski, Michal Jamroz, Maciej Blaszczyk, Andrzej Kolinski, and Sebastian Kmiecik. CABS-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site. *Nucleic Acids Research*, 43 :W419–W424, 2015.

[150] Natalya Kurochkina and Udayan Guha. SH3 domains : modules of protein–protein interactions. *Biophys. Rev.*, 5 :29–39, 2013.

[151] Hannah F. Kyle, Kate F. Wickson, Jonathan Stott, George M. Burslem, Alexander L. Breeze, Christian Tiede, Darren C. Tomlinson, Stuart L. Warriner, Adam Nelson, Andrew J. Wilson, and Thomas A. Edwards. Exploration of the HIF-1α/p300 interface using peptide and Adhiron phage display technologies. *Molecular bioSystems*, 11 :2738–2749, 2015.

[152] Vishuda Laengsri, Chanin Nantasenamat, Nalini Schaduangrat, Pornlada Nuchnoi, Virapong Prachayasittikul, and Watshara Shoombuatong. TargetAntiAngio : A Sequence-Based Tool for the Prediction and Analysis of Anti-Angiogenic Peptides. *International Journal of Molecular Sciences*, 20 :2950, 2019.

[153] Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99 :12562–12566, 2002.

[154] Dilraj Lama, Soo T. Quah, Chandra S. Verma, Rajamani Lakshminarayanan, Roger W. Beuerman, David P. Lane, and Christopher J. Brown. Rational Optimization of Conformational Effects Induced By Hydrocarbon Staples in Peptides and their Binding Interfaces. *Scientific Reports*, 3 :3451, 2013.

[155] Jun Lan, Jiwan Ge, Jinfang Yu, Sisi Shan, Huan Zhou, Shilong Fan, Qi Zhang, Xuanling Shi, Qisheng Wang, Linqi Zhang, and Xinquan Wang. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, 581(7807) :215–220, 2020.

[156] Jolene L. Lau and Michael K. Dunn. Therapeutic peptides : Historical perspectives, current development trends, and future directions. *Bioorganic & Medicinal Chemistry*, 26(10) :2700–2707, 2018.

[157] Erik Laurini, Domenico Marson, Suzana Aulic, Maurizio Fermeglia, and Sabrina Pricl. Computational Alanine Scanning and Structural Analysis of the SARS-CoV-2 Spike Protein/Angiotensin-Converting Enzyme 2 Complex. *ACS Nano*, 14 :11821–11830, 2020.

[158] José Laxio Arenas, Julia Kaffy, and Sandrine Ongeri. Peptides and peptidomimetics as inhibitors of protein–protein interactions involving β-sheet secondary structures. *Curr. Opin. Chem. Biol.*, 52 :157–167, 2019.

[159] Tamas Lazar, Mainak Guharoy, Eva Schad, and Peter Tompa. Unique Physicochemical Patterns of Residues in Protein–Protein Interfaces. *J. Chem. Inf. Model.*, 58 :2164–2173, 2018.

[160] Vincent Le Guilloux, Peter Schmidtke, and Pierre Tuffery. Fpocket : An open source platform for ligand pocket detection. *BMC Bioinformatics*, 10 :168, 2009.

[161] Erinna F. Lee, Anna Fedorova, Kerry Zobel, Michelle J. Boyle, Hong Yang, Matthew A. Perugini, Peter M. Colman, David C.S. Huang, Kurt Deshayes, and W.Douglas Fairlie. Novel Bcl-2 Homology-3 Domain-like Sequences Identified from Screening Randomized Peptide Libraries for Inhibitors of the Pro-survival Bcl-2 Proteins. *J. Biol. Chem.*, 284 :31315–31326, 2009.

[162] Hasup Lee, Lim Heo, Myeong Sup Lee, and Chaok Seok. GalaxyPepDock : a protein-peptide docking tool based on interaction similarity and energy optimization. *Nucleic Acids Research*, 43 :W431–435, 2015.

[163] Ho-Jin Lee and Jie J Zheng. PDZ domains and their binding partners : structure, specificity, and modification. *Cell Commun. Signal.*, 8 :8, 2010.

[164] Marc F. Lensink, Sameer Velankar, and Shoshana J. Wodak. Modeling protein-protein and protein-peptide complexes : CAPRI 6th edition : Modeling Protein-Protein and Protein-Peptide Complexes. *Proteins : Structure, Function, and Bioinformatics*, 85 :359–377, 2017.

[165] Henry Chi-Ming Leung, Man-Hung Siu, Siu-Ming Yiu, Francis Yuk-Lun Chin, and Ken Wing-Kin Sung. Clustering-Based Approach For Predicting Motif Pairs From Protein Interaction Data. *Journal of Bioinformatics and Computational Biology*, 07 :701–716, 2009.

[166] Da-Wei Li, Mina Khanlarzadeh, Jinbu Wang, Shuanghong Huo, and Rafael Bruschweiler. Evaluation of Configurational Entropy Methods from Peptide Folding-Unfolding Simulation. *J. Phys. Chem. B*, 111 :13807–13813, 2007.

[167] Dan Li, Huajie Song, Hong Mei, Erhu Fang, Xiaojing Wang, Feng Yang, Huanhuan Li, Yajun Chen, Kai Huang, Liduan Zheng, and Qiangsong Tong. Armadillo repeat containing 12 promotes neuroblastoma progression through interaction with retinoblastoma binding protein 4. *Nature Communications*, 9 :2829, 2018.

[168] Gen Li, Swagata Pahari, Adithya Krishna Murthy, Siqi Liang, Robert Fragoza, Haiyuan Yu, and Emil Alexov. SAAMBE-SEQ : A Sequence-based Method for Predicting Mutation Effect on Protein-protein Binding Affinity. *Bioinformatics*, page btaa761, 2020.

[169] Xianyi Lian, Shiping Yang, Hong Li, Chen Fu, and Ziding Zhang. Machine-Learning-Based Predictor of Human-Bacteria Protein-Protein Interactions by Incorporating Comprehensive Host-Network Properties. *Journal of Proteome Research*, 18 :2195–2205, 2019.

[170] O. Lichtarge, H. R. Bourne, and F. E. Cohen. An evolutionary trace method defines binding surfaces common to protein families. *Journal of Molecular Biology*, 257(2) :342–358, 1996.

[171] Christopher A. Lipinski, Franco Lombardo, Beryl W. Dominy, and Paul J. Feeney. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.*, 23 :3–25, 1997.

