

TABLE DES MATIERES

INTRODUCTION GENERALE..... - 1 -

CHAPITRE I : MODELISATION ET SIMULATION AU SERVICE DE LA BIOLOGIE - 7 -

Introduction	- 8 -
I-1. Contexte.....	- 8 -
I-1.1. Les enjeux.....	- 9 -
I-1.1.1. Recherche fondamentale	- 9 -
I-1.1.2. Pharmacologie.....	- 10 -
I-1.1.3. Nanobiotechnologies.....	- 11 -
I-1.2. Les biopolymères : une cible privilégiée	- 12 -
I-1.2.1. Les acides nucléiques	- 12 -
I-1.2.2. Les Protéines	- 15 -
I-2. La nature des forces d'interaction.....	- 18 -
I-2.1. Liaisons fortes	- 18 -
I-2.1.1. Liaison covalente	- 19 -
I-2.1.2 .Liaison ionique.....	- 19 -
I-2.2. Liaisons faibles	- 20 -
I-2.2.1. Interaction coulombienne	- 21 -
I-2.2.2. Forces de dispersion et de répulsion	- 22 -
I-2.2.3. Liaison hydrogène	- 24 -
I-2.2.4. L'effet hydrophobe.....	- 26 -
I-3. Modélisation des systèmes biologiques	- 27 -
I-3.1. Les modèles quantiques.....	- 27 -
I-3.2. La mécanique moléculaire	- 29 -
I-3.2.1. Champ de force	- 29 -
I-3.2.2. Optimisation de la structure.....	- 31 -
I-3.2.3. Le solvant.....	- 32 -
I-3.3. Modèles dynamiques	- 34 -
I-3.3.1. Modélisation par dynamique moléculaire	- 34 -
I-3.3.2. Simulation de type Monte-Carlo	- 34 -
I-3.4. Le problème de la flexibilité moléculaire	- 36 -

TABLE DES MATIERES

I-4. Comment prendre en compte la flexibilité moléculaire ?.....	- 38 -
I-4.1. Les méthodes usuelles de simulation	- 39 -
I-4.1.1. Simulations dynamiques	- 39 -
I-4.1.2. Calcul des modes normaux.....	- 39 -
I-4.1.3. Autres méthodes	- 40 -
I-4.1.4. Limitations	- 41 -
I-4.2. La flexibilité dans les logiciels de Docking	- 42 -
I-4.2.1. Verrous	- 43 -
I-4.2.2. Représentation du système.....	- 44 -
I-4.2.3. Les algorithmes de recherche	- 45 -
I-4.2.4. Les fonctions de scoring	- 47 -
I-5. Vers un traitement de la flexibilité relative aux interactions moléculaires.....	- 48 -
I-6. Conclusion	- 52 -

CHAPITRE 2 : METHODOLOGIE..... - 53 -

INTRODUCTION	- 54 -
II-1. Recherche d'un mode d'introduction des déformations dans un calcul de docking.....	- 54 -
II-1.1. Les modes normaux	- 55 -
II-1.1.1. Origine	- 55 -
II-1.1.2. Les modes normaux dans une procédure de docking	- 56 -
II-1.2. Les Modes Statiques	- 58 -
II-2. Le calcul des Modes Statiques	- 60 -
II-3. Validation de la méthode	- 62 -
II-4. Avantages et limites de la méthode	- 64 -
II-5. Les outils développés à partir des modes statiques	- 66 -
II-5.1. Description de la flexibilité moléculaire	- 67 -
II-5.1.1. Optimisation de la force	- 67 -
II-5.1.2. Modes de pincements	- 70 -
II-5.1.3. Grandeurs caractéristiques	- 71 -
II-5.2. Arrimage d'un ion métallique	- 72 -
II-6. Conclusion	- 74 -

CHAPITRE 3 : RESULTATS	- 77 -
Introduction	- 78 -
III-1. Présentation des peptides	- 78 -
III-1-1. Le peptide Amyloïde- β	- 79 -
III-1.2. Le peptide modèle GNNQQNY	- 83 -
III-2. Caractérisation de la flexibilité	- 85 -
III-2.1. Cas du peptide modèle GNNQQNY	- 85 -
III-2.1.1. Calculs sur les monomères	- 85 -
III-2.1.2. Calculs sur les dimères	- 92 -
I-2.2. Cas de l'amyloïde β	- 95 -
III-2.3. Représentation des zones flexibles des molécules	- 97 -
III-2.3.1. Propagation de la déformation	- 101 -
III-2.3.2. Comportement allostérique	- 105 -
III-2.3.3. Relation interactions/oligomérisation.....	- 107 -
III-2.5. Conclusion	- 112 -
III-3. Application au docking	- 113 -
III-3.1. Description des conformations utilisées	- 113 -
III-3.2 Sonde électrostatique	- 116 -
III-3.2.1. Exploration de l'espace conformationnel de la structure A β (1-16) en interaction avec un ion métallique ..	- 116 -
III-3.2.2. Repliement du peptide à partir de la structure sans zinc.....	- 122 -
III-3. Conclusions.....	127
 CONCLUSION GENERALE	 - 128 -
 ANNEXE A.....	 - 131 -
 ANNEXE B.....	 - 134 -
 BIBLIOGRAPHIE.....	 - 137 -

Liste des figures

CHAPITRE 1

Figure 1 - exemple de nucléotide : désoxy(ribo)guanosine monophosphate.....	- 13 -
Figure 2 – schéma d'un acide nucléique. On distingue les deux parties de la structure : le squelette sucre-phosphate et les bases azotées.....	- 14 -
Figure 3 : processus de formation d'une liaison peptidique	- 15 -
Figure 4 : définition d'un angle dièdre. Le plan \mathbf{P}_1 est défini par les atomes A, B, C et le plan \mathbf{P}_2 par B, C, D. L'angle dièdre θ est alors l'angle entre \mathbf{P}_1 et \mathbf{P}_2	- 16 -
Figure 5 : formes résonantes de la liaison peptidique.....	- 17 -
Figure 6 : structure d'une protéine. On peut voir les angles de torsion Φ , Ψ et ω , ainsi que l'on a quatre atomes dans le même plan P, dû à la planarité de la liaison peptidique	- 17 -
Figure 7 : Processus de formation d'une liaison ionique illustré avec l'exemple de NaCl. A) suite à un apport externe d'énergie, l'atome de sodium perd son électron célibataire. B) l'électron libre est alors récupéré par le chlore. C) les deux ions de signes opposés résultant de ce transfert de charge ont alors leur couche de valence remplie et s'attirent. D) on observe alors la formation d'une liaison ionique, exothermique	- 20 -
Figure 8 : Potentiel de Lennard-Jones pour des valeurs de $\epsilon = -0.13$ eV et de $\sigma = 3.24$ Å.....	- 24 -
Figure 9 : la liaison hydrogène est une interaction entre deux dipôles, l'un formé par le donneur et l'atome d'hydrogène, l'autre par l'accepteur et le carbone auquel il est lié.	- 25 -
Figure 10 : structures secondaires des protéines. De gauche à droite : feuillet β antiparallèle ; feuillet β parallèle ; hélice α . Les liaisons hydrogènes sont représentées en pointillé. Les flèches indiquent le sens N-terminal vers C-terminal (30).....	- 26 -
Figure 11 : modèle TIP3P de la molécule d'eau.....	- 33 -
Figure 12 – Schéma des procédures de simulation.....	- 36 -
Figure 13 : surface de Connolly. Les rayons de Van der Waals définissent le volume de Van der Waals sur lequel on fait rouler une sphère. La surface résultante est la surface de Connolly.....	- 44 -

CHAPITRE 2

Figure 14 : figures illustrant les déplacements atomiques pour deux modes normaux de vibration de la protéine BPTI. A) squelette (traits gras) et chaînes latérales (traits fins), les carbones α sont numérotés. B) vecteurs représentant les déplacements atomiques pour un mode de haute fréquence (118.8 cm^{-1}) : les mouvements atomiques sont localisés. C) vecteurs représentant les déplacements atomiques pour un mode de basse fréquence (6.9 cm^{-1}) : les mouvements atomiques sont délocalisés (source ref. 97).	- 56 -
Figure 15 : forces employées lors d'un calcul de modes de pincement.....	- 70 -

CHAPITRE 3

Figure 16 : hypothèse de la cascade amyloïde (source : ref. (140)).	- 80 -
Figure 17 : feuillet β composé de deux peptides GNNQQNY.	- 84 -
Figure 18 : Premières étapes de l'agrégation du peptide modèle GNNQQNY.	- 84 -
Figure 19 : Les 4 structures utilisées dans les calculs sur le peptide modèle GNNQQNY.	- 86 -
Figure 20 : Diagrammes représentant le ΔRMSD sur les atomes du squelette entre les monomères B, R1, R2 et A et un brin du dimère.	- 87 -
Figure 21 : diagrammes représentant le ΔRMSD calculé sur les atomes formant des liaisons H entre les monomères β , R1, R2 et α (respectivement A, B, C et D) et le brin 1 du feuillet.	- 89 -
Figure 22 : feuillets parallèle (à gauche) et antiparallèle (à droite) obtenus avec AMBER8	- 92 -
Figure 23 : calcul du RMSD en fonction des modes de déformation pour les feuillets parallèle et antiparallèle.....	- 93 -
Figure 24 : Evolution du RMSD en fonction de la température lors de la simulation effectuée avec AMBER 8.....	- 94 -
Figure 25 : Les trois conformations utilisées dans cette étude. A) structure en hélice α (noté α). B) peptide seul en conformation β (noté β_{str}). C) feuillet β composé de trois brins. Le brin du milieu (indiqué par la flèche rouge) est noté β_{sh}	- 95 -
Figure 26 : valeurs maximales de déformation de l'angle ϕ_2 pour les trois structures α (A), β_{str} (B) et β_{sh} (C). - 97 -	- 97 -
Figure 27 : Déformations maximales des angles ϕ du squelette pour les trois conformations. De gauche à droite : α , β_{str} , β_{sh}	- 99 -

TABLE DES MATIERES

Figure 28 : Déformations maximales des angles ψ du squelette pour les trois conformations. De gauche à droite : α , β_{str} , β_{sh}	- 100 -
Figure 29 : Résidus 3 à 9 du peptide Amyloïde- β . La coloration représente l'amplitude de la déformation de l'angle ϕ 5. A gauche : conformation β_{str} , sur laquelle on observe 3 liaisons hydrogène entre les résidus Arg5 et Asp7. A droite : conformation β_{sh} , sur laquelle on n'observe plus de liaisons hydrogène.	- 103 -
Figure 30 : peptide GNNQQNY : 2 monomères en feuillet β antiparallèle. A) structure 3D du feuillet. B) représentation des amplitudes de déformation (en Å) lorsque l'on applique une force le long de la liaison hydrogène (pointillé) comme le suggère la flèche.	- 104 -
Figure 31 : zoom sur les diagrammes correspondant aux angles ϕ et ψ de la configuration β_{str}	- 105 -
Figure 32 : représentation des amplitudes de déformation, en degrés, par atome lorsque l'on veut obtenir une variation maximale des angles ϕ 15 (figure de gauche) et ψ 15 (figure de droite).	- 106 -
Figure 33 : A) Répartition dans le plan (ϕ, ψ) des valeurs maximales de déformation pour le résidu Ala2. B) Tracé Ramachandran de tous les résidus exceptés la glycine (182).	- 107 -
Figure 34 : Nombre de modes par résidu déformant les couples (ϕ, ψ) de la conformation α vers des valeurs caractéristiques de la forme β	- 108 -
Figure 35 : Probabilité par résidu de former un feuillet β calculées (a) sur un monomère et (b) sur un dimère. Ces résultats ont été obtenus par Dong et al.	- 109 -
Figure 36 : Représentation simplifiée des diagrammes des figures 24 et 25. La numérotation horizontale correspond au résidu sur lequel la force est appliquée et la numérotation verticale correspond à l'angle dièdre.....	- 110 -
Figure 37 : structures apo (A) et liée (B) du peptide A β (1-16) optimisées avec le logiciel AMBER8 (champ de force ff03).....	- 114 -
Figure 38 : site de fixation du zinc. Seuls les résidus impliqués sont représentés. Les liaisons sont indiquées en orange.	- 115 -
Figure 39 : diagramme des RMSDT dans le plan (ϕ, ψ).	- 117 -
Figure 40 : Evolution de l'énergie du peptide A β (1-16) en fonction de α lorsque l'ion s'accroche sur le site indiqué par la flèche.	- 119 -
Figure 41 : Evolution de l'énergie du peptide A β (1-16) en fonction de α lorsque l'ion s'accroche sur le site indiqué par la flèche.	- 120 -
Figure 42 : Evolution de l'énergie du peptide A β (1-16) en fonction de α lorsque l'ion s'accroche sur le site	

indiqué par la flèche.....	- 121 -
Figure 43 : diagramme représentant l'énergie et l'évolution de la structure apo vers la structure liée. Les croix représentent les 19 structures issues du fichier PDB.....	- 123 -
Figure 44 : Structure correspondant au minimum global d'énergie calculée avec la sonde électrostatique.	- 124 -
Figure 45 :Agrandissement d'une partie du diagramme de la figure 40. A) Les croix représentent les points correspondant aux structures expérimentales. B) Valeurs des énergies de ces structures expérimentales.-	- 124 -
Figure 46 : superposition d'une structure issue de l'expérience (orange) et d'une générée par la sonde électrostatique (vert). Seuls les résidus 6, 11, 13 et 14 sont représentés.	- 125 -

Liste des tableaux

Tableau 1 : paramètres de Lennard-Jones pour les atomes qui composent majoritairement les biomolécules. σ est en Å et ϵ en eV (41).....	- 24 -
Tableau 2 : les trois catégories de liaison hydrogène (6). La longueur est la distance entre l'atome d'hydrogène et l'accepteur ; l'angle est formé par les trois atomes formant la liaison hydrogène : l'accepteur, l'hydrogène et le donneur.	- 25 -
Tableau 3 : séquences des trois peptides que nous avons utilisés pour effectuer nos calculs.	- 79 -
Tableau 4 : Détail des bandes des diagrammes de la figure 22 : nom, numéros des atomes correspondants et leur situation sur le peptide. CL signifie Chaîne Latérale.	- 90 -
Tableau 5 : valeur de la déformation moyenne pour chacune des conformations du peptide Amyloïde- β : α , β_{str} et β_{sh}	- 96 -
Tableau 6 : nombre moyen de résidus possédant au moins un atome participant à la déformation.	- 102 -
Tableau 7 : valeurs des longueurs de liaisons, en Å, entre Zn^{2+} et les atomes auxquels il se fixe, mesurées sur la structure expérimentale et après optimisation de la géométrie.	- 115 -

INTRODUCTION GENERALE

La biologie connaît actuellement un développement extrêmement rapide, notamment dans le domaine de la biologie moléculaire. De plus en plus d'efforts pluridisciplinaires se concentrent sur ces questions, au carrefour « nano-bio-informatique ».

Nos travaux menés au LAAS au sein du groupe N2IS relèvent de cette dynamique. Ils concernent principalement des questions d'interactions molécules à molécules, surface à molécules et leur modélisation informatique. Nous nous appuyons sur le concept des « modes statiques », récemment introduit au laboratoire, pour rendre compte des déformations moléculaires qui sont essentielles dans le développement d'outils de simulation prédictifs performants.

Ces interactions tiennent un rôle fondamental dans tous les processus biologiques : elles permettent aux organismes vivants de réguler leur métabolisme, ce qui signifie que les processus essentiels de la vie reposent sur un ensemble d'interactions entre biomolécules. Il est donc, de toute évidence, important de connaître ces phénomènes responsables du fonctionnement de notre organisme. De nombreux travaux ont été effectués dans ce sens depuis les années 1930, qui ont vu naître les débuts de la biologie moléculaire (1) (2) (3) (4) (5).

L'étude de ces biomolécules commence par la détermination de leur structure tridimensionnelle et des changements conformationnels qu'elles peuvent subir. Ces dernières décennies, le développement de techniques expérimentales telles que la cristallographie par rayons X, ou la résonance magnétique nucléaire, ont permis de déterminer un grand nombre de structures : à titre d'exemple, on ne connaissait à la fin des années 70, qu'environ 70 structures de protéines, alors qu'aujourd'hui on en recense plus de 40000 dans la base de données Protein Data Bank (6). Actuellement, 6000 à 7000 nouvelles structures sont ajoutées chaque année sur cette base de données.

Mais la détermination de ces structures ne suffit pas à comprendre ces mécanismes biologiques complexes d'interaction entre biomolécules. Les phénomènes d'amarrage entre macromolécules, communément désigné par le terme anglais *docking*, sont un sujet d'études fondamentales et appliquées, notamment dans le milieu de la recherche pharmaceutique :

- Les approches expérimentales permettent d'obtenir des informations sur des structures

Introduction Générale

“figées” et les bilans réactionnels. Mais il reste difficile de répondre à un certain nombre de questions fondamentales pour la compréhension des modes évolutifs de ces phénomènes, telles que : quelles sont les mécanismes d’interaction? A quel endroit privilégié ont-ils lieu? Comment se développent-ils ? etc... C’est pour proposer des réponses à ces questions que des outils de simulation informatique ont été développés.

- La simulation numérique en biologie tient alors deux rôles : (i) elle vient compléter le point de vue de l’observateur expérimental (interprétation) et (ii) elle cherche à prévoir les comportements sans passer par l’expérience (prédiction).

Du point de vue de la modélisation, deux voies d’étude sont ouvertes :

- L’étude d’une molécule isolée : l’espace conformationnel d’une molécule est balayé afin de déterminer les états les plus stables de cette dernière et les changements conformationnels qu’elle peut subir (par la dynamique moléculaire par exemple, ou encore par des simulations de type Monte Carlo).
- L’étude de deux (ou plusieurs) molécules en interaction : l’objectif est alors de prédire la formation et la structure tridimensionnelle des complexes formés de deux ou plusieurs molécules, grâce à des algorithmes de docking.

Les outils de modélisation permettent aujourd’hui, grâce au développement des moyens de calcul, une étude de plus en plus précise de l’espace conformationnel d’une molécule : changements conformationnels (7), cinétique de réactions (8) (9), effets dûs au milieu (pH, solvation), chemins réactionnels (10) (11).

En revanche, la prédiction du docking entre macromolécules est plus complexe à mettre en œuvre : ce sont des objets constamment en mouvement, dont la structure dans l’espace change à la moindre interaction avec une autre molécule. Cette flexibilité, caractéristique des molécules biologiques, est une donnée difficile à prendre en compte dans les outils de simulation.

Le premier modèle d’amarrage moléculaire a été pensé par le chimiste allemand Heinz Emil Fischer à la fin du XIXe siècle : le modèle “clef-serrure”. C’est un modèle essentiellement géométrique, où une molécule dans une conformation « clé » s’associe à une molécule dans une conformation « serrure » correspondante. Les deux molécules mises en jeu

étaient alors considérées comme rigides, ce qui s'est avéré ne pas refléter la réalité. Quelques décennies plus tard, le principe "clef-serrure" a en effet été réfuté pour laisser la place à une nouvelle approche plus réaliste : le principe "induced-fit", énoncé en 1958 par le biologiste D.E. Koshland, où les molécules changent de conformation au cours de la phase d'amarrage. Plus récemment, dans un article paru en 2009, Ruth Nussinov et Peter E. Wright ont présenté une méthode alternative (12). Ils ont associé ce principe d'« induced-fit » à un autre principe introduit auparavant par McCammon et Carlson : la sélection conformationnelle. Ce principe postule que toutes les conformations d'une protéine existent et le ligand sélectionne la plus favorable et a été validé par divers travaux de cette même équipe (13) (14).

D'un point de vue de la modélisation, les premiers logiciels de docking développés au début des années 1980, s'appuyaient sur le principe clef-serrure : on considérait les molécules rigides et la tâche consistait à trouver les conformations les plus propices à s'accrocher, en se basant sur des critères géométriques, ainsi que l'orientation qu'elles devaient prendre. Une fois le complexe formé, sa légitimité était testée en se basant sur des critères empiriques et/ou par un calcul énergétique.

Depuis l'élaboration de ces premiers logiciels, des efforts ont été fait pour incorporer la flexibilité au sein des algorithmes de docking. C'est le cas de logiciels tels que AUTODOCK (15), FlexX (16) ou DOCK (17), qui sont aujourd'hui beaucoup utilisés dans le monde académique. Dans le domaine de la recherche pharmaceutique, les programmes privilégiés sont Glide (18) ou encore MOE (<http://www.chemcomp.com/software.htm>). Le protocole opératoire se compose de trois étapes :

- Choix de la représentation des molécules
- Balayage de l'espace conformationnel pour déterminer les structures susceptibles de réagir
- Arrimage des structures retenues et une sélection du composé final, en mesurant la complémentarité par l'intermédiaire d'une fonction de « scoring »

C'est lors des étapes 2. et 3. que nous pourrions introduire la flexibilité du système.

Nous pouvons citer trois catégories de modèles de docking :

- Les modèles rigides, où la compatibilité est testée uniquement par le calcul de l'énergie finale du complexe,
- Les modèles semi-flexibles, où une partie du système, par exemple les sites actifs,

Introduction Générale

« accommodent » leur conformation lors de l'arrimage, en fonction du test des composants,

- Les modèles flexibles, où la flexibilité totale des molécules est prise en compte, ce qui nous concerne plus précisément

Des approches nouvelles font toujours leur apparition sur ce vaste terrain d'investigation, et particulièrement dans le cas du docking protéine-ligand. Nous pouvons citer, par exemple, la technique des structures multiples de protéines, ou MPS (pour Multiple Protein Structure) : plutôt que de considérer la protéine dans une conformation d'équilibre au moment du docking, on considère un ensemble de conformations différentes proches de l'état d'équilibre (19).

En réalité, les molécules mises en jeu subissent des changements conformationnels en permanence au cours du docking. C'est en exploitant l'idée d'intégrer les effets évolutifs au cours du docking, qu'a été élaborée au sein du laboratoire une méthode nouvelle de simulation qui s'appuie sur le calcul des déformations d'une molécule (en termes de déplacement atomique) lorsqu'elle est soumise à des forces extérieures : les Modes Statiques. Ainsi, une nouvelle approche de simulation de docking a été imaginée qui, lorsque les molécules sont mises en contact, permet de calculer les déplacements et les forces en présence, et, par le modèle des « modes statiques », prend en compte les effets des déformations induites. La configuration finale recherchée est alors atteinte, pas à pas, par minimisations successives de l'ensemble.

Ce programme FLEXIBLE développé ces dernières années calcule les « modes statiques » de déformation d'une molécule. Il a déjà permis de créer un certain nombre d'outils utiles pour caractériser la flexibilité des biomolécules (20), qui seront décrits plus tard dans cette thèse. Pour notre part, grâce aux modes statiques, nous tenterons de suivre les changements conformationnels d'un peptide sous l'effet du docking d'un ion métallique. Ces résultats constituent, selon nous, la première étape significative vers le processus complet de docking, tel que nous venons de le décrire. A partir de ces calculs des modes de déformations, nous nous efforçons de montrer qu'il sera possible de développer un logiciel de docking qui répond pleinement au concept « induced fit ».

Dans cette thèse, la méthode des « modes statiques » a d'abord été appliquée à l'étude d'une molécule en particulier : le peptide Amyloïde β , protéine impliquée dans la maladie d'Alzheimer. Ce choix résulte de considérations relatives à l'intérêt que connaît le sujet, mais

il est surtout un choix de première démonstration de l'usage de FLEXIBLE. Ce manuscrit décrivant mon travail se développe selon trois chapitres :

1) Dans une première partie, nous verrons ce que la simulation numérique apporte à la biologie. Afin de bien situer le contexte, les enjeux que cela représente seront présentés, que ce soit dans le domaine de la recherche pharmaceutique ou les biotechnologies émergentes. Puis, afin d'introduire la problématique, les méthodes de simulation et l'état de leur développement au fur et à mesure que les moyens de calcul se sont développés seront présentés. Nous nous pencherons plus particulièrement sur la manière dont est traitée actuellement la flexibilité lors des simulations d'arrimage moléculaire, pour en montrer les limites, avant d'aborder l'approche FLEXIBLE.

2) Le second chapitre sera consacré à présenter la méthodologie originale utilisée pour mes calculs. Après avoir décrit le concept des modes statiques et l'algorithme de calcul, nous verrons comment le programme FLEXIBLE qui permet de les calculer et de les mémoriser a été développé. Nous verrons également dans cette partie les outils développés pour exploiter ces modes de déformations : je pense notamment à la manière de représenter la flexibilité. Finalement, nous décrirons les étapes que nous projetons de franchir pour aller vers un logiciel de docking.

3) Enfin, la troisième partie détaillera les résultats obtenus. Il s'agit tout d'abord d'études menées afin de démontrer la réalité et l'efficacité dans l'approche via FLEXIBLE : nous avons choisi d'effectuer les calculs sur le peptide Amyloïde β , protéine impliquée dans la maladie d'Alzheimer. Cette molécule, présente dans le corps sous une forme non pathogène, peut adopter une conformation la rendant toxique pour les neurones. Les changements de conformation étant dus à l'effet de l'environnement, cela en fait un sujet d'étude intéressant dans notre cas. On peut distinguer deux voies différentes dans les calculs que j'ai effectués : (i) dans un premier temps, je m'intéresserai à la flexibilité du peptide seul. Je présenterai un moyen de cartographier les déformations d'une molécule : ainsi, dans le but de caractériser, en terme de flexibilité, différentes conformations du peptide. (ii) dans un second temps, je m'intéresserai aux changements conformationnels de l'Amyloïde β induits par un ion métallique. Ceci constitue la première étape en vue de l'élaboration d'un logiciel de Docking prenant en compte la flexibilité totale grâce aux modes statiques.

L'approche que nous recommandons pour traiter le docking fera l'objet de réflexions en conclusion de mes travaux.

CHAPITRE I : MODELISATION ET SIMULATION AU SERVICE DE LA BIOLOGIE

SOMMAIRE

I-1. Contexte	- 8 -
I-1.1. Les enjeux	- 9 -
I-1.1.1. Recherche fondamentale	- 9 -
I-1.1.2. Recherche pharmaceutique	- 10 -
I-1.1.3. Nanobiotechnologies	- 11 -
I-1.2. Les biopolymères : une cible privilégiée	- 12 -
I-1.2.1. Les acides nucléiques	- 12 -
I-1.2.2. Les Protéines	- 15 -
I-2. La nature des forces d'interaction	- 18 -
I-2.1. Liaisons fortes	- 18 -
I-2.1.1. Liaison covalente	- 19 -
I-2.1.2. Liaison ionique	- 19 -
I-2.2. Liaisons faibles	- 20 -
I-2.2.1. Interaction coulombienne	- 21 -
I-2.2.2. Forces de dispersion et de répulsion	- 22 -
I-2.2.3. Liaison hydrogène	- 24 -
I-2.2.4. L'effet hydrophobe	- 26 -
I-3. Modélisation des systèmes biologiques	- 27 -
I-3.1. Les modèles quantiques	- 27 -
I-3.2. Le modèle classique	- 29 -
I-3.2.1. Champ de force	- 29 -
I-3.2.2. Optimisation de la structure	- 31 -
I-3.2.3. Le solvant	- 32 -
I-3.3. Modèles dynamiques	- 34 -
I-3.3.1. Modélisation par dynamique moléculaire	- 34 -
I-3.3.2. Simulation de type Monte-Carlo	- 34 -
I-3.4. Le problème de la flexibilité moléculaire	- 36 -
I-4. Comment prendre en compte la flexibilité moléculaire ?	- 38 -
I-4.1. Les méthodes usuelles de simulation	- 39 -
I-4.1.1. Simulations dynamiques	- 39 -
I-4.1.2. Calcul des modes normaux	- 39 -
I-4.1.3. Autres méthodes	- 40 -
I-4.1.4. Limitations	- 41 -
I-4.2. La flexibilité dans les logiciels de Docking	- 42 -
I-4.2.1. Verrous	- 43 -
I-4.2.2. Représentation du système	- 44 -
I-4.2.3. Les algorithmes de recherche	- 45 -
I-4.2.4. Les fonctions de scoring	- 47 -
I-5. Vers un traitement de la flexibilité relative aux interactions moléculaires	- 48 -
I-6. Conclusion	- 52 -

Introduction

L'expérience et la théorie ont permis de comprendre et d'expliquer une grande partie des processus biologiques sur lesquels se fondent les mécanismes du vivant. La combinaison des deux a toujours permis d'apporter les meilleures interprétations. Les fondements de ces mécanismes sont liés à la réactivité intermoléculaire, mais du point de vue de l'analyse prédictive, compte tenu de leur complexité, il reste encore impossible de prédire l'évolution de tels processus.

Or, durant ces dernières décennies, la croissance des ressources informatiques et des capacités de calcul qui en résultent, ont permis d'envisager le développement d'outils de simulation de plus en plus performants. Ils permettent de traiter des macromolécules d'intérêt biologique et sont des moyens prédictifs puissants dans de nombreux domaines : de la conception de médicaments aux biotechnologies, ces outils se rendent très utiles (21) (22). Nous allons nous intéresser ici en particulier à la prédiction de la formation de complexes composés de deux molécules ou processus de « docking ». Après avoir décrit comment sont traitées aujourd'hui les interactions pour de tels systèmes dans les modèles de simulation, nous évoquerons leurs limites et l'importance de prendre en compte la flexibilité des molécules en biologie. Avant de présenter notre problématique et l'approche que nous privilégions, nous détaillerons la manière dont la flexibilité est traitée dans les outils existants.

I-1. Contexte

Le développement de la biologie moléculaire, ainsi que de la biologie structurale, depuis le milieu du siècle dernier a permis de comprendre, par la voie de l'expérience et de l'analyse, les processus biologiques à une échelle de plus en plus petite (23). Les outils de simulation permettent depuis le début des années 1980 d'étudier ces processus sous la forme d'interactions entre biomolécules jusqu'à l'échelle atomique. Grâce à ces avancées scientifiques et techniques, de nombreuses perspectives industrielles ont vu le jour, notamment dans le domaine de la médecine. Dans cette partie, nous allons présenter, du point de vue de la modélisation, l'état de l'art et les chemins d'étude de ces processus biologiques, de plus en plus approfondis.

I-1.1. Les enjeux

La modélisation des fonctions biologiques des molécules nécessite une analyse à des échelles de l'ordre de l'atome, ce qui a donné lieu au développement d'outils de simulation. Leur capacité à expliquer et à prédire leur confère un rôle important dans les domaines de la recherche ou des applications liées à la biologie moléculaire, tels que la conception de médicaments, le développement des biotechnologies et depuis quelques années la création de matériaux artificiels nouveaux, dans le cadre du développement des nanosciences : les matériaux bio-inspirés.

I-1.1.1. Recherche fondamentale

La modélisation des molécules biologiques tient un rôle primordial dans le domaine de la recherche fondamentale. Elle vient compléter les observations issues de l'expérience, qui, historiquement, a toujours été le moyen privilégié d'effectuer des avancées significatives dans le monde de la biologie. Néanmoins, la compréhension de certains phénomènes biologiques passe par l'étude des systèmes à une échelle de plus en plus petite, grâce à la modélisation théorique. Ces outils de simulation apportent alors des réponses complémentaires au travail de l'expérimentateur.

Les molécules biologiques en constante interaction au sein de l'organisme régulent les mécanismes de la vie. L'étude de ces modes d'interaction est donc indispensable dans la compréhension des processus biologiques. Nous sommes alors confrontés à un problème qui se situe au carrefour de la biologie, de la chimie et de la physique, qui ne peut être traité qu'à travers des outils de simulation : les lois de la physique sont appliquées aux biomolécules afin d'en étudier les propriétés.

Depuis les premiers logiciels mis au point aux débuts des années 1980, le développement des moyens de calcul par ordinateur a permis l'analyse et la compréhension de systèmes de plus en plus complexes. Mais le traitement des changements conformationnels subis par les macromolécules en interaction reste une difficulté importante dans ces logiciels de simulation. Prendre en compte ces déformations alourdit énormément le problème, ce qui entraîne une augmentation significative du temps de calcul. Pour améliorer cela, il est nécessaire de faire un certain nombre d'approximations, tout en gardant une précision convenable.

Nous avons au sein du laboratoire une grande expérience de la modélisation à l'échelle atomique, et c'est dans cette idée d' « approximation » évoquée auparavant, que nous avons imaginé cette nouvelle méthode des Modes Statiques (20), à la base de ce travail de thèse.

I-1.1.2. Recherche pharmaceutique

Concevoir de nouveaux médicaments, c'est trouver une molécule qui agit sur une cible biologique, avec des effets secondaires maîtrisés (24). Le médicament est dans la plupart des cas une molécule organique, appelée ligand, qui vient se lier à une macromolécule biologique (cible), impliquée dans la pathologie, et agit sur cette dernière afin d'en moduler les effets.

Les outils de simulation doivent donc permettre de décrire les interactions entre le ligand et la cible biologique, et permettre de tester rapidement plusieurs modes d'opération afin de guider la recherche.

La conception de médicaments conduit à deux axes de travail complémentaires : l'un est basé sur la connaissance préalable du ligand, naturel ou exogène (18), l'autre sur l'identification de la cible (19). La première approche est connue sous la dénomination « Ligand Based Design » (LBD), la seconde étant « Structure Based Design » (SBD) ».

Dans le premier cas, on part du ligand, sans avoir une connaissance précise de la structure de la molécule cible. On cherche alors toutes les molécules cibles susceptibles de réagir avec ce ligand. Le second cas se base sur l'analyse de la structure en trois dimensions de la cible. Celle-ci est déterminée par des techniques expérimentales comme la cristallographie par diffraction des rayons X ou la résonance magnétique nucléaire (RMN). A partir de cette connaissance, il est alors possible d'analyser les sites réactionnels et de déterminer le ligand qui possède les meilleures caractéristiques pour se lier à la cible.

Dans tous les cas, les outils de simulation jouent un rôle important, et plus particulièrement les logiciels de docking (25). En effet, bien que l'intuition des chimistes soit souvent pertinente pour déterminer des ligands potentiels, elle ne permet pas une prédiction fine des ligands les mieux adaptés. Il est donc utile de trouver de nouveaux moyens automatisés et systématiques pour déterminer ou adapter ces ligands potentiels, et comprendre les effets qu'ils induisent, tant d'un point de vue physique que chimique (24).

I-1.1.3. Nanobiotechnologies

Dans les nanobiotechnologies, l'objectif est d'utiliser les connaissances acquises en biologie pour les appliquer sur des procédés et des instruments biochimiques nouveaux (26) ; ou encore, plus récemment, il s'agit de s'inspirer des propriétés des biomolécules pour concevoir de nouveaux matériaux (27) (28).

Il est possible, par exemple, de concevoir des systèmes d'analyses biologiques multiples de la taille d'une puce électronique de quelques millimètres carrés. Ces dispositifs appelés laboratoire sur puce (lab-on-chip) sont l'objet d'un effort de développement important et peuvent devenir des outils de diagnostic très performants (29).

La réalisation de tels dispositifs nécessite souvent le greffage d'une molécule sur une surface. Il faut alors traiter des interactions entre molécules biologiques et surfaces non-biologiques qui posent d'intéressants problèmes scientifiques et technologiques, à la frontière entre le monde inorganique et le monde biologique. L'utilisation d'outils de simulation d'amarrage moléculaire est donc primordiale pour une bonne maîtrise de ces greffages et une analyse des modifications de réactivité de la molécule liées à la proximité de la surface.

Les biotechnologies utilisent les progrès de la connaissance en biologie pour l'appliquer à des technologies nouvelles : les matériaux bio-inspirés sont un exemple d'utilisation des fonctions biologiques comme modèle, afin de développer de nouveaux dispositifs (30). Un exemple est l'exploitation des propriétés structurales de l'ADN pour fabriquer des boîtes nanométriques capables de retenir une protéine et de la relâcher à un endroit stratégique, dans le but d'agir sur un virus (31).

Ce concept est né des deux idées suivantes

- La connaissance de plus en plus précise des structures de biomolécules telles que l'ADN ou les protéines et de leurs propriétés physiques et chimiques permettent de les manipuler et les synthétiser aisément, et donc de les intégrer pour enrichir les technologies que nous connaissons.
- Il existe un fossé, en termes de taille, entre les dispositifs que l'on conçoit actuellement : les technologies ne permettent pas de descendre en dessous de 100-200 nm, et les objets biologiques sont de l'ordre de 2 à 5 nm. Un moyen de faire le lien entre les deux est l'utilisation de macromolécules.

CHAPITRE I : MODELISATION ET SIMULATION AU SERVICE DE LA BIOLOGIE

C'est un domaine de recherche émergeant, où l'imagination et la créativité sont les maîtres mots : on sait qu'il existe énormément d'applications dans ce domaine, mais elles ne sont pas encore toutes connues. Herbert Kroemer, prix Nobel de physique en 2000 et pionnier en la matière, a lui-même déclaré :

« The progress from new technology to new applications is opportunistic rather than deterministic »

Notre laboratoire, qui a développé un savoir-faire dans le domaine des micro- et nano-technologies au cours des quarante dernières années, a décidé d'explorer cette voie pour les années à venir. Dans cette perspective, les modèles de simulation peuvent fournir de très bons résultats dans le cadre d'interactions molécules/molécules, mais il existe peu d'outils prédictifs pour les interactions molécules biologiques/molécules non biologiques ou encore molécules/surface. La méthode des Modes Statiques, que nous allons voir en détail, devra permettre d'effectuer de tels calculs prédictifs.

I-1.2. Les biopolymères : une cible privilégiée

Les biopolymères sont des biomolécules constituées d'une séquence de sous-unités appelées monomères, dont l'ordre joue un rôle dans leur fonction. Ce sont des structures très ordonnées dont la fabrication et la manipulation sont bien maîtrisées, ce qui en fait un objet d'étude privilégié. Il existe quatre types majeurs de biopolymères : les sucres, les lipides, les acides nucléiques et les peptides (ou protéines, suivant la taille de la molécule). Dans cette analyse introductive, nous nous intéresserons uniquement aux deux derniers, qui interfèrent avec nos propres travaux.

I-1.2.1. Les acides nucléiques

Les acides nucléiques sont une catégorie de biopolymères qui jouent un rôle important au sein des cellules. L'ADN (acide désoxyribonucléique) et l'ARN (acide ribonucléique) qui sont responsables du transport et de la transmission des informations génétiques sont des acides nucléiques.

Les monomères constitutants des acides nucléiques sont les nucléotides. Ils sont composés de trois groupes d'atomes liés de manière covalente :

- Une base azotée cyclique : soit une pyrimidine (un cycle) ou une purine (deux cycles)
- Un sucre composé de cinq atomes de carbone – ribose ou désoxyribose
- Un groupement phosphate

Ils se distinguent par leur base azotée : cytosine, guanine, adénine et thymine (remplacée par l'uracile dans l'ARN), ainsi que par le sucre : par exemple, les riboses de l'ADN et de l'ARN se distinguent par l'ajout d'un atome d'oxygène. Un schéma de nucléotide est représenté sur la figure 1.

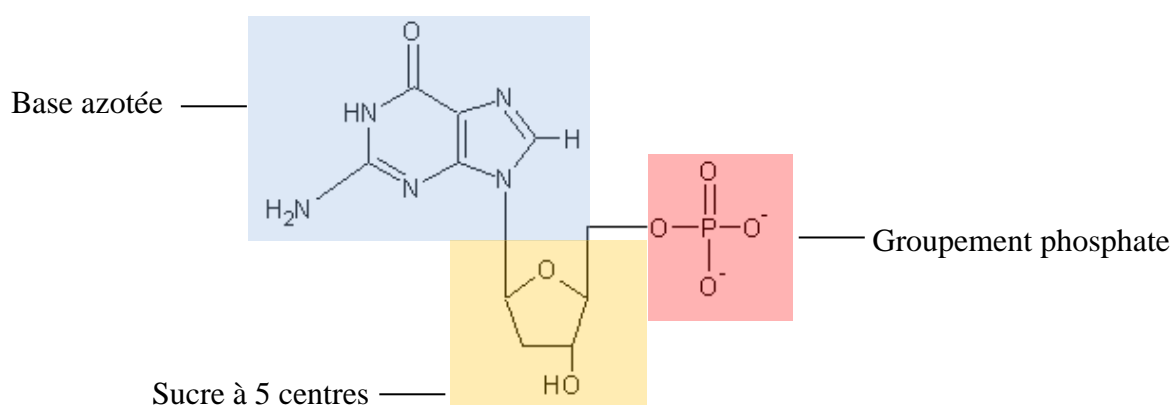


Figure 1 - exemple de nucléotide : désoxy(ribo)guanosine monophosphate

La formation d'un acide nucléique s'effectue par l'intermédiaire de liaisons phosphodiester, qui se forment entre le groupement phosphate et un groupe hydroxyle du sucre. On obtient ainsi une séquence de nucléotides constituée d'un squelette sucre-phosphate et des bases azotées (figure 2).

L'ADN se compose de deux séquences d'acides nucléiques se faisant face et liées par un réseau de liaisons hydrogène. Les deux brins prennent alors une forme hélicoïdale et l'ADN adopte alors sa fameuse forme en double hélice.

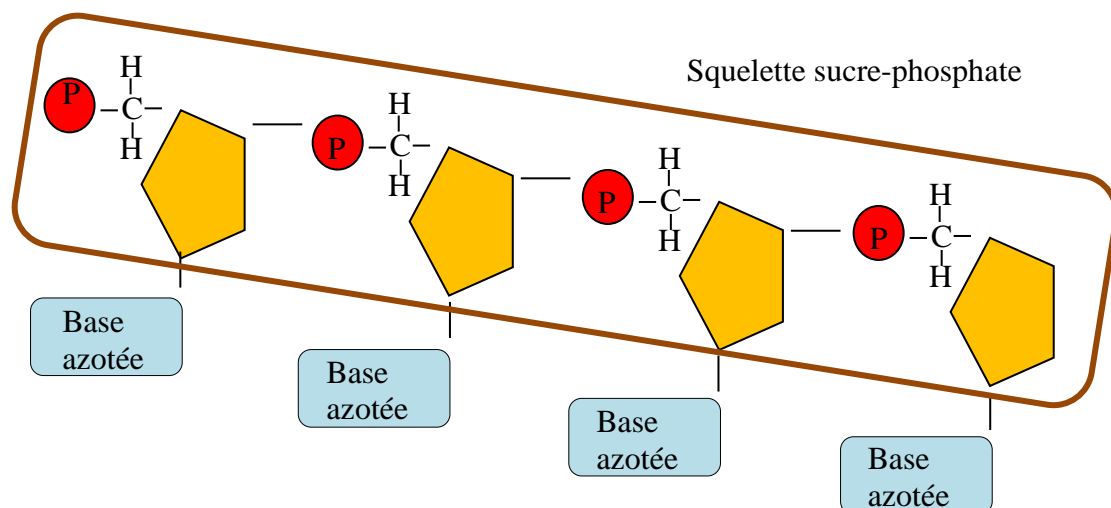


Figure 2 – schéma d'un acide nucléique. On distingue les deux parties de la structure : le squelette sucre-phosphate et les bases azotées.

L'ADN, abréviation d'acide désoxyribonucléique, est la molécule de l'hérédité. Elle contient sous forme codée toutes les informations relatives à la vie d'un organisme vivant, du plus simple au plus complexe, qu'il soit animal, végétal, bactérien ou viral. Sa fonction est de fabriquer les protéines, molécules essentielles à la vie de l'organisme, comme nous le verrons plus tard.

Outre l'intérêt « génétique » que présente l'étude de ces biomolécules à l'échelle atomique, il existe une autre motivation, qui nous intéresse plus particulièrement au sein du laboratoire, qui se situe dans le domaine des biotechnologies. L'ADN possède des caractéristiques qui lui confèrent une place privilégiée dans la conception de matériaux bio-inspirés :

- C'est une molécule simple à synthétiser par des méthodes automatisées, et on peut amplifier une séquence simplement pour passer d'une échelle microscopique à une échelle macroscopique
- Les doubles hélices de petite taille possèdent une grande rigidité mécanique.
- L'ADN possède une stabilité physico-chimique relativement élevée
- Des enzymes telles que les endonucléases ou encore les ligases qui agissent sur l'ADN, permettent un contrôle à l'échelle de l'atome avec une précision de l'ordre de l'Angström (32)
- La maîtrise des nucléotides le désigne comme le matériau « roi » pour la création de

structures 3D à l'échelle nanométrique (27) (33)

I-1.2.2. Les Protéines

Structure

Parmi les quatre types de biopolymères déjà évoqués, ce sont les protéines qui possèdent la structure la mieux comprise (6). C'est un enchaînement d'acides aminés reliés entre eux par des liaisons peptidiques. Le mécanisme de formation d'une liaison peptidique est illustré sur la figure 3.

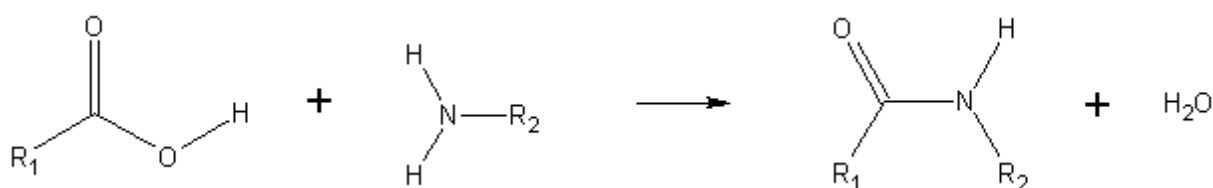


Figure 3 : processus de formation d'une liaison peptidique

D'un point de vue structurel, on peut alors distinguer deux parties de la protéine : le squelette, qui est identique pour toutes les protéines, et les chaînes latérales, qui dépendent de la séquence d'acides aminés. Le terme protéines est employé lorsque la molécule possède une centaine d'acides aminés ou plus. Si elle en comporte moins, elle est qualifiée de peptide, ou de chaîne peptidique.

Il existe une nomenclature pour désigner les atomes constituant les protéines. Sur la figure 6 nous avons explicité les dénominations des atomes du squelette (carbone alpha, carbone prime, oxygène et azote).

On distingue 20 acides aminés, qui possèdent des caractères différents. Ils peuvent être polaires, chargés, ou encore avoir un caractère hydrophobe. L'ensemble de ces caractéristiques déterminent les propriétés de la protéine, la manière dont elle va réagir avec le milieu, etc.... En outre, la séquence d'acides aminés va, en partie, déterminer la structure de la protéine : chaque acide possède un encombrement stérique différent, ce qui restreint l'espace conformationnel autorisé (34).

Pour décrire la structure d'une protéine, il existe trois représentations :

CHAPITRE I : MODELISATION ET SIMULATION AU SERVICE DE LA BIOLOGIE

- Structure primaire : c'est la séquence des acides aminés qui constituent la protéine.
- Structure secondaire : utilisée pour décrire des segments d'une protéine, elle résulte de l'organisation spatiale d'une séquence d'acides aminés qui interagissent entre eux, provoquant des repliements locaux caractéristiques et énergétiquement favorables. Nous verrons par la suite qu'il existe deux formes répandues de structures secondaires : le feuillet beta et l'hélice alpha (figure 10). Une autre forme que l'on retrouve régulièrement est la structure en boucle.
- Structure tertiaire : c'est la représentation tridimensionnelle détaillée de la protéine.
- Structure quaternaire : fait référence à une association de plusieurs chaînes peptidiques liées par des interactions telles que la liaison hydrogène, le pont disulfure ou le pont salé. On recense alors deux types de structure quaternaire : fibreuse ou globulaire.

La structure secondaire est souvent utilisée en biologie structurale : elle représente la forme du squelette. Pour la décrire, les biologistes utilisent les angles de torsion du squelette, aussi appelés angles dièdres. Un angle de torsion est défini par quatre atomes A, B, C et D. C'est l'angle formé par le plan contenant A, B et C et le plan contenant B, C et D (voir figure 4).

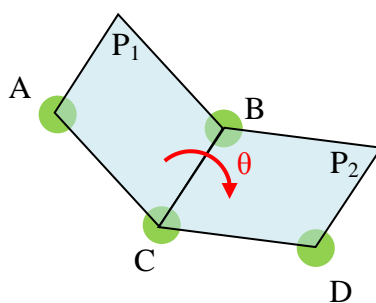


Figure 4 : définition d'un angle dièdre. Le plan P_1 est défini par les atomes A, B, C et le plan P_2 par B, C, D. L'angle dièdre θ est alors l'angle entre P_1 et P_2 .

Ils sont communément désignés par les lettres Φ et Ψ . Grâce à ces valeurs, on peut avoir une idée de l'état de repliement de la molécule. Un troisième angle de torsion est nécessaire pour décrire la totalité du squelette : l'angle ω . Mais le caractère résonant de la liaison peptidique illustré sur la figure 5, fait que cet angle est rigide et plan (35).

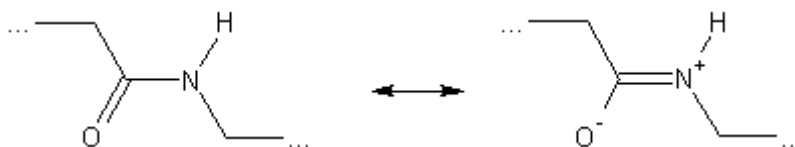


Figure 5 : formes résonantes de la liaison peptidique

La valeur de ces angles dièdres donne la forme du squelette de la protéine, ce qui est une bonne indication de la forme générale de la protéine. Dans la suite du manuscrit, nous utiliserons ces angles dièdres pour décrire la structure des protéines et les variations de ces angles pour les changements de conformation.

