



Nouvelle approche méthodologique pour la prise en compte de la flexibilité dans les interactions entre molécules biologiques: les Modes Statiques

Marie Brut

► To cite this version:

Marie Brut. Nouvelle approche méthodologique pour la prise en compte de la flexibilité dans les interactions entre molécules biologiques: les Modes Statiques. Micro et nanotechnologies/Microélectronique. Université Paul Sabatier - Toulouse III, 2009. Français. NNT: . tel-00377792

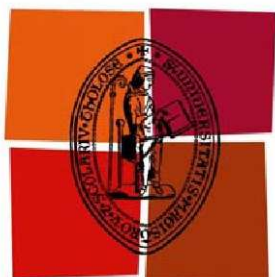
HAL Id: tel-00377792

<https://theses.hal.science/tel-00377792>

Submitted on 23 Apr 2009

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

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



Université
de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITE DE TOULOUSE

Délivré par l'Université Toulouse III - Paul Sabatier

Spécialité : Nanophysique

Présentée et soutenue par

Marie BRUT

Le 5 mars 2009, au LAAS-CNRS

Titre :

Nouvelle approche méthodologique pour la prise en compte de la flexibilité
dans les interactions entre molécules biologiques : Les Modes Statiques

JURY

Mme Françoise NEPVEU, Professeur, Université de Toulouse

M. Jean-Claude LATOMBE, Professeur, Stanford University

M. Jean-Marc VICTOR, Directeur de recherche, LPTMC-CNRS, Paris

Mme Monique ERARD, Directrice de recherche, IPBS-CNRS, Toulouse

Mme Chantal PREVOST, Chargée de recherche, IBPC-CNRS, Paris

M. Frédéric DAUMAS, Directeur du CEEI Théogone, Toulouse

M. Mehdi DJAFARI ROUHANI, Professeur, Université de Toulouse

M. Alain ESTEVE, Chargé de recherche, LAAS-CNRS, Toulouse

Présidente

Rapporteur

Rapporteur

Examinatrice

Examinatrice

Examineur

Directeur

Co-directeur

INVITE

M. Alain Michel BOUDET, Professeur, Université de Toulouse

Invité

Ecole doctorale : Sciences de la Matière ED 482

Unité de recherche : LAAS-CNRS UPR 8001

Directeur de Thèse : Professeur Mehdi DJAFARI ROUHANI

A la mémoire de ma Maraine

A mes parents

A mon Corentin

Remerciements

Le travail de thèse présenté dans ce mémoire a été effectué au sein du groupe Nano Ingénierie et Intégration des Systèmes (N2IS) au Laboratoire d'Analyse et d'Architecture des Systèmes du Centre National de la Recherche Scientifique (LAAS-CNRS).

Je remercie en premier lieu Malik Ghallab et Raja Chatila, directeurs successifs du LAAS, de m'avoir accueillie au sein de leur laboratoire. Je pense sincèrement que l'existence d'un tel établissement, actif, ouvert, et mêlant tant d'acteurs et de sujets de recherches différents est une chance qui a contribué à mon épanouissement personnel et scientifique, en offrant un cadre de travail idéal.

Je remercie également Anne-Marie Gué et Jean-Yves Fourniols, les responsables successifs du groupe MIS/N2IS, groupe fort sympathique qui multiplie initiatives et animations scientifiques pour maintenir un cadre de recherche et d'émulation à la fois performant et convivial.

Bien entendu, un immense et respectueux merci à Mehdi Djafari Rouhani, pour avoir accepté de tenter l'aventure avec moi en encadrant cette thèse. Voici enfin l'occasion de lui dire mon admiration ! J'ai été émerveillée, à chaque conversation, par ce savoureux mélange de savoir, de sagacité, d'humour et de modestie. Une de mes plus grandes satisfactions, à la fin de ces années, est d'avoir essayé d'apprendre un peu de lui (même si j'ai découvert en même temps la déroutante sensation de se sentir toute petite !).

De la même façon, je remercie chaleureusement Alain Estève, mon co-directeur de thèse, qui forme avec Mehdi un duo inattendu et inespéré. Son regard avisé a permis de guider ce travail d'une main de maître, trouvant sans cesse des pistes à explorer ... tantôt sous les sunlights, tantôt à la chandelle ! Je n'ai jamais douté de vouloir m'engager sur ce sujet, mais au-delà, quel bonheur de vous avoir rencontrés tous les deux. Merci de votre patience, de votre engagement, de votre enthousiasme, de votre appui permanent et en toutes circonstances.

Pour procéder par ordre, il me faut à présent remercier Monsieur Georges Landa, dernier membre du tiercé gagnant ! Sa façon si personnelle de tenter de me faire sortir de ma coquille... (comment oublier le fameux "publish or perish ! " ?) m'a marqué sûrement plus qu'il ne croit ! L'application et l'énergie qu'il a déployées pour suivre les voies parfois tortueuses de mon travail m'ont beaucoup touchée. Merci à toi Georges, d'avoir toujours, toujours été là, de mes déboires informatiques à mes interrogations personnelles !

Je tiens à remercier Mme Françoise Nepveu qui m'a fait l'immense honneur de présider mon jury et de porter un regard neuf et enthousiaste sur ce travail. Je remercie particulièrement M. Jean-Claude Latombe et M. Jean-Marc Victor qui ont accepté de rapporter mon travail avec

patience et ouverture, et ont fait avancé, ou même relancé ma réflexion par leurs remarques. Toute ma gratitude va à Mme Monique Erard, pour avoir accepté de participer à ce jury, et plus largement d'avoir répondu présente à chaque sollicitation ; l'intérêt et l'enthousiasme qu'elle a montrés pour ces travaux m'ont apporté un grand réconfort à chaque entretien. Par ailleurs, son regard de biologiste nous est toujours indispensable ! Je tiens enfin à remercier Mme Chantal Prévost qui m'a fait l'honneur et le plaisir d'être membre de ce jury (ce fut l'occasion d'entamer des discussions et de partager certaines préoccupations communes) ainsi que M. Frédéric Daumas et M. Alain Boudet qui, par leurs questions et remarques, ont contribué à la qualité scientifique du débat.

Merci également à Daniel Estève. A l'approche de la banquise, Daniel avance à la manière d'un brise-glace, lorsque de mon côté, j'essaie plutôt de la faire fondre... C'est pour cette raison que nous ne sommes pas toujours très bien compris, mais aussi que j'ai envie de lui dire aujourd'hui toute ma reconnaissance. Le personnage m'a beaucoup impressionné, c'est assez fascinant de voir évoluer Daniel, sur tous les fronts, tel un chef d'orchestre impérieux, magistral. Je lui suis infiniment reconnaissante pour son engagement à faire avancer notre projet, à valoriser ce travail. Daniel, malgré mes efforts, je suis encore très loin d'être aussi fonceuse que vous l'attendez de moi ! Malgré tout, ne doutez pas de ma bonne volonté et de ma gratitude.

Je remercie Juan Cortés, qui a suivi ce travail et m'a fait part de réflexions toujours bien venues. Son approche différente face à notre problème commun et les discussions qui en ont résulté m'ont amenée à me poser de nombreuses questions, et heureusement à en résoudre d'autres... Je pense aussi à Marc Vaisset, qui a accepté sans hésiter d'apporter son savoir informatique à l'amélioration de notre code, ainsi que David Gauchard.

Je tiens également à remercier les permanents du LAAS qui m'ont apporté leur aide, particulièrement dans l'effervescence des derniers instants. Je pense bien sûr à notre secrétaire Nicole Higounet, mais aussi au service de documentation avec Arlette Evrard, à Christian Berthy qui a imprimé ce manuscrit, ainsi qu'à Daniel Daurat et Yves Crouzet qui m'ont aidé pour la mise en oeuvre technique de la soutenance, et ce jusqu'à la dernière minute ! Je n'oublie pas Jean-Michel Pons, créateur de notre site web !

Je voudrais enfin remercier les membres du groupe N2IS qui ont montré de l'enthousiasme pour un travail parfois loin d'eux, et surtout, qui m'ont apporté leur soutien et leur bonne humeur. Je ne tenterai sûrement pas de me lancer dans une liste exhaustive, ce qui est un moyen facile de n'oublier personne, mais je pense bien sûr à Carole parmi les permanents. Et parmi les autres, la fine équipe du bureau 9C, Marine, Hélène et Jean-Marie, avec qui j'ai passé de si bon moments et qui ont toujours répondu présents. Jean-Marie, je n'oublie pas ta contribution à ce manuscrit : il m'aurait sûrement fallu plus de temps pour l'écrire si tu ne m'avais pas harcelé à chaque pause-café ! Merci enfin à mes adorables collègues de bureau, Hugo et Anne, partenaires du stress des derniers instants ! Tout particulièrement, je tiens à témoigner à Anne toute ma reconnaissance pour tout ce qu'elle a pu m'apporter ces derniers mois, de ses blagues du matin à son passage en régie lors de la soutenance. Chère Anne, tu es une fille vraiment, vraiment formidable !

Il y a enfin les amis et proches de l'extérieur, présents dans toutes les aventures, et il devient risqué de citer des noms ! Mais comment ne pas citer Benjamin, associé à tous mes bons souvenirs de fac. Des rencontres formidables au cours de ces dernières années ont pour beaucoup décidé

de ma condition ! Et sans en dire davantage, nos petites soirées et escapades m'ont toujours fait un bien fou ! Merci à toi de me supporter depuis... au moins !

Toute mon amitié également à Noëlla. Elle est toujours à mes côtés, avec patience et tolérance pour mon caractère, depuis 20 ans. Lorsque je nous regarde aujourd'hui, nous et nos vies de "grandes filles", je suis émue de tout ce chemin, mais surtout de notre constance. Que tout continue comme cela longtemps !

Il y a enfin ceux pour lesquels le mot "merci" n'est plus suffisant. Mes parents tout d'abord. Chacuns de vos mots, de vos gestes m'ont construite. Je voudrais vous dire aujourd'hui toute la reconnaissance j'ai pour vous. C'est la plus grande des chances de vivre entourée comme je le suis, par des gens qui rendent la vie plus belle, et les pires moments finalement surmontables. Cette thèse en particulier, n'existe aujourd'hui que par des années de sacrifices et de soutien permanent de votre part. Vous avez ma gratitude éternelle. Je vous aime tant.

Je pense avec tendresse à mon parain, qui est un véritable modèle de courage et d'amour. Je veux lui dire aujourd'hui toute la fierté que j'ai d'être sa nièce et sa filleul. Il est pour moi une raison supplémentaire de ne jamais baisser ni les bras, ni la tête.

Je remercie affectueusement ma grand-mère qui, même de loin, a toujours surveillé ma vie toulousaine et mon travail. Elle m'a toujours soutenue et encouragée, et la savoir toujours derrière moi est un véritable réconfort.

Tout mon amour enfin à mon Corentin. Presque sept années qu'il m'épaule, m'apporte un bonheur immense, et plus que tout, la joie de vivre et le courage d'aller de l'avant. Ce travail ne se serait pas fait si facilement sans lui, mon premier supporter, mon premier lecteur. Je ne te dis pas assez la chance que j'ai de t'avoir, mais je le sais, chaque jour. Merci d'être là !

Un grand merci enfin à Pascale et Jean-Luc, en qui j'ai trouvé un réconfort et un soutien supplémentaire. Merci de m'avoir ouvert les bras avec temps de gentillesse et de votre investissement dans ce travail, par votre lecture tout d'abord, mais surtout par les discussions que nous avons eues, et aurons encore longtemps je l'espère !

Table des matières

Remerciements	i
Table des figures	xi
Liste des tableaux	xvii
Introduction générale	1
Chapitre 1	
La problématique des interactions en biologie	5
1.1 Enjeux socio-économiques	6
1.1.1 La biologie aujourd’hui et ses applications	6
1.1.2 L’importance de la biochimie, une science récente	6
1.1.3 La réponse de la science aux questions de la biochimie	8
1.1.4 Enjeux et impact sur la société	9
1.2 Les objets moléculaires en biologie	10
1.2.1 Les lipides	12
1.2.2 Les glucides	12
1.2.3 Les peptides et les protéines	13
1.2.4 Les acides nucléiques	14
1.2.5 L’eau	15
1.3 Interactions	15
1.3.1 Les liaisons chimiques fortes	16

1.3.2	Les liaisons chimiques faibles	18
1.3.3	Implication des interactions dans la structure et la flexibilité	20
1.4	Conclusion	21

Chapitre 2

Modélisation et simulation des biomolécules

2.1	Les modèles existants	24
2.1.1	Les méthodes quantiques	24
2.1.2	La mécanique moléculaire	28
2.1.3	Les simulations dynamiques	32
2.2	Mise en perspective par rapport à la biologie	34
2.2.1	Un développement parallèle aux progrès informatiques	34
2.2.2	Adaptation des outils aux problèmes de la biologie	35
2.2.3	Prise en compte du solvant	36
2.3	Du docking rigide à l'introduction de la flexibilité	37
2.3.1	Représentation du système	38
2.3.2	Procédure de recherche	38
2.3.3	Le scoring	43
2.4	Conclusion	44

Chapitre 3

La méthode des Modes Statiques

3.1	Origine et concept	48
3.1.1	Faire face aux limites actuelles	48
3.1.2	Limites et légitimité de l'application des Modes Normaux	48
3.1.3	Apports de la méthode des Modes Statiques	52
3.2	La méthode des Modes Statiques	53
3.2.1	Méthodologie	53
3.2.2	Le logiciel FLEXIBLE	55

3.2.3	Discussion	57
3.3	Conclusion	57

Chapitre 4

Résultats

4.1	Quelques outils pour l'étude de la flexibilité	60
4.1.1	Localisation des modes	60
4.1.2	Optimisation des forces	61
4.1.3	Pincements des couples d'atomes	62
4.2	Application des modes statiques à un polymère thermosensible : le PNIPAM . .	63
4.2.1	Présentation du polymère	63
4.2.2	Premières observations sur la flexibilité du polymère	64
4.2.3	Etude du site actif d'un dimère de NIPAM : la liaison hydrogène inter- monomères	65
4.2.4	Extension au cas d'un tétramère de NIPAM	68
4.2.5	Conclusion	70
4.3	Etude de la flexibilité des sucres dans les acides nucléiques	70
4.3.1	Les plissements du sucre : concept de pseudorotation	71
4.3.2	Cas d'un nucléotide : la désoxythymidine monophosphate	72
4.3.3	Etude de l'apport de nucléotides modifiés sur la stabilité de duplex d'acides nucléiques	76
4.3.4	Conclusion	81
4.4	Etude du site actif de la protéase du VIH-1	82
4.4.1	Description de la molécule : structure et mécanismes	82
4.4.2	Evaluation de la stabilité du site actif	84
4.4.3	Exploration du mécanisme d'ouverture	88
4.4.4	Simulation des effets de l'arrimage d'un ligand	92
4.4.5	Conclusion	94

4.5 Conclusions	95
---------------------------	----

Conclusion générale	97
----------------------------	-----------

Annexes

Annexe A

Structure des peptides et des protéines

A.1 Introduction	101
A.2 Différents niveaux de description	101
A.2.1 Structure primaire	101
A.2.2 Structure secondaire	102
A.2.3 Structure tertiaire	104
A.2.4 Structure quaternaire	104
A.3 Acides aminés	104

Annexe B

Structure des acides nucléiques

B.1 Historique	107
B.2 Structure	108
B.3 Configuration	110
B.4 Récapitulatif des bases, nucléosides et nucléotides	112
B.4.1 Purines	112
B.4.2 Pyrimidines	113

Annexe C

Les modes normaux

C.1 Principe	115
C.2 Calcul des modes propres	116

Annexe D**Résolution de systèmes linéaires : méthode de Gauss-Seidel**

D.1	Position du problème	117
D.2	Principe	117
D.2.1	Ecriture du système	117
D.2.2	Décomposition de la matrice A	118
D.3	Méthode de Gauss-Seidel	119
D.3.1	Description de la méthode	119
D.3.2	Condition d'arrêt	119
D.3.3	Algorithme	120
Bibliographie		123

Table des figures

1.1	Exemple d'éléments identifiables par l'expérience sur une molécule biologique et de la technique associée. <i>Source Angew. Chem. Int. Ed.</i> [29]	9
1.2	Exemple de la structure chimique de deux acides gras.	12
1.3	Formation d'une chaîne de polymère par réaction de déshydratation.	13
1.4	Exemple de monosaccharide : le glucose, en représentation linéaire et cyclique.	13
1.5	Exemple d'interactions intervenant dans le repliement d'une chaîne polypeptidique.	14
1.6	De haut en bas : distribution des électrons dans des liaisons covalentes, covalentes polarisées et ioniques.	18
2.1	Représentation des différents potentiels utilisés pour modéliser les termes d'élongation et de variation angulaire.	30
2.2	Exemple d'application de la méthode hybride QM/MM à un système moléculaire.	31
2.3	Dihydrofolate réductase représentée en solvant explicite : une boîte cubique de 64 Å de côté créée par AMBER avec le modèle TIP3P.	36
2.4	Schéma des 3 niveaux de représentation de la flexibilité d'une protéine. a) modèle clé-serrure. b) modèle induced-fit. c) configurations pré-existantes à l'équilibre.	39
3.1	Profils énergétiques d'une molécule passant de l'état libre à l'état lié. a) équilibre de la molécule libre. b) modification de l'énergie durant l'approche du ligand. c) nouvelle énergie de la molécule à l'état lié. <i>Source Structure</i> [180].	51
3.2	Organigramme présentant la procédure actuelle du code Flexible. Un calcul d'énergie totale réalisé au préalable est requis pour fournir la matrice Hessienne et les coordonnées de la molécule en entrée. Flexible fait alors intervenir un certain nombre de sous-programmes dont les principaux sont indiqués ici (RMAT, LCMOD et LinMode). Une fois la résolution achevée, l'ensemble des modes calculés est stocké dans la "matrice des déformations" pour une exploitation ultérieure.	56
4.1	Pincement d'un couple d'atomes par application de deux forces $\vec{F}_i$ et $\vec{F}_j$	62

4.2	Représentation atomique d'un dimère de NIPAM obtenu après minimisation avec GAUSSIAN03 ($-2,128.10^{+6}$ kJ/mol). Le backbone est noté bb et les deux monomères a et b . La liaison hydrogène inter-monomères considérée dans la suite de ce travail est également située. A droite, numérotation des atomes du dimère de NIPAM étudié.	64
4.3	Représentation graphique du calcul de localisation des modes. A droite, nous représentons la déformation de la molécule relative à deux modes du graphe, un délocalisé et un localisé. Ces déformations correspondent respectivement à des contraintes appliquées sur l'atome 22 en direction y et l'atome 12 en direction x, situés tous deux sur la molécule à droite.	65
4.4	Variations de la distance interatomique OH en fonction de l'éloignement entre le site actif (ramené au milieu du segment OH) et les atomes contraints des monomères a , b et du backbone bb respectivement (en Å ² /eV).	66
4.5	A gauche, numérotation des atomes du dimère de NIPAM étudié. A droite, les variations de la distance interatomique OH pour chaque atome sollicité (en Å ² /eV).	67
4.6	A gauche, représentation des forces de direction optimisée et normée pour induire une ouverture maximale de la distance OH. A droite, représentation des mêmes forces avec la norme proportionnelle à l'ouverture engendrée.	67
4.7	Représentation atomique d'un tétramère de NIPAM après minimisation avec GAUSSIAN03 ($-4,045.10^{+6}$ kJ/mol). Le backbone est noté bb et les quatre monomères a , b , c et d . Les trois liaisons hydrogène inter-monomères sont situées et notées 1, 2 et 3.	68
4.8	A gauche sont représentées les variations de l'ouverture de la liaison OH établie entre les groupes CO et NH des monomères b et c . A droite, sur le tétramère de NIPAM étudié, les forces de direction optimisée et de norme proportionnelle à l'ouverture de la liaison OH sont reportées.	69
4.9	A gauche sont représentées les variations de l'ouverture de la liaison OH établie entre les groupes CO et NH des monomères a et b du tétramère. A droite, les forces de direction optimisée et de norme proportionnelle à l'ouverture de la liaison OH sont reportées.	69
4.10	A gauche sont représentées les variations de l'ouverture de la liaison OH établie entre les groupes CO et NH des monomères c et d du tétramère. A droite, les forces de direction optimisée et de norme proportionnelle à l'ouverture de la liaison OH sont reportées.	70
4.11	Représentation du cycle d'un sucre entrant dans la composition des acides nucléiques. Il s'agit d'un furanose, contenant 4 atomes de carbone et 1 atome d'oxygène. La numérotation utilisée est celle de la nomenclature standard [332][333][335].	71
4.12	Cycle de pseudorotation du sucre. Source <i>J. Chem. Inf. Comput. Sci.</i> 2004 [365].	71

4.13	Représentation des conformations préférentielles du sucre : C3'-endo et C2'-endo, respectivement nommées nord et sud. <i>Source Springer-Verlag 1984 [360].</i>	72
4.14	Représentation des deux systèmes étudiés après minimisation avec AMBER8 : à gauche, la désoxythymidine monophosphate (dTMP) seule ($-3,456.10^{+1}$ kJ/mol), et à droite appariée avec une désoxyadénosine monophosphate (dAMP) ($-1,818.10^{+1}$ kJ/mol).	73
4.15	Représentation des variations de phase du sucre d'une dTMP seule à gauche, et appariée avec une dAMP à droite. Chaque point correspond à une maximisation de la phase après optimisation des forces.	74
4.16	Numérotation des atomes de la dTMP.	75
4.17	A gauche, variations de la phase du sucre de la dTMP seule, après optimisation des forces. A droite, deux configurations du sucre, correspondant à l'état d'équilibre (2E) et au plus grand changement de phase (2_1T).	75
4.18	Représentation des forces optimisées pour maximiser le changement de phase du sucre de la dTMP.	75
4.19	Structure chimique d'un dinucléotide thymine-thymine monophosphate (TpT) naturel à gauche et de son homologue contraint à droite. <i>Source Angew. Chem. [320].</i>	77
4.20	Structures des deux duplex d'ADN après minimisation : à gauche le duplex naturel ($-1,691.10^{+2}$ kJ/mol) et à droite le duplex contenant un dinucléotide modifié ($\alpha=g-, \beta=t$) ($-1,481.10^{+2}$ kJ/mol).	78
4.21	Structures des deux hétéroduplex ADN:ARN après minimisation : à gauche le duplex naturel ($-1,348.10^{+2}$ kJ/mol), à droite le duplex contenant un dinucléotide modifié ($\alpha=g-, \beta=t$) ($-1,009.10^{+2}$ kJ/mol).	78
4.22	Numérotation des quatre liaisons hydrogène étudiées.	78
4.23	Structures des deux dinucléotides TpT après minimisation avec GAUSSIAN03 : à gauche, le dinucléotide naturel ($-5,883.10^{+6}$ kJ/mol), à droite, le dinucléotide modifié ($\alpha=g-, \beta=t$) ($-6,086.10^{+6}$ kJ/mol).	80
4.24	Représentation des variations de phase du sucre 5' d'un dinucléotide TpT non modifié à gauche, et modifié à droite. Chaque point correspond à une maximisation de la phase après optimisation des forces.	81
4.25	Représentation de la protéase du VIH et de ses domaines : en vert les volets, en rouge le coeur et en orange l'interface. Sont également indiqués : le site catalytique, les poches de liaison, les leviers et le domaine terminal.	83
4.26	Deux configurations de la protéase : fermée à gauche et ouverte à droite. Ces structures correspondent respectivement aux fichiers 1HHP et 1TW7 de la <i>Protein Data Bank</i> .	84

4.27	Variations maximales Δl de la distance séparant les C_α des résidus Asp25 et Asp25' après optimisation des forces sur chaque atome de la protéase. A droite, les zones de la protéase induisant une réponse du site actif sont représentées en rouge ($\Delta l \geq 0.5 \text{ \AA}$), en orange ($\Delta l \geq 0.05 \text{ \AA}$) et en jaune ($\Delta l < 0.05 \text{ \AA}$).	85
4.28	Représentation graphique des variations de la distance séparant les C_α des résidus Asp25 et Asp25' en réponse au pincement de tous les couples d'atomes de la molécule. La légende à droite indique des variations allant de $-0.8 \text{ \AA}$ à $1.4 \text{ \AA}$ du rouge vers le bleu.	86
4.29	Représentation des couples induisant une réponse du site actif, évaluée par la variation de la distance Asp25-Asp25'. Sont représentées en rouge les séquences correspondant aux quatre bandes de la figure 4.28 numérotées de 1 à 4.	87
4.30	Représentation des couples induisant une réponse du site actif, évaluée par la variation de la distance Asp25-Asp25'. Les trois molécules, identifiées par la notation a, b, c, correspondent aux bandes respectives du graphe 4.28. La séquence représentée en rouge indique les résidus de la bande à pincer avec les séquences bleues.	87
4.31	Comparaison des variations de distance entre les résidus Asp25 et Asp25' de la protéase fermée (à gauche) et ouverte (à droite) après pincement des couples d'atomes.	88
4.32	Variations maximales de la distance séparant les C_α des résidus Ile50 et Asp25 après optimisation des forces pour chaque atome. A droite, les zones de la protéase induisant une ouverture du volet sont représentées en rouge ($\Delta l \geq 1 \text{ \AA}$), en orange ($\Delta l \geq 0.2 \text{ \AA}$) et en jaune ($\Delta l < 0.2 \text{ \AA}$).	89
4.33	Représentation des variations de la distance séparant les C_α des résidus Ile50 et Asp25 en réponse au pincement de tous les couples d'atomes de la molécule. La légende à droite indique des variations allant de $-1.5 \text{ \AA}$ à $2.5 \text{ \AA}$ du rouge vers le bleu.	90
4.34	Représentation des couples induisant une variation de la distance Ile50-Asp25. Sur la première molécule, représentation des couples induisant une variation de la distance Ile50-Asp25. En rouge : les séquences correspondant aux quatre bandes de la figure 4.33 numérotées de 1 à 4. Les molécules, notées de a à h, correspondent aux bandes respectives du graphe 4.33. La séquence rouge indique les résidus de la bande à pincer avec les séquences bleues.	91
4.35	Comparaison des variations de distance entre les résidus Ile50 et Asp25 de la protéase fermée (à gauche) et ouverte (à droite) après pincement des couples d'atomes.	92
4.36	A gauche, représentation de la protéase complexée avec un inhibiteur potentiel type, contenant une urée cyclique (1HVR dans la Protein Data Bank). A droite est présenté un zoom sur le site actif ainsi que les quatre forces considérées entre le ligand et les résidus Ile50, Ile50', Asp25 et Asp25'.	93

4.37	Fermeture des volets obtenue par l'application de quatre forces sur les C_α des résidus Ile50, Ile50', Asp25 et Asp25', en direction d'un inhibiteur potentiel à urée cyclique (voir figure 4.36).	94
A.1	Composition chimique d'un acide aminé.	102
A.2	Formation d'une liaison peptidique entre deux acides aminés.	102
A.3	Angles de torsion Φ et Ψ , caractéristiques d'une liaison peptidique.	103
A.4	Représentation d'une hélice α	103
A.5	Représentation d'un feuillet β	104
B.1	Cliché d'une molécule d'ADN par diffraction de rayons X réalisé par Rosalind Franklin au King's College de Londres en 1953. Il confirme qu'une molécule d'ADN forme une hélice d'un diamètre de 2 nm et d'un pas de 3.4 nm.	108
B.2	Structures du ribose et du désoxyribose.	109
B.3	Composition chimique des bases puriques et pyrimidiques.	109
B.4	Appariement des bases par liaisons hydrogène.	109
B.5	Structure d'une chaîne d'acide nucléique. A droite, localisation des angles de torsion du backbone de α à ζ . <i>Source Springer-Verlag</i> [360].	110
B.6	A gauche : représentation schématique d'une double hélice d'ADN. A droite : structures des formes A, B et Z respectivement.	111
C.1	Schéma d'un système de points matériels représentés en position d'équilibre r_0 et déplacés vers une position r . Leur mouvement est associé à celui d'un oscillateur harmonique représenté par un ressort de constante de raideur k	116

Liste des tableaux

1.1	Exemple de composition moléculaire d'une bactérie. <i>Source de Boeck</i> [44] . . .	11
1.2	Longueurs et énergies d'une liaison covalente simple, double et triple établie entre deux atomes de carbone.	16
1.3	Electrons de valence, électronégativité et possibles liaisons de covalence des atomes d'hydrogène, carbone, oxygène, azote, phosphore et soufre.	17
4.1	Angles du sucre d'une désoxythymidine monophosphate seule et en duplex. .	73
4.2	Moyennes de l'ouverture des quatre liaisons hydrogène sur l'ensemble des couples pincés : à gauche pour le duplex d'ADN naturel, à droite pour le duplex d'ADN contenant le dinucléotide contraint (g-).	79
4.3	Moyennes de l'ouverture des quatre liaisons hydrogène sur l'ensemble des couples pincés : à gauche pour l'hétéroduplex ADN/ARN naturel, à droite pour l'hétéroduplex ADN/ARN contenant le dinucléotide contraint (g-).	79
A.1	Tableau récapitulatif des 20 acides aminés.	105
B.1	Comparaison de quelques caractéristiques structurales des formes A, B et Z. <i>Source Science</i> [380].	112
B.2	Tableau récapitulatif des bases puriques et de leurs composés.	112
B.3	Tableau récapitulatif des bases pyrimidiques et de leurs composés.	113

Introduction générale

Le Projet du Génome Humain (The Human Genome Project [10]) a permis de décrypter la séquence de presque 30 000 gènes mais n'a pas donné d'information directe sur leur fonction. Aujourd'hui, la protéomique et les multiples techniques expérimentales ou calculatoires qui l'accompagnent, prennent le relais pour extraire la fonction des protéines codées, via la connaissance de leur structure et de la manière dont elles interagissent avec d'autres molécules. Comprendre leur fonction et l'intégrer dans l'extraordinaire complexité des processus cellulaires dépend de notre capacité à "cartographier" le réseau des interactions qui existent et sont mises en jeu.

Les interactions intra et intermoléculaires sont décisives lors de la formation des complexes biomoléculaires et pour assurer la viabilité des cellules. En considérant par exemple qu'une protéine peut établir en moyenne 5 interactions différentes, on peut donc estimer à 150 000 au moins le nombre d'interactions dans les cellules humaines [6]. Comprendre et prédire les mécanismes d'interactions macromoléculaires ainsi que leurs possibles modifications avec le temps, le milieu, ou encore les maladies, représente donc un enjeu inestimable dans le domaine de la santé et de la recherche pharmaceutique.

Etant donnée l'ampleur de l'étude, des moyens d'analyse et de caractérisation des interactions à grande échelle sont nécessaires : on parle aujourd'hui d'"interactome" pour décrire l'organisation du protéome [7] jusqu'au niveau cellulaire. En se maintenant au niveau moléculaire, la biologie s'appuie dans un premier temps sur les méthodes expérimentales pour identifier des protéines, leur structure et leur affinité [1][16]. Un exemple de réussite est le recensement en 2003 de 4780 interactions chez la *Drosophila Melanogaster* impliquant 4679 protéines [8]. Une telle approche ne peut cependant pas convenir à l'étude d'un grand nombre d'interactions, toutes les interactions moléculaires ne pouvant être identifiées par des études expérimentales, longues et coûteuses. La voie informatique présente une bonne alternative aux limites expérimentales : la recherche *in silico* est actuellement extrêmement active, son essor laisse présager son importance et son caractère incontournable dans les années à venir.

Les méthodes informatiques utilisées à l'heure actuelle ont deux objectifs : d'une part étudier la structure tridimensionnelle et les changements conformationnels d'une molécule, et d'autre part simuler les mécanismes d'amarrage moléculaire, ou *docking*. Cependant, le défi que se lancent les modélisateurs, à savoir atteindre un niveau de prédiction, demeure pour le moment inaccessible. La simulation se heurte en effet à des problèmes calculatoires (la taille des systèmes biologiques limite les méthodes informatiques) mais également à des problèmes de représentation des systèmes.

La principale difficulté, qui est parallèlement la clé de la compréhension du monde biologique, est l'introduction de la flexibilité moléculaire, en d'autres termes la souplesse conformationnelle, dans les algorithmes. La flexibilité des molécules biologiques résulte de la haute dimensionnalité des systèmes, mais aussi de l'intervention d'un grand nombre d'interactions faibles (van der Waals, liaisons hydrogène, effet hydrophobe, ...). On sait aujourd'hui qu'elle fait partie intégrante de la dynamique des macromolécules et de leur bon fonctionnement. La négliger dans les calculs introduit des erreurs, il convient donc de déterminer la flexibilité intrinsèque d'une molécule mais aussi de pouvoir l'introduire, au moins partiellement, dans une procédure de docking [2][3][4][18].

L'étude du docking est une étape capitale dans la compréhension des réactions biologiques et dans la conception de médicaments. Les tentatives de docking ont longtemps été basées sur le concept *clé-serrure*, ne considérant que des corps rigides capables d'interagir lorsqu'ils présentent une compatibilité géométrique parfaite. Cependant, depuis ces premiers pas [12], une idée s'est peu à peu imposée : les molécules ne sont pas figées, mais peuvent subir des modifications conformationnelles lors d'une interaction. Les procédures ont donc évolué vers une introduction de la flexibilité, au moins partielle, des systèmes mis en jeu.

De nombreuses approches ont été et sont actuellement développées [17][19]. Elles se décomposent toutes en trois phases : représentation du système, algorithme de recherche et algorithme de score. Par simplification, un grand nombre d'algorithmes de docking font l'hypothèse que les molécules sont rigides, mais donnent alors des résultats très approximatifs. Les tentatives nécessaires de prise en compte de la flexibilité augmentent considérablement la complexité du système. On peut définir trois modèles impliquant différents degrés de précision : rigide (les deux molécules sont considérées comme rigides), semi-flexible (une rigide et l'autre flexible), flexible (les deux flexibles). Le modèle semi-flexible est souvent appliqué dans les interactions protéine-ligand, où seul le ligand est considéré comme flexible. L'enjeu majeur est à présent d'introduire les déformations du récepteur, généralement une protéine. Les protéines sont en permanence en mouvement entre des états conformationnels proches en énergie, ce qui rend le traitement du docking extrêmement complexe. Pourtant, la capacité d'accommodation conformationnelle des protéines permet d'accroître l'affinité avec le ligand. Les modèles semi-flexibles peuvent donner de bons résultats pour à peu près 50% des applications [5]. Un tel succès se produit lorsque le site de liaison du récepteur est suffisamment rigide et que la structure de départ est proche de la configuration de la protéine dans le complexe. Cependant, de nombreux systèmes présentent des mouvements significatifs lors de l'interaction et le docking peut être affecté même par de légères déformations, au niveau des chaînes latérales par exemple.

Actuellement, de lourds efforts sont menés pour développer de nouvelles stratégies et l'on peut imaginer que dans le futur, la flexibilité du récepteur sera exprimée dans son intégralité, tout comme celle du ligand l'est aujourd'hui. Parmi ces stratégies, on peut citer l'utilisation de méthodes Monte Carlo [9] ou de dynamique moléculaire [15], les librairies de rotamères [13], ou encore l'utilisation de structures multiples [11][14].

La principale limitation est due à la durée des calculs, trop importante à ce jour pour explorer efficacement et de façon systématique l'intégralité de l'espace des conformations. En outre, de manière générale, on peut constater que certains algorithmes permettent de traiter convenablement un type de docking, mais qu'aucun, à l'heure actuelle, n'est performant sur l'ensemble.

Enfin, l'évaluation des solutions potentielles à l'issue de la procédure de recherche s'effectue par des algorithmes de score basé sur divers critères (complémentarité géométrique, recouvrement inter ou intramoléculaire, liaisons hydrogène, taille de la surface de contact, ...). Ici encore, de nombreuses difficultés apparaissent, au niveau de l'ajustement des critères ou du nombre important de possibles solutions discriminées par l'algorithme, et qui ne contiennent pas toujours les formes natives recherchées.

Ainsi, pour des raisons de rendement de criblage, la tendance est de simplifier la description des systèmes et d'introduire un nombre limité de mouvements possibles afin de conserver un temps de calcul minimum.

"Despite the very promising picture drawn, molecular docking still holds several hidden weaknesses, and the so-called docking problem is far from being solved. The lack of a suitable scoring function, able to efficiently combine both accuracy and speed, is perhaps the most detrimental weakness. The results of a docking experiment should therefore not be taken as the final result of a structural study, but rather as a good starting point for a deeper and more accurate analysis. In this sense, docking must be necessarily fast, enabling large quantities of data to be considered, and reasonable and coherent solutions to be generated. However, the final result (geometry, binding free energy) should always be determined by a more accurate and precise methodology, naturally slower."

Sousa et al.. Proteins, 2006 [17].

Cette thèse a pour but de présenter une alternative à l'ensemble des méthodes utilisées jusqu'alors pour traiter la flexibilité des biomolécules. Il serait aujourd'hui capital de pouvoir prédire un couplage correct, répondant aux observations et aux fonctionnalités biologiques connues des macromolécules, via un outil informatique performant, permettant en un temps raisonnable d'explorer l'ensemble des possibilités de déformation et d'assemblage optimum entre deux molécules.

Dans ce but, nous avons développé une nouvelle méthode : la méthode des Modes Statiques, en partant du constat que les déformations d'une molécule isolée correspondent à celles que l'on retrouve lors de la formation d'un complexe. Nous nous proposons donc de retrouver les états de déformation possibles d'une molécule en nous appuyant sur la propriété de flexibilité des molécules biologiques et recensons l'ensemble des changements de conformation pouvant se produire. Ces derniers, une fois stockés, peuvent être exploités directement pour évaluer la flexibilité d'une molécule, ou utilisés dans une procédure ultérieure de docking.

La technique développée dans ce projet doit permettre une analyse performante des structures et des interactions dans le domaine de la modélisation moléculaire. Les champs d'applications potentiels sont très vastes et notre outil pourra permettre de progresser dans la compréhension, l'analyse ou la prédiction :

- des changements conformationnels des biomolécules
- de la réponse d'un site actif à un stimuli extérieur
- de la cartographie des zones rigides et flexibles d'une molécule

- de la modification possible de la flexibilité, du blocage de certains sites
- de la corrélation entre domaines
- de l'identification de sites allostériques
- ...

Nous présentons nos travaux dans ce manuscrit, au cours de quatre chapitres :

Le premier chapitre introduit notre approche de la biologie à l'heure actuelle, à une époque où les progrès techniques permettent de sonder la matière à l'échelle microscopique, au travers de l'étude des événements atomiques et moléculaires. Nous présentons donc une description des molécules d'intérêt biologique ainsi que des liaisons chimiques impliquées dans leurs structures et leurs interactions. Cette description est accompagnée d'un état des lieux des applications courantes ainsi que d'une perspective de la pénétration des connaissances en biologie moléculaire dans des domaines de pointe émergents tels que les nanotechnologies.

Dans un second chapitre, nous nous intéressons à l'apport de la modélisation atomique et moléculaire à l'étude des mécanismes biologiques. Grâce à des efforts multi-disciplinaires et aux progrès informatiques, en s'appuyant sur des outils existants dédiés au calcul et sur le développement de techniques propres à la biologie, la modélisation devient l'outil de prédilection pour comprendre, voire prédire, les propriétés et le comportement des molécules biologiques. Elle peut ainsi assister et orienter l'expérience longtemps privilégiée, souvent efficace, mais aussi limitée et coûteuse. Nous passons en revue les techniques générales de la recherche *in silico* avant de détailler leurs applications en biologie, particulièrement pour le problème du docking. Nous abordons enfin les difficultés qui limitent la prise en compte de la flexibilité des macromolécules.

Le troisième chapitre est consacré à la présentation de la méthode des Modes Statiques. Nous justifions tout d'abord son intérêt et son apport par rapport aux méthodes pré-existantes de calcul de flexibilité moléculaire, principalement la dynamique moléculaire et l'analyse par modes normaux. Par la suite, nous détaillons le principe de la méthode des Modes Statiques, et développons les étapes de l'algorithme qui a fait l'objet d'un logiciel : FLEXIBLE. Enfin, nous ouvrons la discussion sur les évolutions possibles de l'approche et de l'algorithme actuel.

Dans le quatrième chapitre, nous présentons les résultats obtenus avec le programme FLEXIBLE pendant cette thèse. Ces exemples sont variés : depuis l'étude d'un polymère thermosensible, le PNIPAM, jusqu'à l'exploration des protéines telles que la protéase du VIH-1, en passant par de courtes séquences d'acides nucléiques. Ces cas nous ont permis d'explorer diverses applications des modes statiques sur l'étude de molécules isolées : la cartographie de régions flexibles, l'effet de contraintes extérieures sur cette flexibilité, sur un site actif, la recherche de zones sensibles où appliquer les contraintes pour optimiser une déformation, la recherche de sites allostériques, l'effet de l'amarrage d'un ligand, la corrélation entre domaines d'une protéine, la possibilité de "pinces moléculaires" pour induire une réponse donnée.

Chapitre 1

La problématique des interactions en biologie

Sommaire

1.1	Enjeux socio-économiques	6
1.1.1	La biologie aujourd'hui et ses applications	6
1.1.2	L'importance de la biochimie, une science récente	6
1.1.3	La réponse de la science aux questions de la biochimie	8
1.1.4	Enjeux et impact sur la société	9
1.2	Les objets moléculaires en biologie	10
1.2.1	Les lipides	12
1.2.2	Les glucides	12
1.2.3	Les peptides et les protéines	13
1.2.4	Les acides nucléiques	14
1.2.5	L'eau	15
1.3	Interactions	15
1.3.1	Les liaisons chimiques fortes	16
1.3.2	Les liaisons chimiques faibles	18
1.3.3	Implication des interactions dans la structure et la flexibilité	20
1.4	Conclusion	21

Chaque être vivant est constitué d'unités structurales élémentaires : les cellules [45]. Chaque cellule est elle-même une entité vivante et fonctionnelle, autonome mais coordonnée avec son environnement. Elle a la capacité de grandir, de se reproduire, de recevoir et d'émettre des signaux, de traiter des informations... Toute son activité est régie par de multiples réactions chimiques et de fines interactions inter et intramoléculaires.

C'est à ce niveau qu'interviennent la biologie moléculaire, la biologie structurale et la biochimie : dans la compréhension du vivant, par l'étude d'événements moléculaires, depuis la formation, la structure et la fonction des molécules biologiques jusqu'à leur implication dans les processus chimiques et les assemblages macromoléculaires.

La compréhension de tels phénomènes et leurs applications allant de la médecine aux nanotechnologies, passent donc par une étude approfondie des interactions chimiques et non chimiques mises en jeu lors des processus moléculaires.

1.1 Enjeux socio-économiques

1.1.1 La biologie aujourd'hui et ses applications

La biologie concentrant ses problématiques sur la nature de la vie, son étude a toujours été accompagnée de nombreuses et diverses théories philosophiques et scientifiques, depuis les penseurs grecs jusqu'aux chercheurs modernes.

D'abord fondée sur l'observation, elle a connu des progrès tardifs mais fulgurants depuis la renaissance, avec l'arrivée de nouvelles méthodes expérimentales telles que la microscopie permettant la découverte d'un monde invisible à l'oeil nu.

La biologie ne connaît pas de longue tradition expérimentale. Ses acteurs ont long mené une "enquête naturaliste" basée sur l'observation de phénomènes, de comportements, de l'anatomie... Elle s'est aujourd'hui profondément renouvelée et étendue grâce à l'incorporation de théories (évolution...), de concepts ou de méthodes souvent expérimentales, développés par d'autres disciplines (physique, chimie, informatique).

Au-delà de sa vocation explicative, elle se veut désormais applicative, fonctionnelle et même prédictive. En se diversifiant, elle s'est organisée en différents niveaux d'étude : de la molécule (chimie organique, biochimie, biologie moléculaire) aux écosystèmes et à la biosphère (écologie, phylogénie) en passant par les cellules (biologie cellulaire, microbiologie...), les organismes (biologie des organismes, anatomie...) et les populations (génétique des populations, systématique...).

La diversité de ces domaines rend compte de la variété de nouveaux axes de recherche fondamentale mais aussi de possibles applications présentes dans le quotidien de l'être humain : la biologie se met au service de la santé, de l'agronomie, mais aussi de l'écologie, du développement durable...

1.1.2 L'importance de la biochimie, une science récente

Lors de l'application des découvertes aux multiples applications humaines, une difficulté est la transposition des résultats obtenus sur un organisme à un autre organisme, un autre milieu, voire un milieu inorganique afin de l'intégrer de manière fonctionnelle.

Les progrès de la chimie organique et de la biochimie par exemple, ont permis de sonder la matière afin d'en connaître les propriétés intrinsèques : les connaissances physico-chimiques des molécules impliquées dans les processus du vivant sont aujourd'hui le fil conducteur permettant l'intervention de l'homme dans ces processus, ou la transposition à de nouvelles technologies.

On peut situer le début de la biochimie en 1828, lorsque Wöhler synthétise pour la première fois de l'urée, une substance organique sécrétée par les êtres vivants à partir de cyanate d'argent. La théorie du vitalisme [21][22], jusque là admise, entame son déclin. Distinguant les êtres vivants de la matière inerte en leur associant une force vitale différente de l'âme [28], et en donnant ainsi à la biologie une indépendance par rapport aux autres sciences, le vitalisme laisse place à une conception plus matérialiste de la vie qui lève l'indétermination : on admet désormais que les règles physico-chimiques des êtres vivants sont les mêmes que celles régissant la matière inanimée.

"Reconnaître l'unité des processus physico-chimiques au niveau moléculaire, c'est dire que le vitalisme a perdu toute fonction."

François Jacob. La logique du vivant, 1970 [37].

Cette nouvelle vision modifie totalement le rapport de l'homme aux sciences de la vie, et par conséquent, leur place dans notre société et nos préoccupations [49]. Le XIX^{ème} siècle assiste ainsi à la rationalisation des savoirs et des observations. Les progrès de la physique et de la chimie permettent le triomphe des méthodes expérimentales. En 1864, Claude Bernard publie *L'introduction à la médecine expérimentale* et confirme la déroute du vitalisme. Il propose l'existence de lois naturelles rigoureuses gouvernant la physiologie tout comme la physique et la chimie, et étant la cause de l'existence de tout phénomène vivant.

L'explication physico-chimique de la vie n'est cependant pas encore universellement acceptée, notamment en ce qui concerne la structure et la formation des organismes individuels, on parle encore d'une "idée directrice", de "génération spontanée". Des expériences sur la régénération des cellules, ou encore les expériences de Pasteur sur la fermentation, montrent que les êtres vivants ne se forment pas "spontanément" à partir de matière inorganique. Au contraire, ils se reproduisent, adoptant une structure interne et une forme liées à leur composition chimique, leurs propriétés physiologiques et morphologiques étant transmises par hérédité, bien que les mécanismes de l'hérédité demeurent encore très obscurs.

Aujourd'hui, avec l'avancée des technologies modernes d'exploration et de conception, ainsi que des moyens de calculs, les grandes problématiques de la biologie atteignent l'échelle moléculaire et atomique.

De nombreuses questions trouvent peu à peu leur réponse :

Quelles sont les structures chimiques et tridimensionnelles des molécules biologiques ? Dans quelle mesure leur fonction change-t-elle avec cette structure ? Quels sont les mécanismes de reconnaissance, d'assemblage ? Quels mécanismes inter ou intramoléculaires leur permettent de fonctionner ? Par exemple, comment une enzyme réalise-t-elle la catalyse ? De quelle façon son activité est-elle régulée ? Comment une protéine peut-elle recevoir ou émettre un signal ? Comment peut-on exprimer et transmettre l'information génétique ?

"Les êtres vivants sont des machines chimiques. La croissance et la multiplication de tous les organismes exigent que soient accomplies des milliers de réactions chimiques grâce à quoi sont élaborés les constituants essentiels des cellules."

Jacques Monod. Le hasard et la nécessité, 1970 [49].

1.1.3 La réponse de la science aux questions de la biochimie

Approche expérimentale

Depuis le milieu du XX^{ème} siècle, avec la découverte de la structure de l'ADN puis le séquençage du génome, la quantité de molécules répertoriées ne cesse de croître. On estime qu'il existe environ 100 000 protéines chez l'homme, dont un millier possède une structure tridimensionnelle connue [50], et dont seulement une centaine est employée comme cible pour les médicaments. Les techniques expérimentales permettent d'identifier à haut débit les structures tridimensionnelles des molécules biologiques. Elles contribuent ainsi à l'effort mené sur l'étude de la fonction de ces molécules, de leurs mécanismes d'interactions et de leur implication dans l'activité normale ou pathologique des cellules.

Deux méthodes sont généralement utilisées pour déterminer les structures tridimensionnelles à haute résolution des macromolécules biologiques [24]: la cristallographie par diffraction des rayons X [25][38] et la spectroscopie par résonance magnétique nucléaire (RMN) [40][46]. De très nombreuses structures sont mises à la disposition de l'utilisateur : la *Protein Data Bank* [50] par exemple, recense actuellement près de 55000 structures de molécules, dont 45000 ont été résolues par rayons X et 7500 par RMN [23]. Plus de 80% des structures sont résolues par diffraction des rayons X. Cette méthode, contrairement à la spectroscopie RMN, n'est pas limitée par la taille des molécules. Elle nécessite néanmoins une cristallisation préalable qui peut présenter des difficultés. La RMN quant à elle peut fournir des réponses sur la dynamique. Notons que la résolution de complexes de macromolécules demeure limitée : en 2006, elle ne représentait que 1.5% des structures disponibles [31] [51].

De multiples obstacles doivent être franchis par l'expérimentateur, notamment dans le choix de la méthode qui doit convenir à la propriété recherchée, ainsi qu'à l'échelle de temps du phénomène (qui peut aller de la femtoseconde (10^{-15} s) à l'heure [47]). Il existe donc de nombreuses autres méthodes expérimentales selon la propriété structurale ou physico-chimique que l'on cherche à identifier [29][53] ; la figure 1.1 présente un exemple d'éléments recherchés sur une protéine ainsi que la technique associée.

Approche théorique

La résolution de la structure des molécules biologiques représente une première étape de la connaissance des processus biologiques. Elle montre que ces molécules présentent certaines caractéristiques, comme une structure hautement complexe mais ordonnée, ainsi qu'une fonction spécifique, utilisant des énergies extraites de l'environnement et transformées. Elle ne délivre cependant pas tous les éléments requis pour une connaissance complète des processus biologiques ; le repliement des protéines, par exemple, est trop rapide pour être observé.

Pour cette raison, mais aussi parce que l'expérimentation est lente et coûteuse, l'informatique, par le biais de la modélisation moléculaire, est un nouveau pilier de la biologie structurale. Ces dernières années, les moyens informatiques (mémoire, rapidité, algorithmes, infographie) se sont considérablement développés, permettant la naissance de la bioinformatique, de la chimie et de la biologie informatique.

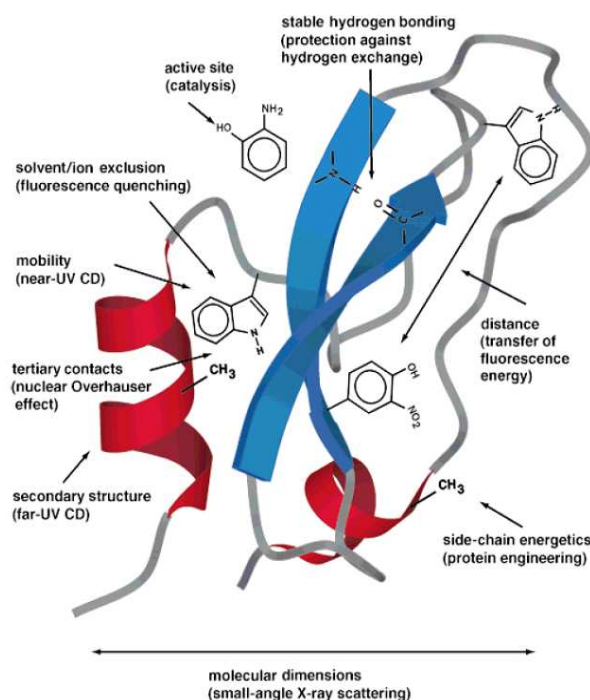


FIG. 1.1 – Exemple d'éléments identifiables par l'expérience sur une molécule biologique et de la technique associée. Source *Angew. Chem. Int. Ed.* [29]

La recherche *in silico* combine de multiples aspects de la physique, de la chimie, des statistiques, de l'infographie, afin, par exemple, d'évaluer et d'interpréter des propriétés physico-chimiques jusque-là inobservables ou incomprises, d'assister ou d'orienter l'expérience. Les approches théoriques et expérimentales deviennent complémentaires. Le grand enjeu de la modélisation demeure bien entendu la prédiction des interactions moléculaires, ou *docking*.

