



**HAL**  
open science

# Un Modèle de Solvatation Semi-Implicite pour la Simulation des Macromolécules Biologiques

Nathalie Basdevant

► **To cite this version:**

Nathalie Basdevant. Un Modèle de Solvatation Semi-Implicite pour la Simulation des Macromolécules Biologiques. Biophysique [physics.bio-ph]. Université d'Evry-Val d'Essonne, 2003. Français. NNT : . tel-00010619

**HAL Id: tel-00010619**

**<https://theses.hal.science/tel-00010619>**

Submitted on 13 Oct 2005

**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.

THÈSE de DOCTORAT de l'UNIVERSITÉ d'ÉVRY-VAL-d'ESSONNE

Spécialité

BIOPHYSIQUE MOLÉCULAIRE

présentée par **Nathalie BASDEVANT (née CAPITAINÉ)**

pour obtenir le titre de

**DOCTEUR de l'UNIVERSITÉ d'ÉVRY-VAL-d'ESSONNE**

## **Un Modèle de Solvation Semi-Implicite pour la Simulation des Macromolécules Biologiques**

soutenue le 14 octobre 2003 devant le jury composé de :

M.	Daniel BORGIS	co-directeur
Mme	Catherine ETCHEBEST	rapporteur
M.	Tap HA DUONG	directeur
M.	Richard LAVERY	président du jury
M.	Thomas SIMONSON	rapporteur
M.	Flavio TOMA	examinateur



*Le travail présenté dans cette thèse a été réalisé au Laboratoire de Modélisation des Systèmes Moléculaires Complexes de l'Université d'Évry-Val-d'Essonne. Je remercie très sincèrement Daniel Borgis, directeur du laboratoire, de m'avoir accueillie, et je lui exprime toute ma gratitude pour avoir initié ce travail et m'avoir suivie tout au long de cette thèse. J'ai pu apprécier à tout moment ses précieux conseils éclairés et ses explications scientifiques toujours précises, ainsi que sa motivation très communicative.*

*Je tiens à remercier vivement Tap Ha Duong, pour m'avoir encadrée pendant ces trois années avec un enthousiasme sans faille. Il m'a fait découvrir le monde de la biophysique moléculaire et de la modélisation et m'en a enseigné les rudiments. Je lui adresse mes plus sincères remerciements pour la confiance qu'il m'a accordée, pour avoir su motiver et valoriser mon travail de recherche, et pour avoir toujours fait preuve d'une grande disponibilité à mon égard.*

*J'exprime ma reconnaissance à Madame Catherine Etchebest et Monsieur Thomas Simonson d'avoir accepté d'être les rapporteurs de cette thèse. Je remercie Monsieur Richard Lavery qui me fait l'honneur de présider ce jury. J'adresse aussi mes remerciements à Monsieur Flavio Toma pour avoir accepté de juger cette thèse et pour avoir porté de l'intérêt à mes travaux.*

*Je tiens à remercier toute l'équipe du Laboratoire de Modélisation des Systèmes Moléculaires Complexes pour son soutien, et en particulier Nicolas Lévy et Cyril Azuara pour leur gentillesse et leur bonne humeur. Je remercie aussi particulièrement Marie-Pierre Gageot pour ses conseils et sa patience pour la relecture de ce mémoire, ainsi que Lionel Gendre pour ses conseils en  $\text{\LaTeX}$ . Je n'oublie pas de remercier Thierry Cartailier, du Laboratoire Analyse et Environnement, pour m'avoir aidée dans les derniers moments de crise.*

*Je souhaiterais également remercier l'ensemble du Département de Physique de l'Université pour avoir contribué à une ambiance de travail chaleureuse, pour leurs conseils, et pour les conversations fructueuses que nous avons eues parfois. Je remercie tout particulièrement Frédérique Augougnon, pour sa très grande disponibilité et son aide précieuse lors de tous les problèmes techniques et humains. Merci à Senem, Laurent, Rosa, Giulia, Jean-Philippe, Stéphane, et tous les autres.*

*Enfin, mes pensées vont à ma famille et à mes amis, pour leur soutien sans faille, et pour m'avoir supportée et réconfortée dans les moments difficiles. Je dédie tout particulièrement ce travail à Lionel et à mes parents.*



# Table des matières

<b>Introduction générale</b>	<b>9</b>
<b>1 Présentation générale et nomenclature des macromolécules biologiques</b>	<b>15</b>
1.1 Les protéines . . . . .	18
1.1.1 Structure chimique des protéines : la chaîne polypeptidique . . . . .	18
1.1.2 Le repliement des protéines . . . . .	21
1.2 Les acides nucléiques . . . . .	26
1.2.1 Structure chimique des acides nucléiques . . . . .	26
1.2.2 La structure des ADN . . . . .	30
1.2.3 La structure des ARN . . . . .	34
<b>2 La solvation des macromolécules biologiques</b>	<b>39</b>
2.1 Mise en évidence du rôle de l'eau par les méthodes expérimentales . . . . .	42
2.1.1 Méthodes structurales . . . . .	43
2.1.2 Méthodes de mesures thermodynamiques . . . . .	47
2.1.3 Conclusion sur les méthodes expérimentales . . . . .	52
2.2 Étude de la solvation par les méthodes théoriques . . . . .	53
2.2.1 Méthodes explicites . . . . .	53
2.2.2 Méthodes implicites . . . . .	55
2.2.3 Méthodes hybrides . . . . .	65
2.2.4 Conclusion sur les méthodes théoriques . . . . .	70

<b>3</b>	<b>Méthodologie de l'étude</b>	<b>73</b>
3.1	Un modèle de solvatation basé sur des pseudo-particules polarisables . . . .	76
3.1.1	Principe général du modèle . . . . .	76
3.1.2	Modèle non local . . . . .	78
3.1.3	Étude de la relation entre $\epsilon$ et $\alpha$ pour le modèle non local . . . . .	86
3.1.4	Modèle local . . . . .	95
3.2	Intégration du modèle en dynamique moléculaire . . . . .	99
3.2.1	Le principe de la dynamique moléculaire . . . . .	99
3.2.2	L'intégration du modèle des pseudo-particules polarisables en dynamique moléculaire . . . . .	106
3.2.3	Évaluation du temps de calcul . . . . .	108
3.3	Le paramétrage du modèle local . . . . .	110
3.3.1	Étude d'un ion . . . . .	110
3.3.2	Le potentiel de force moyenne entre deux ions . . . . .	115
3.3.3	L'énergie de solvatation de petites molécules . . . . .	118
3.3.4	Conclusion relative au paramétrage du modèle local . . . . .	124
Annexe A.3	Énergies libres de solvatation des petites molécules avec le solvant PPP . . . . .	126
<b>4</b>	<b>Application aux peptides et protéines</b>	<b>129</b>
4.1	Application aux peptides . . . . .	131
4.1.1	L'octa-alanine . . . . .	132
4.1.2	L'hélice N-terminale du ribonucléase A . . . . .	140
4.1.3	Conclusion sur l'étude des peptides . . . . .	147
4.2	Application aux protéines . . . . .	147
4.2.1	Les protéines étudiées . . . . .	147
4.2.2	Paramètres de simulation et d'analyse . . . . .	150
4.2.3	Résultats . . . . .	154
4.2.4	Conclusion sur l'étude des protéines . . . . .	168

<b>5</b>	<b>Application aux acides nucléiques</b>	<b>169</b>
5.1	Présentation des molécules étudiées et des simulations . . . . .	171
5.1.1	Les molécules étudiées . . . . .	172
5.1.2	Paramètres des simulations . . . . .	176
5.1.3	Paramètres analysés . . . . .	178
5.2	Résultats obtenus sur l'ARN de transfert . . . . .	180
5.2.1	Analyse dynamique et structurale . . . . .	180
5.2.2	Analyse énergétique . . . . .	190
5.2.3	Analyse de l'hydratation . . . . .	191
5.2.4	Conclusion de la simulation de l'ARNt . . . . .	194
5.3	Résultats obtenus sur les double hélices d'ADN . . . . .	194
5.3.1	Analyse structurale . . . . .	194
5.3.2	Analyse énergétique . . . . .	212
5.3.3	Analyse de l'hydratation . . . . .	216
5.3.4	Conclusion sur les simulations des double hélices d'ADN . . . . .	219
	<b>Conclusion et Perspectives</b>	<b>221</b>



# Introduction générale



# Introduction générale

Afin de mieux comprendre le rôle et la fonction biologique des protéines et acides nucléiques, et en particulier les mécanismes de leurs interactions au sein de la cellule, il est nécessaire de résoudre leurs structures tridimensionnelles et d'étudier leur dynamique au niveau atomique. L'année 2003 a été marquée non seulement par l'anniversaire des 50 ans de la découverte de la structure en double hélice de l'ADN par James Watson et Francis Crick [Watson et Crick, 1953], mais aussi par l'achèvement du séquençage du génome humain, que l'on sait maintenant constitué d'environ 30000 gènes. Comme le nombre de protéines identifiées chez l'homme est nettement supérieur à ce chiffre, l'idée que chaque gène code pour une protéine donnée est maintenant dépassée, et on sait qu'un gène peut coder pour plusieurs protéines, et que les gènes entre eux peuvent avoir des influences croisées. Les différentes problématiques qui se posent d'autant plus aujourd'hui, alors que l'on connaît la séquence linéaire des bases du génome de plusieurs espèces, sont de comprendre la fonction des différents gènes, de prédire la structure tridimensionnelle des protéines pour lesquelles ils codent, et d'étudier les interactions et les mécanismes de reconnaissance moléculaire entre les différentes macromolécules biologiques présentes dans la cellule. Ces macromolécules dans la cellule baignent dans un solvant, l'eau, qui, par des phénomènes dits de solvatation, intervient d'un point de vue stérique et électrostatique dans la stabilisation de leurs structures tridimensionnelles et lors de leurs interactions. L'étude de la solvatation des biomolécules apparaît indispensable pour rationaliser l'influence de l'eau sur les processus biochimiques et représente donc un enjeu majeur de la biologie moderne.

Les méthodes expérimentales, telles que la cristallographie par diffraction des rayons X et la résonance magnétique nucléaire (RMN), permettent aujourd'hui d'identifier un certain nombre de molécules d'eau fortement immobilisées dans les premières couches d'hydratation de nombreux acides nucléiques et protéines [Drew et Dickerson, 1981, Kovacs *et al.*, 1997, Makarov *et al.*, 2002]. Mais ces méthodes ne permettent pas de quantifier ni de rationaliser précisément les effets de la solvatation dans les processus de structuration et de complexation. D'autres méthodes expérimentales basées sur des me-

sure thermodynamiques apportent des informations énergétiques sur l'hydratation des édifices macromoléculaires. Elles permettent en particulier de mesurer l'énergie libre de solvatation, grandeur thermodynamique intervenant dans le repliement et la complexation des biomolécules, ainsi que dans la catalyse enzymatique [Schwabe, 1997, Robinson et Sligar, 1998]. Mais les résultats obtenus par ces méthodes thermodynamiques, dont l'interprétation se révèle souvent ambiguë, sont difficiles à relier aux informations structurales de l'hydratation. Ces différentes méthodes expérimentales, si elles sont de plus en plus précises, ne donnent chacune que des informations partielles sur la solvatation et la dynamique des macromolécules biologiques.

Afin de compléter les informations issues des méthodes expérimentales, les méthodes théoriques de modélisation moléculaire présentent un grand intérêt pour l'étude de la structure et de la solvatation des macromolécules biologiques. En particulier, les simulations de dynamique moléculaire, basées sur des principes de physique statistique, permettent de résoudre numériquement les équations du mouvement de tous les atomes d'un système moléculaire donné en solution, et donc d'étudier son évolution dynamique [Leach, 1996, Frenkel et Smit, 1996]. La puissance croissante des ordinateurs permet aujourd'hui de calculer les trajectoires dynamiques de molécules biologiques de taille importante sur des durées atteignant des dizaines de nanosecondes (et même dans quelques rares cas une microseconde), ce qui représente déjà un temps caractéristique de plusieurs phénomènes biologiques. Afin d'étudier précisément la solvatation des biomolécules en modélisation moléculaire, deux principales familles de méthodes de représentation du solvant se distinguent. D'une part, les méthodes explicites, qui représentent chaque molécule d'eau triatomique autour de la macromolécule, permettent d'étudier la trajectoire précise de certaines molécules d'eau [Jorgensen *et al.*, 1983, Berendsen *et al.*, 1987]. Mais ces méthodes sont très coûteuses en temps de calcul et sont peu précises pour calculer les grandeurs thermodynamiques du système à l'équilibre. D'autre part, les méthodes implicites représentent le solvant comme un milieu continu, ayant les caractéristiques macroscopiques moyennes de l'eau, qui entoure un soluté décrit au niveau atomique [Roux et Simonson, 1999]. Ces méthodes, si elles permettent de calculer plus aisément les grandeurs thermodynamiques du système à l'équilibre, sont parfois difficiles à mettre en œuvre en dynamique moléculaire pour les méthodes les plus exactes [Honig et Nicholls, 1995, Marchi *et al.*, 2001], ou sont délicates à paramétrer pour les méthodes plus approximatives [Still *et al.*, 1990, Hassan *et al.*, 2000]. Dans tous les cas, les méthodes implicites permettent rarement de décrire la structure particulière des couches d'hydratation des biomolécules, à part dans le cas de certaines méthodes hybrides assez coûteuses en temps de calcul et encore assez peu développées [Florian et Warshel, 1997, Ramirez *et al.*, 2002].

Ce travail de thèse propose un nouveau modèle de représentation semi-implicite

du solvant en modélisation moléculaire. Notre méthode consiste à représenter le solvant comme un ensemble de particules microscopiques dont les propriétés électrostatiques découlent de lois électrostatiques macroscopiques [Ha-Duong *et al.*, 2002]. Le solvant ainsi obtenu est, à l'équilibre thermodynamique, un ensemble de particules polarisables par le champ électrique créé par la macromolécule. L'intérêt d'une telle méthode est de représenter à la fois l'aspect microscopique du solvant comme les méthodes explicites, et de calculer efficacement les grandeurs thermodynamiques comme les méthodes implicites. De plus, notre modèle se montre efficace numériquement par rapport aux méthodes explicites, ce qui représente un atout non négligeable. Ce nouveau modèle, après un paramétrage le plus simple et réduit possible, peut être implémenté facilement dans un programme de dynamique moléculaire et nous permet donc d'étudier à la fois la dynamique et l'hydratation de plusieurs protéines et acides nucléiques. L'objectif de notre étude sera d'obtenir, pour la molécule étudiée et ses premières couches d'hydratation, des résultats structuraux en accord avec les méthodes expérimentales et les méthodes théoriques de solvation explicites, et des résultats énergétiques en accord avec ceux des méthodes implicites.

Le premier chapitre de ce mémoire rappellera brièvement les principales caractéristiques structurales des protéines et acides nucléiques, ainsi que les nomenclatures qui seront ensuite utilisées dans les chapitres suivants. Le chapitre 2 fait la revue des différentes méthodes existantes, expérimentales et théoriques, pour étudier précisément la solvation des macromolécules biologiques. Nous présenterons rapidement leurs principes et les informations qu'elles permettent d'obtenir, ainsi que leurs limitations. Ensuite, le chapitre 3 présentera la méthodologie de notre étude, c'est-à-dire les bases théoriques du modèle de solvation semi-implicite utilisé, puis la méthode pour implémenter ce modèle en dynamique moléculaire, ainsi que le paramétrage nécessaire pour le rendre le plus précis possible. Enfin, dans les chapitres 4 et 5, nous présenterons les premières applications de cette nouvelle méthodologie, d'abord à des protéines, puis à des acides nucléiques. Nous avons effectué avec notre nouveau modèle de solvant des dynamiques sur diverses macromolécules, et nous présenterons l'étude des caractéristiques structurales, énergétiques, et d'hydratation de ces molécules, en les comparant aux résultats expérimentaux et théoriques disponibles dans la littérature.



# Chapitre 1

## Présentation générale et nomenclature des macromolécules biologiques



# Présentation générale et nomenclature des macromolécules biologiques

Dans ce chapitre, nous présenterons pour rappel les principales caractéristiques structurales des protéines et acides nucléiques et les nomenclatures associées, dont nous servirons par la suite dans notre étude de l'influence de la solvataion sur la conformation tridimensionnelle des biomolécules. Les macromolécules biologiques sont des polymères constitués d'une succession d'unités élémentaires ou résidus. On distingue principalement deux grandes classes de macromolécules :

- les protéines, composées d'acides aminés ;
- les acides nucléiques, composés de nucléotides.

Ces différentes macromolécules ont des rôles distincts au sein des cellules des organismes vivants. L'information génétique d'un organisme est contenue dans les chromosomes qui se trouvent dans les noyaux de toutes ses cellules. Ces chromosomes sont constitués d'acide désoxyribonucléique (ADN) qui contient dans sa séquence de bases l'information permettant la synthèse des protéines. L'ADN double brin se réplique dans le noyau de la cellule, et est traduit en un autre acide nucléique, l'ARN messager, simple brin, de séquence identique au brin codant de l'ADN (à l'exception de la base thymine remplacée dans l'ARN par la base uracile). L'ARN messager sort du noyau pour transporter l'information génétique dans le cytoplasme de la cellule. Puis, dans les ribosomes (organismes cellulaires constitués d'ARN et de protéines), la séquence de bases est alors traduite en la séquence d'acides aminés de la protéine prévue par le code génétique<sup>1</sup>. C'est un autre acide nucléique, l'ARN de transfert, qui fait la liaison entre l'ARN messager et la protéine. Il va interagir directement avec l'ARN messager par son anticodon complémentaire au codon, et apporter l'acide aminé correspondant au codon dans le ribosome, pour permettre la croissance de la protéine en cours de synthèse.

---

<sup>1</sup>le code génétique est la correspondance entre un triplet de bases nucléiques et un acide aminé.

## 1.1 Les protéines

### 1.1.1 Structure chimique des protéines : la chaîne polypeptidique

Les protéines sont constituées de chaînes polypeptidiques, c'est-à-dire d'un enchaînement d'acides aminés. Un acide aminé est composé d'un atome de carbone (carbone  $\alpha$ ) auquel sont attachés un groupement carboxyle, un groupement amine et une chaîne latérale notée R (cf. figure 1.1). La chaîne latérale R est le plus souvent une chaîne ou un cycle de carbone comportant différents groupements fonctionnels. Cette chaîne peut être réduite à un seul atome d'hydrogène comme pour la glycine. Les structures chimiques des vingt acides aminés sont présentées sur la figure 1.2.

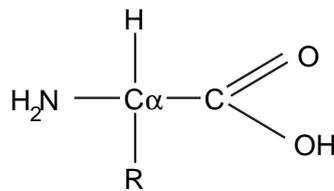


FIG. 1.1 : *Structure chimique d'un acide aminé de chaîne latérale R.*

Les acides aminés sont généralement regroupés en trois classes d'après la nature chimique de leur chaîne latérale [Tooze, 1996] :

- les acides aminés dont la chaîne latérale est hydrophobe : l'alanine (Ala), la valine (Val), la leucine (Leu), l'isoleucine (Ile), la phénylalanine (Phe), la proline (Pro), la méthionine (Met) et la glycine (Gly) ;
- les acides aminés chargés : l'acide aspartique (Asp), l'acide glutamique (Glu), chargés négativement, et la lysine (Lys) et l'arginine (Arg), chargés positivement ;
- les acides aminés dont la chaîne latérale est polaire : la sérine (Ser), la thréonine (Thr), la cystéine (Cys), l'asparagine (Asn), la glutamine (Gln), l'histidine (His), la tyrosine (Tyr) et le tryptophane (Trp).