Sur la figure 6, on peut voir la structure d'une protéine. Les angles de torsion Φ , Ψ et ω sont représentés, ainsi que le plan qui contient les atomes de la liaison peptidique.

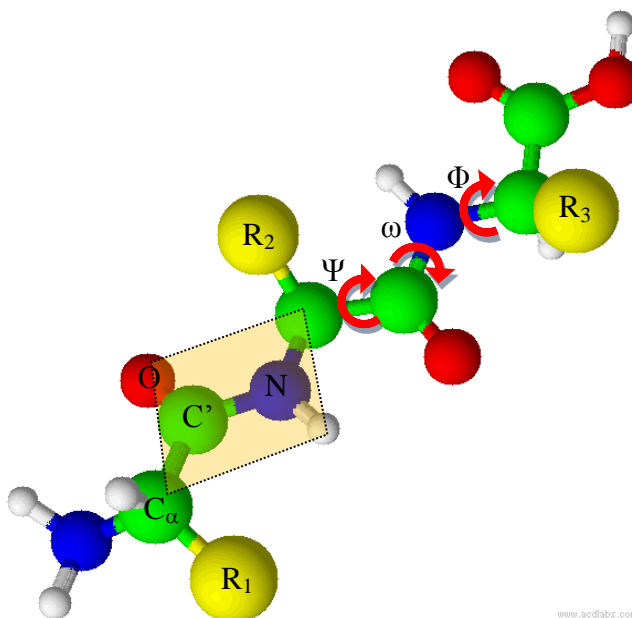


Figure 6 : structure d'une protéine. On peut voir les angles de torsion Φ , Ψ et ω , ainsi que l'on a quatre atomes dans le même plan P , dû à la planarité de la liaison peptidique

Rôle

Les protéines sont un sujet d'étude très important, pour le rôle qu'elles jouent dans les mécanismes du vivant : ces molécules assurent la grande majorité des fonctions cellulaires, ainsi que de nombreuses fonctions au sein de l'organisme (36), parmi lesquelles il faut retenir

leur rôle :

- Structurel : la forme et la résistance des cellules et de certains tissus sont assurés par des protéines comme le collagène ou des protéines dites du cytosquelette
- De transport : parmi les nombreux exemples, on peut citer le rôle de l'hémoglobine qui assure le transport du dioxygène des poumons vers les différents organes de l'organisme
- Immunitaire : les anticorps qui assurent la défense de notre organisme sont des protéines appelées immunoglobulines
- De catalyseur : la plupart des enzymes sont des protéines
- Dans l'expression des gènes : certaines protéines (facteurs de transcription) sont initiatrices de la transcription de l'ADN

Les molécules qui ont été décrites dans ce paragraphe sont indispensables aux processus fondamentaux du vivant. Ces mécanismes sont régulés par des interactions entre les biomolécules, et c'est par l'étude détaillée de ces interactions que passe la compréhension des phénomènes biologiques. Nous allons donc voir dans la suite de ce chapitre quelles sont ces forces d'interaction entre biomolécules.

I-2. La nature des forces d'interaction

Nous avons vu que les interactions entre biomolécules tiennent un rôle essentiel dans la régulation du métabolisme des organismes vivants. Dans cette partie, ces forces intermoléculaires seront décrites, mais également les forces intramoléculaires qui maintiennent la molécule dans un état stable. Parmi ces forces d'interaction, deux classes se distinguent : les liaisons chimiques, qui assurent la stabilité de la molécule, et les liaisons physiques, qui vont régir l'évolution de la molécule dans son milieu (37).

I-2.1. Liaisons fortes

Dans l'organisme, les molécules intervenant dans les processus biologiques peuvent être définies comme des groupements d'atomes ayant une stabilité suffisante pour conserver leur identité lorsqu'elles sont en interaction avec l'environnement (38) (39). Cette stabilité est

assurée par les interactions intramoléculaires qui sont les liaisons chimiques : liaisons covalentes et liaisons ioniques.

I-2.1.1. Liaison covalente

La liaison covalente est le résultat de l'interaction entre deux atomes qui se partagent de manière équiprobable leur paire d'électrons de valence (40). Ce sont des interactions qui ont lieu à courte distance (typiquement entre 0.5 et 2 angströms) et elles possèdent une énergie relativement élevée comparée aux valeurs des autres interactions ou de l'agitation thermique. A titre d'exemple, les valeurs de l'énergie de ces liaisons varient de 24 fois à 350 fois kT (avec $1\text{kJ/mol} \sim 0.4\text{kT}$ par liaison à 300K) (37).

La liaison covalente est le type commun de liaison où la différence d'électronégativité entre les atomes liés est petite ou nulle. Dans ce dernier cas, on parle parfois de liaison purement covalente. Lorsque les atomes se rapprochent, leurs nuages électroniques se recouvrent et les atomes partagent alors leurs électrons de valence. Un équilibre s'établit entre forces attractive et répulsive entre les atomes. Suivant le nombre de paires d'électrons partagé entre les atomes, la liaison sera plus ou moins forte. Les atomes peuvent partager 1, 2, 3 paires d'électrons et peuvent ainsi former des liaisons simple, double ou triple. Par exemple, dans le cas de la liaison O-O, la liaison simple possède une énergie de 145 kJ/mol, alors que celle de la liaison double est de 498 kJ/mol (41). Lors du recouvrement des nuages électroniques, les orbitales atomiques sont totalement perturbées. On ne distingue alors plus les orbitales de chaque atome : une orbitale moléculaire s'est formée. Dans le cadre de la théorie LCAO (Linear Combination of Atomic Orbitals), les orbitales moléculaires sont déterminées comme étant des combinaisons linéaires d'orbitales atomiques.

I-2.1.2 .Liaison ionique

La liaison ionique est une liaison chimique similaire à la liaison covalente, impliquant deux atomes dont la différence d'électronégativité est supérieure à 1,7. L'atome le moins électronégatif cède un ou plusieurs électrons à l'atome le plus électronégatif, formant ainsi des ions positif et négatif respectivement, tous deux possédant alors une configuration électronique stable. Sous l'effet de l'attraction électrostatique, les deux ions se rapprochent

jusqu'à une distance d'équilibre et forment alors une liaison ionique.

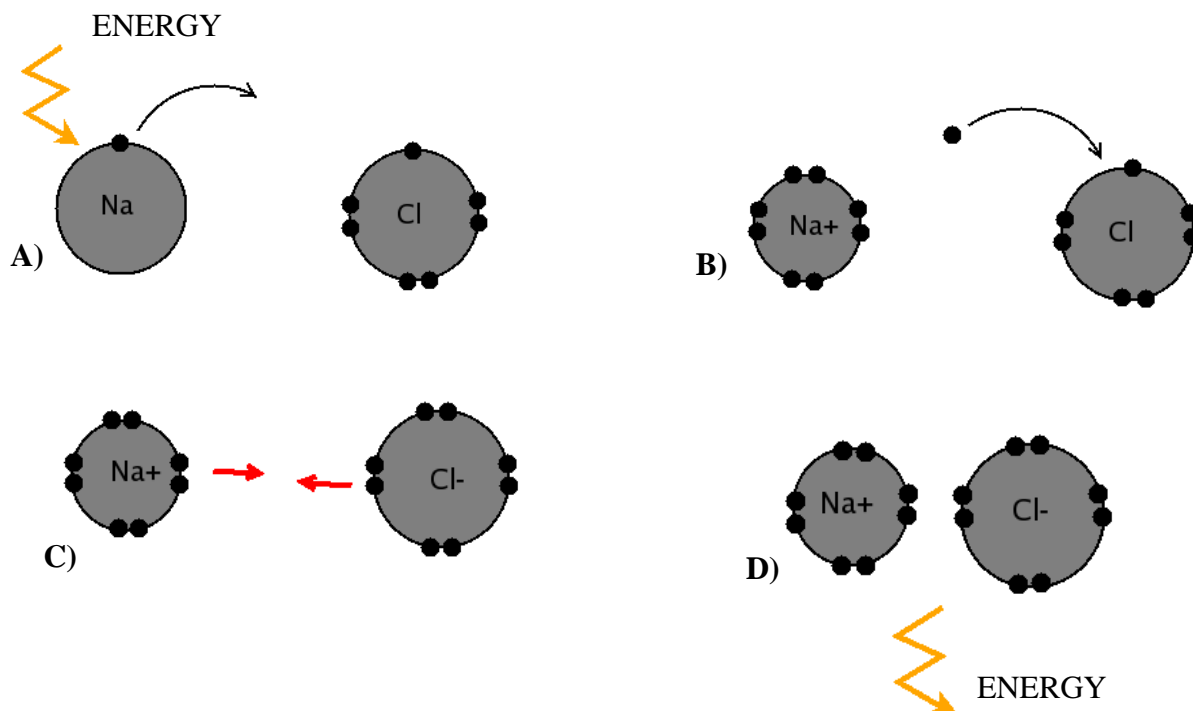


Figure 7 : Processus de formation d'une liaison ionique illustré avec l'exemple de NaCl. **A)** suite à un apport externe d'énergie, l'atome de sodium perd son électron célibataire. **B)** l'électron libre est alors récupéré par le chlore. **C)** les deux ions de signes opposés résultant de ce transfert de charge ont alors leur couche de valence remplie et s'attirent. **D)** on observe alors la formation d'une liaison ionique, exothermique

La liaison ionique résulte de l'attraction entre deux ions de charges opposées, alors que dans le cas de la liaison covalente, les atomes sont liés par le partage d'électrons. Mais en réalité, il n'existe pas de liaison ionique « pure », car théoriquement il faudrait une différence d'électronégativité infinie entre les atomes. Les liaisons ioniques se distinguent par leur pourcentage de covalence.

I-2.2. Liaisons faibles

Les liaisons physiques sont des interactions qui ont lieu pour des distances plus élevées que les liaisons chimiques. Pour la plupart d'entre elles, elles ont une intensité moindre que les liaisons chimiques, mais leur portée est bien plus grande (37). Cela permet aux molécules de réagir avec d'autres éléments éloignés du milieu. **Ce sont ces interactions qui régissent**

l'évolution des biomolécules, et notamment les changements conformationnels. A l'intérieur des cellules par exemple, le repliement des protéines est régi par ces interactions avec le milieu (42) (43). Je vais donc décrire dans ce paragraphe un certain nombre de liaisons physiques qui jouent un rôle important en biologie et dans notre propre démarche.

I-2.2.1. Interaction coulombienne

L'interaction coulombienne est une catégorie de l'ensemble des interactions de nature électrostatique, mettant en jeu deux éléments chargés. Cela peut être l'interaction entre un ion et un autre ion, mais également entre un ion et un dipôle.

Cette interaction est de même nature que celle intervenant dans la liaison ionique, déjà répertoriée parmi les liaisons chimiques. La différence réside dans leurs mécanismes de création qui aboutissent à des distances d'équilibre plus grandes dans le cas des liaisons physiques.

Dans la liaison ionique, les atomes sont neutres au départ. Le transfert d'électron s'opère par effet tunnel lorsque les deux atomes sont proches. Une fois l'attraction amorcée par le transfert de charge, la distance entre atomes continue à diminuer alors que l'effet tunnel est renforcé. Le processus s'auto accélère et n'est arrêté que par la répulsion entre nuages électroniques. La distance d'équilibre est alors très faible et du même ordre de grandeur que celle de la liaison covalente.

Dans la liaison physique, on est en présence d'atomes déjà chargés, à cause de liaisons ioniques intramoléculaires. Pour cette même raison, l'interaction ne s'exerce plus entre un couple unique d'atomes, mais plutôt entre un système d'atomes chargés. De ce fait, les distances d'équilibre sont plus grandes et ne permettent pas de transfert d'électrons supplémentaires.

Dans le vide, l'énergie coulombienne U entre deux particules de charges q_1 et q_2 séparées par une distance r s'écrit :

$$U(r) = \frac{q_1 q_2}{4\pi\epsilon_0 r}$$

Où ϵ_0 représente la permittivité du vide.

Pour un système biologique, l'ensemble des interactions entre les atomes i et j distants de r_{ij} sera donc décrit par la formule :

$$U = \sum_{i < j}^{atomes} \frac{q_i q_j}{4\pi\epsilon r_{ij}}$$

Où ϵ représente la permittivité du milieu.

Dans le cas d'interactions charge/dipôle, l'énergie s'écrit :

$$U(r) = \frac{1}{4\pi\epsilon_0} \cdot q \cdot \frac{\vec{p} \cdot \vec{r}}{r^3}$$

Où q est la charge de l'ion et \vec{p} le dipôle créé par un groupe d'atomes chargés.

Cette interaction est non seulement de plus courte portée que l'interaction entre deux ions, mais elle est aussi directionnelle avec un axe d'anisotropie parallèle au dipôle.

Les interactions coulombiennes sont des forces intenses qui s'exercent sur des grandes distances. Mais dans un milieu physiologique, ces forces sont écrantées par le solvant, ce qui diminue leur champ d'action. Par le biais de ces interactions, la molécule, tout en gardant son identité, va changer de structure, et sa réactivité en sera alors modifiée, comme nous le verrons plus tard.

I-2.2.2. Forces de dispersion et de répulsion

Les forces de dispersion sont des forces attractives qui ont lieu entre tous les atomes, même apolaires, à de très courtes distances. Elles ont une origine quantique, mais elles peuvent tout de même s'expliquer par l'électromagnétisme classique :

L'énergie d'un dipôle \vec{p}_2 dans un champ électrique \vec{E} s'écrit :

$$U = -\vec{E} \cdot \vec{p}_2$$

Si le dipôle est lui-même induit par le champ électrique :

$$\vec{p}_2 = \alpha \cdot \vec{E}$$

où α est la polarisabilité de l'atome.

On en déduit :

$$U = -\alpha E^2$$

Dans la situation qui nous intéresse, le champ électrique est créé par le dipôle \vec{p}_1 d'un

premier atome. Il s'écrit :

$$\vec{E} = -\frac{\vec{p}_1}{r^3} + 3(\vec{p}_1 \cdot \vec{r}) \cdot \frac{\vec{r}}{r^5}$$

Alors que les valeurs moyennes $\langle \vec{p}_1 \rangle$ et $\langle \vec{E} \rangle$ du dipôle et du champ électrique sont nulles, il n'en est plus de même de la valeur moyenne de l'énergie :

$$\langle U \rangle = -\alpha \cdot \langle E^2 \rangle$$

qui est toujours négative. Il s'agit d'un potentiel attractif qui varie comme $1/r^6$ en fonction de la distance interatomique.

Les forces de dispersion attractives entre deux atomes non liés sont contrebalancées par des forces de répulsion. Elles résultent du recouvrement des orbitales électroniques à très courte distance. Selon le principe d'exclusion de Pauli, les électrons ne peuvent pas occuper toute la région de recouvrement. La densité électronique y devient alors plus faible, et les noyaux chargés positivement se retrouvent moins écrantés et se repoussent mutuellement.

Cette interaction a lieu pour de très courtes distances et possède une croissance extrêmement rapide. Cette force de répulsion est souvent représentée par une fonction exponentielle ou une loi de puissance. Lennard-Jones établit une approximation en proposant un terme inversement proportionnelle à la distance interatomique à la puissance 12.

La combinaison de ces forces attractives et répulsives donne lieu à un terme simplifié, mais qui décrit bien la réalité, le terme de Lennard-Jones :

$$U(r) = \frac{A}{r^{12}} - \frac{B}{r^6} = \varepsilon \cdot \left[\left(\frac{\sigma}{r} \right)^{12} - 2 \left(\frac{\sigma}{r} \right)^6 \right]$$

Où ε et σ sont les paramètres du potentiel. ε représente la profondeur du puits de potentiel et σ nous donne la valeur de la distance d'équilibre r_{eq} .

Ce potentiel est représenté sur la figure 8. Des valeurs des paramètres σ et ε pour les atomes C, H, O et N sont également montrés dans le tableau 1.

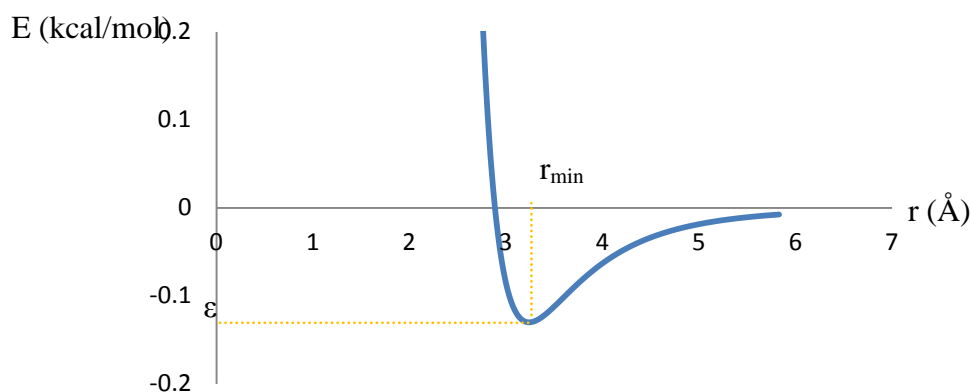


Figure 8 : Potentiel de Lennard-Jones pour des valeurs de $\varepsilon = -0.13$ eV et de $\sigma = 3.24$ Å

Element	ε	σ
C-C	0.06	4.02
H-C	0.03	2.22
O-O	0.19	3.33
N-N	0.13	3.86
H-O	0.03	1.06

Tableau 1 : paramètres de Lennard-Jones pour les atomes qui composent majoritairement les biomolécules. σ est en Å et ε en eV (44)

Lorsque l'on veut modéliser un système biologique, cette forme de potentiel est très utilisée pour décrire les interactions électriques faibles qui ont lieu à courtes distances, aussi appelées forces de Van der Waals.

I-2.2.3. Liaison hydrogène

La liaison hydrogène est une force attractive qui s'opère entre deux groupes d'atomes impliquant un atome d'hydrogène. C'est une liaison physique qui se situe entre la liaison covalente et l'interaction électrostatique. Le caractère covalent de la liaison hydrogène a été démontré à la fin des années 1990, mais la force d'interaction prédominante reste électrostatique (45).

Cette liaison ne s'effectue pas avec n'importe quel atome. Il faut un donneur et un accepteur (sous entendu de proton) :

I-2. La nature des forces d'interaction

- Le donneur est l'atome d'hydrogène lié de manière covalente à un atome électronégatif ; typiquement azote, oxygène, fluor.
- L'accepteur est un autre atome électronégatif ; uniquement azote, oxygène ou fluor.

La liaison hydrogène est donc le résultat d'une interaction dipôle-dipôle, comme le montre la figure 9.

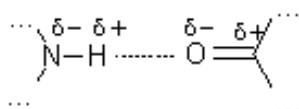


Figure 9 : la liaison hydrogène est une interaction entre deux dipôles, l'un formé par le donneur et l'atome d'hydrogène, l'autre par l'accepteur et le carbone auquel il est lié.

Les liaisons hydrogène sont souvent décrites par deux paramètres : la longueur de la liaison et l'angle formé par les trois atomes. Les valeurs des angles et des distances varient respectivement de 90 à 180 degrés et de 1.2 à 4.0 Å.

Ces paramètres caractéristiques qui dépendent essentiellement du milieu vont déterminer la force des liaisons hydrogènes. On distingue trois catégories répertoriées dans le tableau 2.

Longueur (Å)	Angle (°)	Energie (kcal/mol)	Description
2.2 à 4.0	90 à 150	0.24 à 1	Faible intensité, comparable à Van der Waals. Liaison de type électrostatique
1.5 à 3.2	130 à 180	1 à 4	Liaisons modérées, situées entre les deux extrêmes
1.2 à 2.5	175 à 180	4 à 10	Liaison forte de type covalente

Tableau 2 : les trois catégories de liaison hydrogène (6). La longueur est la distance entre l'atome d'hydrogène et l'accepteur ; l'angle est formé par les trois atomes formant la liaison hydrogène : l'accepteur, l'hydrogène et le donneur.

Ces interactions qui s'établissent entre deux groupes d'atomes, peuvent s'effectuer entre deux molécules différentes (extramoléculaires), mais également au sein d'une même molécule (intramoléculaires). Dans ce dernier cas, elles vont tenir un rôle prépondérant dans la structure de la molécule. Dans le cas des protéines par exemple, les structures secondaires possèdent un réseau bien particulier de liaisons hydrogène qui contribuent à maintenir la

molécule dans cette conformation (v. figure 10).

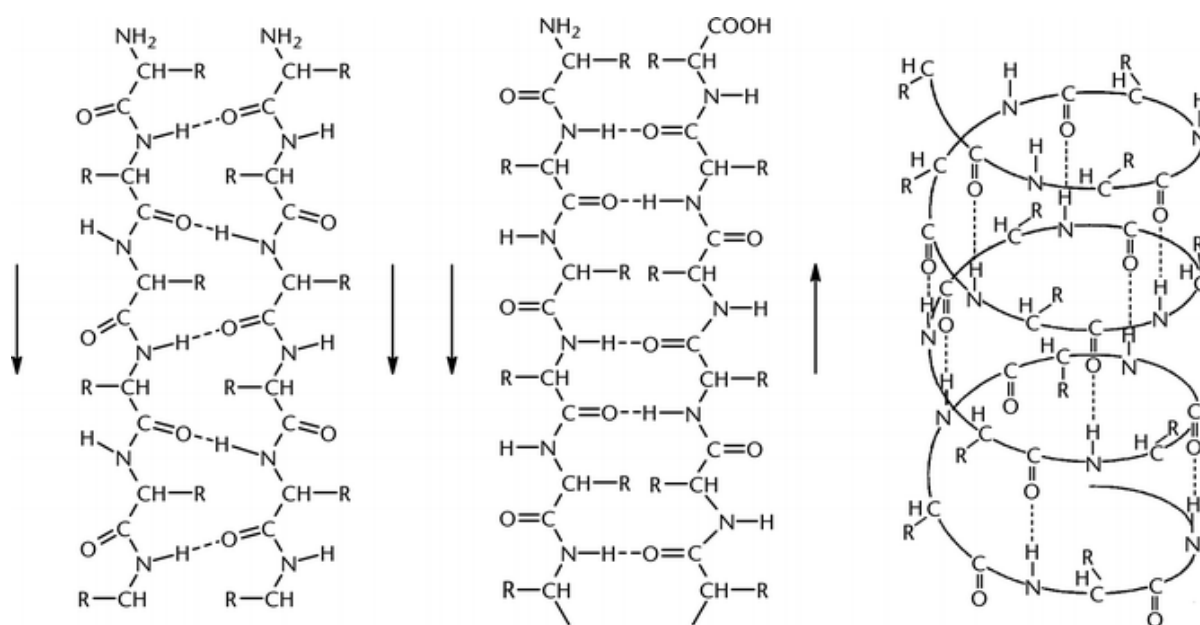


Figure 10 : structures secondaires des protéines. De gauche à droite : feuillet β antiparallèle ; feuillet β parallèle ; hélice α . Les liaisons hydrogènes sont représentées en pointillé. Les flèches indiquent le sens N-terminal vers C-terminal (30).

Ces structures déterminées par Pauling et ses collaborateurs en 1951 (35) (46) se trouvent être des conformations très répandues. En effet, parmi toutes les structures de protéines connues à l'heure actuelle, plus de la moitié des acides aminés sont dans une de ces conformations (47).

L'hélice α possède un nombre plus élevé de liaisons hydrogènes le long du squelette par acide aminé comparé à un brin du feuillet β , ce qui suggère une plus grande stabilité lorsqu'un peptide seul est dans cette conformation. Or, un feuillet β est composé de deux brins ou plus, qui possèdent alors un certain nombre de liaisons hydrogène « inter-brins », qui lui confèrent une plus grande stabilité.

I-2.2.4. L'effet hydrophobe

Les molécules ou groupes d'atomes non polaires ne sont pas capables de former des liaisons hydrogènes et ne peuvent donc pas s'hydrater : pour cette raison, on les nomme substances hydrophobes. L'effet hydrophobe est la tendance qu'ont ces groupes à se

rassembler par coalescence de façon à minimiser les contacts avec l'eau. Les liaisons hydrophobes possèdent une énergie de l'ordre de 7 kcal/mol (6).

Une molécule hydrophobe rompt localement le réseau de liaisons hydrogènes du solvant et crée un état énergétiquement défavorable (perte d'entropie). Les molécules d'eau se réorganisent alors autour des substances hydrophobes de manière à minimiser l'énergie. D'un point de vue thermodynamique, la diminution du nombre de molécules d'eau autour des groupes non-polaires et l'agrégation de ces derniers sous l'effet des forces de dispersion, c'est-à-dire la séparation de deux phases distinctes, organique et aqueuse, s'accompagne d'un gain d'entropie de l'ensemble du solvant et de la molécule non polaires.

Pour cette raison, les régions non polaires sont souvent enfouies à l'intérieur des molécules. Cet effet a d'importantes répercussions sur la conformation et la stabilité des molécules. Elle est à l'origine du processus de repliement des protéines par exemple, ou encore de l'auto-assemblage de biomolécules (48) (49).

I-3. Modélisation des systèmes biologiques

Le point de départ du calcul en modélisation est la constitution d'un modèle statique : on détermine la structure la plus stable de la molécule. On utilise pour cela un champ de force capable de décrire les interactions intra- et inter-moléculaires. Enfin, pour suivre son comportement dans l'espace des phases, on se donne une loi d'évolution appliquée au système de départ et on échantillonne un grand nombre de configurations. A chaque pas on peut ainsi calculer les interactions exactes du système donnant la valeur de l'énergie potentielle du système, et les grandeurs caractéristiques du système sont calculées comme des moyennes sur l'ensemble des configurations.

I-3.1. Les modèles quantiques

La théorie quantique décrit la matière comme un ensemble de noyaux autour desquels gravitent les électrons, sous forme de nuages électroniques, décrits par leur probabilité de présence et représentés par une fonction d'onde.

En appliquant les lois de la mécanique ondulatoire aux électrons, la mécanique quantique permet alors de déterminer l'état électronique d'un ensemble d'atomes, mais aussi l'ensemble

de ses propriétés observables, par le biais d'un opérateur mathématique.

En 1926, Schrödinger établit une équation qui permet de décrire rigoureusement la nature microscopique de la matière (50) :

$$\hat{H}\Psi(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N) = E\Psi(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N)$$

Où $\Psi(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N)$ est la fonction d'onde, les \vec{r}_i représentent les positions électroniques, N est le nombre d'électrons, E est l'énergie du système associée à la fonction d'onde Ψ et \hat{H} est l'opérateur hamiltonien qui contient toutes les contributions énergétiques. Pour un système de M noyaux de masse m_A , de charge $+Z_A e$, de position R_A et de N électrons de masse m_e , de charge $-e$, de position r_i , l'hamiltonien s'écrit :

$$\hat{H} = \hat{T}_N + \hat{T}_e + \hat{V}_{e,e} + \hat{V}_{e,N} + \hat{V}_{N,N}$$

Avec \hat{T}_N et \hat{T}_e les opérateurs énergie cinétique des noyaux et des électrons respectivement, et $\hat{V}_{e,e}$, $\hat{V}_{e,N}$ et $\hat{V}_{N,N}$ les opérateurs énergie potentielle (termes d'interaction coulombienne) respectivement électron-électron, électron-noyau et noyau-noyau.

Cette équation extrêmement complexe est impossible à résoudre de manière analytique pour des systèmes de quelques atomes, comme le suggèrent les travaux de Heitler et London d'une part (51), et de Condon d'autre part (52). Il est donc nécessaire d'effectuer un certain nombre d'approximations, qui donnent tout de même une description précise de la matière :

- Les méthodes ab-initio : ces méthodes s'efforcent à déterminer la fonction d'onde en émettant l'hypothèse que chaque électron se déplace dans un champ moyen dû aux noyaux et aux autres électrons (53). Par une méthode variationnelle, il est alors possible de déterminer la fonction d'onde ψ_0 correspondant à l'état de plus basse énergie E_0 , aussi appelé état fondamental.
- Les méthodes DFT (pour Density Functionnal Theory): dans ce cas, on considère que l'énergie du système est une fonctionnelle de la densité électronique, elle-même fonction de la fonction d'onde (54) (55). Kohn et Sham ont démontré en 1965 que toutes les propriétés d'un système d'électrons peuvent être déterminées par sa densité électronique (56). L'avantage de cette méthode est que la densité électronique ne nécessite que trois variables, contre $3N$ pour la fonction d'onde (pour un système de N électrons).
- Les méthodes semi-empiriques : elles reprennent les principes des méthodes ab-initio,

mais introduisent des paramètres dérivés de l'expérience dans les équations. Cela allège les équations et accélère le temps de calcul.

Les méthodes quantiques permettent donc de sonder la matière de manière très précise, mais on ne peut appliquer ces méthodes à des systèmes trop gros. Ces méthodes ne sont pas très adaptées pour étudier des systèmes biologiques complexes. En revanche, elles seront utiles pour l'étude précise d'un élément de taille raisonnable du système (de l'ordre de la centaine d'atomes).

I-3.2. Le modèle classique

La mécanique classique est utilisée dans le but de déterminer le mouvement d'objets macroscopiques et son application à l'étude des molécules est appelée mécanique moléculaire. En mécanique moléculaire, les atomes sont traités comme des sphères dont la masse dépend de l'élément. Les liaisons chimiques sont assimilées à des ressorts dont la raideur dépend des éléments liés entre eux, et s'il s'agit d'une liaison simple, double ou triple. D'autres types de ressorts sont utilisés pour modéliser les mouvements des angles de liaison, des angles de torsion, etc... Chacun de ces types de ressorts ont leurs propres paramètres, déterminés par l'expérience ou la théorie. D'autres équations issues de la physique classique s'ajoutent au modèle, telle que la loi de Coulomb par exemple, dans le but de décrire les interactions « non-liés » de type électrostatique. Ces termes mis ensembles définissent l'énergie potentielle totale de la molécule et forment ce que l'on appelle un champ de force. Différents champs de forces ont été développés pour décrire différents types de molécules (petites molécules organiques contre macromolécules biologiques). Dans cette partie, nous allons voir plus en détail ces champs de force, ainsi que les méthodes utilisées pour obtenir une description détaillée des molécules.

I-3.2.1. Champ de force

Le champ de force doit décrire au mieux les interactions entre les molécules du système, tout en préservant la nature des molécules mises en jeu. C'est de ce modèle que va dépendre la précision du calcul. Suivant le degré de précision désiré, une fonction plus ou moins sophistiquée est utilisée. La forme la plus simple est le champ de force de classe I :

$$U(R) = \sum_{liaisons} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{dièdres} K_\varphi [1 + \cos(n\varphi - \gamma)] \\ + \sum_{i < j}^{atomes} \varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_{i < j}^{atomes} \frac{q_i q_j}{\varepsilon R_{ij}}$$

Les trois premiers termes de cette expression concernent les atomes liés de manière covalente et correspondent respectivement aux élongations des liaisons, aux angles de liaison et aux angles de torsion. Les deux derniers termes concernent les atomes « non-liés » (sous-entendu de manière covalente) et l'on reconnaît le terme de Lennard-Jones ainsi que le potentiel de Coulomb.

Pour les deux premiers termes, les longueurs et les angles de liaison sont maintenues proches de leur valeur d'équilibre. Ces derniers sont alors décrits par de simples potentiels harmoniques. Dans le cas des longueurs de liaison, K_r représente la constante de force, r la longueur de la liaison et r_{eq} cette longueur à l'équilibre. Pour les angles de liaison, K_θ représente la constante de force, θ l'angle de liaison et θ_{eq} sa valeur à l'équilibre.

La rotation restreinte autour des liaisons covalentes est représentée dans le troisième terme à l'aide d'un potentiel de torsion comprenant un terme en cosinus, où K_φ est la constante de force, φ l'angle de torsion, n sa multiplicité (le nombre des positions angulaires d'équilibre) et γ un terme de phase.

Les deux derniers termes concernent les atomes « non-liés ». On reconnaît les expressions de Lennard-Jones et de Coulomb, où σ_{ij} et ε_{ij} sont les paramètres de Lennard-Jones entre les atomes i et j , R_{ij} la distance entre les atomes i et j , q_i la charge de l'atome i et ε la permittivité du milieu.

Dans certains cas, des champs de force basés sur les polarisabilités des dipôles centrés sur les atomes sont utilisés, en ajoutant alors le terme :

$$E_{pol} = - \sum_i^{atomes} \mu_i \cdot E_i^{(0)}$$

Où μ_i représente un dipôle atomique induit.

Dans le cas des champs de force de classe I, les termes décrivant les mouvements des liaisons, des angles ou des torsions sont des fonctions harmoniques. Des champs de force

plus sophistiqués de classe supérieure (II voire III) sont également disponibles. Ces mouvements sont alors décrits par des fonctions d'ordre supérieure, traduisant les interactions liaison-liaison, liaison-angle, angle-angle, etc...

Ce type de potentiel de mécanique moléculaire dépend fortement de paramètres qui caractérisent les atomes, tels que la charge, les paramètres de Lennard-Jones, ou encore les longueurs de liaisons. Chaque type d'atome possède donc son propre jeu de paramètres, déterminés par l'expérience pour certains d'entre eux et par le calcul *ab initio* pour d'autres¹.

La précision du potentiel dépendra donc à la fois des termes qui définissent le potentiel lui-même, mais également du degré de précision des paramètres caractérisant chaque type d'atomes.

I-3.2.2. Optimisation de la structure

Le potentiel $U(R)$ décrit donc une hypersurface représentative de l'énergie totale de la molécule. Cette énergie, qui décrit les interactions entre tous les atomes de la molécule, est aussi appelée énergie conformationnelle : c'est elle qui décrit la stabilité structurale de la molécule.

Il faut ensuite optimiser la structure, c'est-à-dire trouver la conformation d'équilibre dans laquelle elle se trouve dans un état stable. Cette conformation est alors celle pour laquelle l'énergie est minimale. On part d'une structure initiale (déterminée expérimentalement par exemple) que l'on suppose proche de son état stable et on minimise son énergie pour atteindre un minimum local. Cette région autour de ce minimum est un bassin de potentiel et définit alors un sous-état de la molécule.

Il existe plusieurs techniques pour minimiser l'énergie. Les deux méthodes les plus utilisées dans les logiciels de simulation sont :

- La descente de gradient : cette méthode utilise le gradient de l'énergie potentielle pour déterminer la direction vers le minimum. A chaque étape de la minimisation, l'énergie est calculée et la direction de l'étape suivante est alors donnée par la nouvelle valeur du gradient. Le minimum est atteint lorsque la différence d'énergie de deux étapes

¹ Par exemple, un atome d'oxygène d'une terminaison COO⁻ n'aura pas les mêmes paramètres que l'oxygène d'une molécule d'eau

successives est inférieure à un seuil fixé. Cette méthode n'est pas très efficace et est souvent utilisée en début de minimisation, lorsque la structure est loin du minimum local

- Les gradients conjugués : cette méthode, comme celle la précédente, utilise le gradient de l'énergie comme direction vers le minimum. Mais ici, pour obtenir l'étape suivante, on ne prend pas la direction donnée par le gradient. Il est nécessaire d'effectuer deux calculs intermédiaires : On calcule un premier vecteur perpendiculaire à la direction du gradient puis un deuxième orthogonal à ce dernier, qui sera alors la nouvelle direction à suivre. Ainsi, on évite des effets d'instabilité à proximité d'un minimum et cela permet d'effectuer une optimisation fine de la structure. Le minimum est atteint de la même manière que précédemment.

Les structures minimisées énergétiquement sont la base de tout calcul de modélisation. La méthode des Modes Statiques que nous verrons ultérieurement utilise également des molécules dont la structure a été optimisée par minimisation de l'énergie.

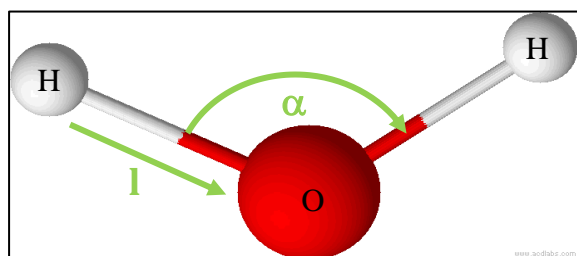
I-3.2.3. Le solvant

Dans leur environnement, les molécules biologiques interagissent en permanence avec les molécules du solvant. Il est donc indispensable de prendre en compte ces interactions dans les modèles de simulation. Pour cela, il y a deux façons de procéder : de manière explicite ou de manière implicite.

La méthode explicite consiste à définir une boîte virtuelle autour de la molécule dans laquelle on dispose des molécules d'eau. Il est également possible d'ajouter des ions positifs ou négatifs afin d'assurer la neutralité du système. Plusieurs modèles différents ont été proposés pour décrire les molécules d'eau, qui se différencient par le nombre de points utilisés pour définir la molécule (atomes plus sites virtuels), suivant la manière dont on veut traiter le solvant (si on considère la structure flexible ou rigide, ou encore si on veut inclure des effets de polarisation).

Le modèle TIP3P (57) modélise simplement et de manière efficace les molécules d'eau, et est de ce fait le plus répandu dans les outils de simulation. C'est un modèle rigide à trois sites correspondant aux atomes et est décrit par un potentiel d'interaction composé d'un terme de Lennard-Jones et d'un terme de Coulomb. Le modèle TIP3P est donc défini par un jeu de

paramètres détaillés sur la figure 11.



Paramètres du modèle TIP3P	
Longueur de liaison l	0.96 Å
Angle de liaison α	104.5 °
Charge de l'oxygène	-0.834 e
Charge de l'hydrogène	0.417 e
Paramètre ϵ de Lennard-Jones	0.1521 kcal/mol
Paramètre σ de Lennard-Jones	3.1506 Å

Figure 11 : modèle TIP3P de la molécule d'eau

Pour traiter le solvant de manière implicite, on utilise une approximation de l'équation de Poisson-Boltzmann : le modèle de Born. L'équation de Poisson-Boltzmann décrit l'environnement électrostatique d'un soluté dans un solvant, mais la résoudre numériquement prend énormément de temps de calcul. L'approximation de Born permet, en rajoutant le terme suivant à la fonction d'énergie potentielle, de traiter les effets du solvant de manière implicite :

$$G_s = \frac{1}{8\pi} \left(\frac{1}{\epsilon_0} - \frac{1}{\epsilon} \right) \sum_{i,j} \frac{q_i q_j}{f_{GB}}$$

avec :

$$f_{GB} = \sqrt{r_{ij}^2 + a_{ij}^2} e^{-D}$$

et

$$D = \left(\frac{r_{ij}}{2a_{ij}} \right), \quad a_{ij} = \sqrt{a_i a_j}$$

Ce terme est la partie « polaire » de l'énergie de solvation via la fonction f_{GB} (58). On peut également rajouter un terme pour la partie apolaire proportionnelle à la surface accessible de la molécule, qui définit la surface de la molécule susceptible d'interagir avec le solvant.

I-3.3. Modèles dynamiques

I-3.3.1. Modélisation par dynamique moléculaire

A partir d'une structure moléculaire de base, la dynamique moléculaire permet de calculer sa trajectoire dans l'espace des phases, c'est-à-dire une suite de conformations, en fonction du temps. Pour cela, il suffit de passer par l'intégration des équations du mouvement de Newton, ou dans certains cas les équations de Langevin dans le cas de simulations dites « stochastiques ».

A partir d'un potentiel d'interaction choisi $u(r_{ij})$ entre les atomes i et j , les forces F_i sont calculées sur l'atome i :

$$F_i = \sum_{i \neq j} \nabla u(r_{ij})$$

Puis, l'équation du mouvement de Newton est utilisée :

$$\gamma_i = \frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i}$$

Ou encore, l'équation du mouvement de Langevin :

$$m_i \gamma_i = m_i \frac{d^2 r_i}{dt^2} = F_i - \xi_i + f_i$$

Où F_i représente la force exercée sur l'atome de masse m_i , γ_i l'accélération, ξ_i une force de frottement fluide et f_i une force supplémentaire aléatoire décrivant les collisions des atomes avec des particules présentes dans son environnement.

Un intervalle de temps Δt est alors défini, sur lequel seront intégrées ces équations pour obtenir une seconde structure. En répétant cette opération N fois, la trajectoire de la molécule sera obtenue dans l'espace des phases pendant un temps $N.\Delta t$. A chaque pas, des valeurs caractérisant l'état du système peuvent être relevées, telle que l'énergie par exemple.

I-3.3.2. Simulation de type Monte-Carlo

Le nom simulation de type Monte Carlo fait référence aux jeux de hasard et est utilisé pour décrire tout algorithme faisant intervenir un processus aléatoire. Développées au milieu du siècle dernier, ces méthodes connaissent un vaste champ d'application. En biologie, elles sont utilisées aussi bien pour décrire des repliements de protéines, effectuer une exploration

stochastique de l'espace des configurations d'une molécule, etc... Dans ce paragraphe, nous allons voir la méthode de Metropolis (59), pionnier de la simulation informatique, qui est la plus répandue pour décrire des systèmes biologiques.

Au lieu de faire évoluer exhaustivement une conformation dans l'espace des phases, la méthode de Monte Carlo Metropolis biaise l'exploration vers les états les plus probables statistiquement, c'est-à-dire ceux qui présentent l'énergie la plus basse. Cette technique se base sur le principe que ces états sont statistiquement les plus probables, donc les plus peuplés, et de ce fait les plus représentatifs du système.

Pour cela, est définie une loi de probabilité basée sur une fonction d'énergie potentielle, équivalente à celle décrite au paragraphe précédent. La différence majeure avec la dynamique moléculaire, est l'absence du temps dans les simulations de type Monte Carlo. De plus, dans le cas d'une simulation Monte Carlo, le système est considéré dans le cadre d'un ensemble canonique, où N (le nombre de particules), V (le volume) et T (la température) restent constants². Dans un tel ensemble, la probabilité p_n d'avoir un échantillon n d'énergie U_N^n est :

$$p_n = \exp(-\beta U_N^n)$$

Avec $\beta = \frac{1}{k_B T}$ où k_B est la constante de Boltzmann.

Les positions atomiques subissent un déplacement Δr aléatoire afin d'obtenir une nouvelle configuration $n+1$. On calcule alors l'énergie U_N^{n+1} de la nouvelle configuration, et on effectue un test d'acceptation basée sur le rapport des probabilités des deux états successifs :

$$\frac{p_{n+1}}{p_n} = \frac{\exp(-\beta U_N^{n+1})}{\exp(-\beta U_N^n)} = \exp(-\beta \Delta U_N)$$

Se présentent alors deux cas de figure :

- $\frac{p_{n+1}}{p_n} > 1$, soit $\Delta U_N < 0$: on conserve la nouvelle configuration et on effectue un nouveau tir aléatoire à partir de cette dernière.
- $\frac{p_{n+1}}{p_n} < 1$, soit $\Delta U_N > 0$: on ne conserve pas la nouvelle configuration et on effectue un nouveau tir aléatoire à partir de l'ancienne configuration.

Nous avons résumé ces deux techniques de simulation (dynamique moléculaire et Monte

² En dynamique moléculaire, on peut faire varier chacun de ces paramètres.

Carlo Metropolis) sur le schéma de la figure 12.

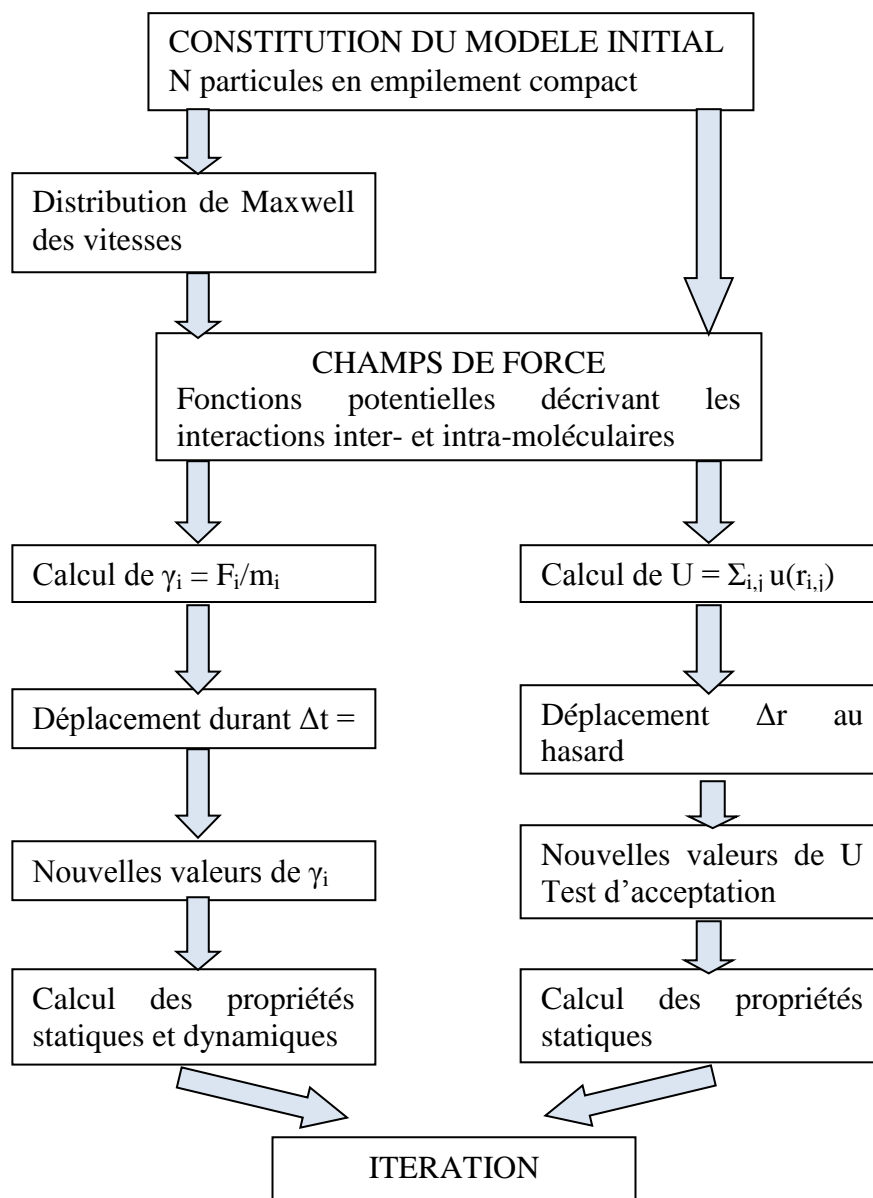


Figure 12 - Schéma des procédures de simulation

I-3.4. Le problème de la flexibilité moléculaire

En biologie, la flexibilité est une propriété cruciale dans l'activité des macromolécules : leur réactivité étant liée à leur conformation, des changements conformationnels peuvent entraîner des changements fonctionnels. Ceci a été mis en évidence expérimentalement au milieu des années 1970 par l'équipe de K. Wutrich (60). Depuis, il a été montré que la flexibilité moléculaire conditionne de nombreuses fonctions telles que la catalyse

enzymatique (61), la régulation de l'activité des protéines (62), ou encore le transport des métabolites (63). Mais également, certaines protéines très flexibles sont impliquées dans le développement de maladies (64).

Cette flexibilité des macromolécules biologiques peut être définie par différents types de mouvement :

- Les fluctuations atomiques se limitent à des vibrations des atomes autour de leur position d'équilibre (sous effet thermique) et n'entraînent pas de changement conformationnel. Ces mouvements ont lieu pour des temps de l'ordre de la picoseconde.
- Dans le cas des protéines, les mouvements des chaînes latérales de surface s'accompagnent en général de changements conformationnels: les chaînes latérales visitent plusieurs sous états du fait de l'agitation thermique et de l'influence d'autres molécules. Ces changements sont de faible amplitude vis-à-vis de l'ensemble de la protéine (0.1 à 1 Angström). On les observe pour des échelles de temps de l'ordre de la nanoseconde ils peuvent être importants pour l'activité ou non
- Toujours dans le cas de protéines, les fluctuations peuvent s'accompagner de modifications structurales de la chaîne principale. Ces modifications peuvent être limitées à des boucles de surface, concerner des mouvements relatifs de domaines « quasi-rigides », ou impliquer la structure protéique entière. Il peut s'agir de déformations de faible amplitude dans un bassin énergétique, ou bien de changements de sous-état conformationnel (passage d'un bassin de potentiel à un autre distinct). Les temps correspondants sont de l'ordre de la microseconde pour des repliements de peptides. Pour des plus grosses protéines, cela peut aller jusqu'à la seconde.
- Dans le cas de l'ADN, on peut également observer des mouvements locaux (entre 2 paires de bases, par exemple), des mouvements globaux de vibration autour d'un sous-état conformationnel (ex, forme B, mouvements d'étirement, d'enroulement) mais également des changements d'état conformationnel impliquant le passage de barrières énergétiques (B-DNA -> A-DNA). Dans tous les cas, l'amplitude du changement global peut être importante du fait de la linéarité de la molécule.

La flexibilité moléculaire peut être mise en évidence par diverses techniques expérimentales :

- La diffraction des rayons X permet de déterminer des déplacements atomiques

moyens (AMSD pour Atomic Mean Square Displacement) ou encore de calculer les facteurs de température. Ces indices rendent compte des fluctuations des positions atomiques, ce qui donne une idée de la flexibilité

- La résonance magnétique nucléaire permet de déterminer un ensemble de conformations proches, à partir desquelles on peut déduire les fluctuations des positions atomiques

Les informations structurales fournies par l'expérience sont une base essentielle pour caractériser les mécanismes, analyser l'évolution et comprendre les fonctions des biomolécules. Mais afin de décrire la flexibilité de ces biomolécules dans leur environnement, il est nécessaire de faire appel à la modélisation.