1.1.4 Enjeux et impact sur la société

Ces développements débouchent sur des progrès quotidiens de la biologie dont l'impact sur la société est de ce fait considérable. La compréhension des mécanismes moléculaires de maladies permet la conception et l'optimisation de produits pharmaceutiques, la compréhension de maladies conformationnelles (Alzheimer, Parkinson, maladies à prions, amyloses...), le dépistage d'anomalies génétiques, mais aussi la modification d'êtres vivants avec des applications dans le domaine agro-alimentaire, le clonage [52]...

Ces avancées deviennent également la source de nouvelles technologies en plein essor : les nanobiosciences [35][36]. Elles se situent à l'interface de la physique, la chimie, la biologie et l'ingénierie et présentent de nombreuses perspectives dans des domaines tels que la médecine (diagnostic avec les biopuces, thérapie, organes artificiels), la visualisation et la manipulation de molécules uniques (nanoparticules magnétiques), l'imagerie biologique (nanoparticules fluorescentes), la microfluidique (nanoleviers, nanosondes...) ou encore le nano-adressage.

Aujourd'hui, les outils bioinformatiques font parler les structures recueillies par les expérimentateurs afin d'en inférer des fonctions biologiques présentant un intérêt (économique, industriel, santé publique, agroalimentaire). Les biomolécules entrent alors en jeu pour la fonctionnalisation de surface (biopuces, biocapteurs...), la détection de cibles impliquées dans les fonctions biologiques (diagnostic précoce). Tous ces dispositifs nécessitent l'élaboration de molécules, parfois même leur conception *de novo* par la biologie de synthèse [54] pour arriver à des systèmes présentant de nouvelles propriétés physico-chimiques adaptées à une fonction dans la nanoingénierie.

Nous nous trouvons donc à un carrefour où l'homme puise dans ses connaissances de la nature pour recréer un monde répondant à ses besoins. Ces indiscutables progrès suscitent des interrogations éthiques, voire des inquiétudes qui sont aujourd'hui largement discutées, aussi bien par la communauté scientifique que par les comités d'éthique ou les philosophes. Ces discussions doivent être prises en compte telles un accompagnement du progrès [30][41][42][48].

Un rapport récent de *National Academies of Sciences* [26] annonce l'intégration des fonctions biologiques dans de nouveaux matériaux comme synonyme d'un apport de qualité de vie et d'avancées en terme technologique mais aussi social et économique. Un exemple d'actualité serait la contribution de la chimie de synthèse au problème de l'énergie afin de réduire notre dépendance aux produits pétroliers.

Un tel progrès a été et ne sera rendu possible que par la poursuite d'investissements et d'efforts multidisciplinaires, de développement d'outils faisant face à la complexité de l'exploration expérimentale (détermination de structure, de dynamique, imagerie de molécule unique...), théorique (complexité fondamentale), informatique (taille des systèmes, simulations multi-échelle), de manipulation (synthèse de nouvelles molécules, manipulation et suivi à l'échelle atomique).

"The exciting state of research in biomolecular processes and materials has been powered by new experimental and computational tools for interrogating complex systems at a high level of details"... "Further advances in the development and application of these tools are crucial to the advancement of the field".

Committee on Biomolecular Materials and Processes, 2008 [26].

1.2 Les objets moléculaires en biologie

Nous avons jusqu'à présent introduit l'importance de l'étude des interactions en biologie pour parvenir à une connaissance approfondie des mécanismes du vivant. Nous nous intéressons à présent aux principaux types de molécules biologiques essentielles à la structure et à la fonction des cellules. Elles sont au coeur des propriétés biologiques de par leurs multiples fonctions, telles que l'apport en énergie, la contribution à l'information, le support et la transmission de l'information génétique.

Les molécules biologiques sont principalement constituées d'atomes de carbone (en raison de leur grande capacité à former différentes liaisons covalentes stables et des structures variées), d'hydrogène, mais aussi d'oxygène, d'azote et de soufre (dans le cas des protéines) ainsi que de phosphore (dans le cas des acides nucléiques).

On peut distinguer quatre grandes familles de molécules biologiques : les glucides, les lipides, les acides aminés et protéines, les nucléotides et acides nucléiques. Elles-mêmes comportent deux types de molécules :

- les petites molécules : sucres, acides gras, acides aminés et nucléotides qui transportent l'énergie, transmettent les signaux et s'assemblent en macromolécules
- les grosses molécules, polymères des ces petites molécules : respectivement les polysaccharides, les lipides, les protéines et les acides nucléiques. Certaines de ces molécules comptent des milliers d'atomes et pour cette raison, sont souvent qualifiées de macromolécules. Elles présentent donc une complexité incroyable, mais peuvent être plus facilement décrites lorsqu'on s'intéresse à leur structure architecturale. En effet, à l'exception des lipides qui forment une famille plus hétérogène, les molécules biologiques sont des polymères, édifices constitués de motifs unitaires et répétés, reliés par liaisons covalentes

Ces éléments unitaires sont polarisés, conférant à la macromolécule une polarité structurale. La séquence d'enchaînement des sous-unités, ainsi que leur orientation, font des molécules biologiques des molécules porteuses d'une information lisible. Elles n'ont cependant pas toutes le même potentiel : les acides nucléiques et les protéines portent des informations particulièrement riches. La séquence est donc un premier élément d'identité. Le second élément capital est l'architecture tridimensionnelle des édifices supramoléculaires. C'est à ce niveau qu'intervient la complexité des molécules biologiques : la chaîne polymérique les constituant peut s'enrouler dans l'espace en créant une architecture très complexe, unique, ordonnée et spécifique, dont découle la fonction et la reconnaissance de la molécule. Avant de nommer les interactions atomiques participant au repliement des chaînes dans la section suivante, nous allons donner une brève description de la nature des molécules biologiques.

Composants	Pourcentage de la masse cellulaire
<i>Inorganiques</i>	
Eau	70
Ions	1
<i>Organiques</i>	
Sucres	3
Acides aminés	0.5
Nucléotides	0.5
<i>Grandes molécules</i>	
Protéines	15
ADN	0.5
ARN	6
Polysaccharides	2
Lipides	2

TAB. 1.1 – Exemple de composition moléculaire d'une bactérie. *Source de Boeck [44]*

1.2.1 Les lipides

Les lipides constituent une famille de molécules hétérogènes [56]. Leur point commun est leur propriété hydrophobe. Leur solubilité est nulle ou faible dans l'eau mais élevée dans les solvants organiques non polaires (méthanol, chloroforme, acétone...). Les lipides jouent un rôle dans le stockage de l'énergie et la structure des membranes cellulaires.

Les lipides sont des molécules soit complètement apolaire, soit bipolaire, avec une tête polaire et une queue apolaire. Ils comptent des molécules variées résultant à un niveau plus ou moins complexe de l'association d'acide gras (chaîne carbonée terminée par un groupe carboxyle, cf. Fig.1.2) et d'alcool. Parmi ces molécules, on peut citer les lipides vrais (condensation d'acides gras avec des alcools) eux-mêmes subdivisés en lipides simples neutres (glycérolipides, cérides, stérides), les lipides complexes (contenant également du phosphore, de l'azote, du soufre...) et les composés à caractère lipidique (dérivés de lipides précédents). Les lipides peuvent s'associer et participer à des édifices supramoléculaires.

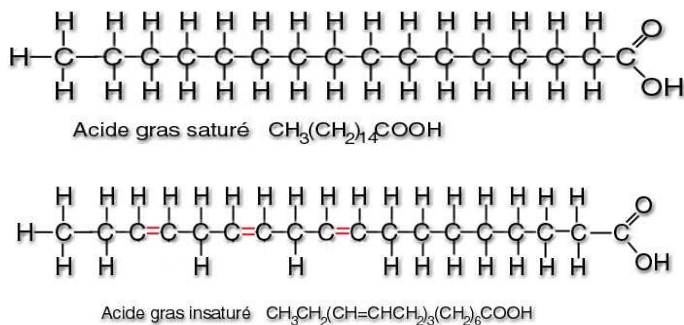


FIG. 1.2 – Exemple de la structure chimique de deux acides gras.

Les autres molécules biologiques sont des polymères de sous-unités structurales identiques ou semblables, les monomères, reliés les uns aux autres par des liaisons covalentes créées par réaction de déshydratation (perte d'une molécule d'eau, cf. Fig. 1.3) :

1.2.2 Les glucides

Les glucides constituent la principale réserve d'énergie pour les êtres vivants [55]. Ils jouent également un rôle structural, par exemple dans les parois des cellules, et participent à de nombreux processus métaboliques. La famille des glucides comprend les sucres simples dits monosaccharides (chaînes carbonées linéaires ou cycliques, de formule $(\text{CH}_2\text{O})_n$, n allant de 3 à 7, cf. Fig. 1.4), et des sucres plus complexes, les polysaccharides, eux-mêmes formés de sucres élémentaires liés par covalence.

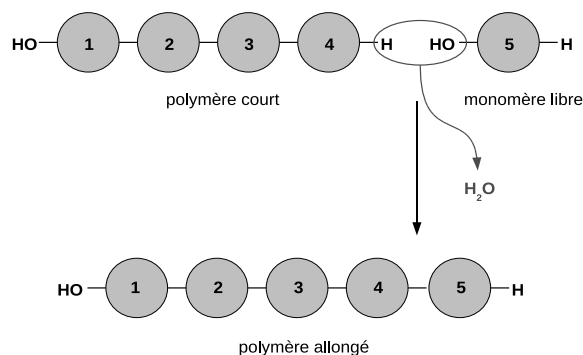


FIG. 1.3 – Formation d’une chaîne de polymère par réaction de déshydratation.

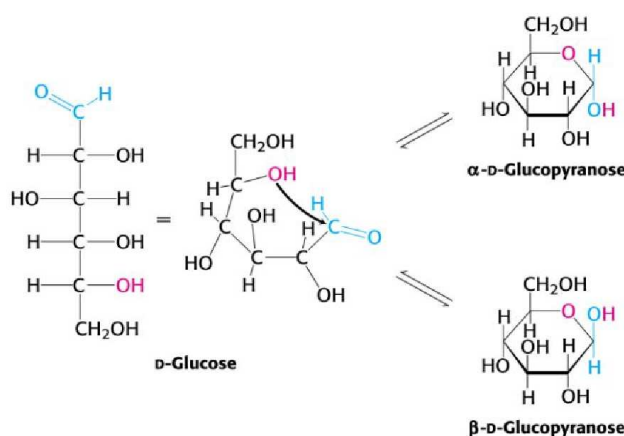


FIG. 1.4 – Exemple de monosaccharide : le glucose, en représentation linéaire et cyclique.

1.2.3 Les peptides et les protéines

Les peptides et les protéines sont les constituants majeurs des tissus vivants¹. Outre leur importance quantitative, ce sont des molécules de premier ordre participant à toutes les fonctions cellulaires (immunité, régulation, transport, mouvement, énergie, catalyse, structure...). La variété de ces fonctions tient à la variété des formes existantes, qui, malgré la complexité des molécules, sont bien définies dans un environnement particulier.

Ces molécules sont formées d’une ou plusieurs chaînes d’acides aminés reliés par liaison peptidique selon une séquence déterminée. On parle de protéine lorsque le nombre d’acides aminés dépasse la centaine, et de peptide en deçà.

¹La structure des protéines est traitée plus en détails dans l’annexe A.

Chaque acide aminé comporte un carbone α asymétrique (à l'exception de la glycine) permettant deux stéréoisomères de configuration, lié à un groupe amine, un groupe carboxyle et un radical propre à chaque acide aminé. Il existe 20 acides aminés pouvant être classés selon la structure ou les propriétés physico-chimiques de leur chaîne latérale. On peut citer les acides aminés polaires, non polaires, chargés ou non chargés. Notons, que les chaînes latérales, de part leur polarité notamment, jouent un rôle capital dans le repliement de la protéine.

L'analyse structurale (par cristallographie à rayons X) a mis en évidence, dans la plupart des protéines, certains motifs réguliers, enroulés ou pliés, dont les hélices α et les feuillets β qui résultent de liaisons hydrogène répétées entre les $-CO$ et les $-NH$ peptidiques. De tels repliements sont à la fois rendus possibles et limités par les capacités de torsion du squelette (encombrement stérique...).

La conformation, ou structure tridimensionnelle des protéines, permet de décrire ces motifs locaux, mais aussi leur arrangement dans l'espace par d'autres repliements permettant d'atteindre un minimum énergétique. Le repliement de la chaîne polypeptidique est stabilisé par de multiples interactions entre les chaînes latérales d'acides aminés éloignés les uns des autres : les liaisons hydrogène, les interactions hydrophobes, de van der Waals, ioniques, les ponts disulfure (Fig.1.5). C'est la conformation qui va donner à la protéine sa fonction biologique.

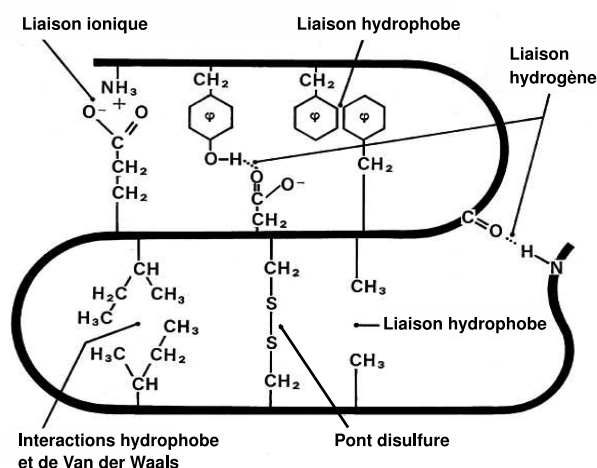


FIG. 1.5 – Exemple d'interactions intervenant dans le repliement d'une chaîne polypeptidique.

1.2.4 Les acides nucléiques

Les acides nucléiques déterminent la structure primaire des protéines en portant et transmettant l'information génétique². Ce sont des polymères de nucléotides, chaque nucléotide se composant d'un groupe phosphate lié par covalence à un sucre (un pentose), lui-même lié à une base azotée. Les nucléotides se lient en créant un squelette sucre-phosphate auquel sont attachées les bases.

²La structure des acides nucléiques est traitée plus en détails dans l'annexe B.

Il existe deux types d'acides nucléiques : l'acide ribonucléique (ARN) et l'acide désoxyribonucléique (ADN) se distinguant par leurs sucres : respectivement le ribose et le désoxyribose. Les bases azotées sont au nombre de quatre : l'adénine, la guanine, la cytosine et la thymine (substituée par l'uracile dans l'ARN). D'autre part, l'ARN est généralement trouvé sous la forme d'un simple brin, souvent replié de manière complexe sur lui-même, alors que l'ADN se présente sous la forme d'un double brin adoptant une structure hélicoïdale. Les deux brins sont alors maintenus par des liaisons hydrogène se formant entre bases complémentaires. D'autres forces, telles que les interactions de van der Waals, contribuent à stabiliser la structure. En outre, les acides nucléiques présentent des propriétés mécaniques intrinsèques d'élasticité et de flexibilité structurale facilitant l'interaction avec d'autres molécules biologiques avoisinantes [43].

1.2.5 L'eau

L'eau est le constituant le plus abondant dans les tissus vivants : elle représente 60 à 95 % de leur masse. C'est un élément simple et familier qui ne doit cependant pas être oublié tant son rôle est important. Ses propriétés particulières font de l'eau une molécule indispensable à la vie [39] : de part sa propriété de molécule polaire, l'eau est un excellent solvant, capable de créer des ponts avec d'autres molécules polaires.

Outre son rôle de solvant, l'eau est aussi le substrat et le produit de multiples réactions chimiques. Elle est également fondamentale dans la structure tridimensionnelle et l'assemblage des molécules biologiques. En effet, les repliements et assemblages macromoléculaires sont stabilisés par l'élimination de l'eau des régions non polaires (effet hydrophobe) et la création de liaisons hydrogène avec certains atomes. Enfin, sa concentration en ions hydrogène H^+ , c'est-à-dire le pH, peut modifier l'état ionique des molécules biologiques, rendant les processus biologiques très sensibles à ces variations.

1.3 Interactions

Certainly no subject or field is making more progress on so many fronts at the present moment than biology, and if we were to name the most powerful assumption of all, which leads one on and on in an attempt to understand life, it is that all things are made of atoms, and that everything that living things do can be understood in terms of the jiggings and wiggings of atoms.

The Feynman Lecture on Physics, 1963 [32].

Mises à part quelques exceptions, les atomes ne sont pas isolés dans les conditions habituelles. Ils établissent entre eux des liaisons chimiques, ou liaisons fortes, afin de former des molécules. Une particularité des molécules biologiques qui rend leur étude d'autant plus complexe est l'importance, aussi bien dans leur structure que lors d'interactions, de la formation de liaisons non chimiques, ou liaisons faibles. Dans cette partie nous allons développer les liaisons chimiques faibles et fortes d'importance biologique.

1.3.1 Les liaisons chimiques fortes

Les liaisons chimiques se forment spontanément entre des atomes pour former une molécule dont l'énergie sera inférieure à celle des atomes pris séparément. Ce sont des interactions à courte portée (typiquement entre 0.7 et 2 Å).

Elles sont assurées par l'interaction des électrons de la couche de valence des atomes (couche la plus externe des atomes) dont les énergies et les propriétés dépendent de l'atome. En connaissant la structure électronique des atomes en jeu, leur électronégativité, leur rayon, on peut définir la nature et la force des liaisons chimiques.

Dans les molécules biologiques, on peut définir trois types de liaisons fortes : covalente, covalente polarisée et ionique [20].

La liaison covalente

La liaison covalente est définie par un partage des électrons de valence de manière équiprobable entre deux atomes. En d'autres termes, les électrons sont distribués symétriquement entre les deux atomes. La liaison est stabilisée par l'attraction que les électrons partagés subissent de la part des deux noyaux positivement chargés.

En fonction de sa structure électronique (*cf.* tableau 1.3), chaque type d'atome peut établir un certain nombre de liaisons covalentes avec d'autres atomes selon une géométrie qui dépend de cette structure ; les liaisons sont simples, doubles ou triples s'il y a une, deux ou trois paires d'électrons partagées.

Liaison	Energie (kJ/mol)	Longueur (pm)
C-C	345.2	154
C=C	609.4	134
C≡C	834.3	120

TAB. 1.2 – Longueurs et énergies d'une liaison covalente simple, double et triple établie entre deux atomes de carbone.

Lorsque deux atomes de même électronégativité se rapprochent, leurs nuages électroniques se recouvrent et peuvent échanger des électrons. Le système est alors stabilisé : il y a formation d'une liaison chimique. Lorsque les forces d'attraction et les forces de répulsion entre les noyaux et entre les électrons se compensent, une distance d'équilibre est atteinte et est caractéristique d'une liaison.

La structure spatiale d'une molécule, c'est-à-dire la disposition relative des atomes qui la constituent, dépend donc de l'ordre d'enchaînement et des types d'atomes qui la constituent. Ceux-ci se placent les uns par rapport aux autres, avec des longueurs et des angles de liaison bien déterminés. D'autre part, la possibilité de rotation autour de la liaison, limitée par le nombre d'électrons partagés ou par l'encombrement stérique, est un élément de flexibilité.

Plus le domaine de recouvrement est important, plus la liaison est forte (cf. tableau 1.2). La force d'une liaison covalente est donnée par l'enthalpie de liaison, c'est-à-dire l'énergie nécessaire à la rupture de la liaison dans la molécule gazeuse pour donner les atomes gazeux. Elle est de l'ordre de la centaine de kJ/mol.

Atomes	Electronégativité	Electrons de valence	Nombre possible de voisins
H	2.2	$\dot{\text{H}}$	1
C	2.55	$\cdot\dot{\text{C}}\cdot$	2, 3 ou 4
O	3.44	$\cdot\ddot{\text{O}}\cdot$	1 ou 2
N	3.04	$\cdot\ddot{\text{N}}\cdot$	2, 3 ou 4
P	2.19	$\cdot\ddot{\text{P}}\cdot$	5
S	2.58	$\cdot\ddot{\text{S}}\cdot$	2, 4 ou 6

TAB. 1.3 – Electrons de valence, électronégativité et possibles liaisons de covalence des atomes d'hydrogène, carbone, oxygène, azote, phosphore et soufre.

La liaison covalente polarisée

Tout comme la liaison covalente, elle résulte de la mise en commun d'électrons célibataires entre deux atomes, mais d'électronégativités différentes. Les électrons sont attirés préférentiellement par l'atome le plus électronégatif. Ce phénomène engendre l'apparition d'une charge partielle sur chaque atome : l'électron est davantage attiré par un atome que par l'autre et se partage entre les deux de manière dissymétrique.

Par conséquent, l'atome le plus électronégatif porte un excès de charge négative et l'atome le moins électronégatif porte une fraction de charge positive. On dit que la liaison covalente est polarisée. L'énergie de la liaison comporte deux contributions : covalente et électrostatique. Elle est comparable à celle d'une liaison covalente pure.

La liaison ionique

Une liaison ionique se crée entre deux atomes d'électronégativités très différentes. Un atome cède un ou plusieurs électrons de valence à l'atome le plus électronégatif, qui porte alors une charge entière négative. Celui qui a cédé le ou les électrons porte une charge entière positive. On obtient deux ions (positif et négatif) liés essentiellement par force électrostatique.

L'énergie d'une liaison ionique est l'énergie nécessaire pour la dissociation d'un cristal ionique en ions gazeux. Appelée énergie réticulaire, elle est comparable à celle d'une liaison covalente pure ou polarisée. Toutefois, la liaison ionique est très affaiblie en milieu aqueux où elle est de l'ordre de 20 kJ/mol.³

³La force coulombienne s'exerçant entre deux charges q et q' distantes de r : $F = \frac{qq'}{4\pi\epsilon r^2}$, dépend de la constante diélectrique du milieu : plus un solvant est polaire, plus ϵ est grand, plus cette force est faible. En milieu aqueux par exemple, la constante diélectrique étant très élevée : 80, la liaison ionique est en fait très faible.

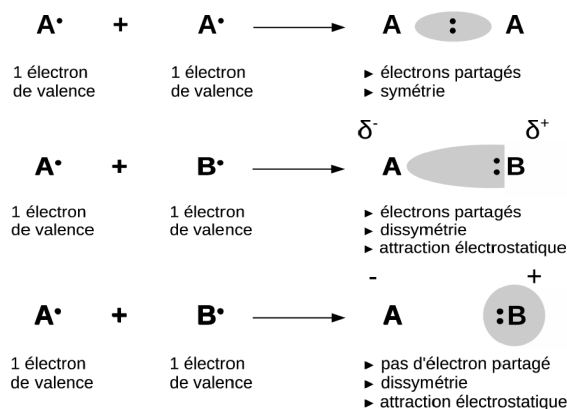


FIG. 1.6 – De haut en bas : distribution des électrons dans des liaisons covalentes, covalentes polarisées et ioniques.

1.3.2 Les liaisons chimiques faibles

Ces liaisons sont essentielles pour expliquer les propriétés des molécules biologiques. Du fait de leur faible énergie (généralement comprise entre 4 et 30 kJ/mol), elles sont en compétition avec les forces de dissociation liées à l'agitation thermique (2.5 kJ/mol à 25°C) et peuvent se rompre et se rétablir très facilement à la température physiologique, permettant ainsi des interactions temporaires entre molécules.

Les liaisons faibles permettent le repliement des molécules, stables en raison du grand nombre de liaisons intramoléculaires accumulées. Elles interviennent aussi bien entre atomes d'une même molécule, qu'entre atomes de molécules voisines, ou encore entre molécules et solvant [33][34].

Les forces de van der Waals

Ces forces résultent de l'interaction des nuages électroniques d'atomes ou de molécules proches les uns des autres. Ces nuages fluctuent, soit de façon permanente, c'est le cas dans les molécules polaires qui se comportent comme un dipôle permanent, soit de façon temporaire, sous l'action d'un champ électrique induit par le voisinage d'un ion, d'une molécule... faisant apparaître un moment dipolaire dit induit.

Les forces de van der Waals naissent de l'interaction électrostatique de ces dipôles et sont la conséquence de trois phénomènes distincts provoquant l'attraction de dipôles :

- *l'effet d'orientation de Keesom* : les forces de Keesom apparaissent entre deux molécules polaires, ce sont des interactions dipôle permanent - dipôle permanent, d'autant plus fortes que les molécules sont polaires. Les dipôles de même signe se repoussent, ceux de sens opposés s'attirent. Ces forces ne s'établissent qu'en l'absence de contrainte extérieure, les dipôles vont s'aligner pour minimiser l'énergie. Celle-ci diminue en fonction de $1/r^3$.

- *l'effet d'induction de Debye* : c'est une interaction de type dipôle permanent - dipôle induit apparaissant entre une molécule polaire et une molécule polaire ou non polaire qui se polarise sous l'effet du champ électrique créé par la molécule polaire. Elle est d'autant plus forte que le moment dipolaire de la molécule polaire est élevé. L'énergie d'interaction décroît en $1/r^5$.
- *l'effet de dispersion de London* : c'est une interaction de type dipôle induit - dipôle induit, s'effectuant entre deux molécules non polaires présentant instantanément un moment dipolaire non nul. L'énergie de ce type d'interaction diminue en fonction de $1/r^6$.

Selon les molécules, la prédominance d'une de ces trois forces par rapport aux autres varie. Les interactions de van der Waals sont généralement faibles, chacune contribuant à la stabilisation du système avec une énergie comprise entre 0.4 et 4 kJ/mol. Leur importance provient du nombre cumulé de liaisons créées.

Outre les forces d'attraction, un phénomène de répulsion apparaît lorsque deux atomes se rapprochent au point que leurs nuages électroniques se recouvrent. La distance d'équilibre des atomes en interaction correspond donc à la distance à laquelle attraction et répulsion se compensent.

Les forces de répulsion stérique sont définies en $1/r^{12}$ et deviennent énormes à très courte distance. On peut donc définir des rayons atomiques, appelés rayons de van der Waals, impénétrables par d'autres atomes et correspondant à la taille du nuage électronique de l'atome.

La liaison hydrogène

Cette liaison, également appelée pont hydrogène, intervient lorsqu'un atome d'hydrogène lié par covalence à un atome électronégatif (le donneur D) est attiré par un autre atome électronégatif (l'accepteur A, généralement un atome d'azote ou d'oxygène).

C'est donc une interaction de type dipôle-dipôle, cas particulier des forces de van der Waals, induite par l'apparition d'une charge partielle positive sur l'hydrogène, son nuage électronique étant attiré par l'atome donneur. Cette charge positive est alors attirée par la charge partielle négative portée par l'accepteur : $\delta^- D - H^{\delta^+} \cdots A^{\delta^-}$.

Ce type de liaison, pouvant être intra ou intermoléculaire, explique en grande partie l'arrangement spatial des macromolécules mais aussi l'importance de l'effet structurant de l'eau. Sa force est de l'ordre de 12 à 30 kJ/mol, soit davantage que celle d'une interaction de van der Waals, et de l'ordre des énergies mises en jeu dans les fluctuations thermiques à la température physiologique. Elle est assez élevée pour avoir des conséquences importantes et assez faible pour être réversible.

Les liaisons hydrogène peuvent ainsi se rompre et se rétablir, conférant aux structures moléculaires à la fois stabilité et souplesse, ainsi que la possibilité d'un réarrangement spatial. D'autre part, elles sont directionnelles, ce qui permet de retrouver des motifs définis (complémentarité des bases dans les acides nucléiques...). Enfin, les ponts hydrogène peuvent opérer des transferts de protons H^+ entre les molécules qu'ils lient, phénomène à l'origine de la réactivité des milieux aqueux.

L'effet hydrophobe

Les molécules ou groupes non polaires ne sont pas capables de former de liaisons hydrogène et ne peuvent donc pas s'hydrater : on les nomme pour cette raison 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. La force des "liaisons" hydrophobes est de l'ordre de 20 à 30 kJ/mol.

Une molécule hydrophobe rompt localement le réseau de liaisons hydrogène du solvant et crée un état énergétiquement défavorable en diminuant le désordre (perte d'entropie). Les molécules d'eau se réorganisent alors autour des molécules 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 de London, 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 polaire.

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 et des assemblages supramoléculaires.

1.3.3 Implication des interactions dans la structure et la flexibilité

Les conditions environnementales (température, pH, salinité...) affectent la stabilité et la flexibilité des biomolécules. Leur rigidité structurale est suffisante pour préserver leur forme native spécifique ; leur flexibilité interne permet leur bon fonctionnement (activité enzymatique des enzymes...). La flexibilité est donc essentielle pour conserver leur activité biologique.

Les liaisons de faible énergie sont nécessaires à l'établissement et au maintien des conformations, de même que pour le fonctionnement des mécanismes des processus biologiques. Leur grand nombre, par effet cumulatif, leur confère un rôle capital au niveau du repliement tridimensionnel et de la stabilité des structures.

En revanche, la facilité qu'ont ces liaisons faibles à se rompre permet les interactions moléculaires caractéristiques des réactions biochimiques dont découlent les propriétés de la vie : en permanence, les molécules se reconnaissent, s'assemblent, réagissent et se dissocient. D'autre part, elles permettent une grande souplesse conformationnelle, clé de la flexibilité des molécules biologiques, et donc des capacités d'adaptation, de reconnaissance : de spécificité.

Il apparaît par conséquent qu'au delà de la séquence, l'information et la fonction que recèlent les molécules biologiques découlent de leur structure. La grande difficulté de l'étude des molécules biologiques réside justement dans la complexité et la flexibilité des structures. En effet, en raison de la longueur des chaînes, des liaisons faibles, des liaisons simples présentes dans les squelettes carbonés et autour desquelles des rotations sont possibles, une molécule peut posséder plusieurs configurations. En outre, pour ces mêmes raisons, ainsi qu'à cause de l'encombrement stérique, ou de phénomènes de répulsion ou d'attraction électrostatique, la molécule ne présente généralement qu'une ou quelques conformations privilégiées, correspondant à la plus grande stabilité thermodynamique possible. Une telle conformation est la forme active d'une

molécule (dite "native" dans le cas d'une protéine), celle qui est fonctionnelle et peut interagir. La fonction biologique d'une protéine est ainsi intimement liée à la conformation adoptée.

Les interactions sont facilitées par la complémentarité structurale et électronique des molécules biologiques. Ces dernières présentent la particularité de pouvoir changer de conformation afin de s'adapter les unes aux autres : elles adoptent des conformations induites. Elles doivent pour cela présenter un maximum d'affinité et de spécificité, par complémentarité. Cette faculté d'adaptabilité conformationnelle est déterminante dans la plupart des phénomènes biologiques et explique simultanément la stabilité et la variété du monde vivant.

La complémentarité structurale est elle-même facilitée par la flexibilité, c'est-à-dire les modifications conformationnelles que peuvent adopter deux molécules lorsqu'elles s'approchent l'une de l'autre. On parle dans ce cas de conformations induites.

La complémentarité électrostatique d'autre part, par jeux d'attraction ou de répulsion à faible distance, intervient favorablement lors de l'approche et de la fixation du ligand sur sa cible. Des molécules présentant de telles complémentarités ont une grande affinité l'une avec l'autre.

1.4 Conclusion

The challenge, in a nutshell, is to understand the cellular information processing system - all of it - from the genome on up.

Committee on Frontiers at the Interface of Computing and Biology, 2005 [27].

Dans ce chapitre, nous avons décrit rapidement les intérêts actuels de la biologie structurale : identifier et comprendre la fonction biologique (et donc pharmacologique) d'une biomolécule via la connaissance de sa structure, de ses mécanismes et de son réseau d'interactions.

Acquérir ce savoir peut en effet permettre d'agir sur l'activité des biomolécules, mais aussi de détourner leur fonction vers des applications de pointe telles que les nanotechnologies. Mais surtout, nous avons introduit le caractère incontournable de l'étude des interactions chimiques pour expliquer les mécanismes biomoléculaires. Plus particulièrement, nous avons souligné l'importance des liaisons faibles, clé de l'adaptabilité conformationnelle et de la réactivité.

Aujourd'hui, si les outils expérimentaux et calculatoires permettent d'ouvrir la voie à l'étude des interactions inter et intra moléculaires, ce projet n'en demeure pas moins ambitieux. L'expérimentation ne permet pas à ce jour d'apporter toutes les réponses et reste lente et coûteuse, malgré ses indiscutables progrès.

Une nouvelle voie s'ouvre alors à nous : celle du calcul, qui, en se basant sur les lois usuelles de la chimie et de la physique, est tout à fait à même de s'appliquer aux problèmes de la biologie. Le calcul se doit néanmoins d'être rapide et rentable, d'une part pour traiter des données abondantes, et d'autre part pour modéliser les systèmes extrêmement complexes de la biologie.

Les méthodes calculatoires sont actuellement entravées par la complexité des macromolécules et la multitude d'interactions fines intervenant entre leurs milliers d'atomes. Elles nécessitent une adaptation des procédures et des algorithmes pour réduire les temps de calculs et rendre accessibles des systèmes de plus en plus grands.

Il ne fait aucun doute aujourd'hui que la prise en compte des interactions, la plus exhaustive possible, est la clé de la réussite de la modélisation, tout d'abord pour reproduire les systèmes, puis pour atteindre des fins prédictives. Le chapitre suivant recense les méthodes générales de la modélisation à l'échelle atomique ainsi que leurs applications aux molécules biologiques.

Chapitre 2

Modélisation et simulation des biomolécules

Sommaire

2.1	Les modèles existants	24
2.1.1	Les méthodes quantiques	24
2.1.2	La mécanique moléculaire	28
2.1.3	Les simulations dynamiques	32
2.2	Mise en perspective par rapport à la biologie	34
2.2.1	Un développement parallèle aux progrès informatiques	34
2.2.2	Adaptation des outils aux problèmes de la biologie	35
2.2.3	Prise en compte du solvant	36
2.3	Du docking rigide à l'introduction de la flexibilité	37
2.3.1	Représentation du système	38
2.3.2	Procédure de recherche	38
2.3.3	Le scoring	43
2.4	Conclusion	44

La modélisation d'un système consiste en l'établissement d'un modèle mathématique décrivant ce système, dans le but de trouver une solution analytique permettant de prévoir son comportement dans certaines conditions par l'introduction de paramètres. La simulation informatique utilise à cette fin des algorithmes de programmation, basés sur des relations arithmétiques et logiques, écrits dans des langages définis. Elle est aujourd'hui présente dans tous les domaines de recherche : physique, chimie, biologie, économie....

Plus particulièrement, la modélisation à l'échelle atomique et moléculaire permet d'explorer la matière, sa réactivité, ses propriétés thermodynamiques, structurales et mécaniques. Elle vise la compréhension et la prédiction du comportement de la matière sous différentes conditions de pression ou de température par exemple. Elle peut ainsi éviter de lourds travaux expérimentaux, apporter des éléments nouveaux de compréhension ou d'orientation de l'expérience. Elle trouve particulièrement bien sa place en biologie, à une période où l'abondance de données impose un rythme de calcul effréné.

La voie de la simulation numérique s'est considérablement développée parallèlement aux progrès informatiques, en termes de rapidité, de mémoire et d'infographie. L'informatique permet une nouvelle approche d'exploration de la matière en s'enrichissant d'outils dédiés au calcul et à la représentation de molécules de plus en plus complexes.

La modélisation moléculaire met en jeu des méthodes de calcul théorique dont les principales font appel à la mécanique moléculaire, la dynamique moléculaire et à la mécanique quantique afin de déterminer la géométrie des atomes d'une molécule ainsi que ses propriétés physico-chimiques [130][169][222]. L'interprétation est facilitée par de nombreux outils de représentation graphique et permet ainsi d'analyser les résultats sous un jour différent de celui de la voie expérimentale, qui demeure néanmoins complémentaire.

2.1 Les modèles existants

Il existe trois principales méthodes de modélisation moléculaire : les méthodes quantiques, la mécanique moléculaire et la simulation dynamique [108].

2.1.1 Les méthodes quantiques

La mécanique quantique décrit la matière comme un ensemble de noyaux atomiques autour desquels gravitent des électrons, eux-mêmes décrits explicitement par leur probabilité de présence en un point et représentés par des fonctions d'onde.

En d'autres termes, en appliquant les lois de la mécanique ondulatoire aux électrons, elle permet de déterminer l'état électronique d'un système d'atomes, mais aussi l'ensemble de ses propriétés observables (structurales : géométries, angles, longueurs... ; énergétiques : énergies de liaison, d'excitation... ; spectroscopiques : fréquences de vibration, spectres UV-visible, IR et micro-onde... ; électroniques, magnétiques et réactionnelles : barrières d'activation...).

Les bases du calcul quantique ont été posées en 1925 par Heisenberg, Born et Jordan, puis finalisées en 1926 par Schrödinger et sa fameuse équation (2.1), dont le formalisme permet de décrire rigoureusement la nature microscopique de la matière [196][225]⁴.

$$\hat{H} \psi(r_1, r_2 \dots r_n) = E \psi(r_1, r_2 \dots r_n) \quad (2.1)$$

où $\hat{H}$ est l'opérateur hamiltonien du système et E l'énergie propre associée à l'état quantique de la particule décrit par la fonction d'onde stationnaire ψ .

Pour une molécule composée 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 $-e$ et de position r_i , l'hamiltonien peut s'écrire sous la forme (2.2) :

$$\hat{H} = \hat{T}_N + \hat{T}_e + \hat{V}_{e,e} + \hat{V}_{e,N} + \hat{V}_{N,N} \quad (2.2)$$

⁴La plupart des articles scientifiques d'archives abordés dans cette partie sont accessibles depuis les sites <http://prola.aps.org> et <http://home.tiscali.nl/physics/HistoricPaper/>

où $\hat{T}_N$ et $\hat{T}_e$ sont les opérateurs énergie cinétique respectivement des noyaux et des électrons explicités par les équations (2.3) :

$$\hat{T}_N = - \sum_A^M \frac{\hbar^2}{2m_A} \nabla_{R_A}^2 \text{ et } \hat{T}_e = - \sum_i^N \frac{\hbar^2}{2m_e} \nabla_{r_i}^2 \quad (2.3)$$

$\hat{V}_{e,e}$, $\hat{V}_{e,N}$ et $\hat{V}_{N,N}$, décrits par les relations (2.4) sont les opérateurs énergie potentielle (termes d'interactions coulombiennes) électron-électron, électron-noyau et noyau-noyau :

$$\hat{V}_{e,e} = \sum_i^N \sum_{j>i}^N \frac{e^2}{|r_j - r_i|}, \hat{V}_{e,N} = - \sum_A^M \sum_i^N \frac{Z_A e^2}{|r_i - R_A|} \text{ et } \hat{V}_{N,N} = - \sum_A^M \sum_{B>A}^M \frac{Z_A Z_B e^2}{|R_B - R_A|} \quad (2.4)$$

La résolution de l'équation peut, en toute rigueur, décrire parfaitement la matière dans toute sa complexité. Cette équation extrêmement complexe a pour première application la molécule de dihydrogène en 1927, menée parallèlement par Heitler et London d'une part [140], et Condon d'autre part [95]. Ces premiers calculs laissent alors entrevoir l'impossibilité d'une résolution analytique exacte pour des systèmes plus complexes et donc la nécessité d'établir des modèles et des approximations.

The underlying laws for the mathematical theory of the whole of chemistry are completely known and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed.

Dirac. Proc. Roy. Soc. Lond., 1929 [106].

On distingue principalement trois niveaux d'approximation évaluant plus ou moins finement les interactions entre électrons, et permettant donc de traiter des systèmes de taille plus ou moins importante [82]. Par ordre d'approximation croissant :

- les méthodes ab-initio (Hartree-Fock...)
- les approximations de champ moyen (DFT)
- les modèles semi-empiriques

Les méthodes ab initio : détermination de la fonction d'onde

Dès 1927, arrivent les premiers modèles simplifiés pour la résolution de l'équation de Schrödinger et visant à simplifier l'hamiltonien : Born et Oppenheimer publient un travail proposant de découpler le mouvement des noyaux, très lent, de celui des électrons [80].

Les nucléons étant beaucoup plus massifs que les électrons (environ 1800 fois), leur inertie est considérée comme nulle ($\hat{T}_N = 0$) et leur énergie potentielle $\hat{V}_{N,N}$ constante. On ne s'intéresse qu'au nuage électronique décrit par l'hamiltonien (2.5) :

$$\hat{H} = \underbrace{\hat{T}_e + \hat{V}_{e,N}}_{\hat{H}_{el}} + \hat{V}_{e,e} \quad (2.5)$$

En 1928, Hartree propose d'écrire une fonction d'onde à N électrons comme le produit de N fonctions d'ondes mono électroniques (2.6), c'est-à-dire N orbitales [138] [184]. Cela implique que chaque électron se déplace dans un champ moyen dû aux noyaux et aux autres électrons, ce qui caractérise les méthodes SCF (*Self Consistent Field*, ou champ auto-cohérent):

$$\Psi(r_1, r_2 \dots r_n) = \Psi_1(r_1) \Psi_2(r_2) \dots \Psi_n(r_n) \quad (2.6)$$

Cependant, cette approximation des orbitales ne satisfait pas à l'exigence d'antisymétrie et d'indiscernabilité des électrons, conformément au principe d'exclusion de Pauli [205] (les fermions ne peuvent occuper le même endroit dans le même état quantique). En 1929, l'introduction du déterminant de Slater [229] permet une expression plus simple d'un système polyélectronique en tenant compte du spin des particules. Le déterminant de Slater (voir l'équation 2.7) permet l'introduction de la notion de spin dans les équations en écrivant la fonction d'onde comme un produit de spin-orbitales tout en satisfaisant à l'exigence d'antisymétrie :

$$\Psi(r_1, r_2 \dots r_n) = \frac{1}{\sqrt{n!}} \begin{vmatrix} \psi_{11} & \psi_{21} & \dots & \psi_{n1} \\ \psi_{12} & \psi_{22} & \dots & \psi_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{1n} & \psi_{2n} & \dots & \psi_{nn} \end{vmatrix} \quad (2.7)$$

Cette avancée contribue largement au développement des méthodes *ab initio* SCF telles que la méthode Hartree-Fock (HF) qui prend en compte la totalité des électrons mis en jeu par chaque atome et calcule explicitement tous les éléments de l'équation de Schrödinger dans le cadre de l'approximation de Born-Oppenheimer.

Il s'agit d'une méthode variationnelle permettant la détermination de la fonction d'onde ψ_0 décrivant l'état de plus basse énergie E_0 , ou état fondamental, en supposant que chaque fonction d'onde d'un système à N fermions peut être approximée par un déterminant de Slater. Par rapport à l'approximation de Hartree, elle fait intervenir un terme d'interaction d'échange. A partir d'une fonction d'essai ψ décrite à l'aide d'un déterminant de Slater, l'énergie moyenne du système, $\langle E \rangle = \langle \psi | \hat{H}_{el} | \psi \rangle$, est calculée. Sur la base du principe variationnel selon lequel l'ensemble des orbitales sera optimal quand $\langle E \rangle$ sera minimale ($E_0 < \langle \psi | \hat{H}_{el} | \psi \rangle$ pour toute fonction ψ solution du système), on résout un système de N équations couplées non linéaires par une procédure itérative jusqu'à converger vers un état auto-cohérent.

La précision du résultat obtenu dépend bien entendu des fonctions d'essai employées par la méthode *ab-initio* [175]. Ce type de calcul donne un résultat précis, cependant le temps de calcul nécessaire est très important et limite les applications à des systèmes comportant un nombre peu important d'atomes. Notons enfin que la méthode Hartree-Fock étant basée sur un modèle de particules indépendantes, elle peut donner une fonction d'onde et une énergie pas totalement satisfaisantes : la différence entre l'énergie exacte et l'énergie HF de la molécule correspond à l'énergie de corrélation non prise en compte et à la surestimation des répulsions électroniques. Ce problème peut être amélioré en incluant la corrélation des mouvements des électrons : c'est ce que font les méthodes *post Hartree-Fock*.

Méthodes DFT : détermination de la densité électronique

Nous venons de voir que les méthodes HF et post-HF nécessitent la mise en œuvre de calculs complexes. Lorsque les systèmes étudiés ont une taille trop importante pour être traités de cette façon, une alternative est la théorie de la fonctionnelle de la densité (DFT, *Density Functional Theory*). On ne considère plus des fonctions d'ondes mais des densités électroniques.

Ce modèle a été introduit par Thomas et Fermi [125][237] en exprimant l'énergie totale d'un système polyélectronique comme une fonctionnelle de la densité électronique totale, puis formalisé en 1964 par Hohenberg et Kohn [144], et Kohn et Sham en 1965 [161]. Ils montrent ainsi que toutes les propriétés d'un système d'électrons peuvent être déterminées par sa densité électronique. La valeur exacte de la densité $\rho(r)$ conduit à l'énergie totale du système dans l'état fondamental $E[\rho]$:

$$E[\rho] = T_s[\rho] + V_{en}[\rho] + J[\rho] + E_{xc}[\rho] \quad (2.8)$$

où la densité d'état est reliée aux fonctions d'ondes par la relation $\rho(\vec{r}) = \sum_{k=1}^N |\psi_k(\vec{r})|^2$.

Les expressions de $T_s[\rho]$ l'énergie cinétique du système, $V_{en}[\rho]$ l'énergie d'attraction électron-noyau, $J[\rho]$ l'énergie de répulsion coulombienne inter électrons et $E_{xc}[\rho]$ l'énergie d'échange et de corrélation sont données ci-dessous :

$$\begin{aligned} T_s[\rho] &= -\frac{1}{2} \sum_k \int \psi_k(\vec{r}) \nabla^2 \psi_k(\vec{r}) d\vec{r} \\ V_{en}[\rho] &= - \sum_i \int \frac{Z_i}{|\vec{R}_i - \vec{r}|} \rho(\vec{r}) d\vec{r} \\ J[\rho] &= \frac{e^2}{2} \int \frac{\rho(\vec{r}) \rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r} d\vec{r}' \\ E_{xc}[\rho] &= \int \epsilon_{xc}(\vec{r}) \rho(\vec{r}) d\vec{r} \end{aligned} \quad (2.9)$$

où $\epsilon_{xc}(\vec{r}) = \epsilon_x(\vec{r}) + \epsilon_c(\vec{r})$ est l'énergie d'échange-corrélation pour une particule d'un gaz homogène d'électrons.

La densité électronique ne nécessite que 3 variables contrairement à la fonction d'onde d'un système à N électrons qui en nécessite 3N. Cependant, le calcul de la fonctionnelle densité nécessite à son tour des approximations portant sur le calcul du terme d'échange-corrélation sur lequel est reportée la difficulté. Citons comme principales approximations :

- l'approximation de la densité locale (LDA) basée sur l'utilisation du modèle d'un gaz uniforme d'électrons (une variante : LSDA permet d'introduire le spin).

- l’approximation du gradient généralisé (GGA) faisant intervenir le gradient de la densité $\nabla\rho(\vec{r})$ afin de rendre compte des variations de distributions dans le nuage électronique (les fonctionnelles les plus répandues sont celles de Becke [77], Perdew et Wang [207][208] et Lee, Yang et Paar [171]).
- les fonctionnelles hybrides, dans lesquelles une partie de l’échange exact, calculé en théorie de Hartree Fock, est introduite.

Enfin, selon une idée introduite par Fermi en 1934 [126], les effets de corrélation électronique dus aux noyaux et à l’effet des électrons de cœur fortement liés et chimiquement inactifs, peuvent être remplacés par un potentiel effectif agissant avec les électrons de valence : c’est l’approximation du pseudopotentiel.

Ainsi, en ne traitant explicitement que les électrons de valence, on obtient un gain important en ressources informatiques tout en ne perdant pas ou peu d’information sur les propriétés physico-chimiques du système.

Méthodes semi-empiriques

Les méthodes semi-empiriques reprennent le principe des méthodes *ab initio*. Toutefois, contrairement à ces dernières qui ne nécessitent aucun paramètre expérimental, mais seulement les lois fondamentales de la mécanique quantique, les méthodes semi-empiriques utilisent des paramètres ajustés avec des résultats expérimentaux afin de simplifier les calculs. En ne prenant en compte que les électrons de la couche de valence ou en négligeant certaines intégrales, elles nécessitent des temps de calcul plus courts et permettent d’accéder à des systèmes de taille plus importante.

On peut citer comme types d’approximations : la représentation limitée aux électrons de valence, l’approximation de l’hamiltonien du système par une somme d’hamiltoniens monoélectroniques non explicités (méthode de Hückel et Hückel étendue [142]) ou l’approximation portant sur les intégrales biélectroniques (équations de Roothaan).

En conclusion, les méthodes quantiques permettent en théorie de sonder la matière et d’en calculer exactement toutes les grandeurs observables. Elles peuvent assister ou orienter l’expérience, voire caractériser des mécanismes jusque-là inaccessibles. Cependant, leur résolution est limitée par leur complexité : celle-ci impose l’étude de systèmes de taille limitée, ainsi qu’un compromis entre rapidité et précision, elle-même dépendante de l’introduction incontournable de diverses approximations et de choix de l’utilisateur.

2.1.2 La mécanique moléculaire

La mécanique moléculaire est apparu en 1930 [66], mais s’est développée à partir des années 1960, avec les progrès d’accessibilité et de performance des ordinateurs. Elle permet de déterminer l’énergie d’une molécule en fonction de ses coordonnées atomiques et de chercher des minima de l’énergie correspondant à des conformères stables [84][160].

Le calcul d'énergie potentielle est fait en fonction de la géométrie des atomes et non par rapport à leur structure électronique.

Cette méthode décrit une molécule comme une série de sphères dures et non interpénétrables, reliées entre elles par des forces mécaniques généralement harmoniques : elle est basée sur les lois de la mécanique classique (newtonienne) et l'utilisation d'un champ de force empirique dont les constantes décrivent les interactions entre atomes liés et non liés, obtenues par étalonnage de résultats expérimentaux.

L'énergie potentielle E d'une molécule peut être approchée par la fonction suivante (2.10), impliquant des atomes liés de manière covalente ou pas :

$$E = E_l + E_\theta + E_\tau + E_{vdW} + E_{el} + E_H \quad (2.10)$$

Termes d'énergies liées

- E_l représente l'énergie d'élongation des liaisons : $E_l = \sum_{liaisons} k_l (r - r_0)^2$, où k_l est la constante de force, r est la longueur de liaison instantanée, r_0 la valeur de liaison à l'équilibre.
- E_θ est l'énergie de variation des angles : $E_\theta = \sum_{angles} k_\theta (\theta - \theta_0)^2$, où k_θ est la constante de force associée, θ est l'angle instantané, θ_0 est la valeur de l'angle à l'équilibre.
- E_τ est l'énergie de torsion des angles dièdres : $E_\tau = \sum_{dièdres} \sum_n V_n (1 + \cos(n\tau - \gamma))$, où V_n est la valeur de la barrière de rotation d'ordre n , γ est la phase et τ est la valeur instantanée du dièdre.

Remarque : Le potentiel de Morse (2.11) peut être utilisé pour représenter les termes d'élongation et de variation angulaire en rendant compte de l'anharmonicité dans les liaisons covalentes [192]. Notons cependant qu'il ne rend pas compte de l'asymétrie de l'anharmonicité, et d'autre part, qu'une même constante a décrit les termes harmoniques et anharmoniques.

$$E_{Morse}(r) = D_e (1 - e^{-a(r-r_0)})^2 \quad (2.11)$$

où D_e est l'énergie potentielle minimum (profondeur du puits) et a est une constante propre à la liaison.

Il est néanmoins complexe et peut être approximé par un potentiel harmonique (2.12) en admettant que l'on s'écarte peu de l'équilibre (Loi de Hooke) :

$$E_l(r) = k_l (r - r_0)^2 \text{ entre deux atomes} \quad (2.12)$$

En cas de non linéarité, l'ajout de termes cubiques, voire d'ordres supérieurs, présentés sur la figure 2.1 permettent de se rapprocher du modèle de Morse (2.13) :

$$E_l(r) = k_l (r - r_0)^2 [1 - k'_l (r - r_0) - k''_l (r - r_0)^2 - k'''_l (r - r_0)^3 - \dots] \quad (2.13)$$

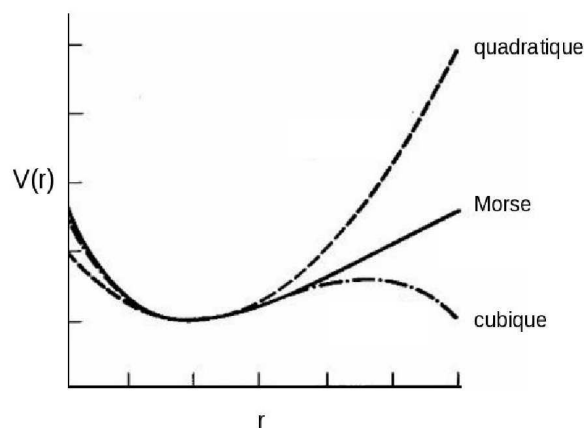


FIG. 2.1 – Représentation des différents potentiels utilisés pour modéliser les termes d'élongation et de variation angulaire.