[172] Siyu Liu, Chuyao Liu, and Lei Deng. Machine Learning Approaches for Protein–Protein Interaction Hot Spot Prediction : Progress and Comparative Assessment. *Molecules*, page 2535, 2018.

[173] Nir London, Barak Raveh, Dana Movshovitz-Attias, and Ora Schueler-Furman. Can Self-Inhibitory Peptides be Derived from the Interfaces of Globular Protein-Protein Interactions ? *Proteins*, 78 :3140–3149, 2010.

[174] Nir London, Barak Raveh, and Ora Schueler-Furman. Druggable protein–protein interactions – from hot spots to hot segments. *Current Opinion in Chemical Biology*, 17 :952–959, 2013.

[175] Haiying Lu, Qiaodan Zhou, Jun He, Zhongliang Jiang, Cheng Peng, Rongsheng Tong, and Jianyou Shi. Recent advances in the development of protein–protein interactions modulators : mechanisms and clinical trials. *Sig. Transduct. Target. Ther.*, 5 :213, 2020.

[176] X.-L. Lu, X. Cao, X.-Y. Liu, and B.-H. Jiao. Recent Progress of Src SH2 and SH3 Inhibitors as Anticancer Agents. *Curr. Med. Chem.*, 17 :1117–1124, 2010.

[177] Jiesi Luo, Yanzhi Guo, Yun Zhong, Duo Ma, Wenling Li, and Menglong Li. A functional feature analysis on diverse protein–protein interactions : application for the prediction of binding affinity. *Journal of Computer-Aided Molecular Design*, pages 619–629, 2014.

[178] Shuai Ma, Huan Zhang, Xiaoyan Dong, Linling Yu, Jie Zheng, and Yan Sun. Head-to-tail cyclization of a heptapeptide eliminates its cytotoxicity and significantly increases its inhibition effect on amyloid β-protein fibrillation and cytotoxicity. *Front. Chem. Sci. Eng.*, 12 :283–295, 2018.

[179] Stephani Joy Y. Macalino, Shaherin Basith, Nina Abigail B. Clavio, Hyerim Chang, Soosung Kang, and Sun Choi. Evolution of In Silico Strategies for Protein-Protein Interaction Drug Discovery. *Molecules*, 23 :1963, 2018.

[180] Craig O. Mackenzie, Jianfu Zhou, and Gevorg Grigoryan. Tertiary alphabet for the observable protein structural universe. *Proceedings of the National Academy of Sciences*, 113 :E7438–E7447, 2016.

[181] Atanu Maity, Sarmistha Majumdar, and Shubhra Ghosh Dastidar. Flexibility enables to discriminate between ligands : Lessons from structural ensembles of Bcl-xl and Mcl-1. *Computational Biology and Chemistry*, 77 :17–27, 2018.

[182] Nawar Malhis and Jörg Gsponer. Computational identification of MoRFs in protein sequences. *Bioinformatics*, 31 :1738–1744, 2015.

[183] Balachandran Manavalan, Tae H. Shin, Myeong O. Kim, and Gwang Lee. AIPpred : Sequence-Based Prediction of Anti-inflammatory Peptides Using Random Forest. *Frontiers in Pharmacology*, 9 :276, 2018.

[184] Taryn L. March, Martin R. Johnston, Peter J. Duggan, and James Gardiner. Synthesis, Structure, and Biological Applications of α-Fluorinated β-Amino Acids and Derivatives. *Chemistry & Biodiversity*, 9(11) :2410–2441, 2012.

[185] Sílvia A. Martins, Marta A. S. Perez, Irina S. Moreira, Sérgio F. Sousa, M. J. Ramos, and P. A. Fernandes. Computational Alanine Scanning Mutagenesis : MM-PBSA vs TI. *Journal of Chemical Theory and Computation*, 9 :1311–1319, 2013.

[186] Carlos Mas-Moruno, Florian Rechenmacher, and Horst Kessler. Cilengitide : the first anti-angiogenic small molecule drug candidate design, synthesis and clinical evaluation. *Anti-Cancer Agents in Medicinal Chemistry*, 10 :753–768, 2010.

[187] Stephen M. Massa, Youmei Xie, and Frank M. Longo. Alzheimer's therapeutics. *Journal of Molecular Neuroscience*, 20 :323–326, 2003.

[188] Dirk Matthes and Bert L. de Groot. Secondary Structure Propensities in Peptide Folding Simulations : A Systematic Comparison of Molecular Mechanics Interaction Schemes. *Biophysical Journal*, 97 :599–608, 2009.

[189] Sean M McHugh, Julia R Rogers, Sarah A Solomon, Hongtao Yu, and Yu-Shan Lin. Computational methods to design cyclic peptides. *Current Opinion in Chemical Biology*, 34 :95–102, 2016.

[190] Haibo Mei, Jianlin Han, Karel D. Klika, Kunisuke Izawa, Tatsunori Sato, Nicholas A. Meanwell, and Vadim A. Soloshonok. Applications of fluorine-containing amino acids for drug design. *European Journal of Medicinal Chemistry*, 186 :111826, 2020.

[191] Haibo Mei, Attila Márió Remete, Yupiao Zou, Hiroki Moriwaki, Santos Fustero, Lorand Kiss, Vadim A. Soloshonok, and Jianlin Han. Fluorine-containing drugs approved by the FDA in 2019. *Chinese Chemical Letters*, 31(9) :2401–2413, 2020.

[192] He Meng and Krishna Kumar. Antimicrobial Activity and Protease Stability of Peptides Containing Fluorinated Amino Acids. *Journal of the American Chemical Society*, 129(50) :15615–15622, 2007.

[193] Michael J. Meyer, Juan Felipe Beltrán, Siqi Liang, Robert Fragoza, Aaron Rumack, Jin Liang, Xiaomu Wei, and Haiyuan Yu. Interactome INSIDER : a structural interactome browser for genomic studies. *Nature Methods*, 15 :107–114, 2018.

[194] Lech-Gustav Milroy, Tom N. Grossmann, Sven Hennig, Luc Brunsveld, and Christian Ottmann. Modulators of protein-protein interactions. *Chemical Reviews*, 114(9) :4695–4748, 2014.

[195] Piotr Minkiewicz, Anna Iwaniak, and Małgorzata Darewicz. BIOPEP-UWM Database of Bioactive Peptides : Current Opportunities. *International Journal of Molecular Sciences*, 20 :5978, 2019.