Deux acides aminés successifs se lient en formant une liaison peptidique entre le groupement amine du premier et le groupement carboxyle du second (cf. figure 1.3), en libérant une molécule d'eau. Un ensemble de plusieurs acides aminés successifs assemblés par liaisons peptidiques forme une chaîne polypeptidique linéaire, appelée protéine lorsqu'elle comporte un certain nombre d'acides aminés, et peptide pour un petit nombre d'acides aminés. Les deux extrémités de la chaîne sont appelées « amino terminale » (ou N-terminale) pour celle qui possède un groupement amine libre, et « carboxy terminale » (ou

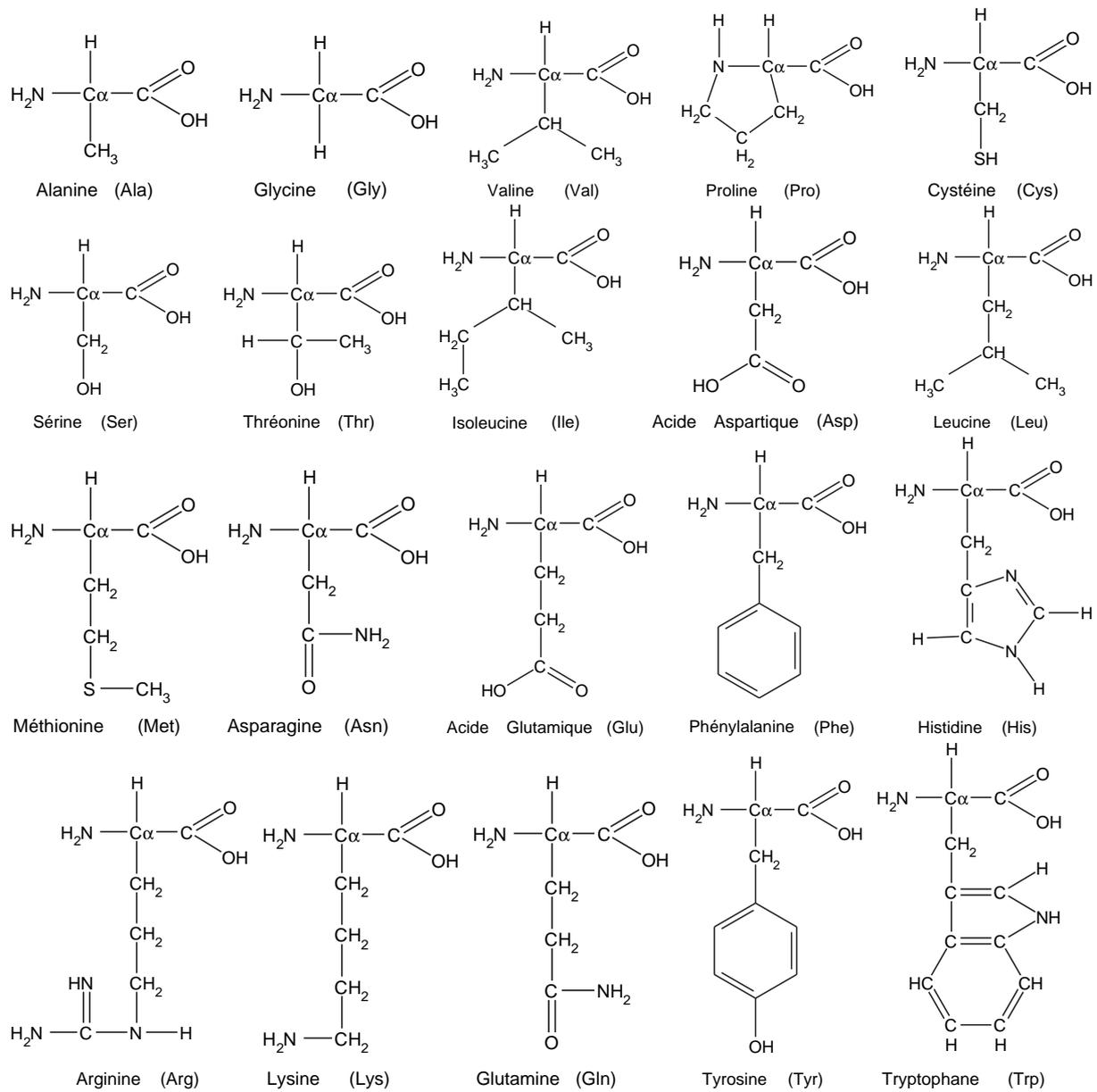


FIG. 1.2 : Structure chimique des vingt acides aminés.

C-terminale) pour celle qui possède un groupement carboxyle libre. La séquence d'acides aminés d'une protéine constitue ce qu'on appelle sa structure primaire. Deux protéines qui présentent de nombreuses similitudes dans leur structure primaire sont dites homologues.

Dans la chaîne polypeptidique, on peut caractériser la déformation de la chaîne principale par deux angles diédraux  $\phi$  et  $\psi$ , représentés sur la figure 1.4 pour une chaîne de trois acides aminés, et définis de la façon suivante entre deux acides aminés  $i$  et  $i + 1$  :

- $\phi$  (pour le résidu  $i$ ) est l'angle de rotation autour de la liaison N-C $\alpha$ , plus préci-

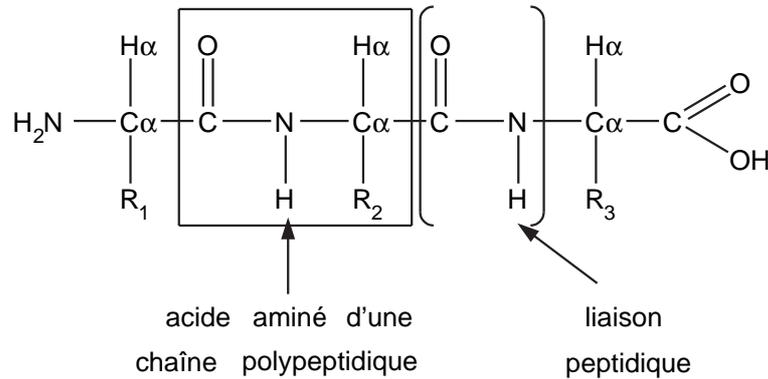


FIG. 1.3 : Chaîne de trois acides aminés successifs de chaînes latérales  $R_1$ ,  $R_2$  et  $R_3$ .

- $\phi$  (pour le résidu  $i$ ) est l'angle dièdre entre les atomes  $C_i-N_i-C\alpha_{i+1}-C_{i+1}$  ;
- $\psi$  (pour le résidu  $i + 1$ ) est l'angle de rotation autour de la liaison  $C\alpha-C$ , plus précisément défini par les atomes  $N_i-C\alpha_{i+1}-C_{i+1}-N_{i+1}$ .

Ces deux angles correspondent aux deux principaux degrés de liberté de rotation entre deux acides aminés. En effet, la liaison peptidique, à cause de son caractère de double liaison partielle qui lui donne une certaine rigidité, reste dans un plan, et on peut considérer que les liaisons et les autres angles entre les atomes de la chaîne principale restent constants le long la chaîne. Ainsi, on peut associer à chaque acide aminé de la chaîne deux angles  $\phi$  et  $\psi$  qui permettront de déterminer totalement la conformation du squelette de la chaîne polypeptidique. À cause des encombrements stériques, les angles  $\phi$  et  $\psi$  adoptent généralement des valeurs privilégiées, correspondant à des minima énergétiques. Pour visualiser cette propriété, on peut représenter sur un graphe, appelé carte de Ramachandran, l'énergie potentielle d'un acide aminé en fonction de  $\phi$  et  $\psi$ , comme on peut le voir sur la figure 1.5.

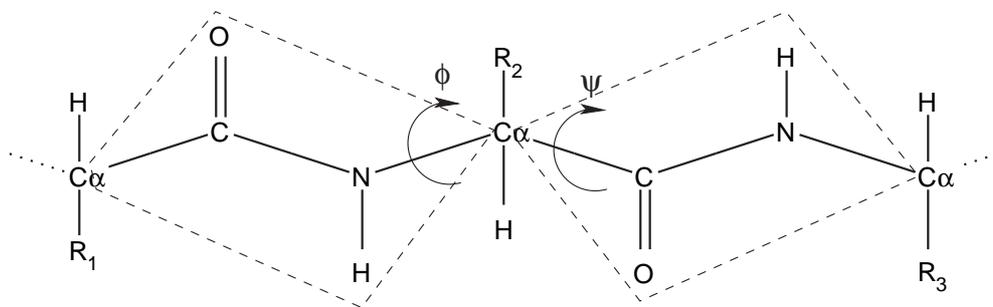


FIG. 1.4 : Représentation des angles diédraux  $\phi$  et  $\psi$  de l'acide aminé de chaîne latérale  $R_2$  dans une chaîne polypeptidique.

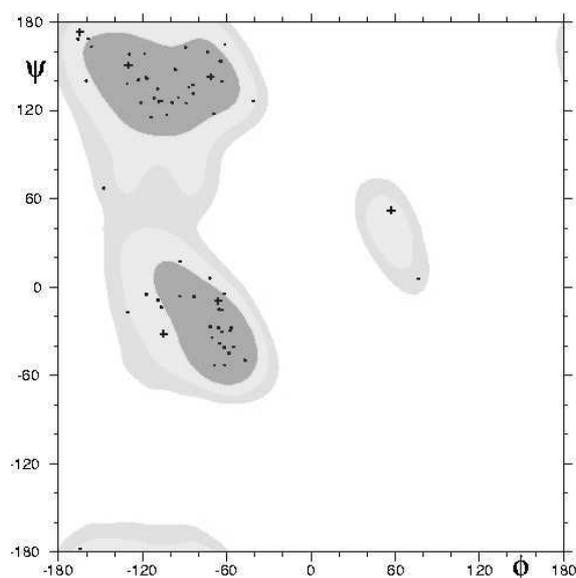


FIG. 1.5 : Carte de Ramachandran pour un acide aminé (ici une alanine), représentant l'énergie potentielle en fonction de  $\phi$  et  $\psi$  : les zones grisées correspondent aux conformations les plus favorables (de plus basses énergies), et les zones blanches aux conformations les plus inaccessibles.

### 1.1.2 Le repliement des protéines

Les chaînes polypeptidiques sont en réalité fortement repliées sur elles-mêmes, et peuvent adopter des structures tridimensionnelles très variées. Une chaîne étendue pourrait atteindre jusqu'à 5000 Å, mais la plupart des protéines repliées atteignent des dimensions de l'ordre de 100 Å [Tooze, 1996]. Le repliement des protéines obéit à certaines contraintes, à cause de l'encombrement stérique des chaînes latérales des acides aminés, et des interactions électrostatiques entre les groupements chargés. De plus, on observe certaines tendances dues aux propriétés chimiques des chaînes latérales des acides aminés, et dues à l'environnement. En effet, alors que les acides aminés polaires ont tendance à se mettre à contact avec l'eau à la surface des protéines, les acides aminés non polaires se regroupent plutôt à l'intérieur de la protéine, formant des chaînes hydrophobes. Ensuite, des liaisons hydrogène se forment entre l'atome d'oxygène du groupement carboxyle d'une liaison peptidique et l'atome d'hydrogène du groupement amine d'une autre liaison. Enfin, des liaisons covalentes peuvent se former entre les groupements SH de deux cystéines (on appelle cette liaison un pont disulfure), contraignant alors de façon importante le repliement.

On voit donc que la structure tridimensionnelle d'une protéine dépend fortement de sa composition en acides aminés. Cependant, la substitution d'un acide aminé polaire par

un autre acide aminé polaire dans la chaîne polypeptidique d'une protéine entraînera peu de changement dans sa structure tridimensionnelle, et de même pour la substitution d'un acide aminé hydrophobe par un autre acide aminé hydrophobe [Freifelder, 1990].

### **La structure secondaire des protéines : les hélices $\alpha$ et feuillets $\beta$**

Dans les structures tridimensionnelles des protéines, on retrouve fréquemment certains motifs particuliers, comme les hélices  $\alpha$  et les feuillets  $\beta$ , qui correspondent à des minima énergétiques des cartes de Ramachandran (cf. figure 1.5). Ces structures particulières sont dues à la formation de liaisons hydrogène entre différentes liaisons peptidiques, plus précisément entre le NH d'une unité peptidique de la chaîne principale et le CO d'une autre. Lorsque les angles  $\phi$  et  $\psi$  de plusieurs résidus successifs sont identiques, des structures régulières se forment, stabilisées par ces liaisons hydrogène. Elles permettent alors de neutraliser les groupements fortement polaires de la chaîne principale. Ces structures sont appelées « structures secondaires » de la protéine, les plus courantes étant les hélices  $\alpha$  et les feuillets  $\beta$ .

#### **• Les hélices $\alpha$ :**

L'hélice  $\alpha$  (cf. figure 1.6) est une hélice stabilisée par les liaisons hydrogène formées entre un groupement peptidique et deux autres groupements peptidiques, l'un situé trois acides aminés en amont, et l'autre trois acides aminés en aval. C'est le CO d'un résidu à la place  $n$  dans la chaîne polypeptidique qui forme une liaison hydrogène avec le NH du résidu  $n + 4$ , et de même son NH forme une liaison hydrogène avec le CO du résidu  $n - 4$ . Cette hélice a des caractéristiques données, avec un pas d'environ 5,4 Å et un diamètre d'environ 2,3 Å. Elle est beaucoup plus étroite que la double hélice d'ADN, et contient 3,6 acides aminés par tour d'hélice. Les angles  $\phi$  et  $\psi$  des résidus impliqués dans une hélice  $\alpha$  varient approximativement autour de  $-60^\circ$  et  $-50^\circ$  [Tooze, 1996].

Les chaînes latérales n'interviennent pas dans la formation des hélices  $\alpha$ , elles sont projetées vers l'extérieur de l'hélice. Mais elles peuvent en bloquer la formation, par exemple dans le cas d'une forte répulsion électrostatique entre les chaînes latérales. L'exemple le plus simple d'hélice  $\alpha$  est la polyglycine. La glycine n'a pas de chaîne latérale, et la conformation en hélice  $\alpha$  est alors énergétiquement la plus favorable [Freifelder, 1990]. La longueur de ces hélices peut varier de 4 ou 5 acides aminés à plus de 40, la longueur moyenne étant d'environ une dizaine de résidus [Tooze, 1996]. L'hélice  $\alpha$  est presque toujours une hélice droite. L'orientation identique des liaisons peptidiques le long d'une hélice  $\alpha$  fait que l'hélice possède un moment dipolaire net orienté vers l'extrémité N-terminale de la chaîne principale.

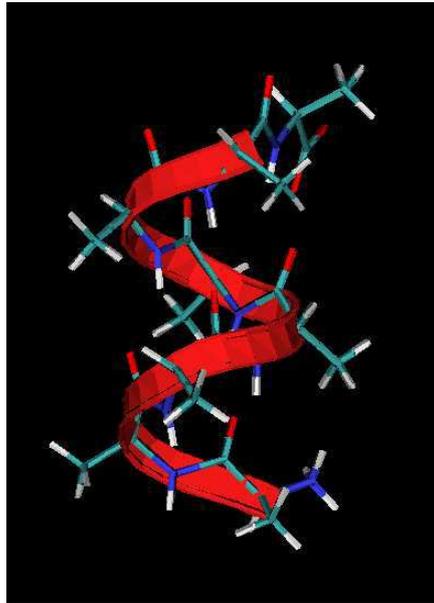


FIG. 1.6 : Structure d'une hélice  $\alpha$ , ici constituée de huit résidus alanine.

Notons que l'on trouve d'autres formes d'hélices que l'hélice  $\alpha$  dans les structures secondaires des protéines. Il existe aussi des hélices appelées  $3_{10}$ , qui se forment également par des liaisons hydrogène entre les groupements peptidiques, mais entre ceux des acides aminés  $n$  et  $n + 3$ . Une hélice  $3_{10}$  sera donc plus étroite qu'une hélice  $\alpha$ .

•Les feuillets  $\beta$  :

Dans les structures en feuillet  $\beta$ , des liaisons hydrogène se forment entre les groupements peptidiques de deux brins placés parallèlement (feuillet  $\beta$  parallèle) ou anti-parallèlement (feuillet  $\beta$  anti-parallèle), d'après le sens d'orientation de la chaîne (cf. figure 1.7). Les chaînes latérales se placent alors alternativement au-dessus et en-dessous du feuillet. La protéine est dans ce cas beaucoup plus étendue que pour les hélices  $\alpha$ . Quand plusieurs brins  $\beta$  sont ainsi alignés, on parle de feuillet  $\beta$ . Chaque brin peut être d'une longueur de 5 à 10 résidus, particulièrement étiré, avec des angles  $\phi$  et  $\psi$  qui peuvent varier largement autour de  $-100^\circ$  et  $+100^\circ$ . Il existe aussi des feuillets  $\beta$  mixtes, où l'on trouve à la fois des brins  $\beta$  parallèles et anti-parallèles. Dans presque tous les feuillets  $\beta$  des protéines connues, les brins ne sont pas complètement plans et possèdent une légère torsion [Tooze, 1996].

Les protéines globulaires, comme les enzymes, qui sont extrêmement compactes, ont des courtes régions en hélice  $\alpha$  et en feuillet  $\beta$ , et d'autres régions appelées boucles où le repliement est plus aléatoire. Les protéines fibreuses, comme le collagène, sont constituées d'une seule de ces structures, et sont assez longues et fines [Freifelder, 1990]. Certaines

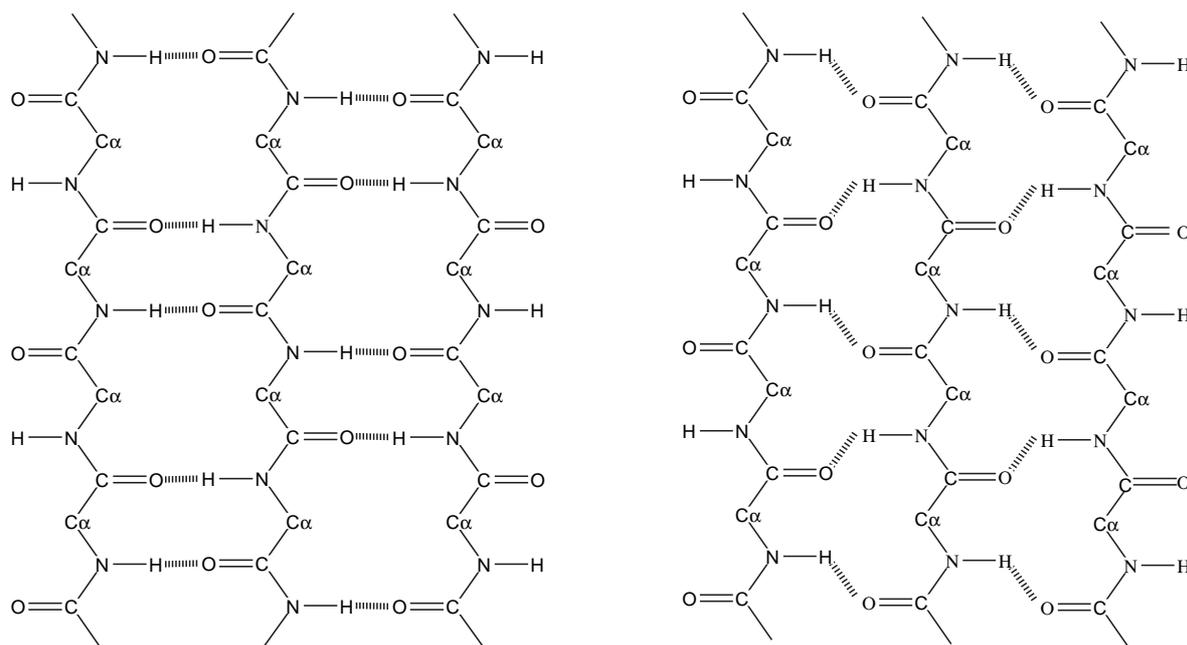


FIG. 1.7 : Structure schématique des feuillets  $\beta$ , anti-parallèles à gauche (les chaînes sont en sens inverse) et parallèles à droite (les chaînes sont orientées dans le même sens). Les chaînes latérales ne sont pas représentées, elles se situent alternativement au-dessus et en-dessous du plan de la feuille.

séquences d'acides aminés peuvent favoriser la formation de l'une ou l'autre des structures secondaires présentées, ou au contraire la formation de boucles. On a représenté schématiquement la structure tridimensionnelle d'une protéine globulaire sur la figure 1.8.

### La structure tertiaire et quaternaire des protéines

La structure tertiaire des protéines représente leurs repliements plus complexes qui résultent des interactions entre les chaînes latérales des acides aminés. Ces interactions peuvent être des liaisons ioniques, des liaisons hydrogène, ou des interactions hydrophobes entre chaînes apolaires [Freifelder, 1990]. Toutes ces interactions peuvent être diverses et complexes, et les structures tridimensionnelles des protéines résultantes peuvent être particulièrement variées. On peut cependant remarquer que, le plus souvent, l'intérieur des protéines est majoritairement constitué d'acides aminés ayant des chaînes latérales hydrophobes, et s'organise en structures secondaires  $\alpha$  et  $\beta$ , alors que les boucles qui relient ces structures secondaires sont situées en surface des protéines.

On distingue dans les protéines globulaires des structures compactes, composées de plusieurs motifs de structures secondaires, appelées domaines. La structure tertiaire

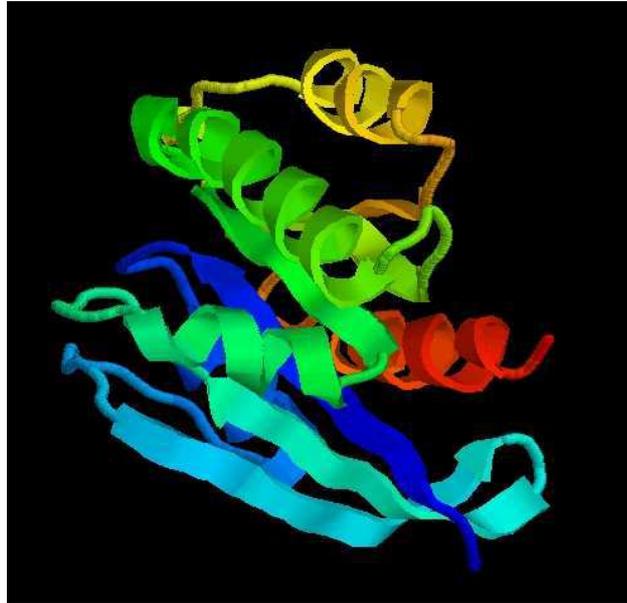


FIG. 1.8 : Structure tridimensionnelle d'une protéine globulaire de type GTP présentant à la fois des structures secondaires en hélice  $\alpha$  et en feuillet  $\beta$  [Nassar et al., 1996].

représente à la fois les arrangements des motifs au sein des domaines, et la caractérisation même de ces domaines dans la structure de la protéine. Un domaine peut être tout ou partie d'une chaîne polypeptidique dont la structure tertiaire est stable. En général, les différents domaines d'une protéine remplissent différentes fonctions biologiques (par exemple un domaine de liaison à l'ADN, ou le domaine d'un site catalytique,...).