Termes d'énergies non liées

- E_{vdW} donne l'énergie d'interaction entre atomes non liés généralement sous la forme d'une fonction de Lennard-Jones "6-12" : $E_{vdW} = \frac{A}{r^{12}} - \frac{B}{r^6}$, où A et B sont des constantes relatives à la profondeur du puits (c'est-à-dire le minimum d'énergie potentielle) et à la distance minimale d'approche entre les deux atomes. Les termes d'ordres 6 et 12 représentent respectivement la dispersion attractive de London et la répulsion.
- E_{elec} représente l'énergie d'interaction électrostatique : $E_{elec} = \frac{q_1 q_2}{4\pi\epsilon_0\epsilon_r r}$ où q_1, q_2 sont les charges des particules en interaction, r la distance les séparant, ϵ_0 la permittivité du vide et ϵ_r la permittivité relative du milieu.
- E_H rend compte des liaisons hydrogène. Cette énergie est souvent traitée sous la forme $E_H = \frac{C}{r^{12}} - \frac{D}{r^{10}}$, ou est incluse dans le terme d'énergie électrostatique.

A l'aide d'un jeu d'équations, de paramètres de champ et des types d'atomes et de liaisons, la molécule peut être soumise à une optimisation géométrique via la minimisation de son énergie par diverses méthodes incrémentales, jusqu'à obtenir un minimum local [145].

Les méthodes itératives couramment usitées reposent sur :

- la dérivée première de l'énergie potentielle : les méthodes de plus grande pente (ou "*steepest descent*") ou du gradient conjugué permettent de faire diminuer rapidement l'énergie dans les premiers cycles mais sont peu convergentes en fin de cycle. Elles sont donc plutôt utilisées au départ d'une minimisation.
- la dérivée seconde de l'énergie potentielle (constante de force) : la méthode de Newton-Raphson est plus coûteuse en temps de calcul et en mémoire car elle fait appel au calcul et à l'inversion de grosses matrices contenant les constantes de force ($3N \times 3N$ pour une molécule de N atomes). Elle est cependant plus convergente et peut être réservée à la phase finale de minimisation.

Notons qu'il faut être vigilant par rapport au résultat de la minimisation qui conduit à un minimum local (et non global car la mécanique moléculaire ne permet pas de franchir de barrière) et parfois à un placement erroné des atomes légers (hydrogène) nécessitant un second passage.

La mécanique moléculaire présente l'important avantage d'être peu coûteuse en temps et moyens de calcul, et ainsi, d'être accessible aux grosses molécules. L'inconvénient majeur néanmoins est la dépendance du résultat au champ de force paramétré pour une famille de molécules et non généralisable. La valeur de l'énergie obtenue n'a donc aucun sens physique : c'est la comparaison entre deux valeurs obtenues pour deux conformations différentes avec un même champ de force qui peut être interprétée.

Depuis les années 1970, de nombreux champs de force ont été développés et commercialisés. Parmi les plus courants, on peut citer MM2 [59], MM3 [60] et MM4 [61] pour les petites molécules, Sybyl [235] pour les molécules organiques, CHARMM [155] pour les protéines, AMBER [248] pour les protéines et les acides nucléiques.

Les méthodes hybrides QM/MM

La mécanique moléculaire ne permet pas de traiter la réactivité des molécules. Quant aux méthodes quantiques, elles sont rapidement limitées par la taille du système étudié.

Le couplage des méthodes de mécanique quantique / mécanique moléculaire (QM/MM) utilise diverses descriptions pour des régions différentes d'une molécule : il est alors possible pour des systèmes de taille conséquente de procéder à l'étude théorique des processus chimiques traités localement par mécanique quantique, alors que le reste des interactions est traité "physiquement" par mécanique moléculaire.

On peut ainsi explorer des systèmes complexes comportant quelques centaines, voire quelques milliers d'atomes [64][116][190][220].

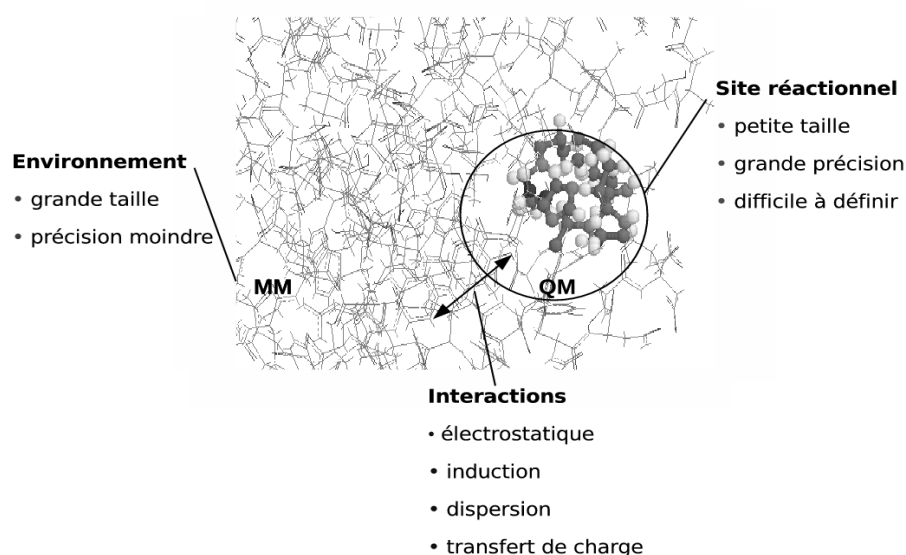


FIG. 2.2 – Exemple d'application de la méthode hybride QM/MM à un système moléculaire.

2.1.3 Les simulations dynamiques

La mécanique moléculaire permet de calculer l'enthalpie d'un système, mais ne peut décrire l'énergie en fonction du temps et de la température (entropie, énergie libre), ni franchir de barrière énergétique.

La dynamique moléculaire et la méthode de Monte Carlo sont des extensions de la mécanique moléculaire. Elles conviennent aux systèmes présentant de nombreux conformères séparés par de faibles barrières énergétiques, comme c'est le cas notamment des biomolécules ou des polymères.

La dynamique moléculaire

Les premières simulations de dynamique moléculaire remontent à la fin des années 1950 [58] et au début des années 1960 pour des systèmes plus importants [210].

Cette technique a pour but de simuler le mouvement des atomes au cours du temps pour une température supérieure à 0 K à partir de l'intégration des équations de Newton ou de Hamilton [183][212]. Ce mouvement correspond à des vibrations autour d'un minimum d'énergie ou au passage d'un minimum à un autre : la dynamique moléculaire permet donc de s'extraire d'un minimum local.

A partir de potentiels $V(r_i)$ extraits de champs de force, à intervalles de temps dt réguliers et très petits, on détermine les forces qui agissent sur chaque atome i grâce à la relation (2.14) :

$$\vec{F}_i = -\vec{\nabla}V(r_i) \quad (2.14)$$

Puis avec la seconde loi de Newton (2.15), on attribue les nouvelles vitesses et positions :

$$\vec{F}_i = m_i \vec{a}_i = m_i \frac{d\vec{v}_i(t)}{dt} = m_i \frac{d^2\vec{r}_i(t)}{dt^2} \quad (2.15)$$

L'intégration numérique fait appel à diverses méthodes de différences finies, l'algorithme le plus utilisé étant celui Verlet [242]. Il suit la procédure suivante :

- calcul des positions à l'instant $t + \delta t$ à partir des positions et des vitesses à t :

$$\vec{r}_i(t + \delta t) = \vec{r}_i(t) + \delta t \vec{v}_i(t) + \frac{\delta t^2}{2m_i} \vec{F}_i(t) + o(\delta t^3)$$
- calcul des forces aux nouvelles positions $\vec{r}_i(t + \delta t)$
- calcul des nouvelles vitesses : $\vec{v}_i(t + \delta t) = \vec{v}_i(t) + \delta t \frac{\vec{F}_i(t)}{m_i} + \frac{\delta t^2}{2m_i} \frac{d\vec{F}_i(t)}{dt} + o(\delta t^3)$

Notons qu'à l'instant initial, les vitesses sont attribuées de manière aléatoire selon une distribution de Boltzmann pour une température donnée.

La dynamique moléculaire nécessite des ajustements fins des divers paramètres induisant une mise en œuvre délicate. D'autre part, l'intégration de l'équation de Newton se faisant dans des intervalles de temps très petits (généralement de l'ordre de la pico à la femtoseconde), la longueur des calculs ne permet pas alors d'explorer des régions de l'espace conformationnel trop éloignées de l'arrangement initial.

La méthode Monte Carlo

Cette autre méthode fait partie des techniques heuristiques : par tirages aléatoires, elle permet d'échantillonner l'espace des conformations.

Son nom fait référence à la ville de Monte Carlo, connue pour ses jeux de hasard [168]. Son développement remonte à la seconde guerre mondiale avec les travaux sur la diffusion des neutrons dans un matériau fissile (Projet Manhattan) de Nicholas Metropolis et ses collaborateurs [188][189]. Le formalisme est publié à la fin des années 1940.

La méthode de Monte Carlo convient particulièrement au calcul des intégrales de grandes dimensions : elle permet de traiter des systèmes trop volumineux pour que les techniques d'approximation numérique puissent conduire à un résultat suffisamment précis dans un temps acceptable. On lui trouve donc de multiples applications dans les domaines ayant recours aux mathématiques appliquées : physique, chimie, biologie, économie, finance, sociologie ...

En modélisation moléculaire, l'idée est de partir d'une configuration donnée et de générer une nouvelle configuration par des déplacements aléatoires des atomes [219]. La nouvelle conformation est acceptée si la différence d'énergie ΔE avec la conformation précédente est négative, ou encore si elle répond à une probabilité p d'être acceptée, selon la relation (2.16) :

$$p = e^{-\frac{\Delta E}{k_B T}} \quad (2.16)$$

La nouvelle conformation sert alors de point de départ pour un nouveau déplacement aléatoire des atomes, sinon on repart de la première ... jusqu'à obtenir une convergence.

Le Monte Carlo cinétique

Le grand avantage des techniques Monte Carlo est leur possible application à des calculs longs sur des systèmes complexes. Elles permettent donc de se rapprocher de la réalité physique de l'observation.

Cependant, elles ne permettent pas de connaître le chemin pour passer d'une conformation à une autre. La méthode de Monte Carlo cinétique [240] permet de décrire l'évolution d'un système en associant à deux mécanismes successifs une énergie d'activation nécessaire pour franchir la barrière permettant de passer de l'un à l'autre. Cette énergie d'activation E_{ac} est liée au temps de résidence dans un état t , lui-même relié à la probabilité de transition p d'un état à un autre par la loi d'Arrhenius : $p = \frac{1}{t} \exp\left(-\frac{E_{ac}}{k_B T}\right)$.

Les techniques Monte carlo, requérant peu de temps de calcul, trouvent de nombreuses applications pour des systèmes de grande taille et nécessitent d'importantes échelles de temps. Par un choix judicieux de paramètres et l'implémentation correcte et complète des nombreuses possibilités de mécanismes, le résultat de la simulation montre un excellent accord avec l'expérience.

2.2 Mise en perspective par rapport à la biologie

I would like to emphasize strongly my belief that the era of computing chemists, when hundreds if not thousands of chemists will go to the computing machine instead of the laboratory for increasing many facets of chemical information, is already at hand.

Robert S. Mulliken, 1966.

2.2.1 Un développement parallèle aux progrès informatiques

Au cours de ces dernières années, le développement important des moyens informatiques (puissance des machines, efforts sur le développement d'algorithmes, infographie) a permis à la modélisation moléculaire d'étendre son champ d'action à des systèmes complexes dont la biologie *in silico* est un excellent exemple.

Le développement de la capacité de stockage informatique croît de façon permanente, la puissance des processeurs a doublé tous les 18 mois pendant 20 ans (loi de Moore), suivie par la croissance des banques de données biologiques. Par ailleurs, les méthodes informatiques sont de plus en plus sophistiquées (réseaux de neurones, algorithmes génétiques ...), de nouvelles architectures matérielles apparaissent (calcul parallèle ...), des grilles de ressources informatiques assurent les besoins en stockage et en puissance de calculs.

L'application des ressources informatiques à la biologie n'est pas limitée par le manque de données, mais par la complexité de la sauvegarde, de la manipulation, du traitement des objets moléculaires. Ainsi, après les fantastiques progrès de la génomique et de la protéomique qui ont fourni de considérables quantités de données, restent à décrypter la fonction des biomolécules et la façon dont elles interagissent entre elles ou avec leur environnement.

Massive computing power, novel modeling approaches, new algorithms and mathematical or statistical techniques, and systematic engineering approaches will provide biologists with vital and essential tools for managing the heterogeneity and volume of the data and for extracting meaning from those data.

Committee on Frontiers at the Interface of Computing and Biology, 2005 [94].

Les méthodes de calcul théorique précédemment citées vont, en manipulant les objets biologiques à des niveaux différents, décrire la structure et les propriétés physico-chimiques des molécules du vivant. Associées à une représentation infographique, elles permettent d'appréhender les systèmes à un niveau de complexité que l'expérience ne peut souvent pas approcher, afin d'interpréter et de prédire des mécanismes au niveau atomique.

En biologie, parmi les nombreux champs d'investigation que la modélisation moléculaire vise à développer, on peut citer :

- l'évaluation de la validité d'une structure moléculaire
- le calcul d'un champ de force empirique pour déterminer les propriétés et les caractéristiques moléculaires

- la prédiction qualitative et la reproduction des propriétés des biomolécules
- la corrélation entre structure et propriétés moléculaires
- l’acquisition d’informations sur la dynamique et les énergies des molécules
- la visualisation des données structurales sous forme de dessin tridimensionnel, la reconnaissance des biomolécules
- la création et l’utilisation de banques de données

2.2.2 Adaptation des outils aux problèmes de la biologie

La modélisation moléculaire est devenue un outil incontournable pour tout développement en chimie médicinale et en biologie structurale [154].

La mécanique quantique est bien entendu inapte à décrire un système de la taille d’une protéine par exemple. Néanmoins, en se restreignant à de petits systèmes ou en ne traitant que certains atomes directement impliqués, elle apporte des renseignements précieux sur la réactivité chimique, notamment avec des méthodes QM/MM, utilisées par exemple pour le traitement de réactions enzymatiques. Une telle utilisation, dont le premier exemple remonte à 1992 [156] est aujourd’hui courante, et permet d’atteindre des échelles de temps plus longues [176]. L’approche par mécanique quantique permet également d’obtenir des paramètres pour les potentiels de la mécanique moléculaire, par exemple pour les charges atomiques partielles ajustées pour reproduire le potentiel électrostatique [227].

La mécanique moléculaire classique est employée quant à elle à l’usage des systèmes de grande taille afin de fournir une description structurale optimale, mais correspondant à un minimum local. Ses applications vont des biomolécules jusqu’à la cellule [71]. Le minimum global peut être atteint par d’autres méthodes, comme les techniques Monte Carlo [118][121][137][177]. Enfin, les structures biologiques n’étant pas figées, la dynamique moléculaire va fournir les informations dynamiques et énergétiques. Pour la première fois en 1977 [187], un article de McCammon *et al.* présente des travaux de dynamique moléculaire sur la BPTI (bovine pancreatic trypsin inhibitor) et remplace définitivement le modèle rigide des protéines par des systèmes dynamiques, animés de mouvements internes résultant de changements conformationnels au rôle crucial pour leur fonction [157].

Notons qu’un grand nombre de champs ont été développés pour les systèmes biologiques. En dynamique moléculaire, trois principaux programmes sont utilisés via une licence payante : CHARMM [91], AMBER [65], et GROMOS [134] ... Ces champs ont été paramétrés pour être appliqués aux macromolécules, particulièrement aux protéines et aux acides nucléiques. Depuis quelques années, des programmes pouvant faire appel aux champs précédents et distribués sans licence payante apparaissent sur internet : GROMACS [133], NAMD [197], TINKER [238], ORAC [203] ... D’autre part, il existe des programmes de construction et de visualisation des biopolymères comme : INSIGHT ou QUANTA [148], Chem3D [92], CACTVS [86] ou encore RasMol [214], VMD [244], Xmol [253], PyMOL [209] ... Certains, VMD par exemple, permettent de visualiser les trajectoires de dynamique moléculaire.

2.2.3 Prise en compte du solvant

Le grand défi imposé par les systèmes biologiques repose sur la complexité et la flexibilité des macromolécules qui ne sont que marginalement stables.

L'autre difficulté est l'intégration dans les modèles de l'interaction avec l'environnement [135]. La prise en compte du solvant est en effet indispensable pour recréer au mieux le milieu dans lequel "baignent" les macromolécules : un bain de solvant, d'ions, de métabolites ...

En raison de l'agitation thermique, les biomolécules entrent en permanence en collision avec les molécules du solvant, provoquant la formation de liaisons faibles. En favorisant la formation de réseaux, elles participent, par effet hydrophobe, à l'arrangement conformationnel des macromolécules.

Le solvant peut être pris en compte de deux façons dans les algorithmes : soit de manière explicite, soit de manière implicite. Le traitement explicite s'opère par le positionnement de molécules d'eau autour du système étudié sous forme d'une "boîte d'eau" polygonale décrite avec des conditions périodiques pour masquer les effets de bord [169]. Il est également possible d'ajouter des ions de manière aléatoire pour assurer la neutralité électronique du système.

Le modèle le plus couramment utilisé, car simple et efficace, est le modèle à trois sites d'interactions correspondant aux trois atomes de la molécule d'eau, chaque atome portant une charge ponctuelle (TIP3P dans AMBER [152]). Il existe des modèles plus complexes, comme le modèle à 4 sites (la charge négative est alors attribuée à un atome fictif), ou encore le modèle à 5 sites (la charge négative est distribuée sur deux atomes fictifs) ... qui améliore la distribution de la charge électrostatique autour de la molécule d'eau.

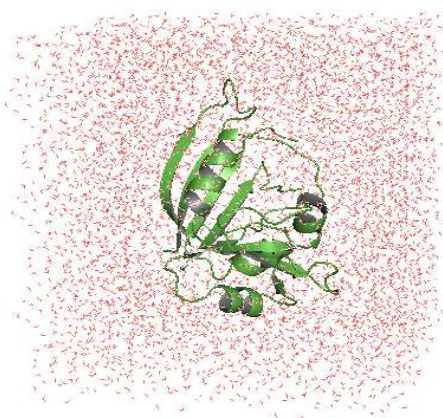


FIG. 2.3 – Dihydrofolate réductase représentée en solvant explicite : une boîte cubique de 64 Å de côté créée par AMBER avec le modèle TIP3P.

La difficulté liée à l'utilisation d'un solvant explicite est bien entendu le coût numérique dû à l'ajout de milliers d'atomes. Une alternative est le recours au solvant implicite, bien plus rapide : les interactions électrostatiques sont traitées de façon à mimer l'effet des molécules d'eau. La méthode adoptée est souvent le modèle de Born généralisé ("*Generalized Born*", GB) traitant le solvant comme un continuum diélectrique [73].

La contribution électrostatique de l'eau sur les charges très enfouies, très accessibles au solvant, ou dans un état intermédiaire est adaptée selon l'équation suivante (2.17) :

$$E_{solv} = -\frac{1}{2} \sum_{i,j} \left(1 - \frac{e^{-\kappa f_{GB}}}{\epsilon} \right) \frac{q_i q_j}{f_{GB}} \quad (2.17)$$

ϵ est la constante diélectrique du solvant.

κ est une constante représentant l'effet d'un sel.

q_i et q_j sont les charges partielles des atomes i et j .

$r_{i,j}$ est la distance entre les deux atomes i et j .

R_i et R_j sont les rayons de Born des atomes i et j .

f_{GB} est une fonction qui représente l'énergie d'interaction coulombienne entre deux charges et mime l'écrantage des charges :

$$f_{GB} = \sqrt{r_{i,j}^2 + R_i R_j \exp\left(\frac{-r_{i,j}^2}{4R_i R_j}\right)}$$

Un tel modèle présente l'avantage de limiter le nombre d'atomes étudiés par rapport à un modèle de solvant explicite tout en tenant compte des effets électrostatiques. Cependant, bien que de grands progrès aient été réalisés depuis la fin des années 1990, il demeure moins précis et fiable.

Outre la question du solvant, l'interaction de la molécule avec son environnement fait ressortir la question de l'interaction avec d'autres molécules, petits ligands ou macromolécules.

2.3 Du docking rigide à l'introduction de la flexibilité

La première problématique de la biologie structurale est le repliement des macromolécules dans une structure tridimensionnelle biologiquement active. La compréhension de la communication moléculaire pour échanger des informations, s'activer ou s'inhiber est la seconde problématique, enjeu majeur de la biologie intégrée [62].

Aujourd'hui, bien qu'un grand nombre de techniques expérimentales soient employées pour étudier la formation de complexes macromoléculaires (système du double hybride en levure, spectroscopie par fluorescence, spectrométrie de masse, résonance des plasmons de surface, calorimétrie ...), peu d'entre elles ont les moyens de donner des informations à l'échelle atomique.

La cristallographie par rayons X fournit un arrangement détaillé de la structure atomique cristallisée, la spectroscopie RMN donne des informations sur la molécule en solution. Les deux ont cependant leurs limites, surtout lorsqu'elles sont appliquées aux complexes biomoléculaires. Le nombre de structures de complexes déposées dans les banques de données est par conséquent très inférieur à celui des molécules à l'état libre.

Complémentaires aux approches expérimentales, les approches théoriques pour l'étude du docking, ou amarrage des molécules, sont en pleine émergence.

Depuis quelques années, les algorithmes de docking se multiplient pour prédire, à partir de la structure de deux molécules isolées, celle du complexe encore irrésolue par RMN ou rayons X [83][88][136][230][252]. Très vite, ces algorithmes ont été adoptés pour la conception de candidats médicaments [63][69][162][167][245].

La plupart des programmes utilisent une approche semblable : une molécule est fixe dans l'espace tandis que la seconde est déplacée par un ensemble de translations et de rotations. Chaque configuration est alors évaluée par une fonction de score en terme de complémentarité de surface, d'interaction électrostatique, de répulsion de van der Waals ...

De manière générale, le protocole de docking peut être décomposé de la façon suivante :

- représentation du système
- procédure de recherche (searching)
- criblage par fonction de score (scoring)

2.3.1 Représentation du système

Il existe trois façons de représenter le système moléculaire : les représentations par atomes, par surface ou par grille [136] :

- la représentation atomique, très complexe, est généralement associée à une fonction d'énergie potentielle et n'intervient que dans le stade final de la procédure de recherche pour affiner le résultat sans trop de dépense informatique [85].
- de nombreuses approches de docking représentent les molécules par une surface, elle-même ramenée à des caractéristiques géométriques : la surface proposée par Conolly en 1983 est construite à partir des rayons de van der Waals des atomes exposés reliés par un réseau de surfaces concaves ou convexes [96][97]. La recherche conformationnelle s'opère ensuite par complémentarité de surface jusqu'à trouver les minima et maxima d'affinité [199] [200]. A cette représentation géométrique, il est possible d'ajouter des termes énergétiques, électrostatiques par exemple, pour compléter le modèle [198][217].
- la représentation de la molécule par une grille d'énergie potentielle est une autre approche très utilisée. Introduite par Goodford en 1985, elle fait généralement appel aux interactions électrostatiques et de van der Waals pour décrire la contribution énergétique du récepteur, qui n'a alors qu'à être lue durant le scoring.

2.3.2 Procédure de recherche

La procédure de recherche doit permettre de couvrir l'espace conformationnel de manière rapide et efficace. Une difficulté du docking réside dans les multiples solutions qu'il est possible de trouver. Deux approches sont possibles : la première est une recherche couvrant tout l'espace des solutions par des rotations et des translations systématiques d'une molécule par rapport à l'autre ; la seconde consiste en une progression graduelle aléatoire dans l'espace des solutions (Monte Carlo, algorithmes génétiques ...) [67], elle peut être guidée selon certains critères (résultats de dynamique moléculaire ...) [243].

Le problème de la reconnaissance moléculaire

Les méthodes initiales étaient basées sur le modèle *clé-serrure* proposé en 1894 par Emile Fischer [128] : la reconnaissance a lieu lorsque le récepteur et son ligand offrent une complémentarité géométrique exacte. Il faut attendre 1958 pour que Koshland [164] propose un nouveau modèle, considérant l'adaptation mutuelle d'un récepteur et de son ligand lors d'une interaction.

L'introduction de la flexibilité, mais aussi de paramètres donnant lieu à une reconnaissance chimique (interactions électrostatiques, liaisons hydrogène...), ont complexifié le problème, mais toujours dans le souci d'établir le cadre d'un modèle rigoureux pour atteindre une grande qualité de prédiction (voir figure 2.4) [2][89][111].

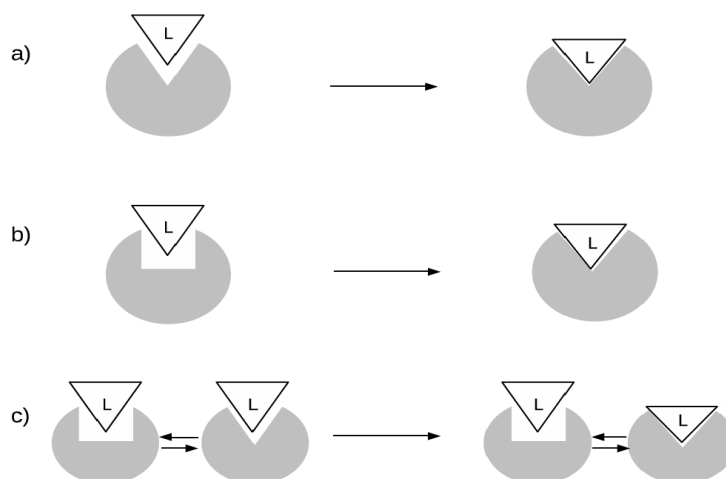


FIG. 2.4 – Schéma des 3 niveaux de représentation de la flexibilité d'une protéine. a) modèle clé-serrure. b) modèle induced-fit. c) configurations pré-existantes à l'équilibre.

Une étape déterminante pour le coût du calcul, et bien entendu pour la qualité de la prédiction, est le choix des degrés de liberté pour chaque procédure. C'est une étape particulièrement difficile qui définit elle-même toute la complexité à venir de l'algorithme.

Le mécanisme de docking a longtemps été perçu comme un simple mécanisme *clé-serrure*. Cette notion introduite par Fisher en 1894 [128] a donné naissance à une longue lignée d'algorithmes depuis 1975 [174] traitant des problèmes de corps rigides, comme DOCK en 1982 [12] [224]. Ces premiers programmes, pour décrire les protéines, fixaient tous les degrés de liberté internes, exceptés les 3 degrés de rotation et 3 de translation. Les atomes étaient traités explicitement à l'aide de potentiels.

En 1978, Wodak et Janin tentent de retrouver la structure expérimentale du complexe protéine-protéine BPTI-trypsine et obtiennent plusieurs configurations possibles énergétiquement non distinguables de la structure native en introduisant des termes non liés, de surface puis de solvant [251]. Greer et Bush [123] approfondissent la complémentarité de surface et utilisent la notion de surface accessible au solvant [216].

Le problème des modèles rigides est qu'ils ne tiennent pas compte de l'*induced-fit*. Koshland est le premier à souligner la capacité d'une enzyme et de son substrat à s'adapter l'un à l'autre en 1958 [164]. Jorgensen [153] étend ce concept aux changements conformationnels que subissent deux molécules durant une interaction. Ce modèle suppose que la reconnaissance est conduite par des forces à longue ou courte portée (respectivement les forces électrostatiques [115], et l'entropie de désolvatation et les forces van der Waals [87]), requérant une certaine complémentarité préalable. Expliquer l'ensemble de l'interaction par l'*induced-fit* aboutirait parfois à une cinétique défavorable [170][185]. Une seconde théorie, basée sur l'existence de multiples conformères en milieu aqueux, est donc envisagée [2][166]. Elle n'est cependant pas indissociable de l'*induced-fit* : la préselection d'un conformère peut précéder l'étape d'ajustement [81]. Une combinaison des deux pourrait même mener à une meilleure description des interactions accompagnées de réarrangements structuraux.

L'introduction de la flexibilité pour prendre en compte un possible réarrangement structural devient alors un des plus grands enjeux du docking. La flexibilité a tout d'abord été introduite en autorisant quelques degrés de liberté au ligand [102], puis étendue à certaines parties des protéines. L'exploration de la flexibilité du ligand est encore aujourd'hui largement explorée. Très vite, les premières applications à la conception de médicaments ont été envisagées [103], un bel exemple est la recherche d'inhibiteurs de la protéase du virus de l'immunodéficience humaine 1 [104].

Outre la complémentarité structurale, l'importance de l'affinité des liaisons chimiques est mise en avant dès 1940 par Pauling et Delbrück [206] : ils postulent que les complexes sont stabilisés par des interactions de van der Waals, des interactions électrostatiques et des liaisons hydrogène. Les effets hydrophobes et de l'entropie, indispensables pour la compréhension de la désolvatation de la surface d'interaction par exemple, sont introduits en 1959 par Kauzmann [158]. On considère actuellement que l'effet hydrophobe est hautement stabilisant pour les complexes macromoléculaires, et que les interactions coulombiennes et les liaisons hydrogène confèrent la spécificité chimique à une interaction.

L'état des lieux que nous allons dresser ici est basé sur les méthodes de prise en compte de la flexibilité dans les algorithmes de docking. C'est un problème toujours très difficile, car contrairement au docking rigide qui ne fait intervenir que 6 degrés de libertés de rotation et de translation, l'introduction de la flexibilité augmente nettement le nombre de degrés de liberté, l'espace de recherche et donc le coût de calcul. On peut distinguer trois niveaux de docking :

- le *docking rigide* est bien sûr le plus simple et demeure encore souvent employé pour l'amarrage protéine-protéine.
- le *docking semi-flexible* est "asymétrique" et généralement utilisé pour l'amarrage protéine-ligand, le ligand étant considéré comme flexible, la protéine gardée rigide.
- le *docking flexible* enfin, traite la flexibilité des deux molécules, mais la flexibilité permise est limitée, simplifiée par des modèles.

La raison pour laquelle la flexibilité est souvent limitée au ligand est directement liée à la capacité informatique nécessaire aux algorithmes. Les tentatives d'extension de la flexibilité aux macromolécules se multiplient aujourd'hui.

Flexibilité du ligand

Trois méthodes peuvent être développées : les méthodes systématiques, les méthodes aléatoires et les méthodes de simulation [83].

Les méthodes de recherche systématique prennent en compte l'ensemble des degrés de liberté du ligand. Pour éviter des problèmes d'explosion du calcul, le ligand peut être placé par fragments sur le site actif et reconstruit par incrémentation, une minimisation étant opérée à chaque pas [249] (c'est le cas par exemple des programmes DOCK [114] et FlexX [213]). D'autres algorithmes font appel à des banques de données de conformations ramenant le problème à celui d'un corps rigide (comme FLOG [159]).

Les méthodes de recherche aléatoire modifient le ligand de manière stochastique et l'évaluent selon des probabilités d'existence. Les techniques employées à cet effet sont le Monte Carlo [202][215] (comme dans AutoDock [202]) et les algorithmes génétiques [107][151][191] (GOLD [132]). Il existe enfin les méthodes dites taboues qui, pour accepter une nouvelle configuration ou non, considèrent les conformations déjà explorées et calculent le *rmsd* (root mean square deviation) entre nouvelles et anciennes conformations [76][250]. PRO_LEADS [250] est un des algorithmes utilisant ces méthodes.

Les méthodes de simulation reposent sur la dynamique moléculaire et la minimisation d'énergie. La dynamique moléculaire est actuellement la plus courante, malgré le fait qu'elle n'atteigne généralement que des minima locaux. Quant à la minimisation d'énergie, elle est rarement utilisée seule, mais plutôt intégrée en complément des techniques précédemment citées.

Flexibilité du récepteur

Plus complexe, son traitement est moins abouti que celui du ligand. D'importants efforts sont actuellement menés pour développer des stratégies efficaces [2][3][57][79][101][112][182]. La flexibilité peut être traitée de deux façons : implicitement (*soft docking* et *cross docking*) ou explicitement. La tâche est facilitée par une connaissance préalable des régions flexibles.

La dynamique moléculaire peut apporter des éléments de réponse à ce problème, mais elle est limitée en ce qui concerne les mouvements de grande amplitude et donc les systèmes de grande taille en raison de la longueur des calculs [181][256]. Une alternative est l'utilisation d'homologues pouvant fournir, par comparaison de structures, des informations sur des mouvements de type charnière par exemple [201]. L'analyse des mouvements collectifs (modes normaux) est actuellement très développée pour décrire les mouvements de grande amplitude [180][236]. L'algorithmique robotique apporte également sa contribution, particulièrement pour la planification de mouvements de séquences protéiques très flexibles (comme les boucles) ou encore de chemins d'arrimage entre macromolécules [67][68][98][105][149][221].

Des services sont proposés sur internet pour générer des conformations et faciliter l'analyse de la flexibilité [72][173][234].

Les traitements implicites comprennent le *soft docking* et le *cross docking*. Le *soft docking* ramène la structure du récepteur à une grille et donc le problème à une recherche de complémentarité de surface. L'algorithme traite des corps rigides, mais permet une certaine flexibilité en autorisant une légère interpénétration des surfaces [150][204]. Cette méthode est peu coûteuse mais ne permet pas toujours d'améliorer le résultat, notamment en raison de l'encombrement stérique [231]. Une flexibilité partielle peut également être introduite en ajustant localement certaines chaînes latérales de la protéine [139]. La localisation des zones flexibles peut être déterminée par dynamique moléculaire, les zones rigides étant conservées dans les diverses conformations calculées [178]. Cette méthode augmente généralement la qualité de la prédiction.

Le *cross docking*, ou *ensemble docking*, réalise le docking rigide de nombreuses structures de départ [11][74][75][146]. Ces structures peuvent être obtenues par RMN [110], dynamique moléculaire [211][124], simulations Monte Carlo [215], échantillonnage thermodynamique [239], algorithmes génétiques [117][172] ... Il existe également des bibliothèques de rotamères contenant les configurations de chaînes latérales les plus favorables [13]. Le *cross docking* permet de trouver davantage de structures natives, mais aussi de mauvaises solutions, rendant le scoring plus long et difficile. Les méthodes de mouvements collectifs (modes normaux) ont également été appliquées à ce problème pour les mouvements de grande amplitude comprenant les atomes du backbone [257] et améliorent les résultats.

Les traitements explicites nécessitent une représentation explicite des molécules. Ils autorisent la flexibilité des chaînes latérales et/ou du backbone et du fait de leur coût, n'interviennent généralement qu'au stade final de la procédure de recherche lors d'une optimisation. Lorsque les changements de conformation sont de faible amplitude, une optimisation des chaînes latérales seules peut être efficace [127][223]. L'introduction de la flexibilité du récepteur et du ligand s'effectue lors d'une étape de raffinement par dynamique moléculaire par exemple (c'est le cas du programme HADDOCK [109]), par simulation Monte Carlo (RosettaDock [122][247]), par intégration d'une bibliothèque de rotamères (pour une représentation réduite des protéines : ATTRACT [255]) ...

Le traitement des changements conformationnels du backbone, complexe, présente davantage de difficultés. Sa prise en compte est pourtant incontournable car les mouvements de grande amplitude sont caractéristiques des systèmes macromoléculaires. Trouver une méthode d'analyse des mouvements de grande amplitude en conservant un coût de calcul raisonnable limite l'utilisation de la dynamique moléculaire qui nécessiterait des simulations très longues.

L'approche des modes normaux qui intègre les déformations harmoniques selon les composantes principales comme des degrés de liberté supplémentaires est aujourd'hui régulièrement intégrée aux approches de docking [179][180]. De nombreux exemples ont en effet montré que les chemins de transition suivis par les biomolécules peuvent souvent être décrits par un ou plusieurs modes normaux de basse fréquence ⁵ [100][141][236], si bien que cette approche est aujourd'hui largement développée [70][129]. Un bilan de l'approche par modes normaux et de ses utilisations est dressé par J. Ma dans un article publié en 2005 [186]. Il y pose notamment la question de la légitimité d'utiliser un modèle harmonique.

⁵L'annexe C apporte quelques précisions sur les modes normaux.

2.3.3 Le scoring

La fonction de score est généralement une fonction contenant des termes d'énergie, capable de discriminer les solutions générées pour en faire ressortir les formes natives. Elle doit également représenter un coût informatique le plus modéré possible pour évaluer rapidement les nombreuses solutions de la recherche conformationnelle. Elle peut être introduite au sein de la procédure de recherche ou implémentée à la fin.

L'établissement d'une bonne fonction de score est un important problème du docking. Il arrive souvent que la solution évaluée comme étant la plus probable ne soit pas la forme native attendue. Ceci peut être dû au fait que le complexe natif n'est pas forcément celui qui présente la plus grande surface d'accès, ou encore le plus grand nombre de liaisons hydrogène disponibles [200]. La solution peut alors être de procéder en deux étapes : la première retenant des candidats, et la seconde, plus complexe refaisant un tri parmi ces candidats [143]. Il existe différents types de fonctions de score selon les critères sur lesquels elles sont basées [136][232].

De nombreux programmes de docking possèdent une fonction de score basée sur la complémentarité géométrique, la distance intermoléculaire, le nombre de liaisons hydrogène établies... d'autres tiennent compte des interactions électrostatiques, de van der Waals ou encore des énergie de solvation.

La complémentarité géométrique est un critère très utilisé, certains programmes n'utilisent que lui. Cette approche fonctionne bien avec le *cross docking*. Avec différentes approches, ce critère est néanmoins insuffisant [200] et d'autres termes plus coûteux doivent être introduits en vue d'un second filtrage. A cette étape, le recouvrement intermoléculaire est souvent introduit et peut permettre une flexibilité partielle s'il est autorisé [204].

L'affinité chimique peut être calculée au travers de l'énergie libre de Gibbs ΔG_l à la température T [163][226] :

$$\Delta G_l = -RT \ln K_{eq}$$

où R est la constante des gaz parfaits et K_{eq} est la constante d'équilibre.

Le calcul de l'énergie libre étant cependant lourd, son application devient impossible lorsqu'il s'agit d'évaluer un grand nombre de complexes. L'estimation de l'affinité repose généralement sur un calcul d'énergie contenant un certain nombre d'approximations :

- *les fonctions de score basées sur un champ de force* calculent, par mécanique moléculaire, l'énergie d'interaction du complexe et l'énergie interne du ligand. G-Score [165] par exemple, fait appel au champ de force de Tripos, et AutoDock [191] à celui d'AMBER. Les interactions entre récepteur et ligand comprennent souvent des termes de van der Waals et électrostatiques. L'énergie interne du ligand est généralement écrite de manière similaire. Les principales limitations des fonctions de score basées sur des champs de forces proviennent du fait qu'elles ont été écrites pour des modèles en phase gazeuse et ne contiennent donc pas d'effet de solvation ni de terme d'entropie. Des extensions incluant un terme d'entropie pour le ligand (dans G-Score) et les liaisons hydrogène protéine-ligand (dans Gold [241] et Autodock) ont été ajoutées récemment.

- les *fonctions de score empiriques* tentent d'ajuster le résultat à des données expérimentales (telles que les énergies de liaison, des informations structurales par rayons X) elles-même approximées par une somme de termes non corrélés [78]. Cette méthode offre l'intérêt d'un calcul plus simple, mais dépendant des données utilisées. D'autres part, les termes d'interactions non liées peuvent être traités de manières très différentes : LUDI [78] par exemple, distingue des termes de liaisons hydrogène neutres ou ioniques alors que d'autres (comme ChemScore [113]) ne font pas de différence. Certaines fonctions traitent les effets de solvation et désolvation, comme la fonction de Fresno par exemple [147][218][228].
- les *fonctions de score "knowledge-based"* sont implémentées pour reproduire des structures expérimentales. Les complexes sont ajustés pour former des paires d'atomes connues pour interagir. De telles fonctions incluent les méthodes Potential of Mean Force (ou PMF [193][194][195], selon laquelle l'hypothèse de Boltzmann convertit les fréquences de trouver l'atome *A* du ligand à une distance *r* de l'atome *B* du récepteur en une énergie d'interaction entre *A* et *B* en fonction de la distance *r*), DrugScore [119]... L'intérêt de ces fonctions est leur grande simplicité, donnant accès au criblage d'un grand nombre de complexes. Leur inconvénient provient de leur propre implémentation qui propose un nombre limité de données.

Pour surmonter les limitations de chaque fonction de score, des fonctions de consensus combinent des éléments de diverses fonctions pré-existantes [90][93][254]. X-CSCORE [246] par exemple, combine des fonctions de score de GOLD, DOCK, Chemscore, PMF et FlexX. Cette approche diminue parfois le nombre de faux positifs.

Actuellement, les voies explorées pour améliorer les fonctions de score portent sur l'intégration des effets de solvation [131][233] et d'entropie de rotation [5].

2.4 Conclusion

Au cours de ce second chapitre, nous avons détaillé les progrès en matière d'informatique et de modélisation ayant permis l'application de la simulation aux problèmes complexes de la biologie. Les méthodes calculatoires sont désormais mises à profit pour déterminer la structure et la dynamique des biomolécules, mais aussi les mécanismes de l'amarrage moléculaire.

Néanmoins, il est actuellement très difficile de caractériser et de prédire une interaction moléculaire sur une base structurale du fait de la grande flexibilité des molécules biologiques. Les tentatives d'introduction de la flexibilité dans les algorithmes que nous avons recensées dans cette partie sont à la fois révélatrices des efforts interdisciplinaires réalisés au cours de ces dernières années, mais également de la difficulté constante de travailler sur des macromolécules.

Les approches de docking développées à l'heure actuelle ne permettent pas de sonder la totalité de l'espace conformationnel en raison de la trop haute dimensionnalité des systèmes. Une préoccupation constante étant la conservation d'un temps de calcul raisonnable, la simplification des modèles est privilégiée, via une prise en compte limitée de la flexibilité moléculaire, surtout en ce qui concerne celle des récepteurs.

Face à ce constat, nous avons développé une nouvelle méthode permettant de prendre en compte la flexibilité totale des macromolécules, la méthode des Modes Statiques, que nous détaillons dans le chapitre suivant.

Chapitre 3

La méthode des Modes Statiques

Sommaire

3.1	Origine et concept	48
3.1.1	Faire face aux limites actuelles	48
3.1.2	Limites et légitimité de l'application des Modes Normaux	48
3.1.3	Apports de la méthode des Modes Statiques	52
3.2	La méthode des Modes Statiques	53
3.2.1	Méthodologie	53
3.2.2	Le logiciel FLEXIBLE	55
3.2.3	Discussion	57
3.3	Conclusion	57

Au cours des deux premiers chapitres, nous avons présenté l'importance des réarrangements structuraux des molécules biologiques, par rapport à leur fonction et à leur activité, mais aussi les limites actuelles liées à leur étude.

Introduire la flexibilité moléculaire dans les algorithmes est un véritable enjeu pour les équipes de modélisation, aussi bien pour accéder aux propriétés intrinsèques des molécules, qu'aux changements conformationnels induits par des facteurs extérieurs (modifications du milieu, approche d'un ligand...).

En réponse à ce problème, nous proposons une alternative aux approches existantes avec une nouvelle méthode : la méthode des Modes Statiques. Elle s'appuie sur le concept "*induced-fit*", c'est-à-dire l'accommodement induit, et permet de cartographier les déformations d'une macromolécule induites par une excitation extérieure [266].

3.1 Origine et concept

L'idée d'origine est de pouvoir explorer un grand nombre de conformations avec un algorithme simple, ce qui est actuellement impossible avec des méthodes telles que la dynamique moléculaire, trop coûteuse en temps.

Dans une optique de simplification, nous calculons les déformations permanentes d'une molécule soumise à une excitation extérieure (c'est-à-dire une interaction), sans tenir compte des états intermédiaires entre deux conformations à l'équilibre.

3.1.1 Faire face aux limites actuelles

La méthode des Modes Statiques est une méthode de traitement de la flexibilité, proposée en alternative à l'utilisation de la dynamique moléculaire ou de l'analyse par modes normaux, jusque-là considérées comme des outils de prédilection.

La dynamique moléculaire (voir paragraphe 2.1.3) est aujourd'hui reconnue comme un outil puissant, précis et adapté au docking, puisqu'en appliquant les lois de la mécanique classique de Newton, elle tient compte de toutes les forces mises en jeu. Cependant, cet atout est aussi sa principale limite, car son application à des systèmes complexes tels que les biomolécules requiert un temps de calcul rédhibitoire pour une exploration complète de l'espace conformationnel, ainsi que pour des mécanismes biologiques longs. D'autre part, la richesse des informations qu'elle fournit rend son exploitation extrêmement difficile. Ces informations sont par ailleurs soumises à d'importantes fluctuations, dues aux traitements statistiques usités par nécessité, mais aussi aux oscillations du système à température ambiante. Rappelons enfin que le résultat d'une dynamique moléculaire est très sensible aux conditions initiales. De très nombreux et longs calculs seraient donc nécessaires pour couvrir tout l'espace conformationnel, ce qui semble impossible dans la pratique et contribue au caractère partiel de la flexibilité explorée. L'utilisation d'un modèle aussi complet et complexe dans sa description doit donc peut-être laisser place à une approche plus simple, ce que nous proposons avec les modes statiques en donnant directement les déformations finales, sans calculer les oscillations transitoires.

L'analyse par modes normaux est également une approche simplifiée : elle ne tient compte que des vibrations spontanées de la molécule, ce qui, par nature, limite ses domaines d'application. Le paragraphe suivant développe les limites des modes normaux et pose la question de la légitimité de leur application à la dynamique des biomolécules.

3.1.2 Limites et légitimité de l'application des Modes Normaux

L'approche par Modes Normaux, adoptée depuis les années 1980 [264], est actuellement très utilisée. Elle consiste à modéliser la dynamique d'une protéine ou d'un complexe macromoléculaire comme la somme de mouvements élémentaires⁶.

⁶L'annexe C introduit les fondements des modes normaux.

Il est indiscutable que l'approche par modes normaux est actuellement un succès. Elle est d'ores et déjà utilisée pour le raffinement de structures expérimentales de basse résolution [100][277], par exemple pour des données obtenues par cryo-microscopie électronique [299][300] et des structures cristallographiques [297]. Des structures ainsi résolues ont d'ailleurs été déposées à la Protein Data Bank [258].

Néanmoins, on peut se questionner au sujet de sa validité et de la légitimité de son utilisation [179][180] :

- pourquoi peut-on utiliser l'approximation harmonique pour décrire des changements de grande amplitude a priori anharmoniques ?
- quel est son domaine de validité ?
- quelle est l'amplitude des mouvements selon les modes normaux ?
- peut-on utiliser des modes harmoniques pour modéliser des conformations de transition (loin des positions d'équilibre) lors du docking ?
- quelles énergies sont impliquées ?
- comment identifier les modes pertinents ?

Au regard des ces interrogations, peut-on alors imaginer utiliser l'approche des Modes Normaux à des fins prédictives [305] ?

Difficultés liées au calcul

Le calcul des modes normaux étant basé sur la diagonalisation de la matrice des dérivées secondes de l'énergie, sa lourdeur a longtemps limité les applications à des systèmes d'assez petite taille (quelques milliers d'atomes). Aujourd'hui, des développements algorithmiques (notamment les modèles "coarse-grained" proposant une représentation simplifiée des protéines ou "elastic network" [141][236][260][261][270][271][280][287][288][289][298][304]) ont permis l'application des modes normaux à la déformation des complexes et des systèmes de très grande taille [263] [279] [281] [306][307]. La méthode des Modes Statiques quant à elle, ne nécessite pas de calcul lourd, l'opération principale étant l'inversion de la matrice Hessienne. Elle n'est donc pas limitée par le nombre d'atomes, et surtout, ne nécessite pas d'approximation.

Outre la lourdeur de la résolution, se pose le problème de la minimisation préalable de la structure étudiée. En effet, pour aboutir à des solutions correctes, particulièrement dans le cas de systèmes complexes, la structure correspondant au minimum d'énergie doit être déterminée avec une grande précision. Si ce n'est pas le cas, cela conduit à trouver des valeurs propres négatives de la matrice Hessienne, et donc des fréquences imaginaires. L'étape de la minimisation est donc un obstacle en elle-même en nécessitant un temps de calcul important. En revanche, atteindre l'équilibre n'est pas une condition nécessaire dans la méthode des Modes Statiques, puisque ces derniers résultent de l'application de forces extérieures induites par le mouvement relatif de deux molécules en évolution et en interaction lors du docking.

Restriction aux modes de basse fréquence

De nombreux travaux se sont attachés à montrer que des modes de basse fréquence (généralement inférieure à 10 cm^{-1}) peuvent décrire les mouvements de grande amplitude des macromolécules [267][268][283][284][285][286][293][294][301][302][303]. Ceci signifierait que la flexibilité des biomolécules est une propriété intrinsèque liée à leur structure, facilitant les changements conformationnels nécessaires à leur fonction [179]. Cette idée est également l'idée de base de la méthode des Modes Statiques, mais où l'on tient également compte de la présence des forces extérieures qui induisent ces déformations.

En revanche, outre la comparaison à l'expérience, quels éléments théoriques permettent-ils de justifier la seule utilisation des modes de basse fréquence ? A température ambiante, seuls les modes normaux de fréquence supérieure à environ 10^{13} Hz sont bien définis, les autres étant mêlés les uns aux autres, en raison de l'agitation thermique⁷. Or, les temps de réaction intervenant lors des amarrages de macromolécules sont plus lents, faisant intervenir des mouvements de grande amplitude, et donc incompatibles avec les vibrations thermiques.

La méthode des Modes Statiques surmonte cette limite en remplaçant les vibrations induites par l'agitation thermique par des déformations permanentes, induites par des forces extérieures appliquées sur la molécule.

Approximation harmonique

De nombreuses expériences ont montré que les mouvements de grande amplitude et de basse fréquence des biomolécules sont hautement anharmoniques [262][269][282][291][295]. Comment alors peut-on utiliser une approximation harmonique et obtenir des résultats cohérents ?

Sur un temps très court, au début d'un changement conformationnel par exemple, la déformation de faible amplitude permet l'approximation des modes normaux [99]. On peut cependant se demander jusqu'à quelle amplitude on peut déformer le long des modes normaux. A quel niveau apparaît l'incohérence entre l'amplitude des vibrations thermiques des oscillateurs harmoniques et celle des biomolécules observées par l'expérimentateur ?

En théorie, la mécanique statistique prévoit que pour un mode normal de basse fréquence calculé selon les méthodes de mécanique moléculaire typiquement utilisées, le carré de l'amplitude moyenne de la fluctuation thermique soit très petit à 300K, soit de l'ordre de 1Å^2 (selon la relation $k_B T/k$, k_B étant la constante de Boltzman, T la température et k la constante de raideur) [265]. Or il s'avère que les amplitudes observées, par exemple lors d'une déformation induite lors du couplage avec un ligand, peuvent être beaucoup plus importantes. On sort en effet du cadre des fluctuations thermiques spontanées de la molécule pour aborder les fluctuations induites par une perturbation extérieure, apportant un gain d'énergie permettant de gagner de l'amplitude.

⁷Selon la statistique de Bose-Einstein [259], le nombre moyen de phonons n de fréquence de vibration ν et présents à l'équilibre thermique à la température T , est donné par $n = \frac{1}{e^{h\nu/k_B T} - 1}$ (où apparaît la constante de Boltzmann : $k_B = 8.617 \cdot 10^{-5} \text{eV} \cdot \text{K}^{-1}$ et la constante de Planck : $h = 4.1357 \cdot 10^{-15} \text{eV} \cdot \text{s}$).

Si $n \ll 1$ ($h\nu \gg k_B T$), alors les phonons sont indépendants les uns des autres. Par contre, si $n \gg 1$ ($h\nu \ll k_B T$), les phonons sont nombreux et interagissent avec des phonons d'autres fréquences.

Comment donc inclure les modes normaux dans une procédure de docking ?

Les modes normaux ne peuvent pas, à proprement parler, convenir à un calcul d'interaction, ils ne sont valides qu'autour de la position d'équilibre. A l'état libre, une macromolécule vibre autour de son état d'équilibre avec un mouvement pouvant alors être décrit par les modes normaux. Son énergie augmente à mesure que la molécule s'éloigne de cet équilibre en suivant la direction du mode normal et atteint une conformation qui ne serait pas stable sans le ligand, c'est-à-dire que la molécule ne pourrait atteindre seule [274][278] (figure 3.1).

Un argument actuellement avancé pour utiliser malgré tout l'analyse vibratoire serait que les mouvements de grande amplitude suivent des trajectoires harmoniques décrites par les modes normaux. En revanche, l'anharmonicité, en majeure partie liée aux nombreuses faibles barrières d'énergie rencontrées lors des transitions conformationnelles (dues notamment aux mouvements rapides des chaînes latérales), n'est pas décrite par les modes normaux.

Sur ce point encore, le problème ne se pose pas dans la méthode des Modes Statiques, qui permet de décrire tous les mouvements sans restriction, des déplacements atomiques, aux grands changements conformationnels.