Les mouvements harmoniques se limitant à un seul bassin de potentiel sont bien appréhendés par les techniques de simulation tel que le calcul des modes normaux de vibration. Le point noir des outils de modélisation vient essentiellement des mouvements anharmoniques de grande amplitude, notamment lorsqu'ils entraînent un changement de sous-état conformationnel. L'interaction entre deux macromolécules peut engendrer de tels mouvements, ce qui peut avoir un impact sur leur réactivité : il est donc indispensable de prendre la flexibilité en compte, et particulièrement lorsque l'on désire simuler des mécanismes de docking, comme le suggère le modèle « induced-fit » évoqué précédemment.

1-4. Comment prendre en compte la flexibilité moléculaire ?

Les mouvements atomiques, engendrant parfois des changements conformationnels, induits par les interactions avec le milieu sont une caractéristique primordiale des biomolécules. Cette flexibilité qui leur est propre leur permet ainsi de remplir leur(s) fonction(s) au sein de l'organisme (65). Une modélisation précise des processus biologiques passe alors par la prise en compte de ces mouvements atomiques et des modifications structurales engendrées.

Nous allons voir dans ce paragraphe quelles sont les techniques employées pour traiter la flexibilité moléculaire. Nous verrons plus particulièrement comment elle est prise en compte lors de mécanismes de « docking ».

I-4.1. Les méthodes usuelles de simulation

Il existe plusieurs modèles pour décrire la flexibilité d'une molécule dans son environnement. Dans ce paragraphe, seront traitées les méthodes les plus utilisées, qui sont les simulations dynamiques (dynamique moléculaire et Monte Carlo) et l'analyse des modes normaux de vibration. Il existe d'autres méthodes récentes plus isolées, également présentées dans cette partie.

I-4.1.1. Simulations dynamiques

Les techniques de simulation décrites dans le paragraphe **I-3.** donnent une représentation très détaillée de la flexibilité des macromolécules. La dynamique moléculaire permet actuellement de simuler l'évolution d'une molécule sur plusieurs microsecondes (66). L'analyse de la trajectoire permet alors de déterminer les régions flexibles.

Il existe plusieurs variantes de la dynamique moléculaire, chacune essayant de faire reculer ses limites :

- Les méthodes qui prennent en compte la mobilité de tous les atomes permettent d'obtenir une description détaillée de la flexibilité, notamment de petits groupes d'atomes, telles que les chaînes latérales des protéines
- Les méthodes qui simulent les mouvements de domaines rigides d'une molécule donnent une idée des mouvements globaux de la molécule.
- Les méthodes dites « gros grains » dans lesquelles on considère des groupements d'atomes (les chaînes latérales par exemple) comme un seul site (67) (68).

I-4.1.2. Calcul des modes normaux

Les vibrations d'une molécule sont caractérisées par l'oscillation des atomes autour de leur position d'équilibre. Ces mouvements atomiques sont les modes de vibration de la molécule et, dans le cadre d'une approximation harmonique de l'énergie potentielle, ils sont donnés par les modes normaux.

Pour une molécule non-linéaire composée de N atomes, chaque atome est libre de se déplacer dans les trois directions de l'espace. Ce qui signifie que la molécule possède $3N$ degrés de liberté, parmi lesquels on a 3 degrés correspondants à la translation de la molécule entière, ainsi que 3 degrés pour sa rotation. Il y a donc $(3N-6)$ modes de vibration, soit $(3N-6)$ modes

normaux³.

Les modes normaux sont obtenus par la diagonalisation de la matrice dynamique (matrice des dérivées secondes de l'énergie potentielle pondérée par les masses) à un minimum local. Cela signifie donc que l'utilisation des modes normaux est limitée à un bassin d'énergie. Chaque vecteur propre représente un mode normal, la valeur propre donnant le carré de la fréquence de vibration. Dans un mode normal, les atomes de la molécule vibrent en phase à la même fréquence. Cette technique permet donc d'obtenir les oscillations autour d'une position d'équilibre, mais pas de trouver un autre sous-état conformationnel.

Toutefois, il a été observé expérimentalement que lorsque deux conformations d'une même molécule sont connues, les modes de basse fréquence, typiquement de l'ordre de 30 cm^{-1} , donnent la direction à suivre pour passer d'une conformation à l'autre. C'est le cas des travaux effectués par Brooks et Karplus (69), Hayward *et al* (70) ou plus récemment Tama *et al* (71).

I-4.1.3. Autres méthodes

En marge des deux approches les plus répandues décrites ci-dessus, des méthodes développées plus récemment apportent également des résultats satisfaisants.

Parmi elles, on trouve des méthodes basées sur le découpage d'une macromolécule en plusieurs domaines, considérés comme des corps rigides, qui bougent les uns par rapport aux autres. Des logiciels ont été développés afin d'effectuer un tel découpage et d'extraire les informations relatives au mouvement de ces groupes d'atomes représenté par les domaines rigides automatiquement. C'est le cas de Domain Parser (72), HingeFinder (73) ou plus récemment Protein Domain Parser (74).

Des données sur le mouvement des atomes peuvent aussi être obtenues à partir de bases de données. La base de données des mouvements macromoléculaires développée par Gerstein et Krebs (75) classe les protéines en fonction de mouvements caractéristiques observés sur des structures cristallographiques. Les informations recueillies traitent uniquement les mouvements dominants (impliquant la chaîne principale). Cette base de données ne fournit

³ Une molécule linéaire ne possède que $(3N-5)$ modes de vibration, la rotation autour de l'axe interatomique étant supprimée

I-4. Comment prendre en compte la flexibilité ?

pas de données sur des mouvements de petite amplitude, tels que ceux opérés par les chaînes latérales d'une protéine.

Un autre outil original servant à décrire les mouvements de grande amplitude des biomolécules est le logiciel BioMove3D (76). Ce programme s'inspire des algorithmes utilisés en robotique en découpant la molécule en plusieurs parties articulées. Cette dernière est soumise à un champ de force classique de la mécanique moléculaire, et l'espace conformationnel peut ainsi être exploré en un temps de calcul bien moins important comparé à la dynamique moléculaire.

On peut également citer les méthodes MC2 (77) et ARTIST (78) qui permettent de prendre en compte la flexibilité des molécules en jeu dans l'interaction. La première combine deux techniques : le mouvement des boucles et des chaînes latérales est géré par une base de données de conformations multiples et une recherche conformationnelle par Monte Carlo est effectuée pour sélectionner les candidats. La seconde utilise la technique ART (79) pour explorer l'espace conformationnel afin de trouver les meilleurs candidats pour former le complexe.

Dans la lignée des méthodes basées sur le calcul des modes normaux, I. Bahar a été à l'initiative de la méthode GNM (80) (pour Gaussian Network Model). Cette technique permet de décrire la molécule comme un réseau de ressorts qui décrit les interactions interatomiques. Une autre méthode basée sur les modes normaux est la technique RTB (71) (pour Rotations and Translations of Blocks). La molécule est divisée en blocs représentant un groupe d'atomes et on décrit les mouvements des blocs les uns par rapport aux autres à l'aide des modes normaux de vibration pertinents, souvent de basses fréquences.

Enfin, il faut également citer une méthode développée récemment par Parrinello, s'appuyant sur la dynamique moléculaire : la métadynamique (81). L'idée est de faire évoluer la molécule dans son milieu, comme en dynamique moléculaire, et lorsque cette dernière se trouve dans un minimum local d'énergie, de « forcer » pour la faire sortir de ce bassin. De plus, l'algorithme permet d'éviter de repasser par ce sous-état conformationnel.

I-4.1.4. Limitations

Bien que toutes ces méthodes aient permis une description satisfaisante des systèmes étudiés, elles possèdent toutefois des limitations :

- La dynamique moléculaire est un outil d'analyse très précis de la flexibilité moléculaire, mais le grand nombre d'atomes impose un temps de calcul très long pour simuler l'évolution du système sur une durée permettant d'observer des mouvements importants de la molécule
- Le calcul des modes normaux donne des mouvements spontanés du système que l'on étudie. Donc, même en calculant les modes de deux molécules en interaction, cette approche ne permet pas de déterminer la réponse d'une molécule à l'action d'une force externe.
- Le fait de considérer la molécule en plusieurs parties rigides ne donne pas une description correcte, car la flexibilité de la molécule entière est à prendre en compte pour décrire la réalité

Chacune de ces approches décrit la flexibilité à différents niveaux de résolution. Lorsqu'on veut prendre en compte la flexibilité d'un système, on choisit donc la méthode en fonction de ce qu'on veut analyser. Le meilleur moyen de considérer la flexibilité globalement serait de combiner ces modèles (82). Zhao *et al* ont développé le *Flexibility Tree* (83) dans ce sens : la molécule est divisée en différentes parties et la flexibilité de chacun de ces morceaux est traitée par une méthode adaptée (modes normaux, dynamique moléculaire, données obtenues par l'expérience...).

I-4.2. La flexibilité dans les logiciels de Docking

Le but d'un logiciel de docking est de prédire la structure tridimensionnelle d'un complexe formé de deux ou plusieurs molécules. Initié au début des années 1980 (84) (85), ce domaine de recherche reste très actif de nos jours, car ces outils sont de plus en plus utilisés en recherche pharmaceutique (86), et particulièrement dans l'analyse de la formation de complexes protéines-ligands, à cause des applications en médecine (87).

Dans cette partie, sera présentée la méthode sur laquelle s'est développée la majorité des logiciels de docking, et de quelle manière la flexibilité y est traitée. Il existe d'autres méthodes pour réaliser l'arrimage de deux molécules qui s'appuient sur les techniques évoquées au paragraphe précédent. Les modes normaux (88) ou encore la dynamique moléculaire (89) sont utilisés pour réaliser ce type de simulation.

I-4.2.1. Verrous

Les deux concepts historiques évoqués dans les paragraphes précédents, les modèles « clef-serrure » et « induced-fit », ont cédé leur place à des théories plus modernes qui s'efforcent de prendre en compte la flexibilité des macromolécules (90) (91). Les molécules ne sont plus considérées comme dans une conformation stable dominante, mais comme un ensemble de conformations (25).

Le protocole de la plupart des logiciels de docking développés à l'heure actuelle se décompose en trois étapes :

- Représentation du système.
- Algorithme de recherche, grâce auquel l'espace conformationnel de chacune des molécules est balayé.
- L'affinité entre les molécules dockées est testée par le biais d'une fonction, appelée fonction de scoring.

De plus, un bon algorithme de docking doit calculer l'énergie de formation du complexe afin de déterminer si la réaction est favorable énergétiquement ou non.

En considérant un système biologique réel, il faut prendre en compte trois éléments : la molécule cible, le ligand et le solvant. Ceci représente un nombre de degrés de liberté très important que les ordinateurs actuels ne peuvent pas prendre en compte.

Afin de réduire le nombre de degrés de liberté, un modèle de solvant implicite est en général utilisé. Diverses techniques permettent alors de diminuer le temps de calcul dû au grand nombre d'atomes. Comme nous l'avons vu plus haut, la molécule peut être « découpée » en plusieurs parties rigides. Ou encore, il existe des modèles, dits gros-grain, qui considèrent un groupe d'atomes comme un seul site (92). Cela permet de gagner du temps de calcul, aux dépens de la précision.

Les premiers logiciels de docking développés au début des années 80 (84), considéraient les molécules rigides. Seuls les 6 degrés de translation et de rotation des molécules étaient pris en compte. Aujourd'hui, la puissance de calcul des ordinateurs permet de traiter la flexibilité du ligand, ainsi que du récepteur, partiellement, dans certains logiciels. Comme nous allons le voir, la flexibilité des protagonistes est prise en compte à différentes étapes de la procédure.

I-4.2.2. Représentation du système

La base des calculs de docking est le choix de la structure de départ, et la manière de la représenter. La conformation initiale de chacun des composés est une structure issue de l'expérience, la cristallographie la plupart du temps, mais également la RMN. Pour utiliser cette structure, il faut ensuite faire le choix de la représentation, qui se résume à trois possibilités :

- Coordonnées des positions atomiques de la molécule.
- Représentation par une surface qui définit l'aire accessible par d'autres molécules.
- Représentation par une grille d'énergie potentielle.

Le premier cas, le plus complexe, impose l'utilisation d'une fonction d'énergie potentielle classique de la mécanique moléculaire (93). Bien qu'en général cette représentation n'est utilisée que pour affiner le complexe final, il existe des programmes de docking qui utilisent cette représentation. C'est le cas par exemple du logiciel DARWIN (94) qui communique avec le logiciel de dynamique moléculaire CHARMM (95) pour calculer l'énergie du système.

Le deuxième choix possible est une façon très répandue de représenter le système dans les logiciels de docking. L'une des façons de définir la surface accessible par les molécules environnantes est d'utiliser la surface de Connolly (96). Elle est définie comme la partie de la surface de Van der Waals (définie par les rayons de Van der Waals des atomes) accessible par une sonde sphérique (figure 13).

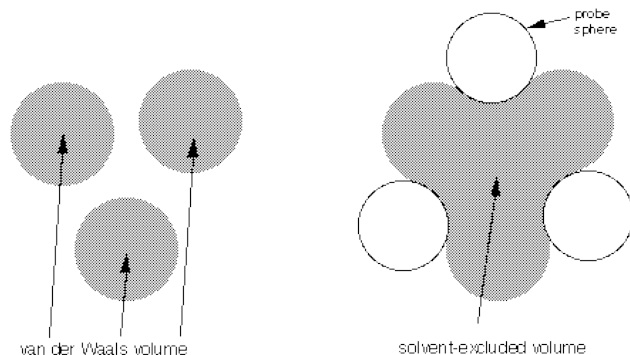


Figure 13 : surface de Connolly. Les rayons de Van der Waals définissent le volume de Van der Waals sur lequel on fait rouler une sphère. La surface résultante est la surface de Connolly.

I-4. Comment prendre en compte la flexibilité ?

La troisième possibilité repose sur le choix d'une grille représentant les interactions « non-liées », typiquement les interactions électrostatiques, Van der Waals et hydrophobe, qui n'ont qu'à être évaluée durant l'étape de scoring. Cette représentation a été imaginée par Goodford en 1985, et est aujourd'hui très répandue dans les programmes de docking.

I-4.2.3. Les algorithmes de recherche

Explorer l'espace conformationnel du système ligand-récepteur est une des difficultés majeures en modélisation, et cela reste vrai lors de la simulation de docking moléculaire. Le nombre de possibilités croît de manière exponentielle en fonction de la taille des composants (97). A titre d'exemple, pour reproduire la structure expérimentale d'un complexe protéine-ligand composés respectivement de 1562 et 449 atomes, en considérant ces derniers rigides, il faut un nombre d'essais de l'ordre de 10^7 (97). En intégrant la flexibilité des molécules, le problème devient encore plus important.

La quasi-totalité des programmes de docking proposent une approche qui prend en compte la flexibilité du ligand, tout en considérant le récepteur rigide. La recherche de candidats potentiels dans le problème du docking, en prenant en compte la flexibilité du ligand, s'appuie sur deux approches différentes :

- La recherche et le test de toutes les conformations possibles de l'espace du ligand
- Une recherche progressive, guidée à travers l'espace conformationnel

La première rassemble toutes les méthodes de recherche systématiques qui tentent d'explorer tous les degrés de liberté du système. Parmi elles, il y a :

- Les méthodes de recherche conformationnelle : les liaisons covalentes du ligand subissent une rotation de 360° avec un incrément fixe, jusqu'à ce que toutes les combinaisons aient été atteintes. Le problème de cette méthode, est que le nombre de combinaisons augmente beaucoup avec le nombre de liaisons. Il existe des moyens de réduire le nombre de conformations à explorer, mais son application reste très limitée.
- Les méthodes par fragmentation : dans un premier temps, le ligand est divisé en plusieurs fragments. Le fragment de base est placé sur le récepteur, puis les autres fragments sont attachés de manière covalente petit à petit. Cette approche est la plus répandue dans les logiciels de docking (FlexX (16), DOCK (17), ADAM (98), LUDI

(99)) pour traiter la flexibilité du ligand (25).

- Les méthodes des bases de données : afin d'éviter le problème d'explosion combinatoire, certains programmes effectuent la recherche conformationnelle des librairies de conformations prédéfinies. C'est le cas par exemple du logiciel FLOG (100) qui fait appel à une base de données de 25 conformations par molécule qui sont ensuite dockées de manière rigide.

La seconde approche englobe deux types de méthodes :

- Les méthodes aléatoires : de type Monte Carlo, utilisées par les logiciels Prodock (101) ou MCDOCK (102) par exemple. Certains logiciels utilisent également des algorithmes génétiques : un ensemble de conformations est considéré au départ, où chacune d'entre elles est définie par un ensemble de variables d'état, définies comme les gènes. L'ensemble de ces variables est donc le génotype, alors que les coordonnées font référence au phénotype. Des opérateurs génétiques (mutation, cross-over, migration) sont ensuite appliqués afin d'explorer l'espace conformationnel. La version 3.0 du logiciel AutoDock (103) utilise ce genre d'algorithme, ainsi que le logiciel GOLD (104) (105).
- Les méthodes de simulation : fondées sur la dynamique moléculaire et la minimisation de l'énergie. Bien que la dynamique soit un outil de simulation puissant, elle possède ses limitations, comme nous l'avons vu dans le paragraphe précédent. Afin de contourner ce problème, de nouvelles techniques font leur apparition, comme la simulation à des températures très élevées ou partir de plusieurs positions du ligand (86). Quant aux méthodes de minimisation, elles ne permettent d'atteindre que des minima locaux. Elle est donc plutôt utilisée en complément des autres méthodes qui ont été décrites.

La flexibilité du ligand étant prise en compte au travers de ces diverses approches, qu'en est-il du récepteur? Les programmes de docking donnent des résultats satisfaisants pour environ la moitié des systèmes étudiés (106) (107). Cela est vrai lorsque le récepteur, considéré rigide, a une conformation « dockée » proche de la structure expérimentale de départ. Mais souvent, le récepteur change de configuration lors de l'arrimage du ligand, ce qui altère les résultats du docking. L'introduction de la flexibilité de la molécule cible dans les logiciels en est encore à ses débuts, mais il existe des techniques pour y parvenir, ou au moins la traiter partiellement :

I-4. Comment prendre en compte la flexibilité ?

- La dynamique moléculaire permet de traiter la flexibilité de la même manière que pour le ligand, mais l'espace conformationnel est alors bien plus important. Lorsqu'on utilise cette technique, seule une partie du récepteur est considérée comme flexible, les temps de calcul étant trop élevés.
- En ne considérant plus une seule conformation du récepteur, mais un ensemble de structures. Dans le cas des protéines, des librairies de rotamères sont utilisées pour traiter les mouvements des chaînes latérales, mais cette approche ne rend pas correctement compte des mouvements du squelette. Certains logiciels, tels que FlexE (108) (extension de FlexX), utilisent cette approche des conformations multiples, mais la question se pose encore sur la manière de choisir ces structures (cristallographie, RMN, simulations).
- Le « docking mou » combine les informations de plusieurs conformations obtenues par l'expérience dans le but de générer une représentation moyenne basée sur l'énergie, qui est ensuite utilisée pour accrocher le ligand.

I-4.2.4. Les fonctions de scoring

Le rôle des fonctions de scoring est d'évaluer et de classer les structures prédites par l'algorithme de recherche. Elles doivent faire la distinction entre les complexes actifs et les complexes aléatoires. Evaluer de manière très rigoureuse chaque composé prendrait énormément de temps de calcul. Il faut donc faire un certain nombre de simplifications pour cette étape d'évaluation. Il y a donc trois types de fonctions de scoring :

- Les fonctions basées sur un champ de force : un champ de force classique est utilisé pour évaluer la somme de l'énergie d'interaction entre le ligand et le récepteur et de l'énergie interne du ligand. Le champ de force est en général composé d'un terme de Lennard-Jones pour les interactions de Van der Waals et d'un terme de Coulomb pour les interactions électrostatiques. L'effet du solvant peut également être pris en compte de manière implicite
- Les fonctions empiriques : elles font appel à des paramètres déterminés par l'expérience, telle que l'énergie de liaison par exemple. Cette approche a l'avantage d'être très rapide, car les termes sont simples à calculer. Mais elle dépend beaucoup des conditions d'expérience pendant la paramétrisation

- Les fonctions basées sur la connaissance : les très nombreuses structures fournies par l'expérience ont permis de calculer des statistiques à partir des fréquences d'interaction de groupes fonctionnels dans la base de données, qui permettent de calculer des potentiels
- déterminer des règles et des principes grâce auxquels ces structures expérimentales peuvent être reproduites. Les fonctions basées sur la connaissance s'appuient sur ce principe et utilisent des potentiels de paire atome-atome en général très simples. Cette méthode permet donc de tester un grand nombre de conformations rapidement, mais elle peut rencontrer des interactions entre atomes impossibles à évaluer.

Ces trois approches ont donc chacune leurs avantages, mais également leurs limites. La meilleure façon d'évaluer les complexes issus de la recherche conformationnelle est donc de combiner les informations obtenues par différentes méthodes. Mais il faut être prudent et bien évaluer la corrélation entre les fonctions de scoring, afin de ne pas amplifier l'erreur sur l'évaluation au lieu de la diminuer.

1-5. Vers un traitement de la flexibilité relative aux interactions moléculaires

Les interactions moléculaires sont les mécanismes fondamentaux du vivant : elles réalisent les fonctions élémentaires intra et intercellulaires largement gérées à partir d'instructions génétiques, codées dans le séquençage ADN. Ces instructions, toujours par ces mécanismes d'interaction complexes, créent des messages moléculaires qui permettent la mise en œuvre de ces processus biologiques essentiels. Ces messages commandent le fonctionnement métabolique du vivant et interviennent dans tous les mécanismes de défense et d'autodéfense de l'organisme. La biologie moléculaire a pour mission d'étudier ces mécanismes : elle connaît une extraordinaire phase de progrès sur l'analyse du génome humain et le développement d'une ingénierie biomoléculaire déjà très performante.

La médecine met en œuvre ces avancées dans la prévention, la guérison et le soulagement des maladies dont les origines se trouvent très souvent dans des altérations moléculaires génétiques ou accidentelles. La médecine utilise alors des molécules actives qui vont intervenir, toujours par la voie d'interactions moléculaires, en ralentissant, inhibant,

détournant les conséquences de ces altérations.

On mesure donc l'extrême importance de la compréhension et de la maîtrise de ces interactions moléculaires : lorsque l'on observe les évolutions récentes de la connaissance, on peut dire que les résultats les plus importants ont été réalisés par l'expérimentation. Mais lorsque l'on considère l'avenir, une fois les premières étapes franchies, la complexité des mécanismes et l'ampleur des expérimentations à mettre en place pour progresser, invitent à construire des outils de compréhension par des modèles et des simulations.

La voie de la modélisation à l'échelle atomique envisagée pour accompagner ces démarches expérimentales reste encore limitée à l'étude des structures et des conformations moléculaires, tirant parti des résultats expérimentaux sur l'analyse structurale par rayons X ou par RMN. Les tentatives nombreuses pour traiter des interactions entre les molécules progressent, sans atteindre toutefois, le caractère prédictif souhaitable : en pratique, l'obstacle majeur au progrès est lié à une modélisation insuffisamment précise de la flexibilité moléculaire :

- Les approches par la dynamique moléculaire sont certainement les plus attractives, mais cela requiert d'importantes ressources informatiques. En effet, le temps de calcul augmente en fonction du nombre d'atomes N : pour les modèles les plus simples, ce temps est proportionnel à N^2 , mais il est possible de diminuer ce temps en utilisant des modèles plus simples pour décrire les interactions tels que Particle Mesh Ewald ($O(N \log N)$) ou P3M ($O(N)$).
- Les approches plus statiques dites de « docking », intéressantes pour le bio-ingénieur, se heurtent justement au fait que l'on ne dispose que de modèles rigides ou semi-rigides pour chaque molécule en interaction. Ceci conduit le plus souvent à des solutions possibles multiples qui ne peuvent être validées que par l'expérience. Bien sûr, de nombreux travaux précédemment décrits, s'efforcent de surmonter ces difficultés sans pour l'instant y parvenir pleinement : notre travail participe à ces efforts en empruntant un chemin original

Dans ce qui suit, nous allons voir quelle est notre approche, en quoi elle apporte une réponse alternative et originale dans le traitement de la flexibilité biomoléculaire *a priori*, que nous détaillerons dans le prochain chapitre.

CHAPITRE I : MODELISATION ET SIMULATION AU SERVICE DE LA BIOLOGIE

La structure d'une molécule peut prendre des conformations innombrables pour des variations énergétiques faibles : c'est sa caractéristique de flexibilité. Cette molécule mise en regard d'une autre molécule va, pour interagir, prendre une conformation particulière, modification conformationnelle évidemment aussi présente dans la deuxième molécule : les chemins conformationnels sont doublement innombrables. Dans ces conditions, la voie qui consiste à pré-calculer des conformations pour analyser et comparer (« scoring ») leur performance d'assemblage paraît impossible à traiter automatiquement.

Conceptuellement, notre approche consiste à déterminer et à guider des changements conformationnels qui seraient relatifs à des interactions (inter ou intramoléculaires). L'originalité de la méthode est liée au fait que l'on va disposer d'une librairie de champs de déformation pré-calculés qui vont être utilisés à façon, en fonction des interactions en action dans le système étudié. Ces champs de déformation correspondent à des déformations permanentes de la molécule qui subit une interaction élémentaire. S'il s'agit, en définitive, de guider un processus de docking en appliquant cette vision de la flexibilité induite, cela suppose alors de résoudre les étapes suivantes :

- 1. Partir d'une structure bien définie de la molécule à étudier : conformation initiale, sites réactionnels, potentiels...
- 2. Choisir le(s) site(s) d'interaction sur chacune des molécules qu'il paraît utile de mettre en contact et d'explorer (c'est au biochimiste et à l'opérateur de les désigner).
- 3. rapprocher les molécules pas à pas en calculant à chaque étape l'énergie totale du complexe moléculaire en calculant les énergies d'interaction « attractives et stabilisantes » ainsi que les énergies de contraintes « déstabilisantes » : si ces calculs sont possibles, à chaque pas, les forces mutuellement exercées lors de l'interaction peuvent induire des champs de déformation qui, une fois appliqués, permettent de faire évoluer les conformations des objets moléculaires en contact.
- 4. On pourra ensuite vérifier la pertinence énergétique du chemin d'interaction calculé de cette façon.

Pour mettre en œuvre cette démarche, il y a des obstacles difficiles à franchir. Certaines de ces étapes figurent dans ce manuscrit : définition des champs de déformation pré-calculés (les Modes Statiques), pertinence de l'approche dans la caractérisation d'une protéine et enfin illustration d'un processus d'interaction élémentaire, une sonde électrostatique.

Notre problématique est donc la suivante :

Notre travail s'insère dans la logique que nous venons de décrire. En partant de l'outil de calcul FLEXIBLE, il vise à considérer le problème du côté des interactions et d'établir un cheminement pouvant conduire à une amélioration des modèles de « docking » actuels. Il s'agit de définir les premiers opérateurs de calcul :

- Caractériser les déformations et les conformations moléculaires
- Calculer les interactions à longue distance et mesurer leur impact sur la conformation moléculaire
- Explorer la manière d'introduire les interactions répulsives à courte distance et traiter l'encombrement stérique
- Préciser la démarche globale de simulation dans la perspective de créer un outil automatique et prédictif

Pour réaliser ces objectifs, nous nous appuierons :

- Sur des étapes de la connaissance déjà bien maîtrisées, pour les calculs de structure : toute une série d'outils informatiques sont proposées et couramment utilisés aujourd'hui. L'opérateur accède à une série de services en ligne qui rendent ces outils de plus en plus conviviaux.
- Sur l'outil FLEXIBLE qui calculera la réponse conformationnelle en termes de variation énergétique et de déplacements relatifs de tous les atomes de la structure, conséquence de l'application d'une force ou d'un mouvement de déplacement sur un atome de la structure dans une direction définie par l'interaction. Le point de départ est le résultat d'un calcul de structure réalisé par un outil référent. FLEXIBLE vient extraire de ce résultat la matrice des dérivées secondes de l'énergie : la matrice hessienne. Comme nous allons l'explicitier dans le second chapitre, les données numériques de cette matrice vont être utilisées pour calculer les variations des positions des atomes lorsque l'un d'entre eux est sollicité. A noter qu'avec les résultats de M. Brut, un outil simple et efficace a été mis à disposition pour défricher le comportement d'une molécule sollicitée en des points choisis ; l'application de FLEXIBLE à la protéine HIV-1 du sida montre que l'outil peut aider efficacement le biologiste dans sa quête des molécules actives puisque, en quelques semaines de

calcul, on retrouve des caractéristiques qui par la voie empirique ont mobilisé des équipes nombreuses et performantes sur plusieurs années.

I-6. Conclusion

Nous avons pu voir tout au long de ce chapitre la place importante qu'occupent actuellement les outils de simulation dans le monde de la biologie moléculaire. Car, si historiquement, c'est par l'expérience que les grandes avancées ont été effectuées dans ce domaine, il est aujourd'hui nécessaire de faire appel à la modélisation afin de comprendre, analyser, prédire les mécanismes essentiels du vivant.

L'accent a été mis dans cette partie sur les interactions entre biomolécules et les changements conformationnels qu'elles peuvent entraîner. Ces mouvements atomiques se trouvent être un problème majeur pour les équipes de modélisation, de par la difficulté à les prendre en compte. Ceci s'applique plus particulièrement lorsque l'on s'intéresse aux mécanismes de « docking » entre macromolécules : les outils existants, bien que performants, possèdent leurs limites qui restreignent leur utilisation.

Nous avons, au sein du laboratoire, imaginé une nouvelle méthode capable de calculer les déformations des macromolécules, en termes de déplacement atomique, en réponse à une force externe : les Modes Statiques. Cette nouvelle approche, que nous allons détailler dans le second chapitre, contourne les principales difficultés rencontrées dans les outils actuels : le temps et la précision du calcul. Ceci en fait une méthode idéale pour la conception d'un programme de « docking ».

CHAPITRE 2 : METHODOLOGIE

SOMMAIRE

INTRODUCTION	- 54 -
II-1. Recherche d'un mode d'introduction des déformations dans un calcul de docking.....	- 54 -
II-1.1. Les modes normaux	- 55 -
II-1.1.1. Origine.....	- 55 -
II-1.1.2. Les modes normaux dans une procédure de docking.....	- 56 -
II-1.2. Les Modes Statiques	- 58 -
II-2. Le calcul des Modes Statiques	- 60 -
II-3. Validation de la méthode	- 62 -
II-4. Avantages et limites de la méthode	- 64 -
II-5. Les outils développés à partir des modes statiques	- 66 -
II-5.1. Description de la flexibilité moléculaire.....	- 67 -
II-5.1.1. Optimisation de la force.....	- 67 -
II-5.1.2. Modes de pincements.....	- 70 -
II-5.1.3. Grandeurs caractéristiques	- 71 -
II-5.2. Arrimage d'un ion métallique	- 72 -
II-6. Conclusion	- 74 -

INTRODUCTION

Les méthodes utilisées actuellement pour introduire la flexibilité macromoléculaire dans les calculs de docking ont fait leur preuve et permettent de résoudre une part significative des besoins. Ces approches restent tout de même limitées, car, comme nous l'avons évoqué dans la partie précédente, les chemins choisis, par principe, restreignent leur utilisation à des cas bien précis.

Une approche particulière nous intéresse qui se fonde sur une analyse préalable de la flexibilité de chacune des molécules en présence par le calcul des modes normaux des structures. Cette méthode consiste à diversifier les conformations d'interaction : les outils les plus récents sont fondés sur cette approche que nous allons considérer plus en détail dans les prochains paragraphes.

Notre ambition est de construire une approche nouvelle basée sur le suivi automatique d'un chemin réactionnel contrôlé. Cette voie originale implique que l'on dispose d'une méthode pour calculer l'énergie de déformation des molécules sous contrainte mécanique. Cette approche peut s'appliquer lorsqu'une molécule est en interaction, et pourra donc s'appliquer au docking des macromolécules.

Nous allons dans un premier temps découvrir les Modes Statiques : quel est le concept, comment les calculer et où se place cette nouvelle approche par rapport à celles utilisées généralement par la communauté. A partir des Modes Statiques, nous avons développé des programmes qui nous ont permis : (i) de décrire la flexibilité de biomolécules et (ii) de mettre en évidence les changements conformationnels d'une molécule suite au passage d'un cation métallique.

II-1. Recherche d'un mode d'introduction des déformations dans un calcul de docking

Afin de faire reculer les limites des méthodes de modélisation du docking évoquées au chapitre précédent, il est nécessaire de trouver un moyen de calculer les déformations des macromolécules à partir de leur position d'équilibre. Une méthode a été largement explorée pour faire cela, basée sur le calcul des modes normaux. Elle a déjà été évoquée au chapitre I,

II-1. Recherche d'un mode d'introduction des déformations dans un calcul de docking et nous allons la discuter un peu plus en détails dans ce paragraphe. L'essentiel est ici consacré à une alternative à ce calcul, qui est la méthode imaginée par notre groupe de recherche : les Modes Statiques.

II-1.1. Les modes normaux

II-1.1.1. Origine

L'idée d'utiliser les modes normaux pour décrire les mouvements atomiques des biomolécules a été suggérée E. B. Wilson dans les années 50 (109). Au début des années 80, les capacités de calcul grandissantes permettent à Nabuhiro Gö et collaborateurs d'exploiter cette idée (110). Leurs travaux, qui s'appuient sur l'étude de la protéine BPTI (pour Bovine Pancreatic Trypsin Inhibitor), montrent que la surface d'énergie potentielle à proximité d'un minimum correspondant à une structure expérimentale, peut être représentée, en première approximation, par une parabole. Cela signifie que l'on peut considérer un potentiel harmonique pour décrire l'énergie conformationnelle de la protéine, lorsque cette dernière est proche de son état natif.

Or, lorsque l'on se place dans l'approximation harmonique, les mouvements collectifs des atomes sont donnés par les modes normaux de vibration, calculés à partir de la matrice hessienne des dérivées secondes de l'énergie. On peut alors envisager d'utiliser les modes normaux pour décrire la dynamique de la molécule à proximité d'un minimum d'énergie.

C'est ce qu'a réalisé cette même équipe peu de temps après cette découverte, en déterminant la dynamique de la protéine BPTI à partir des modes normaux de vibration (111). Ainsi, il a été établi que la molécule se comportait de manière harmonique pour des fréquences supérieures à 50 cm^{-1} , alors que les mouvements collectifs pertinents des atomes ont des fréquences inférieures à 30 cm^{-1} . Ces fréquences représentent des mouvements d'une durée de quelques picosecondes, ce qui est inférieur aux temps sur lesquels se déroulent les mécanismes biologiques, qui varient de la microseconde à la minute. De plus, les modes de basse fréquence correspondent à des mouvements atomiques délocalisés, alors que pour des fréquences supérieures, on observe des mouvements localisés sur des sites de la molécule (voir figure 14). Les auteurs s'interrogent donc sur la validité de cette méthode pour décrire la dynamique de la molécule et concluent en suggérant que les amplitudes des mouvements de

basse fréquence sont telles qu'on ne peut plus considérer l'approximation harmonique. Il est donc impératif de prendre en compte les effets anharmoniques, afin de tester son influence.

Peu de temps après, Brooks et Karplus ont effectué le même calcul, en utilisant un champ de force différent, et l'ont comparé à un calcul de dynamique moléculaire (95). La dynamique moléculaire rendant compte des effets anharmoniques, ils ont pu étudier l'impact de l'approximation harmonique nécessaire au calcul des modes normaux. Leur étude datant de 1983, la durée de la simulation est très court (<100 ps). Néanmoins, le calcul montre que les fluctuations données par la dynamique moléculaire et celles données par les modes normaux coïncident : la dynamique d'une molécule, aux alentours d'un minimum, peut être décrite par les modes normaux de vibration. Cette affirmation est valable pour cette molécule, mais n'est pas toujours vraie pour d'autres molécules.

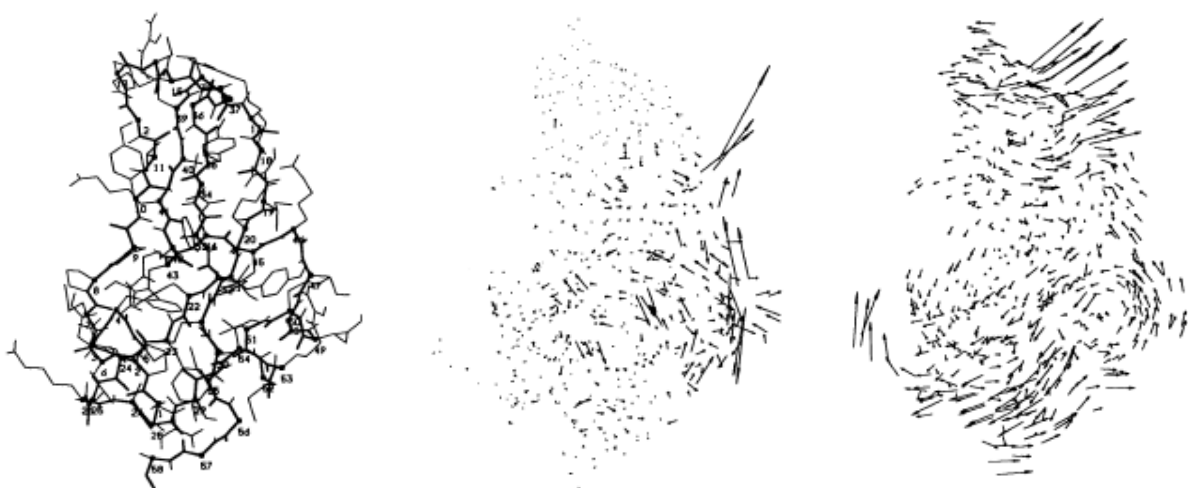


Figure 14 : figures illustrant les déplacements atomiques pour deux modes normaux de vibration de la protéine BPTI. A) squelette (traits gras) et chaînes latérales (traits fins), les carbones α sont numérotés. B) vecteurs représentant les déplacements atomiques pour un mode de haute fréquence (118.8 cm^{-1}) : les mouvements atomiques sont localisés. C) vecteurs représentant les déplacements atomiques pour un mode de basse fréquence (6.9 cm^{-1}) : les mouvements atomiques sont délocalisés (source ref. 97).

II-1.1.2. Les modes normaux dans une procédure de docking

Il est donc possible de traiter les mouvements collectifs d'une biomolécule à l'aide des modes normaux de vibration de basses fréquences. Cela a ouvert de nouvelles perspectives

II-1. Recherche d'un mode d'introduction des déformations dans un calcul de docking

dans le traitement des changements conformationnels. Car, si on peut s'interroger sur la pertinence de la méthode pour décrire la dynamique d'une molécule, l'utilisation des modes normaux est un outil intéressant dans le but de générer un échantillon de conformations à proximité d'un minimum énergétique.

A partir de cette idée, Kolossváry et Guida ont imaginé une méthodologie, basée sur les modes normaux de basse fréquence, afin d'explorer l'espace conformationnel aux alentours d'un minimum local : le programme LMOD (112). La procédure est assez simple : on détermine une conformation dans un minimum local d'énergie, on calcule ensuite la matrice hessienne de la molécule, puis on stocke les modes de basses fréquences, situées en dessous d'un seuil déterminé par l'utilisateur (typiquement 250 cm^{-1}). L'étape suivante consiste à déformer la molécule dans les directions données par chaque mode, jusqu'à ce que l'énergie dépasse une valeur seuil, elle aussi choisie par l'utilisateur. Cela permet donc d'explorer l'espace conformationnel d'une molécule.

A partir de LMOD, cette même équipe a décidé de l'appliquer à un programme de docking, qui s'inscrit dans le schéma classique employé par la plupart des logiciels actuels, que j'ai énoncé au paragraphe **I-4.2.** :

- On construit « à la main » une structure du complexe formé des deux molécules, dont on optimise la structure (par minimisation de l'énergie) en utilisant un logiciel de mécanique moléculaire classique.
- Le complexe est ensuite passé dans LMOD afin de déterminer un certain nombre de structures proches de la position d'équilibre, qui sont testées ensuite par un calcul énergétique

Cette technique permet donc de générer une conformation adaptée des molécules lors du docking. Une étude a été menée en utilisant cette approche sur le peptide HIV integrase, en considérant plusieurs ligands (113). Les résultats coïncident bien avec les observations expérimentales.

Plus récemment, Cavasotto *et al.* ont mis au point un programme de docking, qui s'appuie sur les modes normaux, et qui rentre dans le schéma classique « recherche conformationnel + scoring » (114). L'idée est de sélectionner directement un petit nombre de modes normaux, pertinents, que l'on va utiliser pour déformer la molécule. Les quelques

conformations obtenues sont alors testées par le biais de fonctions de scoring.

Cette technique a été appliquée à la famille de protéines kinase, et grâce à un algorithme qui permet de déterminer les modes de vibration pertinents, le nombre de modes retenu pour la kinase est de 9, et 5 pour le ligand. A partir de ces modes, il est facile de générer plusieurs conformations qui seront ensuite associées et testées par criblage virtuel.

II-1.2. Les Modes Statiques

La méthode des Modes Statiques permet de déterminer les déformations d'une macromolécule à partir d'une conformation initiale. On considère un état déformé permanent (sous l'action d'une force externe) sans considérer de trajectoire continue dans l'espace des configurations. Cette approche peut donc être vue comme une alternative aux modes normaux, mais avec des principes de base complètement différents.

Par le calcul des modes normaux, on regarde les mouvements atomiques autour d'une position d'équilibre dus à l'agitation thermique au sein de l'organisme. **L'idée des Modes Statiques est de calculer les déformations d'une biomolécule en réponse à une force externe.**

Cette force peut être de différentes natures : mécanique, champ électrique, mais ce qui nous intéresse dans notre cas, cette force peut être la conséquence d'une interaction entre deux molécules.

Cette méthode ouvre donc de grandes perspectives dans l'étude des biomolécules à l'échelle atomique. En effet, grâce aux Modes Statiques, il sera possible de prendre en compte pleinement la flexibilité des macromolécules, dans le but de décrire les déformations induites par une action externe sur ces dernières. Il est alors concevable d'incorporer cette nouvelle approche dans une méthode de modélisation d'un mécanisme de docking. Pour le réaliser, nous avons imaginé deux approches.

Une première méthode consiste à incorporer cette nouvelle approche dans un schéma « classique » de docking, de la même manière que les modes normaux, où l'on générerait des conformations, grâce aux modes statiques. Mais, contrairement aux méthodes basées sur les modes normaux, notre approche permet de déterminer des conformations pertinentes en appliquant une force sur un site spécifique. Le protocole serait alors le suivant :

II-1. Recherche d'un mode d'introduction des déformations dans un calcul de docking

- Bien connaître la structure fondamentale de chacune des molécules que l'on veut étudier : conformation initiale, sites réactionnels potentiels...
- Choisir le(s) site(s) d'interaction potentiel(s) sur chacune des molécules sur le(s)quel(s) nous allons appliquer une force
- Appliquer une force sur chacun de ces sites et générer grâce aux modes statiques des conformations déformées pertinentes
- Tester tous les complexes que l'on peut former avec ces différentes conformations par l'intermédiaire d'une fonction de scoring

Nous avons également imaginé notre propre logiciel de docking, dans lequel les molécules se rapprochent pas-à-pas, les déformations étant calculées à chaque pas. La mise en œuvre est donc différente de celle présentée ci-dessus, car la molécule va subir un certain nombre de translations/rotations. La méthode se décompose alors selon les étapes suivantes :

- Bien connaître la structure fondamentale de chacune des molécules que l'on veut étudier : conformation initiale, sites réactionnels potentiels...
- Choisir le(s) site(s) d'interaction sur chacune des molécules qu'il paraît utile de mettre en contact et d'explorer
- Effectuer des pas de translation/rotation en minimisant les forces entre les sites d'interaction, dans le but de rapprocher les deux molécules
- A chaque pas de translation/rotation, on déforme par l'intermédiaire des modes statiques chacune des molécules en utilisant les forces d'interaction mises en jeu
- Vérifier la pertinence du chemin par un calcul énergétique précis donnant la structure du complexe ainsi formé

Cette approche qui détermine les déformations dues aux interactions à chaque pas de translation/rotation des molécules donne une description précise du mécanisme de docking entre macromolécules. Ce chemin suit le concept « induced-fit » et permet alors de décrire les phénomènes de docking proches de la réalité.

Enfin, il est également concevable de construire un modèle de docking autour de FLEXIBLE en ne connaissant pas, ou en connaissant partiellement, la structure initiale des molécules. Dans ce cas, il est donc nécessaire d'ajouter une étape où un certain nombre de conformations

de départ sont générées par FLEXIBLE en déformant les molécules en appliquant un-à-un les modes statiques.

II-2. Le calcul des Modes Statiques

Comme pour le calcul des modes normaux, le calcul des modes statiques est basé sur la matrice hessienne, qui est la matrice des dérivées secondes de l'énergie totale du système par rapport aux coordonnées atomiques. Les éléments de cette matrice sont fonction des paramètres de force utilisés dans les champs de force classiques.

En retirant trois coordonnées pour la translation et trois pour la rotation, on retrouve une matrice de dimension $(3N - 6) \times (3N - 6)$. L'idée est d'appliquer ensuite une contrainte, dans le sens de Lagrange, sur la molécule.

L'énergie totale du système peut s'écrire sous la forme d'une série entière :

$$E = E_0 + H_{ij}x_i x_j + K_{ijk}x_i x_j x_k + L_{ijkl}x_i x_j x_k x_l + \dots \quad \text{Equation 1}$$

où H_{ij} sont les éléments de la matrice hessienne, K, L, \dots les matrices représentant les dérivées d'ordres supérieures de l'énergie et x_i les coordonnées cartésiennes d'un déplacement atomique par rapport à une configuration d'équilibre.

Le vecteur force peut s'exprimer comme le gradient de l'énergie, ce qui donne, au voisinage de l'équilibre :

$$-F_i = \frac{\partial E}{\partial x_i} = \sum_j \left[H_{ij} + \sum_k K_{ijk} x_k + \sum_{k,l} L_{ijkl} x_k x_l \right] x_j = \sum_j H'_{ij} x_j \quad \text{Equation 2}$$

A l'heure actuelle, la méthode des Modes Statiques se place dans l'approximation harmonique. Cela signifie que H' est en fait la matrice hessienne. Le calcul des Modes Statiques passe donc par un calcul préalable d'une structure optimisée, ainsi que de la matrice hessienne. Ces dernières peuvent être calculées soit de manière quantique, via les méthodes décrites au chapitre précédent, soit de manière classique en utilisant un champ de force de la mécanique moléculaire. Des logiciels tels que GAUSSIAN03 ou AMBER8 permettent de

calculer la matrice hessienne. En revanche, le calcul des termes d'ordres supérieurs qui traduisent les effets dus à l'anharmonicité, n'est pas direct et nécessite un calcul supplémentaire.

Revenons à l'équation 2. A l'équilibre et en l'absence de forces extérieures, on a :

$$x_j = 0, \quad \forall j \quad \text{Equation 3}$$

Si on applique une force sur un atome dans une direction de l'espace, c'est-à-dire que tous les F_k sont nuls sauf un, on peut retrouver le déplacement correspondant x_k en résolvant le système Eq. 2. Une autre manière équivalente d'aborder le problème est d'imaginer un déplacement x_k résultant de la force F_k . Dans le premier cas, les déplacements sont normalisés à la force appliquée, alors que dans le second cas, ils sont normalisés au déplacement x_k .

Nous avons opté pour la seconde approche, car cela nous permet de traiter le problème par le formalisme lagrangien. En effet, la contrainte imposée en fixant x_k constitue la contrainte de Lagrange, alors que F_k correspond à une force de réaction que l'on n'est pas obligé de connaître dans le formalisme de Lagrange. Bien que cette approche soit plus complexe, cela permettra de transposer plus facilement la méthode des Modes Statiques aux coordonnées internes, qui peuvent alors être considérées comme des coordonnées généralisées, dans le sens de Lagrange. De plus, en utilisant cette approche, cela évite le calcul direct des forces de répulsion, souvent difficiles à estimer.

D'un point de vue mathématique, on supprime la l -ième ligne de la matrice hessienne, correspondant associée à la contrainte F_l du système Eq.2. D'un autre côté, tous les termes contenant x_l , soit la l -ième colonne, ne sont plus des inconnues. On note cette colonne B et A la matrice hessienne à laquelle on a supprimée une ligne et une colonne. On doit alors résoudre le nouveau système $A.X + B = 0$, où X représente alors un vecteur contenant les déplacements atomiques imposés par la contrainte F_l . Il existe plusieurs façons de résoudre ce système. Nous avons choisi la méthode de Gauss-Seidel, un algorithme bien adapté aux déformations moléculaires, car il permet une propagation régulière des déformations tout le long de la molécule, depuis le point d'excitation.

Le vecteur X solution du système d'équation $A.X + B = 0$ représente les déformations dues à une force appliquée à un endroit donné de la molécule et est appelé un Mode Statique. Tel que l'algorithme est écrit, la force est de 1eV/Å. En éliminant une à une les lignes et les colonnes correspondantes et en résolvant à chaque fois le système $A.X + B = 0$, on obtiendra

tous les vecteurs X que l'on peut alors stocker dans une matrice.