[196] Iain H. Moal, Rudi Agius, and Paul A. Bates. Protein–protein binding affinity prediction on a diverse set of structures. *Bioinformatics*, 27 :3002–3009, 2011.

[197] Ashley E. Modell, Sarah L. Blosser, and Paramjit S. Arora. Systematic Targeting of Protein–Protein Interactions. *Trends in Pharmacological Sciences*, 37 :702–713, 2016.

[198] Julia M. Monkovic, Halle Gibson, Jonathan W. Sun, and Jin Kim Montclare. Fluorinated Protein and Peptide Materials for Biomedical Applications. *Pharmaceuticals*, 15(10) :1201, 2022. Number : 10 Publisher : Multidisciplinary Digital Publishing Institute.

[199] Johann Moschner, Valentina Stulberg, Rita Fernandes, Susanne Huhmann, Jakob Leppkes, and Beate Koksch. Approaches to Obtaining Fluorinated α-Amino Acids. *Chemical Reviews*, 119(18) :10718–10801, 2019.

[200] Yuguang Mu, Phuong H. Nguyen, and Gerhard Stock. Energy landscape of a small peptide revealed by dihedral angle principal component analysis. *Proteins*, 58 :45–52, 2004.

[201] Laura Nevola and Ernest Giralt. Modulating protein-protein interactions : the potential of peptides. *Chemical Communications (Cambridge, England)*, 51 :3302–3315, 2015.

[202] Matilda S. Newton, Yari Cabezas-Perusse, Cher Ling Tong, and Burckhard Seelig. *In Vitro* Selection of Peptides and Proteins—Advantages of mRNA Display. *ACS Synth. Biol.*, 9 :181–190, 2020.

[203] Rod K. Nibbe, Salim A. Chowdhury, Mehmet Koyutürk, Rob Ewing, and Mark R. Chance. Protein–protein interaction networks and subnetworks in the biology of disease. *WIREs Systems Biology and Medicine*, 3(3) :357–367, 2011.

[204] Julian Nomme, Axelle Renodon-Cornière, Yuya Asanomi, Kazuyasu Sakaguchi, Alicja Z. Stasiak, Andrzej Stasiak, Bengt Norden, Vinh Tran, and Masayuki Takahashi. Design of Potent Inhibitors of Human RAD51 Recombinase Based on BRC Motifs of BRCA2 Protein : Modeling and Experimental Validation of a Chimera Peptide. *Journal of Medicinal Chemistry*, 53 :5782–5791, 2010.

[205] Julian Nomme, Yoshimasa Takizawa, Susan F. Martinez, Axelle Renodon-Cornière, Fabrice Fleury, Pierre Weigel, Ken-ichi Yamamoto, Hitoshi Kurumizaka, and Masayuki Takahashi. Inhibition of filament formation of human Rad51 protein by a small peptide derived from the BRC-motif of the BRCA2 protein. *Genes to Cells*, 13 :471–481, 2008.

[206] Anne M. Noonan, Kristen P. Bunch, Jin-Qiu Chen, Michelle A. Herrmann, Jung-Min Lee, Elise C. Kohn, Ciara C. O'Sullivan, Elizabeth Jordan, Nicole Houston, Naoko Takebe, Robert J. Kinders, Liang Cao, Cody J. Peer, W. Douglas Figg, and Christina M. Annunziata. Pharmacodynamic markers and clinical results from the phase 2 study of the SMAC mimetic birinapant in women with relapsed platinum-resistant or -refractory epithelial ovarian cancer. *Cancer*, 122, 2016.

[207] Shuichi Nosé. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.*, 81 :511–519, 1984.

[208] Titia Rixt Oppewal, Ivar D. Jansen, Johan Hekelaar, and Clemens Mayer. A Strategy to Select Macrocyclic Peptides Featuring Asymmetric Molecular Scaffolds as Cyclization Units by Phage Display. *Journal of the American Chemical Society*, 144(8) :3644–3652, 2022.

[209] Cristina Paissoni, Michela Ghitti, Laura Belvisi, Andrea Spitaleri, and Giovanna Musco. Metadynamics Simulations Rationalise the Conformational Effects Induced by N-Methylation of RGD Cyclic Hexapeptides. *Chemistry – A European Journal*, 21 :14165–14170, 2015.

[210] Jyoti Pande, Magdalena M. Szewczyk, and Ashok K. Grover. Phage display : Concept, innovations, applications and future. *Biotechnology Advances*, 28(6) :849–858, 2010.

[211] Poonam Pandey, Vinal Patel, Nithin V. George, and Sairam S. Mallajosyula. KELM-CPPpred : Kernel Extreme Learning Machine Based Prediction Model for Cell-Penetrating Peptides. *Journal of Proteome Research*, 17 :3214–3222, 2018.

[212] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals : A new molecular dynamics method. *J. Appl. Phys.*, 52 :7182–7190, 1981.

[213] Laavanya Parthasarathi, Fergal Casey, Amelie Stein, Patrick Aloy, and Denis C. Shields. Approved Drug Mimics of Short Peptide Ligands from

Protein Interaction Motifs. *Journal of Chemical Information and Modeling*, 48 :1943–1948, 2008.

[214] Linus Pauling. *The Nature of the Chemical Bond and the Structure of Molecules and Crystals : An Introduction to Modern Structural Chemistry*. 1960.

[215] Marta Pelay-Gimeno, Adrian Glas, Oliver Koch, and Tom N. Grossmann. Structure-Based Design of Inhibitors of Protein-Protein Interactions : Mimicking Peptide Binding Epitopes. *Angewandte Chemie (International Ed. in English)*, 54 :8896–8927, 2015.

[216] Lenna X. Peterson, Amitava Roy, Charles Christoffer, Genki Terashi, and Daisuke Kihara. Modeling disordered protein interactions from biophysical principles. *PLOS Computational Biology*, 13 :e1005485, 2017.

[217] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. UCSF Chimera-A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25 :1605–1612, 2004.

[218] Emilie Pihan, Roberto F. Delgadillo, Michelle L. Tonkin, Martine Pugnière, Maryse Lebrun, Martin J. Boulanger, and Dominique Douguet. Computational and biophysical approaches to protein–protein interaction inhibition of Plasmodium falciparum AMA1/RON2 complex. *Journal of Computer-Aided Molecular Design*, 29 :525–539, 2015.