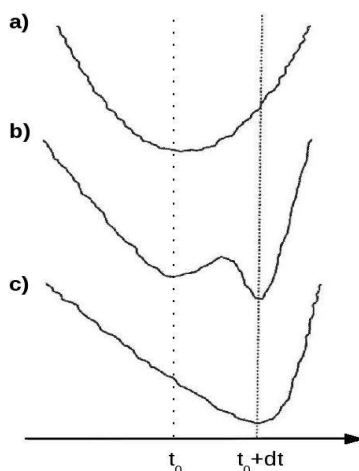


FIG. 3.1 – Profils énergétiques d'une molécule passant de l'état libre à l'état lié. a) équilibre de la molécule libre. b) modification de l'énergie durant l'approche du ligand. c) nouvelle énergie de la molécule à l'état lié. *Source Structure* [180].

Signification et pertinence des Modes Normaux

L'information attribuée aux modes normaux porte surtout sur la direction des mouvements de grande amplitude et non sur leur durée. Ces mouvements sont généralement beaucoup plus lents que l'estimation faite par les valeurs propres des modes normaux. Les utilisateurs attribuent ceci au fait que les calculs sont souvent effectués dans le vide alors que la présence du solvant ralentit les mouvements [276], ou encore au fait que les changements de grande amplitude suivent des trajectoires harmoniques, mais pas les temps d'un oscillateur harmonique libre.

Actuellement, seuls les vecteurs propres sont considérés lors de l'analyse vibrationnelle, ce qui pose problème : comment la diagonalisation peut-elle donner des vecteurs propres corrects et des valeurs propres ne leur correspondant pas ?

Pour avoir les bonnes fréquences, il faut inclure les molécules d'eau dans la matrice Hessienne. Ceci est impossible, car les fréquences vibrationnelles dans les milieux liquides, observées par spectroscopies Raman et infrarouge, ne peuvent pas être déterminées par la diagonalisation de la matrice Hessienne, mais par la transformée de Fourier de la fonction de corrélation des vitesses, calculées par la dynamique moléculaire.

Enfin, parmi les modes calculés, encore faut-il savoir faire le tri et identifier ceux qui peuvent avoir un sens et une existence biologique. La seule certitude est que les mouvements fonctionnels peuvent être décrits à l'aide d'un ou quelques modes de basse fréquence demandant peu d'énergie. La solution réside alors dans la comparaison avec les données expérimentales. Dans cette hypothèse, on ne peut plus parler de modélisation prédictive.

3.1.3 Apports de la méthode des Modes Statiques

En alternative à la dynamique moléculaire et à son coût rédhibitoire pour une exploration complète de la flexibilité, nous proposons de ne calculer que l'état final des déformations, et non les états de transition. En alternative aux limites d'application des modes normaux, nous suggérons de remplacer les vibrations spontanées (thermiques) par des déformations induites par des excitations extérieures.

La méthode des Modes Statiques nous permet ainsi de sortir du cadre de la dynamique intramoléculaire pour considérer celle des interactions intermoléculaires. Cette caractéristique fait d'elle une approche convenant parfaitement à une procédure de docking. Les modes statiques sont conçus pour que la molécule puisse être soumise à une interaction ultérieure. Les déformations possibles sont calculées en fonction de paramètres représentant chacun un site particulier d'interaction.

L'objectif est de pouvoir réaliser le docking, avec un temps minimum de calcul, le gain étant réalisé par un stockage préalable, dans une banque de données, des modes statiques correspondant à chaque molécule. Ainsi, au moment de l'interaction, ces données seront utilisées pour minimiser l'énergie. En fonction des sites d'amarrage moléculaire, toujours en nombre réduit, l'énergie d'interaction ne dépendra que de quelques paramètres.

Nous pouvons en plus envisager, moyennant une augmentation du temps de calcul, d'utiliser des potentiels anharmoniques dans l'expression de l'énergie. Il s'agit d'un autre avantage des modes statiques par rapport aux modes normaux qui sont fondamentalement des oscillateurs harmoniques.

3.2 La méthode des Modes Statiques

3.2.1 Méthodologie

Le calcul des modes statiques est basé sur la matrice Hessienne qui contient les dérivées secondes de l'énergie par rapport aux coordonnées atomiques. Les éléments de cette matrice peuvent être déterminés par un calcul, quantique ou classique, de l'énergie, en fonction de la taille du système à traiter, ou encore de la précision voulue. Un calcul préalable d'énergie totale est donc nécessaire pour fournir les deux éléments de base à la résolution des modes statiques : la matrice Hessienne et les coordonnées atomiques de la molécule étudiée.

Afin d'éliminer les trois translations et les trois rotations qui laissent la molécule invariante, trois atomes (n_i, n_j, n_k) sont fixés respectivement en translation, en azimuth et en longitude. On obtient alors une matrice de taille $(3N-6) \times (3N-6)$, N étant le nombre d'atomes de la molécule.

L'énergie totale E peut être développée en série entière, comme dans l'équation suivante (3.1) :

$$E = E_0 + H_{i,j}x_i x_j + K_{i,j,k}x_i x_j x_k + L_{i,j,k,l}x_i x_j x_k x_l \dots \quad (3.1)$$

où les $H_{i,j}$ sont les éléments de la matrice Hessienne. Les matrices $K, L, \dots$ représentent les dérivées d'ordres supérieurs de l'énergie, soit la contribution anharmonique, et x_i les coordonnées Cartésiennes des déplacements atomiques pour une configuration à l'équilibre d'énergie E_0 . Notons que la matrice H contenant les dérivées secondes de l'énergie peut être fournie par des logiciels de calcul d'énergie (tels que GAUSSIAN03 et AMBER8). La matrice des dérivées troisièmes K (qui traduit l'asymétrie du potentiel anharmonique et n'est pas utilisée actuellement) peut être calculée à l'aide de la méthode des différences finies.

Le vecteur force, gradient de l'énergie, peut alors s'écrire à l'aide des termes de la matrice H' :

$$-F_i = \frac{\partial E}{\partial x_i} = \sum_j H'_{i,j} x_j \quad (3.2)$$

La version actuelle du code se place dans l'approximation harmonique. Dans ce cas, $H = H'$ est la matrice Hessienne. Cependant, la théorie des Modes Statiques permet d'inclure les effets non linéaires. En prenant en compte les termes non linéaires, H' n'est plus la matrice Hessienne, mais contient $K, L, \dots$. Elle dépend des coordonnées atomiques. Le système d'équations (3.2) est toujours correct, mais n'est plus linéaire. L'algorithme que nous avons utilisé jusqu'ici et que nous présentons dans la suite, peut également être utilisé pour la résolution, dans la mesure où la partie anharmonique reste dans le cadre d'une perturbation et ne constitue pas la contribution majeure à l'énergie. Il faut alors envisager un temps de calcul plus long selon l'importance des perturbations, mais aussi une importante augmentation des capacités de stockage.

A ce stade, deux approches peuvent être envisagées pour résoudre le problème. A l'équilibre et en l'absence de forces extérieures, tous les déplacements x_j sont nuls. En appliquant une force sur un seul atome dans une direction x, y, z donnée, c'est-à-dire, en faisant disparaître toutes les forces F_i exceptée une, on peut alors trouver les déplacements x_j en résolvant le système d'équations (3.2). Les déplacements sont alors normés à la force appliquée.

Une autre façon d'aborder le problème est de considérer un déplacement donné x_l , résultant de l'application de la force associée F_l , les déplacements sont alors normés au déplacement x_l .

Les deux approches sont équivalentes. La seconde présente néanmoins certains avantages : d'une part, elle peut se transposer plus aisément en coordonnées internes, associées aux coordonnées généralisées dans le formalisme Lagrangien. Par ailleurs, elle permet d'éviter le calcul direct des forces de répulsion entre atomes proches, difficiles à estimer.

Les forces répulsives présentent en effet des variations importantes et rapides sur de très petits déplacements atomiques. L'anharmonicité est poussée à l'extrême : il suffit d'une légère déformation pour quasiment annuler, ou même inverser, les forces répulsives. En d'autres termes, le potentiel répulsif, en $1/r^{12}$, diverge rapidement pour les petites distances. Il peut alors être remplacé par une droite verticale : c'est l'approximation des sphères dures, qu'il est alors préférable d'utiliser pour éviter les itérations multiples [273][34]. On peut alors supprimer la notion de force pour ne tenir compte que des déplacements ; mais alors, on ne peut plus normer la déformation par rapport à la force.

Nous choisissons donc d'adopter la seconde approche : x_l est associé à la contrainte de Lagrange [272][296], F_l agit comme une force de réaction dont la connaissance n'est pas nécessaire dans le formalisme Lagrangien.

Dans un premier temps, la l ème ligne de la matrice Hessienne associée à la contrainte F_l est éliminée du système d'équations (3.2). Il en va de même pour la l ème colonne, contenant tous les termes en x_l . En nommant cette colonne B , et A le reste de la matrice $(3N-7) \times (3N-7)$, il nous reste alors à résoudre le système $A.X + B = 0$, où le vecteur X contient les déplacements atomiques imposés par la contrainte.

Nous avons choisi la méthode de Gauss-Seidel (voir annexe D), bien adaptée au traitement de ce problème. En effet, outre une convergence rapide, cette méthode présente l'avantage de nécessiter peu de stockage, et donc d'engendrer peu de perte d'information et d'accumulation d'erreur [290]. Elle convient donc particulièrement bien au calcul des déformations et de leur propagation, même faible, loin de la source d'excitation.

La solution X du système, c'est-à-dire la déformation moléculaire correspondant à une contrainte spécifique, est appelée un "mode statique". Ainsi calculé, le mode statique déterminé est normé au déplacement x_l . Par conséquent, en faisant appel au système (3.2), la l ème équation peut être utilisée pour déterminer la force de réaction F_l , à laquelle le mode peut être également normé par la suite.

Enfin, l'énergie E associée à la déformation peut être facilement déduite :

$$E = \sum_{ij} A_{ij} x_i x_j$$

Par élimination successive des lignes de la matrice Hessienne, l'ensemble des modes statiques, ou bien seulement des modes choisis, peuvent être déterminés. Ces modes, caractéristiques d'une molécule unique, peuvent ensuite être stockés dans une banque de donnée pour un usage ultérieur tel que le docking, et ainsi simplifier la procédure en maintenant des temps de calcul modestes.

3.2.2 Le logiciel FLEXIBLE

Dans ce paragraphe, nous détaillons les principales étapes du code logiciel FLEXIBLE que nous avons développé. Ces étapes sont résumées dans l'organigramme présenté sur la figure 3.2.

- La première partie est consacrée à la lecture d'un fichier de sortie issu d'un calcul d'énergie totale. Nous avons jusque-là utilisé les logiciels GAUSSIAN03 pour un modèle quantique et AMBER8 pour un champ de force classique. Cette étape doit nous permettre de récupérer la matrice Hessienne et les coordonnées de la molécule étudiée.
- (1) L'utilisateur choisit trois atomes à fixer en translation ($n_i(0,0,0)$), en azimuth ($n_j(x,0,0)$) et en longitude ($n_k(x,y,0)$). Nous opérons alors un changement de base en nous plaçant par rapport aux atomes fixés. Pour cela, nous considérons les vecteurs $\vec{v}_1$ et $\vec{v}_2$ qui relient respectivement les points n_i et n_j , et n_i et n_k ; ces vecteurs sont alors normés, puis nous calculons $\vec{v}_3 = \vec{v}_1 \times \vec{v}_2$ afin d'obtenir une base orthonormée ($\vec{v}_1, \vec{v}_2, \vec{v}_3$).
- (2) Le premier sous-programme *RMAT* intervient alors et permet :
 - avec *NewMat* :
 - dans *RotateX* : de calculer les coordonnées dans le nouveau repère
 - dans *Rotate* : de calculer les éléments de la nouvelle matrice Hessienne D dans la base $U = (\vec{v}_1, \vec{v}_2, \vec{v}_3)$, 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.
- (3) La fonction *LCMOD* permet de choisir et de recenser les modes à traiter.
- (4) La fonction *LINMODE* permet de résoudre le système $A \cdot X + B = 0$ en suivant la méthode de Gauss-Seidel. Il s'agit d'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 (voir annexe D). Nous calculons x_{k+1} à partir de x_k avec 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 :

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

Nous arrêtons le calcul dès que x_k et x_{k+1} sont très proches, c'est-à-dire, avec ε fixé, lorsque :

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

Enfin, on stocke les déformations obtenues avant de calculer l'énergie correspondante avec l'équation :

$$E = \sum_{ij} A_{ij} x_i x_j$$

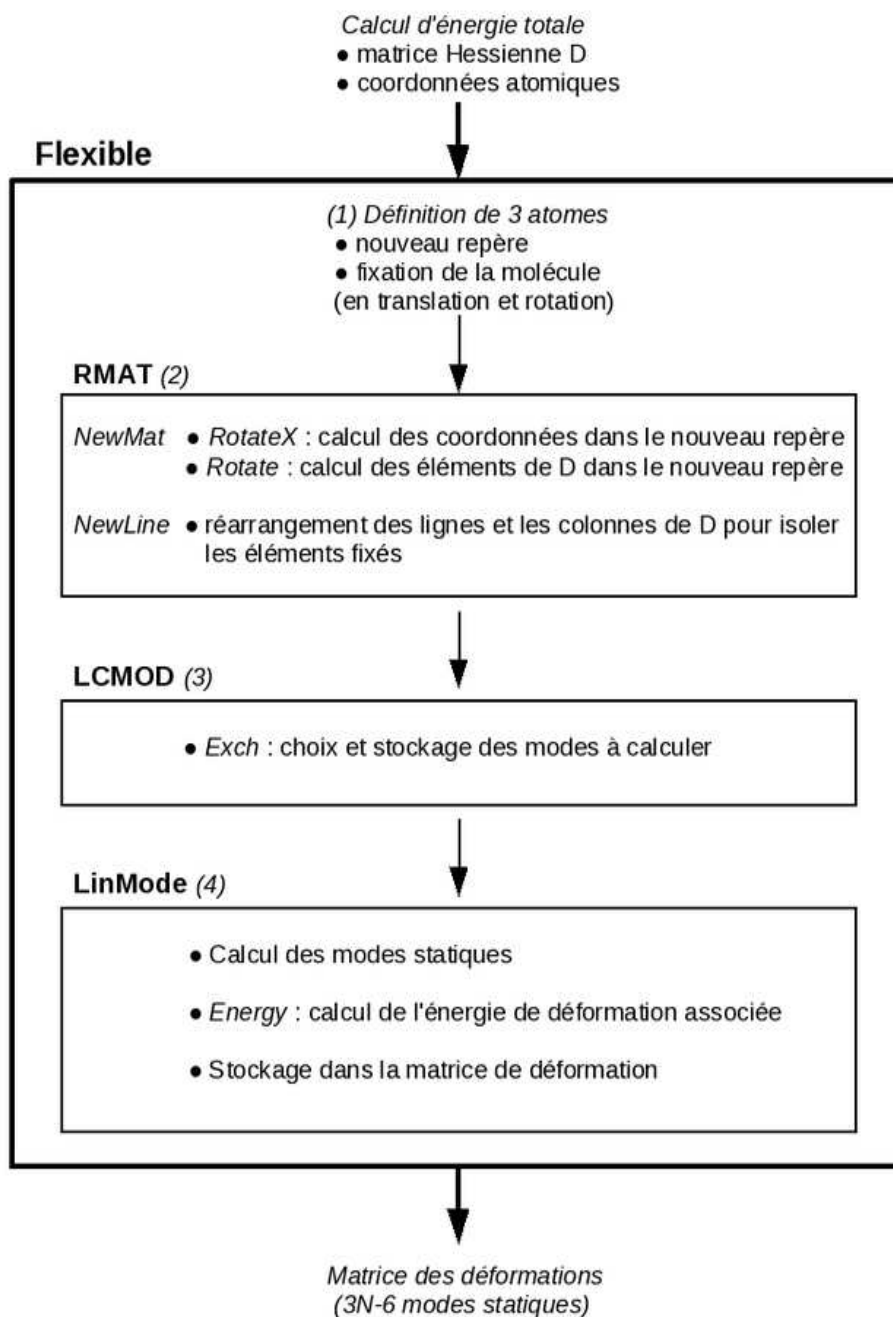


FIG. 3.2 – Organigramme présentant la procédure actuelle du code Flexible. Un calcul d'énergie totale réalisé au préalable est requis pour fournir la matrice Hessienne et les coordonnées de la molécule en entrée. Flexible fait alors intervenir un certain nombre de sous-programmes dont les principaux sont indiqués ici (RMAT, LCMOD et LinMode). Une fois la résolution achevée, l'ensemble des modes calculés est stocké dans la "matrice des déformations" pour une exploitation ultérieure.

3.2.3 Discussion

Une des forces de la méthode des Modes Statiques est de pouvoir prendre en compte l'anharmonicité. Ceci est particulièrement important dans le cas des molécules biologiques puisque les termes non linéaires sont à l'origine de nombreuses conformations existantes [275][292]. La version actuelle du code se place dans l'approximation harmonique. Cependant, nous avons vu que la théorie des modes statiques permet d'inclure les effets non linéaires. L'extension de la méthode des Modes Statiques à des modes non linéaires est envisagée dans de futurs développements.

Le second atout de notre approche est de ne nécessiter qu'un seul calcul pour l'ensemble des déformations, et ce dans un temps très raisonnable. La principale opération à laquelle l'algorithme fait appel est en effet une inversion de matrice, ce qui rend le calcul beaucoup plus aisé que la diagonalisation nécessaire à la détermination des modes normaux. De fait, la résolution des modes statiques ne fait pas intervenir d'approximation. Comme nous le montrons dans le dernier chapitre de cette thèse, l'algorithme actuel utilise moins d'une seconde CPU pour un système de quelques dizaines d'atomes à 8 heures CPU pour un complexe protéine/ligand de 2616 atomes (avec deux processeurs double coeur Intel Xeon à 1.60 GHz). Une seconde version du logiciel, accélérée par un tri préalable des voisins en interaction, est en cours.

Le stockage des modes statiques au fur et à mesure de leur calcul présente un double avantage : le gain de temps et la simplification des procédures ultérieures.

Jusqu'à présent, nous n'avons utilisé la méthode des Modes Statiques que pour l'étude d'une molécule unique dans le cadre de validations préliminaires, c'est-à-dire lors de l'exploration des changements conformationnels.

Le but ultime des modes statiques demeure bien entendu le problème du docking. Dans une procédure adaptée, les modes statiques stockés sont appelés par rapport à des sites spécifiques d'interaction. Au fil de l'évolution des positions relatives des deux molécules l'une par rapport à l'autre, les forces intermoléculaires s'exercent sur les sites d'interaction et induisent ainsi des déformations en faisant appel aux modes statiques pré-stockés. Durant la procédure de docking, les seules variables dynamiques sont donc les coordonnées des atomes impliqués dans l'interaction. Les paramètres concernant les autres atomes sont contenus dans les modes statiques. En effet, les deux molécules sont approchées pas à pas, dans le sens d'une minimisation de l'énergie totale. Seules l'énergie et les forces d'interaction entre molécules sont exprimées explicitement pour toutes les connexions inter-atomiques. En revanche, les énergies et forces internes des molécules elles-mêmes sont traitées au travers des modes statiques. Le temps ainsi gagné permet l'exploration d'un grand nombre de situations et le test de différents sites d'interaction.

3.3 Conclusion

Dans ce troisième chapitre, nous avons présenté la méthode des Modes Statiques, une procédure que nous pensons simple et efficace pour traiter la flexibilité dans les interactions entre biomolécules. Cette nouvelle approche permet la détermination du champ de déformation d'une molécule soumise à une excitation extérieure.

De part la simplicité de l'algorithme, la méthode des Modes Statiques s'avère être une alternative très compétitive par rapport aux méthodes utilisées à l'heure actuelle. D'une part, la flexibilité totale peut être explorée dans un temps de calcul très accessible et sans approximation, mais en plus, le concept permet tout-à-fait d'introduire des termes anharmoniques. Des développements sont prévus afin d'améliorer le code actuel. Une optimisation du code est en cours ; par ailleurs, certaines extensions sont à venir, notamment l'introduction d'effets anharmoniques, ou encore le passage aux modes généralisés qui permettra d'explorer des parties flexibles de la molécule, par rapport à des parties plus rigides, en introduisant des coordonnées internes ou leurs combinaisons.

Une fois les modes statiques calculés, ils peuvent être utilisés directement pour sonder une molécule unique, ou encore dans une procédure de docking. Il s'agit là de l'ambition finale de la méthode. Le dernier chapitre présente les premiers résultats que nous avons obtenus sur des molécules uniques au travers d'exemples variés. Des polymères aux protéines, en passant par l'ADN, nous utilisons les modes statiques pour explorer un site actif, les mécanismes d'activation et d'inhibition, l'allostérie et bien d'autres comportements des systèmes choisis.

Chapitre 4

Résultats

Sommaire

4.1	Quelques outils pour l'étude de la flexibilité	60
4.1.1	Localisation des modes	60
4.1.2	Optimisation des forces	61
4.1.3	Pincements des couples d'atomes	62
4.2	Application des modes statiques à un polymère thermosensible : le PNIPAM	63
4.2.1	Présentation du polymère	63
4.2.2	Premières observations sur la flexibilité du polymère	64
4.2.3	Etude du site actif d'un dimère de NIPAM : la liaison hydrogène inter-monomères	65
4.2.4	Extension au cas d'un tétramère de NIPAM	68
4.2.5	Conclusion	70
4.3	Etude de la flexibilité des sucres dans les acides nucléiques	70
4.3.1	Les plissements du sucre : concept de pseudorotation	71
4.3.2	Cas d'un nucléotide : la désoxythymidine monophosphate	72
4.3.3	Etude de l'apport de nucléotides modifiés sur la stabilité de duplex d'acides nucléiques	76
4.3.4	Conclusion	81
4.4	Etude du site actif de la protéase du VIH-1	82
4.4.1	Description de la molécule : structure et mécanismes	82
4.4.2	Evaluation de la stabilité du site actif	84
4.4.3	Exploration du mécanisme d'ouverture	88
4.4.4	Simulation des effets de l'arrimage d'un ligand	92
4.4.5	Conclusion	94
4.5	Conclusions	95

Dans ce dernier chapitre, nous présentons diverses applications des modes statiques. Les systèmes étudiés sont très variés, des polymères aux protéines, en passant par les acides nucléiques. Leur exploration nous a permis de valider l'algorithme Flexible d'une part, et par

ailleurs de porter un nouveau regard sur leurs propriétés via une approche très différente de celles adoptées jusque-là.

Au cours des paragraphes suivants, nous nous intéressons notamment au comportement d'un polymère thermosensible, à la flexibilité des sucres de l'ADN, à l'apport de nucléotides modifiés pour la stabilité d'acides nucléiques et au mécanisme d'ouverture de la protéase du VIH-1.

4.1 Quelques outils pour l'étude de la flexibilité

Nous commençons par présenter, dans ce premier paragraphe, les outils de calcul qui nous ont permis d'explorer la flexibilité des molécules.

4.1.1 Localisation des modes

Ce critère nous permet de déterminer si les modes sont localisés, c'est-à-dire si les contraintes appliquées aux atomes entraînent le déplacement d'un grand nombre d'atomes ou non. Pour cela, nous partons d'un calcul des moments μ_n (4.1) :

$$\mu_n = \sum_i \left[\sum_{\alpha} (x_{i\alpha})^2 \right]^n \quad (4.1)$$

où les x_i sont les déplacements des atomes i dans la direction α (x, y, z). La localisation l est ensuite déduite à partir de μ_1 et μ_2 par la relation (4.2) [313][367] :

$$l = \frac{\mu_1^2}{\mu_2} \quad (4.2)$$

Si les déplacements $x_{i\alpha}$ sont nuls sauf pour un atome $i \neq 0$, on déduit, à partir de la relation (4.1), que $\mu_n = \left[\sum_{\alpha} (x_{0\alpha})^2 \right]^n$. La localisation vaut alors 1, comme le montrent les équations suivantes :

$$\left. \begin{aligned} \mu_1 &= \sum_{\alpha} (x_{0\alpha})^2 \\ \mu_2 &= \left[\sum_{\alpha} (x_{0\alpha})^2 \right]^2 = \mu_1^2 \end{aligned} \right\} \quad l = 1 \quad (4.3)$$

En revanche, si les déplacements sont équivalents pour tous les atomes, $\sum_{\alpha} (x_{i\alpha})^2$ est indépendant de i , et la relation (4.1) devient $\mu_n = N \left[\sum_{\alpha} (x_{0\alpha})^2 \right]^n$. Dans ce cas, la localisation vaut N :

$$\left. \begin{aligned} \mu_1 &= N \sum_{\alpha} (x_{0\alpha})^2 \\ \mu_2 &= N \left[\sum_{\alpha} (x_{0\alpha})^2 \right]^2 = \frac{\mu_1^2}{N} \end{aligned} \right\} \quad l = N \quad (4.4)$$

On trouve donc, pour la localisation d'un mode, un nombre allant de 1 à N , c'est-à-dire un indicateur de la propagation de la déformation.

4.1.2 Optimisation des forces

Il s'agit d'un outil dont nous nous sommes servis tout au long de cette thèse. On cherche à optimiser la direction de la force à appliquer sur chaque atome de façon à maximiser en réponse les variations d'une distance ou d'un angle par exemple.

Maximisation d'une distance

Nous nous trouvons face à un problème d'optimisation sous contrainte, puisque nous voulons ajuster l'orientation de la force tout en conservant sa norme.

Nous nous plaçons ici dans le cas simple où l'on veut maximiser la distance entre deux atomes N_1 et N_2 . On extrait de la matrice des modes statiques deux 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 de la puissance itérée [342].

- La variation maximale dr_{max} est reliée à la force optimale $|\psi_{max} \rangle$ par la relation : $|dr_{max} \rangle = m|\psi_{max} \rangle$.
- Afin de déterminer $|\psi_{max} \rangle$, on définit tout d'abord un vecteur quelconque $|\psi_0 \rangle$ normé. Une première approximation de la valeur propre associée λ_0 est donnée par la relation :

$$\lambda_0 = \frac{\langle \psi_0 | h | \psi_0 \rangle}{\langle \psi_0 | \psi_0 \rangle} \text{ où } h = m^t m$$

- Chaque nouvelle force est calculée par la relation $|\psi_{n+1} \rangle = h^n |\psi_n \rangle$. Le critère d'arrêt est défini à l'aide de la valeur propre λ_n calculée à chaque itération par l'équation suivante :

$$\lambda_n = \frac{\langle \psi_n | h^n | \psi_n \rangle}{\langle \psi_n | \psi_n \rangle}$$

- On continue jusqu'à ce que $\frac{\lambda_{n+1}}{\lambda_n}$ converge. Lorsqu'on atteint la convergence, on garde $|\psi_{max} \rangle$ comme force optimale.
- Cette force $|\psi_{max} \rangle$ est alors appliquée au système et l'on peut mesurer la variation de notre distance N_1N_2 : $|dr_{max} \rangle = m|\psi_{max} \rangle$. On en déduit enfin l'ouverture de la liaison :

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

$|u \rangle$ étant un vecteur unitaire selon la liaison N_1N_2 .

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 $m^t m$.

Nous nous plaçons ici dans le cas de la maximisation d'un angle dièdre. Nous cherchons alors ϕ tel que $|d\phi^2|$ est associé à la plus grande valeur propre de la matrice $A = m\vec{\nabla}^t \cdot \vec{\nabla} m$, où $\vec{\nabla}$ est l'opérateur Nabla : $\vec{\nabla} = \left(\frac{\partial}{\partial x} \frac{\partial}{\partial y} \frac{\partial}{\partial z} \right)$. Or le déterminant de cette matrice est nul : A possède $n - 1$ valeurs propres nulles et une valeur propre non nulle telle que :

$$\lambda_0 = \sum \lambda_i = \text{Tr}(A) = \sum_i \left(\frac{\partial \phi}{\partial x_i} \right)^2$$

La variation maximale de l'angle est donc directement donnée par les modes.

Chaque atome possède trois modes k_x, k_y, k_z , qui, une fois appliqués, permettent d'obtenir des déformations de l'angle observé égales à $d\phi_x, d\phi_y$ et $d\phi_z$ respectivement. Dans notre cas, chacune de ces valeurs correspond à une composante du vecteur propre cherché, c'est-à-dire de la force optimale à appliquer. D'autre part, la variation angulaire correspondant à l'application du vecteur propre est directement donnée par $d\phi_x^2 + d\phi_y^2 + d\phi_z^2$. 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.

Ce calcul a été appliqué dans le cas particulier des changements de phase du sucre (voir paragraphe 4.3).

4.1.3 Pincements des couples d'atomes

Le procédé présenté ici consiste à pincer un couple d'atomes en leur appliquant un torseur nul, afin d'observer la réponse induite, sur une variation de distance par exemple. 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 que dans le problème du docking, toutes les forces d'interactions intramoléculaires sont par couples et opposées deux à deux. Par conséquent, le choix des atomes fixes n'a aucun effet sur le procédé de docking.

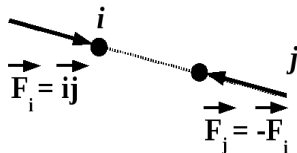


FIG. 4.1 – Pincement d'un couple d'atomes par application de deux forces $\vec{F}_i$ et $\vec{F}_j$.

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 . Comme le montre la figure 4.1, ces forces sont telles que $\vec{F}_i = \vec{i}j$ et $\vec{F}_j = -\vec{F}_i$, pour satisfaire la condition du torseur nul.

On extrait tout d'abord de la matrice des déformations une sous-matrice m de taille $(3N \times 6)$, contenant les 3 modes statiques correspondant à chacun des atomes i et j . Les forces sont par ailleurs calculées de la manière suivante :

$$\vec{F}_i = \begin{pmatrix} \cos \alpha_x \\ \cos \alpha_y \\ \cos \alpha_z \end{pmatrix} \text{ avec } \cos \alpha_l = \frac{\vec{F}_i \cdot \vec{e}_l}{|\vec{F}_i|} = \frac{\vec{i}j \cdot \vec{e}_l}{|\vec{i}j|}$$

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

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

Une fois la déformation engendrée par le pincement calculée, on peut immédiatement déduire la variation induite d'une distance ou d'un angle par exemple.

4.2 Application des modes statiques à un polymère thermosensible : le PNIPAM

4.2.1 Présentation du polymère

Le poly(N-isopropylacrylamide), ou PNIPAM, est un polymère thermosensible présentant une transition de phase hydrophile/hydrophobe réversible à une température caractéristique de 32°C (nommée *low critical solution temperature* ou LCST) [327][361][363].

Lorsque $T < LCST$, c'est-à-dire lorsque le PNIPAM se trouve dans sa phase hydrophile, il est observé sous une forme de "pelote", gonflé par l'eau qui s'insère entre monomères adjacents en formant des liaisons hydrogène. En revanche, si $T > LCST$, le PNIPAM devient hydrophobe et précipite : il expulse l'eau et s'effondre sous une forme de "globule" [325][337].

Ces propriétés font du PNIPAM un polymère "intelligent", un objet d'étude particulièrement intéressant, à l'heure où les efforts se concentrent sur l'essor des nanobiotechnologies. Il présente donc un fort potentiel d'applications dans les milieux aqueux, par exemple :

- en microfluidique pour la fabrication de vannes et d'actionneurs [347]
- pour des applications biopuces : pour adsorber et désorber des molécules biologiques [329], pour délivrer des médicaments [366]...
- pour la conception de dispositifs optiques [350]
- ...

4.2.2 Premières observations sur la flexibilité du polymère

Nous nous attachons dans un premier temps à l'étude d'un dimère de NIPAM. Cette structure a tout d'abord été minimisée avec le logiciel de chimie quantique GAUSSIAN 03 [324] (à l'aide de la fonctionnelle B3LYP [312] et de la base 6-31++G** [323]).

La molécule résultant de la minimisation est représentée sur la figure 4.2.

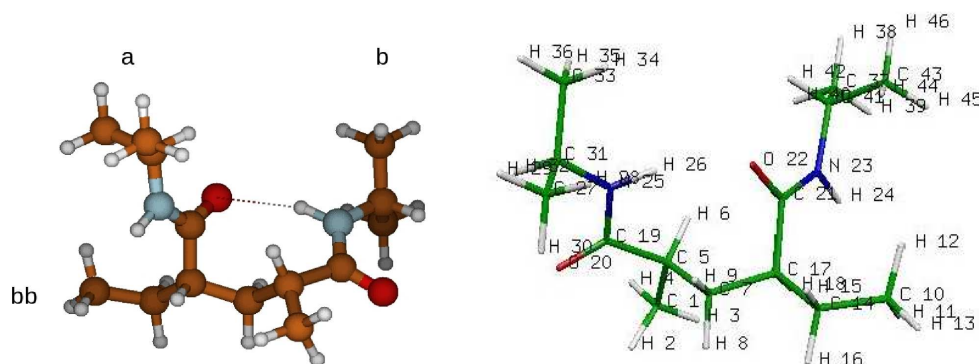


FIG. 4.2 – Représentation atomique d'un dimère de NIPAM obtenu après minimisation avec GAUSSIAN03 ($-2,128.10^{+6}$ kJ/mol). Le backbone est noté **bb** et les deux monomères **a** et **b**. La liaison hydrogène inter-monomères considérée dans la suite de ce travail est également située. A droite, numérotation des atomes du dimère de NIPAM étudié.

Les modes statiques sont ensuite calculés, cette opération prend moins d'une seconde CPU. Ce premier exemple d'application étant de petite taille (46 atomes), nous avons pu explorer la flexibilité de la molécule en détails, au niveau atomique.

Localisation

Nous avons appliqué le calcul de la localisation (voir paragraphe 4.1.1) au dimère de NIPAM (le calcul utilise moins d'une seconde CPU) et présentons les résultats dans la figure 4.3. Ce graphe nous permet de conclure qu'assez peu de modes sont délocalisés. Il s'agit des atomes de la liaison C=O des monomères, mais aussi des atomes de carbone situés au pied des monomères, avec des valeurs de l supérieures à 20. Le reste des atomes, particulièrement les atomes d'hydrogène, donne des valeurs plus faibles de la localisation, autour de 5.

La figure 4.3 présente également l'exemple de deux modes : un délocalisé (correspondant à une contrainte sur l'atome 22 en direction y, soit l'atome d'oxygène d'un monomère) et un localisé (contrainte sur l'atome 12 en direction x, soit un hydrogène du backbone). Grâce à cette représentation, les atomes impliqués dans la déformation sont facilement repérables.

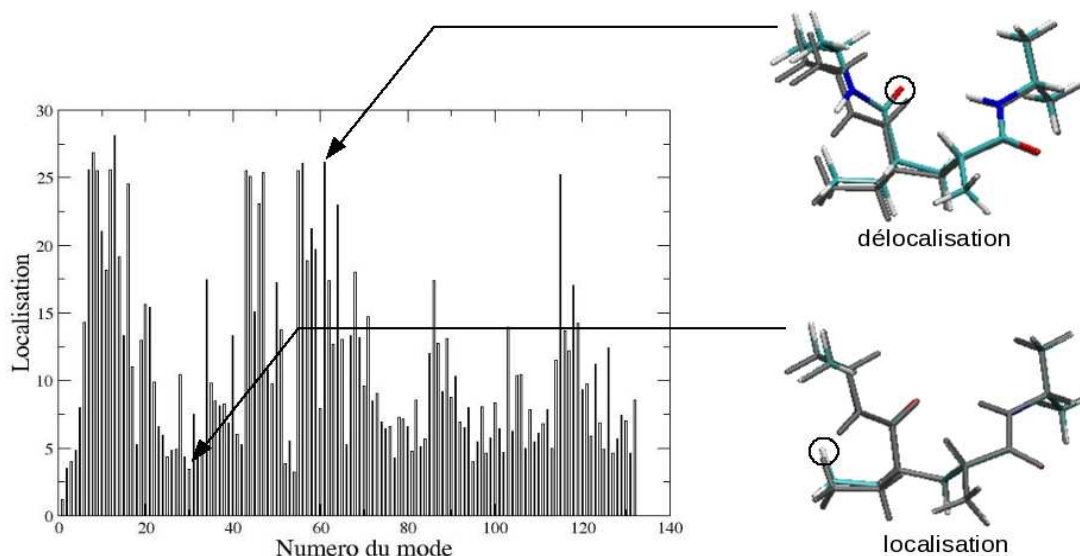


FIG. 4.3 – Représentation graphique du calcul de localisation des modes. A droite, nous représentons la déformation de la molécule relative à deux modes du graphe, un délocalisé et un localisé. Ces déformations correspondent respectivement à des contraintes appliquées sur l'atome 22 en direction y et l'atome 12 en direction x, situés tous deux sur la molécule à droite.

4.2.3 Etude du site actif d'un dimère de NIPAM : la liaison hydrogène inter-monomères

Dans cette étude, notre point de départ est l'exploration de la flexibilité du polymère pour apporter des éléments de compréhension des modifications structurales subies lors de l'arrivée de l'eau. En considérant le site de fixation de l'eau comme un site actif, nous nous sommes donc particulièrement intéressés à la liaison hydrogène formée au niveau de ce site ainsi qu'à sa stabilité [266].

Méthode et procédure : comment casser la liaison hydrogène ?

Nous avons tout d'abord considéré un dimère. La méthode utilisée ici est celle de l'optimisation des forces pour maximiser une variation de distance (cf. paragraphe 4.1.2). Les directions des contraintes sont optimisées pour obtenir la plus grande ouverture possible de la liaison hydrogène, ce qui nécessite moins d'une seconde CPU. Les déformations sont normées à la force.

Nous avons calculé les variations de la distance OH inter-monomères en fonction de l'éloignement entre les atomes contraints et le site actif (ramené au milieu du segment OH).

Ces variations sont représentées sur les graphes 4.4, respectivement pour des forces appliquées sur les atomes des monomères **a**, **b** et du backbone **bb**.

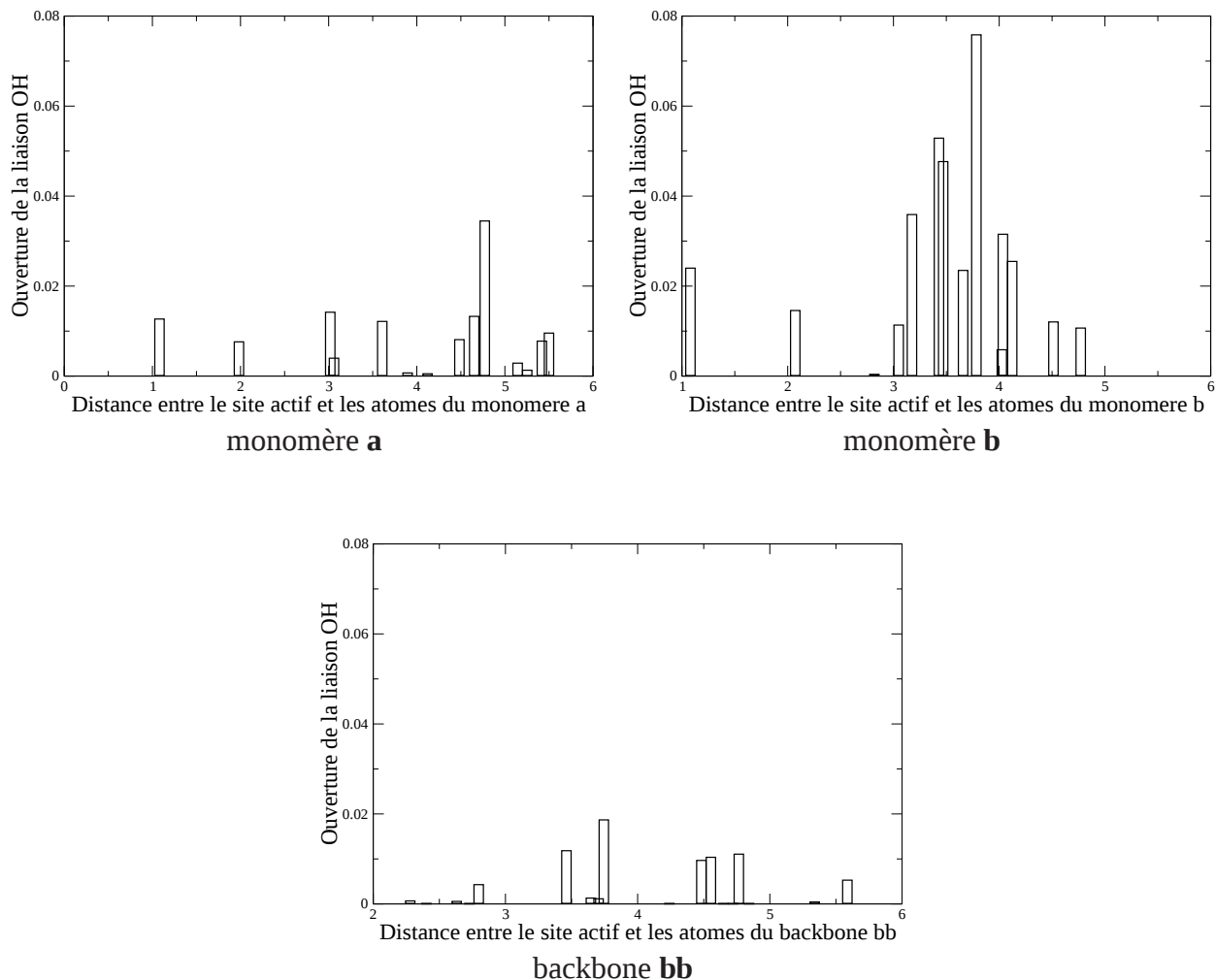


FIG. 4.4 – Variations de la distance interatomique OH en fonction de l'éloignement entre le site actif (ramené au milieu du segment OH) et les atomes contraints des monomères **a**, **b** et du backbone **bb** respectivement (en $\text{\AA}^2/\text{eV}$).

On constate tout d'abord, d'après les deux premiers graphes de la figure 4.4, que les deux monomères induisent des réponses très différentes du site actif : le monomère **a** est beaucoup plus rigide que le **b**, avec des amplitudes maximales respectives d'environ 0.04 et 0.08 $\text{\AA}^2/\text{eV}$. Cette asymétrie de la réponse peut s'expliquer par l'asymétrie chimique du site actif ($\text{C=O} \dots \text{H-N}$). Par ailleurs, selon les résultats présentés sur le dernier graphe de la figure 4.4, les atomes du backbone n'ont pas de rôle sur l'activité du site actif.

D'autre part, les amplitudes d'ouverture maximales sont obtenues pour des contraintes appliquées sur des atomes éloignés du site actif (4.8 et 3.8 $\text{\AA}$), ce qui caractérise l'allostérie des sites sollicités. On remarque également que des contraintes appliquées sur des atomes situés à égale distance du site actif n'induisent pas nécessairement la même réponse.

Une analyse atomique fine est donc essentielle pour comprendre ou prévoir une modification structurale donnée. Les contraintes à appliquer pour obtenir une réponse optimale sont très localisées. La figure 4.5 représente également les variations de la distance OH, mais par rapport aux numéros des atomes et permet donc de localiser les atomes à solliciter. Les principaux pics correspondent à des contraintes appliquées sur les atomes C27, H28, C31, C33 et C37, c'est-à-dire des atomes appartenant aux terminaisons isopropyles.

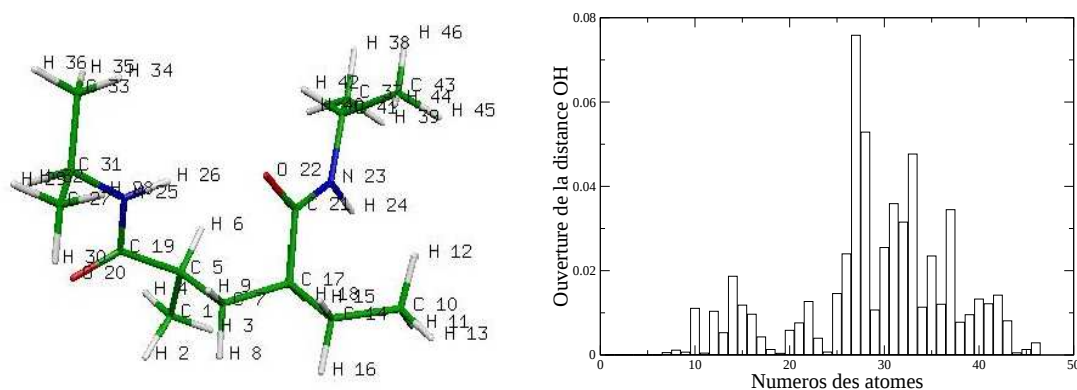


FIG. 4.5 – A gauche, numérotation des atomes du dimère de NIPAM étudié. A droite, les variations de la distance interatomique OH pour chaque atome sollicité (en $\text{\AA}^2/\text{eV}$).

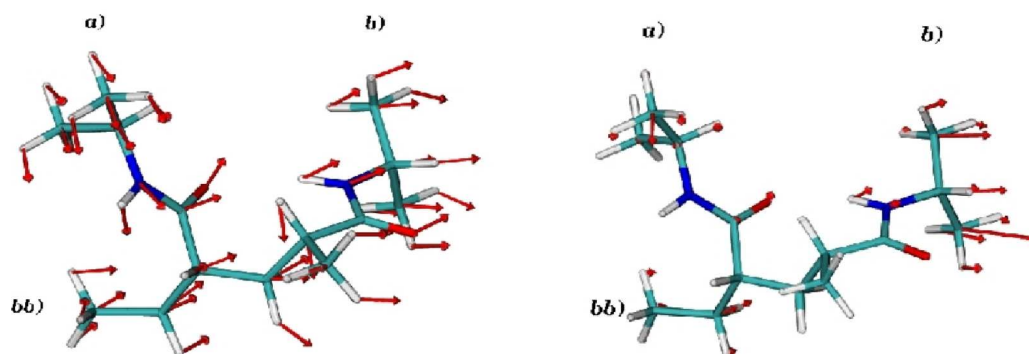


FIG. 4.6 – A gauche, représentation des forces de direction optimisée et normée pour induire une ouverture maximale de la distance OH. A droite, représentation des mêmes forces avec la norme proportionnelle à l'ouverture engendrée.

Par ailleurs, la figure 4.6 permet l'observation directe des forces normées après optimisation des directions. Ces mêmes forces, mais avec une norme proportionnelle à l'ouverture de la liaison OH, sont également représentées.

Cette dernière représentation permet de localiser rapidement les zones de la molécule à solliciter pour obtenir une ouverture optimale des monomères : plus longue est la flèche, plus il est facile d'ouvrir le site actif.

Il apparaît que les modes les plus favorables correspondent à une force appliquée sur les atomes de carbone de la terminaison isopryle du monomère **b**. Ce groupe hydrophobe est donc essentiel lors de l'hydrolyse du polymère. On peut en conclure qu'une contrainte appliquée sur les terminaisons peut entraver, ou au contraire activer, la transition de phase naturelle du PNIPAM.

Conclusion

Ces premiers résultats peuvent s'avérer très utiles pour les expérimentateurs, en révélant une zone particulière du monomère 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.

4.2.4 Extension au cas d'un tétramère de NIPAM

Afin d'étudier la cassure d'une telle liaison hydrogène inter-monomères dans un brin plus long, c'est-à-dire dans un système moins isolé, nous avons considéré un tétramère de NIPAM. Après une minimisation semblable avec GAUSSIAN03, nous avons calculé les modes statiques en moins d'une seconde CPU. La molécule étudiée est présentée sur la figure 4.7. Nous y localisons le backbone **bb**, les quatre monomères **a**, **b**, **c**, **d**, ainsi que les trois liaisons hydrogène inter-monomères.

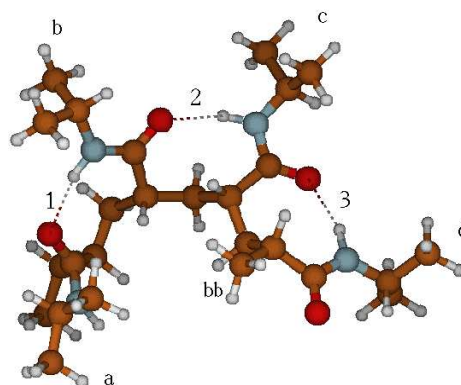


FIG. 4.7 – Représentation atomique d'un tétramère de NIPAM après minimisation avec GAUSSIAN03 ($-4,045 \cdot 10^6$ kJ/mol). Le backbone est noté **bb** et les quatre monomères **a**, **b**, **c** et **d**. Les trois liaisons hydrogène inter-monomères sont situées et notées 1, 2 et 3.

Nous nous intéressons dans un premier temps à la liaison 2, entre les monomères **b** et **c**. La figure 4.8 présente les résultats de la même procédure d'optimisation des forces déjà appliquée au dimère. Ici encore, un calcul d'optimisation de l'ensemble des forces nécessite moins d'une seconde CPU. Les variations d'ouverture de la liaison 2 sont représentées pour chaque atome contraint ainsi que les forces appliquées sur la molécule. La norme de ces forces est représentée comme proportionnelle à l'ouverture induite.

Contrairement au cas du dimère, l'atome d'oxygène de la liaison semble ici jouer un rôle prépondérant. En revanche, si nous ne considérons que les atomes hors du site actif, nous retrouvons bien la dissymétrie des monomères : le monomère **c**, donneur de l'atome d'hydrogène dans la liaison, est essentiel pour casser la liaison 2, contrairement au monomère **b**. D'autre part, la zone à solliciter pour une ouverture optimale se réduit, comme dans le cas du dimère, à la terminaison isopropyle.

Le même travail a été fait pour les liaisons 1 et 3. Les résultats sont présentés respectivement sur les figures 4.9 et 4.10. Il est remarquable et extrêmement intéressant de trouver exactement le même comportement du polymère pour ces liaisons : ce sont toujours les terminaisons isopropyles des monomères portant l'hydrogène de la liaison considérée qui permettent de casser cette dernière le plus facilement.

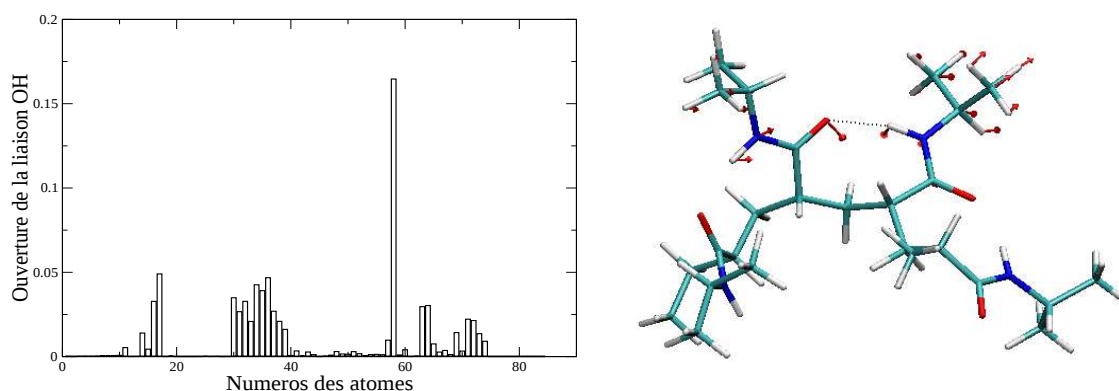


FIG. 4.8 – A gauche sont représentées les variations de l'ouverture de la liaison OH établie entre les groupes CO et NH des monomères **b** et **c**. A droite, sur le tétramère de NIPAM étudié, les forces de direction optimisée et de norme proportionnelle à l'ouverture de la liaison OH sont reportées.

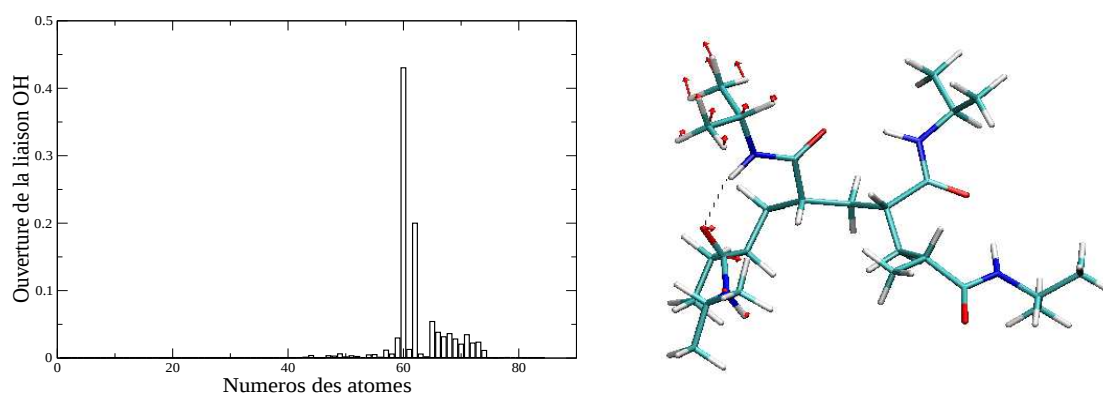


FIG. 4.9 – A gauche sont représentées les variations de l'ouverture de la liaison OH établie entre les groupes CO et NH des monomères **a** et **b** du tétramère. A droite, les forces de direction optimisée et de norme proportionnelle à l'ouverture de la liaison OH sont reportées.