On peut également calculer l'énergie de chaque déformation, grâce à la formule :

$$E = \frac{1}{2} \sum_{i,j} A_{ij} x_i x_j \quad \text{Equation 4}$$

Un programme informatique a été élaboré afin de calculer et de stocker ces Modes Statiques : le programme FLEXIBLE. On peut ainsi déterminer les déformations imposées par l'application d'une force extérieure pour n'importe quelle molécule. Une fois les modes calculés, ils sont stockés dans une matrice des déformations, ainsi que l'énergie de chacune de ces déformations. Le protocole d'utilisation de FLEXIBLE est détaillé en annexe.

II-3. Validation de la méthode

La méthode des Modes Statiques et le programme FLEXIBLE qui les calcule ont été validés par les travaux de thèse de Marie Brut, présentés récemment au LAAS-CNRS. Les principaux résultats de cette thèse, détaillés dans la référence 20, ont démontré que les Modes Statiques pouvaient être utilisés pour décrire des mouvements atomiques pertinents sur différentes molécules.

La méthode a été validée à travers trois exemples portant sur des systèmes variés, et de différentes tailles : un polymère (le PNIPAM), duplex de nucléotides et une protéine (protéase du HIV-1).

Le premier exemple d'application des Modes Statiques porte sur le PNIPAM, polymère thermosensible, qui présente une transition de phase hydrophile/hydrophobe réversible à une température caractéristique de 32°C (115). En dessous de cette température, le polymère est hydrophile et adopte une forme de « pelote », gonflée par l'eau qui s'insère entre monomères adjacents. Au dessus de cette température, le polymère se replie sur lui-même, expulsant l'eau afin de présenter une surface minimale au solvant, et se présente sous la forme de « globules ».

L'étude portait sur un dimère, puis sur un tétramère, qui sont des systèmes de petite taille (46 et 84 atomes respectivement), ce qui a permis d'analyser la flexibilité moléculaire à l'échelle

atomique. La taille des systèmes étant relativement petite, l'optimisation de géométrie s'est faite à l'aide de potentiels quantiques.

Le site actif du polymère peut être identifié comme la liaison hydrogène qui se forme entre les deux monomères. En effet, pour qu'une molécule d'eau s'insère entre les deux brins, il est nécessaire de briser cette liaison hydrogène. L'étude sur le dimère a ainsi permis d'établir le site sur lequel agir afin d'obtenir une élongation maximale de la liaison hydrogène, et donc de la briser. Ce site est une terminaison isopropyle appartenant au monomère qui contient l'atome d'hydrogène de la liaison.

La même étude a été effectuée sur un tétramère, et donc sur les trois liaisons hydrogène inter-brins. Les résultats obtenus sont tous équivalents : pour briser une liaison hydrogène, il faut agir sur cette terminaison isopropyle du monomère possédant l'atome d'hydrogène de la liaison considérée.

Ce résultat s'avère très utile pour les expérimentateurs, révélant une zone particulière ayant une réelle influence sur le site actif, et non connue jusque là. Ce type d'étude peut donc permettre d'identifier de nouveaux mécanismes d'activation ou d'inhibition.

L'exemple suivant portait sur l'étude de la flexibilité des sucres dans les acides nucléiques. Dans la structure des acides nucléiques que nous avons vue au paragraphe **I-1.2.1.**, le sucre joue un rôle de pivot entre les bases azotées et les groupes phosphate, ce qui lui confère un rôle essentiel dans les changements conformationnels des acides nucléiques. Le furanose, connu pour être hautement flexible (116), peut adopter deux types de conformations privilégiées dans un simple brin d'ADN, appelées *nord* et *sud* par les spectroscopistes. En revanche, dans un duplex, le sucre adopte une conformation privilégiée : *nord* dans une conformation de type A, et *sud* dans une conformation de type B.

L'application des Modes Statiques à l'étude du nucléotide désoxythymidine monophosphate (dTMP) seul, met en évidence la flexibilité du sucre dans un simple brin, et notamment cette capacité qu'il a de passer d'une conformation *nord* à une conformation *sud*. Ensuite, le dTMP a été couplé au désoxyadénosine monophosphate (dAMP). Dans ce système, on trouve que le sucre est rigide et reste dans sa conformation *sud*.

Les Modes Statiques ont également été utilisés pour l'étude d'acides nucléiques contraints. Ces derniers sont synthétisés dans le but d'obtenir des duplex plus stables et sont particulièrement étudiés ces dernières années. En se basant sur des résultats expérimentaux

CHAPITRE II : METHODOLOGIE

récents portant sur des ADN contraints nommés D-CNA (117) (118) (119), l'augmentation de la rigidité de ces molécules a été mise en évidence par les modes statiques. De plus, comme pour l'étude précédente, les sites sur lesquels appliquer les forces ont été explicités, ce qui démontre une fois de plus le caractère prédictif de notre approche.

La troisième application porte sur l'étude des sites actifs de la protéase du VIH-1, que l'on peut trouver en détail dans la thèse de Marie Brut (120). Ceci constitue le premier exemple d'analyse d'une macromolécule par les Modes Statiques. Il a été démontré à travers cette application la puissance et la cohérence de notre approche. En effet, grâce à un calcul principal, nous avons retrouvé un nombre important de propriétés connues, obtenues par l'expérimentation ou la simulation durant ces vingt dernières années. En appliquant des forces adéquates sur des sites actifs connus de la protéase, on retrouve par les Modes Statiques, les effets observés lors de précédents travaux. Cela comprend notamment : les changements de flexibilité de deux conformations de la protéine (121), la détection d'un site allostérique (122), les mouvements atomiques obtenus par la simulation de l'arrimage d'un ligand (123)...

Ces premiers résultats obtenus au travers d'exemples variés, mettent en évidence l'étendue et la diversité des applications potentielles de la méthode des Modes Statiques. Tout d'abord, les différentes tailles de ces systèmes montrent qu'il est possible d'utiliser notre méthodologie pour l'analyse précise de mouvements atomiques simples, mais également dans le cas de mouvements plus globaux de plusieurs atomes. Ensuite, la fiabilité de l'approche a été confirmée, car ces résultats coïncident avec les observations des expérimentateurs. Enfin, il a également été mis en évidence que la méthode des Modes Statique s'avère être un outil prédictif puissant.

II-4. Avantages et limites de la méthode

La méthode des Modes Statiques a été énoncée, et nous savons également comment les calculer. Il est donc maintenant intéressant de savoir où se place notre nouvelle approche par rapport aux techniques utilisées actuellement pour traiter la flexibilité macromoléculaire et l'intégrer dans un programme de docking.

Nous avons vu que les Modes Statiques ont été élaborés afin de prédire les changements conformationnels d'une macromolécule, sans calculer les états de transition intermédiaires.

L'analyse des modes normaux de vibration est une autre manière d'aborder ce problème, mais c'est une méthode qui possède un certain nombre de limites qui restreignent son utilisation.

Premièrement, il y a une restriction sur l'état de la structure que l'on veut étudier. En effet, les modes normaux se limitent à une position d'équilibre. En absence de forces externes, ceci signifie que l'on ne peut pas les calculer sur une molécule en interaction. En outre, la détermination des Modes Normaux nécessite une optimisation de la structure très précise. Autrement, cela donne lieu à des valeurs propres négatives de la matrice dynamique, et donc à des fréquences imaginaires. Cette condition d'équilibre de la molécule isolée n'est pas nécessaire pour le calcul des Modes Statiques.

En effet, la théorie repose sur le fait que ces modes sont excités par une force externe. La molécule se retrouve alors dans un état déformé. Le fait qu'elle soit loin de son état d'équilibre au moment où l'on applique la force n'importe pas. Dans ce manuscrit, nous avons effectué tous les calculs sur des structures proches d'un minimum. C'est pourquoi dans le paragraphe **II-2.** nous avons présenté le calcul des modes au voisinage d'un point d'équilibre. Mais dans le cas du calcul des modes statiques, l'optimisation de la structure n'a pas besoin d'être effectuée avec une précision aussi importante que pour les modes normaux, car nous n'avons pas ce problème de fréquence imaginaire résultant de valeurs propres négatives de la matrice hessienne.

Ensuite, il se pose le problème de l'approximation harmonique, indispensable pour le calcul des modes normaux. Or, dans le cas des protéines, cette approximation est vraie aux basses températures. A partir d'une certaine température de transition, entre 100 et 240K suivant la protéine étudiée (124) (125), des mouvements anharmoniques sont présents, ce qui donne lieu à l'apparition de sous-états conformationnels qui peuvent avoir leur importance dans l'activité de la molécule (126) (127) (128). La théorie des Modes Statiques n'impose pas l'utilisation de l'approximation harmonique. En effet, comme nous l'avons vu précédemment (*Eq.1*), le terme H'_{ij} contient d'autres termes qui vont dépendre des coordonnées atomiques. Le système *Eq.2* devient non linéaire, la méthode de Gauss-Seidel peut toujours être utilisée mais le nombre d'itérations requises sera plus important, ce qui impose alors une méthode de résolution plus complexe. La version actuelle de FLEXIBLE ne permet pas de prendre en compte l'anharmonicité, mais des efforts sont faits pour l'incorporer dans les prochaines versions du programme.

CHAPITRE II : METHODOLOGIE

La méthode des Modes Statiques présente également ses limites, qui sont toutefois surmontables. Le principal problème pour le calcul des modes vient du système de coordonnées utilisé. Le fait d'utiliser des coordonnées cartésiennes rend le calcul lourd lorsque l'on veut traiter des macromolécules d'intérêt biologique. La matrice hessienne devient rapidement très grande avec l'augmentation du nombre d'atomes, et son traitement dans FLEXIBLE devient long. Le calcul des modes de l'amyloïde β (1-28) (28 résidus, 438 atomes) prend environ 40 minutes CPU, alors que pour la protéase du VIH (99 résidus, 3128 atomes), il faut compter 15 heures CPU.

Deux solutions sont actuellement envisagées pour réduire le temps de calcul. La première porte sur l'algorithme d'inversion de la matrice hessienne, sans abandonner le recours aux coordonnées cartésiennes. En effet, décrivant les positions atomiques par des coordonnées cartésiennes, nous rencontrons des matrices hessienne de grande taille, dont la majorité des éléments a une valeur nulle, à cause de la portée limitée des interactions atomiques. Ce nouvel algorithme simplifié, actuellement en cours d'évaluation, tient compte de ce nombre limité d'interactions atomiques.

La seconde solution s'intéresse aux effets dus à l'anharmonicité, qui nécessitent le traitement d'un système d'équations non linéaires. Au-delà du temps de calcul qui est plus long, une solution globale ne peut plus se résumer dans une matrice unique, mais demande un ensemble plus ou moins grand de matrices, en fonction du degré de non linéarité et de la précision recherchée. Il faut donc reconsidérer l'approche de FLEXIBLE, notamment dans le système de représentation de la molécule et des déformations.

L'idée est alors de changer la représentation moléculaire en remplaçant les coordonnées cartésiennes par les coordonnées généralisées, au sens de la mécanique de Lagrange. En choisissant ce mode de représentation approprié, on peut alors diminuer de manière très significative le temps de calcul des modes statiques. De plus, cela permet de traiter les effets anharmoniques du système, et donc de générer un plus grand nombre de conformations à partir d'un calcul unique de modes.

II-5. Les outils développés à partir des modes statiques

Une fois les Modes Statiques calculés et stockés dans la matrice des déformations grâce au programme FLEXIBLE, on peut les utiliser pour développer un certain nombre de

programmes. Je vais présenter dans cette partie ceux que j'ai développés pour effectuer mes calculs. Il s'agit dans un premier temps de programmes pour décrire la flexibilité des molécules, puis je parlerai de la sonde électrostatique, qui constitue le premier programme de docking basé sur les modes statiques.

II-5.1. Description de la flexibilité moléculaire

Afin de décrire la flexibilité des biomolécules, un certain nombre de programmes ont été mis au point. Certains nous ont permis de cartographier la flexibilité, comme nous le verrons dans le troisième chapitre.

II-5.1.1. Optimisation de la force

Nous avons conçu un outil capable de déterminer la direction optimale d'une force à appliquer sur un atome lorsque l'on veut maximiser une grandeur caractéristique, telle qu'une distance interatomique ou un angle dièdre.

Maximisation d'une distance

Nous nous trouvons ici dans le cas d'une optimisation sous contraintes, car nous voulons ajuster l'orientation d'une force tout en conservant sa norme.

Dans le cas présent, plaçons-nous dans le cas simple de la maximisation de la distance entre deux atomes N_1 et N_2 . On extrait de la matrice des Modes Statiques les deux sous-matrices (3×3) , m_{N_1} et m_{N_2} , contenant les déformations associées aux atomes N_1 et N_2 par l'application de trois forces dans les trois directions de l'espace, sur un atome N_0 .

La variation de la distance N_1N_2 , dr , est associée à la matrice $m = m_{N_2} - m_{N_1}$. La direction optimisée est alors donnée par la plus grande valeur propre de la matrice $m^t m$, que l'on détermine par la méthode des puissances itérées (129).

La variation maximale dr_{max} est reliée à la force optimale $|\varphi_{max}\rangle$ par la relation :

$$|dr_{max}\rangle = m|\varphi_{max}\rangle$$

Afin de déterminer $|\varphi_{max}\rangle$, on définit tout d'abord un vecteur quelconque $|\varphi_0\rangle$, normé. Une première approximation de la valeur propre associée λ_0 , est donnée par la relation :

$$\lambda_0 = \frac{\langle \varphi_0 | h | \varphi_0 \rangle}{\langle \varphi_0 | \varphi_0 \rangle}$$

$$\text{Où } h = m^t m$$

Chaque nouvelle force est calculée par la relation $|\varphi_{n+1}\rangle = h|\varphi_n\rangle$. Le critère d'arrêt est défini à partir de la valeur propre λ_n calculée à chaque itération par la formule :

$$\lambda_n = \frac{\langle \varphi_n | h | \varphi_n \rangle}{\langle \varphi_n | \varphi_n \rangle}$$

On continue jusqu'à ce que λ_n/λ_{n+1} converge. Lorsque la convergence est atteinte, on garde $|\varphi_{max}\rangle$ comme direction optimale.

Cette direction $|\varphi_{max}\rangle$ est alors appliquée au système et l'on peut mesurer la variation de notre distance N_1N_2 : $|dr_{max}\rangle = m|\varphi_{max}\rangle$. En définissant $|u\rangle$ comme vecteur unitaire le long de l'axe interatomique N_1N_2 , on obtient la distance maximale par :

$$d_{N_1N_2} = \langle u | dr_{max} \rangle$$

Maximisation d'un angle dièdre

Nous venons de voir que pour maximiser une distance d telle que $d^2 = \langle r | r \rangle$, nous devons trouver la plus grande valeur propre de la matrice $m^t m$, où m est la sous-matrice définie précédemment.

Nous nous plaçons ici dans le cas de la maximisation d'un angle dièdre. On peut écrire :

$$\Delta\theta = \sum_{i=1}^{Nat} \sum_{k=1}^3 \frac{\partial\theta}{\partial x_{ik}} dx_{ik}$$

Où les indices i et k représentent les atomes et les coordonnées respectivement. Les dx_{ik} sont les déplacements atomiques sous l'action des forces exercées. Leur expression, sous forme vectorielle, est donnée par :

$$|dx\rangle = M|F\rangle$$

Où M est la matrice des modes statiques et $|F\rangle$ le vecteur force à 3N composantes.

Naturellement, l'angle θ ne dépend des coordonnées de tous les atomes de la molécule. La somme sur les atomes ne dépend donc que de trois atomes, dans le cas d'une flexion, ou sur

II-5. Les outils développés à partir des modes statiques

quatre atomes, dans le cas d'une torsion. Les seuls éléments intéressants de la matrice M sont donc les 9 ou 12 lignes correspondant aux coordonnées cartésiennes des trois ou quatre atomes intervenant dans l'équation 1.

De la même façon, les forces extérieures ne sont exercées que sur un seul atome et le vecteur $|F\rangle$ ne comporte que trois composantes non nulles. Ceci limite à trois le nombre de colonnes intéressantes dans la matrice M des modes statiques. Nous pouvons alors extraire une sous matrice m de la matrice M des modes statiques ne comportant que trois colonnes et 9 ou 12 lignes, respectivement pour une flexion ou une torsion. Nous écrivons :

$$|d\chi\rangle = m|\varphi\rangle$$

Où $|d\chi\rangle$ et $|d\varphi\rangle$ sont les sous vecteurs représentant respectivement les coordonnées et les forces qui nous intéressent. La variation d'angle $\Delta\theta$ peut donc s'exprimer par :

$$\Delta\theta = \langle v^T m | \varphi \rangle$$

Où $|v\rangle$ est le vecteur de composantes :

$$v_i = \frac{\partial \theta}{\partial \chi_i}$$

Alors :

$$|\Delta\theta|^2 = \langle \varphi | m^T v v^T m | \varphi \rangle$$

Nous sommes ainsi amenés à un problème d'optimisation sous contraintes. La solution est donnée par la diagonalisation de la matrice $A = \langle \varphi | m^T v v^T m | \varphi \rangle$.

Nous remarquerons cependant que la matrice A est une matrice de Van der Monde dont toutes les valeurs propres, sauf une, sont nulles. Cette valeur propre s'écrit donc :

$$\lambda_0 = Tr(A)$$

Le vecteur propre associé à λ_0 nous la donne la direction optimale de la force à appliquer.

La direction optimale de la force, pour laquelle la déformation est maximale, peut ainsi être directement déterminée à partir de la matrice des modes statiques, sans passer obligatoirement par un algorithme itératif. Afin de rester cohérent avec l'approximation harmonique, l'évaluation des angles, qui ne sont pas des fonctions linéaires des déplacements en coordonnées cartésiennes, a été linéarisée.

II-5.1.2. Modes de pincements

Le procédé présenté ici consiste à pincer un couple d'atomes en leur appliquant un torseur nul, afin d'observer les déformations induites sur la molécule. L'intérêt d'un tel procédé réside dans le fait qu'il annule toutes les forces de réaction. Ces forces s'exercent normalement sur les atomes fixes, empêchant la translation et la rotation de la molécule. La situation du torseur nul, en annulant les forces de réaction, rend la réponse tout à fait indépendante du choix des atomes fixes. Nous remarquons de plus, que dans le docking moléculaire, toutes les forces d'interaction s'exercent par couples et sont opposées deux à deux. Par conséquent, le choix des atomes fixes n'a aucun effet sur le procédé de docking.

Pour pincer deux atomes i et j , on considère deux forces \vec{F}_i et \vec{F}_j à appliquer respectivement sur les atomes i et j . Ces forces illustrées sur la figure 15 montrent que $\vec{F}_i = \vec{ij}$ et $\vec{F}_j = -\vec{F}_i$, pour satisfaire la condition du torseur nul.

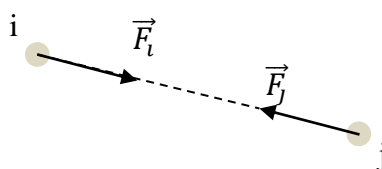


Figure 15 : forces employées lors d'un calcul de modes de pincement

On extrait tout d'abord de la matrice des déformations une sous-matrice m de taille $(3N \times 6)$ contenant les trois modes statiques correspondant à chacun des atomes i et j .

On peut alors appliquer à m les forces \vec{F}_i et \vec{F}_j pour obtenir la déformation engendrée \overrightarrow{def} :

$$\overrightarrow{def} = \begin{pmatrix} m_i & m_j \\ \downarrow & \downarrow \\ \left(\begin{smallmatrix} \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \end{smallmatrix} \right) & \left(\begin{smallmatrix} \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \end{smallmatrix} \right) \end{pmatrix} \begin{pmatrix} F_{ix} \\ F_{iy} \\ F_{iz} \\ F_{jx} \\ F_{jy} \\ F_{jz} \end{pmatrix}$$

Pour conserver une force normée, nous définirons les composantes de la force par :

$$F_{ik} = \frac{\vec{ij} \cdot \vec{e}_k}{|\vec{ij}|}$$

Où les $|\vec{e}_k\rangle$ sont les vecteurs unitaires du repère cartésien.

Une fois la déformation engendrée par le pincement calculée, on peut immédiatement déduire une variation d'angle, de distance, etc.

II-5.1.3. Grandeurs caractéristiques

Pour décrire la flexibilité des macromolécules, nous avons choisi de regarder les variations des angles dièdres. Mais pour une molécule composée de N atomes, nous avons (3N-6) modes. Si on multiplie par le nombre d'angles dièdres, cela fait un grand nombre de données à regarder. Nous avons alors recours à un traitement automatique de ces données pour en extraire les principales caractéristiques. Pour clarifier les résultats, nous avons choisi de regarder deux types de grandeurs pour caractériser la flexibilité de nos structures :

- 1) Des valeurs moyennes de variation d'angles dièdres, calculées comme suit :

$$\langle \Delta\theta \rangle = \frac{1}{N_{dih}} \sum_{i=1}^{N_{dih}} \left(\frac{1}{N_{mod}} \sum_{j=1}^{N_{mod}} (\theta_o^i - \theta_j^i) \right)$$

Où N_{dih} est le nombre d'angles dièdres, N_{mod} le nombre de Modes Statiques, θ_o^i est la valeur de l'angle i de la structure non déformée et θ_j^i la valeur de l'angle i déformé par le mode j .

- 2) Des RMSD (pour Root Mean Square Deviation) calculés sur les angles dièdres afin de comparer les structures déformées avec une structure de référence (noté RMSDT):

$$RMSDT = \left[\frac{1}{N} \sum_{i=1}^N (\theta_i^{def} - \theta_i^{ref})^2 \right]^{\frac{1}{2}}$$

Où N est le nombre de dièdres, θ_i^{def} est la valeur de l'angle dièdre θ de la structure déformée et θ_i^{ref} est la valeur de l'angle dièdre θ de la structure de référence.

- 3) Des RMSD calculés sur les positions atomiques, calculés par la formule (noté RMSD) :

$$RMSD = \left[\frac{1}{N} \sum_{i=1}^N (\vec{r}_i^{def} - \vec{r}_i^{ref})^2 \right]^{\frac{1}{2}}$$

Où les \vec{r} représentent les vecteurs positions des atomes de la molécule.

Une telle formule, appliquée a priori, ne permet pas d'obtenir une valeur optimale, c'est-à-dire minimale, du RMSD. En effet, les positions atomiques définies dans cette formule dépendent des repères de référence, qui peuvent être différents pour les deux molécules. Il faut donc optimiser ces repères avant d'appliquer la formule. Parmi les différentes techniques qui existent pour réaliser cette transformation, nous avons choisi la méthode proposée par W. Kabsch (130), détaillée en annexe.

II-5.2. Arrimage d'un ion métallique

Comme premier exemple de docking, nous avons développé un outil qui calcule les déformations d'une molécule lorsque cette dernière interagit avec un ion métallique. Un algorithme a donc été écrit afin d'accrocher l'ion sur un site réactionnel du peptide, puis d'utiliser la force induite par la présence de l'ion pour déformer la molécule.

Dans un premier temps, on définit un potentiel qui va régir les interactions entre la molécule et le cation. Nous nous plaçons dans le cas simple des interactions électrostatiques de Coulomb :

$$U = \frac{1}{4\pi\epsilon_0} \sum_i \frac{q_i q_{Zn}}{r_{i-Zn}}$$

Où q_i est la charge de l'atome i , q_{Zn} la charge de l'ion Zn^{2+} , r_{i-Zn} la distance entre l'atome i et le cation et ϵ_0 la permittivité du vide.

On translate ensuite l'ion en minimisant les forces d'interaction avec tous les atomes. Lorsque l'ion passe proximité d'un site chargé négativement, il se trouve piégé et s'accroche à cet atome de la molécule.

A proximité du site d'interaction, nous rajoutons un terme de Van der Waals entre l'ion et l'atome sur lequel il s'accroche. Nous avons donc le potentiel :

$$U = \epsilon_{SK} \left[\left(\frac{\sigma_{SK}}{r_0} \right)^{12} - 2 \left(\frac{\sigma_{SK}}{r_0} \right)^6 \right] + \frac{1}{4\pi\epsilon_0} \sum_i \frac{q_i q_{Zn}}{r_{i-Zn}}$$

Où r_0 est la distance entre le cation et l'atome impliquée, ϵ_{SK} et σ_{SK} sont les paramètres déterminés par Stote et Karplus (131), soit :

II-5. Les outils développés à partir des modes statiques

$$\begin{cases} \sigma_{SK} = 2.83 \text{ \AA} \\ \varepsilon_{SK} = 0.0125 \text{ eV} \end{cases}$$

Or, les interactions répulsives qui ont lieu à faible distance ne concernent pas uniquement l'atome sur lequel s'accroche l'ion. Nous devons les considérer également pour les autres atomes de la molécule. Les interactions de Van der Waals représentent une approximation de ces forces de répulsion. On observe néanmoins, qu'à partir d'une certaine distance, les forces de répulsion augmentent très rapidement sur de très faibles variations de distance. Dans notre modèle de flexibilité, où les déformations de la molécule sont très sensibles aux forces extérieures, cette approximation induit des déformations irréalistes de la molécule. C'est pour cette raison que nous avons préféré la remplacer par une approximation de sphère dure où la notion de force est remplacée par la distance. L'ion reste toujours au-delà d'une certaine distance seuil des atomes de la molécule. Cette opération est effectuée dans les dernières étapes de la translation. Du point de vue énergétique, cela revient à considérer une barrière de potentiel infinie à la limite de la sphère dure.

L'étape suivante consiste à déformer la molécule en utilisant les Modes Statiques. Pour cela, on utilise le même champ de force que précédemment pour calculer le vecteur force décrivant l'interaction entre l'ion métallique et les atomes de la molécule. On se sert ensuite de cette force pour déformer la molécule, via les Modes Statiques.

La force d'interaction entre l'ion métallique et un atome i est obtenue par la relation :

$$\vec{F}_i = -\vec{\nabla}_{\vec{r}_i} U$$

Où U est le potentiel défini précédemment, $\vec{r}_i = (x_i, y_i, z_i)$ est le vecteur entre l'ion et l'atome i , et $\vec{\nabla}_{\vec{r}_i} = \left(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial y_i}, \frac{\partial}{\partial z_i} \right)$.

On extrait ensuite de la matrice M des Modes Statiques une sous-matrice m_i de dimension $(3N \times 3)$ composée des 3 colonnes correspondant aux modes statiques associés à l'atome i en interaction :

$$M = \begin{bmatrix} \dots & \dots & \overbrace{\begin{pmatrix} \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \end{pmatrix}}^{m_i} & \dots & \dots \\ \dots & \dots & \vdots & \dots & \dots \\ \dots & \dots & \vdots & \dots & \dots \end{bmatrix}$$

On multiplie ensuite m_i par la force F_i pour obtenir le vecteur Δx contenant les déformations résultant de l'interaction entre l'ion et l'atome i sur toute la molécule : $\Delta x = \vec{F}_i \cdot m_i$

Tous les atomes en interaction avec l'ion vont générer un champ de déformation Δx . Or, l'action de la force électrostatique a lieu en même temps sur ces atomes. Nous utilisons donc tous les Δx pour déterminer la conformation après le passage de Zn^{2+} . Une fois la molécule déformée, on effectue une optimisation de la géométrie par minimisation de l'énergie à l'aide du logiciel AMBER 8 (132). Si l'on désire simuler un nouveau passage de l'ion sur cette nouvelle conformation, il faut alors recalculer la matrice des modes statiques.

Comme nous allons le voir dans le chapitre suivant, la sonde électrostatique a permis de démontrer la puissance des modes statiques dans la modélisation du docking moléculaire. En effet, cet outil valide les deux approches que nous avons imaginées :

- l'algorithme développé pour accrocher l'ion au site réactionnel par rotations/translations successives peut être utilisé pour rapprocher deux molécules lors de la simulation du docking. Ceci constitue la première étape d'un programme de docking « continu » tel que nous l'avons imaginé.
- Nous pouvons déterminer un grand nombre de conformations « déformées » à partir d'une structure de base, générées par une interaction spécifique. Ainsi, on peut se servir de cet outil pour explorer l'espace conformationnel autour d'une position d'équilibre, et utiliser les modes statiques dans une procédure de docking de la même manière que les modes normaux.

II-6. Conclusion

Le processus de docking est un sujet essentiel pour progresser dans la compréhension des mécanismes d'interactions moléculaires et pour le développement d'outils prédictifs important dans les domaines de la nano-bio-ingénierie et de la médecine.

Les méthodes existantes sont principalement tournées vers des techniques de « scoring » et construites sur la génération de conformations multiples pour les molécules en présence. Une des voies de construction de conformations multiples est basée sur l'exploration des données des modes normaux de vibration et nous avons explicités les méthodes les plus récentes consacrées à cette option.

Nous avons présenté une approche originale de docking où les molécules sont approchées pas à pas et les déformations calculées à chaque pas de translation. Cet objectif nécessite la

disponibilité d'un outil de calcul des déformations : les Modes Statiques répondent à cet objectif. Nous avons également précisé dans ce chapitre la possibilité d'utiliser cet outil afin de générer des conformations, dans le but d'incorporer cette méthode dans une procédure de docking à la manière des modes normaux.

Nous avons donné dans ce chapitre la façon d'obtenir les Modes Statiques par le calcul à partir des modèles énergétiques des molécules, ainsi que les résultats validant cette approche pour décrire des mouvements atomiques divers.

Sur cette base, nous proposons plus précisément un outillage pour :

- Optimiser la direction d'une force dans le but de maximiser une distance ou un angle de torsion.
- Déterminer les effets sur la structure d'une molécule lors de l'arrimage d'un ion métallique sur des sites réactionnels.
- Générer des conformations multiples à partir d'une conformation d'équilibre en exerçant des forces pertinentes sur des sites réactifs de la molécule.

Ces outils seront validés par les résultats que nous allons présenter dans le chapitre 3.

CHAPITRE 3 : RESULTATS

SOMMAIRE

CHAPITRE 3 : RESULTATS	- 77 -
SOMMAIRE	- 77 -
Introduction	- 78 -
III-1. Présentation des peptides	- 78 -
III-1.1. Le peptide Amyloïde- β	- 79 -
III-1.2. Le peptide modèle GNNQQNY	- 83 -
III-2. Caractérisation de la flexibilité	- 85 -
III-2.1. Cas du peptide modèle GNNQQNY	- 85 -
I-2.2. Cas de l'amyloïde β	- 95 -
III-2.3. Représentation des zones flexibles des molécules	- 97 -
III-2.3.1. Propagation de la déformation	- 101 -
III-2.3.2. Comportement allostérique	- 105 -
III-2.3.3. Relation interactions/oligomérisation	- 107 -
III-2.5. Conclusion	- 112 -
III-3. Application au docking	- 113 -
III-3.1. Description des conformations utilisées	- 113 -
III-3.2 Sonde électrostatique	- 116 -
III-3.2.1. Exploration de l'espace conformationnel de la structure A β (1-16) en interaction avec un ion métallique	- 116 -
III-3.2.2. Repliement du peptide à partir de la structure sans zinc	- 122 -
III-3. Conclusions	127

Introduction

Nous avons introduit au chapitre précédent les Modes Statiques comme un moyen d'étude de la flexibilité totale des macromolécules dans une procédure générale de docking. Dans ce même chapitre, nous avons également vu la manière dont les Modes Statiques peuvent être utilisés dans le but de décrire les mouvements atomiques des biomolécules. Nous appliquons ici ces outils à l'étude d'un peptide modèle de sept acides aminés (séquence GNNQQNY) dans un premier temps, puis du peptide amyloïde- β dans un second temps.

Ce troisième chapitre se penche sur deux approches de l'utilisation des Modes Statiques. Tout d'abord, il sera question de caractériser et cartographier la flexibilité de différentes conformations de ces peptides. Nous tenterons de confronter et valider certains résultats sur la base de données de la littérature, provenant à la fois de calculs théoriques et de l'expérience. Dans un second temps, nous envisagerons une utilisation des modes statiques au sein de laquelle des interactions successives viendront modifier la nature conformationnelle du peptide.

III-1. Présentation des peptides

La majeure partie des résultats que nous allons découvrir dans ce chapitre, s'appuie sur l'étude d'un peptide en particulier : l'Amyloïde- β , protéine impliquée dans le développement de la maladie d'Alzheimer. Cette molécule hautement flexible, doit sa toxicité à un changement conformationnel qui a lieu pour des raisons encore largement controversées. La cartographie de la flexibilité de ce peptide via les modes statiques devra nous fournir quelques éclairages à cet égard.

Nous avons également utilisé un peptide de taille plus modeste dont la séquence est GNNQQNY, et qui possède également la particularité de s'agréger (133). La petite taille de cette molécule (107 atomes) permet d'effectuer des calculs rapides et nous a servi afin d'explorer les capacités du programme FLEXIBLE. Les séquences de chacun de ces peptides sont présentées dans le tableau 3.

Nom du peptide	Séquence
Peptide modèle	Gly-Asn-Asn-Gln ⁵ -Gln-Tyr
Fragment 1-16 de l'amyloïde β (A β (1-16))	Asp-Ala-Glu-Phe-Arg ⁵ -His-Asp-Ser-Gly-Tyr ¹⁰ - Glu-Val-His-His-Gln ¹⁵ -Lys
Fragment 1-28 de l'amyloïde β (A β (1-28))	Asp-Ala-Glu-Phe-Arg ⁵ -His-Asp-Ser-Gly-Tyr ¹⁰ - Glu-Val-His-His-Gln ¹⁵ -Lys-Leu-Val-Phe-Phe ²⁰ - Ala-Glu-Asp-Val-Gly ²⁵ -Ser-Asn-Lys

Tableau 3 : séquences des trois peptides que nous avons utilisés pour effectuer nos calculs.

Dans ce chapitre, toutes les images représentant des structures moléculaires ont été réalisées à l'aide du logiciel PyMOL (134).

III-1-1. Le peptide Amyloïde- β

La maladie d'Alzheimer est une maladie neuro-dégénérative caractérisée par le dépôt de plaques séniles dans le cerveau (135) (136). Ces plaques sont composées majoritairement de polypeptides de 40 à 42 acides aminés : l'Amyloïde- β (A β), dérivée par traitement protéolytique de la protéine APP (Amyloid Precursor Protein) (137). La théorie privilégiée par la communauté pour décrire le processus entraînant la maladie est l'hypothèse de la cascade Amyloïde : une fois produit, les peptides A β diffusent à l'extérieur de la membrane cellulaire et s'agrègent pour former des fibres, riches en feuillets β , observées dans le cerveau des personnes atteintes de la maladie (138) (139). Des études récentes montrent que les agents neurotoxiques sont de petits agrégats (également appelés oligomères) solubles, et non les fibres détectés dans le cerveau (140) (141) (142). Ce mécanisme, illustré sur la figure 16, est donc associé à un changement conformationnel du peptide de sa forme native vers une conformation riche en feuillet β .

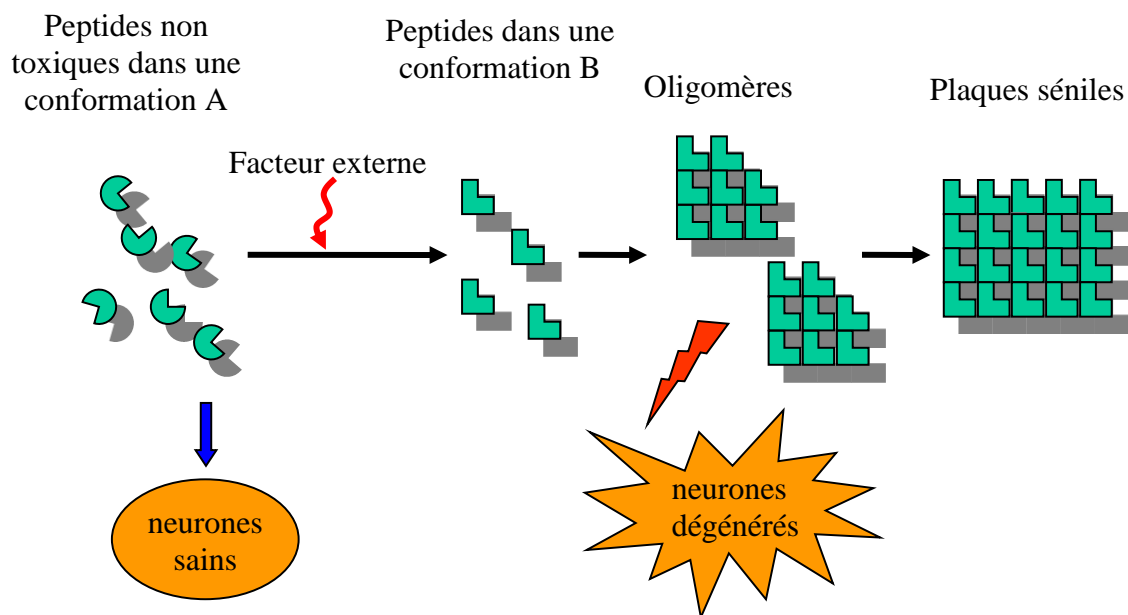


Figure 16 : hypothèse de la cascade amyloïde (source : ref. (143)).

Si l'on veut agir sur le développement de la maladie, il est primordial de comprendre pourquoi le peptide adopte la conformation B plutôt que la conformation A (figure 16). La stratégie à adopter est donc de trouver les facteurs favorisant ce changement dans la structure de la molécule. De plus, ce phénomène d'agrégation est à l'origine d'autres maladies neuro-dégénératives, telles la maladie de Creutzfeld-Jacob, ou encore le diabète de type II (144). Ces peptides présentant cette propriété d'agrégation ont été et sont toujours très étudiés, aussi bien par la voie de l'expérience que par celle de la simulation.

Des travaux menés sur ces molécules qui ont tendance à s'agréger montrent que toutes possèdent une structure secondaire différente dans leur état natif, mais possèdent une forte proportion de résidus dans une conformation en hélice α dans cet état natif, ou dans un état intermédiaire du mécanisme d'agrégation (145) (146) (147). L'étude d'un passage d'une structure en hélice à une structure en feuillet β est donc à mettre au centre des travaux à mener.

Dans le cas de l'amyloïde β par exemple, la partie N-terminale présente une proportion importante de résidus en hélice α . Takeda et Klimov retrouvent ce résultat par des calculs de dynamique moléculaire publiés récemment (148). Ils ont ensuite étudié la transition hélice α \rightarrow feuillet β et arrivent à la conclusion que les interactions entre peptides jouent un rôle

prépondérant. Les résidus impliqués sont notamment ceux situés dans la partie 17-21, dénommée la partie centrale hydrophobe (ou CHC pour Central Hydrophobic Cluster), car elle se compose de résidus hydrophobes. Le CHC est connu pour jouer un rôle important dans l'oligomérisation. Plus généralement, dans le mécanisme d'agrégation, les interactions hydrophobes semblent initier le processus (149) (150). En effet, l'exposition des résidus hydrophobes au milieu extérieur conduit souvent à la dimérisation pour minimiser la surface accessible par le solvant (151) (152) (153). Les résidus impliqués sont l'alanine, la valine ou encore la phénylalanine, cette dernière étant un des peptides les plus amyloïdogéniques (154).

Outre les effets induits par les résidus hydrophobes, la charge des acides aminés est une caractéristique importante lorsque l'on s'intéresse à l'agrégation. La partie N-terminale du peptide amyloïde β contient un grand nombre de résidus chargés susceptibles de réagir avec le milieu via une variation de pH ou l'interaction avec un cation. Dans le cas de l'amyloïde β , cela concerne notamment l'acide aspartique et l'histidine fortement représentés dans la partie N-terminale (v. tableau 3).

Des tests réalisés *in vitro* ont mis en évidence l'influence du pH sur l'oligomérisation, comme le démontrent les travaux réalisés par Kirkitadze *et al.* il y a près de dix ans (147). Plus récemment, Perálvarez-Marín *et al.* ont montré par des mesures de spectroscopie infrarouge sur la partie 1-28 du peptide amyloïde β , qu'en effectuant un saut de pH (de 8.0 à moins de 6.0), on observe une accélération de l'agrégation (155). Ces résultats démontrent bien l'influence du milieu, via des interactions avec les résidus chargés, sur l'oligomérisation.

Outre les effets dus aux variations de pH, les résidus polaires portant une charge négative sur leur chaîne latérale, tels que l'acide aspartique, l'acide glutamique ou l'histidine par exemple, sont susceptibles de réagir avec des cations présents dans l'environnement de la molécule. Ces interactions chaîne latérale/cation entraînent un changement de conformation du peptide. Une des hypothèses avancée depuis quelques années pour expliquer le phénomène d'oligomérisation, est la présence de cations métalliques dans l'environnement de la molécule qui viennent s'accrocher sur des sites bien particuliers et modifient ainsi sa structure (156) (157) (158). Depuis lors, de nombreux travaux basés sur cette hypothèse ont été effectués et un site d'arrimage du cation a été identifié. Il se situe dans la partie N-terminale du peptide et est constitué des histidines 6, 13 et 14. Un quatrième acide aminé est impliqué dans l'arrimage, mais il y a encore débat sur son identité. Plusieurs possibilités ont été proposées, tels que Asp1 (159), Arg5 (160), Tyr10 (161) ou plus récemment Glu11 (162). Les quatre résidus impliqués dans l'arrimage du zinc sont situés dans la partie 1-16 de la chaîne. Cette

CHAPITRE III : RESULTATS

portion est donc la structure « minimale » pour l'étude de l'interaction d'un ion métallique avec l'amyloïde β .

La partie N-terminale (résidus 1 à 28) possède donc un nombre important de résidus susceptibles de réagir avec le milieu via le pH ou des interactions avec des ions métalliques. La partie C-terminale (résidus 29 à 40), quant à elle, se compose principalement de résidus hydrophobes. Nous venons de voir l'importance des interactions hydrophobes dans le phénomène d'oligomérisation. La conformation de cette portion de la molécule peut donc être modifiée par ces interactions. Néanmoins, diverses études s'accordent à dire que c'est la partie N-terminale du peptide qui est à l'origine de l'agrégation. La partie C-terminale intervient tout de même dans des interactions inter- et intra-peptides qui ont un rôle primordial dans l'oligomérisation.

Le peptide amyloïde β possède donc deux parties qui se distinguent par leur caractère. La portion N-terminale (résidus 1 à 28), hydrophile et composée d'une proportion importante de résidus polaires, est susceptible de réagir avec le milieu via le pH ou encore l'interaction avec des cations métalliques. Les résidus Leu17-Ala21 sont hydrophobes et constituent le CHC qui joue un rôle primordial dans l'oligomérisation. Par ailleurs, la substitution de certains résidus situés dans cette partie N-terminale influe sur la vitesse d'agrégation. C'est ce qu'on observe lors des mutations naturelles E22K (163), E22Q (163) (164) ou E22G (165).

Divers modèles du peptide dans sa forme agrégée ont été proposés, mais tous présentent une forme en feuillet β plissé, c'est à dire un enchaînement de feuillets β et de coudes (166). En effet, à plusieurs endroits de la chaîne ont été identifiés des coudes entre les feuillets : la région aux alentours de Gly9, ou encore la partie Val24-Lys28 faisant le lien entre les parties hydrophiles et hydrophobes. En plus des résultats présentés dans la référence 143, ces deux zones ont été identifiées par divers travaux expérimentaux et de simulation (167) (168) (169) (170) (171).

Nous allons nous intéresser dans ce chapitre au changement conformationnel $\alpha \rightarrow \beta$. Nous avons choisi d'effectuer nos calculs sur le fragment (1-28), pour plusieurs raisons évoquées précédemment : i) ce peptide adopte une forme agrégée similaire au peptide complet (172) (173), ii) il contient un nombre important de résidus chargés, qui sont connus pour jouer un rôle dans l'activation ou l'inhibition de l'agrégation (174) (156) (155), ainsi que la partie centrale hydrophobe, qui tient un rôle primordial dans l'oligomérisation, iii) dans un environnement similaire à celui de la membrane, il adopte une structure en hélice α

(175).

Dans la seconde partie des travaux menés sur l'Amyloïde β , nous nous intéresserons à l'interaction du peptide avec un cation métallique. Comme nous l'avons vu, le site de fixation du cation est connu et est situé dans la partie N-terminale du peptide. Nous utiliserons donc le modèle conformationnel établi par Zirah *et al.* (162) qui ont observé que le cation s'arrime sur quatre chaînes latérales : His6, Glu11, His13 et His14.

Auparavant, nous aurons étudié brièvement le cas d'un peptide modèle plus petit (séquence GNNQQNY) permettant de développer rapidement nos outils de caractérisation du logiciel FLEXIBLE.

III-1.2. Le peptide modèle GNNQQNY

La première structure précise d'un agrégat de type amyloïde a été obtenue par Nelson *et al.* (176). Ils ont pu observer ces formes fibrillaires caractéristiques des maladies neuro-dégénératives par diffraction des rayons X d'un peptide, sous formes de microcristaux d'un peptide composé de 7 acides aminés dont la séquence est GNNQQNY. Ce dernier est la partie N-terminale du prion de levure Sup 35. Les microcristaux observés ont une structure similaire aux fibres amyloïdes que l'on retrouve dans de nombreuses maladies neuro-dégénératives. C'est le cas de l'Amyloïde β dans le cas de la maladie d'Alzheimer, qui s'agrège pour former des fibres riches en feuillet β (177) (178) (179).

La structure précise des agrégats de type amyloïde a donc pu être déterminée à partir de ces observations. Cette dernière est détaillée dans la référence (176), mais on peut voir sur la figure 17 un feuillet composé de deux peptides. Cette structure a été construite à partir des coordonnées du fichier PDB (code 1YJP) et des informations relatives à la symétrie du cristal. La géométrie de ce peptide a été optimisée en utilisant le champ de force ff03 de AMBER (180).

La description précise de la structure en feuillet et la petite taille du système en font un objet intéressant pour notre étude.

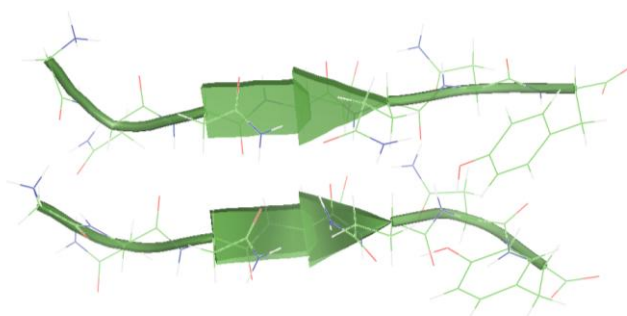


Figure 17 : feuillet β composé de deux peptides GNNQQNY.

Des études récentes ont permis d'établir les premières étapes de l'agrégation du peptide modèle GNNQQNY. Il a été établi expérimentalement que l'arrangement du peptide modèle en microcristaux est précédé par la formation d'un noyau composé de quelques peptides (176). Zheng *et al.* ont ensuite montré, par des simulations de dynamique moléculaire tout atome, que ce noyau est formé de trois ou quatre monomères en feuillet (181), ce qui est en accord avec l'expérience (176).

Dans un premier temps, des dimères se forment dans une configuration en feuillet β antiparallèle, puis les peptides s'agrègent à ce dimère en formant des feuillets parallèles, comme le montre de façon schématisée la figure 18. Le premier complexe antiparallèle se trouve être plus stable que le même complexe dans une configuration parallèle. Ceci constitue la première étape du mécanisme d'agrégation, les autres peptides s'arrimant alors au dimère de manière parallèle pour former les fibres (182).

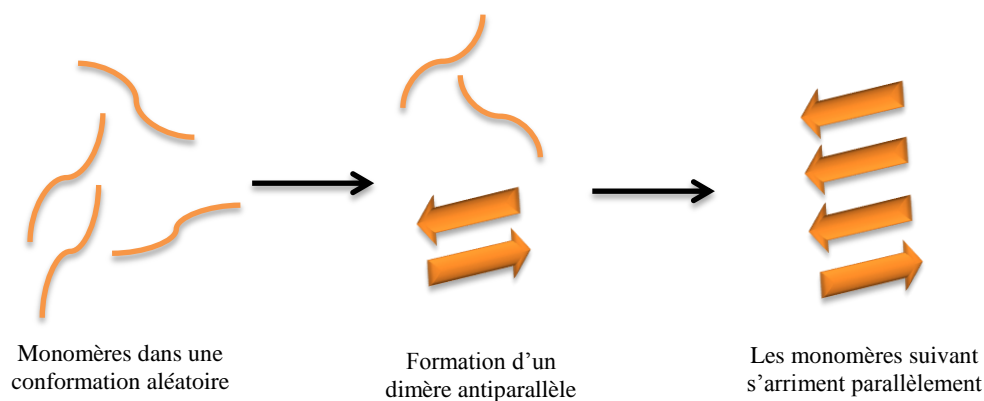


Figure 18 : Premières étapes de l'agrégation du peptide modèle GNNQQNY.

CHAPITRE 3 : RESULTATS

Les interactions qui maintiennent le feuillet stable sont les liaisons hydrogène entre les peptides. La formation de ces liaisons tient donc un rôle primordial dans les premières étapes de l'oligomérisation.

III-2. Caractérisation de la flexibilité

En prenant comme exemples le peptide modèle GNNQQNY d'une part, puis le peptide Amyloïde- β , nous avons utilisé les Modes Statiques pour caractériser les déformations de différentes structures d'un même peptide. Nous avons voulu regarder la flexibilité générale de la molécule en exploitant directement les déformations données par la matrice des modes statiques. C'est pourquoi nous avons utilisé le même poids pour tous les modes.