Les protéines peuvent de plus être constituées de plusieurs sous-unités, c'est-à-dire de plusieurs chaînes polypeptidiques, identiques ou non. On parle alors de protéines multimériques. L'arrangement de ces sous-unités entre elles caractérise la structure quaternaire de la protéine.

À la différence de l'ADN, dont nous verrons que la structure tridimensionnelle en double hélice est relativement bien conservée quelque soit sa séquence, les protéines ont des structures extrêmement complexes et très difficiles à retrouver à partir de la simple séquence des acides aminés, puisque de nombreuses interactions rentrent en jeu, à la fois entre les chaînes principales et les chaînes latérales. Comme la fonction des protéines dans la cellule est intimement liée à leur structure, l'étude de la structure et du repliement des protéines est actuellement un enjeu majeur de la biologie moléculaire. Les structures tridimensionnelles des protéines résolues par cristallographie ou par résonance magnétique nucléaire (ainsi que celles de nombreux complexes) sont disponibles sur le site internet de la Protein Data Bank (appelée PDB, <http://www.rcsb.org/pdb>) [Berman *et al.*, 2000].

## 1.2 Les acides nucléiques

### 1.2.1 Structure chimique des acides nucléiques

#### La chaîne polynucléotidique

Les acides nucléiques sont constitués de chaînes polynucléotidiques. Chaque nucléotide est lui-même composé de trois groupements chimiques [Freifelder, 1990] :

1. un sucre cyclique furanose composé de cinq atomes de carbone, le ribose pour l'acide ribonucléique (ARN) et le désoxyribose pour l'acide désoxyribonucléique (ADN) ;
2. un groupement phosphate chargé négativement, lié au carbone 5' du sucre par une liaison phosphodiester ;
3. une base azotée, purique ou pyrimidique, attachée latéralement au sucre par une liaison N-glycosidique avec l'atome de carbone 1' du sucre.

Deux nucléotides successifs sont attachés par une liaison de type phosphodiester formée entre le groupement phosphate d'un premier nucléotide et le sucre du nucléotide suivant [Watson *et al.*, 1994]. Par convention, la direction de la chaîne nucléotidique est orientée dans le sens 5'–3' (d'après la notation des carbones 5' et 3' des sucres), qui correspond au sens de lecture des séquences d'acides nucléiques (cf. figure 1.9).

#### Les bases des acides nucléiques et la structure en double hélice

Les bases sont des molécules planes cycliques constituées de carbone et d'azote. On les classe en deux catégories : les purines (R) et les pyrimidines (Y). Dans l'ADN, on rencontre pour les purines les bases adénine (A) et guanine (G), et pour les pyrimidines la cytosine (C) et la thymine (T). Dans l'ARN, la thymine est remplacée par l'uracile (U), qui est équivalente à une thymine sans le groupement méthyle à la position C5. On trouve aussi dans l'ARN des bases modifiées qui ne diffèrent des bases usuelles que par quelques groupements d'atomes, comme nous le verrons plus loin.

Les interactions entre ces bases sont de deux sortes :

- les interactions de « stacking », ou d'empilement : ce sont les interactions entre les systèmes électroniques  $\pi$  conjugués des cycles aromatiques des bases. Ces interactions hydrophobes poussent les surfaces des cycles de deux bases adjacentes à se mettre en contact, formant un empilement de bases qui permet de rigidifier les polynucléotides simples brins ;

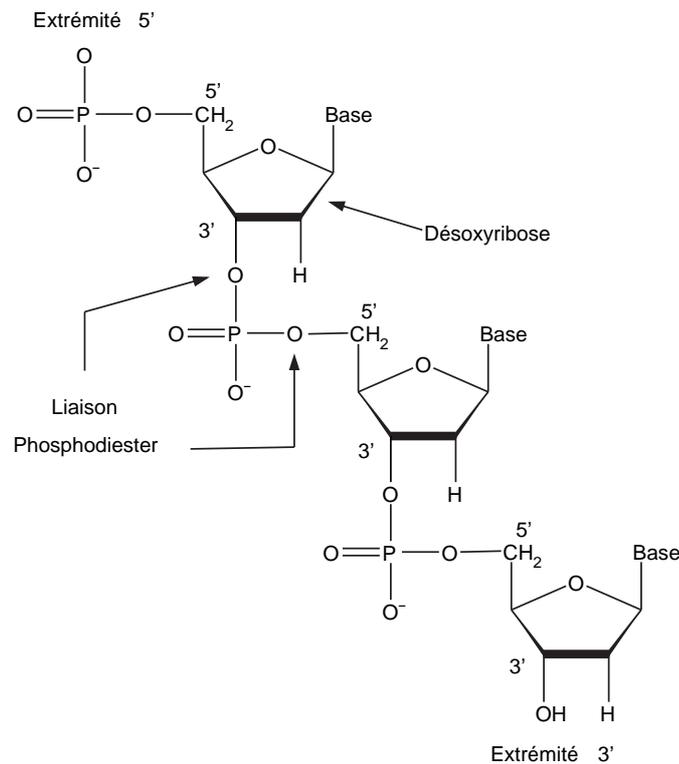


FIG. 1.9 : Structure chimique d'un brin d'ADN.

- les liaisons hydrogène : elles se forment entre bases complémentaires puriques et pyrimidiques, entre G et C d'une part, et entre A et T (ou A et U) d'autre part. La topologie de cet appariement est le plus souvent de type Watson-Crick [Watson et Crick, 1953], représenté sur la figure 1.10. La paire G-C possède trois liaisons hydrogène, alors que la paire A-T (ou A-U) n'en possède que deux. Ces liaisons hydrogène sont majoritairement dues aux interactions électrostatiques.

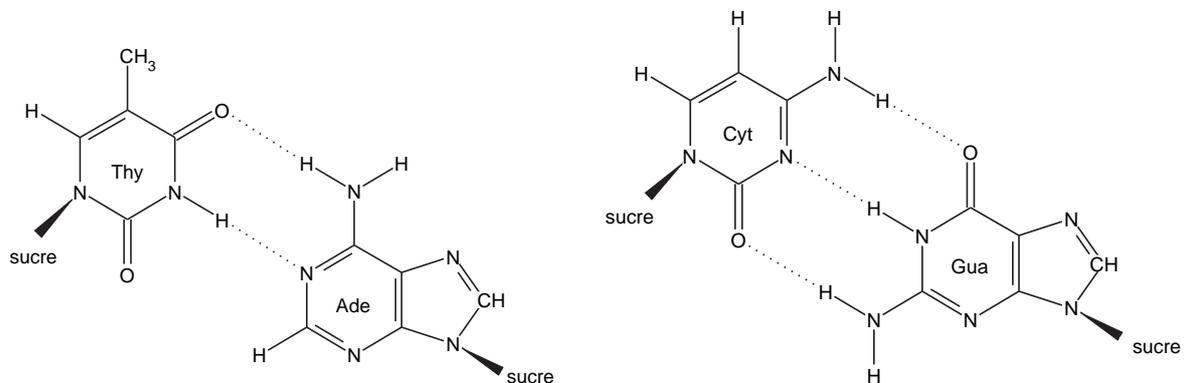


FIG. 1.10 : Appariement des bases complémentaires de l'ADN A-T et C-G.

L'empilement des bases d'un même brin et la complémentarité entre les bases A-T (ou A-U) et G-C sont à l'origine de la stabilité des doubles hélices [Watson et Crick, 1953].

En effet, deux simples brins d'acide nucléique anti-parallèles, c'est-à-dire deux chaînes polynucléotidiques de séquences complémentaires, l'une dans le sens 5'–3' et l'autre dans le sens inverse 3'–5', s'enroulent autour d'un même axe, formant ainsi une double hélice dont la structure est maintenue par les liaisons hydrogène entre bases complémentaires à l'intérieur de l'hélice. Cette double hélice possède deux sillons hélicoïdaux, dont les propriétés diffèrent selon la forme A ou B de l'ADN (cf. figure 1.11 pour la forme B).

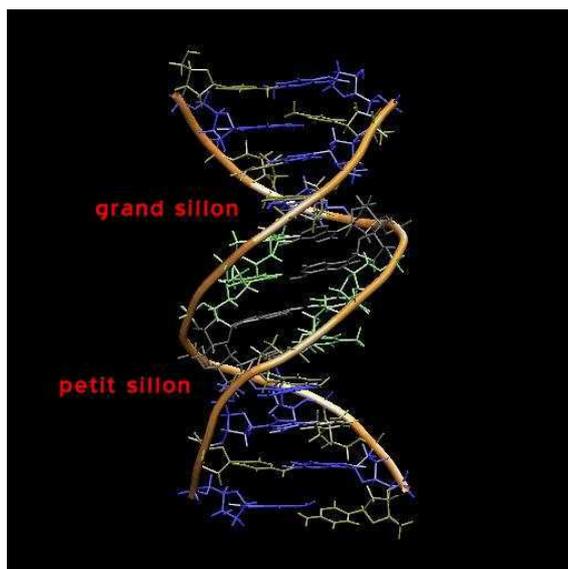
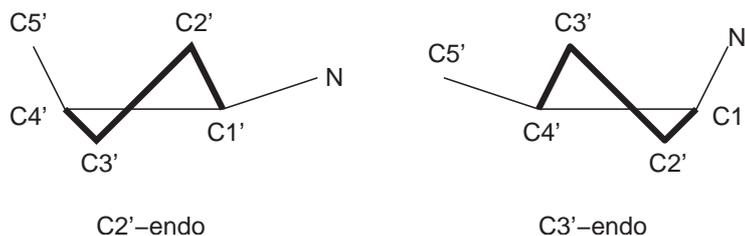


FIG. 1.11 : La double hélice de l'ADN B : le grand sillon est plus large que le petit sillon.

### Le sucre furanose

Le sucre des acides nucléiques (ADN et ARN) est un furanose, cycle à cinq atomes. Un cycle à cinq atomes est généralement « plissé », c'est-à-dire qu'il est non plan. Pour le furanose, on caractérise ce plissement par la nomenclature suivante : on nomme l'atome le plus déplacé hors du plan cyclique moyen, et on le qualifie « endo » s'il pointe vers le côté 5' de la chaîne polynucléotidique, et « exo » selon s'il pointe vers le côté 3'. Par exemple si l'atome C3' est hors du plan vers le côté 5' de la chaîne, on nommera cette conformation « C3'-endo » (cf. figure 1.12).

En pratique, on caractérise principalement la conformation d'un sucre par deux variables (les deux autres variables théoriquement nécessaires pour caractériser complètement la conformation du cycle n'étant que des réajustements mineurs) [Kilpatrick *et al.*, 1947] : la phase  $P$  qui décrit la position et le signe du plissement, et l'amplitude  $\tau_m$  qui quantifie le degré de non planéité. Cette amplitude est définie par l'angle dièdre endo-cyclique de valeur maximale. La phase et l'amplitude d'un sucre sont reliées aux cinq

FIG. 1.12 : Les conformations  $C2'$ -endo (Nord) et  $C3'$ -endo (Sud) du furanose.

angles dièdres endo-cycliques  $\nu_i$ , définis par la figure 1.13, par les relations suivantes :

$$\tan(P) = \frac{(\nu_1 + \nu_4) - (\nu_0 + \nu_3)}{2\nu_2 [\sin(36^\circ) + \sin(72^\circ)]}$$

$$\tau_m = \frac{\nu_2}{\cos(P)}$$

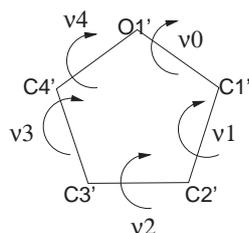
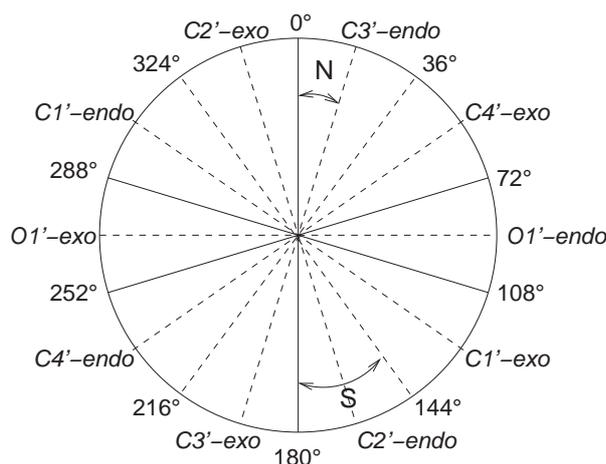


FIG. 1.13 : Les cinq angles dièdres endo-cycliques du furanose.

FIG. 1.14 : Cercle de pseudo-rotation du furanose. Correspondance entre la phase  $P$  en degrés et le nom de la conformation associée.

On caractérise les vingt conformations extrêmes du sucre par une phase multiple de  $18^\circ$  ( $360^\circ/20$ ), et toutes les autres conformations par une phase intermédiaire. On

peut représenter ces conformations sur une figure appelée « cercle de pseudo-rotation », représenté sur la figure 1.14. Dans le cas d'un nucléotide isolé, on observe que les conformations C2'-endo (Sud) et C3'-endo (Nord) sont énergétiquement les plus stables [Levitt et Warshel, 1978].

## 1.2.2 La structure des ADN

### Les paramètres hélicoïdaux des hélices d'ADN

Pour décrire la structure globale d'une double hélice d'ADN, il est nécessaire de connaître non seulement les paramètres de conformation du sucre et les angles dièdres du squelette phosphodiester, mais aussi des paramètres permettant de décrire le motif hélicoïdal de la molécule. Contrairement à l'image habituelle de la double hélice régulière à pas fixe, la double hélice de l'ADN comporte de nombreuses irrégularités structurales qu'il est important de caractériser. Depuis la convention de Cambridge en 1988 [Dickerson *et al.*, 1989], on décrit les structures hélicoïdales des acides nucléiques à l'aide d'un ensemble de paramètres donnés qui sont divisés en trois groupes :

- les paramètres axe-base et axe-paire de bases (cf. figure 1.15) : ils permettent de décrire la position d'une base ou d'une paire de bases par rapport à l'axe hélicoïdal, par deux translations ( $X_{disp}$  et  $Y_{disp}$ ) et deux rotations ( $Inc$  et  $Tip$ ) ;
- les paramètres intra-paire de bases (cf. figure 1.16) : la position d'une base par rapport à sa base complémentaire est repérée à l'aide de six variables :  $Shear$ ,  $Stretch$  et  $Stagger$  pour les translations, et  $Buckle$ ,  $Propeller Twist$  et  $Opening$  pour les rotations ;
- les paramètres inter-bases et inter-paires de bases (cf. figure 1.16) : on repère la position relative de deux bases ou deux paires consécutives à l'aide des six paramètres :  $Shift$ ,  $Slide$  et  $Rise$  pour les translations, et  $Tilt$ ,  $Roll$  et  $Twist$  pour les rotations.

Par convention, l'axe ( $Ox$ ) est dirigé vers le grand sillon, l'axe ( $Oy$ ) du brin  $II$  vers le brin  $I$ , et l'axe ( $Oz$ ) dans le sens 5'-3' du brin  $I$ . Les paramètres axe-paire de bases sont calculés à partir des paramètres axe-bases d'après le tableau 1.1.

Pour calculer ces paramètres, il est donc nécessaire de bien définir l'axe hélicoïdal. Il existe deux façons différentes de calculer cet axe hélicoïdal : la méthode dite locale définit un axe local pour chaque nucléotide et ne caractérise donc pas la structure globale de l'hélice, mais la jonction entre paires de bases successives ; la méthode dite globale définit un axe global pour toute la molécule et tous les paramètres sont calculés par rapport à cet axe. Cette dernière méthode permet de donner une image globale des déformations de

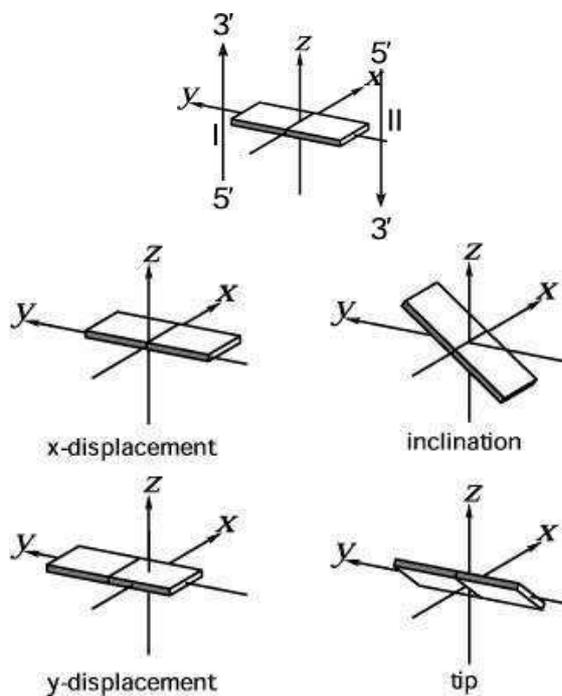


FIG. 1.15 : Paramètres hélicoïdaux axe-paire de bases.

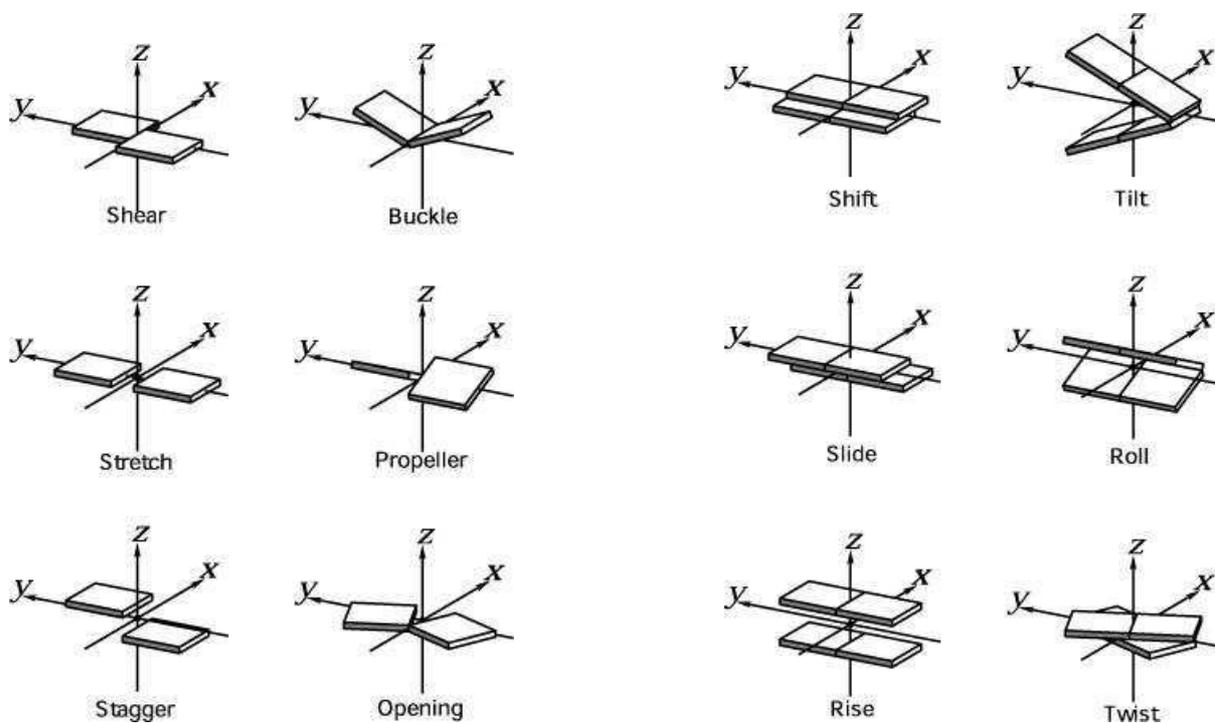


FIG. 1.16 : Paramètres hélicoïdaux intra-paire de bases (à gauche) et inter-paires de bases (à droite).

Nom	Symbole	Définition pour une paire de bases
Xdisp	$dx$	$[dx(I) + dx(II)]/2$
Ydisp	$dy$	$[dy(I) - dy(II)]/2$
Inc	$\eta$	$[\eta(I) + \eta(II)]/2$
Tip	$\theta$	$[\theta(I) - \theta(II)]/2$

TAB. 1.1 : Définitions des paramètres axe-paire de bases.

l'hélice, et aussi de distinguer facilement les différentes formes d'ADN que nous verrons plus loin : A, B (et Z). On préférera donc caractériser la double hélice d'ADN par les paramètres calculés par rapport à l'axe global, même si, dans le cas de molécules d'ADN courbées, cela pose alors le problème du calcul du meilleur axe hélicoïdal courbe [Lavery et Zakrewzka, 1997].