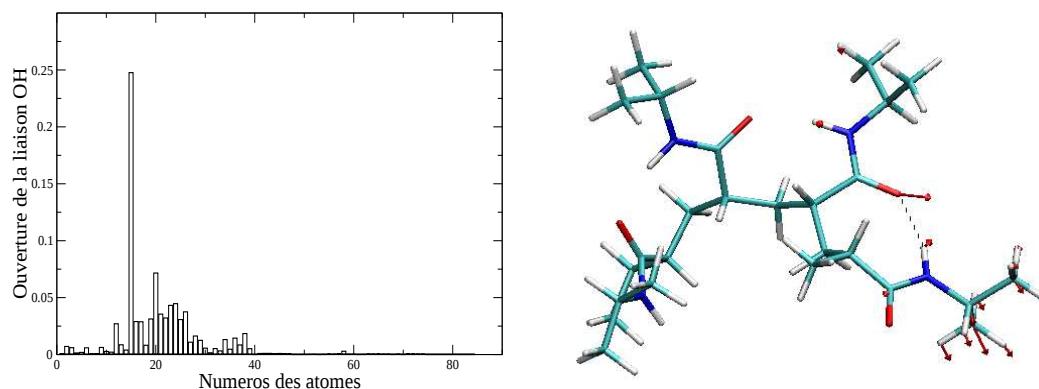


FIG. 4.10 – A gauche sont représentées les variations de l’ouverture de la liaison OH établie entre les groupes CO et NH des monomères **c** et **d** du tétramère. A droite, les forces de direction optimisée et de norme proportionnelle à l’ouverture de la liaison OH sont reportées.

4.2.5 Conclusion

Ce premier exemple, portant sur deux systèmes de petites tailles (respectivement 46 et 84 atomes pour le dimère et le tétramère), nous a permis d’analyser la flexibilité au niveau atomique. L’étude du mécanisme d’ouverture des liaisons hydrogène inter-monomères a apporté des éléments de réponse sur de possibles modifications contrôlées de la transition de phase du PNIPAM.

Dans une optique de prédiction ou d’assistance à l’expérience, ce premier exemple illustre bien le potentiel des Modes Statiques à explorer la flexibilité des molécules, à trouver comment la modifier à des fins définies, à localiser de possibles sites allostériques et à comprendre des relations complexes mais directes entre des sites.

4.3 Etude de la flexibilité des sucres dans les acides nucléiques

Comme nous l’avons vu dans la section 1.2.4, et plus en détails dans l’annexe B, le sucre joue un rôle de pivot entre les bases azotées et les groupes phosphate. Il est donc essentiel lors des changements conformationnels des acides nucléiques. Le furanose est connu pour être extrêmement flexible [346]. En raison des modifications locales ou à grande échelle qu’il peut entraîner, le furanose est un système très étudié, aussi bien expérimentalement [311] que par le calcul [309][349]. Nous avons appliqué la méthode des Modes Statiques au traitement de la flexibilité du sucre dans différents types d’acides nucléiques.

4.3.1 Les plissements du sucre : concept de pseudorotation

Le sucre des acides nucléiques contient un cycle composé de quatre atomes de carbone et d'un atome d'oxygène (voir figure 4.11). Ce cycle, dit furanose, n'est jamais parfaitement plan et peut adopter de multiples conformations, dites "plissements" ou "puckerings". Ces conformations peuvent être schématisées sur un cycle, le cycle de pseudorotation (figure 4.12).

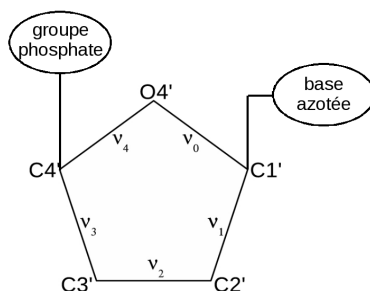


FIG. 4.11 – Représentation du cycle d'un sucre entrant dans la composition des acides nucléiques. Il s'agit d'un furanose, contenant 4 atomes de carbone et 1 atome d'oxygène. La numérotation utilisée est celle de la nomenclature standard [332][333][335].

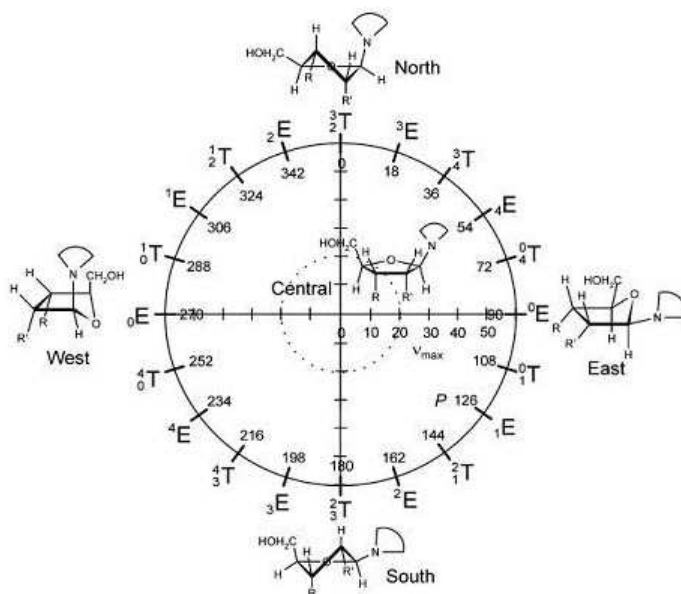


FIG. 4.12 – Cycle de pseudorotation du sucre. Source *J. Chem. Inf. Comput. Sci.* 2004 [365].

Le concept de pseudorotation a été introduit par Sundaralingam *et al.* [308][358]. Il établit une relation entre les quatre angles de torsion endocyclique : v_0 : C4'-O4'-C1'-C2' ; v_1 : O4'-C1'-C2'-C3' ; v_2 : C1'-C2'-C3'-C4' ; v_3 : C2'-C3'-C4'-O4' ; v_4 : C3'-C4'-O4'-C1' (voir figure 4.11).

Chaque mode de plissement est alors défini par deux angles : l'angle de phase P et l'amplitude maximale v_{max} (traduisant la sortie du plan des atomes), décrits par les équations (4.5) et (4.6). P et v_{max} peuvent alors être représentés sur un graphe polaire : le cycle de pseudorotation.

$$P = \arctan \left(\frac{(v_4 - v_0)(v_3 - v_1)}{2 \cdot v_2 \cdot (\sin 36^\circ + \sin 72^\circ)} \right) \quad (4.5)$$

$$v_{max} = \left| \frac{v_2}{\cos P} \right| \quad (4.6)$$

Dans un simple brin d'acide nucléique, le furanose peut théoriquement être trouvé dans différentes conformations. Pourtant, comme le montrent les structures expérimentales, on trouve deux conformations préférentielles, énergétiquement favorables et équivalentes : C3'-endo (appelée *nord* par les spectroscopistes car $-1^\circ \leq P \leq 34^\circ$) et C2'-endo (appelée *sud*, car $137^\circ \leq P \leq 194^\circ$). Ces conformations sont schématisées sur la figure 4.13. Par contre, dans un duplex, le sucre est fixé dans une conformation : nord dans un duplex de forme A, et sud dans un duplex de forme B [358][360].

Comme nous l'avons précisé dans l'annexe B, les duplex d'ARN sont exclusivement trouvés en forme A et possèdent donc des sucres nord. L'ADN, quand à lui, est généralement en forme B, mais peut passer en forme A dans un hétéro duplex ADN:ARN par exemple (le désoxyribose peut donc effectuer des transitions nord/sud).

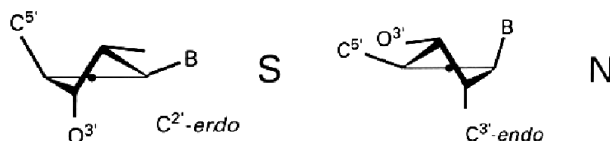


FIG. 4.13 – Représentation des conformations préférentielles du sucre : C3'-endo et C2'-endo, respectivement nommées nord et sud. Source Springer-Verlag 1984 [360].

4.3.2 Cas d'un nucléotide : la désoxythymidine monophosphate

Nous considérons dans un premier temps une désoxythymidine monophosphate (dTMP), étudiée seule et dans un duplex (c'est-à-dire assemblée avec une désoxyadénosine monophosphate (dAMP)).

Les structures que nous présentons dans ce travail ont été minimisées au préalable à l'aide du logiciel AMBER 8 [65], en solvant implicite et avec le champ de force ff94 [316]. A partir de ces structures et de leur matrice Hessienne, nous avons calculé les modes statiques, en moins d'une seconde CPU, et optimisé, pour chaque atome, la direction des forces à appliquer pour maximiser les changements de phase dP du sucre. Cette étape nécessite également moins d'une seconde CPU [266].

Les structures obtenues après minimisation sont présentées sur la figure 4.14. Les valeurs des angles caractéristiques du sucre sont données dans le tableau 4.1. On remarque que le sucre est en conformation sud dans les deux systèmes : $P = 150.86^\circ$ et $v_{max} = 35.29^\circ$ dans le nucléotide seul, $P = 151.52^\circ$ et $v_{max} = 35.21^\circ$ dans le double brin. Ces valeurs correspondent aux valeurs habituellement observées [344][360].

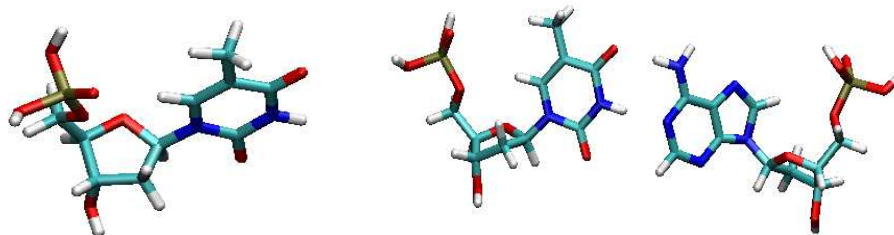


FIG. 4.14 – Représentation des deux systèmes étudiés après minimisation avec AMBER8 : à gauche, la désoxythymidine monophosphate (dTMP) seule ($-3,456.10^{+1}$ kJ/mol), et à droite appariée avec une désoxyadénosine monophosphate (dAMP) ($-1,818.10^{+1}$ kJ/mol).

sucres	v_0	v_1	v_2	v_3	v_4	P	v_{max}	type
dTMP simple brin	-26.59	35.08	-30.82	15.66	6.87	150.86	35.29	C2'-endo
dTMP double brin	-26.24	34.96	-30.95	15.97	6.46	151.52	35.21	C2'-endo

TAB. 4.1 – Angles du sucre d'une désoxythymidine monophosphate seule et en duplex.

La figure 4.15 montre les changements conformationnels du sucre de la dTMP seule et dans un double brin d'ADN, après optimisation des forces. Chaque point correspond donc à la déformation induite par l'application d'une force normée et de direction optimisée, sur un atome.

On observe une différence au niveau des nuages de points autour des points d'équilibre. En effet, le nuage du nucléotide seul est nettement plus étalé, et ce, dans des directions spécifiques. En revanche, le même sucre étudié dans un double brin montre beaucoup moins de flexibilité : son nuage, compact et anisotrope, témoigne de la difficulté de changer de conformation.

La méthode des Modes Statiques peut donc mettre en avant les différences de flexibilité du sucre : flexible et capable de changer de conformation selon des directions spécifiques dans le cas d'un simple brin, et rigidifié dans une conformation sud, dans le cas d'un double brin d'ADN de forme B.

Outre ces conclusions préliminaires, une étude plus poussée du graphe de la dTMP seule peut permettre d'approfondir l'exploration de la flexibilité du sucre par rapport aux changements structuraux globaux de l'ADN.

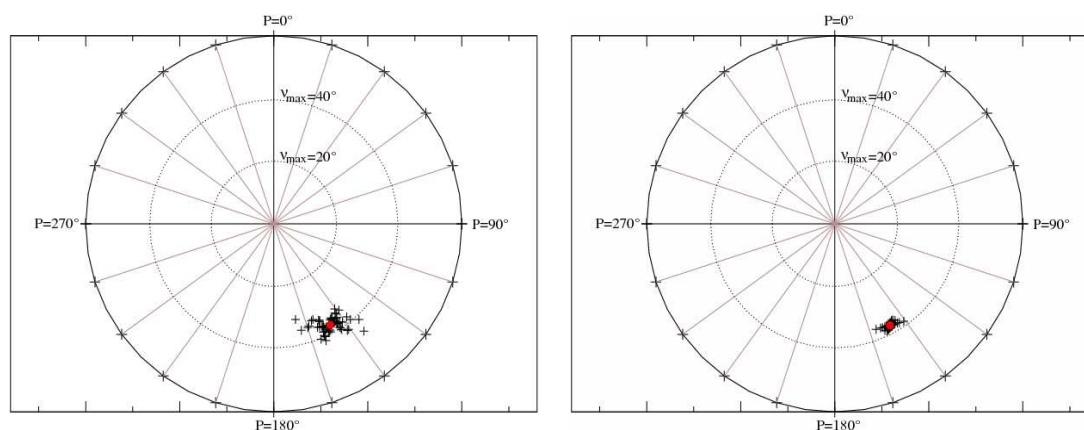


FIG. 4.15 – Représentation des variations de phase du sucre d'une dTMP seule à gauche, et appariée avec une dAMP à droite. Chaque point correspond à une maximisation de la phase après optimisation des forces.

L'observation directe de la figure 4.15 donne tout d'abord des indications sur les tendances du sucre à subir des changements conformationnels. On distingue principalement deux axes, soit deux chemins de transition : un axe est/ouest et un axe nord/sud.

En ce qui concerne le passage est-ouest, la figure 4.17 indique que l'atome contraint qui provoque le plus important changement de phase est l'atome de carbone du groupe méthyle de la thymine (c'est-à-dire l'atome 18, selon la numérotation donnée sur la figure 4.16). Sa sollicitation permet de passer d'une conformation 2E à une conformation 2_1T . Par ailleurs, la plupart des atomes impliqués dans les plus grands changements de phase appartiennent à la base azotée (C18, H19, H20, O23, H25...).

Or, dans un double brin, cette base est appariée, contrainte par des liaisons hydrogène établies avec la base complémentaire (dans le cas d'un brin plus long, elle est également maintenue par des forces de "stacking" par les bases voisines). Les déplacements de ses atomes sont alors restreints, et par conséquent, ne peuvent plus induire les variations de phase observées. Ceci peut expliquer l'absence de grands changements conformationnels dans le cas du double brin.

L'étude du passage nord-sud présente quant à elle un fort intérêt : comprendre le rôle du sucre dans la transition d'une forme B vers une forme A de l'ADN. Si l'on considère le point le plus au nord de cet axe, c'est-à-dire celui qui permet une transition optimale, on trouve qu'il correspond à une excitation de l'atome O3', puis de son hydrogène.

Dans un brin plus long d'ADN, l'atome O3' appartiendrait au squelette sucre-phosphate. On peut donc envisager une relation directe entre le réarrangement global du backbone et les modifications locales du sucre.

Ainsi, pour une transition donnée, on peut déterminer les contraintes à appliquer, et laisser un expert choisir la plus pertinente. La figure 4.18 permet d'observer directement les directions des forces obtenues après optimisation.

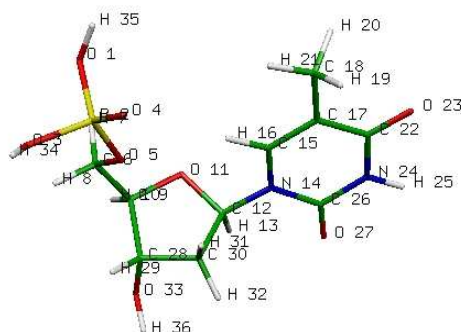


FIG. 4.16 – Numérotation des atomes de la dTMP.

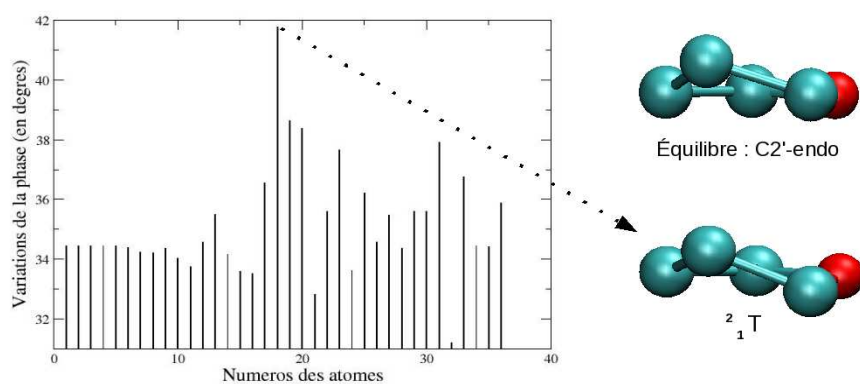


FIG. 4.17 – A gauche, variations de la phase du sucre de la dTMP seule, après optimisation des forces. A droite, deux configurations du sucre, correspondant à l'état d'équilibre (2E) et au plus grand changement de phase (2_1T).

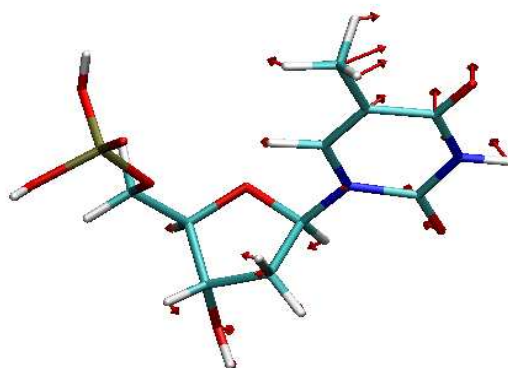


FIG. 4.18 – Représentation des forces optimisées pour maximiser le changement de phase du sucre de la dTMP.

Conclusion

Nous retrouvons dans ces résultats préliminaires certaines caractéristiques essentielles du furanose. D'une part, il présente des différences de flexibilité lorsqu'il est dans un simple ou double brin. D'autre part, sa capacité à passer d'une conformation sud à une conformation nord pour accompagner les modifications structurales à grande échelle de l'ADN. Les modes statiques peuvent indiquer des chemins de transition, et ainsi aider à suivre, pas à pas, le chemin vers une déformation voulue. Cette méthode peut donc proposer à l'expérimentateur des contraintes adéquates à appliquer sur la molécule pour obtenir une réponse physique donnée.

Le paragraphe suivant donne un exemple concret de l'application des modes statiques à des fins prédictives sur les acides nucléiques.

4.3.3 Etude de l'apport de nucléotides modifiés sur la stabilité de duplex d'acides nucléiques

Pour illustrer l'application des modes statiques aux acides nucléiques, nous avons choisi l'étude de dinucléotides contraints. La synthèse de ces systèmes contraints a pour but d'obtenir des duplex plus stables : selon le concept de préorganisation, une molécule rigidifiée dans une conformation acquise dans un complexe, facilite la formation du complexe en raison d'une entropie favorable [317][340]. Les acides nucléiques sont une illustration intéressante de la préorganisation, leur flexibilité intrinsèque leur permettant de s'adapter pour mieux interagir avec diverses molécules.

Les nucléotides contraints, préorganisés pour s'insérer dans les duplex standards d'ADN ou d'ARN, ont été particulièrement étudiés ces dernières années, afin de mesurer leurs effets sur la stabilité des duplex [310][314][345][370].

Rappelons tout d'abord qu'en solution aqueuse, l'ADN double brin est la plupart du temps trouvé en forme B, mais peut également être isolé en forme A (notamment lorsque les conditions d'hydratation changent, ou par exemple, dans un complexe). Les duplex d'ARN adoptent quant à eux toujours la forme A [360].

Une manière de préorganiser les acides nucléiques pour faciliter la formation de duplex, peut donc consister à restreindre le furanose en conformation sud ou nord [345], ou encore d'agir sur la torsion du backbone.

Description du système étudié

Dans cet exemple, nous considérons un dinucléotide thymine-thymine monophosphate (TpT) naturel et son homologue contraint, et comparons leurs effets sur la stabilité de duplex dans lesquels ils sont insérés.

Nous nous référons ici à un travail réalisé par C. Dupouy *et al.* [319][320][321] sur une famille d'acides nucléiques contraints, les D-CNA, dans lesquels les angles dièdres α et β du backbone sont fermés dans une structure dioxaphosphorinane. Ces structures sont décrites sur la figure 4.19.

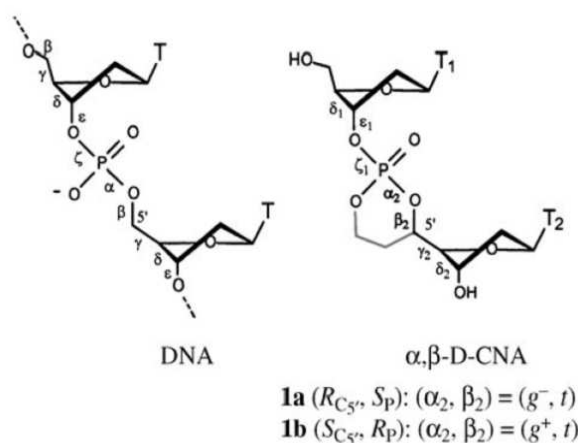


FIG. 4.19 – Structure chimique d'un dinucléotide thymine-thymine monophosphate (TpT) naturel à gauche et de son homologue contraint à droite. *Source Angew. Chem.* [320].

Nous nous intéressons dans cette étude à la forme canonique ($\alpha=g^-$; $\beta=t$), généralement trouvée dans les formes A et B. C. Dupouy *et al.* l'ont synthétisée et insérée dans des duplex pour évaluer ses effets sur la stabilité thermique.

Le travail cité a permis de mettre en évidence, par absorption d'U.V. à 260 nm, que de tels dinucléotides augmentent la stabilité des duplex, et ce de façon plus nette dans un duplex d'ADN (forme B) que dans un hétéroduplex d'ADN:ARN (forme A). La raison identifiée serait la rigidification du sucre du nucléotide contraint (sucre 5') dans une conformation sud.

Stabilité des duplex

Nous avons utilisé quatre structures :

- un homoduplex d'ADN naturel
- un homoduplex d'ADN contenant le dinucléotide contraint ($\alpha=g^-$; $\beta=t$)
- un hétéroduplex ADN:ARN naturel
- un hétéroduplex ADN:ARN contenant le dinucléotide contraint ($\alpha=g^-$; $\beta=t$)

Ces structures ont été minimisées avec AMBER8, en solvant implicite et à l'aide du champ de force ff94. Les structures résultantes sont présentées sur les figures 4.20 et 4.21.

Les modes statiques sont ensuite calculés pour chacune de ces structures, en moins d'une seconde CPU. A partir des modes obtenus, nous évaluons la stabilité de chaque duplex en mesurant l'élongation possible des liaisons hydrogène entre paires de bases.

Pour cela, nous pinçons chaque couple d'atomes (voir paragraphe 4.1.3) et mesurons en réponse la variation de longueur des liaisons hydrogène.

Pour chaque duplex, nous avons quatre liaisons hydrogène, deux par paire de bases : entre les atomes O4 et H6, et H3 et N1 des thymines et des adénines respectivement (selon la nomenclature standard [334]) (voir figure 4.22).

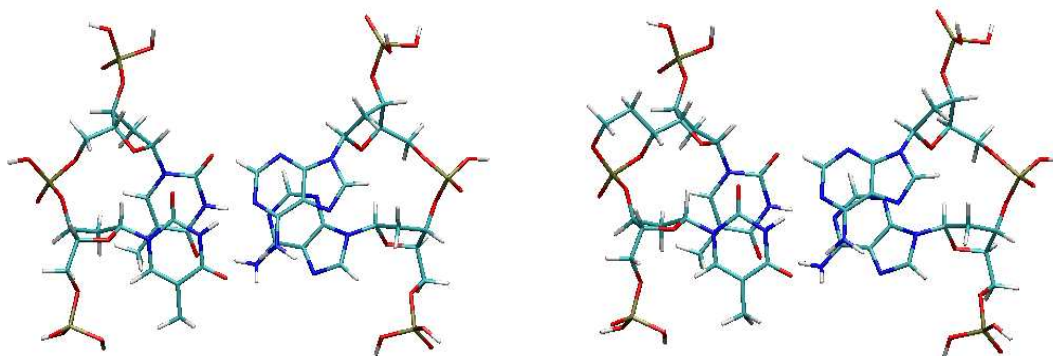


FIG. 4.20 – Structures des deux duplex d'ADN après minimisation : à gauche le duplex naturel ($-1,691.10^{+2}$ kJ/mol) et à droite le duplex contenant un dinucléotide modifié ($\alpha=g-$, $\beta=t$) ($-1,481.10^{+2}$ kJ/mol).

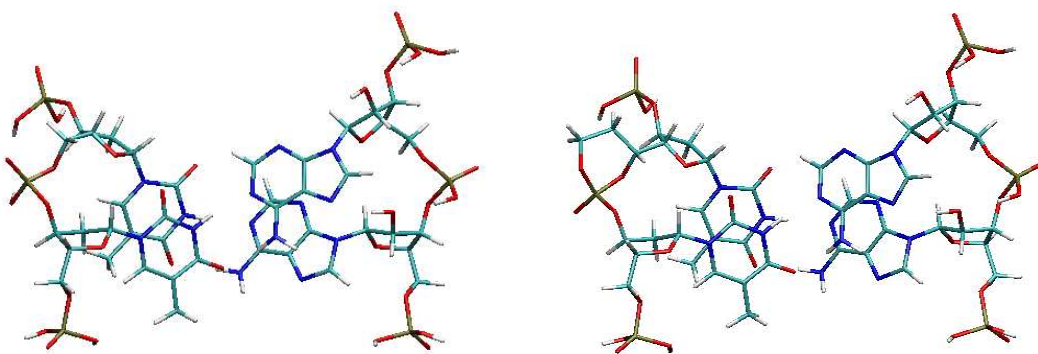


FIG. 4.21 – Structures des deux hétéroduplex ADN:ARN après minimisation : à gauche le duplex naturel ($-1,348.10^{+2}$ kJ/mol), à droite le duplex contenant un dinucléotide modifié ($\alpha=g-$, $\beta=t$) ($-1,009.10^{+2}$ kJ/mol).

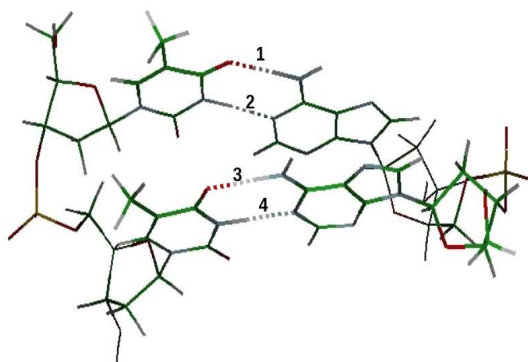


FIG. 4.22 – Numérotation des quatre liaisons hydrogène étudiées.

Dans les tableaux 4.2 et 4.3, nous donnons, pour chaque liaison hydrogène, la moyenne d'ouverture sur l'ensemble des couples pincés (en pourcentage par rapport à la longueur à l'équilibre). Dans le tableau 4.2, nous commençons par comparer l'ouverture des liaisons dans les deux duplex d'ADN (naturel à gauche, et contenant l'analogue contraint à droite). Puis dans le tableau 4.3, nous faisons la même comparaison, mais avec les hétéroduplex ADN:ARN (*idem* naturel à gauche, et contenant l'analogue contraint à droite).

	Δl_H (%) ADN/ADN	Δl_H (%) ADN/ADN(g-)
liaison n° 1	2.81	2.57
liaison n° 2	5.97	5.73
liaison n° 3	2.68	2.62
liaison n° 4	5.19	5.27

TAB. 4.2 – Moyennes de l'ouverture des quatre liaisons hydrogène sur l'ensemble des couples pincés : à gauche pour le duplex d'ADN naturel, à droite pour le duplex d'ADN contenant le dinucléotide contraint (g-).

	Δl_H (%) ARN/ADN	Δl_H (%) ARN/ADN(g-)
liaison n° 1	2.80	2.69
liaison n° 2	6.04	5.57
liaison n° 3	1.62	1.69
liaison n° 4	3.78	3.46

TAB. 4.3 – Moyennes de l'ouverture des quatre liaisons hydrogène sur l'ensemble des couples pincés : à gauche pour l'hétéroduplex ADN/ARN naturel, à droite pour l'hétéroduplex ADN/ARN contenant le dinucléotide contraint (g-).

Dans les deux cas, trois des quatre ouvertures sont plus faibles dans le duplex contenant le dinucléotide modifié que dans le duplex "naturel". Cela signifie que trois liaisons sur quatre sont plus difficiles à ouvrir quand un CNA est inséré dans le brin d'ADN.

Nous pouvons donc prédire, grâce aux modes statiques, que la présence de CNA augmente la stabilité des duplex d'acides nucléiques. A première vue, il paraît cependant difficile d'assurer que le gain de stabilité est plus important dans le cas d'un duplex d'ADN que dans celui d'un hétéroduplex ADN:ARN. L'information serait peut-être davantage visible sur une séquence plus longue.

Néanmoins, on peut s'intéresser à la liaison la plus difficile à ouvrir dans les duplex : on peut ainsi considérer que lorsque celle-ci s'ouvre, les autres sont aussi ouvertes et la fusion est opérée. Dans le cas d'un duplex d'ADN naturel, la liaison la plus difficile à ouvrir est la liaison 3 (2.68 %) ; par ailleurs, elle l'est encore plus avec le CNA (2.62 %). En revanche, dans un duplex ADN:ARN sans modification, la liaison la plus stable, qui est aussi la liaison 3 (1.62 %), perd de sa stabilité en présence du CNA (1.69 %).

Cet élément nous permet donc de conclure que l'effet stabilisant des CNA insérés est moins important dans des duplex de forme A que dans des duplex de forme B.

Dans leur travail, C. Dupouy *et al.* émettent une hypothèse pour expliquer les différences de stabilisation apportées par les CNA aux duplex de formes A et B. Le fait que la stabilisation soit moindre dans les duplex ADN:ARN pourrait s'expliquer par une rigidification du sucre 5' du dinucléotide modifié dans une conformation sud, interdisant le passage vers le nord nécessaire à l'ADN pour transiter d'une forme B vers une forme A lors du couplage avec l'ARN [320]. Dans le paragraphe suivant, nous tentons de vérifier cette hypothèse en explorant la flexibilité du sucre 5' par l'approche des Modes Statiques.

Flexibilité du sucre 5'

Dans cette étude, nous considérons un dinucléotide thymine-thymine monophosphate (TpT) naturel et son analogue contraint. Ces deux structures sont minimisées avec le logiciel de chimie quantique GAUSSIAN 03, en utilisant la fonctionnelle B3LYP et la base 6-31++G**. Les structures résultantes sont présentées sur la figure 4.23.

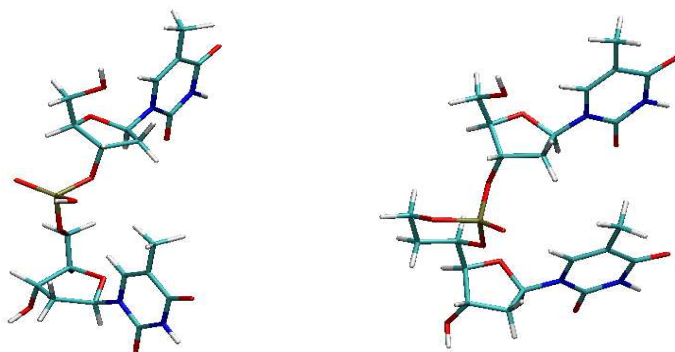


FIG. 4.23 – Structures des deux dinucléotides TpT après minimisation avec GAUSSIAN03 : à gauche, le dinucléotide naturel ($-5,883.10^{+6}$ kJ/mol), à droite, le dinucléotide modifié ($\alpha=g$ -, $\beta=t$) ($-6,086.10^{+6}$ kJ/mol).

Après avoir calculé les modes statiques de ces molécules, nous avons optimisé les forces à appliquer sur leurs atomes pour maximiser le changement de phase favorisant le passage nord/sud (c'est-à-dire la grandeur $v_{max} \cos P$).

La figure 4.24 présente les changements conformationnels du sucre 5' dans le dinucléotide non modifié et modifié. À l'équilibre (point rouge), les deux sucres sont en conformation sud : $P=169.01^\circ$ et $v_{max}=30.91^\circ$ pour le dinucléotide non modifié, $P=167.38^\circ$ et $v_{max}=30.45^\circ$ pour le dinucléotide modifié.

Nous utilisons la même représentation qu'en section 4.3.2 : chaque point des graphes représente un possible changement de conformation du sucre après application, sur un atome donné, d'une force normée et de direction optimisée.

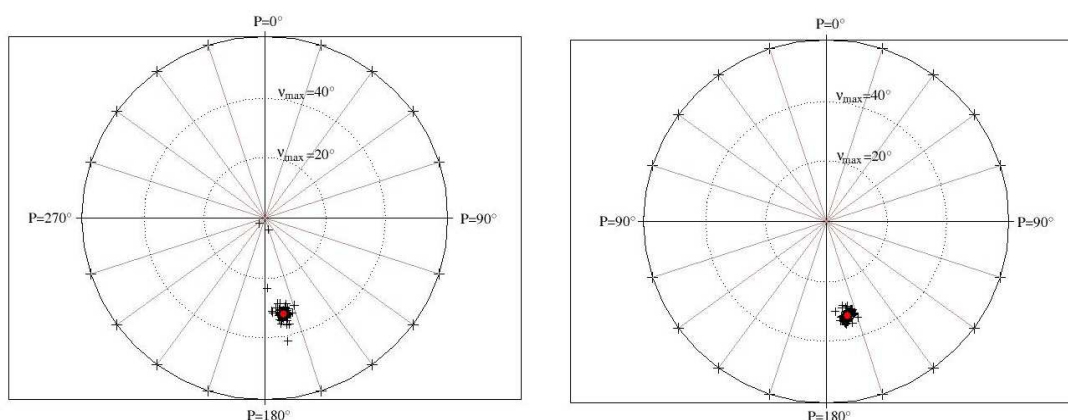


FIG. 4.24 – Représentation des variations de phase du sucre 5' d'un dinucléotide TpT non modifié à gauche, et modifié à droite. Chaque point correspond à une maximisation de la phase après optimisation des forces.

Les graphes de la figure 4.24 présentent tous deux des nuages de points très compacts autour de leur conformation d'équilibre. En revanche, pour le dinucléotide non modifié, on observe nettement quelques conformations éloignées, voire très éloignées, montrant que le sucre 5' peut passer en forme nord.

Les deux points permettant le plus facilement un passage vers le nord correspondent à des contraintes appliquées sur les atomes C5' et O5' (voir figure 4.19). C'est le déplacement de ces atomes, et donc le réarrangement du backbone, qui permet au sucre d'opérer son changement conformationnel.

Par ailleurs, comme le montre la figure 4.19, les atomes C5' et O5' appartiennent au cycle dioxaphosphorinane formé pour contraindre les angles α et β des D-CNA. En figeant ces atomes dans le cycle et en interdisant ainsi leur déplacement, on limite les changements conformationnels du sucre 5'. Il est rigidifié dans sa conformation d'origine, ce que l'on retrouve bien sur le cercle de pseudorotation du dinucléotide modifié.

4.3.4 Conclusion

Cette première application aux acides nucléiques a permis d'aborder d'une façon originale les changements conformationnels du sucre et ses implications sur la structure à grande échelle des duplex. Nous avons ici retrouvé certaines caractéristiques connues de la flexibilité du furanose. En comparant nos résultats à ceux d'un travail expérimental abouti, nous avons surtout fourni une première illustration du caractère prédictif des modes statiques.

4.4 Etude du site actif de la protéase du VIH-1

La protéase du virus de l'immunodéficience humaine de type 1 (PR VIH-1) est une enzyme intervenant dans le cycle viral. Elle participe à la maturation des protéines virales en clivant les produits polypeptidiques des gènes *gag* et *gag-pol* en protéines effectives [322][339][354]. Son inactivation par mutation ou mécanisme inhibiteur conduit à la production de particules virales immatures et non infectieuses.

Elle a donc été considérée comme cible potentielle dès le début des thérapies antivirales dans les années 1990 [330][336][338] et constitue le premier exemple de réussite de conception de médicament par une approche rationnelle [343][371][373].

4.4.1 Description de la molécule : structure et mécanismes

La structure de la protéase est aujourd'hui bien connue, notamment grâce à la cristallographie par rayons X [352][369][372]. Il s'agit d'un homodimère de symétrie C2 [368] composé de deux monomères identiques de 99 résidus. La résolution de structures de la protéase complexée avec un substrat [351] a permis d'identifier le site actif, composé de 6 acides aminés et caractéristique de la famille des protéases aspartiques : ce site actif est composé de deux triades Asp25-Thr26-Gly27 situées à l'interface des monomères [318][348][355].

D'autre part, l'exploration expérimentale et la simulation informatique ont permis de proposer un mécanisme catalytique dès 1987 [364], aujourd'hui très bien décrit [315].

Depuis une vingtaine d'années, l'intérêt pharmaceutique de l'étude de la protéase a permis de recueillir de très nombreuses données structurales, mais aussi de données relatives à la fonction, à l'activité ainsi qu'à la flexibilité de cette protéine.

C'est la raison pour laquelle nous l'avons choisie comme première validation d'une macromolécule. Nous avons confronté les résultats issus des modes statiques à ceux issus d'autres méthodes, expérimentales et calculatoires.

La "géographie" de la protéase du VIH-1 peut être décrite par les 3 domaines présentés sur la figure 4.25 :

- les "volets"
- le domaine de coeur
- le domaine interfacial

Voici quelques propriétés de la protéase que nous avons extraites de la littérature :

- le ligand adopte une orientation spécifique au niveau du site actif, la protéase s'accommodant de manière à optimiser les liaisons pour obtenir un maximum d'affinité [351].
- la protéase possède un site actif enfoui, inaccessible sans l'ouverture des volets. Il existe donc deux conformations, ouverte et fermée comme le montre la figure 4.26 [328], l'amplitude de l'ouverture peut atteindre 7 Å.

- le site actif est clairement identifié, notamment grâce à des simulations de mécanique moléculaire [348]: Asp 25/25' participent directement à la catalyse, Thr 26/26' se lie par des liaisons hydrogène pour stabiliser l'état conformationnel du site actif, Gly 27/27' maintiennent le ligand dans une orientation favorable à la réaction.
- les résidus des "poches de liaison" (81-84) contribuent à maintenir le ligand. Leur modification (par mutations par exemple) induit des changements au niveau du site actif et une perte d'affinité avec le ligand [353][356][362].
- de nombreuses liaisons se créent à l'interface pour maintenir les monomères, y compris au niveau des extrémités C et N-terminales. Dans la forme fermée de la protéase, de telles liaisons (hydrogène en particulier) se forment également entre les volets, au niveau des résidus Ile 50 et Gly 51 [353].
- les formes ouverte et fermée de la protéase présentent des différences de flexibilité : la dynamique moléculaire et la RMN ont montré des fluctuations des résidus de l'interface chez la protéase fermée n'existant pas dans la forme ouverte, probablement pour permettre l'accommodation avec le ligand [331].
- il existe des corrélations entre domaines, particulièrement entre le site actif et les poches de liaison d'une part, ainsi qu'avec l'interface [326][359].
- un site allostérique a été mis en évidence, notamment par dynamique moléculaire sous contrainte : les leviers (résidu 68 et ses voisins) permettent d'agir sur la fermeture de la protéine [357].
- la fermeture de la protéase peut être modélisée par des forces d'interaction entre protéine et ligand lors de l'arrimage du ligand [149].

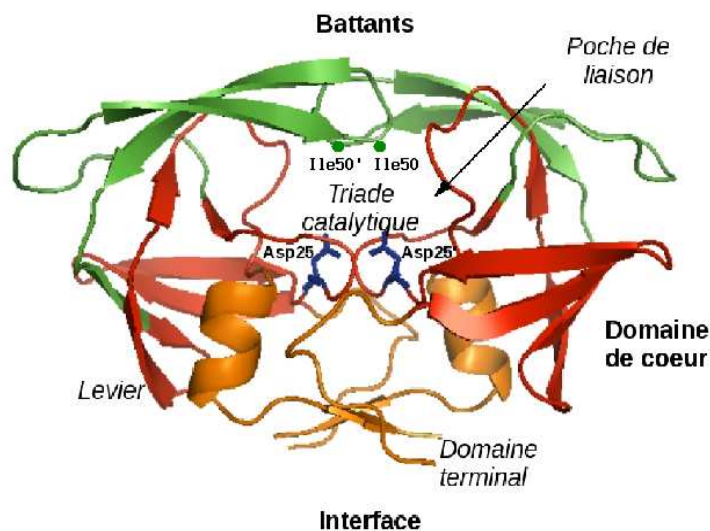


FIG. 4.25 – Représentation de la protéase du VIH et de ses domaines : en vert les volets, en rouge le cœur et en orange l'interface. Sont également indiqués : le site catalytique, les poches de liaison, les leviers et le domaine terminal.

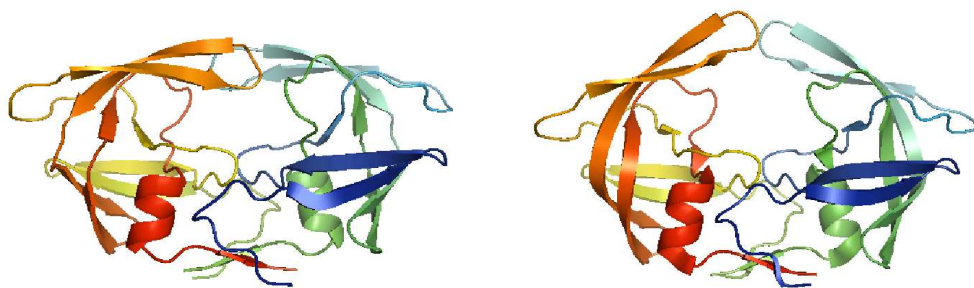


FIG. 4.26 – Deux configurations de la protéase : fermée à gauche et ouverte à droite. Ces structures correspondent respectivement aux fichiers 1HHP et 1TW7 de la *Protein Data Bank*.

Avec l'objectif de valider notre approche, nous nous sommes attachés à retrouver certaines de ces propriétés à partir des modes statiques. Dans nos calculs nous nous sommes particulièrement intéressés au mécanisme d'ouverture et fermeture de la protéase (i) ainsi qu'à la stabilité du site actif (ii) en observant les distances suivantes, respectivement :

- (i) la distance entre les résidus Ile50 (appartenant au volet) et Asp25
- (ii) la distance séparant les résidus Asp25 et Asp25'

La réponse de ces distances aux contraintes extérieures présente en effet un fort intérêt dans la recherche de mécanismes d'inhibition.

4.4.2 Evaluation de la stabilité du site actif

Optimisation des forces pour maximiser la variation des distances

Nous avons tout d'abord étudié la protéine fermée (1HHP dans la Protein Data Bank) : après une minimisation avec le champ de force ff03 de AMBER8, nous avons extrait la matrice hessienne et calculé les modes statiques (cette étape nécessite 15 heures CPU).

Dans ce paragraphe, nous présentons les résultats de l'optimisation des forces permettant d'agir sur les distances Asp25-Asp25'. L'optimisation utilise 1 min 36 sec CPU.

Le graphe 4.27 représente, pour chaque atome sollicité avec une force optimisée et normée à 1eV/Å, la variation de distance C_{α} - C_{α} des résidus Asp25/25' en Å.

Ce diagramme révèle des zones induisant une réponse du site actif bien localisées. Ces zones sont représentées en rouge, orange et jaune sur la protéine, respectivement pour des ouvertures supérieures à 0.5 Å, comprises entre 0.05 et 0.5 Å, et inférieures à 0.05 Å.

Il est alors facilement observable que les zones les plus influentes correspondent aux résidus Asp25 et Asp25' eux-mêmes, ainsi qu'à leurs voisins (résidus 22 à 27 et symétriques). Viennent ensuite les poches de liaisons (résidus 81 à 84), puis l'interface (résidus 4 à 9) et les extrémités C-terminales (résidus 95 à 99).

On peut tout d'abord remarquer que ces zones sont localisées, mais étendues à plusieurs acides aminés. La protéine peut donc être découpée en "domaines", comme nous l'avons décrite : le site actif, les poches de liaison, l'interface et les extrémités terminales. D'autre part, ce premier calcul nous permet de retrouver des propriétés de la protéase d'ores et déjà identifiées par RMN, mutations, dynamique moléculaire ..., à savoir :

- la communication des domaines [326][359]
- la modification de stabilité du site actif par une action sur les résidus des poches de liaison [353][356][362]
- l'interface [326][359] ainsi que les extrémités C-terminales [331][353], impliqués dans l'interaction inter-monomères, contribuent également à la stabilité du site actif

Par ailleurs, l'utilisateur peut pousser l'exploration au niveau atomique, en recherchant par exemple l'atome appartenant aux poches de liaisons (en orange) induisant la plus grande déformation. Il s'agit dans ce cas d'un atome d'hydrogène appartenant au groupe méthyle du résidu Ile84 (et de son homologue dans l'autre monomère).

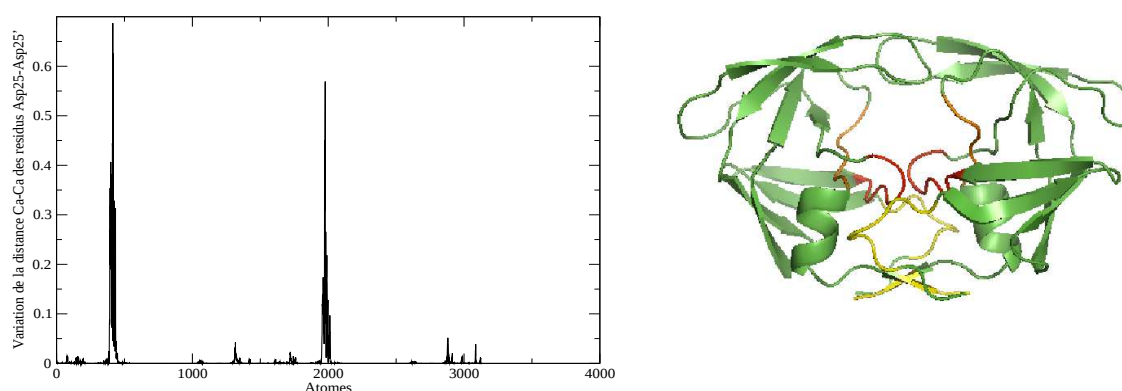


FIG. 4.27 – Variations maximales Δl de la distance séparant les C_{α} des résidus Asp25 et Asp25' après optimisation des forces sur chaque atome de la protéase. A droite, les zones de la protéase induisant une réponse du site actif sont représentées en rouge ($\Delta l \geq 0.5 \text{ \AA}$), en orange ($\Delta l \geq 0.05 \text{ \AA}$) et en jaune ($\Delta l < 0.05 \text{ \AA}$).

Pincement des couples d'atomes et réponse des distances

En travaillant toujours sur la protéase fermée et sur les modes statiques déjà calculés, nous avons simulé les pincements de tous les couples d'atomes de la molécule et à nouveau mesuré les variations de distance séparant les résidus Asp25 et Asp25'. Le graphe 4.28 présente pour tous les couples d'atomes pincés, indiqués en abscisse et en ordonnée, les variations de la distance séparant les C_{α} des aspartates 25 et 25'. L'amplitude de ces variations est définie par la légende, de $-0.8 \text{ \AA}$ à $1.4 \text{ \AA}$ du rouge vers le bleu. Pour faciliter la lisibilité, nous n'avons pas représenté les amplitudes comprises entre -0.1 et $0.1 \text{ \AA}$.

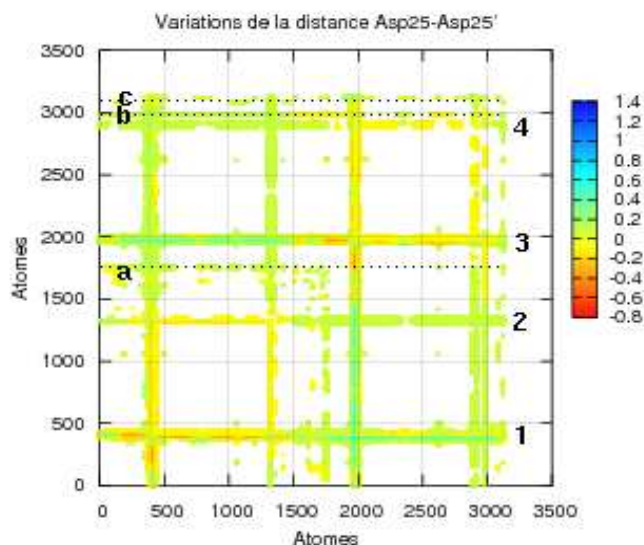


FIG. 4.28 – Représentation graphique des variations de la distance séparant les C_{α} des résidus Asp25 et Asp25' en réponse au pincement de tous les couples d'atomes de la molécule. La légende à droite indique des variations allant de -0.8 Å à 1.4 Å du rouge vers le bleu.

Notons tout d'abord qu'il apparaît quatre bandes pleines numérotées de 1 à 4. Cela signifie que les résidus de ces quatre séquences de la protéine, pincés avec n'importe quel autre résidu, provoquent une variation de la distance $C_{\alpha}(\text{Asp25}) - C_{\alpha}(\text{Asp25}')$.

Ces séquences sont représentées sur la première molécule de la figure 4.29, et correspondent aux aspartates 25 et 25' eux-mêmes ainsi qu'à leur proches voisins (22-27 et 22'-27') et aux poches de liaison (82-85 et 82'-85'). L'optimisation des forces a précédemment permis d'identifier ces séquences influant sur la stabilité des résidus du site actif. Il est donc probable que, quel que soit le pincement, le simple fait de les solliciter entraîne des déplacements des résidus Asp25/25'.

On peut alors s'intéresser à des couples plus spécifiques, caractérisés par quelques lignes "en pointillés", dont les principales sont notées a, b, c. Ces bandes permettent d'identifier trois séquences d'acides aminés qui, pincées avec des résidus particuliers, induisent également une réponse des résidus Asp25/25'.

Le détail des couples est illustré par la figure 4.30. On peut noter que dans ces couples, les trois séquences sollicitées récurrentes appartiennent au domaine interfacial (séquences 109-111, 188-190, 195-198 et leurs symétriques). Le rôle d'une séquence en particulier, correspondant à l'hélice α , n'intervenait pas dans l'optimisation des forces, mais apparaît dans le calcul des pincements.

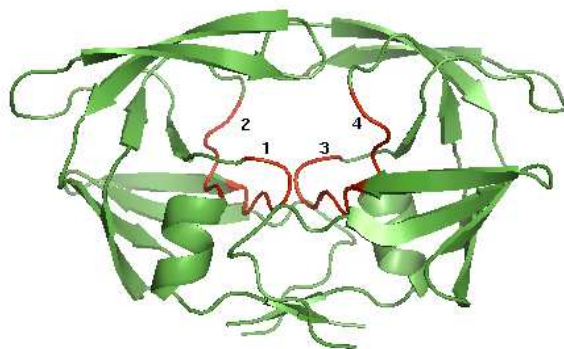


FIG. 4.29 – Représentation des couples induisant une réponse du site actif, évaluée par la variation de la distance Asp25-Asp25'. Sont représentées en rouge les séquences correspondant aux quatre bandes de la figure 4.28 numérotées de 1 à 4.