Nous nous sommes penchés sur des structures bien connues des biologistes : les conformations en hélice α et en feuillet β . En outre, nous avons utilisé des géométries issues de la dynamique moléculaire dans le cas du peptide modèle.

III-2.1. Cas du peptide modèle GNNQQNY

III-2.1.1. Calculs sur les monomères

Quatre conformations différentes du monomère, que l'on peut voir sur la figure 19, ont été utilisées :

- Une hélice α (A)
- Une forme étendue (B) proche de la structure du monomère dans le feuillet β
- Deux structures repliées aléatoires obtenues par la dynamique moléculaire (R1 et R2)

Nous avons également utilisé une conformation en feuillet β parallèle composée de deux brins, illustrée sur la figure 17.

Ces structures ont été optimisées par minimisation de l'énergie en utilisant le logiciel AMBER8, avec le champ de force ff03, particulièrement bien adapté pour le calcul sur des protéines (183). La minimisation de l'énergie se décompose en deux étapes : nous avons effectué 5000 pas de descente de gradient, pour optimiser « grossièrement » la molécule,

CHAPITRE III : RESULTATS

suivis de 5000 nouveaux pas de minimisation par la technique des gradients conjugués, afin d'obtenir une structure plus fine de ces peptides.

Une fois les structures optimisées, nous effectuons une nouvelle étape par l'intermédiaire de ce logiciel : l'extraction de la matrice hessienne. Le module Nmode permet de calculer les modes normaux de vibration des molécules, et cela passe évidemment par la détermination de la matrice hessienne. Le code de la version 8 d'AMBER étant accessible et modifiable, nous avons agi sur le module Nmode, afin d'éditer la matrice hessienne avant sa diagonalisation. Cette dernière est la base, avec les coordonnées atomiques de la géométrie optimisée, du calcul des Modes Statiques.

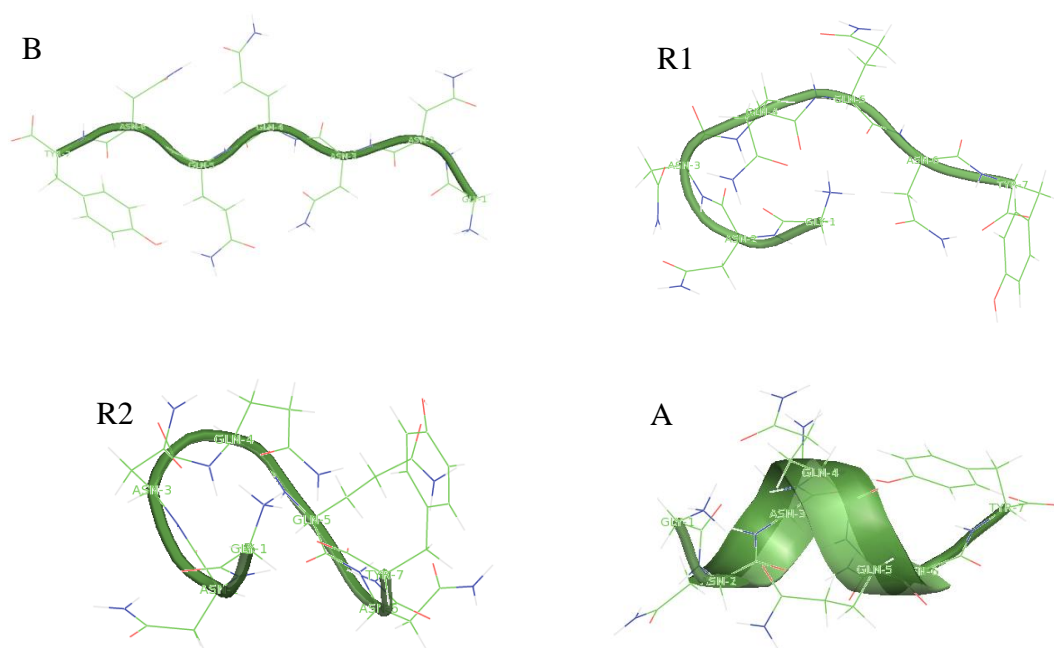


Figure 19 : Les 4 structures utilisées dans les calculs sur le peptide modèle GNNQQNY.

Dans cette partie, nous voulons savoir s'il existe un ou plusieurs mode(s) qui déforme(nt) le monomère pour le rapprocher de la conformation d'un des brins du dimère. Pour cela, nous allons exciter tous les modes de pincement, afin d'éviter les effets dus au choix des atomes fixes. Chaque mode possède le même poids, et pour chacun d'eux, nous avons calculé le RMSD sur les positions atomiques du squelette (les carbone α et les atomes formant les liaisons peptidiques) entre le monomère déformé et un brin du feuillet, qui sera la référence conformationnelle de ces calculs. Le but étant de déterminer les modes déformant la molécule vers sa structure dans le feuillet, donc à une diminution du RMSD lorsque l'on

excite un mode, nous avons représenté sur différents diagrammes les $\Delta RMSD$ donnés par la formule :

$$\Delta RMSD = RMSD_0 - RMSD_i$$

Où $RMSD_0$ est obtenue avec la structure non déformée (que l'on peut voir sur la figure 19) et $RMSD_i$ est obtenue avec la molécule déformée par le mode i . On ne retiendra alors que les valeurs positives du $\Delta RMSD$.

Nous comparons alors chacune des conformations (A, B, R1 et R2) avec un brin du monomère. Nous obtenons alors 4 diagrammes que nous avons représentés sur la figure 20.

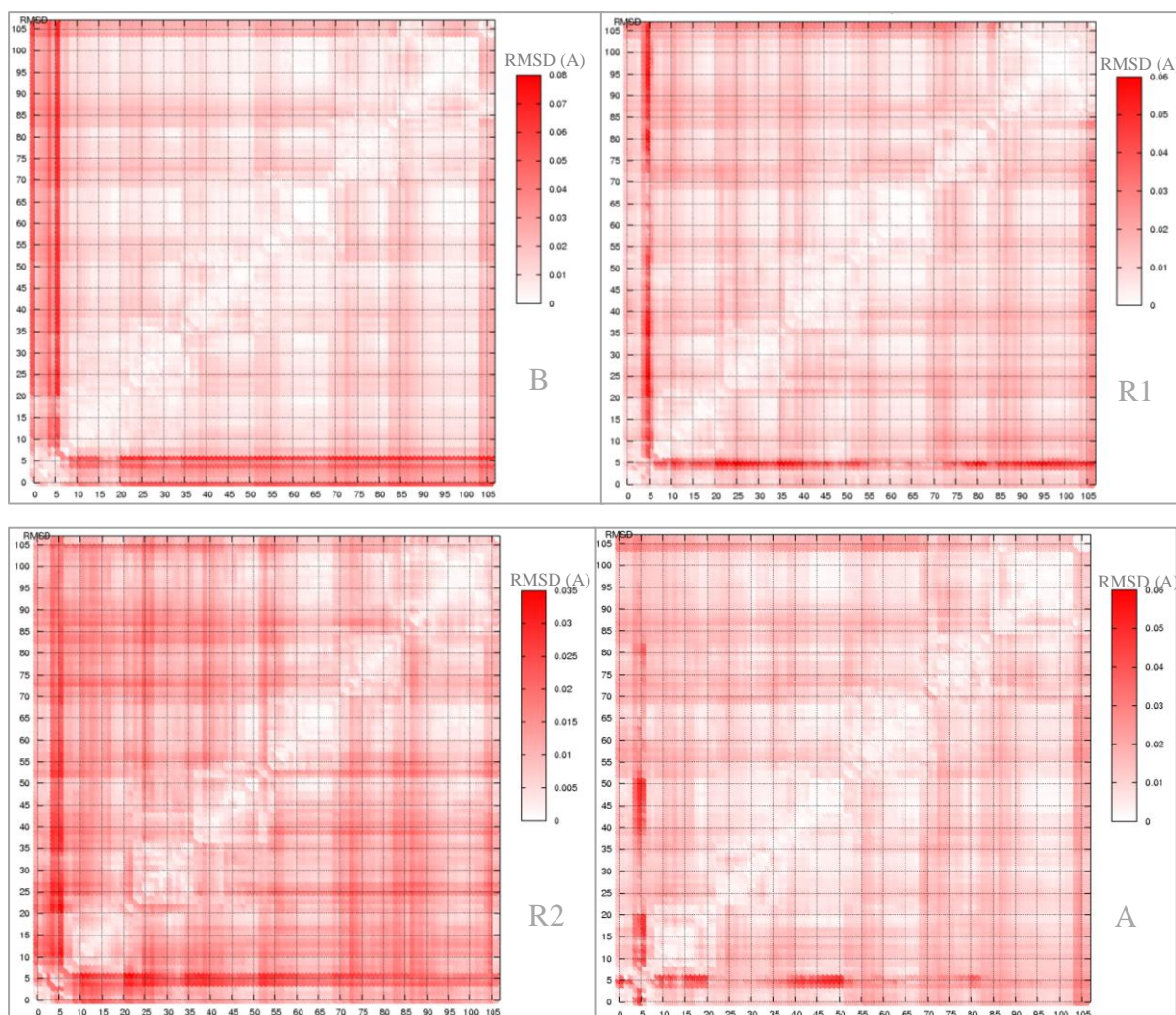


Figure 20 : Diagrammes représentant le $\Delta RMSD$ sur les atomes du squelette entre les monomères B, R1, R2 et A et un brin du dimère.

Nous voyons apparaître des bandes plus foncées qui indiquent des zones privilégiées à exciter pour « déplier » le monomère et le rapprocher d'une conformation β , correspondant à sa

CHAPITRE III : RESULTATS

forme dans le feuillet. Ces variations de RMSD, bien que faibles, nous renseignent sur les zones sensibles de la molécule. Nous constatons tout d'abord qu'une même bande, bien prononcée, se distingue sur chacun des diagrammes impliquant les atomes situés en début de chaîne. Celle-ci indique que la Glycine, située en première position de la chaîne (v. tableau 3), est un élément clef de la flexibilité de la molécule, relative à un déplacement global des atomes du squelette vers une structure de type feuillet β . En effet, sur les trois premiers diagrammes, correspondant respectivement aux conformations B, R1 et R2, les plus fortes variations de RMSD correspondent aux pincements entre les atomes de la glycine et tous les autres atomes du peptide. Ceci donne lieu à l'apparition d'une bande plus foncée dans chacun de ces diagrammes. En revanche, pour la conformation en hélice, les plus fortes variations de RMSD ont lieu pour des pincements entre des atomes de la glycine et quelques atomes seulement. Nous pouvons donc tirer une première conclusion de ces diagrammes qui est de dire que la glycine joue un rôle dans l'oligomérisation du peptide. Ceci est en accord avec les observations de Nelson *et al.* qui ont constaté que lorsqu'on enlève la glycine de la séquence, le peptide NNQQNY ne s'agrège pas⁴ (176).

Toutefois, la glycine n'a pas le même rôle dans les différentes conformations. On voit en effet que sur les diagrammes les bandes présentent des « trous ». Par exemple, dans la conformation A en hélice, on observe ces valeurs maximales du $\Delta RMSD$ pour des pincements entre Gly1 et Asn2 ou Gln4. Or, si l'on regarde de plus près les différentes conformations (figure 19), on voit que la glycine est en interaction directe avec ces deux résidus. Les effets des autres (Asn3, Gln5, Asn6, Tyr7) étant alors écrantés par la présence de Asn2 et Gln4 devant Gly1.

Pour les conformations B et R2, on n'observe pas cet effet d'écran, car la glycine se retrouve en interaction avec tous les autres résidus de la chaîne. En revanche, dans le cas de la conformation R1, on voit apparaître un trou dans cette même bande pour une région comprise entre les atomes 55 et 72. Ces derniers sont les atomes du résidu Gln5, dont les effets ne semblent pas écrantés si l'on se réfère à la figure 19. Cela signifie donc qu'il y a interaction entre ces deux résidus, mais elle ne favorise pas le dépliement de la conformation R1 vers sa structure dans le feuillet.

En plus de cette bande correspondant aux atomes de Gly1, on voit apparaître sur ces diagrammes des bandes plus atténuées. Ce sont les mêmes pour les quatre conformations et

⁴ Il ne s'agrège pas tout seul mais en présence d'ions Zn^{2+} (v.ref. 167).

III-2. Caractérisation de la flexibilité

elles correspondent à des atomes du squelette du peptide. Or, les RMSD sont calculés sur les positions atomiques des atomes du squelette. Il est donc logique d'observer une variation plus importante du RMSD lorsque l'on pince ces atomes en particulier.

Jusqu'à présent, nous avons regardé le mouvement des atomes du squelette du peptide. Nous allons maintenant nous intéresser aux atomes formant des liaisons hydrogène.

Les liaisons hydrogène étant une interaction primordiale pour la formation et la stabilité du feuillet, nous avons calculé des RMSD sur les positions des atomes formant des liaisons H, inter- et intra-peptide, dans le feuillet. Nous obtenons de la même manière que précédemment les quatre diagrammes de la figure 21.

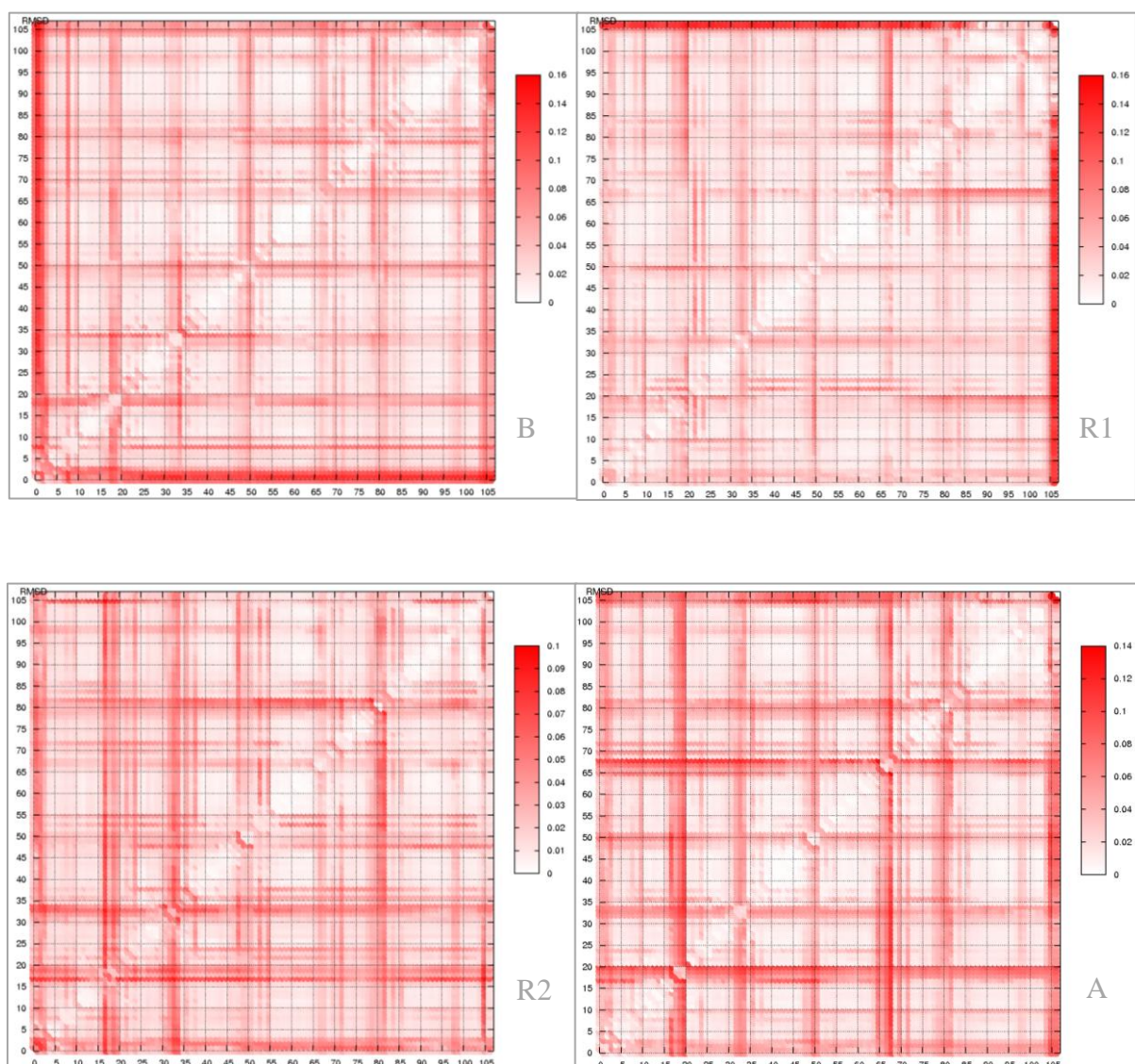


Figure 21 : diagrammes représentant le Δ RMSD calculé sur les atomes formant des liaisons H entre les monomères β , R1, R2 et α (respectivement A, B, C et D) et le brin 1 du feuillet.

CHAPITRE III : RESULTATS

La première observation que nous pouvons faire, est que nous voyons apparaître les mêmes bandes sur les quatre diagrammes. Toutefois, certaines sont plus marquées pour une conformation que pour une autre. Par exemple, le diagramme correspondant à la conformation A montre que les modes impliquant les atomes du NH_3^+ de la partie N-terminale du peptide donnent les valeurs de ΔRMSD les plus élevées. On retrouve bien cette bande sur les autres diagrammes, mais les ΔRMSD correspondant sont plus faibles.

On recense sept bandes sur ces diagrammes que nous noterons a, b, c, d, e, f et g. Elles correspondent à des modes impliquant des atomes que nous avons utilisés pour calculer le RMSD, à l'exception de la bande g. C'est-à-dire que ce sont des atomes qui forment des liaisons hydrogène dans le feuillet. Les sites correspondant à ces bandes sont respectivement : la fonction amine (chargée) du N-terminal, les extrémités des chaînes latérales des résidus Asn2, Asn3, Gln4 et Gln5, ainsi que le groupe carboxyle (ionisée) de la partie C-terminale. Ces bandes sont détaillées dans le tableau 4.

Bande	Atomes	Site
a	0-3	NH_3^+ , N-term
b	16-20	CL, Asn2
c	31-34	CL, Asn3
d	47-51	CL, Gln4
e	65-68	CL, Gln5
f	78-82	CL, Asn6
g	104-106	COO^- , C-term

Tableau 4 : Détail des bandes des diagrammes de la figure 22 : nom, numéros des atomes correspondants et leur situation sur le peptide. CL signifie Chaîne Latérale.

Les plus grandes variations de RMSD sont donc observées pour des pincements impliquant des atomes utilisés dans le calcul du RMSD. Or, pour ce calcul, nous avons également pris des atomes du squelette (les NH et O), et nous ne voyons pas apparaître de bandes au niveau de ces atomes sur les diagrammes de la figure 21. Cela signifie donc que des interactions entre les atomes chargés situés aux extrémités des chaînes latérales favorisent un dépliement du peptide.

Regardons maintenant plus en détail ces bandes, pas toujours uniformes, et qui vont nous révéler les interactions pertinentes, c'est-à-dire qui favorisent le dépliement de la molécule.

Dans la conformation B, la bande la plus prononcée est la a. Cela indique que la terminaison N, chargée, est très sensible aux interactions avec les autres atomes de la chaîne. On voit également une zone plus foncée sur la bande c : les pincements impliquant l'atome 34 (hydrogène situé en bout de la chaîne latérale de Asn3) avec les atomes 10 à 50 (constituant les résidus Asn2, Asn3 et Gln4) donnent des $\Delta RMSD$ plus élevés. Dans sa conformation B, les interactions pertinentes impliquent donc la terminaison NH_3^+ avec tous les atomes de la chaîne, ainsi que l'atome 34 avec les résidus situés à proximité de ce dernier (Asn2, Asn3, Gln4).

Dans la conformation R1, la bande g est la plus prononcée et est uniforme. Ceci indique que les interactions entre la terminaison C et le reste des atomes de la molécule favorisent le dépliement du peptide vers sa conformation dans le feuillet. Nous voyons également que la bande e est plus prononcée dans la zone comprise entre les atomes 70 et 107 (constituant les résidus Asn6 et Tyr7). Les interactions Gln5/Asn6 ou Gln5/Tyr7 vont donc favoriser le dépliement du peptide de sa forme R1 vers sa forme dans le feuillet.

Pour la conformation R2, on voit que la bande f présente des valeurs maximales pour des pincements avec les atomes 53 à 70, qui correspondent aux atomes de Gln5 : des interactions Gln5/Asn6 favorisent le dépliement de la forme R2. Les valeurs situées sur la bande b sont également particulièrement élevées, notamment aux intersections avec les bandes c et g. Dans cette conformation, les interactions Asn2/Asn3 et Asn2/C-terminal semblent donc favoriser également le dépliement du peptide.

Dans la conformation A, on observe les valeurs les plus élevées de $\Delta RMSD$ sur les bandes b, e et g, indiquant ainsi que les interactions impliquant Asn2, Gln5 ou la terminaison COO^- déforment la molécule vers sa forme dans le feuillet. Mais ces bandes ne sont pas totalement uniformes. Par exemple, la bande e présente un « trou » pour des valeurs correspondant aux atomes de Gln4. Des interactions entre Gln4 et Gln5 ne semblent donc pas favoriser le dépliement du peptide. De même, pour la bande g, on observe des trous pour des pincements avec les atomes 70 à 107 (constituant Asn6 et Tyr7).

D'après ces diagrammes, nous voyons donc que les interactions entre les chaînes latérales du peptide modèle, quel que soit sa conformation, favorisent un dépliement du peptide vers sa conformation dans le feuillet. En effet, nous observons le même motif sur chacun des diagrammes de la figure 21. Takeda et Klimov ont observé sur des peptides amyloïdes que les interactions inter-peptide entre chaînes latérales sont à l'origine de la formation de feuillet β

(148). Nous montrons que des interactions intra-moléculaires entre chaînes latérales favorisent le dépliement du peptide vers sa conformation dans le feuillet.

De manière plus générale, si l'on compare les diagrammes de la figure 21 avec ceux de la figure 20, nous constatons que pour des structures équivalentes, les bandes qui apparaissent sont différentes. Cela suggère que le mouvement des atomes du squelette et celui des atomes formant les liaisons hydrogènes sont décorrélés. Or, les atomes formant les liaisons hydrogène sont en majorité situés sur les chaînes latérales. Ces dernières sont connues pour être flexibles en comparaison avec le squelette qui est lui considéré comme semi-rigide. Cela explique les différences entre les diagrammes des figures 20 et 21.

III-2.1.2. Calculs sur les dimères

Nous avons vu au paragraphe **III-1.2.** que la première étape dans l'agrégation du peptide GNNQQNY était la formation d'un feuillet antiparallèle plutôt que parallèle. Nous allons donc nous intéresser à ces deux structures et les comparer afin d'apporter des réponses sur les premières étapes de l'agrégation.

Sur la figure 22 nous avons représenté les feuillets parallèle et antiparallèle obtenus avec AMBER8. Ils possèdent respectivement une énergie de -187.54 kcal/mol et -262.31 kcal/mol. Nous pouvons donc constater que la conformation antiparallèle est plus favorable énergétiquement que le feuillet parallèle, ce qui est en accord avec les travaux précédents.

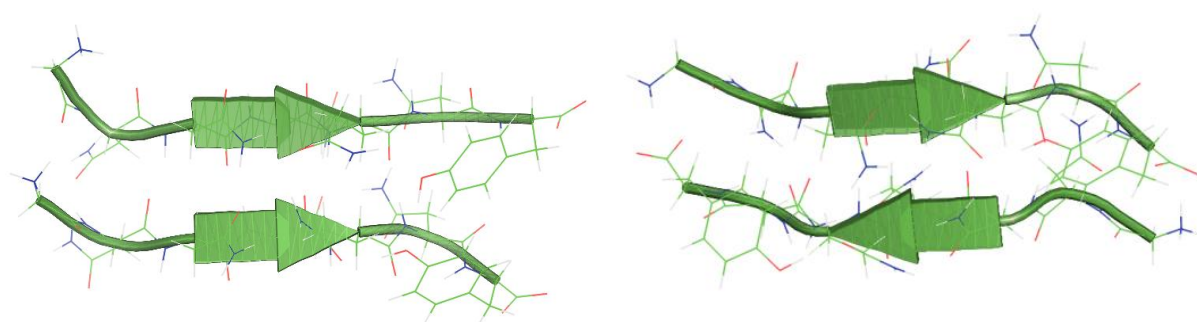


Figure 22 : feuillets parallèle (à gauche) et antiparallèle (à droite) obtenus avec AMBER8

Nous allons à présent tester la stabilité du feuillet afin de déterminer s'il est plus stable dans une conformation que dans l'autre. La liaison hydrogène étant une interaction importante dans la formation et la stabilité du feuillet, nous calculons un RMSD sur les

atomes formant des liaisons H inter-peptide, pour tous les modes de la molécule. Nous excitons un mode, puis le RMSD est calculé entre la structure déformée et la structure non déformée. Nous obtenons alors le graphique de la figure 23.

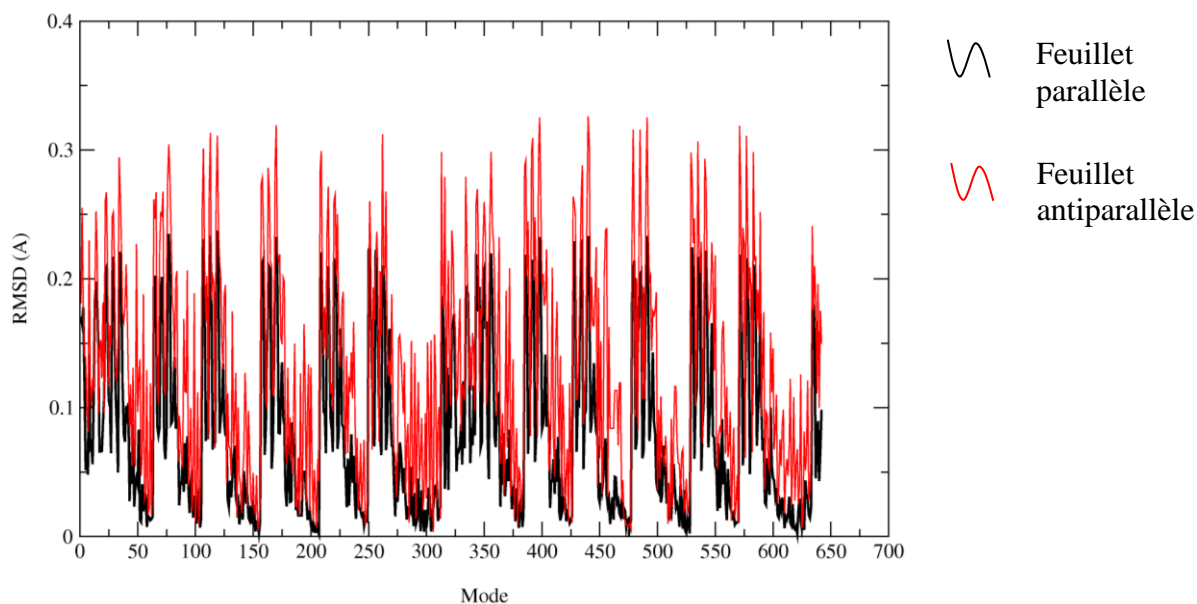


Figure 23 : calcul du RMSD en fonction des modes de déformation pour les feuillets parallèle et antiparallèle.

Nous pouvons donc voir que les valeurs du RMSD calculées sur le feuillet antiparallèle sont plus importantes que celles calculées sur le feuillet parallèle. Cela signifie donc que dans une configuration antiparallèle, le dimère est moins stable que dans une configuration parallèle. Afin de valider ces résultats, nous avons voulu les confronter avec un calcul de dynamique moléculaire. Pour cela, nous avons chauffé les dimères parallèle et antiparallèle de 0 à 600 K en utilisant le logiciel AMBER8. En partant de la structure de la figure 22, nous avons effectué des simulations de 20 ps en augmentant la température de 10 K à chaque étape. En chauffant ainsi le complexe, nous allons briser ces liaisons hydrogène tant importantes dans la stabilité du feuillet. Nous suivons alors l'évolution des structures par le RMSD calculé sur tous les atomes du dimère. Nous pouvons voir le graphique correspondant sur la figure 24.

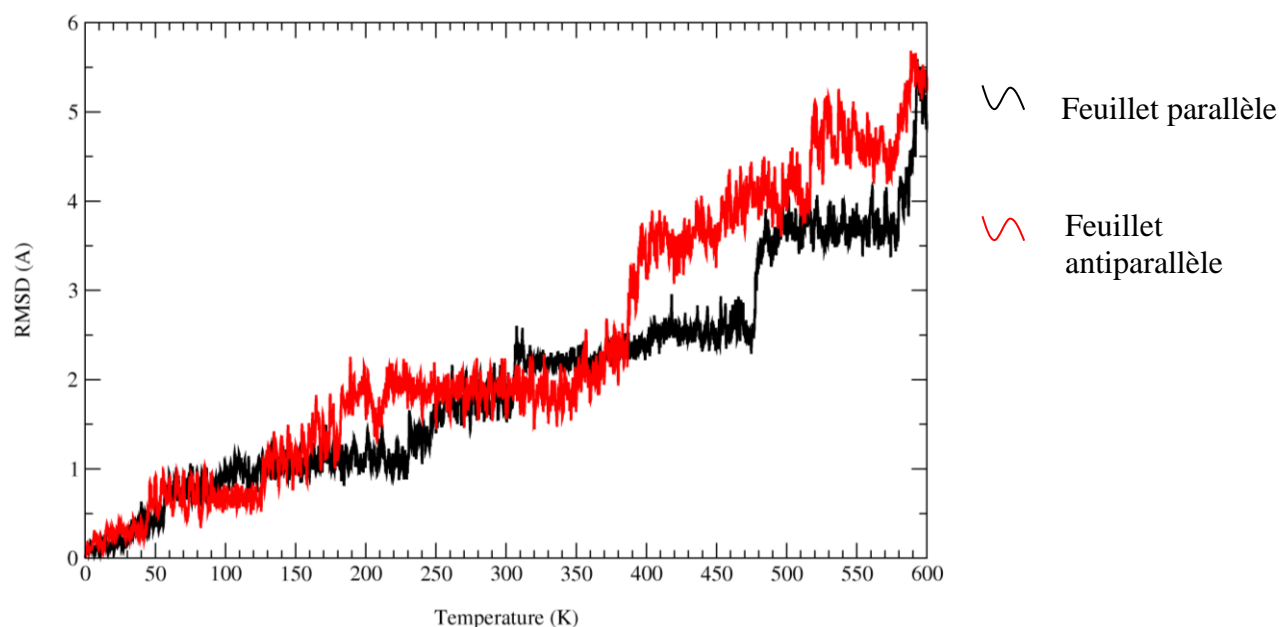


Figure 24 : Evolution du RMSD en fonction de la température lors de la simulation effectuée avec AMBER 8.

On voit sur la figure 24 des paliers se former lorsque la température atteint certaines valeurs. Le premier a lieu pour une valeur de 40K pour le feuillet antiparallèle pour se stabiliser entre des valeurs de 0.5 et 1.0 angström. On observe ce même saut pour le feuillet parallèle à une température de 60K. Ce dernier semble donc être le plus stable des deux. Globalement, les différents paliers sont atteints d'abord par les valeurs du feuillet antiparallèle, suggérant que ce dernier est moins stable que le dimère parallèle. Ceci vient donc confirmer les résultats obtenus avec les modes statiques.

Nous avons donc une structure en feuillet antiparallèle plus favorable énergétiquement que le feuillet parallèle. Cela signifie que deux peptides formant un dimère préfèrent se mettre dans une configuration antiparallèle, ce qui est en accord avec les observations reportées dans la littérature. En revanche, d'un point de vue thermodynamique, le feuillet antiparallèle est moins stable que le feuillet parallèle, comme le suggère la figure 24. Nous retrouvons également ce résultat par les modes statiques sur la figure 23. Le feuillet parallèle étant plus stable cela expliquerait alors pourquoi on trouve le peptide dans cette configuration dans sa forme agrégée.

III-2.2. Cas de l'amyloïde β

Nous nous concentrons à présent sur le peptide amyloïde β , séquence 1-28. Nous nous basons sur trois structures différentes, illustrées sur la figure 25. On retrouve une hélice α , un brin β , c'est-à-dire un monomère en conformation étendue proche de sa forme dans un feuillet, ainsi qu'un feuillet β composé de trois peptides. Dans le cas du feuillet, la structure utilisée fait référence au brin situé au centre du trimère (brin indiqué par la flèche sur la figure 25). Ces trois conformations sont notées respectivement : α , β_{str} et β_{sh} .

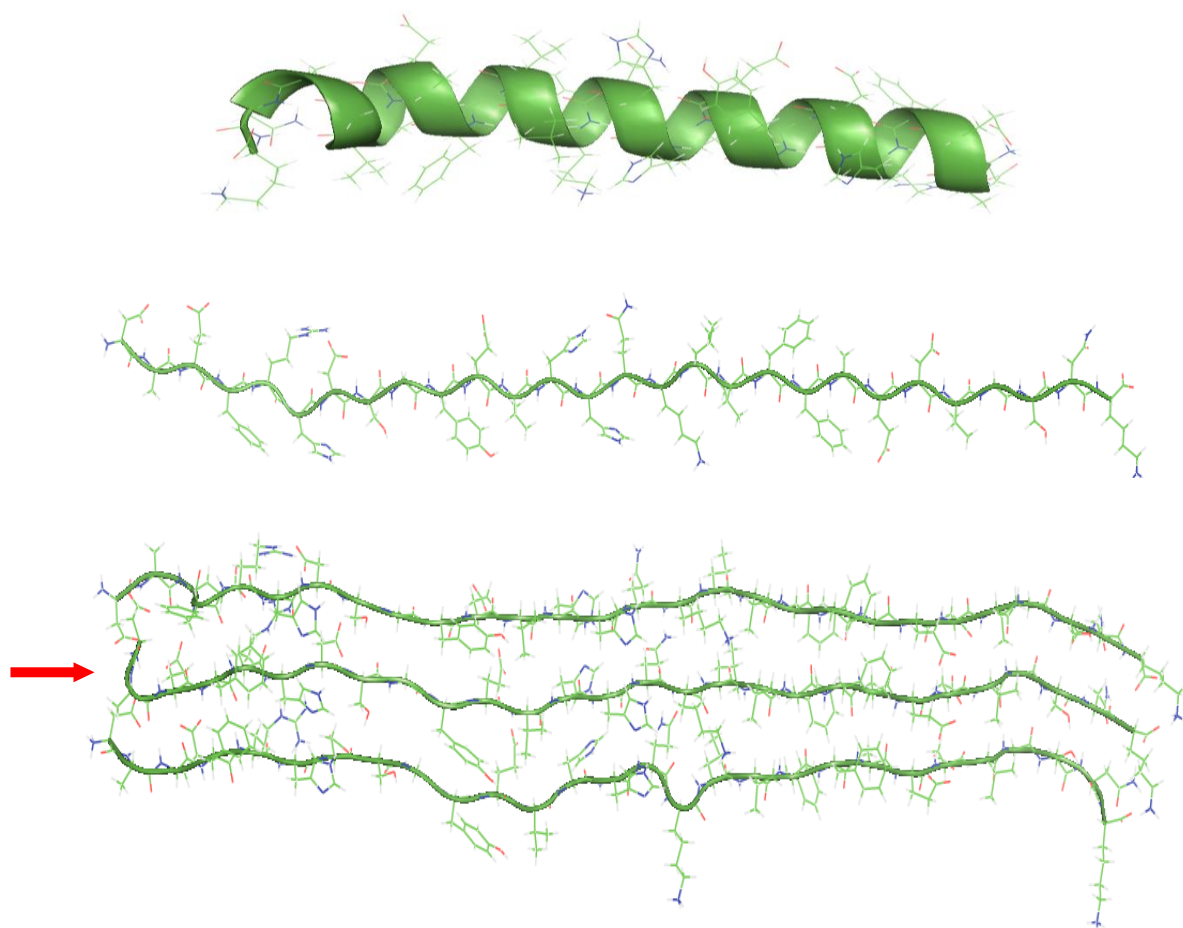


Figure 25 : Les trois conformations utilisées dans cette étude. A) structure en hélice α (noté α). B) peptide seul en conformation β (noté β_{str}). C) feuillet β composé de trois brins. Le brin du milieu (indiqué par la flèche rouge) est noté β_{sh} .

La conformation α est issue de l'expérience et a été déterminée par Talafoous *et al* (175). Elle a été publiée et diffusée sur la base de données Protein Data Bank (184), le code PDB de cette molécule étant 1AMC. Les conformations β_{str} et β_{sh} ont, quant à elles, été générées grâce

CHAPITRE III : RESULTATS

au module LEaP du logiciel de mécanique moléculaire AMBER8. Ces trois structures ont été optimisées de la même manière que pour le peptide GNNQQNY (v. § **III-2.1.1**). Nous avons donc utilisé ces structures dans le but de caractériser les fluctuations atomiques de ces conformations de base des protéines.

Nous avons pu classer ces trois conformations en termes de flexibilité, à partir d'un calcul sur les angles dièdres φ et ψ du squelette. Nous calculons pour chaque angle la déformation engendrée par l'excitation de chaque mode, un à un, et effectuons la moyenne de ces déformations. L'opération est répétée pour tous les angles du squelette, et une nouvelle moyenne est effectuée sur tous les angles. Cela se traduit par la formule :

$$\langle \Delta\theta \rangle = \frac{1}{N_{dih}} \sum_{i=1}^{N_{dih}} \left(\frac{1}{N_{mod}} \sum_{j=1}^{N_{mod}} (\theta_o^i - \theta_j^i) \right)$$

On obtient ainsi une valeur de $\langle \Delta\theta \rangle$ pour chacune des trois conformations, ce qui nous donne une idée de la rigidité du squelette de chacune de ces différentes structures. Ces valeurs sont indiquées dans le tableau 4.

Conformation	$\langle \Delta\theta \rangle$
α	1.84
β_{str}	1.92
β_{sh}	0.96

Tableau 5 : valeur de la déformation moyenne pour chacune des conformations du peptide Amyloïde- β : α , β_{str} et β_{sh}

Ce tableau révèle le fait que la conformation β_{sh} , soit le peptide situé au milieu du feuillet, possède l'indice le plus faible : cela signifie que cette structure est la plus rigide. Quant aux deux autres conformations, elles possèdent toutes deux des indices très proches, suggérant ainsi que le peptide seul dans son environnement est flexible. Toutefois, la valeur de $\langle \Delta\theta \rangle$ correspondant à la conformation α est inférieure à celle calculée sur β_{str} , signifiant que la conformation α est légèrement plus rigide que la conformation β_{str} .

Dans le paragraphe **I-2.2.3.**, nous avons constaté que les conformations α et β des protéines étaient caractérisées par un réseau régulier de liaisons hydrogènes, ce qui nous permettait de les classer suivant leur rigidité. Les valeurs de $\langle \Delta\theta \rangle$ présentées dans le tableau 3 sont en

accord avec ces affirmations précédentes.

III-2.3. Représentation des zones flexibles des molécules

Les conformations α , β_{str} et β_{sh} , couplées avec les outils décrits dans le second chapitre, nous ont permis d'établir une cartographie des déformations. Pour cela, nous avons regardé les déformations des angles dièdres ϕ et ψ du squelette de la protéine. Pour chaque atome de la molécule, la direction de la force a été optimisée afin d'obtenir une déformation maximale des angles ϕ d'une part et ψ d'autre part.

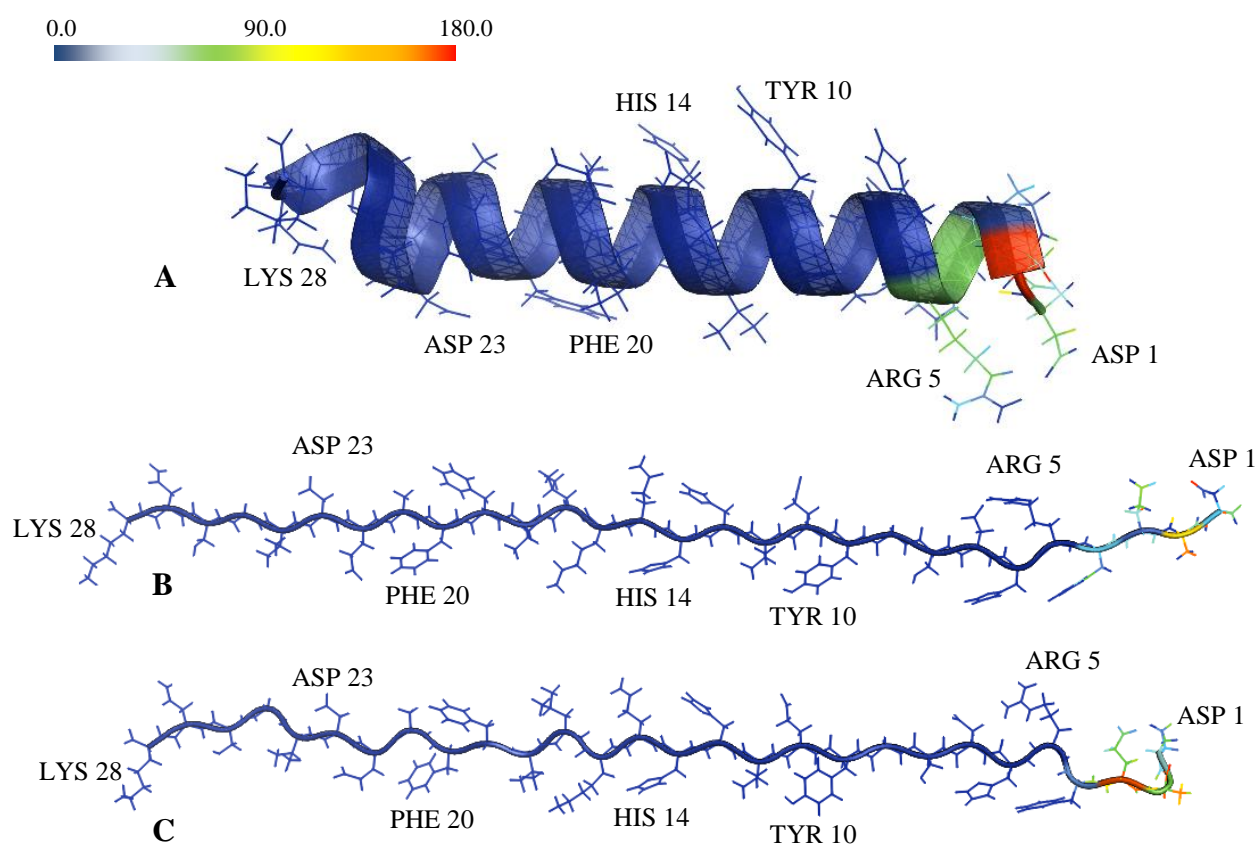


Figure 26 : valeurs maximales de déformation de l'angle ϕ_2 pour les trois structures α (A), β_{str} (B) et β_{sh} (C).

Nous reportons ensuite ces valeurs dans la colonne réservée aux valeurs du facteur de température du fichier pdb. Le logiciel PyMOL permettant de visualiser ces valeurs, nous

CHAPITRE III : RESULTATS

pouvons observer, pour un angle donné, sur chaque atome du peptide la valeur de la déformation lorsqu'on applique une force sur cet atome. Nous avons représenté sur la figure 26 les valeurs des déformations maximales de l'angle ϕ_2 pour les 3 conformations étudiées.

Afin de pouvoir visualiser les valeurs maximales de déformation pour tous les angles sur une unique figure, nous avons construit les diagrammes représentés sur les figures 27 et 28. Chaque résidu i est décrit par deux angles, que nous noterons ϕ_i et ψ_i . Les résidus situés aux extrémités ne peuvent être décrits par deux angles dièdres. En effet, il n'existe pas d'angle ϕ_1 , ni d'angle ψ_{28} , car les atomes utilisés pour les décrire ne sont pas présents. Nous avons donc 27 ϕ et 27 ψ pour cette molécule, et avons choisi de représenter les ϕ sur un premier diagramme et les ψ sur un second. En abscisse, on retrouve le numéro du résidu (donc de l'angle de torsion), en ordonnée nous aurons le numéro de l'atome sur lequel la force est appliquée. L'intensité de la déformation, mesurée en degré, sera alors représentée par un code couleur : 0° = blanc, 120° = bleu et 180° = rouge.

III-2. Caractérisation de la flexibilité

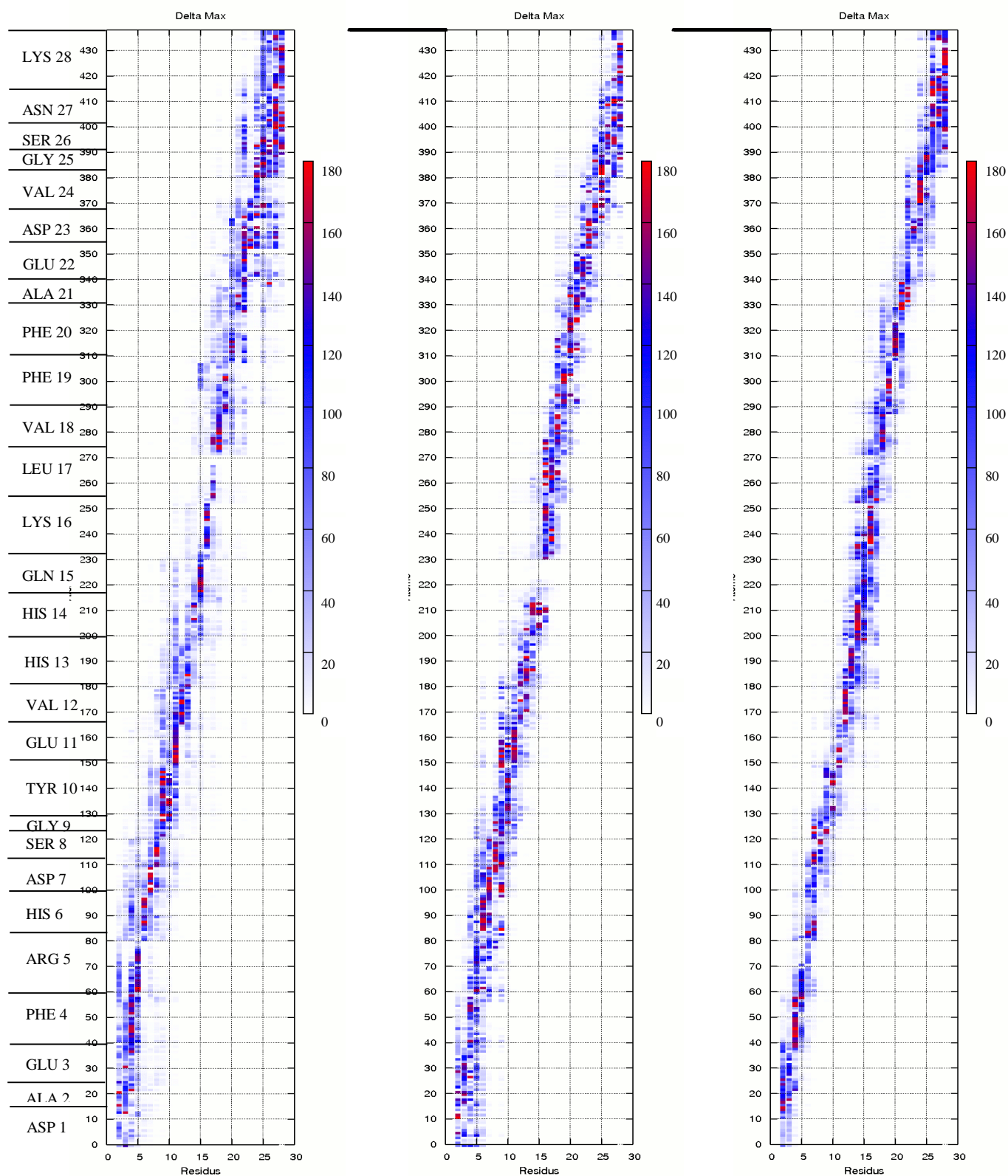


Figure 27 : Déformations maximales des angles ϕ du squelette pour les trois conformations. De gauche à droite : α , β_{str} , β_{sh} .