### Les différentes formes de l'ADN

La double hélice d'ADN peut prendre des formes très différentes selon son environnement, en particulier le degré d'humidité ou la concentration en sels. Grâce aux différentes méthodes expérimentales, comme les études de diffraction aux rayons X sur des fibres et cristaux d'oligonucléotides, et les expériences de RMN en solution, on a pu classer les structures d'ADN en trois grandes familles : B, A et Z. L'ADN Z est généralement instable en solution [Watson *et al.*, 1994], et nous ne détaillerons donc pas cette famille de structures. Les valeurs caractéristiques moyennes non nulles des paramètres structuraux pour les formes A et B sont présentées dans le tableau 1.2.

Paramètre	ADN-A	ADN-B
Xdisp	-5,3 Å	0,0 Å
Inc	20,7°	1,5°
Propeller Twist	-7,5°	-13,3°
Rise	2,6 Å	3,4 Å
Twist	32,7°	36,0°
Phase	18,3°	154,8°
Amplitude	41,6°	39,7°

TAB. 1.2 : Valeurs des paramètres structuraux différents pour les deux familles d'ADN A et B [Hartmann et Lavery, 1996].

● **l'ADN B** (cf. figure 1.17)

C'est la forme dominante dans des conditions de forte humidité [Wilkins *et al.*, 1953,

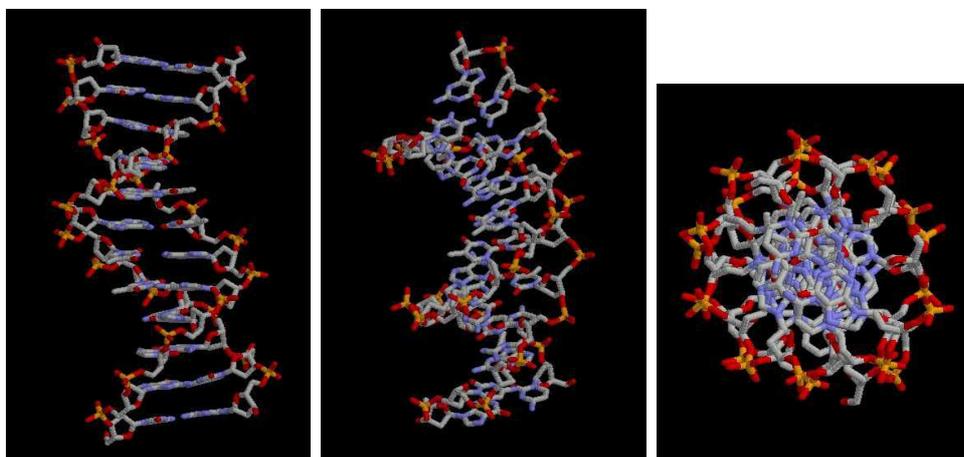


FIG. 1.17 : Vues d'une double hélice d'ADN B.

Franklin et Goslin, 1953], et donc la forme majoritaire de l'ADN dans la cellule. Les paires de bases sont situées au centre de la double hélice, presque perpendiculaires à l'axe hélicoïdal, appariées par des liaisons hydrogène de type Watson-Crick. La valeur du Rise est alors d'environ 3,4 Å à cause des interactions d'empilement entre paires de bases successives. Les groupements phosphates, dont les atomes d'oxygène chargés négativement pointent vers l'extérieur de l'hélice, sont alors stabilisés par leur interaction avec les molécules d'eau et les contre-ions chargés positivement présents dans la cellule. Le petit sillon est deux fois moins large que le grand sillon, mais leurs profondeurs sont comparables. Les sucres de l'ADN-B sont majoritairement en conformation C2'-endo [Ulyanov *et al.*, 1995]. Dans la « Nucleic Acid Database » (NDB, <http://ndbserver.rutgers.edu>), qui recense toutes les structures des acides nucléiques cristallisées et issues d'expériences de résonance magnétique nucléaire, on trouve actuellement environ 250 structures d'ADN B, dont seulement une trentaine d'entre elles résolues à moins de 1,5 Å. Au sein de la famille des ADN B, il existe une grande variabilité structurale, et les paramètres hélicoïdaux ne sont pas fixés sur une valeur unique [Hartmann et Lavery, 1996]. On peut cependant observer des couplages entre certains paramètres hélicoïdaux, ce qui permet de classer grossièrement les différents pas nucléotidiques selon les valeurs de certains de leurs paramètres, en particulier le Twist, le Rise et le Roll. De plus, grâce à des études de résonance magnétique nucléaire, on pense que la séquence de bases aurait une influence plus importante sur le Twist et la phase des sucres [Cheng *et al.*, 1992].

● **l'ADN A** (cf. figure 1.18)

La forme A de l'ADN est dominante dans des conditions de faible humidité, mais peut

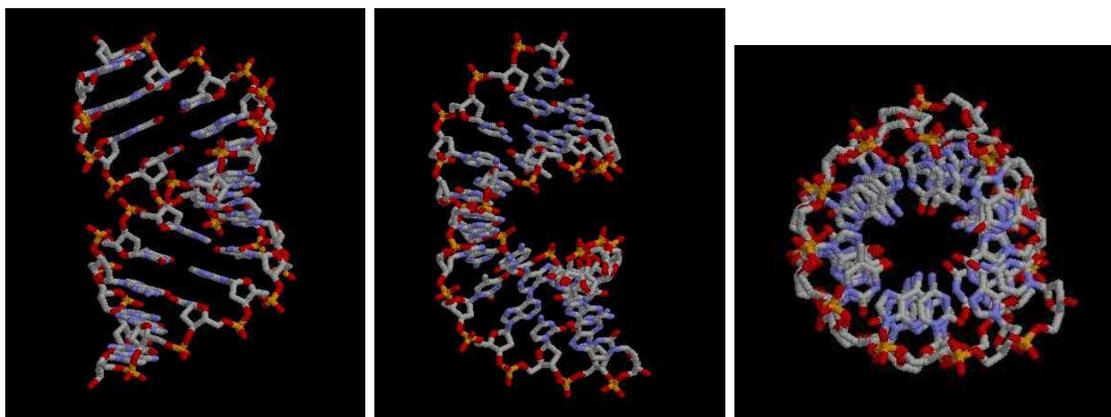


FIG. 1.18 : Vues d'une double hélice d'ADN A.

aussi être favorisée par des séquences de bases riches en paires C–G [Wang *et al.*, 1987]. Dans la cellule, on observe localement des formes A d'ADN entre deux fragments en forme B. Le diamètre de la double hélice de l'ADN en forme A est plus grand que celui de la forme B. Les paires de bases ne sont plus au centre de l'hélice, mais vers le petit sillon, qui est plus large et moins profond que celui de l'ADN B. Les paires de bases sont fortement inclinées par rapport au plan perpendiculaire à l'axe hélicoïdal. Les sucres sont préférentiellement en conformation C3'-endo, quant au Twist et au Rise, ils sont plus faibles que pour la forme B (cf. tableau 1.2). Les groupements phosphates sont plus proches dans le grand sillon, qui est moins large et plus profond que celui de la forme B. La séquence de bases aurait moins d'influence sur les paramètres de la double hélice de forme A que sur celle de forme B. On trouve dans la NDB plus d'une centaine de structures d'ADN en forme A, dont une vingtaine sont résolues à moins de 1,5 Å, soit un peu moins que pour l'ADN B. On trouve aussi fréquemment dans les double hélices d'ADN des structures hybrides entre l'ADN A et l'ADN B [Rao et Kollman, 1985].

### 1.2.3 La structure des ARN

Les molécules d'ARN sont constituées d'un seul brin nucléotidique, mais on peut trouver dans leur structure tridimensionnelle des formes de repliement en double hélice. Ce repliement est dû à la complémentarité entre certaines séquences de bases du même brin. La particularité des ARN par rapport aux ADN est non seulement la nature du sucre de la chaîne (le ribose de l'ARN diffère du désoxyribose de l'ADN par un groupement OH au lieu de H lié au carbone C2' du cycle), mais aussi le remplacement dans leur séquence de la base T par la base U, dont la structure chimique est donnée par la figure 1.19. De plus, notons que l'ARN *in vivo*, quand il est en double brin, est le plus souvent de forme

A [Watson *et al.*, 1994].

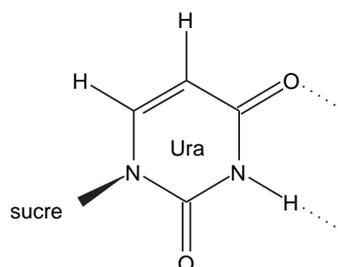


FIG. 1.19 : Structure chimique de l'uracile.

Il faut noter aussi que les ARN ont la particularité de contenir dans leur séquence des bases inhabituelles, formées d'un noyau de purine ou pyrimidine, mais dont certains substituants diffèrent de ceux des bases classiques. Il existe jusqu'à cinquante bases modifiées, la structure chimique de deux d'entre elles étant décrite par la figure 1.20.

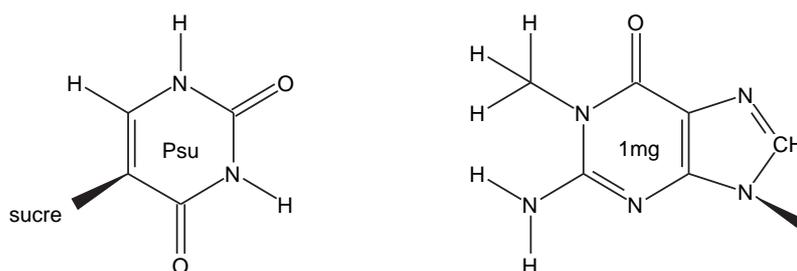


FIG. 1.20 : Structure chimique de la pseudo-uracile ( $\psi$  ou *Psu*), qui possède un noyau d'uracile, et de la 1-méthylguanine ( $m^1G$  ou *1mG*), qui possède un noyau de guanine.

Dans la cellule, on trouve trois classes principales de molécules d'ARN :

- les ARN messagers (ARNm) ;
- les ARN ribosomiques (ARNr) ;
- les ARN de transfert (ARNt).

Comme nous l'avons évoqué précédemment, chacun joue un rôle particulier dans la synthèse protéique. Dans la Nucleic Acid Database, on trouve presque 150 structures d'ARN, dont une cinquantaine sont résolues à mieux que 1,5 Å. On trouve aussi une dizaine de structures de ribozymes, constitués d'ARN ribosomique. On sait depuis quelques années que le site actif (là où sont synthétisées les protéines) de la grande sous-unité du ribosome est exclusivement constitué d'ARN, et non pas de protéines comme on le croyait précédemment [Ban *et al.*, 2000].

Parmi les ARN, nous nous intéresserons dans ce travail plus particulièrement à l'ARN de transfert (ARNt), dont nous allons présenter brièvement quelques propriétés.

L'ARNt est composé d'une seule chaîne de 75 à 95 nucléotides. Sa structure secondaire, représentant ses domaines en boucle et en simple et double brin, est formée de quatre tiges en double hélice et de trois boucles d'ARN monocaténaire simple brin, ce qui lui donne sa forme connue en « feuille de trèfle » (cf. figure 1.21).

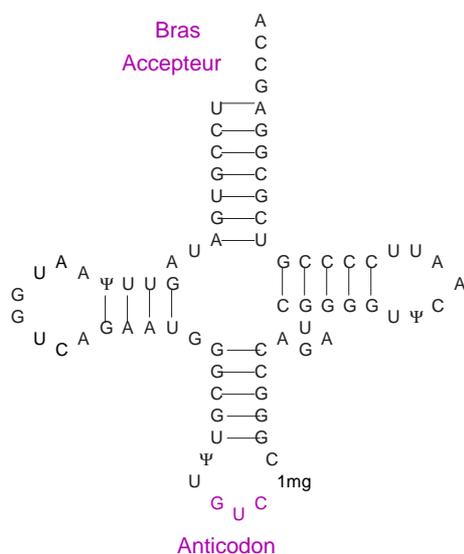


FIG. 1.21 : La structure secondaire de l'ARN de transfert (ici celui de l'acide aspartique) en feuille de trèfle. Seules les bases modifiées  $\psi$  et 1mg sont représentées, les autres sont remplacées par leur noyau (U ou G) pour simplifier le schéma.

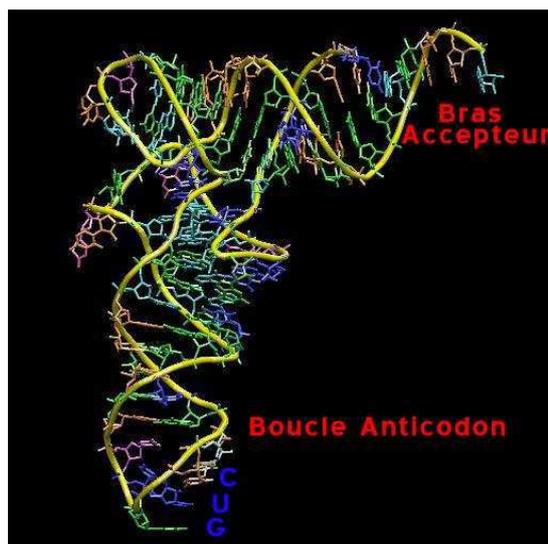


FIG. 1.22 : La structure tertiaire en L de l'ARN de transfert (de l'acide aspartique).

La structure tertiaire de l'ARNt, c'est-à-dire sa structure tridimensionnelle, a la forme d'un L (cf. figure 1.22), avec à une extrémité la boucle comprenant le triplet de

l'anticodon (succession de trois bases), et à l'autre le bras accepteur où se fixe par liaison covalente l'acide aminé, correspondant d'après le code génétique au codon de l'ARN messager complémentaire de l'anticodon.



## **Chapitre 2**

# **La solvataion des macromolécules biologiques**



# La solvatisation des macromolécules biologiques

Nous présenterons dans ce chapitre les différentes méthodes existantes, expérimentales et théoriques, pour étudier la solvatisation des macromolécules biologiques, c'est-à-dire les effets du solvant sur ces molécules. La solvatisation joue un rôle majeur dans la structure et la dynamique des macromolécules biologiques. En effet, dans la cellule, le solvant – l'eau –, qui entoure les molécules, stabilise leurs structures tridimensionnelles et influence les phénomènes de reconnaissance moléculaire, par des effets à la fois stériques et électrostatiques. Les molécules d'eau autour des macromolécules biologiques peuvent être classées en trois catégories selon leur relation avec la macromolécule [Schwabe, 1997], représentées sur la figure 2.1 :

1. les molécules d'eau du « bulk » peu perturbées par la présence du soluté (sensibles au soluté uniquement par des interactions à longue portée), auront un effet moyen sur le soluté ;
2. les molécules d'eau proches du soluté, qui peuvent interagir avec la surface de la macromolécule, forment les premières couches de solvatisation de celle-ci, et s'échangent rapidement avec les molécules du bulk ; ces molécules d'eau vont influencer l'énergie libre de solvatisation de la molécule ;
3. les molécules d'eau fortement liées à la macromolécule, qui se trouvent dans des cavités intérieures de la molécule ou peu accessibles au solvant, s'échangent alors peu avec le bulk ; elles font partie intégrante de la structure tridimensionnelle de la molécule.

L'étude de la solvatisation des macromolécules biologiques permettra dans une certaine mesure de caractériser ces trois catégories de molécules d'eau. D'une part, les méthodes expérimentales (cristallographie des rayons X, RMN et méthodes de mesures énergétiques) permettent de mieux comprendre le rôle de certaines molécules d'eau spécifiques dans les systèmes biologiques et d'accéder à certaines grandeurs thermodynamiques. Les méthodes expérimentales permettront donc de caractériser principalement les molécules

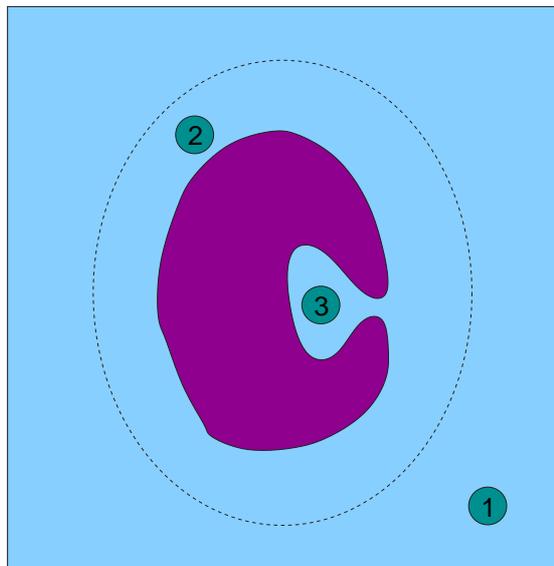


FIG. 2.1 : Le solvant autour d'une macromolécule : le bulk (1), les premières couches de solvation (2), et les molécules d'eau liées dans les cavités (3).

d'eau des cavités et des premières couches de solvation. D'autre part, les méthodes théoriques de modélisation moléculaire, telles que la dynamique moléculaire, vont permettre de comprendre les phénomènes de solvation de façon plus globale, mais elles posent alors le problème de la représentation du solvant dans ce type de modélisation. Cette représentation peut être soit explicite, en représentant chaque molécule d'eau triatomique autour de la macromolécule, soit implicite, en considérant le solvant comme un milieu continu caractérisé par une constante diélectrique élevée, différente de celle du soluté. Les méthodes implicites rendront compte des effets moyens du solvant, c'est-à-dire des molécules d'eau du bulk, alors que les méthodes explicites permettront d'identifier des molécules d'eau dans les cavités et dans les premières couches de solvation. Il existe aussi des méthodes hybrides, implicites, mais qui prennent en compte les aspects moléculaires du solvant, afin de représenter correctement les trois catégories de molécules d'eau.

## 2.1 Mise en évidence du rôle de l'eau par les méthodes expérimentales

Par l'expérience, on peut étudier l'hydratation des macromolécules biologiques à l'aide de plusieurs méthodes qui apportent chacune des informations spécifiques, différentes et complémentaires. On peut distinguer parmi ces méthodes expérimentales deux grandes familles de méthodes :

- les méthodes structurales (cristallographie par diffraction des rayons X et résonance magnétique nucléaire), permettent de détecter la présence (pour la cristallographie) et la dynamique (pour la RMN) de molécules d'eau spécifiques autour de la structure des macromolécules ;
- les méthodes thermodynamiques, reposant sur des mesures de microcalorimétrie, permettent d'accéder à l'énergie libre de solvatation des macromolécules et à ses variations lors de repliement ou de complexation ; ces mesures de microcalorimétrie couplées à des approches volumétriques ou de pression osmotique permettent de caractériser, en terme de nombre de molécules et d'énergie, les premières couches de solvatation.

### 2.1.1 Méthodes structurales

La mobilité des molécules d'eau rendait relativement difficile leur détection par les méthodes structurales classiques telles que la RMN et la cristallographie. Mais le développement et la précision croissante de ces méthodes permettent aujourd'hui de bien identifier les sites préférentiels d'hydratation des protéines et acides nucléiques.

#### La cristallographie par diffraction des rayons X

Les expériences de diffraction des rayons X permettent d'accéder à la densité électronique d'une molécule dans un cristal. Les données obtenues sont moyennées à la fois dans le temps et sur l'ensemble des mailles du cristal. À partir de cette carte de densité électronique complexe et à l'aide de programmes de raffinements cristallographiques qui permettent d'interpréter cette carte, on peut retrouver la structure tridimensionnelle de la molécule étudiée. Les résolutions des structures cristallographiques sont plus ou moins précises, suivant la qualité du cristal et les conditions de l'expérience.

D'après une carte de densité électronique continue du solvant, on identifie des molécules d'eau localisées dans les cristaux au niveau des extrema de la carte de densité. Il faut ensuite vérifier que ces molécules d'eau sont présentes dans différentes formes cristallines de qualité suffisante. En effet, les molécules d'eau identifiées à la surface des molécules ne sont pas toujours observées selon les différents cristaux étudiés de la même molécule [Kovacs *et al.*, 1997]. Notons que les cartes dont on dispose sont moyennées dans le temps et dans le cristal, elles sont continues, et leur interprétation en terme de molécules d'eau ponctuelles peut donc parfois être ambiguë [Makarov *et al.*, 2002].

Pour préciser ensuite les sites d'hydratation de la molécule, on peut étudier quelles molécules d'eau forment des liaisons hydrogène avec la macromolécule cristallographiée. Pour cela, on cherche les molécules d'eau parmi celles identifiées dans le cristal qui peuvent réellement former des liaisons hydrogène avec la macromolécule selon des critères géométriques [Eisenstein et Shakked, 1995]. En effet, théoriquement, une molécule d'eau peut donner ses deux atomes d'hydrogène, et recevoir deux atomes d'hydrogène par les deux doublets électroniques libres de son atome d'oxygène, pour former des liaisons hydrogène avec la macromolécule ou une autre molécule d'eau. La même étude de sites donneurs et accepteurs sur les bases et acides aminés permet de décider, par des critères de distances et d'angles entre les sites donneurs et les sites accepteurs, si la molécule d'eau forme ou non une liaison hydrogène avec la macromolécule. Encore une fois, cette étude est délicate car elle dépend des critères choisis pour considérer s'il y a vraiment liaison hydrogène.