[219] Malak Pirtskhalava, Andrei Gabrielian, Phillip Cruz, Hannah L. Griggs, R. Burke Squires, Darrell E. Hurt, Maia Grigolava, Mindia Chubinidze, George Gogoladze, Boris Vishnepolsky, Vsevolod Alekseev, Alex Rosenthal, and Michael Tartakovsky. DBAASP v.2 : an enhanced database of structure and antimicrobial/cytotoxic activity of natural and synthetic peptides. *Nucleic Acids Research*, 44 :D1104–D1112, 2016.

[220] Srinivas Kumar Ponna, Salla Ruskamo, Matti Myllykoski, Corinna Keller, Tobias M. Boeckers, and Petri Kursula. Structural basis for PDZ domain interactions in the post-synaptic density scaffolding protein Shank3. *J. Neurochem.*, 145 :449–463, 2018.

[221] Alan M Poole and Rama Ranganathan. Knowledge-based potentials in protein design. *Current Opinion in Structural Biology*, 16 :508–513, 2006.

[222] Yanhua Qiao, Yi Xiong, Hongyun Gao, Xiaolei Zhu, and Peng Chen. Protein-protein interface hot spots prediction based on a hybrid feature selection strategy. *BMC bioinformatics*, 19 :14, 2018.

[223] Jin Qiu, Qianqian Fan, Sainan Xu, Dongmei Wang, Juntong Chen, Sainan Wang, Tianhui Hu, Xinran Ma, Yiyun Cheng, and Lingyan Xu. A fluorinated peptide with high serum- and lipid-tolerence for the delivery of siRNA

drugs to treat obesity and metabolic dysfunction. *Biomaterials*, 285 :121541, 2022.

[224] Justin S. Quartararo, Matthew R. Eshelman, Leila Peraro, Hongtao Yu, James D. Baleja, Yu-Shan Lin, and Joshua A. Kritzer. A bicyclic peptide scaffold promotes phosphotyrosine mimicry and cellular uptake. *Bioorganic & Medicinal Chemistry*, 22 :6387–6391, 2014.

[225] H. B. Rao, F. Zhu, G. B. Yang, Z. R. Li, and Y. Z. Chen. Update of PROFEAT : a web server for computing structural and physicochemical features of proteins and peptides from amino acid sequence. *Nucleic Acids Research*, 39 :W385–W390, 2011.

[226] Sarah Rauscher, Vytautas Gapsys, Michal J. Gajda, Markus Zweckstetter, Bert L. de Groot, and Helmut Grubmüller. Structural Ensembles of Intrinsically Disordered Proteins Depend Strongly on Force Field : A Comparison to Experiment. *Journal of Chemical Theory and Computation*, 11 :5513–5524, 2015.

[227] Barak Raveh, Nir London, and Ora Schueler-Furman. Sub-angstrom modeling of complexes between flexible peptides and globular proteins. *Proteins*, 78, 2010.

[228] Asghar M. Razavi, William M. Wuest, and Vincent A. Voelz. Computational Screening and Selection of Cyclic Peptide Hairpin Mimetics by Molecular Simulation and Kinetic Network Models. *Journal of Chemical Information and Modeling*, 54 :1425–1432, 2014.

[229] Jüri Reimand, Shirley Hui, Shobhit Jain, Brian Law, and Gary D. Bader. Domain-mediated protein interaction prediction : From genome to network. *FEBS Letters*, 586 :2751–2763, 2012.

[230] Sébastien Rey, Michael Acab, Jennifer L. Gardy, Matthew R. Laird, Katalin deFays, Christophe Lambert, and Fiona S. L. Brinkman. PSORTdb : a protein subcellular localization database for bacteria. *Nucleic Acids Research*, 33 :D164–D168, 2005.

[231] Michal Richman, Sarah Wilk, Marina Chemerovski, Sebastian K. T. S. Warmlander, Anna Wahlstrom, Astrid Graslund, and Shai Rahimipour. In Vitro and Mechanistic Studies of an Antiamyloidogenic Self-Assembled Cyclic d,l-α-Peptide Architecture. *J. Am. Chem. Soc.*, 135 :3474–3484, 2013.

[232] Kyle E. Roberts, Patrick R. Cushing, Prisca Boisguerin, Dean R. Madden, and Bruce R. Donald. Computational Design of a PDZ Domain Peptide Inhibitor that Rescues CFTR Activity. *PLOS Computational Biology*, 8 :e1002477, 2012.

[233] Carlos H. M. Rodrigues, Yoochan Myung, Douglas E. V. Pires, and David B. Ascher. mCSM-PPI2 : predicting the effects of mutations on protein–protein interactions. *Nucleic Acids Research*, 47 :W338–W344, 2019.

[234] Sandra Romero-Molina, Yasser B. Ruiz-Blanco, Mirja Harms, Jan Münch, and Elsa Sanchez-Garcia. PPI-Detect : A support vector machine model for sequence-based prediction of protein-protein interactions : PPI-Detect : A Support Vector Machine Model for Sequence-Based Prediction of Protein-Protein Interactions. *Journal of Computational Chemistry*, 40 :1233–1242, 2019.

[235] L. Roumen, D.J. Scholten, P. de Kruijf, I.J.P. de Esch, R. Leurs, and C. de Graaf. C(X)CR in silico : Computer-aided prediction of chemokine receptor–ligand interactions. *Drug Discov. Today*, 9 :e281–e291, 2012.

[236] Daniel P Ryan and Jacqueline M Matthews. Protein-protein interactions in human disease. *Current Opinion in Structural Biology*, 15 :441–446, 2005.

[237] Daniel P Ryan and Jacqueline M Matthews. Protein–protein interactions in human disease. *Current Opinion in Structural Biology*, 15(4) :441–446, 2005.

[238] Philip Ryan, Bhautikkumar Patel, Vivek Makwana, Hemant R. Jadhav, Milton Kiefel, Andrew Davey, Tristan A. Reekie, Santosh Rudrawar, and Michael Kassiou. Peptides, Peptidomimetics, and Carbohydrate–Peptide Conjugates as Amyloidogenic Aggregation Inhibitors for Alzheimer's Disease. *ACS Chem. Neurosci.*, 9 :1530–1551, 2018.

[239] Nahid Safari-Alighiarloo, Mohammad Taghizadeh, Mostafa Rezaei-Tavirani, Bahram Goliaei, and Ali Asghar Peyvandi. Protein-protein interaction networks (PPI) and complex diseases. *Gastroenterology and Hepatology From Bed to Bench*, 7(1) :17–31, 2014.