CHAPITRE III : RESULTATS

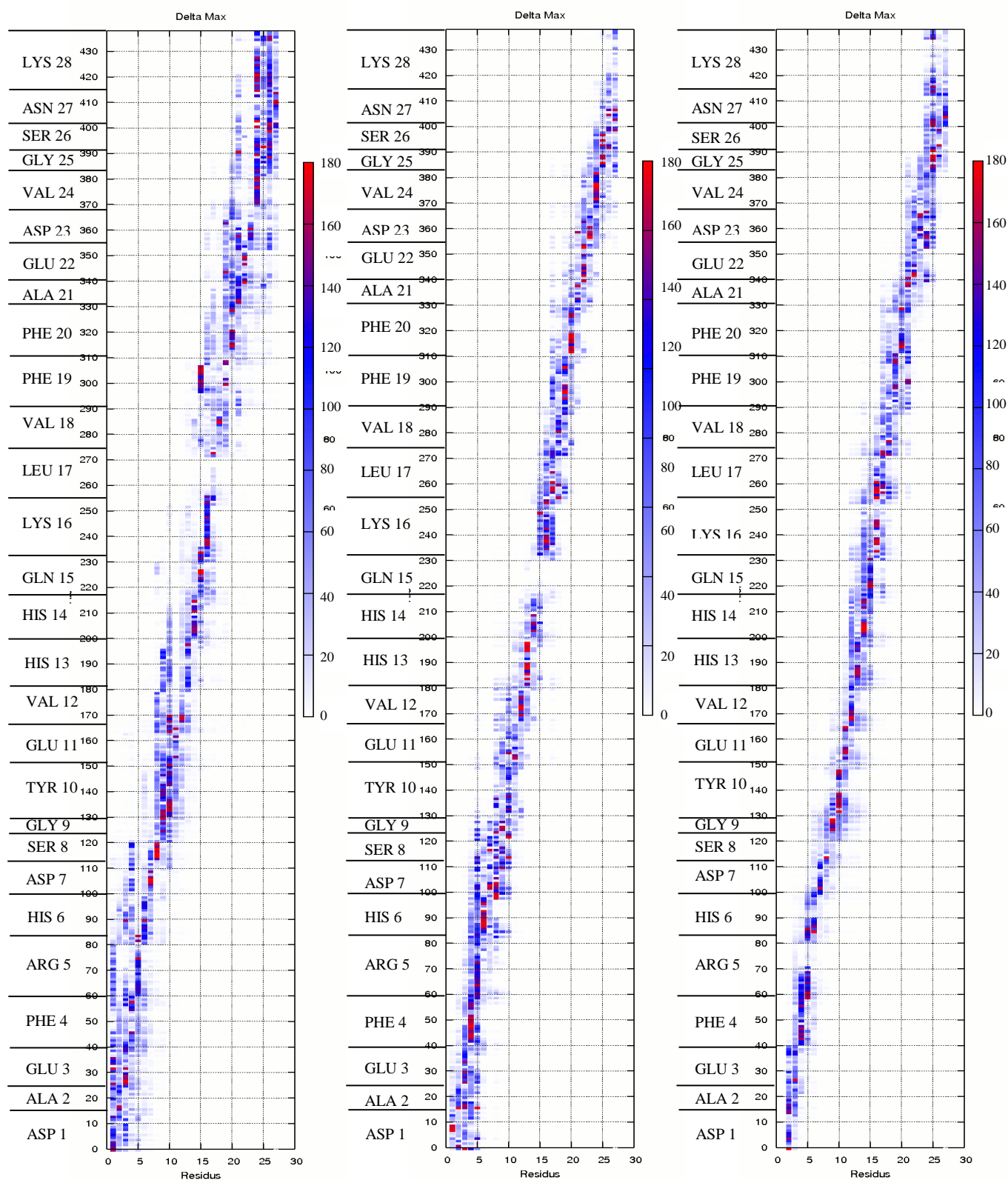


Figure 28 : Déformations maximales des angles ψ du squelette pour les trois conformations. De gauche à droite : α , β_{str} , β_{sh} .

Nous pouvons constater d'après la figure 26, que les atomes sur lesquels appliquer une force afin de déformer un maximum l'angle ϕ_2 sont proches des atomes qui définissent cet angle. De même, sur les diagrammes des figures 27 et 28, nous observons que les points sont situés sur la diagonale, indiquant ainsi que les atomes où appliquer une force afin de déformer au maximum chacun des angles de torsion du squelette sont proches de cet angle.

Ces diagrammes nous procurent des informations sur plusieurs caractéristiques pour ces structures, parmi lesquelles on recense :

- La proportion d'atome participant à la déformation.
- Le positionnement de ces atomes par rapport à l'endroit de la déformation.
- Les zones d'excitation auxquelles la déformation est la plus sensible. Il s'agit d'atomes qui induisent une déformation maximale.

En associant ces différentes observations, nous pouvons déduire des caractéristiques liées à la flexibilité de ces conformations types du peptide Amyloïde- β . Certaines de ces caractéristiques sont maintenant abordées.

III-2.3.1. Propagation de la déformation

Tout d'abord, nous pouvons voir sur ces diagrammes que les points sont groupés autour de la diagonale. Cette observation nous révèle donc que si l'on veut agir sur un angle du squelette, il faut appliquer une force sur un atome situé à proximité de cet angle. Néanmoins, nous pouvons noter que ces points sont en moyenne plus ou moins dispersés suivant la conformation que l'on regarde.

En effet, la différence majeure entre les conformations α et β observée sur ces diagrammes est que, lorsque l'on regarde un angle en particulier (c'est-à-dire les points le long d'une verticale), il semble que les points sont plus étalés dans le cas de l'hélice α en comparaison avec les conformations β . En d'autres termes, il semble que dans le cas d'une conformation en hélice, un plus grand nombre d'atomes peut être sollicité pour déformer le squelette que dans le cas d'une conformation β .

Nous avons donc calculé le nombre moyen de résidus possédant un ou plusieurs atome(s) participant à la déformation. Seules les déformations angulaires supérieures à 20 degrés sont considérées. Les nombres de résidus sont présentés dans le tableau 6.

CHAPITRE III : RESULTATS

Conformation	Nombre de résidus
α	6.04
β_{str}	5.65
β_{sh}	4.93

Tableau 6 : nombre moyen de résidus possédant au moins un atome participant à la déformation.

On peut donc voir que dans la conformation α , plus de résidus sont susceptibles de participer à la déformation du squelette du peptide que dans une conformation β . Nous allons donc à présent déterminer pourquoi on observe cette différence entre ces conformations.

Les protéines sont des molécules composées d'atomes liés de manière covalente. Les déformations se propagent le long de la molécule par ces liaisons covalentes, qui possèdent chacune une rigidité qui leur est propre. Or, il est clair d'après les diagrammes que la déformation ne passe pas uniquement par ces liaisons de covalence. On peut donc s'interroger sur ce fait, et se demander quel phénomène est à l'origine de l'étalement des points observé sur les diagrammes.

Les structures secondaires des protéines sont régies par les interactions de type non-liées, que nous avons décrites au premier chapitre, et notamment par les liaisons hydrogènes qui stabilisent la structure dans une certaine conformation. Nous avons donc émis l'hypothèse que lorsque l'on agit sur un atome de la molécule, la déformation induite se propage via les liaisons covalentes, mais aussi via les liaisons hydrogène.

Pour illustrer cette idée, nous avons appliqué une force sur chaque atome et examiné une déformation particulière de la molécule, choisie à l'avance. La direction de la force est optimisée pour rendre la déformation maximale et son module est normalisé. Nous avons alors reporté, sur chaque atome de la molécule, la déformation maximale provoquée par cette force. Nous avons en fait utilisé la colonne du fichier PDB réservée au facteur de température pour y stocker les valeurs des déformations maximales. Le logiciel PyMOL permet de visualiser ces facteurs de température, qui sont dans notre cas les maxima de déformation. Nous pouvons voir le résultat obtenu sur la figure 29.

Sur cette figure, nous avons choisi un exemple explicite, qui est de représenter les déformations de l'angle ϕ_5 dans les conformations β_{str} et β_{sh} . En effet, dans la conformation β_{str} , on observe trois liaisons hydrogènes entre les chaînes latérales de l'arginine 5 et de

l'acide aspartique 7, alors que dans la conformation β_{sh} , ces liaisons n'existent plus. Nous avons donc observé la contribution des atomes d'Asp7, et il s'avère que ces derniers participent à la déformation de ϕ_5 en présence des liaisons hydrogènes, et ne sont plus impliqués sans liaisons hydrogènes.

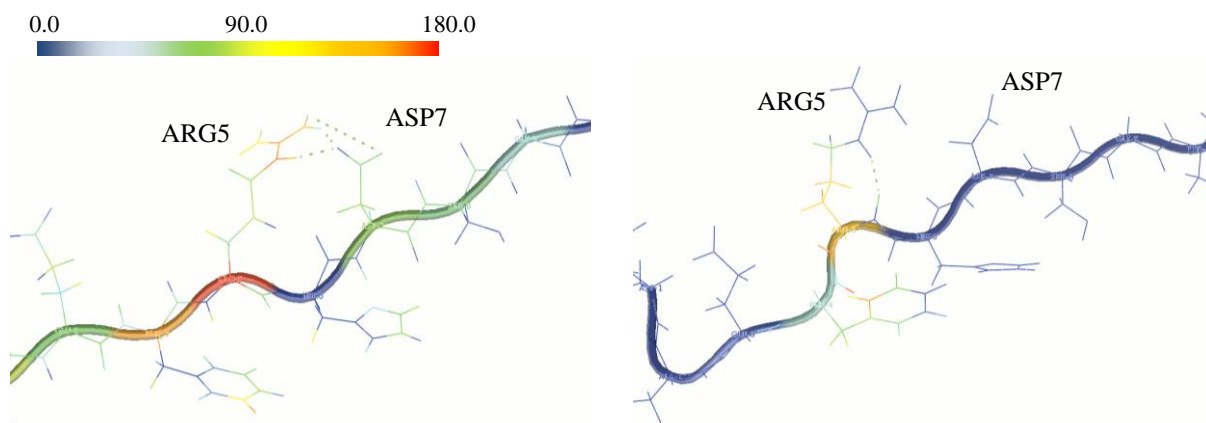


Figure 29 : Résidus 3 à 9 du peptide Amyloïde- β . La coloration représente l'amplitude de la déformation de l'angle ϕ_5 . A gauche : conformation β_{str} , sur laquelle on observe 3 liaisons hydrogène entre les résidus Arg5 et Asp7. A droite : conformation β_{sh} sur laquelle on n'observe plus de liaisons hydrogène.

Ces résultats suggèrent donc que les déformations ne se propagent pas uniquement par les liaisons covalentes de la molécule. Un léger doute peut cependant exister. En effet, dans le cas présenté sur la figure 29, les atomes de la chaîne latérale de Asp7 et les atomes formant l'angle de torsion ϕ_5 sont certes éloignés, mais appartiennent à la même molécule. On peut penser que la déformation a tout de même pu se propager le long des liaisons covalentes. Pour éliminer définitivement cette hypothèse nous avons effectué ce même calcul entre deux groupes d'atomes de deux molécules distinctes, liés par une ou plusieurs liaison(s) hydrogène. C'est ce que nous avons fait sur un feuillet β composé de deux peptides de taille plus modeste : le peptide GNNQQNY présenté au paragraphe précédent.

Sur la figure 30, nous avons représenté le feuillet. Les deux brins sont liés entre eux par 11 liaisons hydrogènes. Nous en avons choisi une, indiquée par la flèche sur la figure 30, et

CHAPITRE III : RESULTATS

appliqué une force dans la direction de la liaison. Nous observons alors la réponse du peptide faisant face à l'atome sur lequel on agit pour voir dans quelle mesure la liaison hydrogène joue un rôle dans la propagation des déformations.

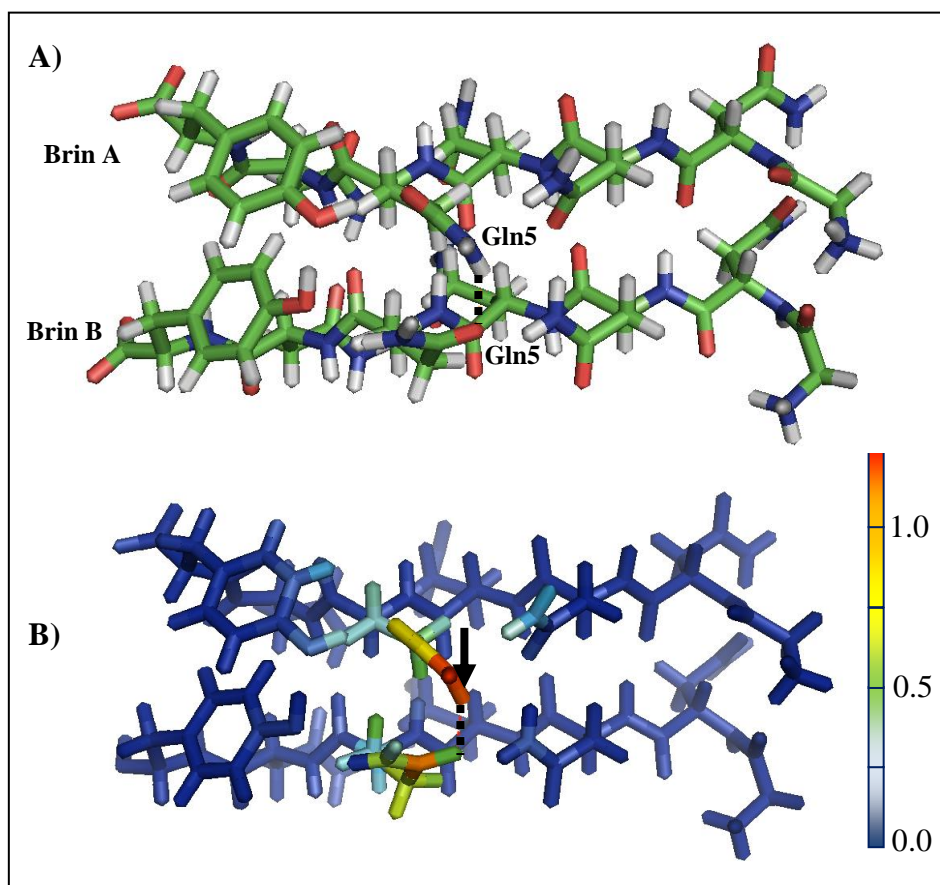


Figure 30 : peptide GNNQQNY : 2 monomères en feuillet β antiparallèle. A) structure 3D du feuillet. B) représentation des amplitudes de déformation (en Å) lorsque l'on applique une force le long de la liaison hydrogène (pointillé) comme le suggère la flèche.

Sur la figure 30, nous avons représenté les variations des positions atomiques lorsque l'on applique une force le long d'une liaison hydrogène entre les résidus Gln5 du brin A et Gln5 du brin B. Les atomes qui présentent l'amplitude la plus élevée sont situés sur le brin A, mais un groupe d'atomes du brin B est également largement affecté par la force appliquée. Certains atomes de ce groupement possèdent parfois même une amplitude plus élevée que certains atomes proche de l'endroit où la force est appliquée. C'est le cas notamment de l'oxygène situé en bout de chaîne latérale de Gln5 du brin A, qui est situé à proximité de

l'endroit où est appliquée la déformation, et auquel correspond une amplitude de 0.8 Å. Alors que le carbone en bout de chaîne latérale de Gln5 du brin B possède une amplitude de 1.1 Å. Ce résultat met en évidence que les déformations peuvent se propager par l'intermédiaire de liaisons physiques telles que la liaison hydrogène.

III-2.3.2. Comportement allostérique

Les diagrammes des figures 27 et 28 présentent tous une diagonale, ce qui signifie que si l'on veut déformer un angle dièdre du squelette du peptide, il faut solliciter un ou plusieurs atomes relativement proche(s) de cet angle. Or, dans la configuration β_{str} (diagramme du milieu), on observe un « trou » dans cette diagonale au niveau du résidu Arg15: agir sur les atomes de ce résidu n'a aucune, ou très peu d'influence sur les angles ϕ_{15} et ψ_{15} . Ceci est illustré par la figure 31, où nous avons zoomé sur la zone d'intérêt.

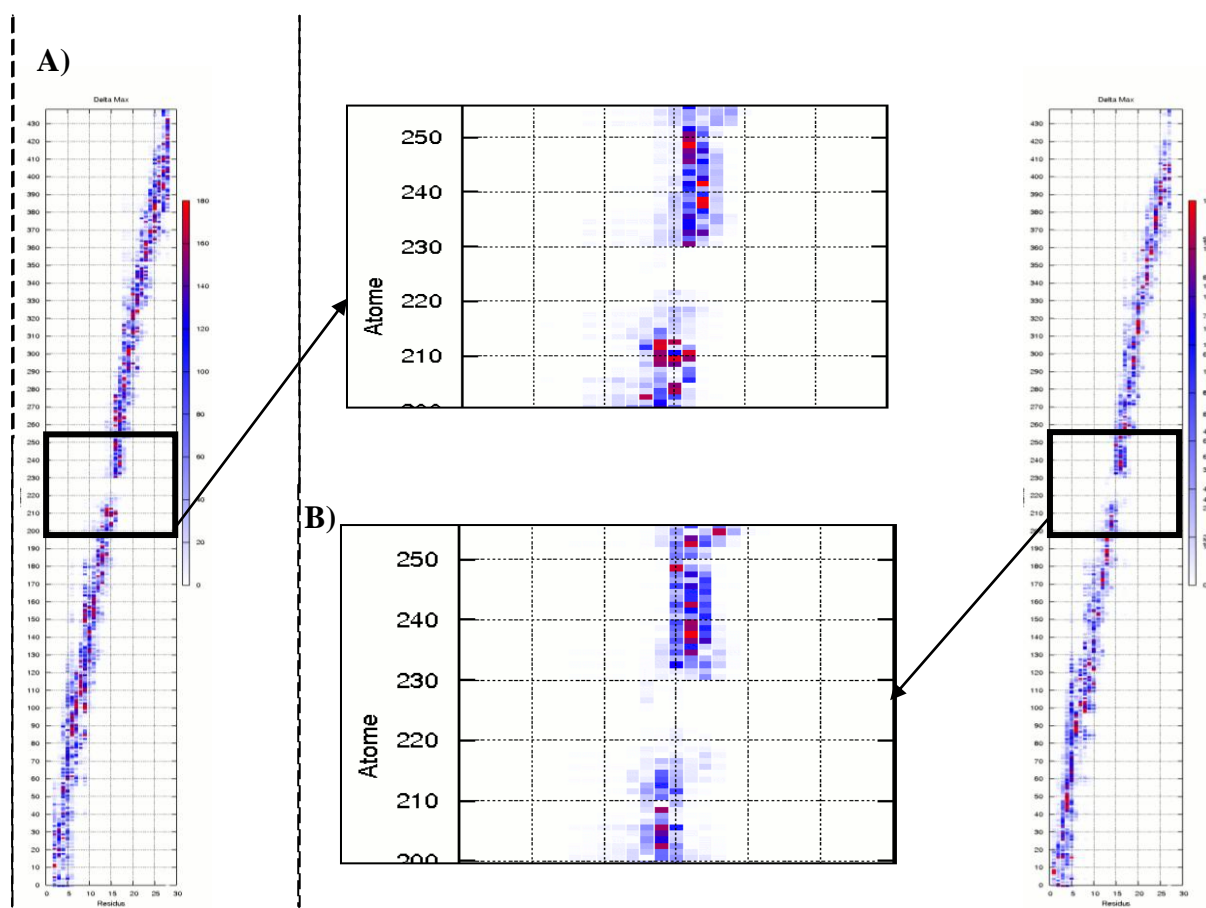


Figure 31 : zoom sur les diagrammes correspondant aux angles ϕ et ψ de la configuration β_{str} .

CHAPITRE III : RESULTATS

Cette discontinuité observée dans la diagonale de ces diagrammes révèle le fait que d'agir sur le résidu 15 n'a aucune incidence sur la déformation du squelette de la molécule à ce même endroit. C'est une sorte de point mou de cette structure du peptide. En revanche, si agir sur les atomes proches de la déformation observée n'a aucune incidence, ce n'est pas vrai pour des atomes plus éloignés. Afin de déterminer quels sont ces atomes, nous avons représenté, sur chaque atome de la molécule soumis à une force, l'amplitude de la déformation engendrée par la force sur les angles ϕ_{15} et ψ_{15} (figure 32).

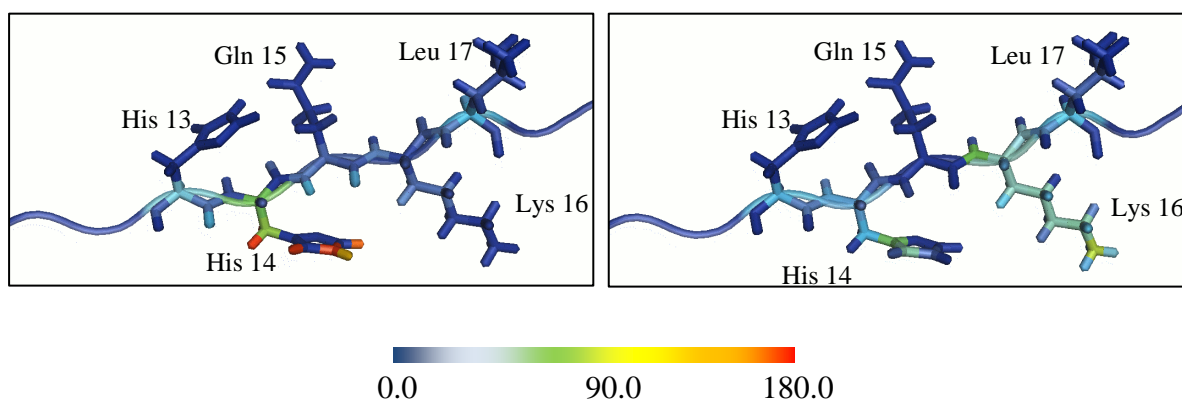


Figure 32 : représentation des amplitudes de déformation, en degrés, par atome lorsque l'on veut obtenir une variation maximale des angles ϕ_{15} (figure de gauche) et ψ_{15} (figure de droite).

On voit bien sur la figure 32 ce « trou » dans la chaîne peptidique. Mais le plus surprenant est d'observer que les forces qui engendrent un maximum de déformation sur l'angle ϕ_{15} doivent être appliquées sur les atomes situés sur la chaîne latérale de l'histidine 14 (figure 32 A). L'histidine 14 est un résidu polaire, susceptible de réagir avec l'environnement dans lequel se trouve la molécule (gain ou perte d'un proton en faisant varier le pH, interaction avec un cation...). En agissant sur ce dernier, on voit bien que le squelette varie fortement au niveau du résidu 15 : ce phénomène, bien que limité dans l'espace ici, rappelle un comportement allostérique.

Nous avons donc pu voir que ces diagrammes révèlent des comportements différents des trois conformations étudiées. Dans la partie suivante, nous allons utiliser ces représentations afin de mieux comprendre le mécanisme d'agrégation du peptide A β (1-28) dans notre vision de déformations induites.

III-2.3.3. Relation interactions/oligomérisation

A partir des valeurs de déformations maximales du squelette obtenues pour la conformation α , nous avons voulu voir, pour chaque résidu, quels sont les modes, c'est-à-dire les déformations pertinentes, qui expriment une « respiration » conformationnelle de l'hélice vers une conformation de type feuillet. Nous avons opté pour la procédure suivante : à partir des valeurs présentées dans les diagrammes des figures 27 et 28, nous groupons les couples (ϕ, ψ) pour chacun des résidu. A titre d'exemple, nous pouvons voir sur la figure 33 (A) ces valeurs de déformations pour Ala2 dans le plan (ϕ, ψ) . Sur la figure 33 (B), est représenté un tracé de type Ramachandran des couples (ϕ, ψ) de tous les acides aminés (exceptée la glycine) relevés sur près de 100.000 résidus (185). Les zones rouges définissent les valeurs pour lesquels les résidus adoptent une forme privilégiée, et plus particulièrement en hélice α (domaine α) ou étendue (domaine β).

Les valeurs situées dans le carré en haut à gauche du graphique 33A) sont considérées comme « allant » vers une conformation β et sont comptabilisées. Nous effectuons ce calcul pour tous les résidus et nous traçons le graphique représentant ces valeurs par résidu (figure 34).

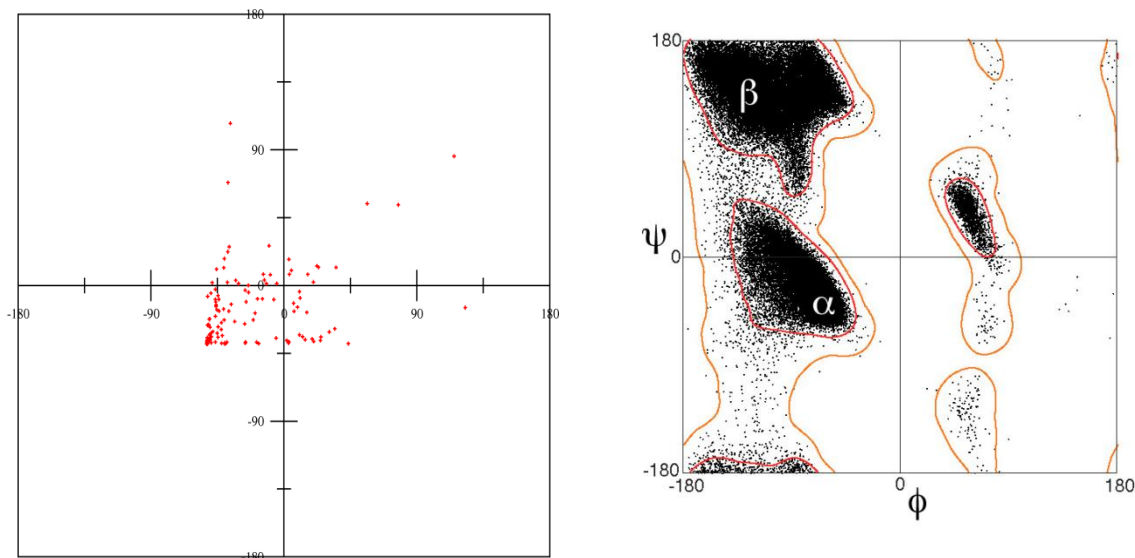


Figure 33 : A) Répartition dans le plan (ϕ, ψ) des valeurs maximales de déformation pour le résidu Ala2. B) Tracé Ramachandran de tous les résidus exceptés la glycine (185).

CHAPITRE III : RESULTATS

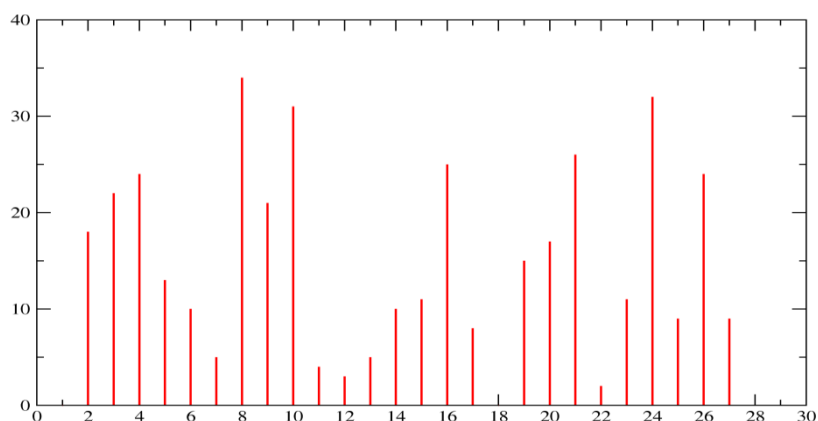


Figure 34 : Nombre de modes par résidu déformant les couples (φ, ψ) de la conformation α vers des valeurs caractéristiques de la forme β .

Sur la figure 35, nous avons reporté les probabilités, par résidu, de former des feuillets β obtenues par Dong *et al.* (186) sur le même peptide A β (1-28) par des simulations gros grains avec la technique ART. Nous constatons une correspondance entre les figures. En effet, par notre approche, nous trouvons que la portion Ala2-Glu3-Phe4 présente une tendance à s'étendre en feuillet β . Or, d'après la figure 35, cette zone a une forte probabilité de former un feuillet avec un second brin. Il en va de même pour les portions Ser8-Tyr10 et Val24-Ser26. Nous voyons toutefois que la partie Phe19-Ala21 présente des valeurs élevées sur la figure 34 et non sur la figure 35. En effet, d'après les travaux décrits dans cet article, le CHC a une faible probabilité de former un feuillet. Or, d'après les travaux de Petkova *et al.* (173) ou encore de Luhers *et al.* (187) sur le peptide entier (42 acides aminés), la partie centrale hydrophobe a tendance à former un feuillet β dans sa forme agrégée. Fernandez-Escamilla *et al.* ont constaté également grâce au logiciel TANGO qu'ils ont développé dans le but de prédire les régions susceptibles de s'agréger, que la partie Leu17-Ala21 présente une forte tendance à l'agrégation (188).

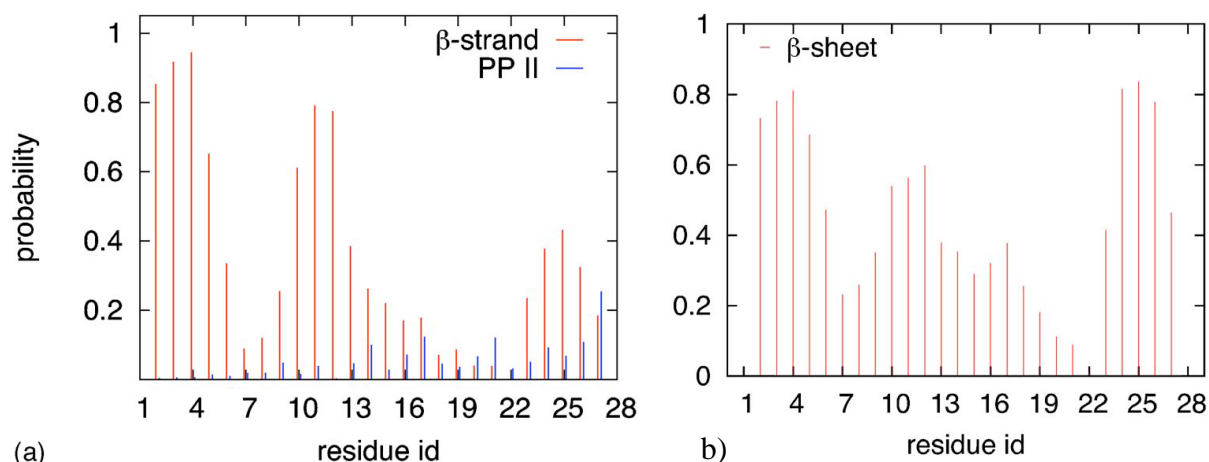


Figure 35 : Probabilité par résidu de former un feuillet β calculées (a) sur un monomère et (b) sur un dimère. Ces résultats ont été obtenus par Dong et al.

Nous allons à présent nous appuyer sur les diagrammes des figures 27 et 28.

Afin de visualiser les résidus les plus impliqués, nous avons simplifié les diagrammes des figures 27 et 28. Pour cela, nous avons comptabilisé, par résidu, le nombre de modes donnant une déformation supérieure ou égale à 100 degrés. C'est-à-dire que nous regardons pour chaque résidu les modes ayant une action forte sur les angles du squelette. Nous effectuons cela pour chaque angle et nous obtenons deux diagrammes par conformation (un pour ϕ , un pour ψ) rendant compte de l'implication de chaque résidu sur chaque angle du squelette. Dans ce paragraphe, nous pouvons voir, afin de visualiser la déformation globale du squelette au voisinage d'un résidu, un seul diagramme par conformation, qui est la somme des valeurs obtenues pour les angles ϕ et ψ (figure 36).

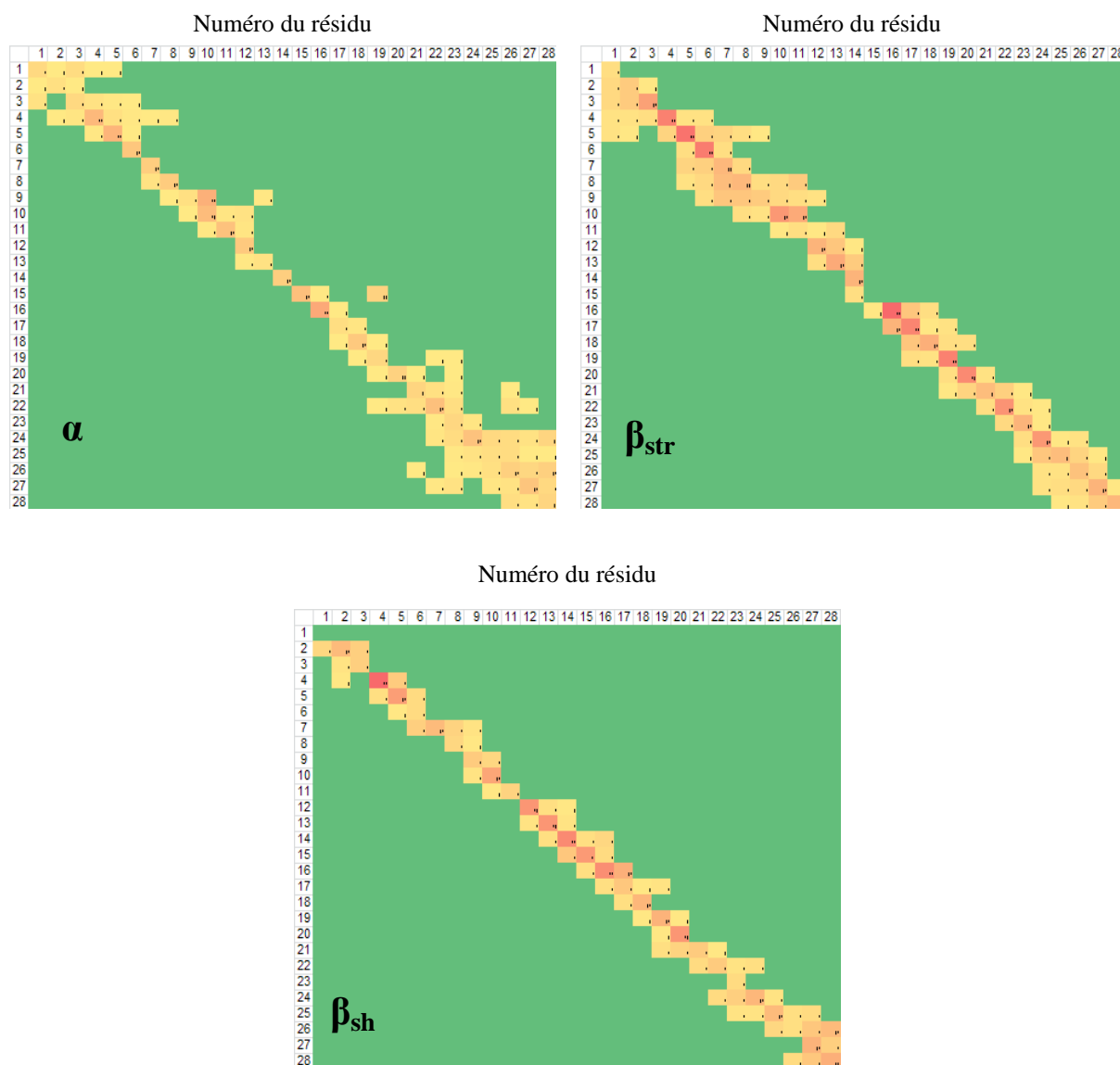


Figure 36 : Représentation simplifiée des diagrammes des figures 24 et 25. La numérotation horizontale correspond au résidu sur lequel la force est appliquée et la numérotation verticale correspond à l'angle dièdre.

De la même manière que sur les figures 27 et 28, nous voyons sur la figure 36 que les valeurs sont groupées le long de la diagonale et s'étalent plus ou moins suivant le résidu. On distingue deux types d'étalement : i) horizontal, qui va refléter la sensibilité du squelette et ii) vertical, qui va nous indiquer l'impact de l'excitation d'un résidu sur l'ensemble du squelette.

Avant de détailler les particularités de ces diagrammes, nous pouvons faire une comparaison globale. Les valeurs de la conformation β_{str} se trouvent plus étendues le long des lignes horizontales, en comparaison avec les autres conformations. Ceci reflète une sensibilité du

squelette plus élevée lorsque le peptide seul est dans sa forme étendue. Cette observation est en accord avec le résultat présenté dans le tableau 6. Ensuite, on observe un étalement le long des verticales plus important dans la conformation en hélice. Cela confirme que les déformations se propagent davantage dans la conformation α . Enfin, le diagramme correspondant à la forme β_{sh} présente des valeurs groupées autour de la diagonale, indiquant que dans cette conformation, le peptide est rigide et peu sensible aux interactions venant du milieu extérieur. Nous allons maintenant nous intéresser en détail à ces diagrammes.

On voit tout d'abord dans la conformation α les valeurs s'étalant horizontalement au niveau du résidu 4. Ceci reflète une sensibilité élevée du squelette à cet endroit de la molécule. Nous avons vu sur la figure 34 qu'à cet endroit de la molécule le squelette a tendance à former un feuillet β .

Dans la conformation β_{str} (figure 36), nous constatons une sensibilité du squelette au niveau de Gly9, révélé par un étalement horizontal des valeurs. La glycine, dont la chaîne latérale est composée d'un unique atome d'hydrogène, a tendance à rendre le squelette plus flexible. Dans cette conformation, la plus flexible des trois (tableau 6), cette sensibilité du squelette autour du résidu 9 ressort fortement, reflétant la flexibilité de la chaîne induite par la présence de la glycine. De plus, nous avons vu dans le paragraphe **III-1.1.** qu'à cet endroit de la molécule se forme un coude. Cette zone sensible révélée par le diagramme de la figure 36 représente donc une zone charnière du peptide. De plus, on observe cette partie sensible de la molécule uniquement dans cette conformation. Cela suggère donc que le peptide seul se déplie pour se mettre dans une conformation étendue, et est ensuite susceptible de se plier au niveau de Gly9.

La figure 36 indique également une forte sensibilité de la partie 24-28 du peptide en conformation α , connue pour être flexible. On retrouve une fois encore la glycine (position 25), qui a tendance à rendre la chaîne flexible.

La zone comprise entre ces deux extrémités semble donc être plus rigide, moins facile à perturber. On retrouve notamment la partie hydrophobe comprise entre les résidus 17 et 21, qui est connue pour être particulièrement rigide (189) (152). Cette portion du peptide tient un rôle important dans l'initiation et le contrôle de l'oligomérisation (190) (191). Or, nous constatons qu'il n'est pas facile de perturber le squelette à cet endroit de la molécule. Les résidus Phe19, Glu22 et Asp23 semblent tout de même jouer un rôle, comme le montre l'étalement vertical des valeurs correspondant à ces résidus. Des travaux récents ont

CHAPITRE III : RESULTATS

également montré que les interactions impliquant les anneaux aromatiques de la phénylalanine et de la tyrosine sont importantes lors de l'oligomérisation (192) (193). L'implication de Phe19 observée sur la figure 36, conformation α , semble donc avoir son importance dans les premières étapes de l'oligomérisation, et particulièrement dans le mécanisme de passage hélice $\alpha \rightarrow$ feuillet β .

Mais le résidu qui semble jouer le rôle le plus important est Asp23. En effet, lorsqu'il entre en interaction, il peut déformer le squelette de la molécule à proximité de 9 résidus. Son rôle dans l'oligomérisation et plus particulièrement dans le mécanisme de passage d'une conformation en hélice à une conformation en feuillet a été démontré expérimentalement (147).

III-2.5. Conclusion

Dans ce paragraphe, nous avons présenté une première application des modes statiques, qui est la caractérisation de la flexibilité moléculaire. A partir de calculs de RMSD sur un peptide modèle de sept acides aminés, nous avons pu établir des zones flexibles de la molécule et clefs dans la formation d'un feuillet, comme la glycine située en début de séquence. De plus, les diagrammes présentés sur les figures 20 et 21, indiquent les interactions intra-peptidiques pertinentes dans l'optique de se déplier vers sa conformation dans le feuillet.

Nous avons ensuite étudié le cas d'une molécule plus grosse : l'amyloïde β (1-28). En regardant les mouvements de torsion du squelette, nous avons cartographié sur des diagrammes les déformations de différentes conformations du peptide. Ces diagrammes nous ont permis d'expliquer le mode de propagation des déformations dans les biomolécules, et notamment l'importance des interactions faibles. De plus, nous avons relié ces données aux premières étapes de l'oligomérisation. En effet, en recoupant les informations fournies dans les diagrammes avec les données de la littérature, nous avons trouvé des interactions pertinentes pour la formation du peptide en feuillet β .

III-3. Application au docking

Nous avons vu au paragraphe précédent un exemple d'application dans le but de caractériser les déformations d'une molécule. Nous allons maintenant nous intéresser à la manière d'aborder le difficile problème du docking.

Nous avons évoqué deux manières d'aborder le problème :

- Utiliser les modes statiques pour générer des conformations du ligand et du récepteur pour les soumettre ensuite à des tests de scoring.
- Développer notre propre logiciel de docking en rapprochant pas-à-pas les molécules et calculer à chaque pas les déformations dues aux interactions.

Nous verrons, par l'intermédiaire de la sonde électrostatique, comment ces deux chemins peuvent être abordés.

Cette fois encore, nous nous sommes appuyés sur le peptide Amyloïde- β pour ces calculs. Mais dans cette étude, nous avons utilisé la portion (1-16) de la molécule, car cette partie du peptide est connue pour accrocher l'ion Zn^{2+} . De plus, des structures expérimentales, avec et sans le cation, ont été déterminées. Nous avons donc utilisé ces conformations expérimentales pour développer une première version d'un programme de docking.

III-3.1. Description des conformations utilisées

La partie 1-16 du peptide Amyloïde- β , notée $A\beta(1-16)$ et dont la séquence est explicitée dans le tableau 5, est connue pour être la partie initiatrice du processus d'agrégation de la molécule entière. De plus, c'est sur cette portion qu'ont été identifiés les sites d'accrochage de l'ion Zn^{2+} . Les structures apo (sans zinc) et liée (avec zinc) ont été publiées, et sont disponibles sur la *Protein Data Bank* (162).

Sur ces structures, le zinc se lie à quatre résidus du peptide : trois histidines (numérotées 6, 13 et 14), ainsi qu'à l'acide glutamique numéro 11.

A partir des structures de la Protein Data Bank (codes 1ZE7 pour la structure apo et 1ZE9 pour la structure liée), nous avons optimisé leur géométrie à l'aide du logiciel AMBER8 avec le champ de force ff03 : 5000 pas de descente de gradients suivis de 5000 pas de gradients conjugués. Nous obtenons alors les structures présentées sur la figure 37.

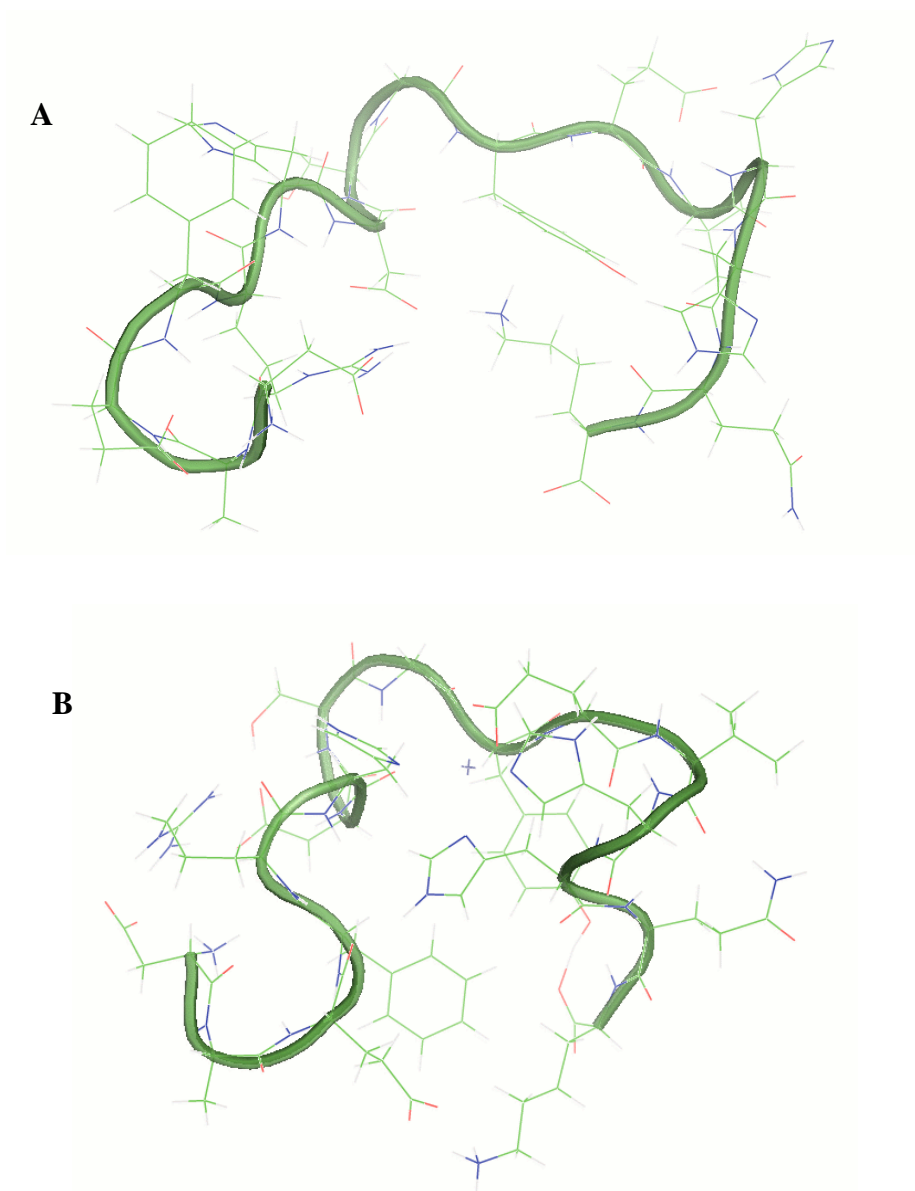


Figure 37 : structures apo (A) et liée (B) du peptide $A\beta(1-16)$ optimisées avec le logiciel AMBER8 (champ de force ff03)

Afin de décrire correctement les effets non-liés du cation métallique, nous avons utilisé une charge de + 2, ainsi que les paramètres de Lennard-Jones de Stote et Karplus mentionnés au paragraphe II-5.3. Les longueurs de liaison entre le zinc et les atomes auxquels il se lie ont été mesurées afin de vérifier si les valeurs trouvées après minimisation correspondent bien aux valeurs expérimentales. Sur la figure 38, nous avons représenté le site de fixation du zinc, avec une croix, et nous avons explicitée ses liaisons.

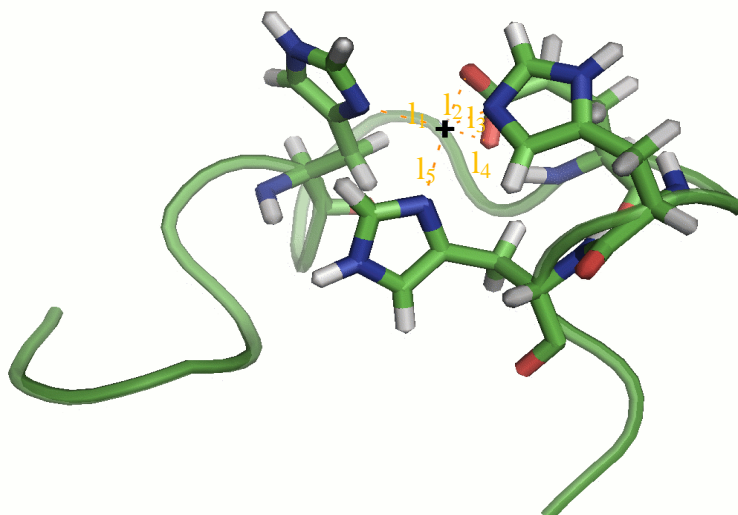


Figure 38 : site de fixation du zinc. Seuls les résidus impliqués sont représentés. Les liaisons sont indiquées en orange.

Dans le tableau 7, on retrouve les valeurs de ces liaisons pour la structure expérimentale, mesurées sur la structure PDB, et celles de la structure optimisée de la figure 38.

Liaison	Expérience	AMBER
l_1	2.11	2.14
l_2	2.11	1.93
l_3	2.15	2.05
l_4	2.11	1.92
l_5	2.29	2.18

Tableau 7 : valeurs des longueurs de liaisons, en Å, entre Zn^{2+} et les atomes auxquels il se fixe, mesurées sur la structure expérimentale et après optimisation de la géométrie.

Nous constatons que les longueurs mesurées après optimisation sont proches des valeurs expérimentales. En effet, l'erreur relative entre les valeurs expérimentales et celles calculées avec AMBER est inférieure à 10% dans chacun des cas. Cela signifie que les paramètres utilisés pour calculer l'énergie sont corrects, et nous pouvons donc les utiliser pour la suite de nos calculs.

III-3.2 Sonde électrostatique

III-3.2.1. Exploration de l'espace conformationnel de la structure A β (1-16) en interaction avec un ion métallique

Afin de sonder les effets de l'ion Zn^{2+} sur la structure apo du peptide A β (1-16), nous avons développé un premier programme permettant d'envisager des modifications conformationnelles utilisant les modes statiques : la sonde électrostatique. La méthode, décrite en détail au paragraphe **II-5.2.**, consiste à accrocher le cation sur des sites de la molécule susceptibles de réagir avec le cation, en minimisant l'énergie totale du système, par annulation des forces d'interaction. L'énergie totale contient à la fois l'énergie interne de déformation de la molécule et son énergie d'interaction avec le cation. L'algorithme de calcul commence par guider l'ion vers le site d'interaction en fonction des forces d'attraction. Les forces internes et la déformation moléculaire sont introduites dans l'algorithme quand le cation se trouve suffisamment proche de son site d'interaction. Les forces d'interaction entre le site et le cation sont alors considérées comme des forces d'excitation qui engendrent les déformations de la molécule à travers la matrice des modes statiques.