## La Résonance Magnétique Nucléaire

Les expériences de résonance magnétique nucléaire (RMN) sur les macromolécules biologiques, contrairement aux expériences de cristallographie, peuvent être effectuées en solution, permettant alors une étude de l'hydratation des macromolécules. La spectroscopie par RMN sur les macromolécules biologiques en solution permet de mesurer les effets nucléaires Overhauser (NOE) intermoléculaires entre les protons de l'eau et ceux de l'acide nucléique ou de la protéine. Ce sont des effets de relaxation croisée qui correspondent à un transfert d'aimantation entre des noyaux proches, couplés dipolairement. La RMN en solution permet également de mesurer la dispersion de la relaxation magnétique nucléaire des résonances (NMRD)  $^2\text{H}$  et  $^{17}\text{O}$  de l'eau [Sunnerhagen *et al.*, 1998]. Ces deux méthodes complémentaires permettent de connaître les atomes du soluté exposés au solvant et d'estimer le temps de résidence des molécules d'eau à différentes échelles de temps suivant la méthode utilisée :

- la méthode NOE, par le signe des pics de son spectre, permet d'identifier les molécules d'eau ayant un temps de résidence autour de 500 ps, mais seulement celles qui se trouvent proches des protons de la molécule (et pas, par exemple, des groupements phosphates de l'ADN) ;
- la méthode NMRD, par la mesure des fréquences de dispersion, peut déterminer les molécules d'eau liées à la molécule pour un temps de résidence de l'ordre de la nanoseconde jusqu'à plusieurs centaines de microsecondes.

Les méthodes de RMN permettront donc difficilement de caractériser les molécules d'eau de surface qui s'échangent avec le bulk, ou dont le temps de résidence est inférieur à 500 ps. Elles permettent cependant d'identifier la présence et le nombre de molé-

cules d'eau dans une cavité moléculaire ou à l'interface d'un complexe ADN-protéine ou protéine-protéine, ainsi que leurs temps de résidence. Grâce à l'information de ce temps de résidence, elles peuvent nous indiquer si les molécules d'eau sont fortement ordonnées dans ces cavités.

### Résultats obtenus avec les méthodes structurales

Les études par cristallographie ou par RMN d'acides nucléiques et de protéines sont nombreuses. Elles ne sont pas toujours en accord entre elles, puisque certaines molécules d'eau présentes dans les cristaux ne sont pas retrouvées dans les études de RMN, et inversement.

Pour les acides nucléiques, des études de cristallographie sur des molécules d'ADN en forme B ont montré la présence d'un motif d'hydratation appelé « épine d'hydratation » dans le petit sillon [Drew et Dickerson, 1981, Berman, 1994]. Cette épine est constituée de plusieurs molécules d'eau ordonnées qui stabiliseraient la structure en formant des liaisons hydrogène avec les sites donneurs et accepteurs laissés libres des bases et sucres des nucléotides de l'ADN. Les études de RMN ont pu confirmer l'existence d'épines d'hydratation pour les ADN B en solution, puisque quelques molécules d'eau avec un temps de résidence supérieur à 0,5 ns ont pu être identifiées dans le petit sillon [Liepinsh *et al.*, 1994]. Ces études n'ont pas permis de retrouver dans le grand sillon, plus large, un tel ordonnancement de molécules d'eau. De plus, le motif de l'épine d'hydratation serait privilégié par une séquence riche en paires A-T, car le petit sillon est alors plus profond et plus étroit, et les atomes donneurs et accepteurs libres des bases A et T (l'atome N3 de l'adénine et l'atome O2 de la thymine) favoriseraient la formation de cette épine [Kochoyan et Leroy, 1995]. Des études similaires ont été effectuées sur des cristaux d'ADN A [Eisenstein et Shakked, 1995], moins nombreux que ceux d'ADN B. Ces études ont montré que le grand sillon de l'ADN A était très hydraté, que des molécules d'eau formaient des ponts aqueux entre les groupements phosphates, et que ces molécules étaient plus mobiles que celles liées aux bases. Ces études ont également montré que l'arrangement des molécules d'eau autour de l'ADN était différent suivant sa séquence nucléotidique, et que ce schéma d'hydratation pourrait donc intervenir lors de la reconnaissance moléculaire de l'ADN par une protéine [Eisenstein et Shakked, 1995]. Des études cristallographiques de l'hydratation des paires d'ARN ont aussi été effectuées [Auffinger et Westhof, 1998a], identifiant huit sites d'hydratation dans le plan des paires Watson-Crick A-U et G-C, ainsi que des sites d'hydratation des sucres et des phosphates. D'autres études cristallographiques sur les paires d'ADN et d'ARN ont montré que la structure d'hydratation des paires de bases permettait de distinguer les paires Watson-Crick classiques des paires mal

appariées, comme par exemple la paire « wobble » G–U [Sundaralingam et Pan, 2002]. Notons que l’hydratation des acides nucléiques est souvent un processus dynamique, les temps de résidence des molécules d’eau étant assez courts, et les études de RMN sont donc moins nombreuses et moins précises que pour l’hydratation des protéines.

L’hydratation des protéines a été étudiée par de nombreuses expériences de cristallographie et de RMN. Des études se sont intéressées à l’hydratation de la surface des protéines, et en particulier aux effets hydrophobes [Kovacs *et al.*, 1997]. Mais seules quelques molécules d’eau ont été identifiées à la surface de certaines protéines avec les méthodes expérimentales [Otting *et al.*, 1991, Zhang et Matthews, 1994], et, de plus, correspondent rarement d’une méthode à l’autre, empêchant de tirer des conclusions générales sur ce sujet. Comme à la surface des acides nucléiques, les molécules d’eau forment des liaisons hydrogène avec les acides aminés de la protéine, mais encore une fois les temps de résidence sont assez courts [Kovacs *et al.*, 1997]. Les études expérimentales sont plus concluantes pour les molécules d’eau placées dans les cavités des protéines, comme par exemple pour la myoglobine [Makarov *et al.*, 2000], où des molécules d’eau à long temps de résidence ont été localisées par RMN dans des cavités, sauf lorsque les cavités sont en contact avec le bulk [Lounnas et Pettitt, 1994]. Une autre étude de RMN sur différentes protéines [Denisov et Halle, 1996] a cependant montré que, même dans les cavités, les molécules d’eau peuvent être désordonnées, en particulier dans les cavités non polaires. Il est alors difficile de les observer par cristallographie, et seule la RMN permet de les identifier.

Ces méthodes expérimentales permettent aussi d’étudier des complexes protéine-protéine et ADN-protéine. Les molécules d’eau se trouvant à l’interface peuvent avoir une grande importance en agissant comme médiateurs de la reconnaissance moléculaire. On observe que ces molécules d’eau à l’interface des complexes, d’une part complètent les espaces laissés libres lors de la reconnaissance, et d’autre part peuvent stabiliser le complexe en formant des liaisons hydrogène entre les deux molécules, appelées ponts aqueux [Schwabe, 1997]. Des études cristallographiques des molécules seules et complexées pour le complexe du répresseur du *trp* lié à l’ADN [Sunnerhagen *et al.*, 1998] montrent en effet que le schéma d’hydratation de chaque molécule peut être spécifique, car certaines molécules d’eau localisées sont présentes à la fois dans les structures cristallographiques des molécules isolées et du complexe. Les temps de résidence de certaines molécules ordonnées à l’interface mesurés par RMN peuvent varier de la microseconde à la milliseconde, comme l’a montré l’étude du complexe protéine-protéine HIV-1 Protéase/DMP323 [Wang *et al.*, 1996]. Pour d’autres complexes étudiés, comme celui de l’homéodomaine lié à l’ADN [Labeots et Weiss, 1997], les molécules d’eau sont désordonnées et ne servent qu’à remplir l’espace libre, en agissant comme un lubrifiant moléculaire, et seule la RMN peut alors les identifier. Certaines études cristallographiques et de RMN s’intéressent aux molécules

d'eau à l'interface de complexes spécifiques et non spécifiques (par exemple où la molécule cible a subi des mutations ponctuelles dans sa séquence de bases ou d'acides aminés). Le nombre de molécules d'eau observées permet alors de différencier les complexes spécifiques des complexes non spécifiques. C'est le cas par exemple pour le complexe du domaine de liaison à l'ADN du récepteur du glucocorticoïde lié à sa séquence d'ADN cible (GRDBD-ADN), où aucune molécule d'eau ordonnée n'a été identifiée à l'interface, alors que pour un complexe non spécifique, une étude cristallographique a trouvé sept molécules d'eau ordonnées à l'interface [Gewirth et Sigler, 1995]. Pour le cas du complexe spécifique du récepteur de l'œstrogène (ERDBD-ADN), trois molécules d'eau ordonnées ont été identifiées à l'interface [Kosztin *et al.*, 1997].

L'hydratation joue donc un rôle-clé dans la stabilisation de la structure tridimensionnelle des acides nucléiques et protéines, ainsi que lors de la reconnaissance moléculaire, mais des études uniquement structurales ne permettent pas de quantifier et de décrire précisément ces phénomènes. En effet, les méthodes de cristallographie et de RMN ne permettent de détecter que des molécules d'eau fortement localisées (pour la cristallographie aux rayons X), ou fortement immobilisées (pour la RMN), alors que les phénomènes d'hydratation sont souvent des processus très dynamiques, impliquant de nombreuses molécules d'eau, telles que la déshydratation lors d'une complexation.

## 2.1.2 Méthodes de mesures thermodynamiques

### Principe des mesures thermodynamiques

Afin de quantifier et de prédire la spécificité de la formation de complexes macromoléculaires, il est important de savoir mesurer et calculer l'énergie libre de tels systèmes biomoléculaires. En effet, lors de l'interaction entre deux molécules, l'énergie libre du système doit diminuer pour que la formation du complexe soit favorable. Or, l'énergie libre de liaison comprend une contribution importante due aux interactions du soluté avec le solvant et à la perturbation du solvant par le soluté, c'est l'énergie libre de solvatation du complexe. L'énergie libre de liaison s'écrit  $\Delta G = \Delta H - T\Delta S$ , où  $\Delta H$  est l'enthalpie de la réaction,  $T$  la température du système en Kelvin et  $\Delta S$  la variation de l'entropie du système. Par exemple, l'expulsion de molécules d'eau ordonnées lors de la complexation va entraîner une augmentation de l'entropie et aura donc une influence favorable sur la réaction en diminuant l'énergie libre. Au contraire, la mobilité réduite des molécules due à la complexation aura un effet contraire, entropiquement défavorable. Mais, d'autre part, l'enthalpie de réaction qui mesure les changements des interactions non covalentes (donc les interactions électrostatiques et de van der Waals) va diminuer lors de la formation de

liaisons hydrogène entre la molécule et les molécules de solvant. En effet, dans le vide, la formation d'une liaison hydrogène a un effet enthalpique de  $\Delta H$  entre  $-3$  et  $-6$  kcal/mol, alors que l'immobilisation d'une molécule d'eau à 300 K a un coût entropique de  $T\Delta S$  entre 0 et 2 kcal/mol [Schwabe, 1997]. Donc la formation d'une liaison hydrogène peut compenser l'immobilisation d'une molécule d'eau à l'interface. Si on veut étudier l'influence de la solvation sur la complexation des biomolécules, il apparaît donc nécessaire de pouvoir accéder à l'énergie libre de solvation et à ses variations lors de processus biologiques.

### Les mesures de microcalorimétrie

Les études de microcalorimétrie sont particulièrement adaptées pour mesurer ces variations enthalpiques et entropiques lors de la complexation des biomolécules. Avec de telles méthodes, les expérimentateurs mesurent la chaleur échangée, qui donne la variation enthalpique  $\Delta H$ , et ils mesurent également la constante d'affinité  $K_a$  (ou parfois la constante de dissociation  $K_d = 1/K_a$ ), reliée à l'énergie libre de liaison par la relation  $\Delta G = -RT \ln K_a$ . Connaissant  $\Delta H$  et  $\Delta G$ , on peut donc en déduire  $\Delta S$  et ainsi distinguer les effets enthalpiques des effets entropiques lors de la complexation [Fields *et al.*, 1996]. Les études effectuées à différentes températures pour un même complexe montrent que  $\Delta G$  dépend peu de la température, mais que les contributions enthalpiques et entropiques en dépendent beaucoup [Schwabe, 1997]. Des études thermodynamiques de différents complexes, comme sur le complexe du répresseur met lié à l'ADN [Hyre et Spicer, 1995], ont montré qu'à basse température les effets entropiques prédominent, alors qu'à haute température ce sont les effets enthalpiques qui sont majoritaires. Ces contributions diffèrent aussi selon les complexes, et il ne semble pas y avoir de règle générale. Certaines complexations s'opèrent en éjectant les molécules d'eau ordonnées à la surface des macromolécules, elles sont alors entropiquement favorables, mais la perte de liaisons hydrogène est enthalpiquement défavorable. C'est le cas du complexe GRDBD-ADN, où aucune molécule d'eau ordonnée n'a été identifiée à l'interface, et pour lequel des mesures de microcalorimétrie ont montré que la formation du complexe, globalement favorable, était entropiquement favorable mais enthalpiquement défavorable [Lündback et Härd, 1996]. Au contraire, d'autres complexations s'effectuent en ordonnant un certain nombre de molécules d'eau à l'interface, elles sont alors entropiquement défavorables, mais enthalpiquement favorables grâce à la formation de liaisons hydrogène, comme pour le complexe du répresseur trp lié à l'ADN [Ladbury *et al.*, 1998].

La compensation de ces deux effets contraires est relativement difficile à prévoir, et les énergies libres de liaison obtenues avec ces méthodes de microcalorimétrie dépendent

fortement des conditions de l'expérience. De plus, on ne peut avec ces seules mesures distinguer précisément la contribution de la solvation dans l'énergie libre de liaison. Il semble donc nécessaire de coupler ces méthodes avec d'autres approches afin de rationaliser ces effets de solvation sur la complexation.

### Les mesures basées sur la pression osmotique

Les mesures de calorimétrie peuvent être couplées avec des approches basées sur le principe de la pression osmotique. On peut analyser expérimentalement l'effet du stress osmotique sur une interaction entre deux macromolécules. En effet, dans une cavité aqueuse inaccessible aux cosolutés neutres (appelés osmolytes, tels que le glycérol, l'éthanol,...), l'activité de l'eau<sup>1</sup> est contrôlée par la pression osmotique, qui peut être régulée par la concentration en cosolutés. Une augmentation de la pression osmotique, encore appelée stress osmotique, engendre une diminution de l'activité de l'eau, les molécules d'eau liées dans les cavités sont alors rejetées dans le bulk, et les conformations macromoléculaires sont déplacées vers leur état le plus déshydraté. On peut aussi observer ce phénomène dans le cas où l'eau est localisée à la surface des macromolécules, car les surfaces sont préférentiellement hydratées et excluent les cosolutés neutres. Ces mesures sont également fréquemment effectuées sur des complexes spécifiques et non spécifiques, afin de mesurer l'influence de l'eau dans la spécificité de la reconnaissance moléculaire. Les méthodes basées sur la pression osmotique doivent être couplées à des mesures énergétiques, par calorimétrie ou par analyse de migrations sur des gels par électrophorèse, afin de mesurer par exemple la constante de dissociation du complexe à différentes pressions osmotiques. Puis, par interpolation, on trouve une relation entre la constante d'affinité (ou de dissociation) et la pression osmotique. On peut déduire de cette relation le changement de volume  $\Delta V$  lié à la formation du complexe, d'après la formule reliant  $\Delta V$  à la pression osmotique  $\pi$ , à la constante de dissociation  $K_d$  et à la température  $T$  :  $\partial \ln K_d / \partial \pi = \Delta V / (RT)$ , où  $R$  est la constante des gaz parfaits. Puis on relie directement  $\Delta V$  au nombre de molécules d'eau relâchées d'après la densité de l'eau [Robinson et Sligar, 1998].

Souvent le stress osmotique va favoriser la formation d'un complexe, car la déshydratation accompagne généralement la complexation, comme pour le complexe du répresseur gal lié à l'ADN, pour lequel de nombreuses molécules d'eau sont relâchées lors de la formation du complexe [Garner et Rau, 1995]. Mais, dans les cas plus rares où la liaison de molécules d'eau accompagne la formation du complexe, le stress osmotique peut défavoriser l'interaction, comme on le verra dans les deux exemples suivants. Ainsi, des mesures

---

<sup>1</sup>l'activité de l'eau est la proportion d'eau non liée chimiquement aux cosolutés, donc libre et disponible pour l'hydratation de la molécule.

de pressions osmotiques ont été effectuées pour étudier l'enzyme de restriction endonucléase EcoRI lié à l'ADN, qui coupe l'ADN à un endroit précis repéré par une séquence de bases spécifique [Robinson et Sligar, 1993]. Ce complexe a été étudié avec l'ADN cible canonique, mais aussi avec un ADN de séquence ne différant de la séquence spécifique que par une paire de bases, appelé ADN alternatif, et enfin avec un ADN non spécifique de séquence complètement différente [Robinson et Sligar, 1998]. Les mesures de la constante de dissociation à différentes pressions osmotiques pour les trois complexes ont montré que le nombre de molécules d'eau relâchées lors de l'interaction était très différent. La séquence spécifique liée à l'enzyme relâche plus de molécules d'eau que la séquence non spécifique, ce qui confirme que la déshydratation permettrait la reconnaissance. Mais la complexation de la séquence spécifique avec EcoRI relâche 70 molécules d'eau de moins que la séquence alternative, et l'interface avec la séquence alternative est donc plus déshydratée que la séquence spécifique lors de la complexation. Certaines molécules d'eau seraient donc indispensables pour la reconnaissance entre les deux molécules du complexe spécifique [Robinson et Sligar, 1998]. Cette étude renforce l'idée que l'eau liée à l'interface des complexes joue un rôle très important pour la spécificité de la reconnaissance moléculaire. Le complexe le plus déshydraté n'est pas forcément le complexe spécifique, et des molécules d'eau peuvent donc faire partie intégrante de la reconnaissance et de l'interaction.

Le même genre d'étude a été effectué sur des complexes protéine-protéine tels que des systèmes antigène/anticorps [Goldbaum *et al.*, 1996]. Pour deux complexes spécifiques composés chacun d'un anticorps différent lié au même lysozyme, des analyses des constantes d'affinité et des enthalpies de réaction ont été effectuées à différentes pressions osmotiques. La diminution de l'activité de l'eau par stress osmotique entraîne une diminution de la constante d'association et de l'enthalpie de réaction pour l'un des deux complexes spécifiques, alors que l'on observe l'effet contraire pour l'autre complexe, pour lequel des molécules d'eau sont relâchées lors de la complexation. Pour le premier complexe spécifique, les molécules d'eau semblent donc contribuer à la stabilité du complexe en se plaçant dans des poches hydrophiles à l'interface. Pour le second, la formation du complexe s'accompagne d'une déshydratation à l'interface [Goldbaum *et al.*, 1996]. L'eau à l'interface peut donc avoir un rôle très différent pour différents complexes spécifiques.

Enfin, des études similaires utilisant la pression hydrostatique associée à la migration sur gels par électrophorèse peuvent apporter le même genre d'informations. D'après la relation donnant la constante de dissociation  $K_d$  en fonction de la pression hydrostatique, on en déduit le volume  $\Delta V$  accompagnant la complexation par le même type de relation que pour la pression osmotique [Lynch et Sligar, 2002]. Selon le signe de  $\Delta V$ , on en déduit si des molécules d'eau sont relâchées ou au contraire liées pendant la complexation.

Les méthodes de pression osmotique permettent donc de compter le nombre de molécules d'eau relâchées lors de la complexation, mais elles sont moins utiles pour étudier l'hydratation des molécules isolées. D'autres méthodes, comme les mesures volumétriques, vont tenter de mesurer les effets de la solvatation sur des molécules non complexées.

### Les études volumétriques

Les études volumétriques effectuées à l'aide de méthodes très sensibles permettent d'obtenir de nouvelles informations sur le rôle de l'hydratation pour les macromolécules [Chalikian et Breslauer, 1998]. Ces études volumétriques reposent sur des mesures du volume molaire partiel  $V^\circ$ , volume apparent occupé par une mole de soluté dans une dilution infinie, et sur les mesures de la compressibilité adiabatique molaire partielle  $K_S^\circ$ , compressibilité adiabatique apparente d'une mole de soluté infiniment diluée. On décompose  $V^\circ$  en deux termes,  $V_M$ , le volume intrinsèque de soluté (somme des volumes de van der Waals de tous les atomes du soluté et des vides) et  $\Delta V_h$ , le changement de volume du solvant induit par les interactions soluté-solvant dues à l'hydratation du soluté. Ce volume est évalué par des mesures différentielles de densité sur le soluté dans le solvant, et sur le solvant seul. Quant à  $K_S^\circ$ , elle peut être décomposée en deux termes de façon analogue, et évaluée par des mesures différentielles de la vitesse du son sur le soluté en solution et sur le solvant seul. L'interprétation des résultats de ces expériences apporte des informations sur l'existence et les propriétés de réseaux d'hydratation à la surface des macromolécules, plus structurés que le bulk [Chalikian et Breslauer, 1998].