[240] Ruben Sanchez-Garcia, C O S Sorzano, J M Carazo, and Joan Segura. BIPSPI : a method for the prediction of partner-specific protein-protein interfaces. *Bioinformatics*, 35 :470–477, 2019.

[241] Brianda L. Santini and Martin Zacharias. Rapid in silico Design of Potential Cyclic Peptide Binders Targeting Protein-Protein Interfaces. *Front. Chem.*, 8 :573259, 2020.

[242] Debasree Sarkar, Tanmoy Jana, and Sudipto Saha. LMDIPred : A webserver for prediction of linear peptide sequences binding to SH3, WW and PDZ domains. *PLOS One*, 13 :e0200430, 2018.

[243] Matthew P. Sarnowski, Kyle P. Pedretty, Nicole Giddings, H. Lee Woodcock, and Juan R. Del Valle. Synthesis and β-sheet propensity of constrained N-amino peptides. *Bioorganic & Medicinal Chemistry*, 26 :1162–1166, 2018.

[244] R. M. Scarborough, M. A. Naughton, W. Teng, J. W. Rose, D. R. Phillips, L. Nannizzi, A. Arfsten, A. M. Campbell, and I. F. Charo. Design of potent and specific integrin antagonists. Peptide antagonists with high specificity for glycoprotein IIb-IIIa. *Journal of Biological Chemistry*, 268 :1066–1073, 1993.

[245] Robert M. Scarborough. Development of eptifibatide. *American Heart Journal*, 138 :1093–1104, 1999.

[246] Nalini Schaduangrat, Chanin Nantasenamat, Virapong Prachayasittikul, and Watshara Shoombuatong. Meta-iAVP : A Sequence-Based Meta-Predictor for Improving the Prediction of Antiviral Peptides Using Effective Feature Representation. *International Journal of Molecular Sciences*, 20 :5743, 2019.

[247] Christian E. Schafmeister, Julia Po, and Gregory L. Verdine. An All-Hydrocarbon Cross-Linking System for Enhancing the Helicity and Metabolic Stability of Peptides. *Journal of the American Chemical Society*, 122 :5891–5892, 2000.

[248] Christina E.M. Schindler, Sjoerd J. de Vries, and Martin Zacharias. Fully Blind Peptide-Protein Docking with pepATTRACT. *Structure*, 23 :1507–1515, 2015.

[249] J. Schlitter. Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem. Phys. Lett.*, 215 :617–621, 1993.

[250] Karsten Schrör and Artur-Aron Weber. Comparative Pharmacology of GP IIb/IIIa Antagonists. *Journal of Thrombosis and Thrombolysis*, 15 :71–80, 2003.

[251] Yuval Sedan, Orly Marcu, Sergey Lyskov, and Ora Schueler-Furman. Peptiderive server : derive peptide inhibitors from protein–protein interactions. *Nucleic Acids Research*, 44 :W536–W541, 2016.

[252] Xiaojian Shao, Chris S. H. Tan, Courtney Voss, Shawn S. C. Li, Naiyang Deng, and Gary D. Bader. A regression framework incorporating quantitative and negative interaction data improves quantitative prediction of PDZ domain–peptide interaction from primary sequence. *Bioinformatics*, 27 :383–390, 2011.

[253] Oz Sharabi, Jason Shirian, and Julia M. Shifman. Predicting affinity- and specificity-enhancing mutations at protein–protein interfaces. *Biochemical Society Transactions*, 41 :1166–1169, 2013.

[254] Chunquan Sheng, Guoqiang Dong, Zhenyuan Miao, Wannian Zhang, and Wei Wang. State-of-the-art strategies for targeting protein–protein interactions by small-molecule inhibitors. *Chemical Society Reviews*, 44 :8238–8259, 2015.

[255] Chunquan Sheng, Guoqiang Dong, Zhenyuan Miao, Wannian Zhang, and Wei Wang. State-of-the-art strategies for targeting protein–protein interactions by small-molecule inhibitors. *Chem. Soc. Rev.*, 44 :8238–8259, 2015.

[256] Timothy R. Siegert, Michael J. Bird, Kamlesh M. Makwana, and Joshua A. Kritzer. Analysis of Loops that Mediate Protein–Protein Interactions and Translation into Submicromolar Inhibitors. *J. Am. Chem. Soc.*, 138 :12876–12884, 2016.

[257] Stuart A. Sievers, John Karanicolas, Howard W. Chang, Anni Zhao, Lin Jiang, Onofrio Zirafi, Jason T. Stevens, Jan Münch, David Baker, and David Eisenberg. Structure-based design of non-natural amino-acid inhibitors of amyloid fibril formation. *Nature*, 475 :96–100, 2011.

[258] Thomas Simonson, Thomas Gaillard, David Mignon, Marcel Schmidt am Busch, Anne Lopes, Najette Amara, Savvas Polydorides, Audrey Sedano, Karen Druart, and Georgios Archontis. Computational protein design : the Proteus software and selected applications. *Journal of Computational Chemistry*, 34 :2472–2484, 2013.

[259] Diana P. Slough, Hongtao Yu, Sean M. McHugh, and Yu-Shan Lin. Toward accurately modeling N-methylated cyclic peptides. *Physical Chemistry Chemical Physics*, 19 :5377–5388, 2017.

[260] Colin A. Smith and Tanja Kortemme. Structure-Based Prediction of the Peptide Sequence Space Recognized by Natural and Synthetic PDZ Domains. *Journal of Molecular Biology*, 402 :460–474, 2010.

[261] George P. Smith. Phage Display : Simple Evolution in a Petri Dish (Nobel Lecture). *Angew. Chem. Int. Ed.*, 58 :14428–14437, 2019.

[262] George P. Smith and Valery A. Petrenko. Phage Display. *Chemical Reviews*, 97(2) :391–410, 1997.

[263] Miguel A. Soler, Alex Rodriguez, Anna Russo, Abimbola Feyisara Adedeji, Cedrix J. Dongmo Foumthuim, Cristina Cantarutti, Elena Ambrosetti, Loredana Casalis, Alessandra Corazza, Giacinto Scoles, Daniela Marasco, Alessandro Laio, and Sara Fortuna. Computational design of cyclic peptides for the customized oriented immobilization of globular proteins. *Phys. Chem. Chem. Phys.*, 19 :2740–2748, 2017.