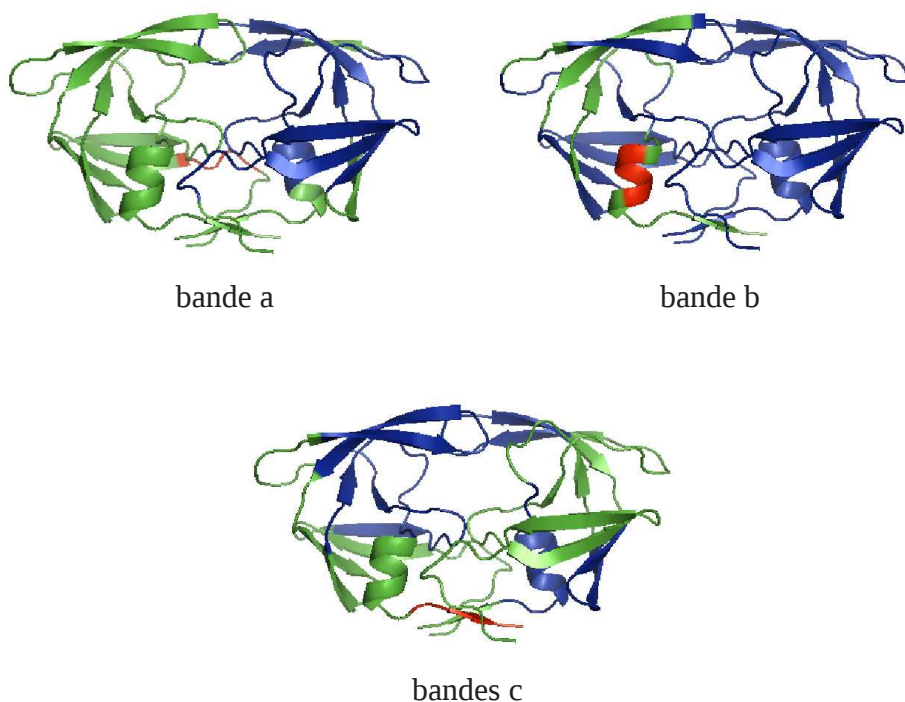


FIG. 4.30 – Représentation des couples induisant une réponse du site actif, évaluée par la variation de la distance Asp25-Asp25'. Les trois molécules, identifiées par la notation a, b, c, correspondent aux bandes respectives du graphe 4.28. La séquence représentée en rouge indique les résidus de la bande à pincer avec les séquences bleues.

Comparaison des réponses de la protéase en configuration fermée et ouverte

La même procédure de minimisation avec AMBER8, puis de calcul des modes statiques avec Flexible, a été appliquée sur la protéase fermée (1TW7 dans la Protein Data Bank). Nous avons ensuite recalculé pour l'ensemble des pincements, les variations de distance Asp25/Asp25' et Asp25/Ile50, ce qui nécessite 14 heures CPU.

La figure 4.31 permet de comparer les variations de distance C_{α} - C_{α} des résidus Asp25/25', en réponse aux pincements de tous les couples d'atomes, pour la protéine en conformation fermée (à gauche) et ouverte (à droite).

Bien que les deux graphes paraissent très semblables, quelques différences apparaissent au niveau des bandes que nous avons notées a, b, c sur la figure 4.28 et qui disparaissent pour la forme ouverte. Ces bandes correspondant à des séquences de l'interface, on peut conclure que le rôle de l'interface, prépondérant en conformation fermée, s'atténue lorsque la protéine s'ouvre.

Il a en effet été montré, notamment par des études de RMN [331], que les séquences de l'interface présentent des fluctuations en conformation fermée, que l'on ne retrouve pas en conformation ouverte, probablement pour conférer au site actif une flexibilité propice à l'accommodation au ligand.

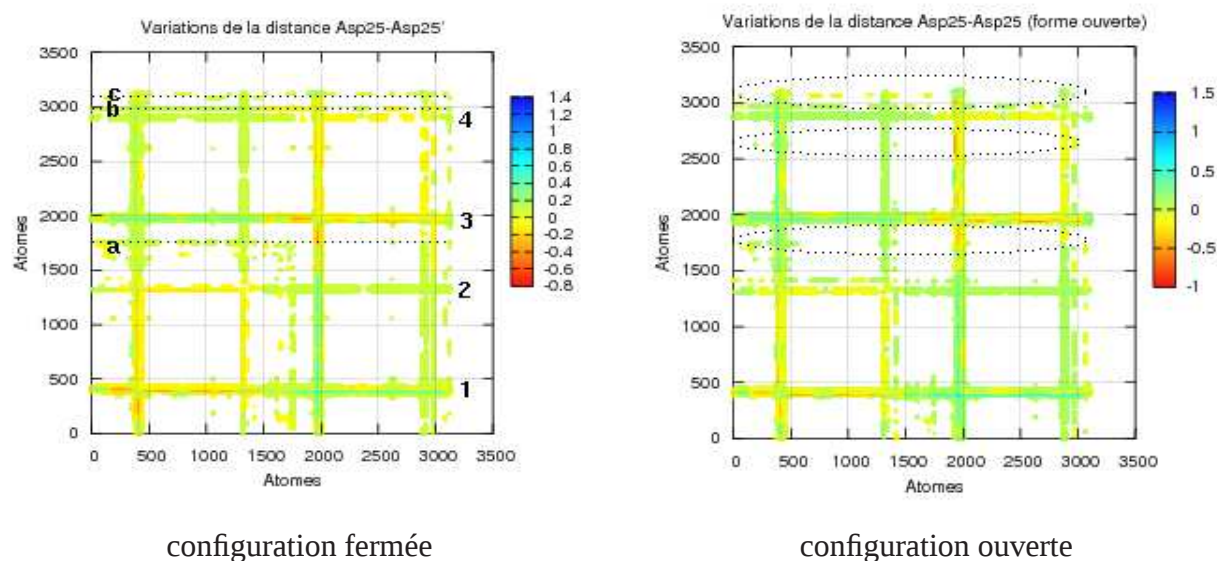


FIG. 4.31 – Comparaison des variations de distance entre les résidus Asp25 et Asp25' de la protéase fermée (à gauche) et ouverte (à droite) après pincement des couples d'atomes.

4.4.3 Exploration du mécanisme d'ouverture

Cette section se décompose exactement comme la précédente. Nous avons utilisé les mêmes systèmes et les mêmes procédures pour calculer les variations de la distance séparant les résidus Asp25 et Ile50, c'est-à-dire l'optimisation des forces et les pincements des couples d'atomes sur protéase fermée puis ouverte.

Optimisation des forces pour maximiser la variations des distances

Le graphe 4.32 permet d'évaluer quels acides aminés de la protéase, après sollicitation, induisent une augmentation de la distance séparant les C_{α} des résidus Ile50 et Asp25. Ce calcul peut ainsi donner des informations sur le mécanisme d'ouverture et de fermeture, mais aussi sur des possibilités de contrôler l'accessibilité au site actif. Comme dans le paragraphe précédent, nous avons cartographié sur la molécule les zones auxquelles correspondent les pics visibles sur le graphe, pour une meilleure lisibilité.

A nouveau, des domaines apparaissent : les plus fortes amplitudes, supérieures à 1 Å, sont obtenues après sollicitation des résidus Asp25 et Ile 50 ainsi que de leurs voisins (respectivement 22 à 27 et 47 à 53), représentés en rouge sur la molécule. En orange, apparaît le second volet (résidus 47' à 53') avec une amplitude atteignant 0.4 Å, puis en jaune, avec une amplitude inférieure à 0.2 Å, les poches de liaison, l'interface et les extrémités C-terminales. Enfin, on peut noter qu'une contrainte sur l'extrémité du levier (Glu65-Ile66) cause également une ouverture du volet de 0.1 Å. Les poches de liaison et l'interface étant fortement impliqués dans les mouvements du résidu Asp25, nous n'en tenons pas compte pour l'analyse de l'ouverture de la distance Asp25-Ile50 : en effet, dans ce cas, nous observerons davantage les déplacements du C_{α} de Asp25 que l'ouverture du volet.

En revanche, on peut déduire les comportements suivants :

- une contrainte sur le volet opposé entraîne le mouvement du volet observé (résidus 47' à 53'), ce qui peut s'expliquer par les interactions établies entre ces deux boucles et participant à la cohésion inter-monomères [353]
- il existe un site allostérique au niveau des résidus Glu65 et Ile66, permettant, par un mouvement de levier, d'agir sur l'ouverture du volet [326][357]

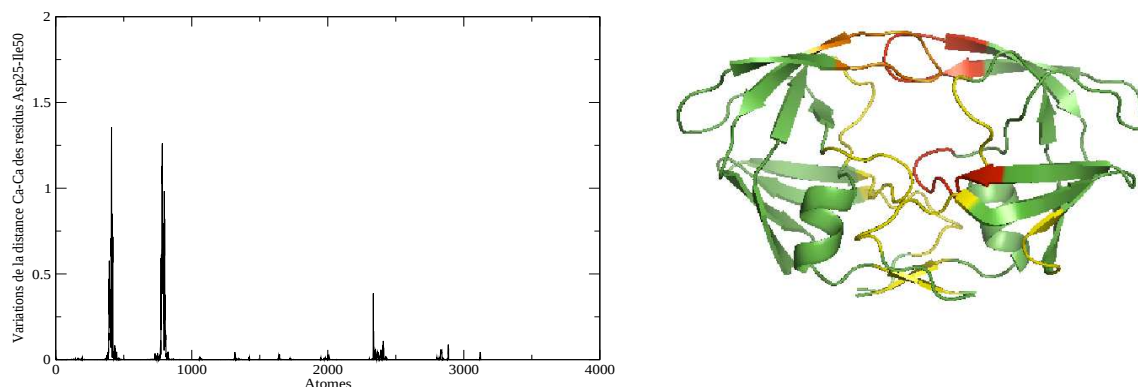


FIG. 4.32 – Variations maximales de la distance séparant les C_{α} des résidus Ile50 et Asp25 après optimisation des forces pour chaque atome. A droite, les zones de la protéase induisant une ouverture du volet sont représentées en rouge ($\Delta l \geq 1$ Å), en orange ($\Delta l \geq 0.2$ Å) et en jaune ($\Delta l < 0.2$ Å).

Pincement des couples d'atomes et réponse des distances

La procédure de pincement de tous les couples d'atomes est à nouveau appliquée en évaluant cette fois les fluctuations de la distance C_{α} - C_{α} des résidus Ile50-Asp25. Les résultats sont reportés sur le graphe 4.33, présentant des variations allant de -1.5 Å à 2.5 Å du rouge vers le bleu. Les amplitudes comprises entre -0.1 et 0.1 Å ne sont pas représentées.

Comme dans le cas précédent, nous obtenons quatre bandes continues, numérotées de 1 à 4, présentant quatre séquences déjà identifiées lors de l'optimisation des forces. La première molécule de la figure 4.34 permet de représenter ces séquences : les résidus voisins du site actif (22 à 27), les deux volets (47-53 et 47'-53'), ainsi que l'extrémité C-terminale du monomère considéré.

En revanche, un grand nombre d'autres couples, référencés de a à h sur les figures 4.33 et 4.34 apparaissent. Chaque molécule de la figure 4.34 correspond à une de ces bandes : nous avons représenté en rouge la séquence à pincer avec les résidus des séquences bleues pour obtenir une ouverture du volet.

Cet outil pourra peut-être servir un jour à créer des "pinces moléculaires" si la technologie le permet. Pour l'instant, on peut conclure ici que l'étude de ces couples n'apporte pas d'information supplémentaire par rapport à l'optimisation des forces. En revanche, la comparaison des résultats obtenus sur protéine fermée et ouverte fournit de nouvelles conclusions : c'est l'objet du paragraphe suivant.

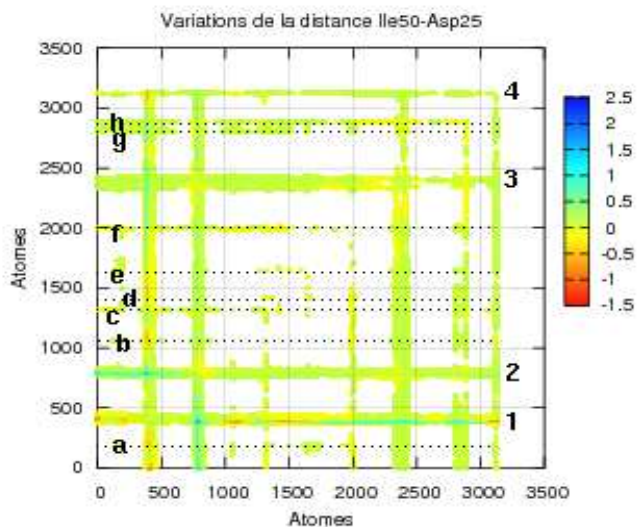


FIG. 4.33 – Représentation des variations de la distance séparant les C_{α} des résidus Ile50 et Asp25 en réponse au pincement de tous les couples d'atomes de la molécule. La légende à droite indique des variations allant de -1.5 Å à 2.5 Å du rouge vers le bleu.

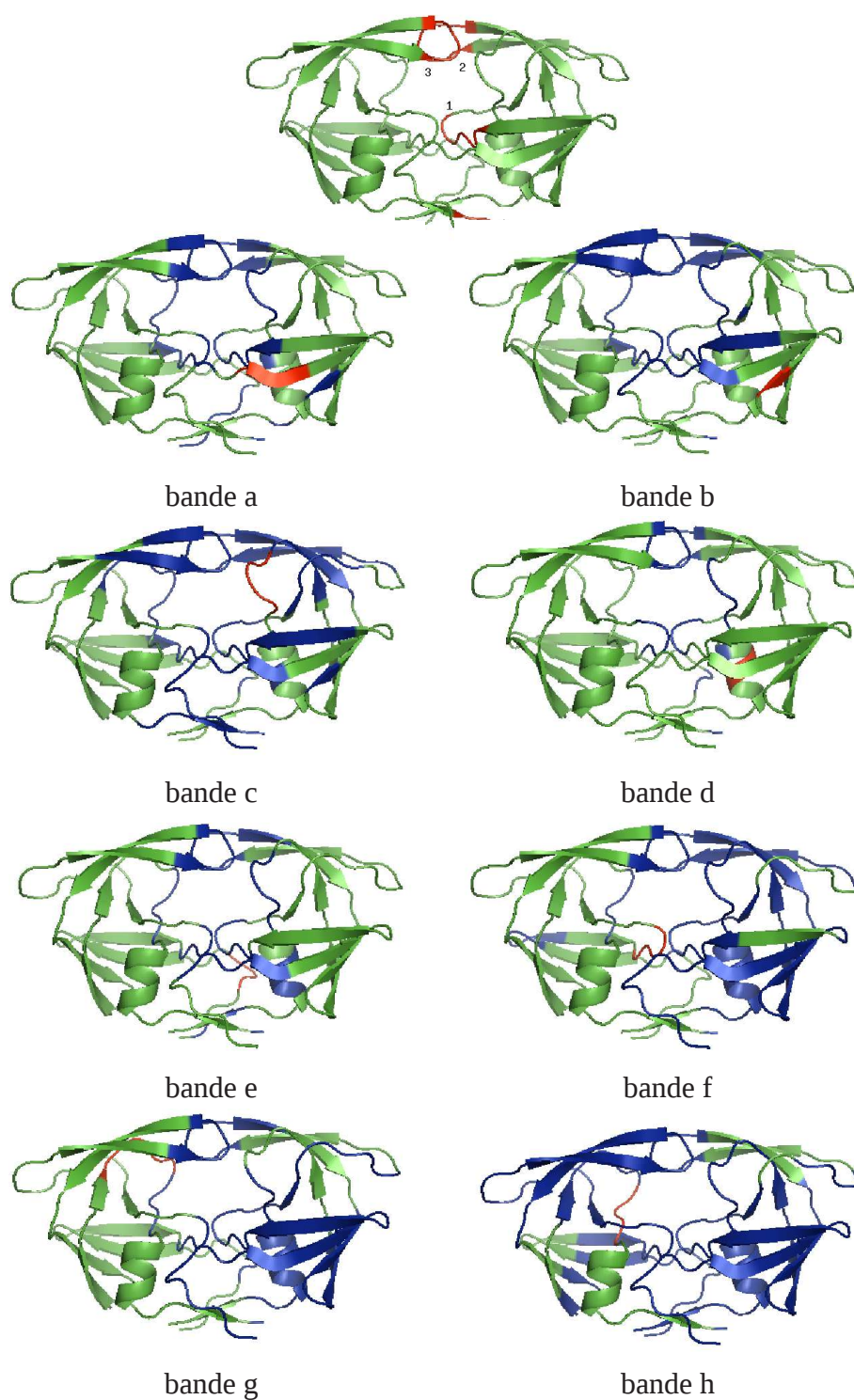


FIG. 4.34 – Représentation des couples induisant une variation de la distance Ile50-Asp25. Sur la première molécule, représentation des couples induisant une variation de la distance Ile50-Asp25. En rouge : les séquences correspondant aux quatre bandes de la figure 4.33 numérotées de 1 à 4. Les molécules, notées de a à h, correspondent aux bandes respectives du graphe 4.33. La séquence rouge indique les résidus de la bande à pincer avec les séquences bleues.

Comparaison des réponses de la protéase en configuration fermée et ouverte

Nous comparons les résultats avec la protéase ouverte, comme nous l'avons fait dans la section précédente. La figure 4.35 nous permet d'évaluer les séquences jouant un rôle dans les mouvements du volet en conformation fermée (à gauche) et ouverte (à droite).

On observe nettement qu'une grande partie des couples présents en forme fermée disparaissent en forme ouverte : principalement les bandes notées f, g, h sur la figure 4.33 appartenant au monomère opposé à celui que nous observons, ainsi que l'ensemble du "quart nord-est" du graphe, c'est-à-dire les pincements intra-monomère de ce même monomère opposé.

Ainsi, dans la forme ouverte, nous perdons l'influence d'un monomère sur l'autre, ce qui peut s'expliquer par une perte d'affinité au niveau de l'interface, notamment au niveau de l'interaction des volets.

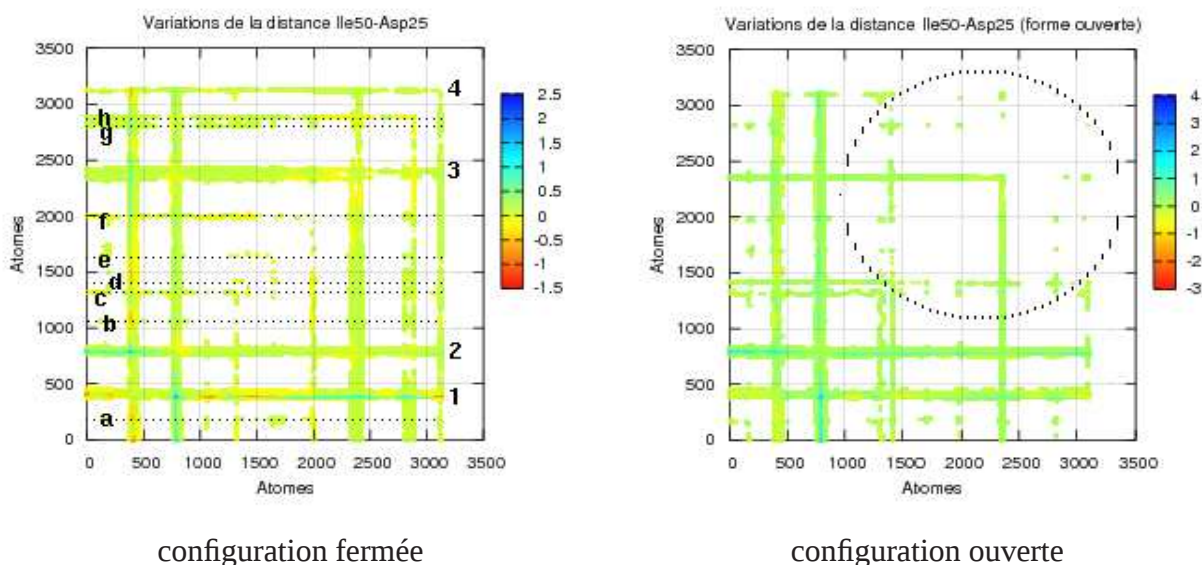


FIG. 4.35 – Comparaison des variations de distance entre les résidus Ile50 et Asp25 de la protéase fermée (à gauche) et ouverte (à droite) après pincement des couples d'atomes.

4.4.4 Simulation des effets de l'arrimage d'un ligand

Pour ce travail, nous nous sommes appuyés sur un exemple utilisé par d'autres modélisateurs. Pour valider leur algorithme de mouvements de grande amplitude des protéines (basé sur une approche mécanique et robotique du backbone), Jagodzinski *et al.* ont simulé la fermeture de la protéase lors de l'arrimage d'un ligand [149].

En partant de la conformation ouverte 1TW7 et des modes déjà calculés, nous avons appliqué 4 forces sur les C_{α} des résidus Ile50, Ile50', Asp25 et Asp25' en direction des atomes d'oxygène d'un possible ligand à urée cyclique, comme le montre la figure 4.36.

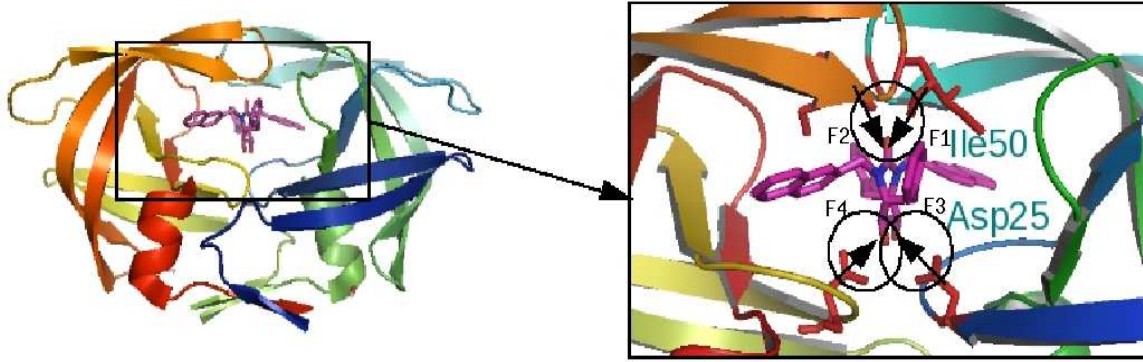


FIG. 4.36 – A gauche, représentation de la protéase complexée avec un inhibiteur potentiel type, contenant une urée cyclique (1HVR dans la Protein Data Bank). A droite est présenté un zoom sur le site actif ainsi que les quatre forces considérées entre le ligand et les résidus Ile50, Ile50', Asp25 et Asp25'.

Bien entendu, de nombreuses autres forces interviennent lors du docking, mais il a été montré que ces 4 forces peuvent suffire à décrire l'interaction dans un modèle simplifié [149][341].

A l'aide des coordonnées de notre protéase en conformation ouverte et du ligand correctement orienté, les forces s'écrivent :

$$\vec{F}_1 = \begin{pmatrix} 1.954 \\ -1.958 \\ 7.041 \end{pmatrix}; \vec{F}_2 = \begin{pmatrix} -5.821 \\ -0.362 \\ 3.164 \end{pmatrix}; \vec{F}_3 = \begin{pmatrix} -10.273 \\ -0.936 \\ 6.624 \end{pmatrix}; \vec{F}_4 = \begin{pmatrix} -6.263 \\ -3.482 \\ -3.106 \end{pmatrix}$$

Ces forces sont appliquées respectivement sur les C_α des résidus Ile50, Ile50', Asp25 et Asp25' par l'opération suivante :

$$M'_i = \sum_{j=0}^3 m_{1ij} \dot{F}_{1j} + \sum_{j=0}^3 m_{2ij} \dot{F}_{2j} + \sum_{j=0}^3 m_{3ij} \dot{F}_{3j} + \sum_{j=0}^3 m_{4ij} \dot{F}_{4j}$$

m_1, m_2, m_3 et m_4 étant les matrices $3N \times 3$ contenant les trois modes statiques correspondant à chacun des C_α .

On obtient ainsi le vecteur déformation correspondant M'_i , que l'on applique aux anciennes coordonnées pour calculer la nouvelle conformation.

Le résultat est visible sur la figure 4.37 : nous obtenons une fermeture des volets de 2.6 Å.

Cette application simple et directe des modes statiques laisse présager des applications cohérentes au docking. Elle met en évidence le fait que les changements conformationnels, même de grande amplitude, peuvent être décrits à l'aide des modes de déformations spécifiques.

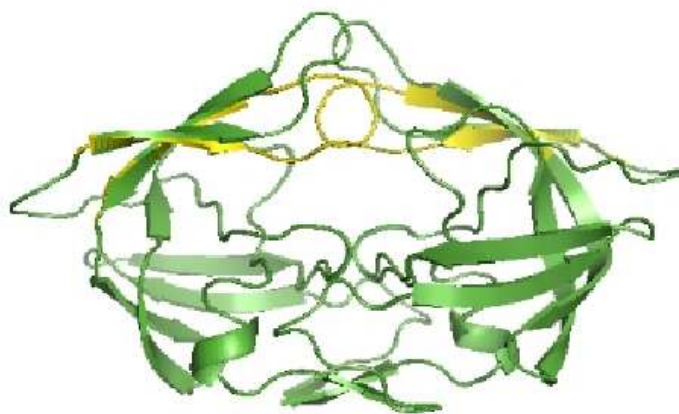


FIG. 4.37 – Fermeture des volets obtenue par l'application de quatre forces sur les C_{α} des résidus Ile50, Ile50', Asp25 et Asp25', en direction d'un inhibiteur potentiel à urée cyclique (voir figure 4.36).

Notons que la figure 4.36 présente une déformation, une tendance à passer d'une conformation ouverte à une conformation fermée. Nous n'atteignons en aucun cas une conformation fermée à l'équilibre. Le passage définitif vers un état stable nécessiterait une seconde étape de minimisation de l'énergie. Dans cette optique, par itération de ce type de procédure, il pourrait être envisagé d'explorer des chemins réactionnels.

4.4.5 Conclusion

Ce premier exemple macromoléculaire permet de valider la cohérence et la puissance des modes statiques. Grâce à un calcul principal, nous avons retrouvé un nombre important de propriétés connues, obtenues au cours de ces 20 dernières années par l'expérimentation ou le calcul :

- l'existence de domaines principaux : les volets, le site actif, l'interface, les poches de liaisons, qui communiquent entre eux
- l'influence des poches de liaison et de l'interface, notamment au niveau des extrémités C-terminales, sur la stabilité du site actif
- l'existence d'un site allostérique au niveau du levier agissant sur l'ouverture des volets
- l'importance des interactions inter-monomères au niveau des volets
- les différences de flexibilité au niveau de l'interface entre les conformations ouverte et fermée pour favoriser l'arrimage du ligand
- l'affaiblissement des rôles mutuels d'un monomère sur l'autre en conformation ouverte
- la fermeture des volets provoquée par l'arrimage d'un ligand

4.5 Conclusions

Au cours de ce quatrième chapitre, nous avons présenté les premières applications de la méthode des Modes Statiques au cas de l'étude des propriétés mécaniques de molécules uniques.

Nous avons montré qu'une fois les modes statiques calculés et stockés, leur exploitation peut être facilement adaptée par l'utilisateur, à l'aide d'outils plus ou moins complexes, selon les propriétés qu'il souhaite explorer. L'excitation optimisée et systématique des atomes par exemple, pour mesurer une réponse donnée, s'avère être un moyen rapide et efficace pour cartographier des régions ayant une action à courte ou longue portée sur cette réponse.

La variété des exemples que nous avons traités dans cette thèse montre tout d'abord l'étendue et la diversité des applications potentielles de la méthode des Modes Statiques. Par ailleurs, elle nous a permis d'aborder différentes échelles et donc différentes problématiques des biomolécules, d'un point de vue neuf et original.

En premier lieu, le cas de la transition de phase du PNIPAM, via l'étude des liaisons hydrogène inter-monomères, a apporté de nouveaux éléments sur l'importance des groupes isopropyles dans le mécanisme d'ouverture des monomères. Caractérisé par l'identification des modes les plus pertinents permettant l'activation ou l'inactivation du site actif, cet exemple peut être abordé comme celui d'un site allostérique, et trouver des applications expérimentales.

L'exploration des changements conformationnels des sucres des acides nucléiques a tout d'abord permis de retrouver des caractéristiques connues, et donc de valider la pertinence de l'approche par Modes Statiques pour ce type d'étude. En outre, le cas plus poussé des acides nucléiques contraints a nécessité une confrontation directe des Modes Statiques à l'expérience. Les résultats de notre travail ont non seulement convergé vers les conclusions de travaux antérieurs sur la stabilité de duplex contenant des D-CNA, mais surtout apporté des éléments d'explication sur le lien entre la structure atomique des D-CNA, la rigidification des sucres, et son effet sur la stabilité des duplex. Nous avons ainsi pu expliquer dans quelle mesure les changements conformationnels peuvent être décrits en termes de modes de déformation spécifiques, et apporter un éclairage nouveau sur la relation entre des modifications locales des nucléotides et des réarrangement globaux de duplex d'acides nucléiques.

Enfin, le traitement de la protéase du VIH-1 a permis de proposer un premier exemple de macromolécule : celui d'une enzyme allostérique. Toujours par une application optimisée et systématique des contraintes, nous avons évalué quelles séquences de la protéine peuvent avoir une influence sur le mécanisme d'ouverture des volets, ou encore la stabilité du site actif. Cette étape nous a permis de retrouver un nombre important de propriétés connues recensées par l'expérience et le calcul. Nous avons également pu identifier certains sites allostériques comme des cibles potentiellement accessibles. La méthode des Modes statiques, en caractérisant certains comportements moléculaires, peut donc offrir la possibilité de localiser de nouvelles régions actives dans une optique de conception de médicaments.

Conclusion générale

La compréhension et le traitement des interactions inter et intra-moléculaires en biologie représentent un enjeu majeur pour l'exploration et l'utilisation des biomolécules à des fins thérapeutiques ou technologiques.

Ces dernières années, nous faisons face à une masse de données expérimentales toujours plus abondante qui vient nourrir des bases de données dont l'utilisation par les chercheurs et industriels du monde entier est devenue quotidienne. Cependant, ces données "statistiques" ne permettent pas aujourd'hui de comprendre les mécanismes fondamentaux d'interaction ou de repliement des biomolécules, passage obligé de la mise en oeuvre d'une modélisation prédictive.

Parallèlement à l'essor des moyens de calcul, nous avons assisté au développement de multiples algorithmes de modélisation des macromolécules. Néanmoins, la difficulté de modéliser de tels systèmes est importante (notamment en raison de leur taille et de la complexité des interactions nécessaires pour les définir), si bien que ces mêmes algorithmes présentent certaines limites :

- soit ils tentent de tenir compte avec précision de la complexité des biomolécules et ne peuvent être utilisés pour un traitement à haut débit (c'est le cas par exemple de la dynamique moléculaire qui apportent des informations au niveau atomique, mais moyennant des temps de calcul inaccessibles pour certains processus biologiques)
- soit ils comportent au contraire trop d'approximations et ne peuvent refléter la réalité des systèmes dans son intégralité (c'est le cas des modes normaux qui décrivent des mouvements collectifs d'atomes, et sont inadaptés au calcul de mouvements anharmoniques)

Alors que la communauté s'accorde aujourd'hui sur l'importance d'intégrer la flexibilité biomoléculaire dans les algorithmes de calculs, les possibilités demeurent actuellement limitées, surtout lorsqu'il s'agit de macromolécules.

La méthode des Modes Statiques que nous avons développée et que nous présentons dans cette thèse offre une alternative compétitive aux approches utilisées jusqu'ici. Son originalité est double :

- au niveau conceptuel, sa force réside dans le fait qu'elle traduit par le calcul les effets "induced-fit" ou "main et gant" qui ont été proposés pour expliquer les interactions entre biomolécules, au-delà du modèle "clé-serrure" pourtant à la base des premiers algorithmes de calcul. Ainsi, cette méthode peut être adaptée avantageusement à l'exploration des propriétés mécaniques et biochimiques d'une molécule seule, ou bien à une procé-

dure de docking. Elle permettra certainement de mieux appréhender les interactions au sens mécanistique.

- au niveau du calcul, les modes statiques sont pré-calculés et organisés au sein d'une base de données qui peut être ensuite utilisée pour générer des déformations sur les édifices moléculaires avec une grande efficacité.

Plus précisément, la méthode des Modes Statiques permet de prendre en compte la flexibilité totale du système, d'étudier les changements conformationnels au travers des déplacements atomiques. L'objectif est d'obtenir, en un calcul, l'ensemble des déformations d'une molécule en réponse à l'excitation systématique de tous les atomes dans toutes les directions.

Pour ce faire, on s'appuie sur la matrice Hessienne d'une biomolécule, son calcul renferme l'ensemble des informations relatives aux interactions intramoléculaires. Par ailleurs, chaque mode statique calculé traduit l'effet d'une excitation externe, c'est-à-dire d'une interaction inter-moléculaire, et représente un champ de déformation associé à la molécule. Notre méthode convient donc parfaitement à un calcul ultérieur de docking.

L'algorithme qui sous-tend le calcul des Modes Statiques, en reposant sur l'inversion de la matrice Hessienne, reste simple et efficace. Il offre un gain de temps important, en fournissant et en stockant dans la matrice des déformations, l'ensemble des modes statiques de la molécule en un seul calcul. Cette matrice peut alors être réutilisée à volonté, pour explorer la flexibilité et les propriétés de la molécule unique, ou pour être intégrée à un processus de docking.

Cette thèse a permis de développer un code : Flexible, et de l'appliquer à des calculs préliminaires ou plus aboutis sur diverses molécules d'intérêt biologique.

Les premiers calculs ont porté sur de courtes séquences de polymère : un dimère et un tétramère de NIPAM. L'intérêt de traiter cette molécule était d'une part sa taille, permettant de nombreux calculs "tests" et l'exploration de mécanismes au niveau atomique. D'autre part, elle présentait un certain nombre de propriétés connues et utilisées par les technologies actuelles, relatives notamment à des changements de conformation et d'hydratation. Nous avons considéré le site d'hydratation comme un site actif, et optimisé les contraintes à appliquer à la molécule pour explorer les mécanismes d'activation/inactivation de ce site. Nous avons ainsi pu mettre en évidence le rôle prépondérant des terminaisons isopropyles de certains monomères des molécules considérées. Nous voyons ici, un premier exemple de l'apport des modes statiques à la compréhension de mécanismes moléculaires fins et pensons pouvoir apporter de nouvelles orientations à l'expérimentateur, comme ici, une nouvelle voie pour intervenir dans la transition de phase naturelle du PNIPAM.

Notre travail s'est ensuite axé sur la flexibilité du sucre des acides nucléiques. Dans une première partie, nous nous sommes intéressés aux modifications structurales du sucre d'un nucléotide en simple et double brin afin d'identifier les différences de flexibilité et d'émettre des hypothèses sur leurs origines. Cette étape nous a permis de retrouver des propriétés de base du sucre, telles que sa rigidification en conformation sud dans un double brin d'ADN B, mais aussi d'établir un lien entre le réarrangement du "backbone" de l'ADN et les changements conformationnels du sucre. Dans ce même axe d'étude, nous avons également appliqué la méthode des Modes Statiques au cas de nucléotides modifiés. Nous avons pu confronter directement nos

résultats à des mesures récentes de la stabilité thermique de duplex comportant des CNA. Nous avons effectivement retrouvé que des modifications (α =g- ; β =t) entraînent un gain de stabilité des duplex, et ce davantage dans le cas de duplex d'ADN que dans celui d'hétéroduplex ADN:ARN, ce qui confirme les capacités prédictives des modes statiques. Nous avons également pu confirmer l'hypothèse selon laquelle la rigidification du sucre 5' en conformation sud expliquerait le faible gain de stabilité dans les hétéroduplex ADN:ARN. En effet, l'exploitation des modes statiques calculés a montré que les sites sollicités favorisant la transition nord/sud du sucre se trouvent immobilisés dans le cycle dioxaphosphorinane créé lors de la synthèse des D-CNA. De tels calculs peuvent éclairer le lien entre les modifications locales et à grande échelle des acides nucléiques. Par ailleurs, on peut imaginer aborder le problème à l'envers : les modes statiques auraient pu indiquer où modifier chimiquement le backbone pour rigidifier le sucre.

Enfin, nous sommes passés à un premier cas de macromolécule avec l'étude de la protéase du VIH-1. Déjà largement traitée comme cible thérapeutique, nous avons recueilli un grand nombre de ses propriétés établies par le calcul ou l'expérience ces 20 dernières années. Par le même type de calculs que précédemment, nous avons retrouvé l'ensemble de ces propriétés, portant notamment sur les mécanismes d'accessibilité au site catalytique, l'existence de sites allostériques, la corrélation inter-monomères, ou encore la stabilité du site actif et la réponse de la molécule à l'arrivée d'un ligand. Ces derniers calculs ont présenté l'avantage de fournir des éléments de validation solides pour l'utilisation de l'approche des Modes Statiques. En outre, avec le calcul des pincements, nous avons pu fournir des exemples novateurs de cartographie de la molécule à des fins de conception de médicaments multi-domaines, qui commencent à trouver leurs premières applications.

Les Modes Statiques ont été proposés à des fins d'intégration de la flexibilité dans les procédures de docking. Une première perspective de développement est donc un algorithme de docking faisant appel aux modes statiques. Cette étape est en cours. Par ailleurs, des efforts restent à mener sur l'optimisation du code, ce qui est en cours également. Enfin, nous prévoyons d'inclure les effets des termes non-linéaires dans la version actuelle du code, qui se place pour le moment dans l'approximation harmonique.

Nous avons décrit une nouvelle méthode que nous pensons prometteuse pour de multiples et diverses applications, de la biologie structurale à la conception de médicaments, en passant par des technologies émergentes de pointe, telles que les nanobiotechnologies. Nous avons présenté des exemples préliminaires qui, sur la base de validations, ont tous permis d'apporter des éléments nouveaux de compréhension des mécanismes moléculaires. Par ailleurs, ils nous ont permis de démontrer les capacités prédictives des modes statiques. L'exploitation des modes statiques a nécessité la mise en place d'un certain nombre d'outils, comme l'optimisation des forces ou les pincements d'atomes. Ce type d'outils peut bien entendu être développé et adapté par l'utilisateur selon les mécanismes qu'il cherche à observer. La variété des outils possiblement existants reflète la diversité des mécanismes moléculaires d'intérêt des objets biologiques.

Annexe A

Structure des peptides et des protéines

A.1 Introduction

Les peptides sont des chaînes d'acides aminés reliés entre eux par liaison peptidique. On parle de protéine lorsque la taille d'un telle chaîne dépasse une centaine d'acides aminés. Leur découverte remonte à 1835, lorsque le chimiste Gerhard Mulder identifie dans le blanc d'oeuf une substance constituée essentiellement d'atomes de carbone, d'oxygène, d'hydrogène et d'azote, mais aussi de phosphore et de soufre. Il montre également que cette substance se retrouve dans tous les tissus vivants, propriété qui lui confère son nom, tiré du grec *prôtos* : premier, essentiel.

Depuis, l'étude des protéines a largement montré leur implication dans tous les processus cellulaires (structure, transport, communication, métabolisme, immunité...). On sait également que leur fonction est étroitement liée à leur conformation qui contient les informations nécessaires pour initier, mener et réguler l'activité protéique. Par conséquent, un repliement anormal de la protéine peut avoir des conséquences néfastes, telles que le développement de maladies conformationnelles (Creutzfeld-Jacob...).

A.2 Différents niveaux de description

La structure des protéines, bien que constituées de briques élémentaires répétées, nécessite plusieurs niveaux de description en raison de la complexité des interactions qui interviennent.

A.2.1 Structure primaire

La structure primaire est la structure chimique, c'est-à-dire la séquence d'acides aminés, unités de base des protéines. Chaque acide aminé est constitué d'un carbone C_{α} asymétrique (à l'exception de la glycine) portant deux groupes fonctionnels : un groupe carboxyle et un groupe amine (dont l'état d'ionisation dépend du pH du milieu), ainsi qu'un radical propre à l'acide aminé (cf Fig.A.1).

La nature de ce radical permet de classer les acides aminés en différents groupes qui influencent le repliement de la séquence : on peut citer les acides aminés non polaires (hydrophobes), polaires (hydrophiles), chargés positivement, chargés négativement.

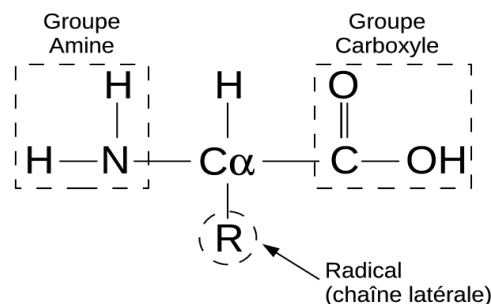


FIG. A.1 – Composition chimique d'un acide aminé.

L'assemblage des acides aminés se fait par une liaison covalente dite peptidique entre le groupe carboxyle d'un acide aminé et le groupe amine du suivant, comme l'illustre la figure Fig.A.2. Par convention, la chaîne commence du côté de l'extrémité ayant un groupe amine libre, dite N-terminale, et se termine à l'extrémité ayant un groupe carboxyle libre, dite C-terminale.

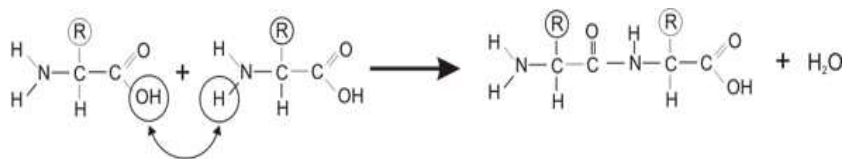


FIG. A.2 – Formation d'une liaison peptidique entre deux acides aminés.

A.2.2 Structure secondaire

Une protéine ne se présente pas sous la simple forme déployée d'une séquence. Elle est au contraire compacte, repleyée dans une structure native et fonctionnelle, imposée par des interactions intramoléculaires (facilitées par la structure primaire) et environnementales (solvant, pH...).

La structure secondaire a été décrite en 1951 par Linus Pauling [375][376]. Elle résulte de l'organisation spatiale d'une séquence d'acides aminés qui interagissent entre eux, provoquant des repliements locaux caractéristiques et énergétiquement favorables.

De tels repliements sont à la fois rendus possibles et limités par les capacités de torsion de la chaîne principale autour de deux axes : $N - C_{\alpha}$ et $C_{\alpha} - C$. Ces torsions sont décrites par deux angles ϕ et ψ tels que le montre la figure Fig.A.3. En excluant les configurations non permises du fait de l'encombrement stérique, un grand nombre de conformations reste possible par des rotations libres autour des liaisons covalentes de la chaîne du squelette ; cependant, seules quelques unes ont une signification biologique.

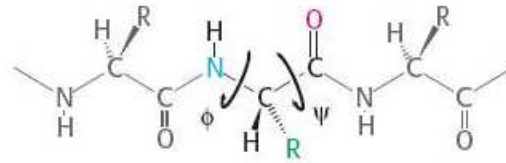


FIG. A.3 – Angles de torsion Φ et Ψ , caractéristiques d'une liaison peptidique.

Il en existe principalement deux structures secondaires, ou conformations locales du squelette [368]: les hélices *alpha* et les feuillets *beta*, conséquences de différents échaffaudages de liaisons hydrogène se créant entre les groupements amide (-NH) et carbonyle (-CO) du squelette peptidique.

- l'hélice α est un repliement hélicoïdal périodique dû à des liaisons hydrogène établies entre le groupement carbonyle d'un résidu n et le groupement amide d'un résidu $n + 4$. En moyenne, un tour d'hélice contient 3,6 résidus et mesure 0.54 nm. Le coeur de l'hélice est très compact, et renforcé par des interactions de van der Waals à travers l'hélice, les résidus sont rejetés à l'extérieur.

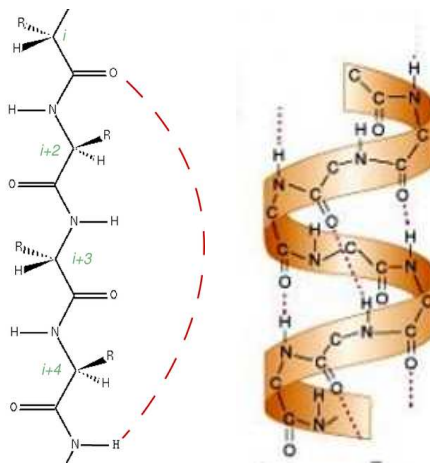


FIG. A.4 – Représentation d'un hélice α .

- le feuillet β est stabilisé par le plus grand nombre de liaisons hydrogène possible entre les atomes d'hydrogène de l'azote et les atomes d'oxygène carbonyles des séquences de polypeptides adjacents les uns aux autres. Ils sont dits parallèles ou antiparallèles selon la direction des chaînes.

Le reste des acides aminés se trouve dans des régions de coudes, de boucles ou de structures irrégulières.

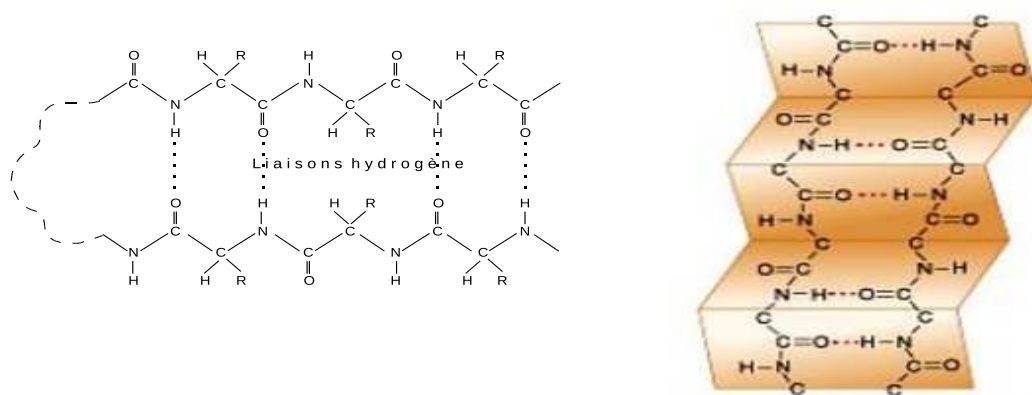


FIG. A.5 – Représentation d'un feuillet β .

A.2.3 Structure tertiaire

La structure tertiaire décrit le repliement dans l'espace de la chaîne polypeptidique [374]. Elle est stabilisée par :

- des liaisons hydrogène, notamment entre les groupements polaires et les molécules d'eau, mais aussi entre les résidus aminoacyles entre eux
- des interactions hydrophobes mettant en jeu des chaînes latérales non polaires
- des interactions électrostatiques (ponts salins) qui se forment entre des groupes de charges opposées
- des interactions des van der Waals

A.2.4 Structure quaternaire

La structure quaternaire résulte de plusieurs séquences protéiques sous forme d'un ensemble de sous-unités (monomères), associées entre elles par des liaisons non covalentes, les liaisons hydrophobes étant prépondérantes.

A.3 Acides aminés

On peut classer les acides aminés, en fonction de leur chaîne latérale, par groupes polaires, non polaires, chargés et non chargés. Le tableau A.1 réunit les 20 acides aminés les plus répandus.

Cas à part Glycine (G, Gly) $\begin{array}{c} \text{COO}^- \\ \\ ^+\text{H}_3\text{N}-\text{C}-\text{H} \\ \\ \text{H} \end{array}$					
Sérine (S, Ser) $\begin{array}{c} \text{COO}^- \\ \\ ^+\text{H}_3\text{N}-\text{C}-\text{H} \\ \\ \text{CH}_2\text{OH} \end{array}$ Hydroxyl-	Thréonine (T, Thr) $\begin{array}{c} \text{COO}^- \\ \\ ^+\text{H}_3\text{N}-\text{C}-\text{H} \\ \\ \text{CHOH} \\ \\ \text{CH}_3 \end{array}$	Asparagine (N, Asn) $\begin{array}{c} \text{COO}^- \\ \\ ^+\text{H}_3\text{N}-\text{C}-\text{H} \\ \\ \text{CH}_2 \\ \\ \text{C}=\text{O} \\ \\ \text{H}_2\text{N} \end{array}$ Amido-	Glutamine (Q, Gln) $\begin{array}{c} \text{COO}^- \\ \\ ^+\text{H}_3\text{N}-\text{C}-\text{H} \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{C}=\text{O} \\ \\ \text{H}_2\text{N} \end{array}$	Polaires et non-chargés	
Acide glutamique (E, Glu) $\begin{array}{c} \text{COO}^- \\ \\ ^+\text{H}_3\text{N}-\text{C}-\text{H} \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{COO}^- \end{array}$ Négativement	Acide aspartique (D, Asp) $\begin{array}{c} \text{COO}^- \\ \\ ^+\text{H}_3\text{N}-\text{C}-\text{H} \\ \\ \text{CH}_2 \\ \\ \text{COO}^- \end{array}$	Lysine (K, Lys) $\begin{array}{c} \text{COO}^- \\ \\ ^+\text{H}_3\text{N}-\text{C}-\text{H} \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{H}_3\text{N}^+ \end{array}$	Arginine (R, Arg) $\begin{array}{c} \text{COO}^- \\ \\ ^+\text{H}_3\text{N}-\text{C}-\text{H} \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{NH} \\ \\ \text{C}=\text{NH}_2^+ \\ \\ \text{NH}_2 \end{array}$	Histidine (H, His) $\begin{array}{c} \text{COO}^- \\ \\ ^+\text{H}_3\text{N}-\text{C}-\text{H} \\ \\ \text{CH}_2 \\ \\ \text{C}-\text{NH} \\ \\ \text{C}-\text{N}^+ \\ \\ \text{H} \end{array}$ Positivement	Polaires et chargés
Méthionine (M, Met) $\begin{array}{c} \text{COO}^- \\ \\ ^+\text{H}_3\text{N}-\text{C}-\text{H} \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{S} \\ \\ \text{CH}_3 \end{array}$ Contenant du soufre	Cystéine (C, Cys) $\begin{array}{c} \text{COO}^- \\ \\ ^+\text{H}_3\text{N}-\text{C}-\text{H} \\ \\ \text{CH}_2 \\ \\ \text{SH} \end{array}$	Tryptophane (W, Trp) $\begin{array}{c} \text{COO}^- \\ \\ ^+\text{H}_3\text{N}-\text{C}-\text{H} \\ \\ \text{CH}_2 \\ \\ \text{C}=\text{CH} \\ \\ \text{NH} \\ \\ \text{C}_6\text{H}_5 \end{array}$	Phénylalanine (F, Phe) $\begin{array}{c} \text{COO}^- \\ \\ ^+\text{H}_3\text{N}-\text{C}-\text{H} \\ \\ \text{CH}_2 \\ \\ \text{C}_6\text{H}_5 \end{array}$	Tyrosine (Y, Tyr) $\begin{array}{c} \text{COO}^- \\ \\ ^+\text{H}_3\text{N}-\text{C}-\text{H} \\ \\ \text{CH}_2 \\ \\ \text{C}_6\text{H}_4 \\ \\ \text{OH} \end{array}$ Aromatiques	Non-polaires
Alanine (A, Ala) $\begin{array}{c} \text{COO}^- \\ \\ ^+\text{H}_3\text{N}-\text{C}-\text{H} \\ \\ \text{CH}_3 \end{array}$	Valine (V, Val) $\begin{array}{c} \text{COO}^- \\ \\ ^+\text{H}_3\text{N}-\text{C}-\text{H} \\ \\ \text{CH} \\ \\ \text{CH}_3 \quad \text{CH}_3 \end{array}$	Leucine (L, Leu) $\begin{array}{c} \text{COO}^- \\ \\ ^+\text{H}_3\text{N}-\text{C}-\text{H} \\ \\ \text{CH}_2 \\ \\ \text{CH} \\ \\ \text{CH}_3 \quad \text{CH}_3 \end{array}$	Isoleucine (I, Ile) $\begin{array}{c} \text{COO}^- \\ \\ ^+\text{H}_3\text{N}-\text{C}-\text{H} \\ \\ \text{HC}-\text{CH}_3 \\ \\ \text{CH}_2 \\ \\ \text{CH}_3 \end{array}$	Proline (P, Pro) $\begin{array}{c} \text{COO}^- \\ \\ \text{HN}-\text{C}-\text{H} \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \end{array}$	Non-polaires
Aliphatiques					

TAB. A.1 – Tableau récapitulatif des 20 acides aminés.

Annexe B

Structure des acides nucléiques

B.1 Historique

En 1866, les travaux du moine et botaniste Gregor Mendel sur l'hybridation des végétaux lui permettent d'établir les premières considérations sur la transmission héréditaire de caractères physiques dans un article qui passe alors inaperçu [384]. Ses conclusions selon lesquelles les caractéristiques héréditaires sont contenues en double dans chaque être vivant (allèles) et qu'une seule des deux est transmise au descendant par chaque parent seront cependant reprises au début du XXème siècle sous le nom des "lois de Mendel", constituant les bases de la génétique.

Mais quel est donc le support de l'hérédité ?

Sa découverte remonte en 1869, lorsque Friedrich Miesher, menant alors des travaux sur le noyau des cellules, isole pour la première fois une substance très riche en phosphore qu'il baptise "nucléine" sans pour autant déterminer son rôle [385]. Sans le savoir, F. Miesher vient de découvrir le support de l'information génétique.