Pour les protéines, même pour les groupements atomiques exposés au solvant, les études volumétriques de petits composés, plus simples à mettre en œuvre, ne peuvent pas être généralisées, car des réseaux complexes et structurés de molécules d'eau se forment à la surface polaire des protéines, réseaux qu'on ne peut pas retrouver sur les groupements atomiques seuls [Chalikian et Breslauer, 1998]. D'autres études volumétriques ont montré que l'eau des premières couches d'hydratation des protéines globulaires était moins compressible que celle du bulk, et que sa densité était plus élevée à la surface des protéines que dans le bulk [Murphy *et al.*, 1998]. Sur les acides nucléiques, les études sont plus rares mais ont montré que les double hélices d'ADN les moins hydratées étaient celles contenant environ 55 à 60 % de paires A-T [Chalikian *et al.*, 1994b]. L'alternance des paires A-T et G-C serait à l'origine de cette faible hydratation, car les séquences répétées de paires A-T ou G-C entraînent des structures d'hydratation connues, comme par exemple l'épine d'hydratation dans le petit sillon. Ces études tendent donc à montrer que l'on ne peut pas estimer l'hydratation de l'ADN comme une somme pondérée de contributions des paires A-T et G-C isolées [Chalikian *et al.*, 1994b]. Des études volumétriques sur les complexes

acides nucléiques/ligands permettent également, comme les études de pression osmotique, de quantifier le nombre de molécules d'eau relâchées à lors de la formation du complexe [Chalikian *et al.*, 1994a].

Les études volumétriques sont donc intéressantes pour mieux décrire la structure d'hydratation des premières couches de solvation des macromolécules. Elles tendent à prouver que ces structures d'hydratation sont spécifiques à la conformation tridimensionnelle globale de chaque macromolécule, et que les propriétés thermodynamiques de ces premières couches d'hydratation diffèrent de celles du bulk [Chalikian et Breslauer, 1998].

### 2.1.3 Conclusion sur les méthodes expérimentales

Toutes les méthodes expérimentales présentées ici apportent de nombreuses informations diverses et complémentaires sur l'hydratation des acides nucléiques, des protéines, et de leurs complexes. Elles confirment que les molécules d'eau sont indispensables à la stabilité des structures tridimensionnelles des biomolécules, et jouent aussi un rôle majeur lors des phénomènes de reconnaissance moléculaire. Alors que les méthodes structurales caractérisent quelques molécules d'eau peu mobiles à la surface ou à l'interface des macromolécules, les méthodes énergétiques donnent des informations globales sur la thermodynamique de ces systèmes. Mais les observables thermodynamiques donnent des grandeurs macroscopiques dont l'interprétation microscopique, c'est-à-dire en terme de molécules d'eau individuelles, n'est pas forcément évidente et dépend du modèle choisi pour décomposer les grandeurs mesurées. C'est pourquoi il faut souvent combiner ces différentes méthodes entre elles pour pouvoir tirer des conclusions précises sur chaque système étudié. Précisons que toutes les méthodes énergétiques ne peuvent s'appliquer qu'à des complexes très stables à différentes conditions expérimentales, et sont difficiles à mettre en œuvre pour des complexes peu stables.

Les méthodes expérimentales permettant d'étudier l'hydratation sont actuellement de plus en plus précises. Cependant, elles posent des difficultés d'interprétation, et elles ne permettent le plus souvent que d'obtenir des informations partielles sur l'hydratation des macromolécules biologiques. Il est donc nécessaire d'envisager des méthodes théoriques permettant d'étudier de façon plus exhaustive les processus de solvation et de confirmer et compléter les résultats issus des méthodes expérimentales.

## 2.2 Étude de la solvation par les méthodes théoriques

Les méthodes théoriques, telles que la modélisation moléculaire [Leach, 1996, Frenkel et Smit, 1996, Allen et Tidesley, 1987], permettent aujourd’hui de préciser au niveau atomique et de mieux rationaliser la relation structure/fonction des systèmes biologiques complexes dont les études expérimentales ne sont pas toujours simples à mettre en œuvre, ni évidentes à interpréter. Comme nous l’avons évoqué précédemment, l’eau joue un rôle très important dans la structure des édifices biomoléculaires et il apparaît donc essentiel de bien représenter le solvant autour des solutés dans les études de modélisation moléculaire. Deux grandes stratégies se distinguent pour représenter le solvant en modélisation, les méthodes explicites, qui représentent le solvant de façon microscopique, et les méthodes implicites, qui représentent les effets du solvant de façon macroscopique. Plus récemment, de nouvelles méthodes implicites « hybrides » sont apparues, exploitant l’efficacité des méthodes implicites sans négliger l’aspect moléculaire et microscopique des particules de solvant.

### 2.2.1 Méthodes explicites

Les méthodes explicites représentent chaque molécule d’eau autour du soluté comme une molécule triatomique  $\text{H}_2\text{O}$ , de charges et rayons atomiques donnés, comme on l’a représenté schématiquement sur la figure 2.2. L’avantage d’une telle méthode est que soluté et solvant sont tous deux représentés de façon cohérente, par une description atomique. La dynamique des molécules d’eau peut alors être calculée par les méthodes classiques de dynamique moléculaire que nous verrons plus précisément dans le chapitre suivant. Une telle méthode permet d’étudier précisément la trajectoire de toutes les molécules d’eau en parallèle à l’évolution de la structure du soluté. De plus, elle permet d’observer directement les liaisons hydrogène que chaque molécule d’eau peut former avec le soluté, puisque chaque atome H et O est représenté explicitement. La représentation explicite a aussi l’avantage de déterminer naturellement la frontière entre le soluté et le solvant par les rayons de van der Waals des atomes du système.

Il existe différents modèles explicites pour représenter l’eau en modélisation moléculaire, car divers paramètres peuvent être ajustés : le nombre de sites (de 3 jusqu’à 5 sites, afin de mieux représenter la structure tétraédrique de l’eau à cause des deux doublets électroniques de son atome d’oxygène [Mahoney et Jorgensen, 2000]), les charges et rayons de ces sites, la longueur des liaisons et la valeur des angles de valence entre les

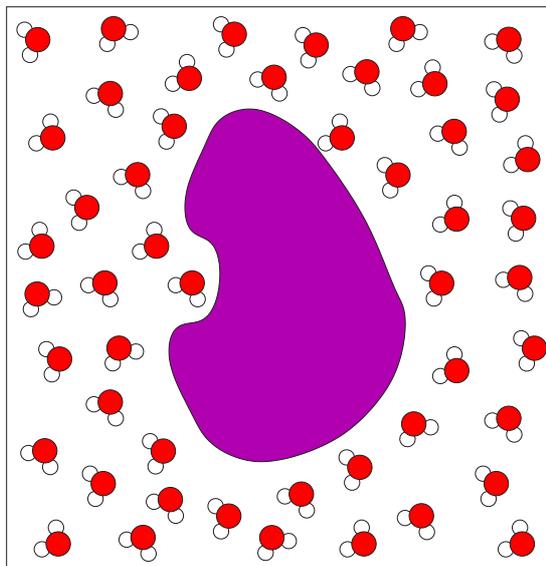


FIG. 2.2 : Représentation explicite du solvant : chaque molécule d'eau triatomique est représentée autour du soluté.

sites [Jorgensen *et al.*, 1983, Berendsen *et al.*, 1987]. Les molécules d'eau peuvent être considérées flexibles ou rigides, selon le degré de précision désirée.

Les deux modèles les plus utilisés pour les simulations de dynamique moléculaire des acides nucléiques et protéines sont TIP3P et SPC (ou SPC/E), deux modèles plans à trois sites, le plus souvent rigides. Les paramètres de charges, de rayons et d'angles sont légèrement différents pour les deux modèles, quelques-uns de ces paramètres étant présentés dans le tableau 2.1. Le paramétrage est choisi pour reproduire des données expérimentales sur l'eau telles que la fonction de distribution radiale  $g_{OO}$  issue des expériences de rayons X, l'énergie de dimérisation, ou le nombre de liaisons hydrogène que la molécule d'eau peut former [Jorgensen *et al.*, 1983]. Ces deux modèles utilisés à température et pression ambiante sont généralement en bonne adéquation avec les données expérimentales. Néanmoins, l'analyse de plusieurs simulations de systèmes comportant un nombre différent de molécules d'eau et des conditions de simulation différentes (densité, traitement des interactions à longue distance) montre quelques différences notables entre les modèles TIP3P et SPC quant au coefficient de diffusion et à la constante diélectrique effective [van der Spoel *et al.*, 1998] (autour de 60–70 pour SPC/E et autour de 90 pour TIP3P). Les deux modèles reproduisent bien l'arrangement des molécules d'eau dans les premières couches d'hydratation autour des macromolécules biologiques, qui se différencient des molécules d'eau du bulk [Bizzarri et Cannistraro, 2002]. Récemment, l'étude comparée des deux modèles TIP3P et SPC/E pour la simulation du lysozyme [Marchi *et al.*, 2002] montre des différences minimales entre les deux modèles quant au coefficient de diffusion ou au

temps de relaxation à la surface de la protéine. Ces deux modèles semblent donc donner des résultats assez comparables pour les simulations de macromolécules biologiques.

<i>Modèle</i>	$r_{OH}$	$q_0(e)$	$q_H(e)$	<i>angle HOH</i>	<i>Valeur du dipôle</i>
<b>SPC</b>	1,0 Å	-0,82	0,41	109,47°	2,27 Debye
<b>SPC/E</b>	1,0 Å	-0,8476	0,4238	109,47°	2,39 Debye
<b>TIP3P</b>	0,9572 Å	-0,834	0,417	104,52°	2,35 Debye

TAB. 2.1 : Valeurs de quelques paramètres des modèles SPC, SPC/E et TIP3P de l'eau.

Les modèles de solvant explicites permettent de représenter l'eau de façon réaliste, grâce à des modèles paramétrés en accord avec l'expérience. Cependant, pour un système avec des conditions aux bords périodiques comme on utilise habituellement en modélisation pour simuler un milieu infini, on devra considérer un très grand nombre de molécules d'eau, jusqu'à plusieurs dizaines de milliers. Le calcul des interactions solvant/solvant va prendre la majorité du temps de calcul par rapport aux interactions soluté/solvant et soluté/soluté, alors que la trajectoire exacte de toutes les molécules d'eau n'est pas l'information la plus intéressante des simulations.

De plus, en ce qui concerne les propriétés diélectriques et thermodynamiques, comme par exemple l'énergie libre de solvation, il est nécessaire de faire des moyennes sur un très grand nombre de configurations différentes afin d'avoir une statistique suffisante pour obtenir une moyenne représentative des systèmes étudiés, et il faut utiliser des méthodes de physique statistique évoluées (méthodes de perturbation ou « umbrella sampling » [Kollman, 1993, Frenkel et Smit, 1996]) pour calculer les grandeurs thermodynamiques. Or la convergence de ces propriétés est connue pour être très lente [Bottcher, 1973], et l'étude des propriétés thermodynamiques d'un système biomoléculaire par les méthodes de solvation explicites sera donc assez peu précise et très coûteuse en temps de calcul.

## 2.2.2 Méthodes implicites

Afin de calculer les grandeurs thermodynamiques de solvation du système de façon beaucoup plus avantageuse, des méthodes de solvation appelées implicites ont été développées [Roux et Simonson, 1999]. Elles reposent sur le fait que les effets de l'eau sur les molécules sont principalement d'ordre électrostatique, et qu'ils peuvent être moyennés dans le temps et l'espace. Le solvant est alors représenté comme un milieu continu, sans structure microscopique, de constante diélectrique  $\epsilon_s$  élevée, dont les propriétés diélectriques découlent des lois macroscopiques de l'électrostatique [Honig *et al.*, 1993, Simon-

son, 2001, 2003]. Le soluté est une cavité de faible constante diélectrique immergée dans le solvant, avec des charges partielles fixées (cf. figure 2.3). L'intérêt de ces méthodes est qu'elles calculent des grandeurs moyennes et permettent donc de traiter directement les interactions soluté/solvant en terme d'énergie libre. Dans le cas de trajectoires de dynamique moléculaire avec un solvant continu, le but n'est plus de reproduire des trajectoires réalistes comme avec un solvant explicite, mais d'explorer l'espace conformationnel de façon efficace, en plus du calcul de l'énergie libre de solvation.

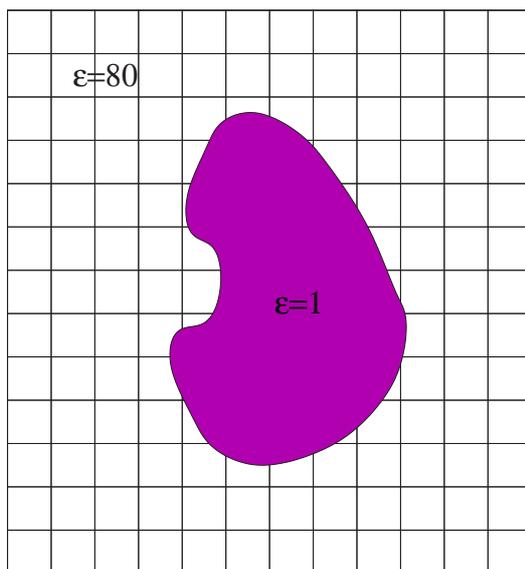


FIG. 2.3 : Représentation implicite du solvant : le solvant est un milieu continu de constante diélectrique  $\epsilon_s = 80$  et le soluté un milieu de constante diélectrique  $\epsilon_i = 1$ .

Les méthodes de solvation implicites permettent donc de calculer l'énergie libre de solvation  $\Delta G_S$ , que l'on peut décomposer en deux contributions de la façon suivante :  $\Delta G_S = \Delta G^p + \Delta G^{np}$ . Le terme  $\Delta G^p$  est la contribution électrostatique ou polaire à l'énergie libre de solvation, dont nous présenterons les différentes méthodes de calcul par la suite. Le terme  $\Delta G^{np}$  représente la contribution non polaire à l'énergie libre de solvation, il comprend donc les interactions de van der Waals  $\Delta G_{vdW}$  entre le solvant et le soluté, et l'énergie libre nécessaire pour former la cavité dans le solvant  $\Delta G_{cav}$ . La méthode la plus répandue pour calculer  $\Delta G^{np}$  repose sur le calcul de la surface accessible au solvant de la macromolécule étudiée (méthode appelée SASA pour « Solvent-Accessible Surface Area ») [Honig et Nicholls, 1995, Roux et Simonson, 1999, Fologari *et al.*, 2001]. Cette méthode considère que l'énergie de solvation non polaire provient de la première couche de solvation, et que le nombre de molécules d'eau de cette première couche est proportionnel à la surface accessible au solvant  $A$  de la macromolécule étudiée. Cette surface est calculée en tenant compte des volumes de van der Waals des atomes de la

molécule exposés au solvant, ainsi que des volumes des molécules d'eau de la première couche de solvation. On exprime alors l'énergie de solvation non polaire par la formule :

$$\Delta G^{np} = \Delta G_{vdW} + \Delta G_{cav} = \gamma A + b \quad (2.1)$$

Les paramètres  $\gamma$  (représentant la tension de surface microscopique) et  $b$  sont obtenus par interpolation linéaire à partir du calcul d'énergie de solvation de différents hydrocarbures dont l'énergie de solvation est majoritairement non polaire et connue expérimentalement.

Parmi les méthodes implicites qui permettent le calcul de la contribution électrostatique de l'énergie libre de solvation, on distingue plusieurs approches, dont nous présenterons brièvement les principes fondamentaux :

1. les méthodes de continuum considèrent le solvant comme un milieu continu de constante diélectrique élevée : on trouve d'une part les méthodes basées sur l'équation de Poisson-Boltzmann, et d'autre part celles basées sur les interactions de paires effectives ;
2. les méthodes hybrides prennent en compte l'aspect moléculaire du solvant tout en restant implicites : nous verrons d'une part celles des dipôles de Langevin, qui traitent le solvant comme un fluide dipolaire, dont les équations sont basées sur les propriétés de polarisation du solvant, et d'autre part les méthodes reposant sur la théorie de la fonctionnelle de la densité des liquides, qui prennent en compte les variations de la densité du solvant.

### Méthodes basées sur l'équation de Poisson-Boltzmann

Pour les méthodes basées sur l'équation de Poisson-Boltzmann, le solvant est considéré comme un milieu continu de constante diélectrique  $\epsilon_s = 80$ , le soluté étant une cavité de constante diélectrique  $\epsilon_i = 1$  à 4, et l'ensemble obéissant à la théorie macroscopique de Poisson-Boltzmann [Jackson, 1999]. Lorsque qu'on ne considère aucun ion en solution, le potentiel électrostatique en tout point de l'espace  $\phi(\mathbf{r})$  obéit à l'équation de Poisson, donnée par l'équation 2.2, qui relie ce potentiel à la densité de charges fixes du soluté  $\rho(\mathbf{r})$  et à la constante diélectrique (qui peut varier en fonction de la distance)  $\epsilon(\mathbf{r})$  [Jackson, 1999] :

$$\nabla \cdot [\epsilon(\mathbf{r}) \nabla \phi(\mathbf{r})] + 4\pi \rho(\mathbf{r}) = 0. \quad (2.2)$$

Lorsque des ions mobiles sont présents en solution, c'est-à-dire lorsque la densité de charge ionique est non nulle, la distribution de ces ions est alors modélisée par une

statistique de Boltzmann, et le potentiel électrostatique du système obéit alors à l'équation de Poisson-Boltzmann, donnée par l'équation 2.3 dans le cas d'un sel monovalent ( $z^+ = z^- = 1$ ) :

$$\nabla \cdot [\varepsilon(\mathbf{r})\nabla\varphi(\mathbf{r})] - \varepsilon(\mathbf{r})\kappa(\mathbf{r})^2 \sinh[\varphi(\mathbf{r})] + \frac{4\pi e}{kT}\rho(\mathbf{r}) = 0, \quad (2.3)$$

où  $\varphi = e\phi/kT$  est le potentiel électrostatique sans dimension,  $e$  la charge du proton,  $k$  la constante de Boltzmann,  $T$  la température du système, et  $\kappa$  est la constante de Debye-Hückel (inverse de la longueur d'écrantage de Debye-Hückel), donnée par la relation :

$$\kappa(\mathbf{r})^2 = \frac{8\pi I e^2}{\varepsilon(\mathbf{r})kT}, \quad (2.4)$$

où  $I = \frac{1}{2} \sum_{ions} [i] \cdot z_i^2$  est la force ionique ( $[i]$  représente la concentration de l'ion d'espèce  $i$  de charge  $z_i$ ). Notons qu'en l'absence d'ions, la force ionique est évidemment nulle. On a donc  $\kappa = 0$  et on retrouve alors l'équation de Poisson donnée par 2.2.

Dans le cas d'une faible concentration ionique, on préférera utiliser la forme linéarisée de l'équation de Poisson-Boltzmann ( $\sinh(\varphi) \simeq \varphi$ ), plus facile à manipuler que la forme exacte :

$$\nabla \cdot [\varepsilon(\mathbf{r})\nabla\phi(\mathbf{r})] - \varepsilon(\mathbf{r})\kappa(\mathbf{r})^2\phi(\mathbf{r}) + 4\pi\rho(\mathbf{r}) = 0. \quad (2.5)$$

L'énergie libre de solvation du soluté de densité de charges  $\rho(\mathbf{r})$  est calculée à partir du potentiel  $\phi(\mathbf{r})$  solution de l'équation de Poisson-Boltzmann et du potentiel  $\phi_0(\mathbf{r})$  créé par le soluté dans le vide, par la formule :

$$\Delta G^p = \frac{1}{2} \left( \int \rho(\mathbf{r})\phi(\mathbf{r})d\mathbf{r} - \int \rho(\mathbf{r})\phi_0(\mathbf{r})d\mathbf{r} \right). \quad (2.6)$$

L'équation de Poisson-Boltzmann, si elle est un bon modèle pour le potentiel électrostatique, ne peut être résolue analytiquement que pour des cas à géométrie simple pour le soluté (sphère, cylindre,...). Pour le cas général, cette équation ne peut être résolue que par des méthodes numériques approximatives. Ces méthodes doivent d'abord être testées sur les cas à géométrie simple dont la solution analytique est connue afin d'être validées [Harvey, 1989]. Les méthodes numériques utilisées pour résoudre cette équation sont celles des différences finies (« Finite Difference Poisson-Boltzmann » - FDPB) ou des éléments finis (« Finite Element Method » - FEM).