[264] Andrea Spitaleri, Michela Ghitti, Silvia Mari, Luca Alberici, Catia Traversari, Gian-Paolo Rizzardi, and Giovanna Musco. Use of Metadynamics in the Design of isoDGR-Based αvβ3 Antagonists To Fine-Tune the Conformational Ensemble. *Angewandte Chemie International Edition*, 50 :1832–1836, 2011.

[265] Yerukala Srinivasulu, Jyun-Rong Wang, Kai-Ti Hsu, Ming-Ju Tsai, Phasit Charoenkwan, Wen-Lin Huang, Hui-Ling Huang, and Shinn-Ying Ho. Characte-

rizing informative sequence descriptors and predicting binding affinities of heterodimeric protein complexes. *BMC Bioinformatics*, 16 :S14, 2015.

[266] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314 :141–151, 1999.

[267] Haiying Sun, Zaneta Nikolovska-Coleska, Chao-Yie Yang, Dongguang Qian, Jianfeng Lu, Su Qiu, Longchuan Bai, Yuefeng Peng, Qian Cai, and Shaomeng Wang. Design of Small-Molecule Peptidic and Non-Peptidic Smac Mimetics. *Accounts of chemical research*, 41(10) :1264–1277, 2008.

[268] Thomas Szekely, Cécile Caumes, Olivier Roy, Sophie Faure, and Claude Taillefumier. α-Peptoïdes et composés apparentés : synthèse et contrôle de la conformation. *Comptes Rendus Chimie*, 16 :318–330, 2013.

[269] Ghazaleh Taherzadeh, Yuedong Yang, Tuo Zhang, Alan Wee-Chung Liew, and Yaoqi Zhou. Sequence-based prediction of protein-peptide binding sites using support vector machine. *Journal of Computational Chemistry*, 37 :1223–1229, 2016.

[270] Soon-Heng Tan, Willy Hugo, Wing-Kin Sung, and See-Kiong Ng. A correlated motif approach for finding short linear motifs from protein interaction networks. *BMC Bioinformatics*, 7 :502, 2006.

[271] Yaw Sing Tan, David P. Lane, and Chandra S. Verma. Stapled peptide design : principles and roles of computation. *Drug Discovery Today*, 21 :1642–1653, 2016.

[272] Masakazu Tanaka. Design and Synthesis of Chiral α,α-Disubstituted Amino Acids and Conformational Study of Their Oligopeptides. *Chemical and Pharmaceutical Bulletin*, 55 :349–358, 2007.

[273] Hua Tang, Zhen-Dong Su, Huan-Huan Wei, Wei Chen, and Hao Lin. Prediction of cell-penetrating peptides with feature selection techniques. *Biochemical and Biophysical Research Communications*, 477 :150–154, 2016.

[274] Cinzia Tarsia, Alberto Danielli, Francesca Florini, Paolo Cinelli, Stefano Ciurli, and Barbara Zambelli. Targeting Helicobacter pylori urease activity and maturation : In-cell high-throughput approach for drug discovery. *Biochimica et Biophysica Acta*, 1862 :2245–2253, 2018.

[275] Chuan Tian, Koushik Kasavajhala, Kellon A. A. Belfon, Lauren Raguette, He Huang, Angela N. Migues, John Bickel, Yuzhang Wang, Jorge Pincay, Qin Wu, and Carlos Simmerling. ff19SB : Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.*, 16 :528–552, 2020.

[276] Linh Tran, Tobias Hamp, and Burkhard Rost. ProfPPIdb : Pairs of physical protein-protein interactions predicted for entire proteomes. *PLOS One*, 13 :e0199988, 2018.

[277] Daniela Trisciuoglio, Maria Grazia Tupone, Marianna Desideri, Marta Di Martile, Chiara Gabellini, Simonetta Buglioni, Matteo Pallocca, Gabriele Alessandrini, Simona D'Aguanno, and Donatella Del Bufalo. BCL-XL overexpression promotes tumor progression-associated properties. *Cell Death & Disease*, 8(12) :1–15, 2017. Number : 12 Publisher : Nature Publishing Group.

[278] Oleg Trott and Arthur J. Olson. AutoDock Vina : improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J. Comput. Chem.*, 31(2) :455–461, 2010.

[279] Nurcan Tuncbag, Attila Gursoy, and Ozlem Keskin. Identification of computational hot spots in protein interfaces : combining solvent accessibility and inter-residue potentials improves the accuracy. *Bioinformatics*, pages 1513–1520, 2009.

[280] Chun-Wei Tung, Matthias Ziehm, Andreas Kämper, Oliver Kohlbacher, and Shinn-Ying Ho. POPISK : T-cell reactivity prediction using support vector machines and string kernels. *BMC Bioinformatics*, 12 :446, 2011.

[281] Eugene Varfolomeev, John W. Blankenship, Sarah M. Wayson, Anna V. Fedorova, Nobuhiko Kayagaki, Parie Garg, Kerry Zobel, Jasmin N. Dynek, Linda O. Elliott, Heidi J. A. Wallweber, John A. Flygare, Wayne J. Fairbrother, Kurt Deshayes, Vishva M. Dixit, and Domagoj Vucic. IAP antagonists induce autoubiquitination of c-IAPs, NF-kappaB activation, and TNFalpha-dependent apoptosis. *Cell*, 131 :669–681, 2007.

[282] Daniel F. Veber, Stephen R. Johnson, Hung-Yuan Cheng, Brian R. Smith, Keith W. Ward, and Kenneth D. Kopple. Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. *J. Med. Chem.*, 45 :2615–2623, 2002.

[283] Alexander A. Vinogradov, Yizhen Yin, and Hiroaki Suga. Macrocyclic Peptides as Drug Candidates : Recent Progress and Remaining Challenges. *J. Am. Chem. Soc.*, 141 :4167–4181, 2019.

[284] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0 : Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17 :261–272, 2020.

[285] Randi Vita, Swapnil Mahajan, James A Overton, Sandeep Kumar Dhanda, Sheridan Martini, Jason R Cantrell, Daniel K Wheeler, Alessandro Sette, and Bjoern Peters. The Immune Epitope Database (IEDB) : 2018 update. *Nucleic Acids Research*, 47 :D339–D343, 2019.

[286] Amanda E. Wakefield, William M. Wuest, and Vincent A. Voelz. Molecular Simulation of Conformational Pre-Organization in Cyclic RGD Peptides. *Journal of Chemical Information and Modeling*, 55 :806–813, 2015.