Pour déterminer les sites susceptibles d'accrocher le zinc, nous avons choisi des atomes électronégatifs du peptide ayant une charge supérieure à 0.5e, en valeur absolue. Cela représente 35 atomes sur la molécule, positionnés aussi bien sur le squelette que sur les chaînes latérales. Je rappelle la procédure suivie :

- Une fois les sites réactionnels déterminés, on rapproche le cation en minimisant les forces d'interaction via un potentiel attractif de Coulomb
- A proximité du site réactionnel, c'est-à-dire après un certain nombre d'itérations défini par l'utilisateur, on teste les collisions entre l'ion et le site d'une part, en introduisant un terme de type Lennard-Jones, et entre l'ion et les atomes proches par l'approximation des sphères dures.
- Une fois que l'on a testé les collisions, on déforme la molécule en utilisant le vecteur force induit par la présence de l'ion comme excitation en termes de modes statiques. Le vecteur force est l'opposé du gradient du potentiel, le champ de déformation induit est donné par les modes statiques.
- Une fois la molécule déformée, on optimise la structure en minimisant l'énergie à travers le logiciel AMBER8.

Pour suivre la conformation de la molécule, nous avons calculé le RMSDT présenté au paragraphe **II-5.1.3.** qui s'écrit :

$$RMSDT = \left[\frac{1}{N} \sum_{i=1}^N (\theta_{def}^i - \theta_{ref}^i)^2 \right]^{\frac{1}{2}}$$

Où N représente le nombre d'angles dièdres, θ est l'un des trois angles φ , ψ ou ω , les indices *def* et *ref* désignent respectivement la structure déformée et la structure de référence, qui est la structure liée de la figure 37. Les valeurs du RMSDT qui vont nous intéresser sont celles qui se rapprochent de zéro. Nous obtenons donc des triplets ($RMSDT_{\varphi}$, $RMSDT_{\psi}$, $RMSDT_{\omega}$) pour chacune des 35 structures déformées que nous représentons sur un diagramme (figure 39). Les fluctuations étant faibles dans le cas de ω , une fois de plus à cause de la rigidité de la liaison peptidique, on se limitera au plan (φ, ψ).

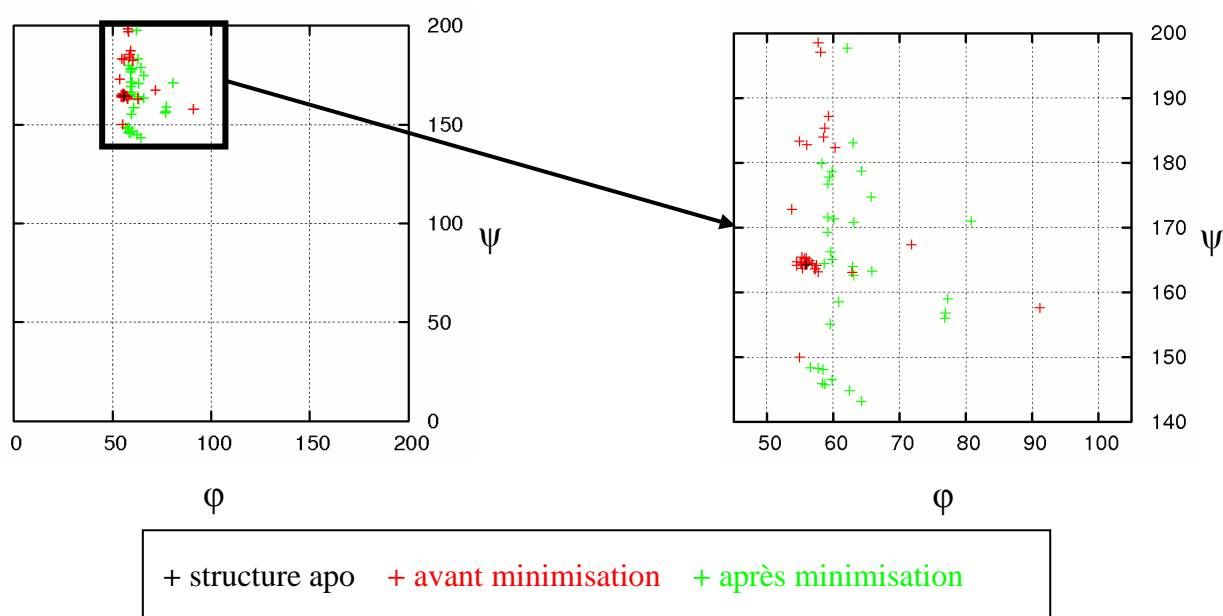


Figure 39 : diagramme des RMSDT dans le plan (φ, ψ).

Cette approche nous permet d'explorer l'espace conformationnel de la molécule, en appliquant une action bien spécifique sur cette dernière. De plus, nous pouvons voir sur ces figures que pour certains sites, le RMSDT diminue à la fois pour φ et pour ψ , pour des structures minimisées, mais également avant minimisation. Cela signifie que lorsque le cation

CHAPITRE III : RESULTATS

Zn^{2+} s'accroche sur certains sites, la structure apo du peptide se déforme vers la structure liée.

Afin d'observer les changements conformationnels occasionnés, nous avons voulu déformer le peptide pas-à-pas, c'est-à-dire en appliquant un coefficient α devant le vecteur force. L'introduction de ce coefficient est une manière implicite de tenir compte de la présence du solvant qui, à cause de sa permittivité élevée, écranter les interactions électrostatiques. Précédemment, pour réaliser les diagrammes de la figure 36, nous avons :

$$\Delta x = [M].\vec{F}$$

Où Δx représente le déplacement atomique, \vec{F} le vecteur force et $[M]$ la matrice des déformations.

Maintenant, nous appliquons :

$$\Delta x = \alpha[M].\vec{F}$$

Où l'on fait varier α de 0 à 2.

Pour chaque valeur de α nous obtenons alors une conformation déformée du peptide. Chacune de ces structures est alors optimisée par 1500 étapes de descente de gradient suivies de 1500 étapes de gradients conjugués, et nous reportons alors les valeurs de l'énergie sur un graphique. Ces graphiques représentent donc le profil énergétique suivie par la molécule lorsqu'on applique une force de plus en plus élevée. On recueille alors des informations sur la structure du peptide.

Nous avons sélectionné trois cas de figure, chacun représentant une situation particulière :

- La structure ne change pas lors de l'interaction avec le cation. La présence de l'ion déforme bien le peptide, mais le site d'interaction est tel qu'on ne peut pas aboutir à une déformation permanente. Lorsque la structure de l'ensemble est optimisée, la molécule retrouve sa conformation initiale.
- Un changement conformationnel est mis en évidence par une variation soudaine de l'énergie. Les forces d'excitation sont capables d'engendrer un changement de conformation. Le peptide ne retrouve plus sa structure initiale après l'optimisation, et son énergie interne est modifiée.

- Une variation progressive de l'énergie correspondant au repliement du peptide au fur et à mesure que l'on augmente la force

Nous allons voir en détail ces différentes situations en nous appuyant sur les figures 40, 41 et 42 ci-dessous.

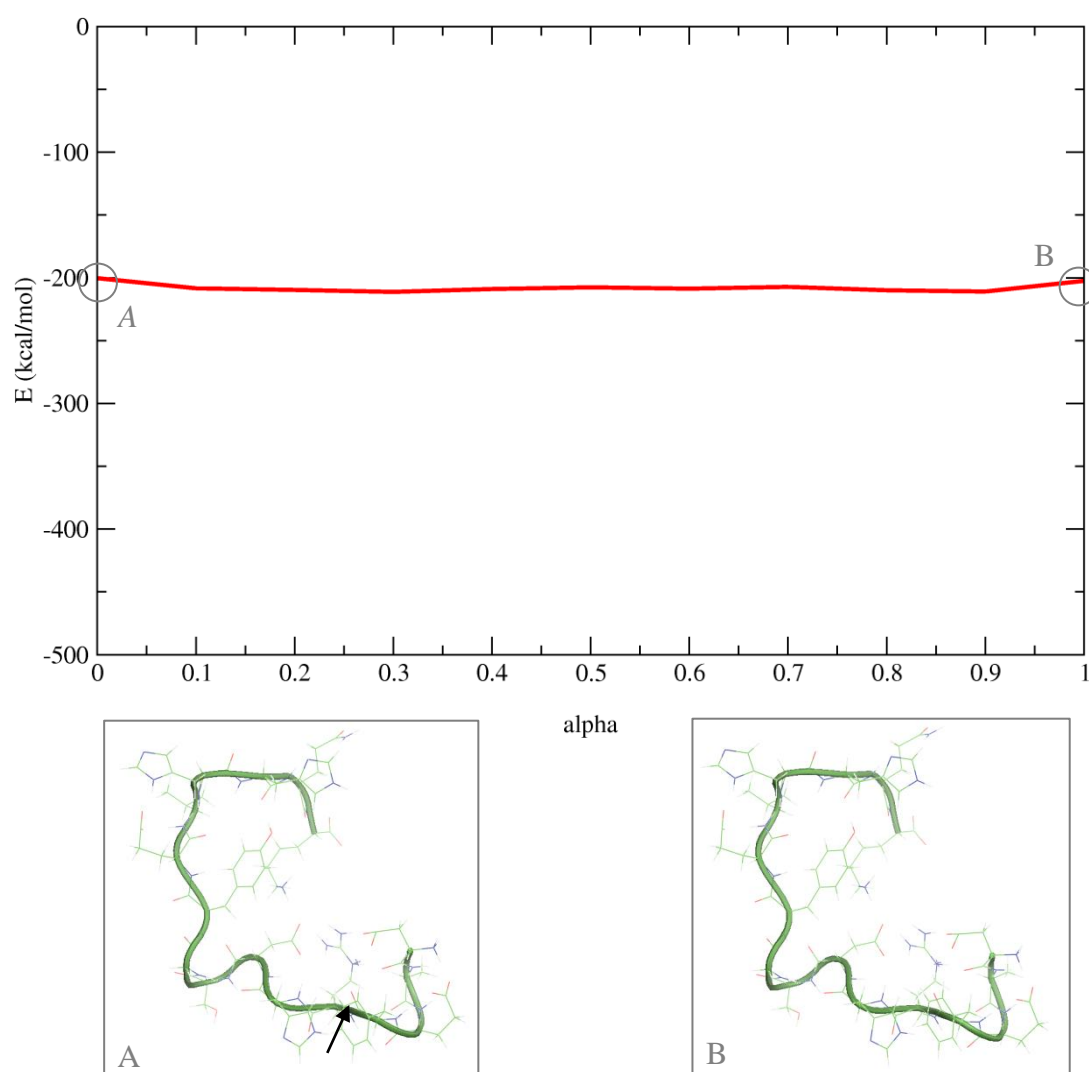


Figure 40 : Evolution de l'énergie du peptide $A\beta(1-16)$ en fonction de α lorsque l'ion s'accroche sur le site indiqué par la flèche.

CHAPITRE III : RESULTATS

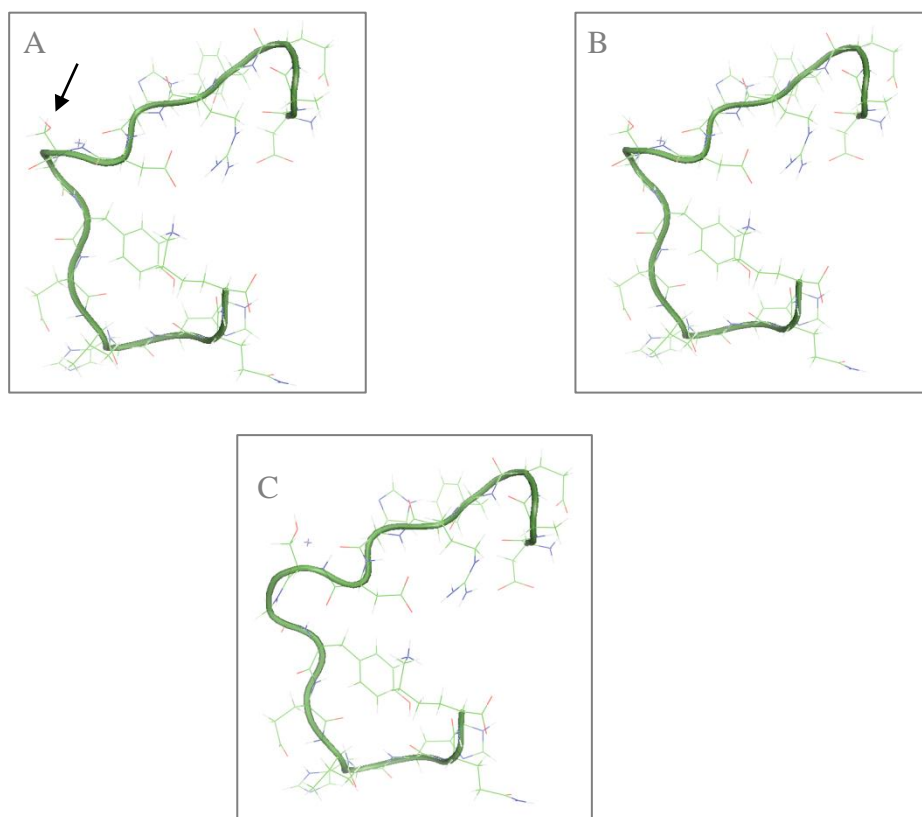
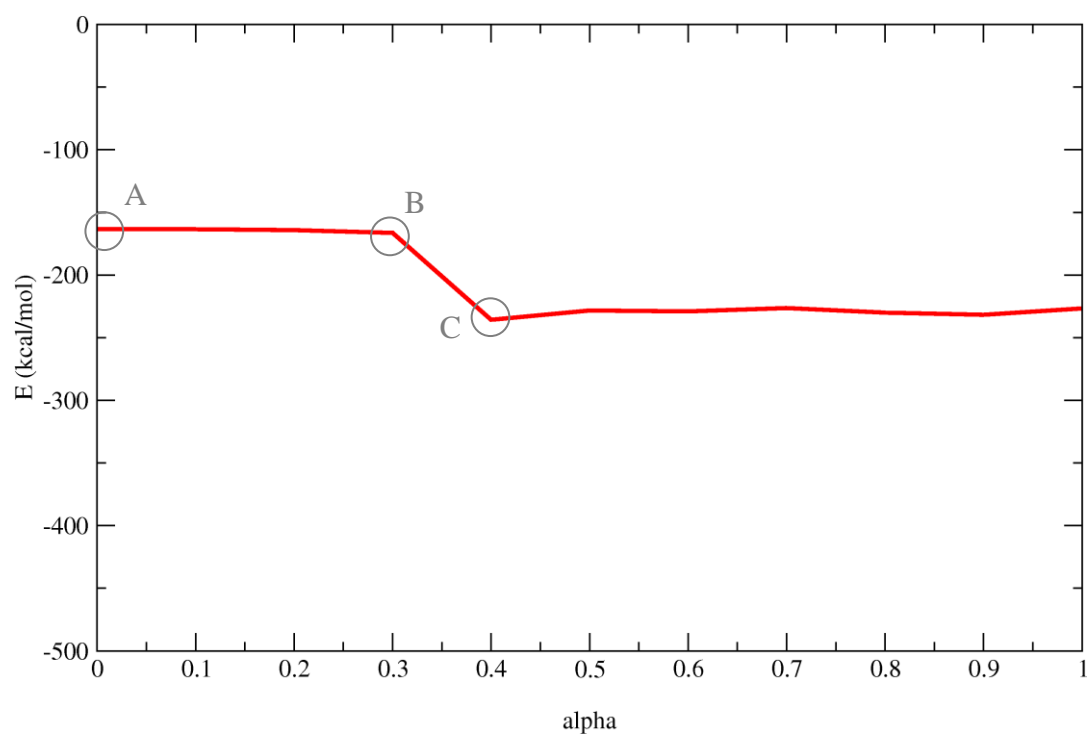


Figure 41 : Evolution de l'énergie du peptide $A\beta(1-16)$ en fonction de α lorsque l'ion s'accroche sur le site indiqué par la flèche.

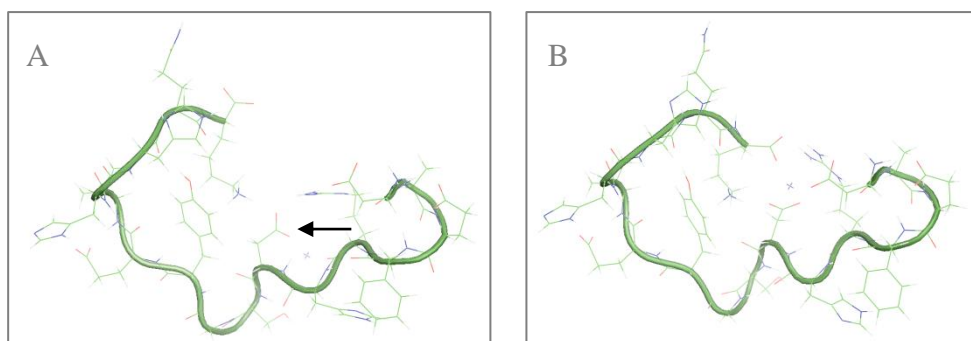
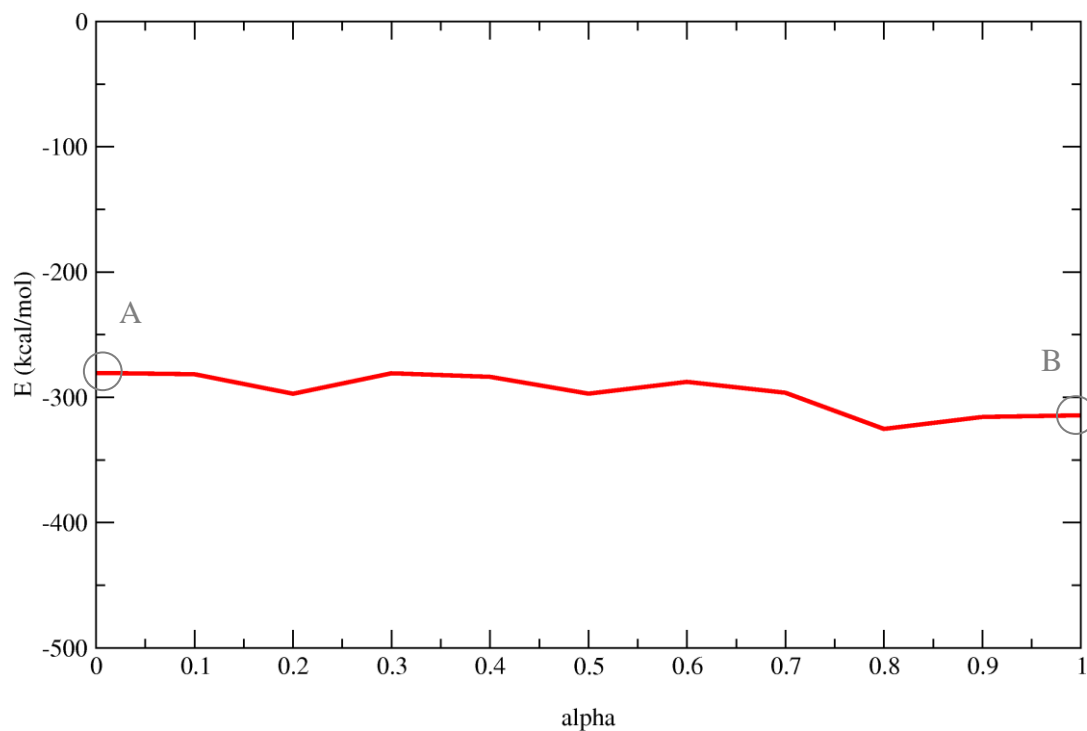


Figure 42 : Evolution de l'énergie du peptide $A\beta(1-16)$ en fonction de α lorsque l'ion s'accroche sur le site indiqué par la flèche.

Sur la figure 40, nous pouvons voir que l'énergie totale du peptide reste quasi-constante lorsque l'on augmente la force imposée par l'ion Zn^{2+} . Cela signifie, après examen des structures sur cette même figure 40, que la conformation ne change pas en dépit de la force exercée.

En revanche, on peut voir sur la figure 41 que l'énergie reste constante jusqu'à une certaine valeur de α , à partir de laquelle l'énergie de la molécule diminue. Après minimisation, on observe un changement conformationnel, comme le suggère la structure C.

Nous venons donc de mettre en évidence des changements conformationnels « brutaux » qui se traduisent par des marches sur le profil énergétique de la molécule, tel qu'il est représenté

CHAPITRE III : RESULTATS

sur les graphiques ci-dessus. Sur la figure 42, nous constatons que l'énergie diminue au fur et à mesure que l'on augmente la force générée par l'ion, sans montrer de palier comme nous l'observons sur la figure 41. D'un point de vue conformationnel, on voit tout d'abord que l'ion s'accroche sur un site situé au cœur de la molécule. En augmentant alors la force due à ce dernier, on attire petit-à-petit les extrémités de la molécule qui sont électronégatives : nous avons mis en évidence le repliement de la molécule lorsque l'on place l'ion au cœur du peptide.

III-3.2.2. Repliement du peptide à partir de la structure sans zinc

Nous avons vu que l'on pouvait explorer les changements conformationnels du peptide A β (1-16), sous l'action d'une force spécifique, grâce aux modes statiques. Dans cette partie, nous allons voir qu'il est possible d'utiliser cette même sonde électrostatique pour suivre le repliement du peptide seul.

Comme précédemment, nous déformons le peptide sous l'action du zinc en relation avec les interactions potentielles sur les 35 sites de la molécule. Nous obtenons ainsi, après minimisation, 35 molécules déformées, et nous retenons celle qui possède la structure la plus proche de la structure finale. Pour cela, nous calculons des RMSD pour chacune des molécules déformées et retenons les valeurs les plus faibles. Ensuite, nous recalculons les modes statiques de cette dernière, et répétons l'opération jusqu'à atteindre une conformation la plus proche possible de la structure liée.

Afin de choisir la structure la plus proche de la structure finale, nous avons fait le choix de comparer les valeurs du RMSD calculé sur les positions des atomes du squelette et les valeurs du RMSDT calculé sur les angles dièdres χ qui rendent compte de l'orientation des chaînes latérales du peptide. Ainsi, nous avons des informations sur le mouvement du squelette et des chaînes latérales lors du passage de ;'ion métallique.

Nous avons simulé une centaine de passage de l'ion, soit autant de calculs de modes statiques, en générant à chaque fois 35 conformations. Cela représente quelques 3000 structures dont la valeur de l'énergie est représentée sur le diagramme de la figure 43, dans le plan défini par les RMSD et RMSDT cités ci-dessus.

La figure 43 montre qu'en partant de la structure sans zinc, nous avons obtenu après plusieurs passages de Zn²⁺ une conformation proche de celle observée expérimentalement. En termes

de RMSD, la valeur la plus petite atteinte pour les atomes du squelette est de 1.24 Angström, ce qui est très raisonnable. En revanche, nous voyons que les minima d'énergie correspondent à des valeurs de RMSD bien plus grandes (aux alentours de 3.7 Angströms). Le minimum global de cette figure est représenté sur la figure 44 et possède une énergie de -124.2 kcal/mol. Avec la sonde électrostatique, nous trouvons donc un minimum global d'énergie loin de la structure expérimentale.

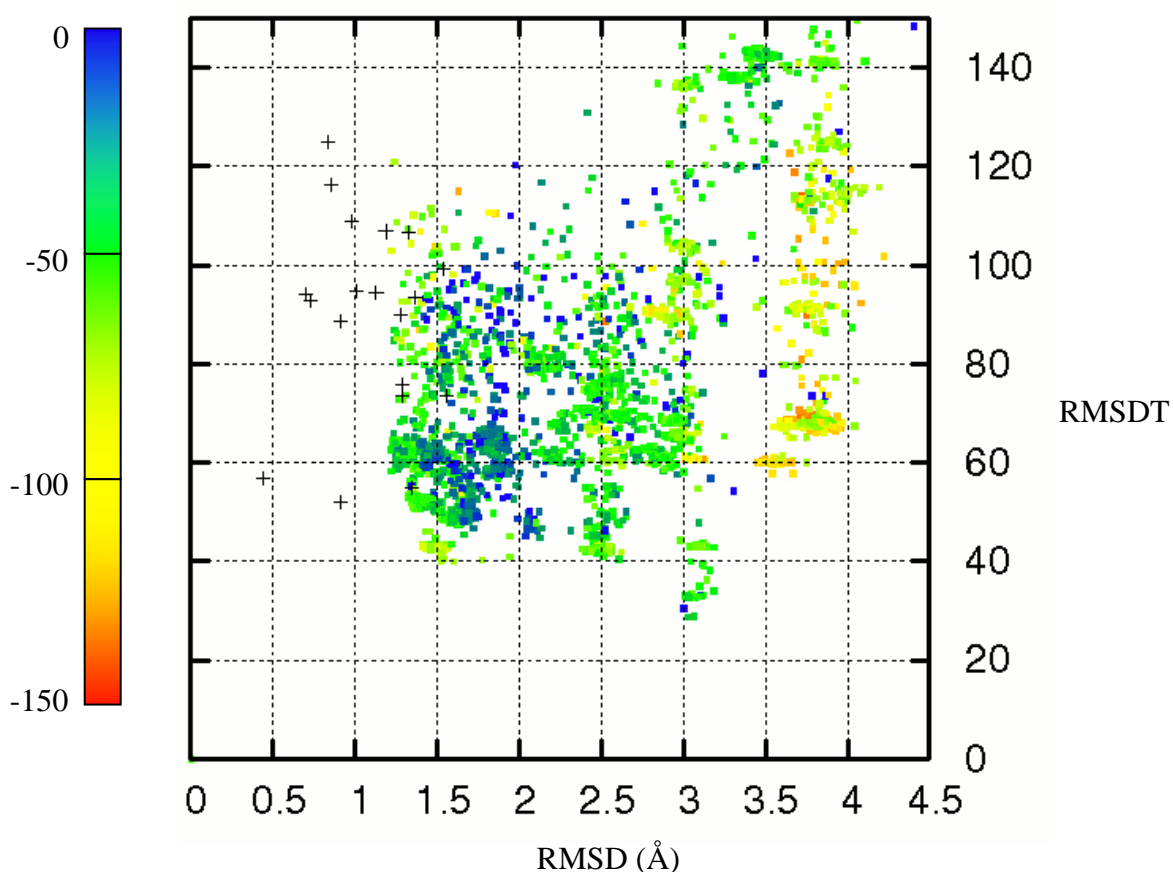


Figure 43 : diagramme représentant l'énergie et l'évolution de la structure apo vers la structure liée. Les croix représentent les 19 structures issues du fichier PDB.

La structure de référence correspondant à l'origine du diagramme de la figure 43, est issue du fichier PDB 1ZE9. Or, ce dernier contient 20 structures expérimentales différentes du peptide, dont la partie contenant les résidus liés au zinc ne change pas (ou très peu), mais dont les extrémités ont des structures différentes. Nous avons donc reporté sur la figure 43 les points correspondant à ces différentes conformations du peptide dans sa forme liée. Nous pouvons alors observer que certains de ces points se rapprochent, voire touchent le nuage de

CHAPITRE III : RESULTATS

points générés sous l'action de Zn^{2+} .

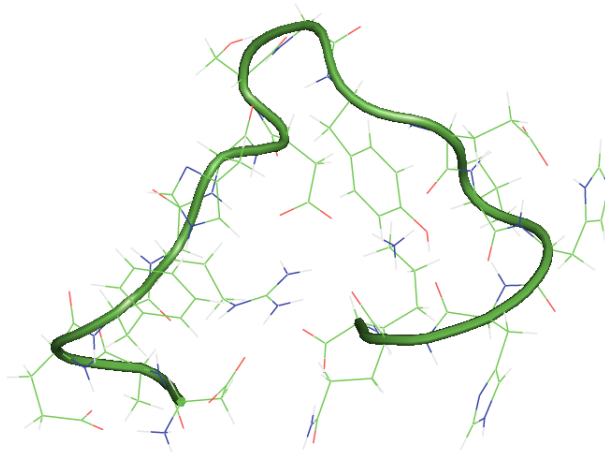


Figure 44 : Structure correspondant au minimum global d'énergie calculée avec la sonde électrostatique.

Sur la figure 45, nous représentons une partie du diagramme de la figure 43 (RMSD compris entre 1 et 2 Å et RMSDT compris entre 40 et 120 degrés). Nous pouvons y voir plus précisément la coïncidence des points expérimentaux avec ceux générés par la sonde. Nous avons également représenté les énergies correspondant à ces points expérimentaux.

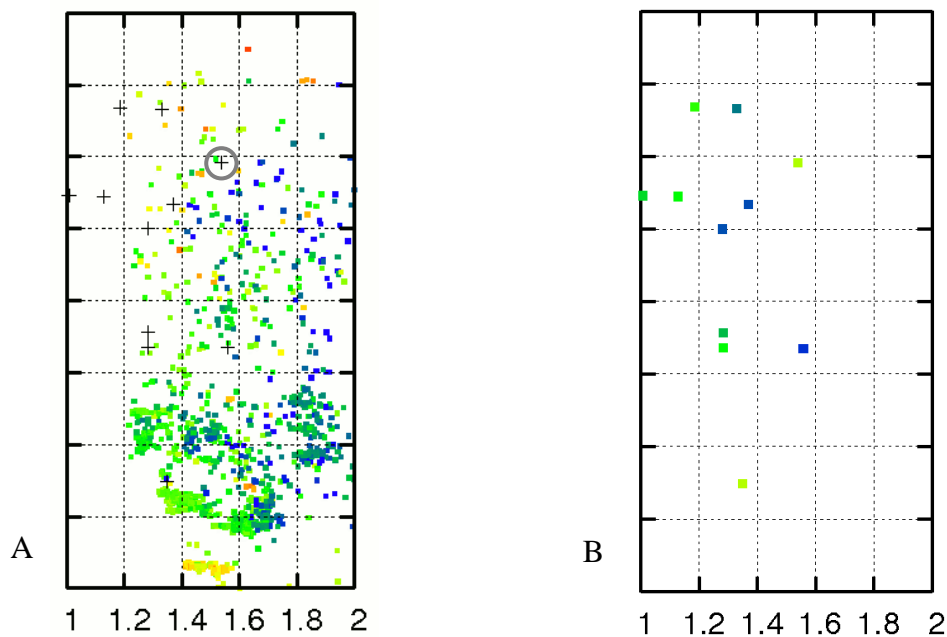


Figure 45 : Agrandissement d'une partie du diagramme de la figure 40. A) Les croix représentent les points correspondant aux structures expérimentales. B) Valeurs des énergies de ces structures expérimentales.

Pour les calculs de RMSD, nous avons choisi une structure de référence parmi les 20 fournies par le fichier PDB. Nous voyons sur la figure 43 que les points expérimentaux sont dispersés et parfois assez éloignés de la référence. De plus, nous constatons que le nuage formé par ces points et celui généré par la sonde électrostatique coïncident. Nous retrouvons donc une dispersion similaire entre l'expérience et notre simulation. Cela vient en outre justifier l'arrêt de notre simulation de sonde électrostatique.

Nous voyons sur la figure 45 que des points générés par la sonde électrostatique sont très proches de certains points expérimentaux. Or, le fait que deux points soient situés à proximité ne signifie pas forcément que les structures correspondantes sont identiques. Nous avons tout de même comparé les conformations expérimentales avec celles obtenues par le calcul possédant des RMSD proches. Nous trouvons alors que les deux structures dont les RMSD correspondent aux points contenus dans le cercle représenté sur la figure 45A) sont très proches. Nous avons représenté sur la figure 46 la superposition de ces deux structures. Le RMSD calculé sur les positions des carbones alpha a une valeur de 2.11 Angströms.

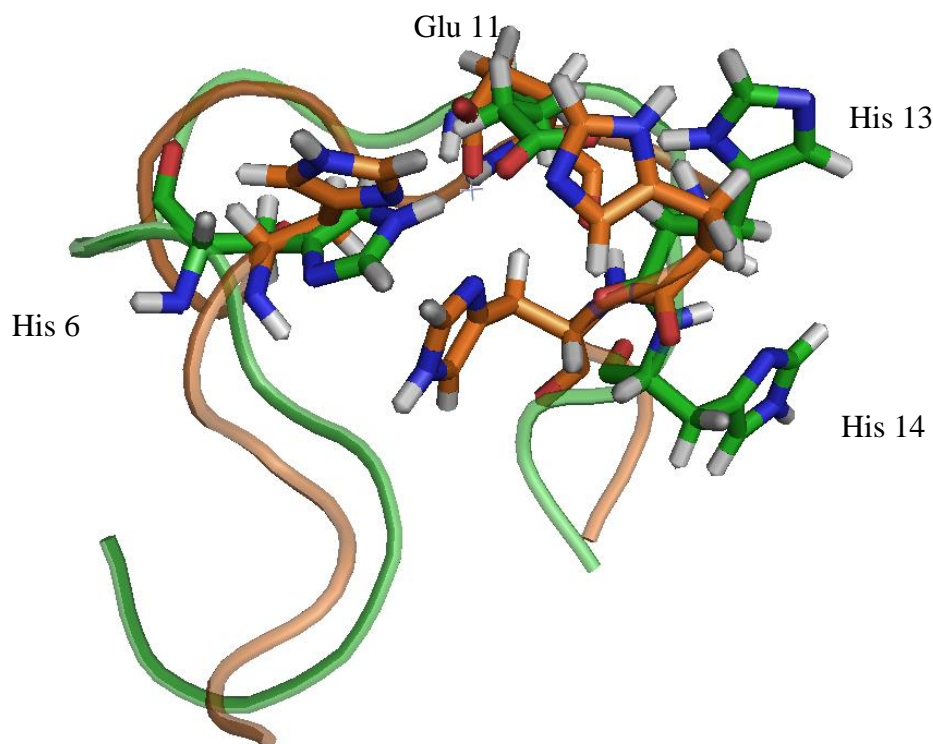


Figure 46 : superposition d'une structure issue de l'expérience (orange) et d'une générée par la sonde électrostatique (vert). Seuls les résidus 6, 11, 13 et 14 sont représentés.

CHAPITRE III : RESULTATS

Sur la figure 46, seules les chaînes latérales des résidus composant le site de fixation du Zinc sont représentées. Nous constatons que les histidines 6, acides glutamiques 11 et histidines 13 se superposent, confirmant ainsi les observations faites sur le diagramme de la figure 45. En revanche, les histidines 14 ne se superposent pas. Cela est dû à la différence de forme du squelette des deux structures au niveau de ce résidu. L'angle χ_{14} de la conformation expérimentale et celui de la conformation générée par la sonde ont donc des valeurs proches, mais la différence de forme entre les squelettes au niveau de l'histidine 14 fait que les deux chaînes latérales ne se superposent pas. Si l'on agit sur cette partie du peptide, afin de plier le squelette dans une forme se rapprochant de celle issue de l'expérience, nous obtiendrons une superposition des chaînes latérales du site d'accrochage.

A travers cette application des modes statiques, que nous avons qualifiée de sonde électrostatique, nous avons validé deux points importants dans l'usage que nous souhaitons faire des modes statiques. Tout d'abord, l'algorithme d'arrimage pas-à-pas de l'ion sur le peptide s'est montré efficace. Ce programme est le premier pas vers une utilisation des modes statiques permettant d'aborder des changements conformationnels de grande amplitude relatifs à des interactions spécifiques. Dans une procédure de docking classique, on peut ainsi imaginer les déformations à chaque pas de translation/rotation. Ensuite, nous avons montré l'efficacité des modes statiques pour balayer l'espace conformationnel d'une molécule, en mettant en évidence l'évolution de la structure du peptide sous l'impulsion d'une interaction spécifique, qui est dans notre cas l'influence d'un ion métallique.

Le paysage énergétique que l'on peut voir sur la figure 43 montre que grâce à la sonde électrostatique, nous avons pu replier le peptide A β (1-16) sans zinc vers une conformation proche d'une structure avec zinc issue de l'expérience. En effet, le nuage de points défini par les croix sur la figure 43 coïncide avec celui généré par la sonde électrostatique. Il est à noter de plus qu'à chaque étape, la sélection de la structure la plus proche de la structure finale est effectuée par l'utilisateur, ce qui ralentit grandement la procédure. Ce programme peut donc être optimisé en l'automatisant, ce qui ferait gagner un temps précieux et le rendrait alors encore plus performant.

III-4. Conclusions

En considérant un peptide modèle simple composé de sept acides aminés, nous avons pu voir quelques applications directes des modes statiques pour caractériser les fluctuations atomiques. Ces premiers calculs nous ont permis d'émettre des hypothèses sur les premières étapes du mécanisme d'agrégation. Nous nous sommes ensuite penchés sur le peptide Amyloïde β , et nous avons mis au point un système de représentation des zones flexibles de la molécule. Nous avons vu comment ces diagrammes permettaient de différencier les structures en feuillet des structures en hélice. Mais également, nous avons retrouvé au travers de ces représentations un certain nombre de résultats obtenus précédemment par le calcul ou l'expérience.

Nous avons ensuite traité l'interaction entre l'Amyloïde β et un ion métallique. Pour cela, nous avons imaginé la sonde électrostatique, qui utilise le même algorithme de base que celui du docking moléculaire. Les résultats obtenus démontrent la puissance des modes statiques pour leur utilisation dans une procédure « classique » de docking : en appliquant une force précise sur une molécule, nous la déformons et pouvons explorer ainsi l'espace conformationnel de cette dernière. Cette approche peut s'apparenter à la « steered dynamics », dans le sens où l'on applique une contrainte sur un domaine de la molécule qui engendre (ou pas) un changement conformationnel. Mais dans notre approche, le temps n'entre pas en jeu, les déformations étant données directement par la matrice des modes statiques. On s'affranchit alors du temps de calcul nécessaire pour la résolution des équations de Newton dépendantes du temps.

CONCLUSION GENERALE

Durant ces dernières décennies, nous avons assisté à une progression colossale en génétique et biologie moléculaire. La compréhension et l'interprétation des processus fondamentaux de la vie ont ouvert la porte à d'innombrables applications dans le domaine de la santé, et participent à l'amélioration de notre quotidien.

D'un autre côté, les techniques de fabrication dans le domaine des micro-nano-technologies permettent de travailler à des échelles de taille se rapprochant de la taille de macromolécules biologiques que nous savons manipuler aisément. On envisage donc de faire des dispositifs associant le monde électronique et le monde biologique. Des applications existent déjà dans ce domaine : les biopuces sont des micro-dispositifs de quelques millimètres carrés, capables d'effectuer de multiples analyses biologiques.

Si l'on se place du point de vue des applications, qu'elles soient médicales ou en biotechnologie, il faut trouver rapidement de nouvelles avancées. Pour faire vite, il est nécessaire de développer des outils pour avancer virtuellement, par prototypage informatique. Cela passe par la modélisation des mécanismes élémentaires qui régissent l'activité des biomolécules, et donc par une description précise des interactions intermoléculaires.

Il n'existe pas aujourd'hui de modèles satisfaisant tous les critères pour décrire correctement ces interactions. Le point noir des outils de simulation est de prendre en compte les fluctuations atomiques des biomolécules : les modèles dynamiques précis ne sont pas capables de décrire ces mouvements sur une durée pertinente, et les modèles simplifiés ne sont pas adaptés pour décrire tous les systèmes.

Nous avons proposé dans cette thèse une approche nouvelle pour décrire ces fluctuations et prendre pleinement en compte la flexibilité des macromolécules dans les modèles de simulation. La méthode des modes statiques est un outil très utile pour la description des déformations moléculaires causées par une force d'interaction externe. Il est donc envisageable d'utiliser cette approche originale dans un modèle de simulation de docking moléculaire.

C'est un défi difficile, qui a soulevé un certain nombre de problèmes. Le principal souci étant le nombre d'atomes, la description de la molécule entière devient alors compliquée et longue en termes de temps de calcul. Malgré cela, nous avons vu au travers de l'exemple « simple »

de la sonde électrostatique, que les modes statiques rendaient compte des déformations occasionnées par une interaction spécifique. De plus, il est possible d'incorporer la méthode dans un schéma « classique » de docking : on utilise les modes statiques pour générer un grand nombre de conformations, que l'on teste à l'aide de fonctions de scoring.

Cette approche originale pour rendre compte de la flexibilité des macromolécules a donc été présentée dans cette thèse.

Le premier chapitre a introduit le contexte dans lequel se sont déroulés ces travaux. En effet, nous avons pu voir l'importance de la simulation pour l'analyse, la compréhension et l'interprétation de ces phénomènes biologiques. Dans un domaine à la frontière entre biologie, chimie et physique, il est primordial de décrire les interactions biomoléculaires de manière précise, afin de développer des outils prédictifs puissants.

Ensuite, nous avons pu découvrir dans le second chapitre, le concept et les idées qui ont mené au développement des modes statiques, et en quoi l'approche était originale. L'idée est de développer une méthode capable de donner les déformations totales d'une molécule, sans passer par des états de transition intermédiaires. Ce qui en fait une approche, est que nous partons du principe que les déformations sont causées par une force externe. Cela place la méthode comme candidate idéale pour l'intégrer dans un programme de docking. Dans ce chapitre, nous avons également pu voir que les modes statiques ont été validés par les travaux de thèse réalisés par Marie Brut.

Enfin, le troisième chapitre montre comment utiliser les modes statiques pour rendre compte des déformations d'une protéine. Nous avons pu voir un moyen original de décrire globalement les déformations d'une molécule, et découvrir les zones d'intérêt où il est alors possible d'agir. Une autre application originale des modes statiques est la sonde électrostatique, qui a montré des possibilités d'exploration de l'espace conformationnel. De plus, cet exemple constitue le premier programme de docking s'appuyant sur les modes statiques.

C'est justement cette application de la sonde électrostatique qui a également révélé les limites du modèle. Le principal problème vient du mode de représentation du système. Les modes devant être recalculés à chaque fois que la molécule est déformée, il devient très long d'effectuer du docking dans un système décrit en coordonnées cartésiennes. Le développement de la prochaine génération de FLEXIBLE se fera dans un système de

CONCLUSION GENERALE

coordonnées simplifié, où l'on considèrera uniquement les angles de torsion comme variables : ce sont les modes généralisés. De plus, ce moyen de représentation devra permettre d'établir un modèle multi-harmonique des modes statiques, et l'on pourra alors, grâce à un calcul unique pour une molécule, générer un très grand nombre de conformations.

Cette nouvelle version des modes généralisés devrait voir le jour dans quelques mois, et elle devra alors être validée de la même manière que la première version l'a été. Elle permettra alors le développement d'outils de simulation puissants pour l'interprétation, l'analyse, mais surtout la prédiction de processus biologiques, pour des applications médicales, mais également de processus entre monde biologique et non-biologique pour le développement de nouvelles technologies bio-inspirées.

ANNEXE A

LE LOGICIEL FLEXIBLE

Dans cette partie, nous allons détailler les différentes étapes du logiciel FLEXIBLE, capable de déterminer et stocker les modes statiques d'une molécule.

1- Calcul préalable

Afin de calculer les modes statiques, le logiciel FLEXIBLE a besoin de deux fichiers d'entrée : l'un contient les coordonnées cartésiennes des positions atomiques, l'autre est la matrice hessienne des dérivées secondes. Ces derniers sont issus d'un calcul d'énergie totale de la molécule, effectué à l'aide d'un logiciel adéquat. Dans notre cas, nous avons utilisé le logiciel GAUSSIAN03 pour les calculs quantiques, et AMBER8 pour les calculs classiques. Contrairement à AMBER8, GAUSSIAN03 ne permet pas d'obtenir directement la matrice hessienne des dérivées secondes. Nous sommes obligés de la calculer à partir des vecteurs propres et de valeurs propres issues du calcul des modes normaux de vibration.

2- Choix des atomes fixes

Une fois le calcul énergétique préalable effectué, le programme FLEXIBLE peut alors être lancé avec comme entrée les fichiers de coordonnées et des dérivées secondes de l'énergie. Afin de fixer les six degrés de liberté liés à la translation et à la rotation de la molécule, l'utilisateur est tenu de choisir trois atomes à fixer en translation ($n_i(0,0,0)$), en azimuth ($n_j(x,0,0)$) et en longitude ($n_k(x,y,0)$). Un changement de base est alors effectué pour se placer dans le repère défini par les atomes fixes. Pour cela, nous considérons les vecteurs $\vec{\mu}_1$ et $\vec{\mu}_2$ qui relient respectivement les points n_i , n_j et n_i , n_k . Ces vecteurs sont alors normés, puis nous calculons le vecteur $\vec{\mu}_3 = \vec{\mu}_1 \times \vec{\mu}_2$ afin d'obtenir la base orthonormée $(\vec{\mu}_1, \vec{\mu}_2, \vec{\mu}_3)$.

3- Sous-programme *RMAT*

ANNEXE A : LE LOGICIEL FLEXIBLE

Le premier sous-programme appelé est *RMAT* et permet :

- Avec *NewMat* :
 - Dans *RotateX* : de calculer les coordonnées atomiques dans le nouveau repère défini par les atomes fixes.
 - Dans *Rotate* : de calculer les éléments de la nouvelle matrice hessienne D dans la nouvelle base $U = (\vec{\mu}_1, \vec{\mu}_2, \vec{\mu}_3)$, en agissant bloc par bloc (sur des blocs de 3×3 éléments) avec la relation :

$$D \rightarrow U^{-1} D U$$

- Avec *Newline* : d'échanger les lignes et les colonnes dans la matrice D , afin de séparer les éléments fixés de ceux que nous allons déformer.

4- La fonction *LCMOD*

On peut ensuite choisir les modes que l'on désire calculer, grâce à la fonction *LCMODE*. Par défaut, le logiciel calcule tous les modes de déformation. Mais on peut envisager de n'en calculer que quelques uns. Sur une très grosse molécule, par exemple, il est possible de choisir de calculer les modes correspondants à des atomes situés à proximité du ou des site(s) d'interaction.

5- La fonction *LINMODE*

L'étape suivante est celle qui consomme le plus de temps CPU. Elle fait appel à la fonction *LINMODE* qui permet de résoudre le système $A.X + B = 0$ par la méthode de Gauss-Seidel. C'est une méthode itérative basée sur la construction d'une suite de vecteurs $x_k = (x_1^k \dots x_n^k)$ qui converge vers la solution du système. Nous calculons x_{k+1} à partir de x_k grâce à la relation :

$$x_i^{k+1} = - \left(\frac{B_i + \sum_{j \neq i} A_{ij} x_j^k}{A_{ii}} \right)$$

Puis nous optimisons la solution en réutilisant x^{k+1} à la place des x^k dès qu'ils sont calculés :

$$x_i^{k+1} = \frac{B_i + \sum_{j=1}^{i-1} A_{ij} x_j^{k+1} - \sum_{j=i+1}^n A_{ij} x_j^k}{A_{ii}}$$

Le calcul est arrêté lorsque la différence entre x_k et x_{k+1} est très faible. Cela se traduit par fixer une valeur ε et arrêter le calcul lorsque :

$$\max |x_i^k - x_i^{k+1}| < \varepsilon$$

Enfin, on stocke les déformations obtenues avant de calculer l'énergie correspondante par la relation :

$$E = \frac{1}{2} \sum_{i,j} A_{ij} x_i x_j$$

L'intégralité du programme et des sous-routines a été programmée en C. Il a également été nécessaire de mettre en place des bibliothèques annexes dans le but de traiter les fichiers d'entrée et de sortie. Elles ont servi notamment pour la lecture des fichiers d'entrée (coordonnées et matrice hessienne) ou pour l'écriture de fichiers de sortie (coordonnées au format XYZ, matrice des modes statiques...). Une bibliothèque « mathématique » a également été créée dans laquelle on retrouve des calculs qui reviennent régulièrement dans le logiciel, tels que le produit scalaire, le produit vectoriel, le calcul de la distance euclidienne, etc...

Le temps nécessaire pour le calcul des modes est évidemment proportionnel au nombre d'atomes du système étudié. Pour une molécule de moins de 100 atomes, cela requiert moins d'une seconde CPU sur un processeur Intel Xeon Dual Core 1.6 GHz. Lorsque l'on calcule les modes sur de plus grosses molécules, on atteint des temps de plusieurs heures. Par exemple, pour la protéase du HIV-1 composée de 3128 atomes, il faut 7 heures et 20 minutes pour effectuer l'opération. Une version parallélisée du code est aussi disponible, ce qui permet de réduire le temps effectif pour le calcul sur de gros systèmes.

ANNEXE B

MINIMISATION DU RMSD

Le calcul du RMSD tient une place importante en biologie moléculaire. En effet, cette grandeur sert à comparer deux structures d'une molécule et est souvent utilisée pour décrire l'espace conformationnel de la molécule. La formule générale du RMSD est :

$$RMSD = \left[\frac{1}{N} \sum_{i=1}^N (\vec{r}'_i - \vec{r}_i)^2 \right]^{\frac{1}{2}}$$

Où les \vec{r}_i et \vec{r}'_i représentent les positions atomiques de la molécule initiale et de la molécule cible, respectivement.

On cherche alors à superposer les deux structures de manière la plus précise possible, afin d'obtenir un RMSD minimal. Cela revient à aligner les jeux de vecteurs \vec{r}_i et \vec{r}'_i en effectuant une succession de translations et de rotations. Une méthode pour faire cela a été élaborée par le biophysicien Kabsch.