La méthode des différences finies est la plus couramment utilisée et elle a été implémentée dans plusieurs programmes commerciaux tels que DelPhi [Honig *et al.*, 1993,

Honig et Nicholls, 1995] et UHBD [Gilson *et al.*, 1995]. Elle consiste à découper le système par un maillage cubique régulier appliqué sur le soluté et le solvant. En chaque point de la grille, on assigne une valeur à tous les paramètres de l'équation de Poisson-Boltzmann (densité de charge, constante diélectrique, force ionique), et on détermine le potentiel électrostatique en chaque point en fonction du potentiel des proches voisins par une méthode itérative. Pour une telle méthode, il faut déterminer précisément la frontière soluté/solvant, grâce aux données des rayons de van der Waals des atomes du soluté. Le choix de la finesse de la grille doit être un compromis entre la précision des résultats et un temps de calcul raisonnable. La technique de focalisation permet d'optimiser le choix du maillage selon la taille du système étudié.

L'approche basée sur l'équation de Poisson-Boltzmann néglige la nature moléculaire du solvant, ainsi que la taille finie des ions et les effets de corrélations entre ions, mais elle donne pourtant des résultats tout à fait cohérents avec les résultats expérimentaux pour les énergies de solvation. Cela confirme l'idée que beaucoup des effets de la solvation peuvent être traités précisément en ignorant la nature moléculaire du solvant. Les propriétés du solvant sont décrites par des valeurs moyennes, alors que le soluté est traité avec ses détails atomiques. Cependant, il faut souvent déterminer la constante diélectrique à l'intérieur du soluté qui ne vaut pas toujours 1, et cela nécessite un paramétrage délicat, en particulier dans le cas des protéines [Harvey, 1989, Simonson, 2003].

Le principal désavantage de ces méthodes est qu'elles sont limitées à l'étude de la solvation de molécules figées dans une structure donnée. Pour pouvoir étudier la dynamique des macromolécules par ces méthodes, il faut dériver les forces associées aux énergies de solvation issues de la résolution numérique de l'équation de Poisson-Boltzmann. La méthode UHBD a été implémentée dans un programme de dynamique moléculaire [Gilson *et al.*, 1995], où on ne peut pas réactualiser l'énergie de solvation à chaque pas afin de gagner du temps de calcul. D'autres méthodes analogues existent, par exemple celles basées sur l'évaluation des forces par différences finies [Im *et al.*, 1998]. Mais le problème de toutes ces méthodes est qu'elles conduisent souvent à une instabilité de l'énergie du système, due à la discontinuité des forces induite par la discrétisation du maillage.

Les méthodes basées sur l'équation de Poisson-Boltzmann constituent aujourd'hui une référence pour le calcul des énergies libres électrostatiques de solvation, car elles ont été testées et comparées aux valeurs expérimentales avec succès. Généralement, les autres méthodes de solvation continues sont comparées avec les résultats issus des FDPB avant d'être validées. Cependant, elles sont utilisées principalement pour des conformations statiques de macromolécules, puisque leur implémentation dans les programmes de dynamique moléculaire pose de nombreux problèmes. C'est pourquoi d'autres méthodes

de solvation continues doivent être envisagées pour être implémentées de façon efficace dans les programmes de dynamique moléculaire.

### Méthodes basées sur une fonctionnelle d'énergie libre

La méthode de solvation implicite développée par Marchi *et al.* [2001] n'est pas directement basée sur l'équation de Poisson-Boltzmann, mais sur la théorie électrostatique de Marcus [Marcus, 1956]. Cette théorie repose sur une fonctionnelle  $\mathcal{F}_{el}$  d'énergie libre électrostatique de la densité de polarisation du milieu  $\mathbf{P}(\mathbf{r})$ , qui est minimale à l'équilibre thermodynamique. L'avantage de cette méthode est que la fonctionnelle minimisée, donc à l'équilibre, donnera directement l'expression de l'énergie libre de solvation du soluté. Le soluté est considéré comme une distribution de charges gaussiennes  $\rho(\mathbf{r})$ , créant dans le vide un potentiel  $\phi_0(\mathbf{r})$ . La fonctionnelle s'écrit alors [Marchi *et al.*, 2001] :

$$\begin{aligned} \mathcal{F}_{el}[\mathbf{P}(\mathbf{r})] = & \frac{1}{2} \int \rho(\mathbf{r})\phi_0(\mathbf{r})d\mathbf{r} + \frac{1}{2} \int \mathbf{P}(\mathbf{r}) \cdot (\chi(\mathbf{r})^{-1})\mathbf{P}(\mathbf{r}) d\mathbf{r} \\ & - \int (\nabla \cdot \mathbf{P}(\mathbf{r}))\phi_0(\mathbf{r})d\mathbf{r} + \frac{1}{2} \int \int \frac{(\nabla \cdot \mathbf{P}(\mathbf{r}))(\nabla \cdot \mathbf{P}(\mathbf{r}'))}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' \quad (2.7) \end{aligned}$$

La minimisation de la fonctionnelle par rapport à la polarisation donne des équations constitutives qui permettent de retrouver l'équation de Poisson. Cette méthode est donc équivalente aux méthodes basées sur l'équation de Poisson. L'intérêt de cette fonctionnelle est qu'elle a un sens hors équilibre, et qu'elle peut être généralisée pour tenir compte de la polarisation orientationnelle et électronique du milieu. Nous verrons également un peu plus loin une méthode hybride dont la fonctionnelle est très proche et qui prend en compte la densité du solvant. De plus, le fait de chercher le minimum de la fonctionnelle va permettre d'utiliser des techniques de type Car-Parinello [Car et Parinello, 1985] pour la dynamique, au lieu des méthodes classiques de résolution itérative.

Afin de minimiser la fonctionnelle, la polarisation du solvant  $\mathbf{P}(\mathbf{r})$  est développée selon une expansion pseudo-spectrale sur une grille en séries discrètes de Fourier. On écrit  $\mathbf{P}(\mathbf{u}) = \sum_{\mathbf{g}} \mathbf{P}(\mathbf{g})e^{-i\mathbf{g}\cdot\mathbf{u}}$ , où  $\mathbf{u} = \mathbf{r}_1, \dots, \mathbf{r}_N$  et  $\mathbf{g} = \mathbf{g}_1, \dots, \mathbf{g}_N$  sont des ensembles de points régulièrement espacés sur une grille respectivement dans l'espace direct et dans l'espace réciproque. Cette décomposition permet d'utiliser des algorithmes de transformées de Fourier rapides pour résoudre le problème électrostatique. Ainsi, les coefficients  $\mathbf{P}(\mathbf{g})$  de l'expansion de  $\mathbf{P}(\mathbf{r})$  vont devenir les paramètres variationnels de la minimisation de la fonctionnelle discrétisée sur grille. Dans le but d'une intégration dans un programme de dynamique moléculaire, la polarisation est traitée comme une variable dynamique du système avec une masse virtuelle, ses coordonnées dans l'expansion pseudo-spectrale étant traitées par une méthode de type Car-Parinello [Car et Parinello, 1985]. À la limite de

l'évolution adiabatique des variables du champ de polarisation, cette méthode permet de donner une solution du problème continu à l'équilibre alors que les coordonnées du soluté évoluent. Le milieu diélectrique autour du soluté sera donc à chaque pas à l'équilibre thermodynamique.

Cette méthode prometteuse ne permet pas encore de traiter les effets de force ionique de type Poisson-Boltzmann, et est assez coûteuse en temps de calcul à chaque pas de dynamique. Cependant, on peut considérer que le temps des algorithmes d'intégration n'a pas la même valeur que pour les simulations en solvant explicite, puisque le solvant est toujours à l'équilibre thermodynamique. Les énergies de solvation obtenues avec cette méthode sont en très bon accord avec celles issues des méthodes FDPB. De plus, cette méthode a permis d'effectuer des dynamiques conservatives sur de petits peptides en solution en reproduisant correctement leurs propriétés électrostatiques et structurales [Lévy, 2002, Borgis *et al.*, 2003].

### Méthodes d'interactions de paires effectives

Les méthodes d'interactions de paires effectives reposent sur la propriété d'écrantage par l'eau des interactions électrostatiques du soluté. En effet, il est connu que deux charges immergées dans un milieu diélectrique ont une interaction coulombienne modifiée par la constante diélectrique du milieu. Nous présenterons ici les deux principales catégories de méthodes d'interaction de paires effectives, les méthodes de « diélectrique dépendant de la distance » (DDD), et la méthode de Born Généralisé.

#### •Les méthodes de Diélectrique Dépendant de la Distance (DDD)

Les méthodes de « Diélectrique Dépendant de la Distance » (DDD) représentent directement et empiriquement la propriété d'écrantage de l'eau en terme d'une constante diélectrique qui dépend de la distance  $D(r)$ . Cette fonction modifie l'interaction coulombienne entre les atomes du soluté, qui devient alors  $q_i q_j / D(r_{ij}) r_{ij}$  entre deux atomes  $i$  et  $j$ . La fonction  $D(r)$  représente la modification des interactions électrostatiques dans le solvant par rapport aux interactions dans le vide. La forme la plus intuitive de la fonction  $D(r)$  pour l'eau est représentée par une fonction sigmoïdale qui tend vers 80 à grande distance, et qui vaut 1 à faible distance [Hingerty *et al.*, 1985, Lavery *et al.*, 1986]. Cette fonction s'exprime de la façon suivante :

$$D(r) = D - \frac{D-1}{2} [(Sr)^2 + 2Sr + 2] e^{-Sr}, \quad (2.8)$$

où  $D$  est la valeur du plateau ( $D = 80$  pour l'eau) et  $S$  la pente qui vaut en pratique le plus souvent  $S = 0,356$ , et  $S = 0,16$  pour la modélisation de certains ADN B en solution

[Lavery, 1988]. Les courbes des fonctions sigmoïdales  $D(R)$  pour les deux valeurs de la pente  $S$  sont représentées sur la figure 2.4. Cette méthode simple a permis de modéliser de façon satisfaisante la structure et la dynamique interne de nombreux acides nucléiques [Lavery *et al.*, 1986].

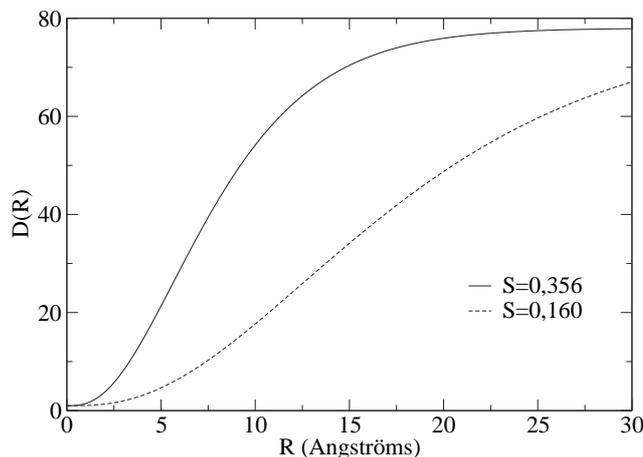


FIG. 2.4 : Courbe représentant la fonction sigmoïdale pour l'eau ( $D = 80$ ) pour deux pentes  $S$  différentes.

La méthode plus récente et plus réaliste des « Screened Coulomb Potentials » [Hassan *et al.*, 2000] fait également partie des méthodes DDD, et elle a été implémentée avec succès dans des programmes de dynamique moléculaire et de Monte-Carlo. Elle définit une fonction d'écrantage  $D(r)$  non linéaire de forme sigmoïdale, en accord avec la théorie classique de Lorenz-Debye-Sack (LDS) des fluides dipolaires [Ehrenson, 1989]. Les paramètres de cette fonction dépendent à la fois du type de l'atome concerné et du groupe d'atomes auquel il appartient. Avec ce modèle, l'énergie électrostatique de solvation va s'écrire selon la formule :

$$\begin{aligned} \Delta G^p = & \frac{1}{2} \sum_{i \neq j}^N \frac{q_i q_j}{r_{ij}} \left[ \frac{1}{D_s(r_{ij})} - \frac{1}{D_v(r_{ij})} \right] \\ & + \frac{1}{2} \sum_i^N q_i^2 \left( \frac{1}{R_{i,Bs}} \left[ \frac{1}{D_s(R_{i,Bs})} - 1 \right] - \frac{1}{R_{i,Bv}} \left[ \frac{1}{D_v(R_{i,Bv})} - 1 \right] \right), \quad (2.9) \end{aligned}$$

où  $i$  et  $j$  sont les atomes du soluté de charges  $q_i$  et  $q_j$  à la distance  $r_{ij}$ , la fonction  $D_v(r)$  représente la fonction d'écrantage des charges dû aux autres atomes de la molécule dans le vide,  $D_s(r)$  la fonction d'écrantage dû au solvant,  $R_{i,Bv}$  et  $R_{i,Bs}$  sont les rayons de Born de l'atome  $i$  de la molécule respectivement dans le vide et dans le solvant (la définition des rayons de Born est donnée plus loin par l'équation (2.11)). Les résultats obtenus pour les énergies de solvation, après une paramétrisation délicate et systématique pour coller aux

valeurs expérimentales, sont en assez bonne corrélation avec les résultats des méthodes FDPB, même si la droite de corrélation ne passe pas par zéro et que sa pente n'est pas égale à un [Hassan *et al.*, 2000]. Cependant, de telles méthodes ont l'avantage évident d'être faciles à implémenter et d'être surtout extrêmement peu coûteuses en temps de calcul, puisqu'elles sont aussi rapides que des simulations dans le vide. Cette méthode DDD a permis récemment de modéliser correctement la dynamique de peptides et de protéines [Hassan *et al.*, 2003]. Parmi les méthodes théoriques de représentation du solvant, les méthodes DDD sont donc de loin les plus rapides en temps de calcul.

### •La méthode de Born Généralisé

La méthode de Born Généralisé (« Generalized Born » - GB) permet d'exprimer l'énergie libre de solvation électrostatique en terme d'interactions de paires effectives [Still *et al.*, 1990]. Pour un ensemble de  $N$  atomes de charges  $q_i$  dans un milieu de constante diélectrique  $\varepsilon$ , l'énergie de solvation polaire du système peut s'exprimer par la formule :

$$\Delta G^p = -\frac{1}{2} \left(1 - \frac{1}{\varepsilon}\right) \sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{\sqrt{r_{ij}^2 + a_{ij} e^{-\frac{r_{ij}^2}{(2a_{ij})^2}}}}, \quad (2.10)$$

avec  $a_{ij} = \sqrt{a_i a_j}$ , où  $a_i$  est le rayon de Born de l'atome  $i$ ,  $q_i$  sa charge, et  $r_{ij}$  la distance entre l'atome  $i$  et l'atome  $j$ .

Une telle formulation nécessite donc de calculer les rayons de Born  $a_i$  des atomes du soluté, qui dépendent de la position de l'atome par rapport au solvant et donc de la conformation du soluté. Le calcul de ces rayons est effectué en calculant l'énergie libre de solvation propre  $\Delta G_i^p$  de chaque atome, qui est l'énergie libre de solvation électrostatique de la molécule étudiée, mais où tous les atomes ont une charge nulle sauf l'atome  $i$  de charge  $q_i$ . Ensuite, le rayon de Born de l'atome  $i$  est calculé grâce à la formule de Born reliant l'énergie de solvation et le rayon d'un ion isolé :

$$a_i = -\frac{1}{2} \left(1 - \frac{1}{\varepsilon}\right) \frac{q_i^2}{\Delta G_i^p}. \quad (2.11)$$

Le calcul numérique de l'énergie libre de solvation propre de chaque atome, par exemple par des méthodes de différences finies pour résoudre Poisson-Boltzmann (FDPB), permet d'obtenir un jeu de rayons de Born satisfaisant, donnant des énergies de solvation en accord avec l'expérience. Cependant, pour un soluté composé de  $N$  atomes,  $N$  calculs numériques sont préalablement nécessaires pour obtenir l'énergie de solvation de la molécule, et ce pour chaque configuration de la molécule. Un traitement analytique est donc nécessaire pour pouvoir implémenter de façon moins coûteuse la méthode GB en

dynamique moléculaire [Schaefer et Karplus, 1996, Qiu *et al.*, 1997]. Ces méthodes analytiques ne sont réalisables qu'en faisant plusieurs approximations pour accéder à l'énergie de solvation propre de chaque atome. Tout d'abord elles nécessitent d'écrire rigoureusement l'énergie de solvation propre de l'atome  $i$  avec la formule [Schaefer et Karplus, 1996] :

$$\Delta G_i^p = \frac{\tau}{8\pi} \int_{\mathbb{R}^3} \mathbf{D}_i(\mathbf{r})^2 \cdot \mathbf{H}_s(\mathbf{r}) d\mathbf{r} - \tau \frac{q_i^2}{2R_i}, \quad (2.12)$$

où  $\tau = 1/\varepsilon_i - 1/\varepsilon_s$ , avec  $\varepsilon_i$  et  $\varepsilon_s$  les constantes diélectriques respectives du soluté et du solvant,  $R_i$  le rayon de van der Waals de l'atome  $i$ ,  $\mathbf{H}_s(\mathbf{r})$  est la fonction de densité moléculaire qui vaut 1 si  $\mathbf{r}$  appartient au soluté et 0 sinon, et  $\mathbf{D}_i(\mathbf{r})$  est le déplacement diélectrique en  $\mathbf{r}$  dû à l'atome  $i$ .

Dans ces méthodes analytiques, on approxime le déplacement diélectrique par sa forme dite « locale » (comme nous le préciserons dans le chapitre suivant), qui consiste à considérer que le déplacement électrique en tout point est égal au champ électrique créé par le soluté dans le vide :

$$\mathbf{D}_i(\mathbf{r}) = q_i \frac{(\mathbf{r} - \mathbf{r}_i)}{|\mathbf{r} - \mathbf{r}_i|^3}. \quad (2.13)$$

Enfin, pour exprimer la fonction de densité moléculaire  $\mathbf{H}_s(\mathbf{r})$ , on fait l'approximation que c'est une somme de fonctions atomiques  $\mathbf{H}_k(\mathbf{r})$  qui vaut 1 si  $\mathbf{r}$  appartient à l'atome  $k$  et 0 sinon. Il ne reste ainsi qu'à développer une somme d'intégrales  $\Delta G_{ik}^p$  :

$$\Delta G_{ik}^p = \int_{\mathbb{R}^3} \mathbf{D}_i(\mathbf{r})^2 \cdot \mathbf{H}_k(\mathbf{r}) d\mathbf{r}. \quad (2.14)$$

La grande difficulté de cette méthode est de trouver les volumes atomiques qui permettent de bien reproduire les résultats issus des méthodes basées sur l'équation de Poisson-Boltzmann, puisque l'utilisation des rayons de Lennard-Jones des champs de force habituels ne donne pas de résultats satisfaisants [Dominy et Brooks, 1999]. Ce paramétrage est empirique et parfois fastidieux pour les systèmes importants tels que les macromolécules biologiques. Il doit en effet être effectué spécifiquement pour l'étude d'une famille de molécules similaires. La corrélation entre les énergies de solvation issues des méthodes GB analytiques et celles issues des méthodes FDPB dépend beaucoup de ce paramétrage.

La méthode de Born Généralisé analytique a été implémentée dans divers programmes de dynamique moléculaire tels qu'AMBER [Tsui et Case, 2001], et appliquée à la simulation de nombreux acides nucléiques et protéines [Dominy et Brooks, 1999, Calimet *et al.*, 2001]. En temps de calcul, cette méthode est particulièrement avantageuse

par rapport aux méthodes de solvant explicites. Elle est toujours comparée aux méthodes FDPB pour les énergies de solvation, et aux méthodes explicites pour l'étude conformationnelle. Les résultats obtenus, à la condition d'un paramétrage précis, sont en bon accord à la fois avec les méthodes explicites et Poisson-Boltzmann. Cependant, le temps de calcul gagné par rapport aux méthodes explicites avec l'utilisation de cette méthode ne prend pas en compte le temps de paramétrage préalable, qui peut être long, afin d'obtenir des résultats corrects, comme pour la simulation de la protéine 1PGB où de nombreux essais de volumes atomiques ont été effectués avant de trouver les meilleurs paramètres [Calimet *et al.*, 2001].

Les méthodes d'interaction de paires effectives, que ce soit celles basées sur Born Généralisé ou sur la constante diélectrique dépendant de la distance, sont très performantes en terme de temps de calcul, et peuvent donner des résultats comparables à celles basées sur l'équation de Poisson-Boltzmann tout en étant plus simples à implémenter dans les programmes de dynamique moléculaire. Cependant, elles nécessitent un paramétrage empirique qui doit convenir à chaque système étudié, et de ce paramétrage dépend fortement la précision des résultats obtenus.

### 2.2.3 Méthodes hybrides

Les méthodes de solvation continues présentées précédemment négligent complètement la nature moléculaire du solvant, et ne peuvent donc pas rendre compte de propriétés telle que la structure des couches d'hydratation des macromolécules étudiées. Si le bulk peut en effet être considéré comme un milieu continu, cette approximation est plus discutable pour les molécules d'eau proches du soluté, même si elle permet de calculer correctement l'énergie de solvation. Pour palier ce manque, des méthodes de solvation hybrides ont été développées, tentant de prendre en compte certains des aspects microscopiques et structuraux des premières couches de solvation.