[287] Rike Wallbrecher, Luc Depré, Wouter P. R. Verdurmen, Petra H. Bovée-Geurts, Richard H. van Duinkerken, Mariët J. Zekveld, Peter Timmerman, and Roland Brock. Exploration of the Design Principles of a Cell-Penetrating Bicylic Peptide Scaffold. *Bioconjugate Chemistry*, 25(5) :955–964, 2014.

[288] C. Paul Wang, Neil R. Hartman, Raman Venkataramanan, Ian Jardine, Fu-Tyan Lin, Joseph E. Knapp, Thomas E. Starzl, and Gilbert J. Burckart. Isolation of 10 Cyclosporine Metabolites from Human Bile. *Drug Metab. Dispos.*, 17 :292–296, 1989.

[289] Conan K Wang and David J. Craik. Cyclic peptide oral bioavailability : Lessons from the past : Cyclic Peptide Oral Bioavailability. *Biopolymers*, 106 :901–909, 2016.

[290] Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case. Development and testing of a general amber force field. *J. Comput. Chem.*, 25 :1157–1174, 2004.

[291] Lin Wang, Zhi-Ping Liu, Xiang-Sun Zhang, and Luonan Chen. Prediction of hot spots in protein interfaces using a random forest model with hybrid features. *Protein Engineering, Design & Selection*, 25 :119–126, 2012.

[292] Lincong Wang, Yaqin Hou, Haihua Quan, Weiwei Xu, Yongli Bao, Yuxin Li, Yuan Fu, and Shuxue Zou. A compound-based computational approach for the accurate determination of hot spots. *Protein Science*, 22 :1060–1070, 2013.

[293] Qiuming Wang, Guizhao Liang, Mingzhen Zhang, Jun Zhao, Kunal Patel, Xiang Yu, Chao Zhao, Binrong Ding, Ge Zhang, Feimeng Zhou, and Jie Zheng. De Novo Design of Self-Assembled Hexapeptides as β-Amyloid (Aβ) Peptide Inhibitors. *ACS Chem. Neurosci.*, 5 :972–981, 2014.

[294] Xiaoshan Shayna Wang, Peng-Hsun Chase Chen, J. Trae Hampton, Jeffery M. Tharp, Catrina A. Reed, Sukant K. Das, Duen-Shian Wang, Hamed S. Hayatshahi, Yang Shen, Jin Liu, and Wenshe Ray Liu. A Genetically Encoded, Phage-Displayed Cyclic-Peptide Library. *Angew. Chem. Int. Ed.*, 58 :15904–15909, 2019.

[295] Andrew M. Watkins and Paramjit S. Arora. Structure-based inhibition of protein-protein interactions. *European Journal of Medicinal Chemistry*, 94 :480–488, 2015.

[296] Leyi Wei, PengWei Xing, Ran Su, Gaotao Shi, Zhanshan Sam Ma, and Quan Zou. CPPred-RF : A Sequence-based Predictor for Identifying Cell-Penetrating Peptides and Their Uptake Efficiency. *Journal of Proteome Research*, 16 :2044–2053, 2017.

[297] Leyi Wei, Chen Zhou, Huangrong Chen, Jiangning Song, and Ran Su. ACPred-FL : a sequence-based predictor using effective feature representation to improve the prediction of anti-cancer peptides. *Bioinformatics*, 34 :4007–4016, 2018.

[298] Leyi Wei, Chen Zhou, Ran Su, and Quan Zou. PEPred-Suite : improved and robust prediction of therapeutic peptides using adaptive feature representation learning. *Bioinformatics*, 35 :4272–4280, 2019.

[299] Andrew M. White and David J. Craik. Discovery and optimization of peptide macrocycles. *Expert Opinion on Drug Discovery*, 11 :1151–1163, 2016.

[300] Kanin Wichapong, Hessel Poelman, Bogac Ercig, Johana Hrdinova, Xiaosong Liu, Esther Lutgens, and Gerry AF Nicolaes. Rational modulator design by exploitation of protein–protein complex structures. *Future Medicinal Chemistry*, 11 :1015–1033, 2019.

[301] William G.T. Willats. Phage display : practicalities and prospects. *Plant Molecular Biology*, 50(6) :837–854, 2002.

[302] Beili Wu, Ellen YT Chien, Clifford D. Mol, Gustavo Fenalti, Wei Liu, Vsevolod Katritch, Ruben Abagyan, Alexei Brooun, Peter Wells, and F. Christopher Bi. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science*, 330 :1066–1071, 2010.

[303] Hongxi Wu, Yalan Liu, Mingrong Guo, Jingli Xie, and XiaMin Jiang. A Virtual Screening Method for Inhibitory Peptides of Angiotensin I–Converting Enzyme. *Journal of Food Science*, 79 :C1635–C1642, 2014.

[304] Xianghong Wu, Punit Upadhyaya, Miguel A. Villalona-Calero, Roger Briesewitz, and Dehua Pei. Inhibition of Ras–effector interactions by cyclic peptides. *Med. Chem. Commun.*, 4 :378–382, 2013.

[305] Jun-Feng Xia, Xing-Ming Zhao, Jiangning Song, and De-Shuang Huang. APIS : accurate prediction of hot spots in protein interfaces by combining protrusion index with solvent accessibility. *BMC bioinformatics*, 11 :174, 2010.

[306] Jian-Ping Xiong, Thilo Stehle, Rongguang Zhang, Andrzej Joachimiak, Matthias Frech, Simon L. Goodman, and M. Amin Arnaout. Crystal Structure of

the Extracellular Segment of Integrin αVβ3 in Complex with an Arg-Gly-Asp Ligand. *Science*, 296 :151–155, 2002.

[307] Lei Xu, Guangmin Liang, Longjie Wang, and Changrui Liao. A Novel Hybrid Sequence-Based Model for Identifying Anticancer Peptides. *Genes*, 9 :158, 2018.

[308] Li C Xue, Drena Dobbs, and Vasant Honavar. HomPPI : a class of sequence homology based protein-protein interface prediction methods. *BMC Bioinformatics*, 12 :244, 2011.

[309] Chengfei Yan, Xianjin Xu, and Xiaoqin Zou. Fully Blind Docking at the Atomic Level for Protein-Peptide Complex Structure Prediction. *Structure*, 24 :1842–1853, 2016.

[310] Yumeng Yan, Zeyu Wen, Xinxiang Wang, and Sheng-You Huang. Addressing recent docking challenges : A hybrid strategy to integrate template-based and free protein-protein docking. *Proteins : Structure, Function, and Bioinformatics*, 85 :497–512, 2017.