En 1889, Richard Altmann parvient, à partir de la nucléine, à séparer des protéines d'une substance acide qu'il baptise "acide nucléique" [377]. Par la suite, Albrecht Kossel, avec l'aide de Levene et Jacobs [383] décrit l'acide nucléique comme une suite séquentielle d'unités composées chacune d'un phosphate, d'un sucre et d'une base azotée. Il identifie cinq bases azotées : la guanine (notée G par la suite) en 1882, l'adénine (A) en 1886, la thymine (T) en 1893, la cytosine (C) en 1894 et l'uracile (U) en 1900. D'autre part, il décrit deux types d'acides nucléiques se distinguant par leurs oses, identifiés par Levene et Jacobs : le ribose en 1908 et le désoxyribose en 1929, qui confèrent respectivement leur nom à l'acide ribonucléique (ARN) et l'acide désoxyribonucléique (ADN). Il obtient le prix Nobel de physiologie ou médecine en 1910 pour "sa contribution à la connaissance de la chimie de la cellule et en particulier pour ses travaux sur les protéines et les substances nucléiques".

Il faudra attendre 1944 pour que Oswald T. Avery, Colin MacLeod et Maclyn McCarty [378] suggèrent que l'ADN puisse fonctionner comme support du matériel génétique, une théorie confirmée en 1952 par Alfred Hershey et Martha Chase grâce à leurs travaux sur l'infection d'une cellule par un virus [382].

Quant à la structure de l'ADN, une première hypothèse est émise en 1943 par William Astbury qui la décrit, grâce à ses expériences par diffraction de rayons X, comme régulière et périodique, "like a pile of pennies". En 1953, Rosalind Franklin et Maurice Wilkins réalisent le cliché d'une molécule d'ADN par diffraction de rayons X (fig. B.1) qui permet d'établir la structure hélicoïdale de l'ADN [388]. C'est en se basant sur cette découverte, ainsi que sur celle de Erwin Ghargaff sur la complémentarité des bases (il montre en 1950 que le rapport A+T/C+G est variable selon les espèces, mais constant entre membres d'une même espèce et que les rapports C/G et A/T sont toujours égaux à 1 [379]) que James Watson et Francis Crick présentent le modèle en double hélice de l'ADN en 1953 [386]. Ils recevront en 1962 le prix Nobel de médecine pour cette découverte qui est à la base de la compréhension du code génétique, des mécanismes de traduction, de réplication...

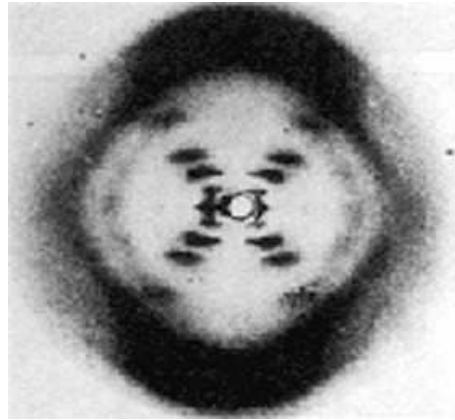


FIG. B.1 – Cliché d'une molécule d'ADN par diffraction de rayons X réalisé par Rosalind Franklin au King's College de Londres en 1953. Il confirme qu'une molécule d'ADN forme une hélice d'un diamètre de 2 nm et d'un pas de 3.4 nm.

B.2 Structure

Les acides nucléiques sont donc des polymères de nucléotides [360][387]. Il en existe deux types : l'ADN, support de l'information génétique, et l'ARN, vecteur de l'expression de cette information jusqu'à la synthèse des protéines.

Chaque nucléotide est constitué d'un groupe phosphate, d'un sucre et d'une base azotée :

- *le groupement phosphate* est entièrement ionisé au pH physiologique
- *le sucre* est un pentose, plus précisément un furanose, cycle composé de quatre atomes de carbone et d'un atome d'oxygène. Il n'est pas plan et peut adopter des conformations particulières impliquant des modifications à grande échelle dans la structure des acides nucléiques. Le sucre est une des deux différences chimiques existant entre l'ADN et l'ARN : le premier contient un désoxyribose et le second un ribose, comportant respectivement un hydrogène (H) ou un groupe hydroxyle (OH) en position 2' comme le montre la figure Fig.B.2.

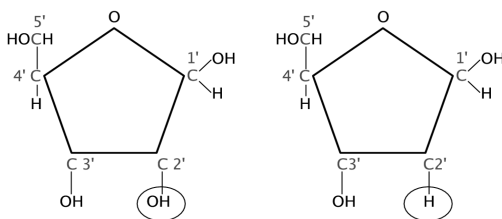


FIG. B.2 – Structures du ribose et du désoxyribose.

- les bases azotées ou nucléiques sont de deux types comme le montre la figure Fig.B.3: les bases puriques (adénine et guanine) et les bases pyrimidiques (cytosine et thymine). Notons que dans l'ARN, l'uracile se substitue à la thymine, c'est la seconde différence chimique entre ADN et ARN. Complémentaires, elles peuvent s'apparier deux à deux : l'adénine avec la thymine (dans l'ADN) ou l'uracile (dans l'ARN) à l'aide de deux liaisons hydrogène et la guanine avec la cytosine à l'aide de trois liaisons hydrogène (Fig.B.4).

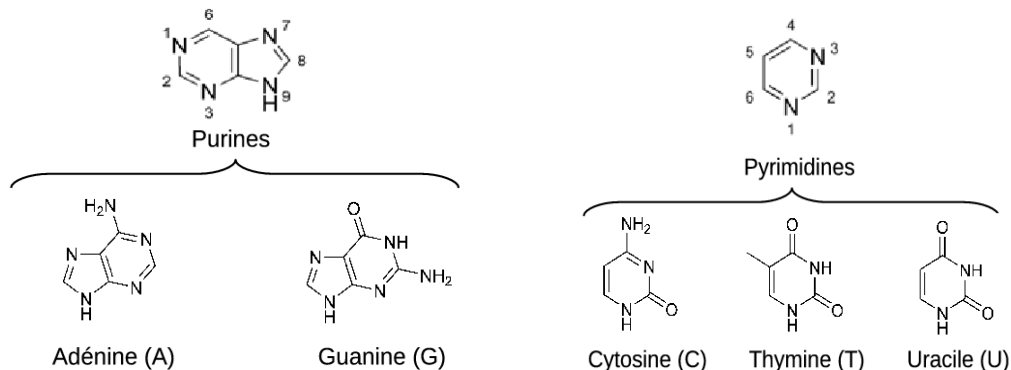


FIG. B.3 – Composition chimique des bases puriques et pyrimidiques.

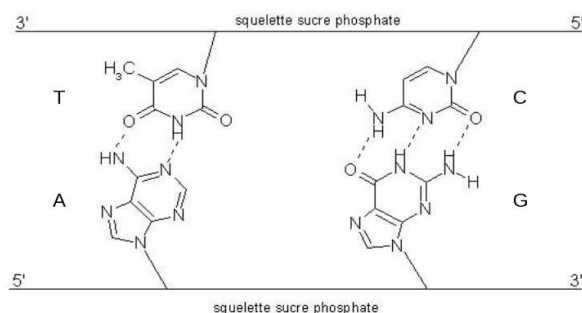


FIG. B.4 – Appariement des bases par liaisons hydrogène.

Un sucre et une base, liés par une liaison covalente N-glycosidique impliquant le carbone C1' de l'ose et l'azote N1 de la pyrimidine ou N9 de la purine, forment un nucléoside. Ce nucléoside lui-même lié par liaison covalente phosphoester avec un groupe phosphate sur le carbone C5' du pentose constitue un nucléotide, ou nucléoside monophosphate.

Ainsi, l'enchaînement des nucléosides liés les uns aux autres par liaison phosphodiester - le groupe phosphate se liant au carbone C5' du sucre du premier nucléoside et au carbone C3' du sucre du nucléoside suivant - confère aux acides nucléiques une structure séquentielle dont la brique unitaire est un 3',5'-nucléotide (Fig.B.5).

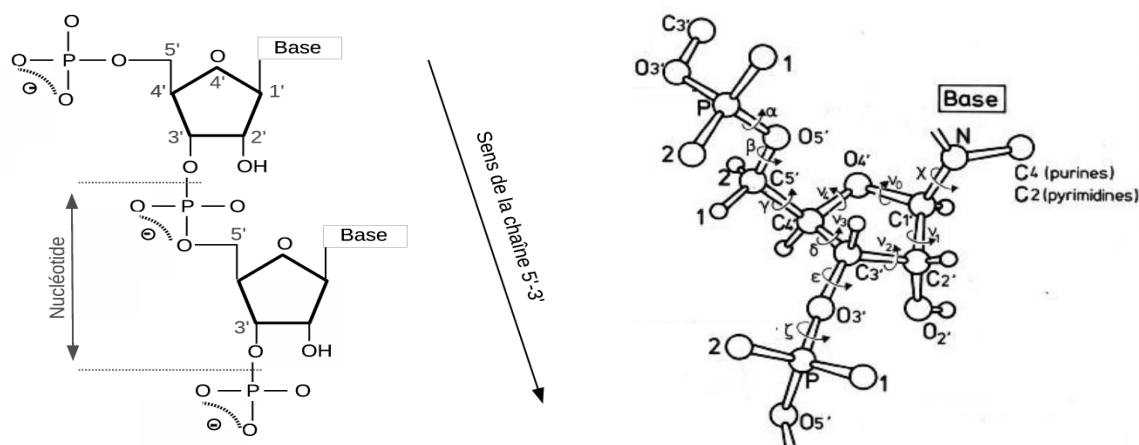


FIG. B.5 – Structure d’une chaîne d’acide nucléique. A droite, localisation des angles de torsion du backbone de α à ζ . *Source Springer-Verlag [360].*

B.3 Configuration

ADN et ARN ne diffèrent donc chimiquement que par leur sucre et une base azotée. On trouve cependant l'ADN sous la forme d'une hélice double brin et l'ARN généralement en simple brin. Si une double hélice d'ARN se forme, c'est souvent par le repliement d'un simple brin sur lui-même. Néanmoins, les singularités chimiques de l'ARN ne lui ôtent pas sa faculté de former une double hélice. On le trouve d'ailleurs dans cette configuration en formant une hélice complémentaire avec une hélice d'ADN.

La structure en double hélice se forme par l'appariement de deux séquences complémentaires et antiparallèles via des liaisons hydrogène se formant entre les bases. La stabilité thermodynamique de l'hélice est due à ces liaisons hydrogène, mais aussi aux interactions de stacking (ou empilement) entre les bases. Ces forces sont développées dans la partie 1.3.

Il existe principalement trois formes de doubles hélices : les formes B et A dans lesquelles l'hélice tourne à droite et la forme Z dans laquelle l'hélice tourne à gauche [381], [44]. Leurs caractéristiques sont résumées dans le tableau B.1.

- la *forme B* est la forme la plus répandue de l'ADN. L'axe de l'hélice est perpendiculaire au plan des bases et centré sur celles-ci. En moyenne, chaque pas est de 3.4 nm, soit 10,4 paires de bases (pb) avec une rotation de 34.6° entre chaque sucre. Elle présente un petit sillon et un grand sillon avec lequel peuvent s'opérer des interactions avec les protéines qui se lient à des séquences spécifiques d'ADN.
- la *forme A* est celle de l'ARN double brin et des duplex ADN-ARN apparaissant lors de la transcription (l'ADN passe alors de la forme B vers la forme A pour se lier à l'ARN qui est de type A, les groupes 2'-hydroxyles empêchant l'adoption de la forme B). Le type A est plus compact que le B (11 paires de bases par tour), les bases sont inclinées par rapport à l'axe de l'hélice. Le grand sillon est peu accessible, plus profond, avec des bases enfouies sous les groupements phosphate. Le petit sillon est superficiel.
- la *forme Z* tire son nom de la disposition en zigzag du squelette sucre-phosphate. L'ADN-Z serait une forme transitoire de l'ADN B *in vivo*, favorisée par l'alternance de purines et de pyrimidines. L'hélice présente 12 paires de bases par tour, inclinées par rapport à l'axe de l'hélice. Elle possède un seul sillon profond.

Quelques caractéristiques de l'ADN-B ainsi que les structures respectives des formes A, B et Z sont représentées sur la figure Fig.B.6. Un récapitulatif des différences structurales des ADN-A, B et Z est présenté dans le tableau B.1.

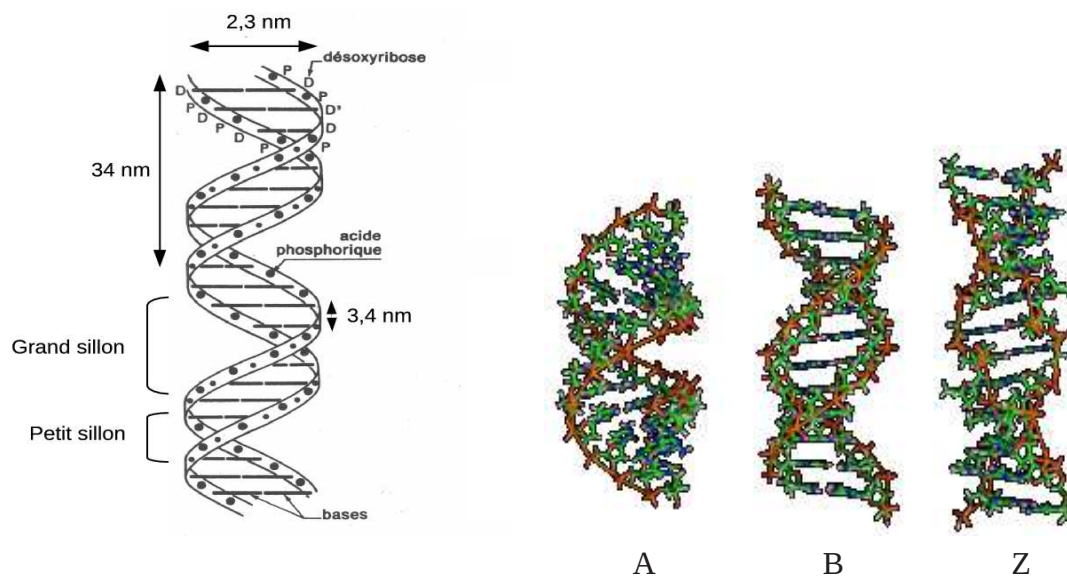


FIG. B.6 – A gauche : représentation schématique d'une double hélice d'ADN. A droite : structures des formes A, B et Z respectivement.

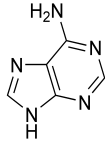
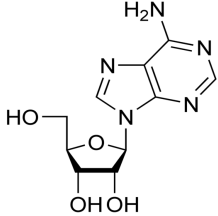
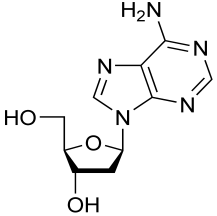
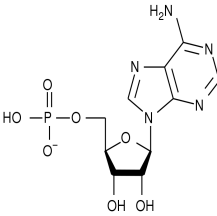
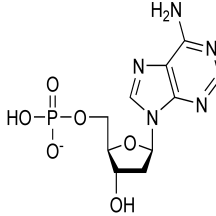
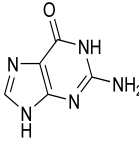
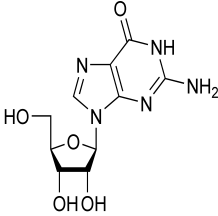
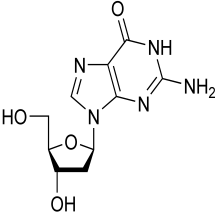
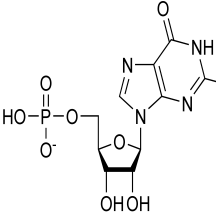
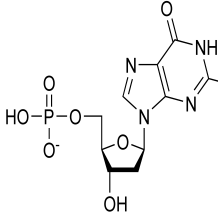
	A	B	Z
pb/tour	11	10	12
rotation/pb	33.6°	35.9°	-30°
inclinaison des pb	+19°	-1.2°	-9°
pas par tour	24.6 Å	43.2 Å	35.6 Å
diametre de l'hélice	25.5 Å	13.7 Å	18.4 Å

TAB. B.1 – Comparaison de quelques caractéristiques structurales des formes A, B et Z. *Source Science* [380].

B.4 Récapitulatif des bases, nucléosides et nucléotides

B.4.1 Purines

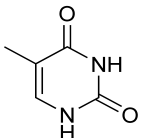
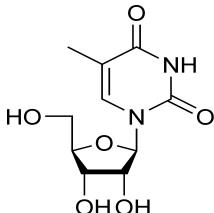
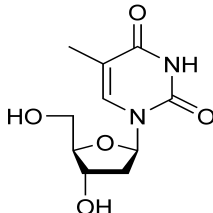
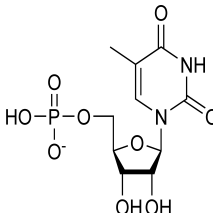
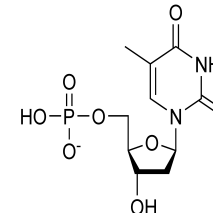
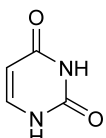
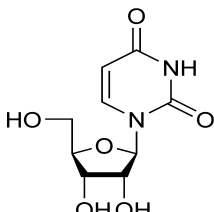
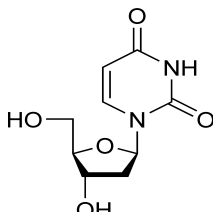
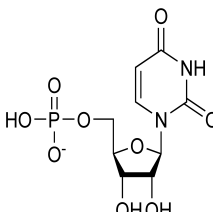
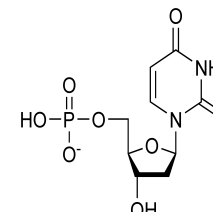
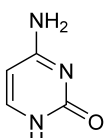
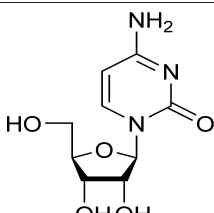
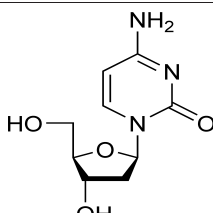
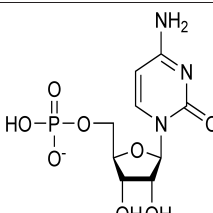
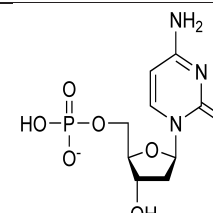
Le tableau B.2 présente les bases puriques et leurs dérivés.

Base nucléique	Nucléoside	Désoxynucléoside	Nucléotide	Désoxynucléotide
 Adénine	 Adénosine A	 Désoxyadénosine dA	 Adénosine monophosphate AMP	 Désoxyadénosine monophosphate dAMP
 Guanine	 Guanosine G	 Désoxyguanosine dG	 Guanosine monophosphate GMP	 Désoxyguanosine monophosphate dGMP

TAB. B.2 – Tableau récapitulatif des bases puriques et de leurs composés.

B.4.2 Pyrimidines

Le tableau B.3 présente les bases pyrimidiques et leurs dérivés.

Base nucléique	Nucléoside	Désoxynucléoside	Nucléotide	Désoxynucléotide
 Thymine	 Ribothymidine T	 Désoxythymidine dT	 Thymidine monophosphate TMP	 Désoxythymidine monophosphate dTMP
 Uracile	 Uridine U	 Désoxyuridine dU	 Uridine monophosphate UMP	 Désoxyuridine monophosphate dUMP
 Cytosine	 Cytidine C	 Désoxycytidine dC	 Cytidine monophosphate CMP	 Désoxycytidine monophosphate dCMP

TAB. B.3 – Tableau récapitulatif des bases pyrimidiques et de leurs composés.

Annexe C

Les modes normaux

C.1 Principe

Il est possible de décomposer le mouvement compliqué et apparemment désordonné d'un système de points matériels en mouvements plus simples, en supposant que ces mouvements sont de petite amplitude. A chaque instant, la position du système de points est donnée par la somme géométrique des positions qu'aurait le système s'il était seulement soumis à des mouvements élémentaires.

Ces mouvements élémentaires sont nommés modes normaux de vibration, ou modes propres. Ils sont indépendants les uns des autres et traduisent les vibrations spontanées du système. Dans le cas d'une molécule, si elle arrive à vibrer dans un de ces modes (ce qui est impossible pendant longtemps en raison des collisions moléculaires) aucune autre vibration n'est excitée sans perturbation extérieure et elle continue à vibrer dans ce mode. Dans chaque mode normal, les atomes vibrent en phase et avec la même fréquence. L'énergie totale du système peut s'écrire comme la somme de l'énergie de chacun des modes normaux excités. Chaque mode normal a sa fréquence d'excitation propre.

L'analyse des modes normaux est actuellement un outil de prédilection pour étudier les mouvements collectifs des atomes d'un système placé dans une approximation harmonique : chaque atome est soumis à des vibrations de faible amplitude autour de sa position d'équilibre. Si les oscillateurs sont en interaction assez importante, ils sont animés de mouvements collectifs caractérisés comme un couplage d'oscillateurs harmoniques indépendants.

Dans le cas d'une molécule polyatomique non linéaire de N atomes, chaque atome a 3 degrés de liberté car il peut se déplacer indépendamment selon chacun des trois axes. La molécule dans son ensemble compte donc $3N$ degrés de liberté. Parmi ces $3N$ degrés de liberté, 3 correspondent à une translation de la molécule entière le long des trois axes et trois correspondent à une rotation de la molécule autour de chacun des axes. Ceci laisse $3N - 6$ degrés de liberté de vibration et autant de modes normaux de vibration. Une molécule linéaire n'en possède que $3N - 5$ car elle a seulement 2 degrés de liberté de rotation (on ne tient pas compte de la rotation autour de l'axe internucléaire).

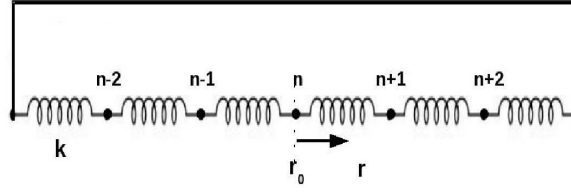


FIG. C.1 – Schéma d'un système de points matériels représentés en position d'équilibre r_0 et déplacés vers une position r . Leur mouvement est associé à celui d'un oscillateur harmonique représenté par un ressort de constante de raideur k .

C.2 Calcul des modes propres

Considérons une molécule de N atomes. Dans ce cas, ses équations de mouvements dans un potentiel harmonique sont données par :

$$M.\ddot{r} = -K.(r - r_0) \quad (C.1)$$

où M est une matrice diagonale $3N \times 3N$ contenant les masses des atomes, K est la matrice des constantes de raideur, r la position et r_0 la position à l'équilibre. Ces équations sont caractéristiques d'un oscillateur harmonique tridimensionnel. Si l'on introduit des grandeurs normées par rapport aux masses :

$$\begin{aligned} \tilde{r} &= \sqrt{M}.r \\ \tilde{r}_0 &= \sqrt{M}.r_0 \\ \tilde{K} &= \sqrt{M}^{-1}.K.\sqrt{M} \end{aligned}$$

les équations du mouvements C.1 peuvent s'écrire :

$$\ddot{\tilde{r}} = \tilde{K}.(\tilde{r} - \tilde{r}_0) \quad (C.2)$$

Les solutions du système d'équations C.2 sont de la forme :

$$\tilde{r}(t) = \tilde{r}_0 + A_i \cos(\omega_i t + \delta_i) \quad i = 1, \dots, 3N$$

où δ_i est un facteur de phase arbitraire et ω_i et A_i sont les solutions de l'équation aux valeurs propres :

$$K.A_i = \omega_i A_i$$

ω_i est la fréquence de vibration. A_i , son amplitude, indique jusqu'où on peut déformer dans la direction du mode normal correspondant, cette grandeur est relative à l'énergie nécessaire pour déformer.

Lors de l'analyse par modes normaux, les six premières valeurs propres (correspondant aux 6 mouvements de corps rigide de translation et de rotation) sont nulles : ces mouvements ne requièrent pas d'énergies. Les modes dits de plus basse énergie, utilisés lors de l'analyse vibrationnelle des biomolécules sont les premiers modes d'énergie non nulle.

Annexe D

Résolution de systèmes linéaires : méthode de Gauss-Seidel

D.1 Position du problème

La résolution de systèmes linéaires par des méthodes directes dépend en particulier de la capacité du calculateur. Au delà d'un certain nombre d'équations à un certain nombre d'inconnues, les méthodes directes deviennent inappropriées. On peut alors avoir recours aux méthodes itératives qui permettent de converger vers la solution du système, à une erreur donnée près.

Nous décrivons dans cette annexe la méthode de Gauss-Seidel, qui, tout comme la méthode de Jacobi [290], est basée sur une décomposition additive de matrice.

Pour ce faire, nous résolvons le système $A \cdot X = B$, où A est une matrice inversible, en construisant une suite de vecteurs $\vec{X} \in R^n$, où $\vec{X} = [x_1 \dots x_n]^T$. Le vecteur optimal est généralement obtenu après un certain nombre d'itérations lorsqu'on atteint une précision souhaitée ε appelée également critère d'arrêt.

D.2 Principe

D.2.1 Ecriture du système

Les méthodes itératives consistent à utiliser un vecteur initial $X^0 = [x_1^0 \dots x_n^0]^T$ afin de produire une suite de vecteurs du type :

$$X^{k+1} = F^k \left(x_1^{k-1} \dots x_n^{k-1} \right)$$

Le système $A \cdot X = B$ où A est une matrice carrée d'ordre n peut s'écrire sous une autre forme identique $(M - N) \cdot X = B$ en décomposant $A = M - N$.

A l'aide d'un vecteur initial X^0 et de cette décomposition, on peut donc générer une suite de la façon suivante :

$$\begin{cases} X^1 = M^{-1} \cdot N \cdot X^0 + M^{-1} \cdot B \\ X^2 = M^{-1} \cdot N \cdot X^1 + M^{-1} \cdot B \\ \dots \\ X^{k+1} = M^{-1} \cdot N \cdot X^k + M^{-1} \cdot B \end{cases}$$

Cette suite est représentée par la relation itérative suivante :

$$X^{k+1} = T \cdot X^k + V$$

où $T = M^{-1} \cdot N$ et $V = M^{-1} \cdot B$.

Cependant, nous ne pouvons pas savoir si le vecteur estimé se dirige vers la solution optimale si un critère de convergence n'est pas défini. Pour cela, un vecteur d'erreur est établi par la relation :

$$\epsilon^k = X^k - X^\bullet = T \cdot \epsilon^{k-1}$$

Autrement dit, la convergence existe si l'erreur tend vers 0 lorsqu'on se rapproche de la solution optimale :

$$X^\bullet \approx X^k \text{ si } \lim_{k \rightarrow \infty} \epsilon^k = 0$$

Remarque : La convergence se produit lorsque le résidu ou la précision ϵ^k reste inférieur à ϵ^{k-1} .

D.2.2 Décomposition de la matrice A

La décomposition de la matrice A est nécessaire pour assurer la convergence de la méthode. Nous créons trois matrices D, L, U , telles que $A = D - L - U$.

D est une matrice diagonale :

$$D = \begin{bmatrix} a_{1,1} & 0 & \dots & \dots & 0 \\ 0 & \ddots & \ddots & 0 & \vdots \\ \vdots & \ddots & a_{i,j} & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & a_{n,n} \end{bmatrix}$$

L est une matrice inférieure :

$$L = \begin{bmatrix} 0 & 0 & \dots & \dots & 0 \\ a_{2,1} & \ddots & \ddots & 0 & \vdots \\ \vdots & \ddots & 0 & \ddots & \vdots \\ \vdots & a_{i,j} & \ddots & \ddots & 0 \\ a_{n,1} & \dots & \dots & a_{n,n-1} & 0 \end{bmatrix}$$

U est une matrice supérieure :

$$U = \begin{bmatrix} 0 & a_{1,2} & \dots & \dots & a_{1,n} \\ 0 & \ddots & \ddots & a_{i,j} & \vdots \\ \vdots & \ddots & 0 & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & a_{n-1,n} \\ 0 & \dots & \dots & 0 & 0 \end{bmatrix}$$

D.3 Méthode de Gauss-Seidel

D.3.1 Description de la méthode

La méthode de Gauss-Seidel est une des méthodes itératives permettant de résoudre ce type de système.

On pose $A = M - N$ où $M = D - L$ et $N = U$. Dans ces conditions :

$$T = (D - L)^{-1} \cdot U \cdot X^{k+1} =^T (D - L)^{-1} \cdot U \cdot X^k + (D - L)^{-1} \cdot B$$

On peut donc écrire le système $A \cdot X = B$ sous la forme itérative suivante :

$$(D - L) \cdot X^{k+1} = U \cdot X^k + B \text{ soit } X^{k+1} = (D - L)^{-1} \cdot U \cdot X^k + (D - L)^{-1} \cdot B$$

Remarque : Les pivots $a_{i,j}$ doivent être non nuls. Dans le cas contraire il suffit d'intervertir les lignes pour remplir la condition nécessaire.

D.3.2 Condition d'arrêt

On note r un vecteur résidu tel que $r^k = b - A \cdot X^k$, de sorte que le critère d'arrêt soit :

$$\frac{\|r^k\|}{\|B\|} < \varepsilon \text{ avec } \varepsilon \text{ choisi petit} \quad (\text{D.1})$$

Une autre technique consiste à utiliser un autre test d'arrêt basé sur :

$$\frac{\|X^k - X^{k-1}\|}{\|X^k\|} < \varepsilon \quad (\text{D.2})$$

Lorsque l'optimum est voisin de 0, on se contente du critère d'arrêt suivant :

$$\|X^k - X^{k-1}\| < \varepsilon \quad (\text{D.3})$$

D.3.3 Algorithme

On choisit un vecteur initial X^0 et un critère d'arrêt ε , et on crée :

– D avec $d_{i,j} = a_{i,j}$

– L avec :

$$\begin{cases} l_{i,j} = -a_{i,j} & \text{pour } i > j \\ l_{i,j} = 0 & \text{pour } i \leq j \end{cases}$$

– U avec :

$$\begin{cases} u_{i,j} = -a_{i,j} & \text{pour } i < j \\ u_{i,j} = 0 & \text{pour } i \geq j \end{cases}$$

Enfin, tant que les conditions (D.1), (D.2) et (D.3) sont vérifiées, on calcule $r^k = B - A \cdot X^k$ et

$$X^{k+1} = (D - L)^{-1} \cdot U \cdot X^k + (D - L)^{-1} \cdot B$$

Remarque : Cette méthode présente deux avantages : une convergence rapide tout d'abord, et par ailleurs, une économie de stockage. En effet, elle ne nécessite pas l'introduction de valeurs additionnelles pour X , puisque chaque nouvelle valeur X^{k+1} trouvée remplace directement X^k dans la suite de l'itération.

Bibliographie

Références de l'introduction générale

- [1] Ameziane N., Bogard M., Lamoril J., *Principes de biologie moléculaire en biologie clinique*. Elsevier Masson, 2006.
- [2] Carlson H.A., McCammon J.A., *Accommodating protein flexibility in computational drug design*. Pharmacology 57, 213-218, 2000.
- [3] Carlson H.A., *Protein flexibility and drug design: how to hit a moving target*. Curr. Opin. Chem. Biol. 6, 447-452, 2002.
- [4] Carlson H.A., *Protein flexibility is an important component of structure-based drug discovery*. Curr. Pharm. Des. 8, 1571-1578, 2002.
- [5] Erickson J.A., Jalaie M., Robertson D.H., Lewis R.A., Vieth M., *Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy*. J. Med. Chem. 47, 45-55, 2004.
- [6] Figeys D., *Novel approaches to map protein interactions*. Curr. Opin. Biotechnol. 14, 119-125, 2003.
- [7] Gavin A.C., Superti-Furga G., *Protein complexes and proteome organization from yeast to man*. Curr. Opin. Chem. Biol. 7, 21-27, 2003.
- [8] Giot L., Bader J.S., Brouwer C., Chaudhuri A., Kuang B., Li Y., Hao Y.L., Ooi C.E., Godwin B., Vitols E., Vijayadamodar G., Pochart P., Machineni H., Welsh M., Kong Y., Zerhusen B., Malcolm R., Varrone Z., Collis A., Minto M., Burgess S., McDaniel L., Stimpson E., Spriggs F., Williams J., Neurath K., Ioime N., Agee M., Voss E., Furtak K., Renzulli R., Aanensen N., Carrolla S., Bickelhaupt E., Lazovatsky Y., DaSilva A., Zhong J., Stanyon C.A., Finley R.L., White K.P., Braverman M., Jarvie T., Gold S., Leach M., Knight L., Shimkets R.A., McKenna M.P., Chant J., Rothberg J.M.E., *A protein interaction map of Drosophila melanogaster*. Science 302, 1727-1736, 2003.
- [9] Hart T.N., Read R.J., *A multiple-start Monte Carlo docking method*. Proteins 13, 206-222, 1992.
- [10] Human Genome Project . http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml
- [11] Knetgel R.M.A., Kuntz I.D., Oshiro C.M., *Molecular docking to ensembles of protein structures*. J. Mol. Biol. 266, 424-440, 1997.

- [12] Kuntz I.D., Blaney J.M., Oatley S.J., Langridge R., Ferrin T.E., *A geometric approach to macromolecule-ligand interactions*. J. Mol. Biol. 161, 269-288, 1982.
- [13] Leach A.R., *Ligand docking to proteins with discrete sidechain flexibility*. J. Mol. Biol. 235, 245-356, 1994.
- [14] Osterberg F., Morris G.M., Sanner M.F., Olson A.J.,Goodsell D.S., *Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock*. Proteins 46, 34-40, 2002.
- [15] Pak Y.S., Wang S., *Application of a molecular dynamics simulation method with a generalized effective potential to the flexible molecular docking problems*. J. Phys. Chem. B 104, 354-359, 2000.
- [16] Sobott F., Robinson C.V., *Protein complexes gain momentum*. Curr. Opin. Struct. Biol. 12, 729-734, 2002.
- [17] Sousa S.F., Fernandes P.A., Ramos M.J., *Protein-ligand docking: current status and future challenges*. Proteins 65, 15-26, 2006.
- [18] Teague S.J., *Implications of protein flexibility for drug discovery*. Nature rev. drug Discov. 2, 527-541, 2003.
- [19] Warren G., Andrews C., Capelli A-M., Clarke B., LaLonde J., Lambert M.H., Lindvall M., Nevins N., Semus S.F., Senger S., Tedesco G., Wall I.D., Woolven J.M., Peishoff C.E., Head M.S., *A Critical assessment of docking programs and scoring functions*. J. Med. Chem. 49, 5912-5931, 2006.

Références du chapitre 1

- [20] Atkins P.W., *Eléments de chimie physique*. de Boeck Université, chap. 12, 1998.
- [21] Barteaux P.J., *Nouveaux éléments de la science de l'homme*. Réed. augmentée de l'ouvrage de 1778, Baillière, Paris, 1806.
- [22] Bergson H., *L'évolution créatrice*. Presses Universitaires de France, 1907 ; 86ème Ed., 2007.
- [23] Berman H., Henrick K., Nakamura H., Markley J.L.. *The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data*. Nucleic Acids Research 35, 301-303, 2007.
- [24] Branden C., Tooze J., *Introduction à la structure des protéines*. de Boeck Université, chap. 17, 1996.
- [25] Cavarelli J., *Cristallographie des macromolécules biologiques*. Techniques de l'ingénieur, P1090, 2000.
- [26] Committee on Biomolecular Materials and Processes. National Research Council of the National Academies. *Inspired by biology: from molecules to materials to machines*. Nat. Ac. Press, 2008. http://books.nap.edu/openbook.php?rec ord_id=12159&page=11
- [27] Committee on Frontiers at the Interface of Computing and Biology, N.R.C., *Catalyzing Inquiry at the Interface of Computing and Biology*. Nat. Ac. Press, 2005, p. 24.

-
- [28] Descartes R., *Lettre à Morus (extrait)* http://www.caute.lautre.net/article.php3?id_article=1431,1649.
 - [29] Dobson C.M., Sali A., Karplus M., *Protein folding: a perspective from theory and experiment*. Angew. Chem. Int. Ed. 37, 868-893, 1998.
 - [30] Dupuy J-P., Roure F., *Les nanotechnologies : éthique et prospective industrielle*. C.G. des Mines, C.G. des technologies de l'information, sect. "innovation et entreprise", 2004. <http://www.cgm.org/themes/deveco/develop/nanofinal.pdf>
 - [31] Dutta S., Berman H.M., *Large macromolecular complexes in the Protein Data Bank: a status report*. Structure 13, 381-388, 2005.
 - [32] Feynman R.P., Leighton R.B., Sands M., *The Feynman lectures on physics Vol. 1*. Addison-Wesley, 1963.
 - [33] Garrett R.H., Grisham C.M., *Biochimie*. de Boeck Université, chap. 1, 2000.
 - [34] Gilbas C., Jambeck P., Dauchel H., *Introduction à la bioinformatique*. O'Reilly France, chap. 9, 2002.
 - [35] Grognet J-M., *La nanobiotechnologie, ou l'apport des nanotechnologies aux sciences du vivant*. Découverte 312, 42-50, 2003.
 - [36] Grognet J-M., *Nanobiotechnologies : présent, promesses et enjeux*. MURS, 47, 38-48, 2006.
 - [37] Jacob F., *La logique du vivant*. Gallimard, 1970.
 - [38] Jeannin Y., *Résolution d'une structure cristalline par rayons X*. Techniques de l'ingénieur, P1075, 1998.
 - [39] Karp G., Bouharmont J., Wissocq J-C., *Biologie cellulaire et moléculaire, 2ème ed.* de Boeck Université, chap. 2, 2007.
 - [40] Kay L.E., *NMR methods for the study of protein structure and dynamics*. Biochem. Cell Biol. 75, 1-15, 1997.
 - [41] Khushf G., *The ethics of nanotechnology. Vision and values for a new generation of science and engineering*. Nat. Ac. Eng. (Emerging technologies and ethical issues in engineering), 2004.
 - [42] Laurent L., *Technologies convergentes et questions nouvelles*. Clefs CEA 52, été 2005.
 - [43] Leger J-F., *L'ADN : une flexibilité structurale adaptée aux interactions avec les autres macromolécules de son environnement*. Thèse de Doctorat, Université Louis Pasteur, Strasbourg, 1999.
 - [44] Lewin B., *Gene VI*. de Boeck Université, chap. 1, 1998.
 - [45] Lodish H., Berk A., Matsudaira P., Kaiser C.A., Krieger M., Scott M.P., Zipursky S.L., Darnell J., *Biologie moléculaire de la cellule*. de Boeck Université, chap. 1, 2005.
 - [46] Malliavin T., Dardel F., *Structure des protéines par RMN*. Techn. Ing., AF6608, 2002.
 - [47] McCammon J.A., Harvey S.C., *Dynamics of proteins and nucleic acids*. Cambridge University Press, 1987.
 - [48] Mnyusiwalla A., Daar A.S., Singer P., *"Mind the gap": science and ethics in nanotechnology*. Nanotechnology 14, R9-R13, 2003.

- [49] Monod J., *Le hasard et la nécessité : essai sur la philosophie naturelle de la biologie moderne*. Seuil, 1970.
- [50] Protein Data Bank. <http://www.rcsb.org/pdb/home/home.do>
- [51] Radaev S., Li S., Sun P.D., *A survey of protein-protein complex crystallizations*. Acta Cryst. D62, 605-612, 2006.
- [52] Ronsin C., *L'histoire de la biologie moléculaire : héros et pionniers*. de Boeck Université, 2005.
- [53] Sali A., Glaeser R., Earnest T., Baumeister W., *From words to literature in structural proteomics*. Nature 422, 216-225, 2003.
- [54] Synthetic Biology. <http://syntheticbiology.org/>
- [55] Voet D., Voet J.G., *Biochimie (2th Ed.)*. de Boeck Université, chap. 11, 2005.
- [56] Voet D., Voet J.G., *Biochimie (2th Ed.)*. de Boeck Université, chap. 12, 2005.

Références du chapitre 2

- [57] Abagyan R., Totrov M., *High-throughput docking for lead generation*. Curr. Op. Chem. Biol. 5, 375-382, 2001.
- [58] Alder B.J., Wainwright T.E., *Phase transition for a hard sphere system*. J. Chem. Phys. 27, 1208, 1957.
- [59] Allinger N.L., *MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms*. J. Am. Chem. Soc. 99, 8127, 1977.
- [60] Allinger N.L., Yuh Y.H., Lii J-H., *The MM3 force field for hydrocarbons*. J. Am. Chem. Soc. 111, 8522, 1989.
- [61] Allinger N.L., Chen K., Lii J-H., *An improved force field (MM4) for saturated hydrocarbons*. J. Comp. Chem. 17, 642, 1996.
- [62] Aloy P., Russell R.B., *Structural systems biology: modelling protein interactions*. Nat. Rev. Mol. Cell Biol. 7, 188-197, 2006.
- [63] Alvarez J.C., *High-throughput docking as a source of novel drug leads*. Curr. Op. Chem. Biol. 8, 365-370, 2004.
- [64] Amara P., Field M., *Combined quantum mechanics and molecular mechanics potentials*. Enc. Comp. Chem. 1, Ed. P.v.R. Schleyer, Wiley & sons, 431-437, 1998.
- [65] AMBER. <http://www.amber.ucsf.edu/>
- [66] Andrews D.H., *The relation between the Raman spectra and the structure of organic molecules*. Phys. Rev. 36, 544-554, 1930.
- [67] Apaydin M.S., Guestrin C., Varma V., Brutlag D.L., Latombe J.C., *Studying protein-ligand interactions with stochastic roadmap simulation*. Bioinformatics 18, 18-26, 2002.
- [68] Apaydin M.S., Brutlag D.L., Guestrin C., Hsu D., Latombe J.C., Varma C., *Stochastic Roadmap Simulation: an efficient representation and algorithm for analyzing molecular motion*. J. Comp. Biol. 10, 257-281, 2003.

-
- [69] Bajorath J., *Integration of virtual and high-throughput screening*. Nature Rev. Drug Discov. 1, 882-894, 2002.
- [70] Bahar I., Rader A.J., *Coarse-grained normal mode analysis in structural biology*. Curr. Op. Struct. Biol. 15, 586-592, 2005.
- [71] Bao G., Suresh S., *Cell and molecular mechanics of biological materials*. Nature materials 2, 715-725, 2003.
- [72] Barrett C.P., Noble M.E.M., *Dynamite extended: two new services to simplify protein dynamic analysis*. Bioinformatics 21, 3174-3175, 2005.
- [73] Bashford D., Case D., *Generalized Born models of macromolecular solvation effects*. Annu. Rev. Phys. Chem. 51, 129-152, 2000.
- [74] Bastard K., Thureau A., Lavery R., Prévost C., *Docking macromolecules with flexible segments*. J. Comp. Chem. 24, 1910-1920, 2003.
- [75] Bastard K., Prévost C., Zacharias M., *Accounting for loop flexibility during protein-protein docking*. Proteins 62, 956-969, 2006.
- [76] Baxter C.A., Murray C.W., Clark D.E., Westhead D.R., Eldridge M.D., *Flexible docking using tabu search and an empirical estimate of binding affinity*. Proteins 33, 367-382, 1998.
- [77] Becke A.D., *Density-functional exchange-energy approximation with correct asymptotic behavior*. Phys. Rev. A 38, 3098, 1988.
- [78] Böhm H-J., *LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads*. J. Comput. Aided Mol. Des. 6, 593-606, 1992.
- [79] Bonvin A.M.J.J., *Flexible protein-protein docking*. Curr. Opin. Struct. Biol. 16, 194-200, 2006.
- [80] Born M., Oppenheimer R., *Zur quantentheorie der molekeln*. Ann. Phys. 84, 457-484, 1927.
- [81] Bosshard H.R., *Molecular recognition by induced fit : how fit is the concept ?*. News Physiol. Sci. 16, 171-173, 2001.
- [82] Boucekkine G., *Méthodes de la chimie quantique*. Techn. Ing., AF6050, 2007.
- [83] Brooijmans N., Kuntz I.D., *Molecular recognition and docking algorithms*. Annu. Rev. Biophys. Biomol. Struct. 32, 335-373, 2003.
- [84] Burkert U., Allinger N.L., *Molecular mechanics*. Am. Chem. Soc. monograph, Washington, DC, 1982.
- [85] Burnett R.M., Taylor J.S., *DARWIN: a program for docking flexible molecules*. Proteins 41, 173-191, 2000.
- [86] CACTVS. <http://www2.ccc.uni-relangen.de/software/cactvs/>
- [87] Camacho C.J., Kimura S.R., DeLisi C., Vajda S., *Kinetics of desolvation-mediated protein-protein binding*. Biophys. J. 78, 1094-1105, 2000.
- [88] Camacho C.J., Vajda S., *Protein-protein association kinetics and protein docking*. Curr. Opin. Struct. Biol. 12, 36-40, 2002.

- [89] Cavasotto C.N., Orry A.J.M., Abagyan R.A., *The challenge of considering receptor flexibility in ligand docking and virtual screening*. Curr. Comput. Aided Drug Des. 1, 423-440, 2005.
- [90] Charifson P.S., Corkery J.J., Murckov, Walters W.P., *Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins*. J. Med. Chem. 42, 5100-5109, 1999.
- [91] CHARMM. <http://yuri.harvard.edu/>
- [92] Chem3D. <http://www.camsoft.com/>
- [93] Clark R.D., Strizhev A., Leonard J.M., Blake J.F., Matthew J.B., *Consensus scoring for ligand/protein interactions*. J. Mol. Graph. Modell. 20, 281-295, 2002.
- [94] Committee on Frontiers at the Interface of Computing and Biology, N.R.C., *Catalyzing Inquiry at the Interface of Computing and Biology*. Nat. Ac. Press, 2005, p. 31.
- [95] Condon E.U., *Wave mechanics and the normal state of the hydrogen molecule*. Proc. Nat. Ac. Sc. U.S.A 13, 466-470, 1927.
- [96] Connolly M.L., *Solvent-accessible surfaces of proteins and nucleic acids*. Science 221, 709-713, 1983.
- [97] Connolly M.L., *Analytical surface molecular calculation*. J. Appl. Cryst. 16, 548-558, 1983.
- [98] Cortés J., Siméon T., Ruiz de Angulo V., Guieysse D., Remaud-Siméon M., Tran V., *A path planning approach for computing large-amplitude motions of flexible molecules*. Bioinformatics 21, 116-125, 2005.
- [99] Deak J., Chiu H.L., Lewis C.M., Miller R.J.D., *Ultrafast phase grating studies of heme proteins: observation of the low-frequency modes directing functionally important protein motions*. J. Phys. Chem. 102, 6621-6634, 1998.
- [100] Delarue M., Dumas P., *On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models*. PNAS 101, 6957-6962, 2004.
- [101] Deremble C., Lavery R., *Macromolecular recognition*. Curr. Opin. Struct. Biol. 15, 171-175, 2005.
- [102] Desjarlais R.L., Sheridan R.P., Dixon J.S., Kuntz I.D., Venkataraghavan R., *Docking flexible ligands to macromolecular receptors by molecular shape*. J. Med. Chem. 29, 2149-2153, 1986.
- [103] Desjarlais R.L., Sheridan R.P., Seibel G.L., Dixon J.S., Kuntz I.D., Venkataraghavan R., *Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensionnal structure*. J. Med. Chem. 31, 722-729, 1988.
- [104] Desjarlais R.L., Seibel G.L., Kuntz I.D., Furth P.S., Alvaraz J.C., Ortiz de Montellano P.R., Decamp D.L., Babe L.M., Craik C.S., *Structure-based design of non peptide inhibitors specific for human immunodeficiency virus 1 protease*. Proc. Natl. Acad. Sci. USA 87, 6644-6648, 1990.
- [105] Dhanik A., Yao P., Marz N., Propper R., Kou C., Liu G., van den Bedem H., Latombe J.C., *Efficient algorithms to explore conformation spaces of flexible protein loops*. WABI 2007, Philadelphia, 265-276.

-
- [106] Dirac P.A.M., *Quantum mechanics of many-electron systems*. Proc. Roy. Soc. Lond. Series A, 123, 714-733, 1929.
- [107] Dixon J.S., Oshiro C.M., *Flexible ligand docking using a genetic algorithm*. J. Comput. Aided Mol. Des. 9, 113-130, 1995.
- [108] Dodziuk H., *Modern conformational analysis. Elucidating novel exciting molecular structures*. Wiley-VCH, chap. 3, 1995.
- [109] Dominguez C., Boelens R., Bonvin A., *HADDOCK: a protein-protein docking approach based on biochemical or biophysical data*. J. Am. Chem. Soc. 125, 1731-1737, 2003.
- [110] Dominguez C., Bonvin A., Winkler G., van Schaik F., Timmers H., Boelens R., *Structural model of the UbcH5B/CNOT4 complex revealed by combining NMR, mutagenesis, and docking approaches*. Structure 12, 633-644, 2004.
- [111] Ehrlich L.P., Nilges M., Wade R.C., *The impact of protein flexibility on protein-protein docking*. Proteins 58, 126-133, 2005.
- [112] Elcock A.H., *Modeling supramolecular assemblages*. Curr. Opin. Struct. Biol. 12, 154-160, 2002.
- [113] Eldridge M.D., Murray C.W., Auton T.R., Paolini G.V., Mee R.P., *Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes*. J. Comput. Aided Mol. Des. 11, 425-445, 1997.
- [114] Ewing T.J.A., Makino S., Skillman A.G., Kuntz I.D., *DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases*. J. Comp. Aided Mol. Des. 15, 411-428, 2001.
- [115] Frisch C., Fersht A.R., Schreiber G., *Experimental assignment of the structure of the transition state for the association of barnase and barstar*. J. Mol. Biol. 308, 69-77, 2001.
- [116] Gao J., *Methods and applications of combined quantum mechanical and molecular mechanical potentials*. Lipkowitz K.B, Boyd D.B., Eds., ; Rev. Comp. Chem. 7, VCH, 119-185, 1996.
- [117] Gardiner E.J., Willett P., Artymiuk P.J., *Protein docking using a genetic algorithm*. Proteins 44, 44-56, 2001.
- [118] Ghomi M., Victor J.-M., Henriët C., *Monte Carlo simulations on short single-stranded oligonucleotides. I. Application to RNA trimers*. J. Comp. Chem. 15, 433-445, 1994.
- [119] Gohlke H., Hendlich M., Klebe G., *Knowledge-based scoring function to predict protein-ligand interactions*. J. Mol. Biol. 295, 337-356, 2000.
- [120] Goodford P.J., *A computational procedure for determining energetically favorable binding sites on biologically important macromolecules*. J. Med. Chem. 28, 849-857, 1985.
- [121] Grassberger P., *The Monte Carlo approach to biopolymers and protein folding*. Hardback, 1998.
- [122] Gray J.J., Moughon S., Wang C., Schueler-Furman O., Kuhlman B., Rohl C.A., Baker D., *Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations*. J. Mol. Biol. 331, 281-299, 2003.
- [123] Greer J., Bush B.L., *Macromolecular shape and surface maps by solvent exclusion*. Proc. Natl. Acad. Sci. USA 75, 303-307, 1978.

- [124] Grünberg R., Leckner J., Nilges M., *Complementarity of structure ensembles in protein-protein binding*. Structure 12, 2125-2136, 2004.
- [125] Fermi E., *Un metodo statistico per la determinazione di alcune proprietà dell'atomo*. Atti. Accad. Naz. Lincei 6, 602-607, 1927.
- [126] Fermi E., Nuovo Cimento 11, 157, 1934.
- [127] Fernandez-Recio J., Totrov M., Abagyan R., *ICM-DISCO docking by global energy optimization with fully flexible side-chains*. Proteins 52, 113-117, 2003.
- [128] Fischer E., *Einfluss der configuration auf die Wirkung der Enzyme*. Ber. Dtsch. Chem. Ges. 27, 2985-2993, 1894.
- [129] Floquet N., Marechal J., Badet-Denisot M., Robert C., Dauchez M., Perahia D., *Normal mode analysis as a prerequisite for drug design: Application to matrix metalloproteinase inhibitors*. FEBS Letters 580, 5130-5136, 2006.
- [130] Frenkel D., Smit B., *Understanding molecular simulation: from algorithms to applications*. Academic Press, London, 2nd ed., 2002.
- [131] Ghosh A., Rapp C.S., Friesner R.A., *A generalized Born model based on a surface integral formulation*. J. Phys. Chem. B 102, 10983-10990, 1998.
- [132] GOLD. <http://www.ccdc.cam.ac.uk/>
- [133] GROMACS. <http://www.gromacs.edu/>
- [134] GROMOS. <http://www.igc.ethz.ch/gromos>
- [135] Gunsteren W.F.V., Luque F.J., Timms D., Torda A.E., *Molecular mechanics in biology: from structure to function, taking account of solvation*. Annu. Rev. Biophys. Biomol. Struct. 23, 847-863, 1994.
- [136] Halperin I., Ma B., Wolfson H., Nussinov R., *Principles of docking: an overview of search algorithms and a guide to scoring functions*. Proteins 47, 409-443, 2002.
- [137] Hansmann U.H.E., Okam Y., *New Monte Carlo algorithms for protein folding*. Curr. Op. Struct. Biol. 9, 177-183, 1999.
- [138] Hartree D.R., *The wave mechanics of an atom with a non-coulomb central field. Part I. II. III.* Proc. Camb. Phil. Soc. 24, 89-110, 111-132, 426-437, 1928.
- [139] Heifetz A., Eisenstein M., *Effect of local shape modifications of molecular surfaces on rigid-body protein-protein docking*. Prot. Engin. 16, 179-185, 2003.
- [140] Heitler W., London F., *Wechselwirkung neutraler Atome und homöpolare Bindung nach der Quantenmechanik*. Zeitschrift für Physik A Hadrons and Nuclei, 44, 455-472, 1927.
- [141] Hinsen K., *Analysis of domain motions by approximate normal mode calculations*. Proteins 33, 417-429, 1998.
- [142] Hoffmann R., *An extended Hückel theory. I. Hydrocarbons*. J. Chem. Phys. 39, 1397-1412, 1963.
- [143] Hoffmann D., Kramer B., Washio T., Steinmetzer T., Rarey M., Lengauer T., *Two-stage method for protein-ligand docking*. J. Med. Chem. 42, 4422-4433, 1999.
- [144] Hohenberg P., Kohn W., *Inhomogeneous electron gas*. Phys. Rev. 136, 864-871, 1964.