L'algorithme de Kabsch permet de déterminer la matrice de rotation optimale pour minimiser le RMSD. Cela suppose que la molécule que l'on veut aligner soit placée dans le même repère que celui de la molécule cible. Une manière simple de faire cela est de superposer les deux centres de masse. Nous avons alors les deux jeux de vecteurs des positions atomiques dans le même repère, et la différence entre ces derniers peut être calculée par l'expression :

$$E = \sum_{i=1}^N (\vec{r}'_i - \vec{r}_i)^2$$

On retrouve alors le RMSD en effectuant l'opération simple :

$$RMSD = \sqrt{\frac{E}{N}}$$

On cherche alors à minimiser l'expression E . Pour cela, on applique une matrice de rotation U au jeu de vecteurs \vec{r}'_i . L'expression de E s'écrit dans ce cas :

$$E = \sum_{i=1}^N (\overrightarrow{Ur_i} - \vec{r}_i)^2$$

La résolution effectuée initialement par Kabsch nécessite l'introduction de facteurs de Lagrange et est un peu fastidieuse. Nous allons utiliser une démonstration plus simple faisant appel à l'algèbre linéaire et correspondant à l'algorithme utilisé dans nos programmes.

Dans un premier temps, on développe E , qui est fonction de U :

$$E = \sum_{i=1}^N (|x_i|^2 + |y_i|^2) - 2 \sum_{i=1}^N x_i U y_i = E_0 - 2L$$

La partie E_0 est invariante par rapport à une rotation U . La variation se situe dans le terme L . Le but de minimiser le RMSD est de trouver la matrice U_{min} qui donne une valeur maximale L_{max} de L , avec laquelle on obtient :

$$RMSD = \sqrt{\frac{E_{min}}{N}} = \sqrt{\frac{E_0 - 2L_{max}}{N}}$$

L'astuce consiste ensuite à fabriquer les matrices $(N \times 3)$ X et Y à partir des jeux de vecteurs x_i et y_i , respectivement. Nous pouvons alors écrire L comme une trace :

$$L = Tr(X^t U Y)$$

Où X^t est une matrice $(3 \times N)$ et Y $(N \times 3)$. En jonglant avec les matrices au sein de l'expression L , on obtient :

$$L = Tr(X^t U Y) = Tr(U Y X^t) = Tr(U R)$$

On ne sait pas quelle est la matrice U , mais on peut étudier la matrice $R = Y X^t$, qui est la matrice de corrélation entre X et Y . D'après le théorème de décomposition en valeurs simples, R peut s'écrire comme le produit :

$$R = V S W^t$$

On a alors les matrices (3×3) V , W (qui sont orthogonales) et S (qui est diagonale positive et composée des éléments σ_i). On insère alors ces matrices dans l'expression L :

$$L = Tr(U V S W^t) = Tr(S W^t U V)$$

Comme S est diagonale, si l'on groupe les matrices $W^t U V$ ensemble pour former la matrice T , on obtient une expression simple de la trace en fonction des termes σ_i :

ANNEXE B : MINIMISATION DU RMSD

$$L = Tr(ST) = \sigma_1 T_{11} + \sigma_2 T_{22} + \sigma_3 T_{33}$$

Or, d'après l'expression qui définit la matrice T , on voit que cette dernière est un produit de matrices diagonales (U , W^t et V). T doit être également diagonale. Dans ce cas, la plus grande valeur des éléments de T est 1, ou $T_{ij} < 1$. Maintenant L est un produit de termes linéaires des éléments diagonaux de T , et on a donc la valeur maximale de L pour $T_{ii} = 1$:

$$L_{max} = Tr(ST_{max}) = \sigma_1 + \sigma_2 + \sigma_3 = Tr(S)$$

En reprenant l'équation donnant le RMSD et en remplaçant L_{max} par l'expression ci-dessus, on obtient :

$$RMSD = \sqrt{\frac{E_0 - 2(\sigma_1 + \sigma_2 + \sigma_3)}{N}}$$

Dans sa démonstration, Kabsch suggère de résoudre la matrice $R^t R$, qui est réelle et symétrique donc plus simple à résoudre, afin d'obtenir les σ_i . En développant l'expression $R^t R$, on trouve que terme de valeur singulière de $R^t R$ est S^2 . En d'autres termes, les valeurs singulières de $R^t R$ (λ_i) sont liées aux valeurs singulières de R (σ_i). par la relation :

$$\lambda_i = \sigma_i^2$$

Comme $R^t R$ est symétrique, les valeurs propres sont identiques aux valeurs singulières λ_i . On peut alors utiliser l'algorithme standard de la transformée de Jacobi pour trouver les valeurs propres de $R^t R$, qui sont les mêmes que les λ_i . On obtient donc l'expression suivante pour le RMSD :

$$RMSD = \sqrt{\frac{E_0 - 2(\sqrt{\lambda_1} + \sqrt{\lambda_2} + \sqrt{\lambda_3})}{N}}$$

BIBLIOGRAPHIE

1. *Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction isolated from Pneumococcus Type III.* **Avery O T, MacLeod C M, MacCarthy M.** 2, 1944, J. Exp. Med., Vol. 79, pp. 137-158.
2. *Genetic Control in Biochemical Reactions in Neurospora.* **Beadle W, Tatum E L.** 1941, PNAS, Vol. 27, pp. 499-506.
3. *Independent Functions of Viral Protein and Nucleic Acid in Growth of Bacteriophage.* **Hershey A D, Chase M.** 1952, J. Gen. Physiol.
4. *Genetic Regulatory Mechanisms in the Synthesis of Proteins.* **Jacob F, Monod J.** 1961, J. Mol. Biol., Vol. 3, pp. 318-356.
5. *A Structure for Desoxyribose Nucleic Acid.* **Watson J D, Crick F H C.** 1953, Nature, Vol. 171, pp. 737-738.
6. **Schulz G E, Schirmer R H.** *Principles of Protein Structure.* s.l. : Springer-Verlag, 1978.
7. *Modeling the Alpha-Helix to Beta-Hairpin Transition Mechanism and the Formation of Oligomeric Aggregates of the Fibrillogenic Peptide Abeta(1-16): Insights from All-Atom Molecular Dynamics Simulations.* **Simona F, Tiana G, Broglia R A, Colombo G.** 2004, J. Mol. Graph. Mod., Vol. 23, pp. 263-273.
8. *Thermodynamics and Kinetics of Aggregation for the GNNQQNY Peptide.* **Strodel B, Whittleston C S, Wales D J.** 2007, JACS, Vol. 129, pp. 16005-16014.
9. *Interpreting the Aggregation Kinetics of Amyloid Peptides.* **Pellarin R, Caflisch A.** 2006, J. Mol. Biol., Vol. 360, pp. 882-892.
10. *Studies of Folding and Misfolding using Simplified Models.* **Dokholyan, N V.** 2006, Curr. Op. Struct. Biol., Vol. 16, pp. 79-85.
11. *The Yin and Yang of Protein Folding.* **Jahn T R, Radford S E.** 2005, FEBS J., Vol. 272, pp. 5962-5970.
12. *The Role of Dynamic Conformational Ensembles in Biomolecular Recognition.* **Boehr D D, Nussinov R, Wright P E.** 2009, Nat. Chem. Biol., Vol. 5, pp. 789-796.
13. *Method for Including the Dynamic Fluctuations of a Protein in a Computer-Aided Drug Design.* **Carlson H A, Mordasini T, McCammon J A.** 49, 1999, J. Phys. Chem. A, Vol. 103, pp. 10213-10219.
14. *Accommodating Protein Flexibility in Computational Drug Design.* **Carlson H A, McCammon J A.** 2,

BIBLIOGRAPHIE

2000, Mol. Pharmacol., Vol. 57, pp. 213-218.

15. *Automated Docking using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function.* **Morris G M, Goodsell D M, Halliday R S, Huey R, Hart W E, Belew R K, Olson A J.** 1998, J. Comp. Chem., Vol. 19, pp. 1639-1662.

16. *A Fast Flexible Docking Method Using an Incremental Construction Algorithm.* **Rarey M, Kamer B, Lengauer T, Klebe G.** 1996, J. Mol. Biol., Vol. 261, pp. 470-489.

17. *Critical Evaluation of Search Algorithm for Automated Molecular Docking and Database Screening.* **Erwing T J A, Kuntz I D.** 1997, J. Comp. Chem., Vol. 18, pp. 1175-1189.

18. *GLIDE: A New Approach for Rapid, Accurate Docking and Scoring 1. Method and Assessment of Docking Accuracy.* **Friesner R A, Banks J L, Murphy R B, Halgren T A, Klicic J J, Mainz D T, Repasky M P, Knoll E H, Shelley M, Perry J K, Shaw D E, Francis P, Shenkin P S.** 7, 2004, J. Med. Chem., Vol. 47, pp. 1739-1749.

19. *Complementarity in Structure Ensembles in Protein-Protein Docking.* **Grunberg R, Leckner J, Nigels M.** 12, 2004, Structure, Vol. 12, pp. 2125-2136.

20. *The Static Modes: An Alternative Approach for the Treatment of Macro- and Bio-Molecular Induced-fit Flexibility.* **Brut M, Esteve A, Landa G, Renvez G, Djafari Rouhani M.** 2009, Eur. Phys. J. E, Vol. 28, pp. 17-25.

21. *Managing Protein Flexibility in Docking and its Application.* **B-Rao C, Subramanian J, Sharma S D.** 2009, Drug Discovery Today, Vol. 14, pp. 394-400.

22. *Toward Accurate Screening in Computer-Aided Enzyme Design.* **Roca M, Vardi-Kilshtain A, Warshel A.** 14, 2009, Biochemistry, Vol. 48, pp. 3046-3056.

23. *Molecular Biology or Ultrastructural Biology?* **Astbury, W.** 1961, Nature, Vol. 190, p. 1124.

24. **Madsen U, Krosgaard-Larsen P, Liljefors T.** *Textbook of Drug Design and Discovery.* Washington, DC : Taylor & Francis, 2002.

25. *Protein-Ligand Docking: Current Status and Future Challenges.* **Sousa S F, Fernandes P A, Ramos M J.** 2006, PROTEINS: Structure, Functions and Bioinformatics, Vol. 65, pp. 15-26.

26. **Crueger W, Crueger A, Brock T D.** *Biotechnology : A Textbook of Industrial Microbiology.* s.l. : Sinauer Associates Inc, 1990.

27. *Self-Assembly of a Nanoscale DNA Box with a Controllable Lid.* **Andersen E S, Dong M, Nielsen M M, Jahn K, Subramani R, Mamdouh W, Golas M M, Sander B, Stark H, Oliveira C L P, Pedersen J S, Birkedal V, Besenbacher F, Gothelf K V, Kjems J.** 7243, 2009, Nature, Vol. 459, pp. 73-U75.

28. *Aptamers-Based assays for Diagnostics, Environmental and Food Analysis*. **Tombelli S, Minnuni M, Mascini M.** 2, 2007, *Biomol. Engin.*, Vol. 24, pp. 191-200.
29. *Microfluidic Diagnostic Technologies for Global Public Health*. **Yager P, Edwards T, Fu E, Helton K, Tarn M R, Weigl B H.** 7101, 2006, *Nature*, Vol. 442, pp. 412-418.
30. *Nanofabrication by DNA Self-Assembly*. **Li H, Carter J D, LaBean T H.** 5, 2009, *Materials Today*, Vol. 12.
31. *Molecular Access to Structurally Switchable 3D Discrete DNA Assembly*. **Aldaye F A, Sleiman H F.** 44, 2007, *JACS*, Vol. 129, pp. 13376-13377.
32. *Nanoparticles, Proteins and Nucleic Acids: Biotechnology Meets Materials Science*. **Niemeyer, C.** 2001, *Angew. Chem. Int. Ed.*, Vol. 40, pp. 4128-4158.
33. *Folding DNA Origami from a Double-Stranded Source of Scaffold*. **Hogberg B, Liedl T, Shih W M.** 26, 2009, *J. Am. Chem. Soc.*, Vol. 131, pp. 9154-+.
34. *Stereochemistry of Polypeptide Chain Configurations*. **Ramachandran N G, Ramakrishnan C, Sasisekharan V.** 1968, *Adv. Prot. Chem.*, Vol. 23, pp. 283-437.
35. *The Structure of Proteins : Two Hydrogen Bonded Helical Configurations of the Polypeptide Chain*. **Pauling L, Corey R B, Branson H R.** 1951, *PNAS*, Vol. 37, pp. 205-211.
36. **Petsko G A, Ringe D.** *Protein Structure and Function*. s.l. : Blackwell Publishing Ltd, 2004.
37. **Gershel, A.** *Liaison intermoléculaires*. [ed.] InterEditions/CNRS Editions. s.l. : Savoirs Actuels, 1995.
38. *Conformational Changes Associated with Protein-Protein Interactions*. **Goh C S, Milburn D, Gerstein M.** 2004, *Curr. Op. Struct. Biol.*, Vol. 14, pp. 104-109.
39. *Protein Dynamism and Evolvability*. **Tokuriki N, Tawfik D S.** 203, 2009, *Science*, Vol. 324, pp. 203-207.
40. *The Arrangement of Electrons in Atoms and Molecules*. **Langmuir, I.** 1919, *JACS*, Vol. 41, pp. 868-934.
41. **J, Clark.** *Edexcel IGCSE CHEMISTRY Syllabus*. s.l. : Pearson, 2009.
42. *Folding versus Aggregation: Polypeptide conformations on Competing Pathway*. **Jahn T R, Radford S E.** 2008, *Biochem. and Biophys.*, Vol. 468, pp. 100-117.
43. *Protein Dynamics Tightly Connected to the Dynamics of Surrounding and Internal Water Molecules*. **Helms, V.** 2007, *ChemPhysChem*, Vol. 8, pp. 23-33.
44. *Lennard-Jones Parameters for the Combined QM/MM Method using the B3LYP/6-31G*/AMBER*

BIBLIOGRAPHIE

- Potential*. **Freindorf M, Shao Y, Furlani T R, Kong J.** 12, 2005, J. Comp. Chem., Vol. 26, pp. 1270-1278.
45. *Covalency of the Hydrogen Bond in Ice: A direct X-Ray Measurement*. **Isaacs E D, Shukla A, Platzman P M, Hamann D R, Barbielleini B R, Tulk C A.** 1999, Phys. Rev. Lett., Vol. 82, pp. 600-603.
46. *Configurations of Polypeptide Chains with Favoured Orientations Around Single Bonds: Two Pleated Sheets*. **Pauling L, Corey R B.** 1951, Proc. Natl. Ac. Sci. USA, Vol. 37, pp. 729-740.
47. *The Protein Databank*. **Berman H M, Westbrook J, Feng Z, Gilliland G, Bhat T N, Weissig H, Shindyalov I N, Bourne P E.** 2000, Nucl. Ac. Res., Vol. 28, pp. 235-242.
48. *Investigation of the Mechanism of Beta-Amyloid Fibril Formation by Kinetic and Thermodynamics Analyses*. **Lin M S, Chen L Y, Tsai H T, Wang S S, Chang Y, Higuchi A, Chen W Y.** 2008, Langmuir, Vol. 24, pp. 5802-5808.
49. *Response of a Protein Structure to Cavity-Creating Mutations and its Relation to the Hydrophobic Effect*. **Eriksson A, Baase W, Zhang X, Heinz D, Blaber M, Baldwin E, Matthews B.** 1992, Science, Vol. 255, pp. 178-183.
50. **Schrödinger, E.** *Mémoires sur la Mécanique Ondulatoire*. Paris : Felix-Alcan, 1933.
51. *Wechselwirkung neutraleur atome und homöpolare bindung nach dem quantenmechanik*. **Heitler W, London F.** 1927, Zeitschrift für Physik A Hadrons and Nuclei, Vol. 44, pp. 455-472.
52. *Wave Mechanics and the Mormal State oh the Hydrogen Molecule*. **Condon, E U.** 1927, Proc. Natl. Ac. Sci. USA, Vol. 13, pp. 466-470.
53. *Zur Quantentheorieder Molekeln*. **Born M, Oppenheimer R.** 1927, Ann. Phys., Vol. 84, pp. 457-484.
54. *Un Metodo Statistice per la Determinazione di Alcuna Proprieta dell'atomo*. **Fermi, E.** 1927, Atti. Accad. Naz., Vol. 6, pp. 602-607.
55. *The calculation of Atomic Fields*. **Thomas, L H.** 1927, Proc. Cambridge Phils. Ac., Vol. 23, p. 542.
56. *Self-Consistent Equations Including Exchange and Correlation Effects*. **Kohn W, Sham L J.** 1965, Phys. Rev., Vol. 140, pp. 1133-1138.
57. *Comparison of Simple Potential Function for Simulating Liquid Water*. **Jorgensen W L, Chandrasekhar J, Madura J, Klein M L.** 1983, J. Chem. Phys., Vol. 79, pp. 926-935.
58. *Semianalytical treatment of Solvation for Molecular Mechanics and Dynamics*. **Still W C, Tempczyk A, Hawley R C, Hendrickson T.** 1990, J. Am. Chem. Soc., Vol. 112, pp. 6127-6129.
59. *Equation of State Calculations by Fast Computing Machines*. **Metropolis N, Rosenbluth A W, Rosenbluth N M, Teller A H, Teller E.** 1953, J. Chem. Phys., Vol. 21, p. 1087.

60. *Dynamics of the Aromatic Amino Acid Residues in the Globular Conformation of the Basic Pancreatic Trypsin Inhibitor. I. 1H NMR Study.* **Wagner G, DeMarco A, Wutrich K.** 1976, Biophys. Struct. Mech., Vol. 2, pp. 139-158.
61. *Glucose-Induced Conformational Change in Yeast Hexokinase.* **Bennett W S, Steitz T A.** 10, 1978, PNAS, Vol. 75, pp. 4848-4852.
62. *Stereochemistry of Cooperative Effect in Haemoglobin.* **Perutz, M F.** 5273, 1970, Nature, Vol. 228, p. 726.
63. *The 2.3-A Resolution Structure of the Maltose-Binding or Maltodextrin-Binding Protein a Primary Receptor of Bacterial Active Transport and Chemotaxis.* **Spurlino J C, Lu G Y, Quijcho F A.** 8, 1991, J. Biol. Chem., Vol. 266, pp. 5202-5219.
64. *Core Structure of the gp41 from the HIV Envelope Glycoprotein.* **Chan D C, Fass D, Berger J M et al.** 2, 1997, Cell, Vol. 89, pp. 263-273.
65. *Flexibility and Packing in Proteins.* **Halle, B.** 3, 2002, PNAS, Vol. 99, pp. 1274-1279.
66. *Ten-Microsecond Molecular Simulation of a Fast-Folding WW Domain.* **Freddolino P L, Liu F, Gruebele M, Schulten K.** 10, 2008, Biophys. J., Vol. 94, pp. 75-77.
67. *Coarse Grained Protein Molecular Dynamics Simulations.* **Derreumaux P, Mousseau N.** 2, 2007, J. Chem. Phys., Vol. 126.
68. *The MARTINI Force Field : Coarse Grained Model for Biomolecular Simulations.* **Marrink S J, Risselada H J, Yefimov S, Tieleman D P, de Vries A H.** 27, 2007, J. Phys. Chem. B, Vol. 111, pp. 7812-7824.
69. *Normal Modes for Specific Motions of Macromolecules: Application to the Hinge-Bending Mode of Lysozyme.* **Brooks B, Karplus M.** 1985, PNAS, Vol. 82, pp. 4995-4999.
70. *Model-Free Methods of Analyzing Domain Motions in Proteins from Simulation: A comparison of Normal Mode Analysis and Molecular Dynamics Simulation of Lysozyme.* **Hayward S, Kitao A, Berendsen H J C.** 1997, Proteins, Vol. 27, pp. 425-437.
71. *Building-Block Approach for Determining Low Frequency Normal Modes of Macromolecules.* **Tama F, Gadea F X, Marques O, Sanejouand Y H.** 2000, Proteins, Vol. 41, pp. 1-7.
72. *Protein Domain Decomposition Using a Graph-Theory Approach.* **Xu Y, Xu D, Gabow H N.** 2000, Bioinformatics, Vol. 16, pp. 1091-1104.
73. *Protein Domain Movements: Detection of Rigid Domains and Visualisation of Hinges in Comparisons with Atomic Coordinates.* **Wriggers W, Schulten K.** 1997, Proteins, Vol. 29, pp. 1-14.

BIBLIOGRAPHIE

74. *PDP: Protein Domain Parser*. **Alexandrov N, Shindyalov I**. 2003, *Bioinformatics*, Vol. 19, pp. 429-430.
75. *A Database of Macromolecular Motions*. **Gerstein M, Krebs W**. 1998, *Nucl. Acid Res.*, Vol. 26, pp. 4280-4290.
76. *A Path Planning for Computing Large Amplitude Motions of Flexible Molecules*. **Cortés J, Siméon T, Ruiz de Angulo V, Guieysse D, Remaud-Siméon M, Tran V**. 2005, *Bioinformatics*, Vol. 21, pp. 116-125.
77. *Docking Macromolecules with Flexible Ligands*. **Bastard K, Thureau A, Lavery R, Prévost C**. 2003, *J. Comp. Chem.*, Vol. 24, pp. 1910-1920.
78. *ARTIST: an Activated Method in Internal Coordinate Space for Sampling Protein Energy Landscapes*. **Yun M R, Lavery R, Mousseau N, Zakrzewska K, Derreumaux P**. 2006, *Prot. Struct. Gen. Bioinfo.*, Vol. 63, pp. 967-975.
79. *Event-Based Relaxation of Continuous Disordered Systems*. **Barkema G T, Mousseau N**. 1996, *Phys. Rev. Lett.*, Vol. 77, pp. 4358-4361.
80. *Direct Evaluation of Thermal Fluctuations in Proteins Using a Single Parameter Harmonic Potential*. **Bahar I, Atilgan A R, Erman B**. 1997, *Fold. Des.*, Vol. 2, pp. 173-181.
81. *Escaping Free Energy Minima*. **Laio A, Parrinello M**. 20, 2002, *PNAS*, Vol. 99, pp. 12562-12566.
82. *Flexible Protein-Protein Docking*. **Bonvin, A**. 2006, *Opinion in Struct. Biol.*, Vol. 16, pp. 194-200.
83. *Hierarchical and Multi-Resolution Representation of Protein Flexibility*. **Zhao Y, Stoffer D, Sanner M**. 22, 2006, *Bioinformatics*, Vol. 22, pp. 2768-2774.
84. *A Geometric Approach to Macromolecule-Ligand Interactions*. **Kuntz I D, Blaney J M, Oatley S J, Langridge R, Ferrin T E**. 1982, *J. Mol. Biol.*, Vol. 161, pp. 269-288.
85. *Computer-Analysis of Protein-Protein Interaction*. **Wodak S J, Janin J**. 2, 1978, *J. Mol. Biol.*, Vol. 124, pp. 323-342.
86. *Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications*. **Kitchen D B, Decornez H, Furr J R, Bajorath J**. 2004, *Nat. Rev. Drug Discov.*, Vol. 3, pp. 935-948.
87. *Small Molecule Docking and Scoring*. **Muegge I, Rarey M**. 2001, *Rev. Comp. Chem.*, Vol. 17, pp. 1-60.
88. *Harmonic Modes as Variables to Approximately Account for Receptor Flexibility in Ligand-Receptor Docking Simulations: Application to DNA Minor Groove Ligand Complex*. **Zacharias M, Sklenar H**. 3, 1999, *J. Comp. Chem.*, Vol. 20, pp. 287-300.

89. *Molecular Modeling, Dynamics and Docking Studies of Purine Nucleoside Phosphorylase from Streptococcus Pyogenes.* **Timmers L F S M, Caceres R A, Dias R, et al.** 1-3, 2009, Biophys. Chem., Vol. 142, pp. 7-16.
90. *ProteinFlexibility and Drug Design: How to Hit a Moving Target.* **Carlson, H A.** 2002, Curr. Op. Chem. Biol., Vol. 6, pp. 447-452.
91. *Implications of Protein Flexibility for Drug Discovery.* **Teague, S J.** 2003, Nat. Rev. Drug. Discov., Vol. 2, pp. 527-541.
92. *Coarse-Grained Molecular Dynamics and the Atomic Limit of Finite Element.* **Rudd R E, Broughton J Q.** 10, 1998, Phys. Rev. B, Vol. 58, pp. R5893-R5896.
93. *Principles of Docking: an Overview of Search Algorithms and a Guide to Scoring Functions.* **Halperin I, Ma B, Wolfson H, Nussinov R.** 2002, Proteins, Vol. 47, pp. 409-443.
94. *DARWIN: A Program for Docking Flexible Molecules.* **Burnett R M, Taylor J S.** 2000, Proteins, Vol. 41, pp. 173-191.
95. *CHARMM: A Program for Macromolecular Energy, Minimization and Dynamics Calculations.* **Brooks B R, Bruccoleri R E, Olafson B D, States D J, Swaminathan S, Karplus M.** 1983, J. Comp. Chem., Vol. 4, pp. 187-217.
96. *Computation of Molecular Volume.* **Connolly, M L.** 1985, JACS, Vol. 107, pp. 1118-1124.
97. *Molecular Docking Using Shape Descriptor.* **Shoichet B K, Bodian D L, Kuntz I D.** 1992, J. Comp. Chem., Vol. 13, pp. 380-397.
98. *Rotational Automatic Search Method for Stable Docking Models of Protein and Ligand.* **Mizutani M Y, Tomioka N, Itai A.** 1994, J. Mol. Biol., Vol. 243, pp. 310-326.
99. *The Computer Program LUDI: A New Method for the De Novo Design of Enzyme Inhibitor.* **Bohm, H J.** 1992, J. Comput. Aided Mol. Des., Vol. 6, pp. 61-78.
100. *FLOG: A System to Select "Quasi-Flexible" Ligands Complementarity to a Receptor of Known Three Dimensional Structure.* **Miller M D, Kearsley S K, Underwood D J, Sheridan R P.** 1994, J. Comput. Aided Mol. Des., Vol. 8, pp. 153-174.
101. *Prodock: Software Package for Protein Modeling and Docking.* **Trosset J Y, Scheraga H A.** 1999, J. Comp. Chem., Vol. 20, pp. 412-427.
102. *A Monte Carlo Simulation Approach to the Molecular Docking Problem.* **Liu M, Wang S.** 1999, J. Comp. Aided Mol. Des., Vol. 13, pp. 435-451.

BIBLIOGRAPHIE

103. *Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function.* **Morris G M, Goodsell D L, Halliday R S, Huey R, Hart W E, Belew R K, Olson A J.** 1998, J. Comp. Chem., Vol. 19, pp. 1639-1662.
104. *Molecular Recognition of Receptor Sites Using a Genetic Algorithm with a Description of Dessolvation.* **Jones G, Willet P, Glen R C.** 1995, J. Mol. Biol., Vol. 245, pp. 43-53.
105. *Development and Validation of a Genetic Algorithm for Flexible Docking.* **Jones G, Willet P, Dlen R C, Leach A R, Taylor R.** 1997, J. mol. Biol., Vol. 267, pp. 727-748.
106. *Lessons in Molecular Recognition: the Effects of Ligand and Protein Flexibility in Molecular Docking Accuracy.* **Erickson J A, Jalaie M, Robertson D H, Lewis R A, Vieth M.** 2004, J. Med. Chem., Vol. 47.
107. *Automatic Docking to Multiple Target Structures: Incorporation of Protein Mobility and Structural Water Heterogeneity in AutoDock.* **Osterberg F, Morris G M, Sanner M F, Olson A J, Goodsell D S.** 2002, Proteins, Vol. 46, pp. 34-40.
108. *FlexE: Efficient Molecular Docking Considering Protein Structure Variation.* **Claussen H, Buning C, Rarey M, Lengauer T.** 2001, J. Mol. Biol., Vol. 308, pp. 377-395.
109. **Wilson E B, Delcius J C, Cross P C.** *Molecular Vibrations : the Theory of Infrared and Raman Vibrational Spectra.* New York : McGraw-Hill, 1955.
110. *Collective Variable Description of Small Amplitude Conformational Fluctuations in a Globular Protein.* **Noguti T, Gö N.** 1982, Nature, Vol. 296, pp. 776-778.
111. *Dynamics of a Small Globular Protein in Terms of Low Frequency Vibrational Modes.* **Gö N, Noguti T, Nishikawa T.** 1983, Proc. Natl. Ac. Sci. USA, Vol. 80, pp. 3696-3700.
112. *Low Mode Search. An Efficient, Automated Computational Method for Conformational Analysis: Application to Cyclic and Acyclic Alkanes and Cyclic Peptides.* **Kolossváry I, Guida W C.** 1996, J. Am. Chem. Soc., Vol. 118, pp. 5011-5019.
113. *Fully Flexible Low-Mode Docking: Application to Induced-fit in HIV Integrase.* **Keresû G M, Kolossváry I.** 2001, J. Am. Chem. Soc, Vol. 123, pp. 12708-12709.
114. *Representing Receptor Flexibility in Ligand Docking through Relevant Normal Modes.* **Cavasotto C N, Kovacs J A, Abagan R A.** 2005, J. Am. Chem. Soc., Vol. 127, pp. 9632-9640.
115. *Solution Properties of poly(N-isoproylacrylamide).* **Heskins M, Guillet J E.** 1968, J. Macro-mol. Sc. Chem., Vol. A2, pp. 1441-1445.
116. *A Statistical Stereochemical Model of The Flexible Furanose Ring.* **Bartenev V N, Kameneva N G, Lipanov A A.** 1987, Acta Crystal., Vol. B43, pp. 275-280.

117. *Watson-Crick Base-Pairing Properties of Nucleic Acid Analogues with Stereocontrolled and Torsion Angles (alpha,beta D-CNAs)*. **Dupouy C, Iché-Tarrat N, Durrieu M-P, Rodriguez F, Escudier J-M, Vigroux A.** 2006, *Angew. Chem.*, Vol. 118, pp. 3705-3709.
118. *alpha,beta D-CNA Induced Rigidity within Oligonucleotides*. **Dupouy C, Iché-Tarrat N, Durrieu M-P, Vigroux A, Escudier J-M.** 2008, *Org. Biomol. Chem.*, Vol. 6, pp. 2849-2851.
119. *Diastereoselective Synthesis of Conformationally Restricted Dinucleotides Featuring Canonical and Non-Canonical alpha/beta Torsion Angle Combination (alpha/beta D-CNA)*. **Dupouy C, Le Clézio I, Lavedan P, Gornitzka H, Escudier J M, Vigroux A.** 2006, *Eur. J. Org. Chem.*, Vol. 12, pp. 1935-1941.
120. **Brut, M.** *Nouvelle Approche Méthodologique pour la Prise en Compte de la Flexibilité dans les Interactions entre Molécules Biologiques : Les Modes Statiques*. Toulouse : s.n., 2009.
121. *Flap Opening and Dimer-Interface Flexibility in the Free and Inhibitor-Bound HIV Protease, and their Implications for Function*. **Ishima R, Freedberg D I, Wang Y-X, Louis J M, Torchia D A.** 1999, *Structure*, Vol. 7, pp. 1047-1055.
122. *Restrained Molecular Dynamics Simulations of HIV-1 Protease: the First Step in Validating a New Target for Drug Design*. **Perryman A L, Lin J-H, McCammon J A.** 2006, *Biopolymers*, Vol. 82, pp. 272-284.
123. *Towards a Mechanistic View of Protein Motion*. **Jagodzynski F, Brock O.** New Orleans, LA, USA : s.n., 2007. Proceedings of the 46th IEEE Conference on Decision and Control.
124. *Temperature Dependence of Protein Dynamics: Computer Simulation Analysis of Neutron Scattering Properties*. **Hayward J A, Smith J C.** 2002, *Biophys. J.*, Vol. 82, pp. 1216-1225.
125. *Crystalline Ribonuclease A Loses Function Below Dynamical Transition at 220K*. **Rasmussen B F, Stock A M, Ringe D, Petsko G A.** 1992, *Nature*, Vol. 357, pp. 423-424.
126. *Evidence for the Correlation between the Photoinduced Electron Transfer and Dynamic Properties of the Chromatophore Membranes from Rhodospirillum Rubrum*. **Parak F, Frolov E N, Kononenko A A, Mossbauer R L, Goldanskii V I, Rubin A B.** 1980, *FEBS Lett.*, Vol. 117, pp. 368-372.
127. *Thermal Motions and Function of Bacteriorhodopsin in Purple Membranes: Effects of Temperature and Hydration Studied by Neutron Scattering*. **Ferrand M, Dianoux A J, Petry W, Zaccai G.** 1993, *Proc. Natl. Acad. Sci. USA*, Vol. 90, pp. 9668-9672.
128. *Ligand Binding and Conformational Motions in Myoglobin*. **Ostermann A, Waschipky R, Parak F G, Nienhaus G U.** 2000, *Nature*, Vol. 404, pp. 205-208.
129. **Lay, D C.** *Algèbre Linéaire : théorie, exercices et applications*. de Boeck Université, Chap.5 : s.n.,

BIBLIOGRAPHIE

2007.

130. *A Solution for the Best Rotation to Relate Two Sets of Vectors*. **Kabsch, W.** 1976, Acta Cryst., Vol. A32, pp. 922-923.

131. *Zinc Binding in Proteins and Solution: A simple But Accurate Representation*. **Stote R H, Karplus M.** 1995, Proteins, Vol. 23, pp. 12-31.

132. **Case D A, Darden T A, Cheatham T E, Simmerling C L, Wang J, Duke R E, Luo R, Merz K M, Wang B, Pearlman D A, Crowley M, Brozell S, Tsui V, Gohlke H, Mongan J, Nornak V, Cui G, Beeroza P, Schafmeister C, Caldwell J W, Ross W S, Kollman P A.** AMBER 8. University of California, San Francisco : s.n., 2004.

133. *An Amyloid-Forming Peptide from Yeast Prion Sup35 Reveals a Dehydrated Beta Sheet Structure for Amyloid*. **Balbirnie M, Grothe R, Eisenberg D S.** 2001, Proc. Nat. Ac. Sci. USA, Vol. 98, pp. 2375-2380.

134. PyMOL. <http://www.pymol.org>. [Online]

135. *The Toxicity In-Vitro of Beta Amyloid Protein*. **Iversen L L, Mortishire-Smith R J, Pollack S J, Shearman M S.** 1995, Biochem. J., Vol. 311, pp. 1-16.

136. *The Amyloid Hypothesis of Alzheimer's Disease: Progress and Problems to the Road of Therapeutics*. **Hardy J, Selkoe D.** 2002, Science, Vol. 297, pp. 353-356.

137. *Novel Precursor of Alzheimers-Disease Amyloid Protein Shows Protease Inhibitory Activity*. **Kitaguchi N, Takahashi Y, Tokushima Y, Shiojiri S, Ito H.** 6156, 1988, Nature, Vol. 331, pp. 530-532.

138. *Secondary Structure of Amyloid Beta Peptides Correlates with Neurotoxic Activity In-Vitro*. **Simmons L K, May P C, Tomaselli K J, Rydel R E, Fuson K S, Brigham E F, Wright S, Liederburg I, Brems D M, Li W.** 3, 1994, Mol. Pharmacol., Vol. 45, pp. 373-379.

139. *Naturally Secreted Oligomers of Amyloid Beta Protein Potently Inhibit Hippocampal Long-Term Potentiation In-Vitro*. **Walsh D M, Klyubin I, Fadeeva J V, Cullen W K, Anwyl R, Wolfe M S, Rowan M J, Selkoe D J.** 2002, Nature, Vol. 416, pp. 535-538.

140. *Water-Soluble A β (n-40, n-42) Oligomers in Normal and Alzheimer's Disease Brains*. **Kuo YM, Emmerling MR, Vigo-Pelfrey C, Kasunic TC, Kirkpatrick JB, Murdoch GH, Ball MJ, Roher AE.** 1996, J. Biol. Chem., Vol. 271, pp. 4077-4081.

141. *Protofibrillar Intermediates of Amyloid Beta Protein Induce Acute Electrophysiological Changes and Progressive Neurotoxicity in Cortical Neurons*. **Hartley D M, Walsh D M, Ye C P, Diehl T, Vasquez S, Vassilev P M, Teplow D B, Selkoe D J.** 20, 1999, J. Neurosci., Vol. 19, pp. 8876-8874.

142. *Oligomeric and Fibrillar Species of Amyloid- β Peptides Differentially Affect Neuronal Viability*.

- Dahlgren KN, Manelli AM, Stine BW, Jr, Baker LK, Krafft GA, La Du MJ.** 2004, J. Biol. Chem., Vol. 279, pp. 198-204.
143. *Bioinorganic Chemistry of Copper and Zinc Ions Coordinated to Amyloid-Beta Peptide.* **Faller P, Hureau C.** 2009, Dalton Transactions, Vol. 7, pp. 1080-1094.
144. *Amyloids, Prions and Inherent Infectious Nature of Misfolded Protein Aggregates.* **Soto C, Estrada L, Castilla J.** 3, 2006, Trends in Biochem. Sci., Vol. 31, pp. 150-155.
145. *Amyloid Beta Protein and the Genetics of Alzheimer's Disease.* **Selkoe, D J.** 31, 1996, J. Biol. Chem., Vol. 271, pp. 18295-18298.
146. *Prion Diseases and the BSE Crisis.* **Prusiner, S B.** 5336, 1997, Science, Vol. 278, pp. 245-251.
147. *Identification and Characterization of Key Kinetic Intermediates in Amyloid Beta-Protein Fibrillogenesis.* **Kirkitadze M D, Condrón M M, Teplow D B.** 5, 2001, J. Mol. Biol., Vol. 312, pp. 1103-1119.
148. *Interpeptide Interactions Induce Helix to Strand Transition in A β Peptides.* **Takeda T, Klimov K.** 2009, Proteins, Vol. 77, pp. 1-13.
149. *Protein Aggregation : Folding Aggregates, Inclusion Bodies and Amyloid.* **Fink, A L.** 1998, Fold. Des., Vol. 3, pp. R9-R23.
150. *Dissecting the Assembly of A β (16-22) Amyloid Peptides into Antiparallel β Sheets.* **Klimov D K, Thirumalai D.** 2003, Structure, Vol. 11, pp. 295-307.
151. *Kinetic Control of Dimer Structure Formation in Amyloid Fibrillogenesis.* **Hwang W, Zhang S, Kamm R D, Karplus M.** 2004, Proc. Nat. Ac. Sci. , Vol. 101, pp. 21916-12921.
152. *Probing the Initial Stage of Aggregation of the A β (10-35)-Protein: Assessing the Propensity for Peptide Dimerization.* **Tarus B, Straub J E, Thirumalai D.** 2005, J. Mol. Biol., Vol. 345, pp. 1141-1156.
153. *Elongation of Ordered Peptide Aggregate of an Amyloidogenic Hexapeptide NFGAIL Observed in Molecular Dynamics Simulations with Explicit Solvent.* **Wu C, Lei H, Duan Y.** 2005, J. Am. Chem. Soc., Vol. 127, pp. 13530-13537.
154. *Casting Metal Nanowires Within Discrete Assembled Peptide Nanotubes.* **Reches M, Gazit E.** 2003, Science, Vol. 300, pp. 625-627.
155. *Time-Resolved Infrared Spectroscopy of pH-Induced Aggregation of the Alzheimer A-Beta(1-28) Peptide.* **Peralvarez-Marín A, Barth A, Graslund A.** 3, 2008, J. Mol. Biol., Vol. 379, pp. 589-596.
156. *Aluminium, Iron and Zinc Ions Promote Aggregation of Physiological Concentrations of Beta-Amyloid*

BIBLIOGRAPHIE

Peptide. **Mantyh P W, Ghilardi J R, Rogers S, Demaster E, Allen C J, Stimson E R, Maggio J E.** 3, 1993, J. Neurochem., Vol. 61, pp. 1171-1174.

157. *Histidine-13 is a Crucial Residue in the Zinc-Ion Induced Aggregation of the A β Peptide of Alzheimer's Disease*. **Liu S T, Howlett G, Barrow C J.** 29, 1999, Biochemistry, Vol. 38, pp. 9373-9378.

158. *Examining the Zinc Binding Site of Amyloid- β Peptide*. **Yang D S, McLaurin J, Qin K, Westaway D, Fraser P E.** 22, 2000, Eur. J. Biochem., Vol. 267, pp. 6692-6698.

159. *Characterization of the ZnII Binding to the Peptide Amyloid-beta1-16 linked to Alzheimer's Disease*. **Mekmouche Y, Coppel Y, Hochgräfe K, Guilloreau L, Talmard C, Mazarguil H, Faller P.** 2005, ChemBioChem, Vol. 6, pp. 1663-1671.

160. *Zinc Binding Properties of the Amyloid Fragment A β (1-16) Studied by Electrospray-Ionization Mass Spectrometry*. **Zirah S, Rebuffat S, Kozin S A, Debey P, Fournier F, Lesage D, Tabet J-C.** 2003, Int. J. Mass Spectrom., Vol. 228, pp. 999-1016.

161. *Metal Binding Modes of Alzheimer's Amyloid β -Peptide in Insoluble Aggregates and Soluble Complexes*. **Miura T, Suzuki K, Kohata N, Takeuchi H.** 23, 2000, Biochemistry, Vol. 39, p. 7024.

162. *Structural Changes of Region 1-16 of the Alzheimer Disease Amyloid Beta-Peptide upon Zinc Binding and In Vitro Aging*. **Zirah S, Kozin S A, Mazur A K, Blond A, Cheminant M, Segalas-Milazzo I, Debey P, Rebuffat S.** 4, 2006, J. Biol. Chem., Vol. 281, pp. 2151-2161.

163. *Neurotoxicity and Physicochemical Properties of A β Mutant Peptides from Cerebral Amyloid Angiopathy: Implication for the Pathogenesis of Cerebral Amyloid Angiopathy and Alzheimer's Disease*. **Murakami K, Irie K, Morimoto A, Ohigashi H, Shindo M, Nagao M, Shimizu T, Shirazawa T.** 2003, J. Biol. Chem., Vol. 278, pp. 46179-46187.

164. *Toxicity of Various Amyloid Beta Peptide Species in Cultured Human Blood-Brain Barrier Endothelial Cells: Increased Toxicity of Dutch-Type Mutant*. **Eisenhauer P B, Johnson R J, Wells J M, Davies T A, Fine R E.** 2000, J. Neurosc. Res., Vol. 60, pp. 804-810.

165. *The 'Arctic' APP Mutation (E693G) causes Alzheimer's Disease by Enhanced Abeta Protofibril Formation*. **Nilsberth C, Westlind-Danielson A, Eckman C B, Condon M M, Axelman K, Forsell C, Sten C, Luthman J, Teplow D B, Younkin S G, Naslund J, Lannfelt L.** 2001, Nat. Neurosci., Vol. 4, pp. 887-893.

166. *A Molecular Dynamics Approach to the Structural Characterization of Amyloid Aggregation*. **Cecchini M, Curcio R, Pappalardo M, Melki R, Caflisch A.** 2006, J. Mol. Bio., Vol. 357, pp. 1306-1321.

167. *Elucidating β -Amyloid Folding and Assembly: A Multidisciplinary Approach*. **Teplow D B, Lazo N D,**

- Bitá G, Bernstein S, Wyttenbach T, Towers M T, Baumketner A, Shea J E, Urbanc B, Cruz L, Borreguero J, Stanley H E.** 2006, *Acc. Chem. Res.*, Vol. 39, pp. 635-645.
168. *On the Nucleation of Amyloid β -Protein Monomer Folding.* **Lazo N D, Grant M A, Condrón M M, Rigby A C, Teplow D B.** 2005, *Prot. Sci.*, Vol. 14, pp. 1581-1596.
169. *Molecular Dynamics Simulations of Alzheimer's β -Amyloid Protofilaments.* **Buchete N V, Tycko R, Hummer G.** 2005, *J. Mol. Bio.*, Vol. 353, pp. 804-821.
170. *The Conformations of the Amyloid- β (21-30) Fragment can be Described by Three Families in Solution.* **Chen W, Mousseau N, Derreumaux P.** 2006, *J. Chem. Phys.*, Vol. 125, pp. 084911-084918.
171. *Solvent and Mutation Effects on the Nucleation of Amyloid β -Protein Folding.* **Cruz L, Urbanc B, Borreguero J M, Lazo N D, Teplow D B, Stanley H E.** 2005, *J. Am. Chem. Soc.*, Vol. 127, pp. 18258-18263.
172. *High-Resolution NMR Spectroscopy of the Beta-Amyloid (1-28) Fibril Typical for Alzheimer's Disease.* **Mikros E, Benaki D, Humpfer E, Spraul M, Loukas S, Stassinopoulou C I, Pelecanou M.** 19, 2001, *Angew. Chem. Int. Ed.*, Vol. 40, p. 3603.
173. *A Structural Model for Alzheimer's Beta-Amyloid Fibrils Based on Experimental Constraints from Solid State NMR.* **Petkova A T, Ishii Y, Balbach J J, Antzutkin O N, Leapman R D, Delaglio F, Tycko R.** 26, 2002, *PNAS*, Vol. 99, pp. 16742-16747.
174. *Conformation and Fibrillogenesis of Alzheimer A-Beta Peptides with Selected Substitution of Charged Residues.* **Fraser P E, MacLachlan D R, Surewicz W K, Mizzen C A, Snow A D, Nguyen J T, Kirschner D A.** 1, 1994, *J. Mol. Biol.*, Vol. 224, pp. 64-73.
175. *Solution Structure of Residues 1-28 of the Amyloid Beta-Peptide.* **Talafous J, Marcinowski K J, Klopman G, Zagorski M G.** 25, 1994, *Biochem.*, Vol. 33, pp. 7788-7796.
176. *Structure of the Cross- β Spine of Amyloid-Like Fibrils.* **Nelson R, Sawaya M R, Balbirnie M, Madsen A O, Riek C, Eisenberg D.** 2005, *Nature*, Vol. 435, pp. 773-778.
177. *Models of Amyloid Seeding in Alzheimer's Disease and Scrapie: Mechanistic Truth and Physiological Consequences of Time-Dependent Solubility of Amyloid Proteins.* **Harper J D, Lansbury P T.** 1997, *Ann. Rev. of Biochem.*, Vol. 66, pp. 385-407.
178. *Pathologic Conformations of Prion Proteins.* **Cohen F E, Prusiner S B.** 1998, *Ann. Rev. of Biochem.*, Vol. 67, pp. 793-819.
179. *Prion Protein Biology.* **Prusiner S B, Scott M R, De Armond S J, Cohen F E.** 1998, *Cell*, Vol. 93, pp. 337-348.

BIBLIOGRAPHIE

180. *A Point-Charge Force Field for Molecular Mechanics Simulations for Proteins Based on Condensed Phase Quantum Mechanical Calculations*. **Duan Y, Wu C, Chowdhury S, Lee M C, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P**. 2003, J. Comp. Chem., Vol. 24, pp. 1999-2012.
181. *Structural Stability and Dynamics of an Amyloid-Forming Peptide GNNQQNY from the Yeast Prion Sup35*. **Zheng J, Ma B, Tsa C J, Nussinov R**. 2006, Biophys. J., Vol. 91, pp. 824-833.
182. *Molecular Dynamics Simulations on the Oligmer Formation Process of the GNNQQNY Peptide from Yeast Prion Protein Sup35*. **Zhang Z, Chen H, Bai H, Lai L**. 2007, Biophys. J., Vol. 93, pp. 1484-1492.
183. *Force Fields for Protein Simulations*. **Ponder J W, Case D A**. 2003, Adv. Prot. Chem., Vol. 66, pp. 27-78.
184. Protein Data Bank. <http://www.rcsb.org/pdb/home/home.do>. [Online]
185. *Structure Validation by Ca Geometry: ϕ , ψ and $C\beta$ Deviation*. **Lovell S C, Davis I W, Arendall W B, de Bakker P I W, Word J M, Prisant M G, Richardson J S, Richardson D C**. 2003, Proteins, Vol. 50, pp. 437-450.
186. *Energy Landscape of the Monomer and Dimer of the Alzheimer's peptide $A\beta(1-28)$* . **Dong X, Chen W, Mousseau N, Derreumaux P**. 2008, J. Chem. Phys, Vol. 125, pp. 125108-125118.
187. *3D Structure of Alzheimer's Amyloid- $\beta(1-42)$ Fibrils*. **Luhers T, Ritter C, Adrian M, Riek-Loher D, Bohrmann B, Dobeli H, Schubert D, Riek R**. 2005, Proc. Nat. Ac. Sci., Vol. 102, pp. 17342-17347.
188. *Prediction of Sequence-Dependant and Mutational Effects on the Aggregation of Peptides and Proteins*. **Fernandez-Escamilla A M, Rousseau F, Schymkowitz J, Serrano L**. 10, 2004, Nature Biotech., Vol. 22, pp. 1302-1306.
189. *Simulation Study of the Structure and Dynamics of the Alzheimer's Amyloid Beta Congener in Solution*. **Massi F J, Peng J W, Lee J P, Straub J E**. 2001, Biophys. J., Vol. 80, pp. 31-44.
190. *Structural and Kinetic Features of Amyloid β -Protein Fibrillogenesis*. **Teplow, D B**. 1998, Amyloid, Vol. 5, pp. 121-142.
191. *Characterization of Chemical Exchange Between Soluble and Aggregated States of β -Amyloid by Solution-State NMR upon Variation of Salt Conditions*. **Narayanan S, Reif B**. 2005, Biochemistry, Vol. 44, pp. 1444-1452.
192. *Role of Aromatic Interactions in Amyloid Formation by Peptides Derived from Human Amylin*. **Tracz S M, Abedini A, Driscoll M, Raleigh D P**. 2004, Biochemistry, Vol. 43, pp. 15901-15908.
193. *Interspecies Sequence Variations Affect the Kinetics and the Thermodynamics of Amyloid Formation:*

Peptide Models of Pancreatic Amyloid. **Ashburn T T, Lansbury Jr., P T.** 1993, J. Am. Chem. Soc., Vol. 115, pp. 11012-11013.