[311] Wei Yang, Qi Zhang, Changsheng Zhang, Annan Guo, Yanyan Wang, Hantian You, Xiaoling Zhang, and Luhua Lai. Computational design and optimization of novel d-peptide TNFα inhibitors. *FEBS Letters*, 593 :1292–1302, 2019.

[312] Xue Qing Yang, Ji Yuan Liu, Xian Chun Li, Mao Hua Chen, and Ya Lin Zhang. Key Amino Acid Associated with Acephate Detoxification by Cydia pomonella Carboxylesterase Based on Molecular Dynamics with Alanine Scanning and Site-Directed Mutagenesis. *Journal of Chemical Information and Modeling*, 54 :1356–1370, 2014.

[313] Eugene Yedvabny, Paul S. Nerenberg, Clare So, and Teresa Head-Gordon. Disordered Structural Ensembles of Vasopressin and Oxytocin and Their Mutants. *Journal of Physical Chemistry B*, 119 :896–905, 2015.

[314] Hongtao Yu and Yu-Shan Lin. Toward structure prediction of cyclic peptides. *Physical Chemistry Chemical Physics*, 17 :4210–4219, 2015.

[315] Zhipeng Yu, Yue Fan, Wenzhu Zhao, Long Ding, Jianrong Li, and Jingbo Liu. Novel Angiotensin-Converting Enzyme Inhibitory Peptides Derived from Oncorhynchus mykiss Nebulin : Virtual Screening and In Silico Molecular Docking Study. *Journal of Food Science*, 83 :2375–2383, 2018.

[316] Zhipeng Yu, Ruotong Kan, Sijia Wu, Hui Guo, Wenzhu Zhao, Long Ding, Fuping Zheng, and Jingbo Liu. Xanthine oxidase inhibitory peptides derived from tuna protein : virtual screening, inhibitory activity, and molecular mechanisms. *Journal of the Science of Food and Agriculture*, 2020.

[317] K. Yugandhar and M. Michael Gromiha. Feature selection and classification of protein–protein complexes based on their binding affinities using machine learning approaches. *Proteins : Structure, Function, and Bioinformatics*, 82 :2088–2096, 2014.

[318] K Yugandhar and M Michael Gromiha. Analysis of Protein-Protein Interaction Networks Based on Binding Affinity. *Current Protein and Peptide Science*, 17 :72–81, 2016.

[319] Marlon S. Zambrano-Mila, Karen Elizabeth Sánchez Blacio, and Nelson Santiago Vispo. Peptide Phage Display : Molecular Principles and Biomedical Applications. *Ther. Innov. Regul. Sci.*, 54 :308–317, 2020.

[320] Changsheng Zhang, Qi Shen, Bo Tang, and Luhua Lai. Computational Design of Helical Peptides Targeting TNFα. *Angewandte Chemie International Edition*, 52 :11059–11062, 2013.

[321] Ning Zhang, Yuting Chen, Haoyu Lu, Feiyang Zhao, Roberto Vera Alvarez, Alexander Goncearenco, Anna R. Panchenko, and Minghui Li. MutaBind2 : Predicting the Impacts of Single and Multiple Mutations on Protein-Protein Interactions. *iScience*, 23 :100939, 2020.

[322] Ren Zhang, Hong-Yu Ou, and Chun-Ting Zhang. DEG : a database of essential genes. *Nucleic Acids Research*, 32 :D271–D272, 2004.

[323] Xuan Zhang, Dinesh Thummuri, Xingui Liu, Wanyi Hu, Peiyi Zhang, Sajid Khan, Yaxia Yuan, Daohong Zhou, and Guangrong Zheng. Discovery of PROTAC BCL-XL degraders as potent anticancer agents with low on-target platelet toxicity. *European Journal of Medicinal Chemistry*, 192 :112186, 2020.

[324] Yuqi Zhang and Michel F. Sanner. AutoDock CrankPep : combining folding and docking to predict protein-peptide complexes. *Bioinformatics*, 35 :5121–5127, 2019.

[325] Yuqi Zhang and Michel F. Sanner. Docking Flexible Cyclic Peptides with AutoDock CrankPep. *J. Chem. Theory Comput.*, 15 :5161–5168, 2019.

[326] Juan Zhao, Baohua Yin, Haibo Sun, Laixue Pang, and Jianzhong Chen. Identifying hot spots of inhibitor-CDK2 bindings by computational alanine scanning. *Chemical Physics Letters*, 747 :137329, 2020.

[327] Zijuan Zhao, Zhenling Peng, and Jianyi Yang. Improving Sequence-Based Prediction of Protein–Peptide Binding Residues by Introducing Intrinsic Disorder and a Consensus Method. *Journal of Chemical Information and Modeling*, 58 :1459–1468, 2018.

[328] Fan Zheng, Heather Jewell, Jeremy Fitzpatrick, Jian Zhang, Dale F. Mierke, and Gevorg Grigoryan. Computational Design of Selective Peptides to Discriminate between Similar PDZ Domains in an Oncogenic Pathway. *Journal of Molecular Biology*, 427 :491–510, 2015.

[329] Fan Zheng, Jian Zhang, and Gevorg Grigoryan. Tertiary Structural Propensities Reveal Fundamental Sequence/Structure Relationships. *Structure*, 23 :961–971, 2015.

[330] Pei Zhou, Bowen Jin, Hao Li, and Sheng-You Huang. HPEPDOCK : a web server for blind peptide–protein docking based on a hierarchical algorithm. *Nucleic Acids Research*, 46 :W443–W450, 2018.

[331] Jian Zhu, Sen Wei, Linchen Huang, Qi Zhao, Haichao Zhu, and Anwei Zhang. Molecular modeling and rational design of hydrocarbon-stapled/halogenated helical peptides targeting CETP self-binding site : Therapeutic implication for atherosclerosis. *Journal of Molecular Graphics and Modelling*, 94 :107455, 2020.

[332] Katja C. Zimmermann and Douglas R. Green. How cells die : Apoptosis pathways. *Journal of Allergy and Clinical Immunology*, 108(4, Supplement) :S99–S103, 2001.

[333] Alessandro Zorzi, Kaycie Deyle, and Christian Heinis. Cyclic peptide therapeutics : past, present and future. *Curr. Opin. Chem. Biol.*, 38 :24–29, 2017.