-
- [145] Höltje H-D., Sippl W., Folkers G., Rognan D., *Molecular modeling: basic principles and applications*. Wiley-VCH, chap. 2, 2003.
- [146] Huang S-Y., Zou X., *Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking*. *Proteins: Structure, Function, and Bioinformatics* 66, 399-421, 2007.
- [147] Huo S., Wang J., Cieplak P., Kollman P.A., Kuntz I.D., *Molecular dynamics and free energy analyses of cathepsin D-inhibitor interactions: insight into structure-based ligand design*. *J. Med. Chem.* 45, 1412-1419, 2002.
- [148] INSIGHT, QUANTA. <http://www.accelrys.com/>
- [149] Jagodzinski F., Brock O., *Towards a Mechanistic View of Protein Motion*. Proceedings of the 46th IEEE Conference on Decision and Control, New Orleans, LA, USA 2007.
- [150] Jiang F., Kim S.H., "Soft docking": matching of molecular surface cubes. *J. Mol. Biol.* 219, 79-102, 1991.
- [151] Jones G., Willet P., Glen R.C., Leach A.R., Taylor R., *Development and validation of a genetic algorithm for flexible docking*. *J. Mol. Biol.* 267, 727-748, 1997.
- [152] Jorgensen W.L., Chandrasekhar J., Madura J., Klein M.L., *Comparison of simple potential functions for simulating liquid water*. *J. Chem. Phys.* 79, 926-935, 1983.
- [153] Jorgensen W.L., *Rusting of the lock and key model for protein-ligand binding*. *Science* 254, 954-955, 1991.
- [154] Jorgensen W.L., *The many roles of computation in drug discovery*. *Science* 303, 1813-1818, 2004.
- [155] Karplus M.J., *CHARMM: a program for macromolecular energy, minimization, and dynamics calculations*. *Comput. Chem.* 4, 187, 1983.
- [156] Karplus M., Evanseck J.D., Joseph D., Bash P.A., Field M.J., *Simulation analysis of triose phosphate isomerase: conformational transition and catalysis*. *Faraday Discuss.* 93, 239-248, 1992.
- [157] Karplus M., McCammon M.A., *Molecular dynamics simulations of biomolecules*. *Nature structural biology* 9, 646-652, 2002.
- [158] Kauzmann W., *Some factors in the interpretation of protein denaturation*. *Adv. Protein Chem.* 14, 1-63, 1959.
- [159] Kearsley S.K., Underwood D.J., Sheridan R.P., Miller M.D., *Flexibases: a way to enhance the use of molecular docking methods*. *J. Comput. Aided Mol. Des.* 8, 565-582, 1994.
- [160] Keserü G.M., Kolossváry I., *Molecular mechanics and conformational analysis in drug design*. Blackwell Publishing, 1999.
- [161] Khon W., Sham L.J., *Self-consistent equations including exchange and correlation effects*. *Phys. Rev.* 140, 1133-1138, 1965.
- [162] Kitchen D.B., Decornez H., Furr J.R., Bajorath J., *Docking and scoring in virtual screening for drug discovery: methods and applications*. *Nature Rev. Drug Discov.* 3, 935-950, 2004.

- [163] Kollman P.A., *Free energy calculations: applications to chemical and biochemical phenomena*. Chem. Rev. 93, 2395-2417, 1993.
- [164] Koshland D.E. Jr., *Application of a theory of enzyme specificity to protein synthesis*. Proc. Natl. Acad. Sci. USA 44, 98-104, 1958.
- [165] Kramer B., Rarey M., Lengauer T., *Evaluation of the FlexX incremental construction algorithm for protein-ligand docking*. Proteins 37, 228-241, 1999.
- [166] Kumar S., Ma B., Tsai C.J., Sinha N., Nussinov R., *Folding and binding cascades: dynamic landscapes and population shifts*. Protein Sci. 9, 10-19, 2000.
- [167] Langer T., Hoffmann R.D., *Virtual screening: an effective tool for lead structure discovery*. Curr. Pharm. Design 7, 509-527, 2001.
- [168] Laurencelle L., *Hasard, nombres aléatoires et méthode Monte Carlo*. PUQ, 2001.
- [169] Leach A.R., *Molecular modeling. Principles and applications*. Prentice Hall, Harlow, 2nd ed., 2001.
- [170] Leder L., Berger C., Bornhauser S., Wendt H., Ackermann F., Jelesarov I., Bosshard H.R., *Spectroscopic, calorimetric and kinetic demonstration of conformational adaptation in peptide-antibody recognition*. Biochem. 34, 16509-16518, 1995.
- [171] Lee C., Yand W., Parr R.G., *Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density*. Phys. Rev. B 37, 785, 1988.
- [172] Legrand S., Merz K., *The application of genetic algorithm to conformational search*. FASEB 6, 132-132, 1992.
- [173] Lei M., Zavodszky M.I., Kuhn L.A., Thorpe M.F., *Sampling protein conformations and pathways*. J. Comput. Chem. 25, 1133-1148, 2004.
- [174] Levinthal C., Wodak S.J., Kahn P., Dadivanian A.K., *Hemoglobin intercation in sickle cell fibers. I. Theoretical approaches to the molecular contacts*. Proc. Natl. Acad. Sci. USA 72, 1330-1334, 1975.
- [175] Lewars E., *Computational chemistry: Introduction to the theory and applications of molecular and quantum mechanics* Springer, chap.5, 2003.
- [176] Liu H., Elstner M., Kaxiras E., Frauenheim T., Hermans J., Yang W., *Quantum mechanics simulation of protein dynamics on long timescale*. Proteins: Structure, Function, and Genetics 44, 484-489, 2001.
- [177] Lotan I., Schwarzer F., Halperin D., Latombe J.C., *Algorithm and data structures for efficient energy maintenance during monte carlo simulation of proteins*. J. Comp. Biol. 11, 902-932, 2004.
- [178] Ma B., Elkayam T., Wolfson H., Nussinov R., *Protein-protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces*. PNAS 100, 5772-5777, 2003.
- [179] Ma J., *New advances in normal mode analysis of supermolecular complexes and applications to structural refinement*. Curr. Protein Pept. Sci. 5, 119-123, 2004.
- [180] Ma J., *Usefulness and limitations of Normal Mode Analysis in modeling dynamics of biomolecular complexes*. Structure 13, 373-380, 2005.

-
- [181] Mangoni M., Roccatano D., Di Nola A., *Docking of flexible ligands to flexible receptors in solution by molecular dynamics simulation*. Proteins 35, 153-162, 1999.
- [182] McCammon J.A., *Target flexibility in molecular recognition*. Bioch. Biophys. Acta 1754, 221-224, 2005.
- [183] Malliavin T., *Simulations de dynamique moléculaire en biochimie*. Techn. Ing., AF6043, 2003.
- [184] Marchildon L., *Mécanique quantique*. de Boeck université, chap. 9, 2000.
- [185] Mason D.W., Williams A.F., *Handbook of Experimental Immunology (4ème ed.)*. edited by Weir D.M., Herzenberg L.A., Blackwell C., Herzenberg L.A. Oxford: Blackwell Scientific, 38.1-38.17, 1986.
- [186] May A., Zacharias M., *Accounting for global protein deformability during protein-protein and protein-ligand docking*. Bioch. Biophys. Acta 1754, 225-231, 2005.
- [187] McCammon J.A., Gelin B.R., Karplus M., *Dynamics of folded proteins*. Nature 267, 585-590, 1977.
- [188] Metropolis N., Ulam S., *The Monte Carlo Method*. J. Am. Stat. Assoc. 44, 335-341, 1949.
- [189] Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., Teller E., *Equation of state calculations by fast computing machines*. J. Chem. Phys. 21, 1087, 1953.
- [190] Monard G., Merz K.M. Jr., *Combined quantum mechanics and molecular mechanics methodologies applied to biomolecular systems*. Acc. Chem. Res. 32, 904-911, 1999.
- [191] Morris G.M., Goodsell D.S., Halliday R.S., Huey R., Hart W.E., *Automated docking using a Lamarckian genetic emphasis on quantitative analysis of protein-ligand algorithm and an empirical free energy function*. J. Comput. Chem. 19, 1639-1662, 1998.
- [192] Morse P.M., *Diatomic molecules according to the wave mechanics. II. Vibrational levels*. Phys. Rev. 34, 57-64, 1929.
- [193] Muegge I., *A knowledge-based scoring function for protein-ligand interactions: probing the reference state*. Perspect. Drug Discov. Des. 20, 99-114, 2000.
- [194] Muegge I., *Effect of ligand volume correction on PMF scoring*. J. Comput. Chem. 22, 418-425, 2001.
- [195] Muegge I., Martin Y.C., *A general and fast scoring function for protein-ligand interactions: a simplified potential approach*. J. Med. Chem. 42, 791-804, 1999.
- [196] Nagasawa M., *Schrödinger equations and diffusion theory*. Birkhäuser, 1993.
- [197] NAMD. <http://www.ks.uiuc.edu/Research/namd/>
- [198] Nicholls A., Bharadwaj R., Honig B., *Grasp - Graphical Representation and Analysis of Surface-Properties*. Biophys. J. 64, A166-A166, 1993.
- [199] Norel R., Lin S.L., Wolfson H.J., Nussinov R., *Shape complementarity at protein-protein interfaces*. Biopolymers 34, 933-940, 1994.
- [200] Norel R., Petrey D., Wolfson H.J., Nussinov R., *Examination of shape complementarity in docking of unbound proteins*. Proteins 36, 307-317, 1999.

- [201] O'Hearn S.D., Kusalik A.J., Angel J.F., *MolCom: a method to compare protein molecules based on 3-D structural and chemical similarity*. Protein Eng. 16,169-178, 2003.
- [202] Olson A.J., Goodsell D.S., *Automated docking in crystallography: analysis of the substrates of aconitase*. Proteins 17, 1-10, 1993.
- [203] ORAC. <http://www.chim.unifi.it/orac/index.html>
- [204] Palma P.N., Krippah L., Wampler J.E., Moura J.J.G., *BIGGER: A new (soft) docking algorithm for predicting protein interactions*. Proteins 39, 372-384, 2000.
- [205] Pauli W., *Über den zusammenhang des abschlusses der elektronengruppen im atom mit der komplexstruktur der spektren*. Z. Phys. 31, 765-785, 1925.
- [206] Pauling L., Delbrück M., *The nature of the intermolecular forces operative in biological processes*. Science 92, 77-79, 1940.
- [207] Perdew J.P., Wang Y., *Accurate and simple density functional for the electronic exchange energy: generalized gradient approximation*. Phys. Rev. B, 33, 8800-8802, 1986.
- [208] Perdew J.P., Burke K., Wang Y., *Generalized gradient approximation for the exchange-correlation hole of a many-electron system*. Phys. Rev. B, 54, 16533-16539, 1996.
- [209] PyMOL. <http://pymol.sourceforge.net/>
- [210] Rahman A., *Correlations in the motion of atoms in liquid argon*. Phys. Rev. 136, 405-411, 1964.
- [211] Rajamani D., Thiel S., Vajda S., Camacho C.J., *Anchor residues in protein-protein interactions*. PNAS 101, 11287-11292, 2004.
- [212] Rapaport D.C., *The art of molecular dynamics simulation*. Cambridge University Press, 2nd ed., 2004.
- [213] Rarey M., Kramer B., Lengauer T., Klebe G., *A fast flexible docking method using an incremental construction algorithm*. J. Mol. Biol. 261, 470-489, 1996.
- [214] RASMOL. <http://www.umass.edu/microbio/rasmol/>
- [215] Read R.J., Hart T.N., *A multiple-start Monte Carlo docking method*. Proteins 13, 206-222, 1992.
- [216] Richards F.M., *Areas, volumes, packing, and protein structure*. Annu. rev. Biophys. Bioeng. 6, 151-176, 1977.
- [217] Richie D.W., Kemp G.J., *Protein docking using spherical polar Fourier correlations*. Proteins 39, 178-194, 2000.
- [218] Rognan D., Lauemoller S.L., Holm A., Buus S., Tschinke V., *Predicting binding affinities of protein ligands from three-dimensional models: application to peptide binding to class I major histocompatibility proteins*. J. Med. Chem. 42, 4650-4658, 1999.
- [219] Rubino G., Tuffin B., *Simulations et méthodes de Monte Carlo*. Techn. Ing. AF600, 2007.
- [220] Ruiz-López M.F., Rivail J.L., *Combined quantum mechanics and molecular mechanics approaches to chemical and biochemical reactivity*. Enc. Comp. Chem. 1, Ed. P.v.R. Schleyer, Wiley & sons, 437-448, 1998.

-
- [221] Sandak B., Wolfson H.J., Nussinov R., *Flexible docking allowing induced fit in proteins: insights from an open to closed conformational isomers*. Proteins: Structure, Function, and Genetics 32, 159-174, 1998.
- [222] Santiso E.E., Gubbins K.E., *Multi-scale molecular modeling of chemical reactivity*. Molecular Simulation 30, 699-748, 2004.
- [223] Sherman W., Day T., Jacobson M.P., Friesner R.A., Farid R., *Novel procedure for modeling ligand/receptor induced fit effects*. J. Med. Chem. 49, 534-553, 2006.
- [224] Schneidman-Duhovny D., Inbar Y., Nussinov R., Wolfson H.J., *Geometry-based flexible and symmetric protein docking*. Proteins: Struct., Funct., Bioinform. 60, 224-231, 2005.
- [225] Schrödinger E., *Mémoires sur la mécanique ondulatoire*. Félix-Alcan, Paris, 1933. Rééd. Jacques Gabay, 1988. (a)E. Schrödinger, Ann. Phys. 1926, 79, 361. (b)E. Schrödinger, Ann. Phys. 1926, 79, 489.
- [226] Simonson T., Archontis G., Karplus M., *Free energy simulations come of age: protein-ligand recognition*. Acc. Chem. Res. 35, 430-437, 2002.
- [227] Singh U.C., Kollman P.A., *An approach to computing electrostatic charges for molecules*. J. Comp. Chem. 5, 129-145, 1984.
- [228] Sitkoff D., Sharp K.A., Honig B., *Accurate calculation of hydration free energies using macroscopic continuum models*. J. Phys. Chem. 98, 1978-1983, 1998.
- [229] Slater J.C., *The theory of complex spectra*. Phys. Rev. 34, 1293-1322, 1929.
- [230] Smith G.R., Sternberg M.J., *Prediction of protein-protein interactions by docking methods*. Curr. Opin. Struct. Biol. 12, 28-35, 2002.
- [231] Smith G.R., Sternberg M.J., Bates P.A., *The relationship between the flexibility of proteins and their conformational states on forming protein-protein complexes with an application to protein-protein docking*. J. Mol. Biol. 347, 1077-1101, 2005.
- [232] Stahl M., Rarey M., *Detailed analysis of scoring functions for virtual screening*. J. Med. Chem. 44, 1035-1042, 2001.
- [233] Still W.C., Tempczyk A., Hawley R.C., Hendrickson T., *Semianalytical treatment of solvation for molecular mechanics and dynamics*. J. Am. Chem. Soc. 112, 6127-6129, 1990.
- [234] Suhre K., Sanejouand Y-H., *ElNémo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement*. Nucleic Acids Res. 32, 610-614, 2004.
- [235] Sybyl. <http://www.tripos.com/> Tripos Inc., St. Louis, USA.
- [236] Tama F., Sanejouand Y-H., *Conformational change of proteins arising from normal mode calculations*. Prot. Eng. 14, 1-6, 2001.
- [237] Thomas L.H., *The calculation of atomic fields*. Proc. Cambridge Philos. Soc, 23, 542, 1927.
- [238] TINKER. <http://dasher.wustl.edu/tinker/>
- [239] Totrov M., Abagyan R., *Flexible protein-ligand docking by global energy optimization in internal coordinates*. Proteins Suppl. 1, 215-220, 1997.

- [240] Toulhoat H., *Modélisation moléculaire. Bases théoriques (partie 3)*. Tecnh. Ing. J1013, 2007.
- [241] Verdonk M.L., Cole J.C., Hartshorn M.J., Murray C.W., Taylor R.D., *Improved protein-ligand docking using GOLD*. Proteins 52, 609-623, 2003.
- [242] Verlet L., *Computer "experiments" on classical fluids. I Thermodynamical properties of Lennard-Jones molecules*. Phys. Rev. 159, 98-103, 1967.
- [243] Vieth M., Hirst J.D., Dominy B.N., Daigler H., Brooks C.L., *Assessing search strategies for flexible docking*. J. Comp. Chem. 19, 1623-1631, 1998.
- [244] VMD. <http://www.ks.uiuc.edu/Research/vmd/>
- [245] Walters W.P., Stahl M.T., Murcko M.A., *Virtual screening - an overview*. Drug Discov. Today 3, 160-178, 1998.
- [246] Wang R., Lai L., Wang S., *Further development and validation of empirical scoring functions for structure-based binding affinity prediction*. J. Comput. Aided Mol. Des. 16, 11-26, 2002.
- [247] Wang C., Schueler-Furman O., Baker D., *Improved side-chain modeling for protein-protein docking*. Protein Sci. 14, 1328-1339, 2005.
- [248] Weiner P.K., Kollman P.A., *AMBER: assisted model building with energy refinement. A general program for modeling molecules and their interactions*. J. Comput Chem. 2, 287, 1981.
- [249] Welch W., Ruppert J., Jain A.N., *Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites*. Chem. & Biol. 3, 449-462, 1996.
- [250] Westhead D.R., Clark D.E., Murray C.W., *A comparison of heuristic search algorithms for molecular docking*. J. Comput. Aided Mol. Des. 11, 209-228, 1997.
- [251] Wodak S.J., Janin J., *Computer analysis of protein-protein interactions*. J. Mol. Biol. 124, 323-342, 1978.
- [252] Wolfson H.J., Shatsky M., Schneidman-Duhovny D., Dror O., Shulman-Peleg A., Ma B., Nussinov R., *From structure to function: methods and applications*. Curr. Prot. Pept. Sc. 6, 171-183, 2005.
- [253] XmMol. <http://condor.urbb.jussieu.fr/logiciels/>
- [254] Yang J-M., Chen Y-F., Shen T-W., Kristal B.S., Hsu D.F., *Consensus scoring criteria for improving enrichment in virtual screening*. J. Chem. Inf. Model. 45, 1134-1146, 2005.
- [255] Zacharias M., *Protein-protein docking with a reduced protein model accounting for side-chain flexibility*. Protein Sci. 12, 1271-1282, 2003.
- [256] Zacharias M., *Rapid protein-ligand docking using soft modes from molecular dynamics simulations to account for protein deformability: binding of FK506 to FKBP*. Proteins 54, 759-767, 2004.
- [257] Zavodszky M.I., Lei M., Thorpe M.F., Day A.R., Kuhn L.A., *Modeling correlated main-chain motions in proteins for flexible molecular recognition*. Proteins 57, 243-261, 2004.

Références du chapitre 3

- [258] Abergel C., Chenivresse S., Byrne D., Suhre K., Arondel V., Claverie J.M., *Mimivirus tyrrs: preliminary structural and functional characterization of the first amino-acyl tRNA synthetase found in a virus*. Acta Cryst. F 61, 212-215, 2005.
- [259] Ashcroft N.W., Mermin N.D., *Physique des solides*. EDP Sciences, chap. 23, 2002.
- [260] Atilgan A.R., Durell S.R., Jernigan R.L., Demirel M.C., Keskin O., Bahar I., *Anisotropy of fluctuation dynamics of proteins with an elastic network model*. Biophys. J. 80, 505-515, 2001.
- [261] Bahar I., Atilgan A.R., Erman B., *Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential*. Fold. Des. 2, 173-181, 1997.
- [262] Balog E., Becker T., Oetl M., Lechner R., Daniel R., Finney J., Smith J.C., *Direct determination of vibrational density of states change on ligand binding to a protein*. Phys. Rev. Lett. 93, 028103, 2004.
- [263] Beuron F., Flynn T.C., Ma J., Kondo H., Zhang X., Freemont P.S., *Motions and negative cooperativity between p97 domains revealed by cryoelectron microscopy and quantized elastic deformational model*. J. Mol. Biol. 327, 619-629, 2003.
- [264] Brooks B., Karplus M., *Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme*. Proc. Natl. Acad. Sci. USA 82, 4995-4999, 1985.
- [265] Brooks C.L., Karplus M., Pettitt B.M., *Proteins: a theoretical perspective of dynamics, structure, and thermodynamics*. Adv. Chem. Phys. 71, 1-249, 1988.
- [266] Brut M., Estève A., Landa G., Renvez G., Djafari Rouhani M., *The Static Modes: an alternative approach for the treatment of macro- and bio-molecular induced-fit flexibility*. Eur. Phys. J. E 28, 17-25, 2009.
- [267] Cui Q., Li G., Ma J., Karplus M., *A normal mode analysis of structural plasticity in the biomolecular motor F(1)-ATPase*. J. Mol. Biol. 340, 345-372, 2004.
- [268] Cui Q., Bahar I., *Normal mode analysis theoretical and applications to biological and chemical systems*. Chapman and Hall/CRC (Mathematical and Computational Biology Series).
- [269] Cusack S., Smith J., Finney J., Tidor B., Karplus M., *Inelastic neutron scattering analysis of picosecond internal protein dynamics. Comparison of harmonic theory with experiment*. J. Mol. Biol. 202, 903-908, 1988.
- [270] Delarue M., Sanejouand Y.H., *Simplified normal modes analysis of conformational transitions in DNA-dependent polymerases: the Elastic Network Model*. J Mol. Biol. 320, 1001-1024, 2002.
- [271] Doruker P. , Jernigan R.L., Bahar I., *Dynamic of large proteins through hierarchical levels of coarse-grained structures*. J. Comput. Chem. 23, 119-127, 2002.
- [272] Frenkel D., Smit B., *Understanding molecular simulation: from algorithms to applications*. Academic Press, London, 2nd ed., appendiceA, 2002.

- [273] Gerschel A., *Liaisons intermoléculaires. Les forces en jeu dans la matière condensée.* EDP Sciences Ed., chap.2, 1995.
- [274] Goh C.S., Milburn D., Gerstein M., *Conformational changes associated with protein-protein interactions.* Curr. Opin. Struct. Biol. 14, 104-109, 2004.
- [275] Hayward S., Kitao A., Go N., *Harmonicity and anharmonicity in protein dynamics: a normal mode analysis and principal component analysis.* Proteins 23, 177-189, 1995.
- [276] Hinsén K., Petrescu A.J., Dellerue S., Bellissent-Funel M.C., Kneller G.R., *Harmonicity in slow protein dynamics.* Chem. Phys. 261, 25-37, 2000.
- [277] Hinsén K., Reuter N., Navaza J., Stokes D.L., Lacapere J.J., *Normal mode-base fitting of atomic structure into electron density maps: application to sarcoplasmic reticulum Ca-ATPase.* Biophys. J. 88, 818-827, 2005.
- [278] Kern D., Zuiderweg E.R., *The role of dynamics in allosteric regulation.* Curr. Opin. Struct. Biol. 13, 748-757, 2003.
- [279] Keskin O., Bahar I., Flatow D., Covell D.G., Jernigan R.L., *Molecular mechanisms of chaperonin GroEL-GroES function.* Biochemistry 41, 491-501, 2002.
- [280] Kirillova S., Cortés J., Stefaniu A., Siméon T., *An NMA-guided path planning approach for computing large-amplitude conformational changes in proteins.* Proteins 70, 131-143, 2008.
- [281] Kong Y., Ming D., Wu Y., Stoops J.K., Zhou Z.H., Ma J., *Conformational flexibility of pyruvate dehydrogenase complexes: a computational analysis by quantized elastic deformational model.* J. Mol. Biol. 330, pp. 129-135, 2003.
- [282] Kottalam J., Case D.A., *Langevin modes of macromolecules: applications to crambin and DNA hexamers.* Biopolymers 29, 1409-1421, 1990.
- [283] Krebs W.G., Alexandrov V., Wilson C.A., Echols N., Yu H., Gerstein M., *Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic.* Proteins 48, 682-695, 2002.
- [284] Li G., Cui Q., *A coarse-grained normal mode approach for macromolecules: an efficient implementation and application to Ca(2+)-ATPase.* Biophys. J. 83, 2457-2474, 2002.
- [285] Ma J., Karplus M., *Ligand-induced conformational changes in ras p21: a normal mode and energy minimization analysis.* J. Mol. Biol. 274, 114-131, 1997.
- [286] Ma J., Karplus M., *The allosteric mechanism of the chaperonin GroEL: a dynamic analysis.* Proc. Natl. Acad. Sci. USA 95, 8502-8507, 1998.
- [287] Ming D., Kong Y., Lambert M., Huang Z., Ma J., *How to describe protein motion without amino-acid sequence and atomic coordinates.* Proc. Natl. Acad. Sci. USA 99, 8620-8625, 2002.
- [288] Ming D., Kong Y., Wu Y., Ma J., *Substructure synthesis method for simulating large molecular complexes.* Proc. Natl. Acad. Sci. USA 100, 104-109, 2003.
- [289] Miyashita O., Onuchic J.N., Wolynes P.G., *Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins.* Proc. Natl. Acad. Sci. USA 100, 12570-12575, 2003.

-
- [290] Rappaz M., Bellet M., Deville M., *Modélisation numérique en science et génie des matériaux*. PPUR presses polytechniques, 1998.
- [291] Reat V., Dunn R., Ferrand M., Finney J.L., Daniel R.M., Smith J.C., *Solvent dependence of dynamic transitions in protein solutions*. Proc. Natl. Acad. Sci. USA 97 , 9961-9966, 2000.
- [292] Roh J.H., Novikov V.N., Gregory R.B., Curtis J.E., Chowdhuri Z., Sokolov A.P., *Onsets of anharmonicity in protein dynamics*. Phys. Rev. Lett. 95, 038101, 2005.
- [293] Seno Y., Go N., *Deoxymyoglobin studied by the conformational normal mode analysis II. The conformational change upon oxygenation*. J. Mol. Biol. 216, 111-126, 1990.
- [294] Seno Y., Go N., *Deoxymyoglobin studied by the conformational normal mode analysis. I. Dynamics of globin and the heme-globin interaction*. J. Mol. Biol. 216, 95-109, 1990.
- [295] Smith J.C., *Protein dynamics: comparison of simulations with inelastic neutron scattering experiments*. Q. Rev. Biophys. 24, 227-291, 1991.
- [296] Stewart J., *Analyse : fonctions de plusieurs variables*. De Boeck Université, chap. 11, 2005.
- [297] Suhre K., Sanejouand Y-H., *On the potential of normal mode analysis for solving difficult molecular replacement problems*. Act. Cryst. D 60, 796-799, 2004.
- [298] Tama F., Wriggers W., Brooks C.L., *Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory*. J. Mol. Biol. 321, 297-305, 2002.
- [299] Tama F., Valle M., Frank J., Brooks C.L., *III, Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy*. Proc. Natl. Acad. Sci. USA 100, 9319-9323, 2003.
- [300] Tama F., Miyashita O., Brooks C.L., *Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis*. J. Mol. Biol. 337, 985-999, 2004.
- [301] Thomas A., Field M.J., Mouawad L., Perahia D., *Analysis of the low frequency normal modes of the T-state of aspartate transcarbamylase*. J. Mol. Biol. 257, 1070-1087, 1996.
- [302] Thomas A., Field M.J., Perahia D., *Analysis of the low-frequency normal modes of the R state of aspartate transcarbamylase and a comparison with the T state modes*. J. Mol. Biol. 261, 490-506, 1996.
- [303] Thomas A., Hinsen K., Field M.J., Perahia D., *Tertiary and quaternary conformational changes in aspartate transcarbamylase: a normal mode study*. Proteins 34, 96-112, 1999.
- [304] Tirion M.M., *Large amplitude elastic motions in proteins from a single-parameter, atomic analysis*. Phys. Rev. Lett. 77, 1905-1908, 1996.
- [305] Valadie H., Lacapere J.J., Sanejouand Y-H., Etchebest C., *Dynamical properties of the MscL of Escherichia Coli: a normal mode analysis*. J Mol. Biol. 332, 657-674, 2003.
- [306] Wang Y., Rader A.J., Bahar I., Jernigan R.L., *Global ribosome motions revealed with elastic network model*. J. Struct. Biol. 147, 302-314, 2004.
- [307] Xu C., Tobi D., Bahar I., *Allosteric changes in protein structure computed by a simple mechanical model: hemoglobin T-R2 transition*. J. Mol. Biol. 333, 153-168, 2003.

Références du chapitre 4

- [308] Altona C., Sundaralingam M., *Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation*. J. Am. Chem. Soc. 94, 8205-8212, 1972.
- [309] Arora K., Schlick T., *Deoxyadenosine sugar puckering pathway simulated by the stochastic difference equation algorithm*. Chem. Phys. Let. 378, 1-8, 2003.
- [310] Barbe S., Le Bret M., *Ab initio determination of the flexibility of 2'-aminoribonucleosides and 2'-aminoarabinonucleosides inserted in duplexes*. J. Comput. Chem. 29, 1353-1363, 2008.
- [311] Bartenev V.N., Kamenewa N.G., Lipanov A.A., *A statistical stereochemical model of the flexible furanose ring*. Acta Cryst. B43, 275-280, 1987.
- [312] Becke A.D., *Density-functional thermochemistry. III. The role of exact exchange*. J. Chem. Phys. 98, 5648, 1993.
- [313] Bell R.J., Dean P., Hibbins-Butler D.C., *Localization of normal modes in vitreous silica, germania and beryllium fluoride*. J. Phys. C 3, 2111-2118, 1970.
- [314] Bolli M., Lubini P., Leumann C., *Nucleic-acid analogs with restricted conformational flexibility in the sugar-phosphate backbone ("Bicyclo-DNA"). Part 5. Synthesis, characterization, and pairing properties of oligo- α -D-(bicyclodeoxynucleotides) of the bases adenine and thymine (α -Bicyclo-DNA)*. Helv. Chim. Acta 78, 2077-2096, 1995.
- [315] Brik A., Wong C-H., *HIV-1 protease: mechanism and drug discovery*. Org. Biomol. Chem. 1, 5-14, 2003.
- [316] Cornell W.D., Cieplak P., Bayly C.I., Gould I.R., Merz K.M. Jr., Ferguson D.M., Spellmeyer D.C., Fox T., Caldwell J.W., Kollman P.A., *A second generation force field for the simulation of proteins, nucleic acids, and organic molecules*. J. Am. Chem. Soc. 117, 5179-5197, 1995.
- [317] Cram D.J., *The benign of molecular hosts, guests, and their cineplexes*. Angew. Chem. Int. Ed. Engl. 27, 1009-1020, 1988.
- [318] Davies D.R., *The structure and function of aspartic proteinases*. Annu. Rev. Biophys. Chem. 19, 189-215, 1990.
- [319] Dupouy C., Le Clézio I., Lavedan P., Gornitzka H., Escudier J-M., Vigroux A., *Diastereoselective synthesis of conformationally restricted dinucleotides featuring canonical and non canonical α/β torsion angle combinations (α,β -D-CNA)*. Eur. J. Org. Chem. 12, 1935-1941, 2006.
- [320] Dupouy C., Iché-Tarrat N., Durrieu M-P., Rodriguez F., Escudier J-M., Vigroux A., *Watson-Crick base-pairing properties of nucleic acid analogues with stereocontrolled and torsion angles (α,β -D-CNAs)*. Angew. Chem. 118, 3705-3709, 2006.
- [321] Dupouy C., Iché-Tarrat N., Durrieu M-P., Vigroux A., Escudier J-M., *α,β -D-CNA induced rigidity within oligonucleotides*. Org. Biomol. Chem. 6, 2849-2851, 2008.
- [322] Erickson-Viitanen S., Manfredi J., Viitanen P., Tribe D.E., Tritch R., Hutchison C.A., *Cleavage of HIV-1 gag polyprotein synthesized in vitro: sequential cleavage by the viral protease*. AIDS Res. Hum. Retroviruses 5, 577-591, 1989.

-
- [323] Estève A., Bail A., Landa G., Dkhissi A., Brut M., Djafari Rouhani M., Sudor J., Gue A-M., *A new insight into the understanding of the collapsed form of poly(N-isopropylacrylamide) molecules*. Chem. Phys. 340, 12-16, 2007.
- [324] Frisch M.J., Trucks G.W., Schlegel H.B., Scuseria G.E., Robb M.A., Cheeseman J.R., Montgomery J.A., Vreven T., Kudin K.N., Burant J.C., Millam J.M., Iyengar S.S., Tomasi J., Barone V., Mennucci B., Cossi M., Scalmani G., Rega N., Petersson G.A., Nakatsuji H., Hada M., Ehara M., Toyota K., Fukuda R., Hasegawa J., Ishida M., Nakajima T., Honda Y., Kitao O., Nakai H., Klene M., Li X., Knox J.E., Hratchian H.P., Cross J.B., Bakken V., Adamo C., Jaramillo J., Gomperts R., Stratmann R.E., Yazyev O., Austin A.J., Cammi R., Pomelli C., Ochterski J.W., Ayala P.Y., Morokuma K., Voth G.A., Salvador P., Dannenberg J.J., Zakrzewski V.G., Dapprich S., Daniels A.D., Strain M.C., Farkas O., Malick D.K., Rabuck A.D., Raghavachari K., Foresman J.B., Ortiz J.V., Cui Q., Baboul A.G., Clifford S., Cioslowski J., Stefanov B.B., Liu G., Liashenko A., Piskorz P., Komaromi I., Martin R.L., Fox D.J., Keith T., Al-Laham M.A., Peng C.Y., Nanayakkara A., Challacombe M., Gill P.M.W., Johnson B., Chen W., Wong M.W., Gonzalez C., Pople J.A., *Gaussian03 Rev C.01* Gaussian, Inc.: Pittsburgh, 2003.
- [325] Graziano G., *On the temperature-induced coil to globule transition of poly-N-isopropylacrylamide in dilute aqueous solutions*. Int. J. Biol. Macromol. 27, 89-97, 2000.
- [326] Harte W.E., Swaminathan S., Mansuri M.M., Martin J.C., Rosenberg I.E., Beveridge D.L., *Domain communication in the dynamical structure of human immunodeficiency virus 1 protease*. Proc. Nat. Ac. Sc. USA 87, 8864-8868, 1990.
- [327] Heskins M., Guillet J.E., *Solution properties of poly(N-isopropylacrylamide)*. J. Macromol. Sc. Chem. A2, 1441-1455, 1968.
- [328] Hornak V., Okur A., Rizzo R.C., Simmerling C., *HIV-1 protease flaps spontaneously open and reclose in molecular dynamics*. PNAS 103, 915-920, 2006.
- [329] Huber D.L., Manginell R.P., Samara M.A., Kim B-I., Bunker B.C., *Programmed adsorption and release of proteins in a microfluidic device*. Science 301, 352-354, 2003.
- [330] Huff R.H., *Perspective HIV protease: a novel chemotherapeutic target for AIDS*. J. Med. Chem. 34, 2305-2314, 1991.
- [331] Ishima R., Freedberg D.I., Wang Y-X., Louis J.M., Torchia D.A., *Flap opening and dimer-interface flexibility in the free and inhibitor-bound HIV protease, and their implications for function*. Structure 7, 1047-1055, 1999.
- [332] IUPAC Definitive rules for nomenclature of organic chemistry, *Section A. Hydrocarbons. Section B. Fundamental heterocyclic systems*. J. Am. Chem. Soc. 82, 5545-5574, 1960.
- [333] IUPAC-IUB Commission on Biochemical Nomenclature (CNB), *Abbreviations and symbols for nucleic acids, polynucleotides and their constituents*. Eur. J. Biochem. 15, 203-208, 1970.
- [334] IUPAC-IUB Commission on Biochemical Nomenclature (CNB), *Abbreviations and symbols for description of conformation of polypeptide chains*. Pure Appl. Chem. 40, 291-308, 1974.
- [335] IUPAC-IUB Joint commission on biochemical nomenclature, *Abbreviations and symbols for the description of conformations of polynucleotide chains*. Eur. J. Biochem. 131, 9-15, 1983.

- [336] Katoh I., Yasunaga T., Ikawa Y., Yoshinaka Y., *Inhibition of retroviral protease activity by an aspartyl proteinase inhibitor*. Nature 329, 654-656, 1987.
- [337] Katsumoto Y., Tanaka T., Sato H., Ozaki Y., *Conformational change of Poly(N-isopropylacrylamide) during the coil-globule transition investigated by attenuated total reflection/infrared spectroscopy and density functional theory calculation*. J. Phys. Chem. A, 106, 3429-3435, 2002.
- [338] Kent A., Sepkowitz M.C., *AIDS - The first 20 years*. N. Engl. J. Med. 344, 1764-1772, 2001.
- [339] Kohl N.E., Emini E.A., Schleif W.A., Davis L.J., Heimbach J.C., Dixon R.A.F., Scolnick E.M., Sigal I.S., *Active human immunodeficiency virus protease is required for viral infectivity*. Proc. Natl. Acad. Sci. USA 85, 4686-4690, 1988.
- [340] Kool E.T., *Preorganization of DNA: design principles for improving nucleic acid recognition by synthetic oligonucleotides*. Chem. Rev. 97, 1473-1487, 1997.
- [341] Lam P.Y., Jadhav P.K., Eyermann C.J., Hodge C.N., Ru Y., Bacheler L.T., Meek J.L., Otto M.J., Rayner M.M., Wong Y.N., Chong-Hwan C., Weber P.C., Jackson D.A., Sharpe T.R., Erickson-Vitanen S., *Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors*. Science 263, 380-384, 1994.
- [342] Lay D.C., *Algèbre linéaire : théorie, exercices & applications*. de Boeck Université, chap. 5, 2007.
- [343] Lebon F., Ledecq M., *Approaches to the design of effective HIV-1 protease inhibitors*. Curr. Med. Chem. 7, 455-477, 2000.
- [344] de Leeuw H.P.M., Haasnoot C.A.G., Altona C., *Empirical correlations between conformational parameters in β -D-furanoside fragments derived from a statistical survey of crystal structures of nucleic acid constituents: full description of nucleoside molecular geometries in terms of four parameters*. Isr. J. Chem. 20, 108-126, 1980.
- [345] Leumann C.J., *DNA analogues: from supramolecular principles to biological properties*. Bioorg. Med. Chem. 10, 841-854, 2002.
- [346] Levitt M., Warshel A., *Extreme conformational flexibility of the furanose ring in DNA and RNA*. A. J. Am. Chem. Soc. 100, 2607-2613, 1978.
- [347] Lopez G., Chilkoti A., Atanassov P., Goparaju V., *Stimuli responsive hybrid materials containing molecular actuators and their applications*. US patent 6491061B1, 2003.
- [348] Mager P.P., *The active site of HIV-1 protease*. Med. Res. Rev. 21, 348-353, 2002.
- [349] Meints G.A., Karlsson T., Drobny G.P., *Modeling furanose ring dynamics in DNA*. J. Am. Chem. Soc. 123, 10030-10038, 2001.
- [350] Mias S., Sudor J., Camon H., *PNIPAM: a thermo-activated nano-material for use in optical devices*. Microsyst. Technol. 14, 691-695, 2008.
- [351] Miller M., Schneider J., Sathyanarayana B.K., Toth M.V., Marshall G.R., Clawson L., Selk L., Kent S.B., Wlodawer A., *Structure of complex of synthetic HIV-1 protease and substrate-based inhibitor at 2.3 Å resolution*. Science 246, 1149-1152, 1989.

-
- [352] Navia M.A., Fitzgerald P.M.D., McKeever B.M., Leu C., Heimbach J.C., Herber W.K., Sigal I.S., Darke P.L., Springer J.P., *3-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1*. Nature 337, 615-620, 1989.
- [353] Nicholson L.K., Yamazaki T., Torchia D.A., Grzesiek S., Bax A., Stahl S.J., Kaufman J.D., Wingfield P.T., Lam P.Y.S., Jadhav P.K., Hodge C.N., Dommaille P.J., Chang C-H., *Flexibility and function in HIV-1 protease*. Nature Struct Biol 2, 274-280, 1995.
- [354] Oroszlan S., Luftig R.B., *Retroviral proteinases*. Curr. Top. Microbiol. Immunol. 157, 153-185, 1990.
- [355] Pear L.H., Taylor W.R., *A structural model for the retroviral proteases*. Nature 329, 351-354, 1987.
- [356] Perryman A.L., Lin J-H., McCammon J.A., *HIV-1 protease molecular dynamics of a wild-type and of the V82F/I84V mutant: possible contributions to drug resistance and a potential new target site for drugs*. Protein Science 13, 1108-1123, 2004.
- [357] Perryman A.L., Lin J-H., McCammon J.A., *Restrained molecular dynamics simulations of HIV-1 protease: the first step in validating a new target for drug design*. Biopolymers 82, 272-284, 2006.
- [358] Rao S., Westhof E., Sundaralingam M., *Exact method for the calculation of pseudorotation parameters P , τ_m and their errors. A comparison of the Altona-Sundaralingam and Cremer-Pople treatment of puckering of five-membered rings*. Acta Cryst. A37, 421-425, 1981.
- [359] Rose R.B., Craik C.S., Stroud R.M., *Domain flexibility in retroviral proteases: structural implications for drug resistant mutations*. Biochemistry 37, 2607-2621, 1998.
- [360] Saenger W., *Principles of nucleic acid structure*. Springer-Verlag, 1984.
- [361] Schild H.G., *Poly (N-isopropylacrylamide): experiment, theory and application*. Prog. Polym. Sc. 17, 163-249, 1992.
- [362] Short G.F., Laikhter A.L., Lodder M., Shayo Y., Arslan Y., Hecht S.M., *Probing the S1/S1' substrate binding pocket geometry of HIV-1 protease with cyclic HIV protease inhibitors: synthesis, conformational analysis, P2/P2'*. Biochemistry 39, 8768-8781, 2000.
- [363] Snowden M.J., Murray M.J., Chowdry B.Z., *Some like it hot! Thermo-sensitive polymers*. Chemistry and industry 14, 531-534, 1996.
- [364] Suguna J.K., Padlan E.A., Smith C.X., Carlson W.D., Davies D.R., *Binding of a reduced peptide inhibitor to the aspartic proteinase from Rhizopus chinensis: implications for a mechanism of action*. Proc. Nat. Ac. Sc. USA 84, 7009-7013, 1987.
- [365] Sun G., Voigt J.H., Filippov I.V., Marquez V.E., Nicklaus M.C., *PROSIT: Pseudo-Rotational Online Service and Interactive Tool, applied to a conformational survey of nucleosides and nucleotides*. J. Chem. Inf. Comput. Sci. 44, 1752-1762, 2004.
- [366] Twaites B., de las Heras Alarcón C., Alexander C., *Synthetic polymers as drugs and therapeutics*. J. Mater. Chem. 15, 441-455, 2005.
- [367] Visscher W., *Localization of electron wave functions in disordered systems*. J. Non Cryst. Solids, 8-10, 477-484, 1972.

- [368] Voet D., Voet J.G., *Biochimie (2th Ed.)*. de Boeck Université, chap. 8, 2005.
- [369] Vondrasek J., Wlodaver A., *HIVdb: a database of the structures of human immunodeficiency virus protease*. *Proteins: Struct. Funct. Genet.* 49, 429-431, 2002.
- [370] Wengel J., Petersen M., Nielsen K.E., Jensen G.A., Hakansson A.E., Kumar R., Sørensen M.D., Rajwanshi V.K., Bryld T., Jacobsen J.P., *LNA (locked nucleic acid) and the diastereoisomeric alpha-L-LNA: conformational tuning and high-affinity recognition of DNA/RNA targets*. *Nucleosides, Nucleotides and Nucleic Acids* 20, 389-396, 2001.
- [371] West M.L., Fairlie D.P., *Targeting HIV-1 protease: a test of drug-design methodologies*. *TiPS* 16, 67-75, 1995.
- [372] Wlodaver A., Miller M., Jaskolski M., Sathyanarayana B.K., Baldwin E., Weber I.T., Selk L.M., Clawson L., Schneider J., Kent S.B.H., *Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease*. *Science* 245, 616-621, 1989.
- [373] Wlodawer A., Vondrasek J., *Inhibitors of HIV-1 protease: a major success of structure-assisted drug design*. *Ann. Rev. Biophys. Biomol. Struct.* 27, 249-284, 1998.

Références de l'annexe A

- [374] Murray R.K., Granner D.K., Mayes P.A., Rodwell V.W., *Biochimie de Harper (2th Ed.)*. de Boeck Université, chap. 6, 2003.
- [375] Pauling L., Corey R.B., Branson H.R., *The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain*. *Proc. Nat. Acad. Sci.*, 37, 205-211, 1951.
- [376] Pauling L., Corey R.B., *The pleated sheet, a new layer configuration of polypeptide chains*. *Proc. Nat. Acad. Sci.*, 37, 251-256, 1951.

Références de l'annexe B

- [377] Altmann R., *Ueber nucleinsäuren*. *Arch. f. Anatomie u. Physiol*, 524-536, 1889.
- [378] Avery O.T., MacLeod C. M., McCarty M., *Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a deoxyribonucleic acid fraction isolated from pneumococcus type III*. *J. Exp. Med.* 149, 297-326, 1979 (Reprint of 1944 paper).
- [379] Chargaff E., *Chemical specificity of nucleic acids and mechanism of their enzymatic degradation*. *Experientia* 6, 201, 1950.
- [380] Dickerson R.E., *The anatomy of A-, B-, and Z-DNA*. *Science* 216, 475-485, 1982.
- [381] Garrett R.H., Grisham C.M., *Biochimie*. de Boeck Université, chap. 12, 2000.
- [382] Hershey A.D., Chase M., *Independent functions of viral protein and nucleic acid in growth of bacteriophage*. *J. Gen. Physiol.* 36, 39-56, 1952.
- [383] Levene A.T., *Darstellung und analyse einiger nucleinsäuren*. *Z. physiol. Chem.* 39, 4-8, 1903.

-
- [384] Mendel G., *Versuche uber Pflanzen-Hybriden*. Verhandlungen des naturforschenden Vereins in Brünn, 4, 3-47, 1866. *Recherches sur les hybrides végétaux, trad. en français par A. Chappellier*. Bulletin Scientifique de la France et de la Belgique, 41, 371-419, 1907.
- [385] Miescher F., *Ueber die chemische Zusammensetzung der Eiterzellen*. Med. Chem. Unters. 4, 441-460, 1871.
- [386] Watson J.D., Crick F., *A structure for deoxyribose nucleic acid*. Nature 171, 737, 1953.
- [387] Watson J.D., Baker T.A., Bell S.P., Gann A., Levine, Losick R., *Molecular biology of the gene (5th Ed.)*. Pearson, Benjamin Cummings, CSHL Press, 2004.
- [388] Wilkins M.H.F., Stokes A.R., Wilson H. R., *Molecular Structure of Deoxypentose Nucleic Acids*. Nature 171, 738-740, 1953.

Nouvelle approche méthodologique pour la prise en compte de la flexibilité dans les interactions entre molécules biologiques : Les Modes Statiques

Résumé

La connaissance des propriétés structurales et physico-chimiques des biomolécules passe par une compréhension fine des interactions intra et inter-moléculaires. L'abondance des données et les enjeux socio-économiques nécessitent la mise en place de moyens d'analyse à haut débit, tels que la recherche *in silico*. Les algorithmes développés demeurent cependant peu satisfaisants pour le traitement de la flexibilité conformationnelle des macromolécules.

Nous proposons dans cette thèse une nouvelle méthode, les Modes Statiques, afin d'obtenir un algorithme prenant en compte la flexibilité totale du système. Cette approche, fondée sur le concept de déformations induites entre deux macromolécules en interaction, vise à déterminer tous les modes de déformation possibles d'une molécule. Chaque déformation, appelé Mode Statique, résulte d'une excitation extérieure sur un atome de la biomolécule. Ces modes sont calculés en fonction des constantes de force contenues dans le modèle énergétique. Ils sont ensuite stockés pour des utilisations futures, du docking en particulier. Le problème du docking se réduit alors à des interactions entre sites, les déformations moléculaires émanant des Modes Statiques pré-calculés. Ce travail a donné lieu au développement d'un code, Flexible, dont nous présentons les premières applications aux propriétés de diverses molécules uniques : des polymères thermosensibles aux enzymes allostériques, en passant par les acides nucléiques.

Mots-clés : biomolécules, interactions moléculaires, flexibilité conformationnelle, modélisation à l'échelle atomique

New methodological approach for the treatment of flexibility in the interactions between biological molecules: The Static Modes

Abstract

Knowledge of structural and physical-chemical properties of biomolecules requires a precise understanding of intra and inter-molecular interactions. Abundance of data and socio-economic issues require the development of powerful analytical tools, such as *in silico* research. Developed algorithms remain unsatisfactory, particularly for the treatment of conformational flexibility of macromolecules.

We propose in this thesis a new method, the Static Modes, to obtain an algorithm taking into account the flexibility of the entire system. This approach, based on the "induced-fit" concept, aims to identify all possible modes of deformation of a molecule. Each deformation, called a Static Mode, results from an external excitation on one atom of the biomolecule. These modes are calculated according to force constants in the energy model. They are then stored for future uses, docking in particular. The problem of docking is then reduced to interactions between sites, molecular deformations arising from pre-calculated Static Modes. This work has led to the development of a code: Flexible. We present here the first applications to the properties of various unique molecules : thermosensitive polymer, allosteric enzymes, nucleic acids.

Keywords: biomolecules, molecular interactions, conformational flexibility, atomic scale modeling

THESE
Préparée en vue de l'obtention du
DOCTORAT DE L'UNIVERSITE DE TOULOUSE
Délivré par l'Université Toulouse III - Paul Sabatier

AUTEUR : Marie BRUT

TITRE : Nouvelle approche méthodologique pour la prise en compte de la flexibilité dans les interactions entre molécules biologiques : Les Modes Statiques

DIRECTEUR DE THESE : Mehdi DJAFARI ROUHANI

CO-DIRECTEUR : Alain ESTEVE

LIEU ET DATE DE SOUTENANCE :

Laboratoire d'Analyse et d'Architecture des Systèmes - CNRS, le 5 mars 2009

RESUME :

La connaissance des propriétés structurales et physico-chimiques des biomolécules passe par une compréhension fine des interactions intra et inter-moléculaires. L'abondance des données à traiter et l'importance des enjeux socio-économiques nécessitent actuellement la mise en place de moyens d'analyse à haut débit, tels que la recherche *in silico*. Les algorithmes développés jusque-là demeurent cependant peu satisfaisants, particulièrement pour le traitement de la flexibilité conformationnelle des macromolécules.

Nous proposons dans cette thèse une nouvelle méthode, les Modes Statiques, afin d'obtenir un algorithme prenant en compte la flexibilité totale du système. Cette approche, fondée sur le concept de déformations induites entre deux macromolécules en interaction, vise à déterminer tous les modes de déformation possibles d'une molécule. Chaque déformation, appelé Mode Statique, résulte d'une excitation extérieure sur un atome de la biomolécule. Ces modes sont calculés en fonction des constantes de force contenues dans le modèle énergétique. Ils sont ensuite stockés pour des utilisations futures, du docking en particulier. Le problème du docking se réduit alors à des interaction entre sites, les déformations moléculaires émanant des Modes Statiques pré-calculés.

Ce travail a donné lieu au développement d'un code, Flexible, dont nous présentons les premières applications aux propriétés de diverses molécules uniques : des polymères thermosensibles aux enzymes allostériques, en passant par les acides nucléiques.

MOTS CLEFS :

biomolécules, interactions moléculaires, flexibilité conformationnelle, modélisation à l'échelle atomique

DISCIPLINE ADMINISTRATIVE :

Sciences de la Matière

INTITULE ET ADRESSE DU LABORATOIRE :

Laboratoire d'Analyse et d'Architecture des Systèmes - CNRS

7, avenue du Colonel Roche, 31077 TOULOUSE Cedex 4 - FRANCE