#### Méthodes des dipôles de Langevin

La méthode des dipôles de Langevin peut être classée parmi les méthodes hybrides car elle prend en compte l'aspect microscopique du solvant [Warshel et Levitt, 1976, Warshel et Russel, 1984, Florian et Warshel, 1997, Papazyan et Warshel, 1997]. En effet, ce modèle peut être considéré comme pseudo-microscopique car il représente le solvant en terme de polarisation macroscopique sur une grille tridimensionnelle cubique autour du soluté, avec un dipôle ponctuel en chaque point de la grille représentant la polarisation

moyenne d'une molécule d'eau à cette position (cf figure 2.5). Ce modèle permet donc de représenter la polarisation électrostatique du solvant dans le sens d'une représentation microscopique, même si elle ne représente pas la géométrie moléculaire du solvant.

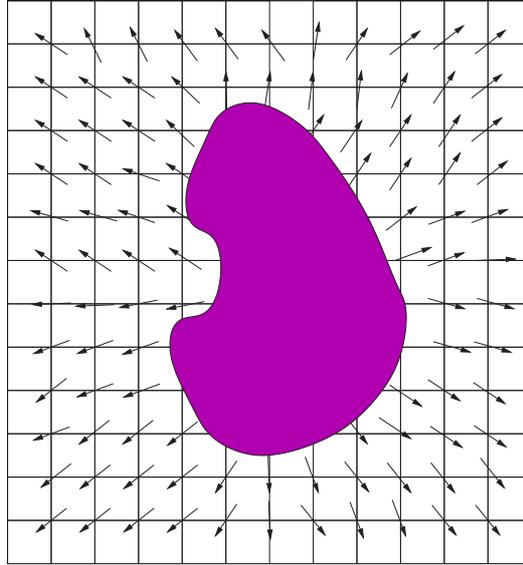


FIG. 2.5 : Représentation des dipôles sur grille.

Les dipôles ponctuels de la grille sont polarisables par le champ extérieur créé par les atomes du soluté et par les autres dipôles. Les effets de saturation diélectrique sont pris en compte par une fonction de Langevin  $\mathcal{L}$ . Ainsi, en chaque point  $j$  de la grille, on a un dipôle  $\mathbf{p}_j$  :

$$\mathbf{p}_j = \mu_0 \mathcal{L} \left( \frac{C \mu_0 E_j}{kT} \right) \frac{\mathbf{E}_j}{E_j}, \quad (2.15)$$

où  $\mathcal{L}(x) = \coth(x) - 1/x$  est la fonction de Langevin,  $\mu_0$  est la valeur du dipôle de saturation qui varie en fonction de la finesse de la grille, donc de la densité des dipôles,  $C$  est un paramètre ajustable indiquant la résistance des dipôles à leur orientation dans un champ extérieur,  $\mathbf{E}_j$  est le champ électrostatique total créé par le soluté et les autres dipôles au point  $j$  de la grille, et  $E_j$  son module.

Pour le calcul du champ électrostatique  $\mathbf{E}_j$  utilisé dans le calcul des dipôles, deux méthodes se distinguent :

- la méthode non itérative (NLD), correspondant à l'introduction de l'approximation locale, ne calcule pas les interactions des dipôles entre eux : le champ électrique est alors le champ créé par le soluté en un point  $j$  de la grille écranté par une fonction  $d(r_{ij})$  :

$$\mathbf{E}_j^{NLD} = \sum_i \frac{q_i \mathbf{r}_{ij}}{d(r_{ij}) r_{ij}^3}, \quad (2.16)$$

avec la fonction d'écrantage :

$$d(r_{ij}) = \frac{\sqrt{2 + r_{ij}}}{1.7}, \quad (2.17)$$

où  $i$  représente un atome du soluté de charge  $q_i$ ,  $j$  le point de la grille,  $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$  est le vecteur distance entre  $i$  et  $j$ , et  $r_{ij}$  son module ;

- la méthode itérative (ILD) prend en compte les interactions des dipôles entre eux : la méthode est alors itérative, puisque le champ électrostatique au point  $j$  dépend des autres dipôles dont les caractéristiques dépendent du dipôle en  $j$ . Le champ au point  $j$  est alors :

$$\mathbf{E}_j^{ILD} = \mathbf{E}_j^0 + \mathbf{E}_j^L - \mathbf{E}_j^C, \quad (2.18)$$

où :

$$\mathbf{E}_j^0 = \sum_i \frac{q_i \mathbf{r}_{ij}}{r_{ij}^3}$$

est le champ électrique créé par le soluté dans le vide,  $\mathbf{E}_j^C$  est le champ dû aux dipôles les plus proches, retiré pour des raisons de convergence de la méthode, et  $\mathbf{E}_j^L$  est le champ électrique créé par tous les autres dipôles, déterminé de façon itérative, qui vaut à la  $n$ ième itération :

$$\mathbf{E}_j^{L(n)} = \sum_{k \neq j} \frac{r_{jk}^2 \mathbf{P}_k^{(n-1)} - (\mathbf{r}_{jk} \cdot \mathbf{P}_k^{(n-1)}) \mathbf{r}_{jk}}{r_{jk}^5}. \quad (2.19)$$

Dans les deux cas, l'énergie de solvation électrostatique du soluté est alors calculée simplement par les formules :

$$\Delta G^{p(NLD)} = -\frac{1}{2} k^{NLD} \sum_{j=1}^N \mathbf{P}_j^{NLD} \cdot \mathbf{E}_j^{NLD}, \quad (2.20)$$

$$\Delta G^{p(ILD)} = -\frac{1}{2} k^{ILD} \sum_{j=1}^N \mathbf{P}_j^{ILD} \cdot \mathbf{E}_j^0, \quad (2.21)$$

où  $k^{NLD}$  et  $k^{ILD}$  sont des paramètres ajustables.

Les tailles de grille utilisées pour ces deux méthodes sont de l'ordre de 1 à 3 Å. Notons que pour la méthode des dipôles de Langevin, la contribution de van der Waals ( $\Delta G_{vdW}$ ) à l'énergie de solvation n'est pas calculée par la méthode de SASA d'après l'équation (2.1) mais par une somme de potentiels de répulsion de London en  $1/r^9$  [Florian et Warshel, 1997]. Précisons enfin qu'afin de traiter le solvant et le soluté de façon cohérente, la méthode des dipôles de Langevin est souvent utilisée avec l'introduction d'une polarisation électronique sur les atomes du soluté, c'est-à-dire que des dipôles ponctuels sont placés

sur chaque atome du soluté, traités dans les équations de façon analogue aux dipôles du solvant, mais avec une polarisabilité atomique différente à déterminer selon l'atome et la molécule à laquelle il appartient. Cette méthode appelée « Protein Dipole Langevin Dipole » a été implémentée dans le programme de mécanique moléculaire POLARIS, en particulier pour étudier les protéines [Florian et Warshel, 1997]. Le paramétrage de la méthode des dipôles de Langevin doit être fait de façon systématique pour la détermination des paramètres  $C$ , vu à l'équation (2.15), et  $k$  des équations (2.20) et (2.21), mais aussi pour les charges et polarisabilités atomiques des macromolécules étudiées.

Ces méthodes itératives ou non des dipôles de Langevin donnent des résultats intéressants car elles peuvent décrire certaines particularités dans l'orientation des dipôles des premières couches d'hydratation des protéines grâce aux variations des dipôles ponctuels correspondant à des variations diélectriques locales du solvant. Elles donnent également des valeurs correctes d'énergies libres de solvation [Harvey, 1989, Florian et Warshel, 1997, Warshel et Papazyan, 1998]. Cependant, pour obtenir des résultats cohérents, ces méthodes demandent un grand travail de paramétrage parfois empirique. De plus, comme toutes les méthodes sur grille, elles sont délicates à implémenter dans des programmes de dynamique moléculaire à cause des discontinuités dans l'expression des forces qui empêchent la conservation de l'énergie.

### Méthodes basées sur la fonctionnelle de la densité des liquides (DFT) et la théorie des équations intégrales

La prise en compte des effets moléculaires du solvant peut être faite grâce à la théorie de mécanique statistique des liquides, l'idée générale de ces méthodes hybrides étant d'introduire une densité de position pour le solvant  $\rho(\mathbf{r})$ , qui varie à courte distance du soluté et qui tend vers  $\rho_0$ , densité du fluide homogène, à grande distance. L'avantage de ces méthodes hybrides est qu'elles vont permettre de calculer simultanément et de façon cohérente des grandeurs microscopiques et macroscopiques du solvant.

La théorie de mécanique statistique des liquides peut être appliquée à la solvation des molécules biologiques. Le solvant est considéré comme un fluide de particules de densité de position et d'orientation  $\rho(\mathbf{x})$  (où  $\mathbf{x} \equiv (\mathbf{r}, \boldsymbol{\Omega})$  représente le vecteur position et le vecteur orientation de la molécule), qui interagissent entre elles par un potentiel de paires. Ces particules sont soumises à un potentiel extérieur  $V_{ext}(\mathbf{x})$ , qui est le potentiel de perturbation dû au soluté. Dans le fluide homogène, on définit la fonction de corrélation de paire  $h(\mathbf{r}_1, \mathbf{r}_2) = g^{(2)}(\mathbf{r}_1, \mathbf{r}_2) - 1$ , où  $g^{(2)}(\mathbf{r}_1, \mathbf{r}_2)$  est la fonction de distribution de paire qui représente la probabilité qu'il y ait une particule en  $\mathbf{r}_2$  sachant qu'il y a une particule

en  $\mathbf{r}_1$  [Hansen et McDonald, 1986]. Comme le fluide est homogène, cette probabilité tend vers 1 à grande distance, et on a donc :  $\lim_{|\mathbf{r}_1 - \mathbf{r}_2| \rightarrow \infty} h(\mathbf{r}_1, \mathbf{r}_2) = 0$ . On définit la fonction de corrélation directe  $c$  du fluide homogène par rapport à la fonction de corrélation  $h$  grâce à la relation intégrale d'Ornstein-Zernike [Hansen et McDonald, 1986], qui décompose la fonction de corrélation en une partie directe et une partie indirecte :

$$h(\mathbf{x}_1, \mathbf{x}_2) = c(\mathbf{x}_1, \mathbf{x}_2) + \rho_0 \int c(\mathbf{x}_1, \mathbf{x}_3)h(\mathbf{x}_3, \mathbf{x}_2)d\mathbf{x}_3. \quad (2.22)$$

L'approche de Ramirez *et al.* [2002], basée sur la théorie de la fonctionnelle de la densité [Chandler *et al.*, 1986, Teixeira et da Gama, 1991, Frodl et Dietrich, 1992] appliquée à ce fluide de particules de solvant perturbé par la présence d'un soluté, permet d'écrire une fonctionnelle d'énergie libre en fonction de la densité  $\rho$  du fluide de la façon suivante :

$$\begin{aligned} \mathcal{F}[\rho] = & \beta^{-1} \int d\mathbf{x} \left[ \rho(\mathbf{x}) \ln \left( \frac{\rho(\mathbf{x})}{\rho_0} \right) - \rho(\mathbf{x}) + \rho_0 \right] + \int d\mathbf{x} V_{ext}(\mathbf{x})\rho(\mathbf{x}) \\ & - \frac{\beta^{-1}}{2} \int \int d\mathbf{x}_1 d\mathbf{x}_2 c(\mathbf{x}_1, \mathbf{x}_2) \Delta\rho(\mathbf{x}_1) \Delta\rho(\mathbf{x}_2), \end{aligned} \quad (2.23)$$

où  $\beta = 1/kT$  et  $\Delta\rho(\mathbf{x}) = \rho(\mathbf{x}) - \rho_0$ .

À l'équilibre thermodynamique, cette fonctionnelle est minimale par rapport à  $\rho$ , on a  $\partial\mathcal{F}/\partial\rho = 0$ , ce qui revient à l'expression suivante pour la densité  $\rho$  du fluide à l'équilibre :

$$\rho(\mathbf{x}_1) = \rho_0 \exp \left( -\beta V_{ext}(\mathbf{x}_1) + \int d\mathbf{x}_2 c(\mathbf{x}_1, \mathbf{x}_2) \Delta\rho(\mathbf{x}_2) \right). \quad (2.24)$$

La méthode de DFT présentée ici permet, d'une façon analogue aux équations intégrales, soit par une minimisation de la fonctionnelle par des méthodes numériques, soit par une résolution itérative des équations (2.22) et (2.24), d'obtenir la fonctionnelle d'énergie libre à l'équilibre qui correspond à l'énergie libre de solvation du soluté.

Lorsque cette méthode est appliquée à un fluide dipolaire tel que l'eau, la fonctionnelle peut s'écrire plus simplement en fonction de la densité de polarisation  $\mathbf{P}(\mathbf{r})$  et de la densité de position  $\rho(\mathbf{r})$ . Il suffit alors de minimiser la fonctionnelle par rapport à la densité de particules et par rapport à la densité de polarisation pour obtenir toutes les caractéristiques du solvant à l'équilibre autour du soluté. Cette méthode a été testée pour des ions et des petits composés organiques rigides et a permis de reproduire correctement les couches de solvation de ces molécules [Ramirez *et al.*, 2002]. Elle est cependant assez coûteuse en temps de calcul et n'a pas encore été implémentée en dynamique moléculaire.

Notons que l'équation (2.24) est très similaire à l'équation de fermeture HNC (« hypernetted chain ») utilisée pour la théorie des équations intégrales [Pettitt et Karplus,

1986, Beglov et Roux, 1997, Roux et Simonson, 1999]. La théorie des équations intégrales permet, pour un système composé d'un mélange de plusieurs espèces de particules sans potentiel extérieur, de résoudre des systèmes d'équations de même type que les équations (2.22) et (2.24) sur une grille tridimensionnelle par des méthodes de transformées de Fourier rapides. Elle permet d'obtenir des informations sur la variations de densité du solvant autour d'une molécule, mais est coûteuse en temps de calcul et ne pourrait permettre que difficilement d'effectuer des dynamiques et de calculer l'énergie libre de solvation.

## 2.2.4 Conclusion sur les méthodes théoriques

Nous avons vu que, malgré les différentes méthodes existant pour représenter le solvant en modélisation moléculaire, il est difficile d'obtenir avec la même méthode des informations à la fois dynamiques et thermodynamiques sur le système étudié. En effet, les méthodes explicites, si elles donnent une description précise des trajectoires des molécules d'eau, sont beaucoup moins adaptées pour calculer les grandeurs thermodynamiques moyennes, qui sont pourtant des données très importantes pour comprendre les processus biochimiques. Quant aux méthodes implicites, les méthodes basées sur l'équation de Poisson-Boltzmann donnent de façon précise ces grandeurs thermodynamiques, mais souvent pour des configurations statiques des molécules. En effet, leur implémentation dans des programmes de dynamique moléculaire est complexe car il n'y a pas d'expression analytique des forces. Les méthodes de type diélectrique dépendant de la distance sont plus simples à implémenter en dynamique moléculaire, mais elles demandent souvent un grand travail de paramétrage. De plus, ces méthodes continues ne donnent pas d'informations sur la structure moléculaire du solvant, puisqu'elles le considèrent comme un milieu continu. Nous avons vu que les méthodes de solvation hybrides étaient une avancée importante dans la description des phénomènes de solvation, puisqu'elles ont les avantages des méthodes continues mais donnent plus d'informations microscopiques sur les structures des couches d'hydratation. Elles aussi sont cependant délicates à implémenter en dynamique moléculaire, et relativement coûteuses en temps de calcul. Le coût en terme de temps de calcul est en effet un véritable enjeu pour la représentation du solvant en modélisation moléculaire, et doit malheureusement être le plus souvent un compromis entre la précision des résultats voulue et un temps de calcul non prohibitif pour les machines actuelles.

Dans le but de modéliser la dynamique des structures des couches d'hydratation des biomolécules et l'énergie libre de solvation, il serait très utile d'avoir un modèle entièrement analytique qui pourrait décrire le solvant en terme de particules tout en se basant sur des équations électrostatiques macroscopiques, à l'image des modèles hybrides,

afin de calculer efficacement les grandeurs thermodynamiques moyennes. C'est dans cette optique de méthodes « mésoscopiques », à la frontière entre les méthodes macroscopiques et microscopiques, que s'inscrit notre démarche, dont la méthodologie est présentée de façon détaillée dans le chapitre suivant.



## **Chapitre 3**

# **Méthodologie de l'étude**



# Méthodologie de l'étude

Dans ce chapitre, nous développerons la méthodologie de l'étude effectuée, qui sera présentée aux chapitres suivants. Dans un premier temps, nous présenterons la théorie du modèle de solvatation utilisé ensuite dans l'étude. Il s'agit d'un nouveau modèle électrostatique simple sur particules qui reproduit en moyenne une représentation exacte ou approximative des lois macroscopiques de l'électrostatique. À partir d'une fonctionnelle d'énergie libre de polarisation d'un milieu continu, projetée sur des particules de Lennard-Jones, on obtient un fluide de particules polarisables. Ce modèle peut être décliné avec ses deux approches non locales et locales, dont les caractéristiques seront décrites dans la partie 3.1. L'approche non locale, basée sur une représentation initiale exacte de l'électrostatique, est aussi plus coûteuse en temps de calcul. L'approche locale, qui consiste à faire une approximation sur la densité de polarisation, est très performante en terme de temps de calcul, et donne des résultats très convenables. Dans un deuxième temps, après avoir rappelé les principes généraux de la dynamique moléculaire dans la partie 3.2, nous montrerons comment notre nouveau modèle peut être implémenté simplement dans un algorithme de dynamique moléculaire par un hamiltonien spécifique. Enfin, dans un troisième temps, nous reviendrons sur le modèle local et nous présenterons dans la partie 3.3 le paramétrage de la méthode afin qu'elle soit en meilleur accord possible avec l'électrostatique. Un tel paramétrage est nécessaire pour toute nouvelle méthode qui n'a pas encore été testée et comparée aux méthodes existantes. Nous effectuerons ce paramétrage de la façon la plus universelle possible, afin de pouvoir utiliser notre modèle sur des molécules différentes sans avoir à trouver de nouveaux paramètres spécifiques pour chaque nouvelle étude.

## 3.1 Un modèle de solvation basé sur des pseudo-particules polarisables

### 3.1.1 Principe général du modèle

Nous proposons dans cette partie un nouveau modèle pour étudier la solvation des macromolécules biologiques, développé au Laboratoire de Modélisation des Systèmes Moléculaires Complexes [Ha-Duong *et al.*, 2002]. Cette méthode repose sur une fonctionnelle de la densité de polarisation, correspondant à l'équilibre à l'énergie libre de solvation électrostatique. Cette fonctionnelle est calculée non pas sur une grille régulière, mais sur des particules de Lennard-Jones polarisables. Les particules ont leurs trois degrés de liberté de translation, alors que l'orientation de leur dipôle considéré à l'équilibre thermodynamique est calculée en minimisant la fonctionnelle. Les particules de solvant désordonnées représentent ainsi une grille large et irrégulière sur laquelle évaluer la fonctionnelle (cf. figure 3.1), et les grandeurs thermodynamiques moyennes sont obtenues en moyennant sur plusieurs configurations des particules de solvant.

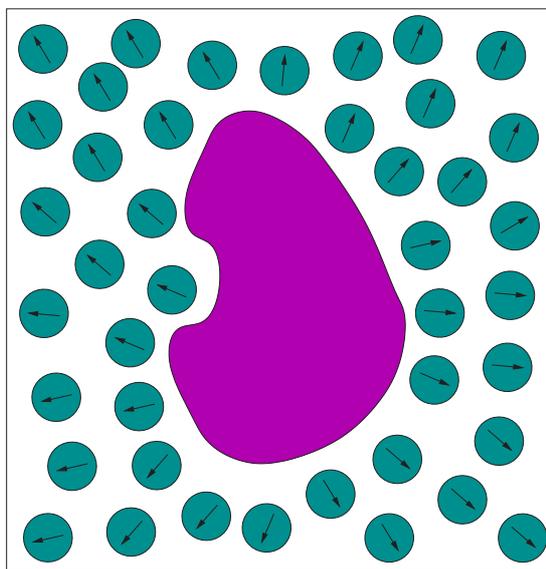


FIG. 3.1 : Représentation des pseudo-particules polarisables autour d'une macromolécule.

Cette méthode peut être qualifiée de semi-implicite, ou mésoscopique, puisqu'elle découle de propriétés électrostatiques macroscopiques et implicites appliquées à des particules microscopiques et explicites. Les avantages de ce traitement semi-implicite du solvant sont divers :

1. la frontière entre le soluté et le solvant est alors déterminée de façon naturelle,

puisqu'elle est définie par les rayons de van der Waals des atomes et des pseudo-particules de solvant ;

2. le solvant et le soluté sont traités de manière cohérente, comme dans le cas des solvants explicites, et le système peut donc être étudié simplement en dynamique moléculaire par un champ de force atomique ;
3. le gain de temps de calcul par rapport à un solvant explicite devrait être important : en effet, chaque molécule d'eau est traitée comme un seul site (au lieu de trois pour l'eau explicite habituelle) avec un moment dipolaire induit, sans charges ponctuelles. De plus, comme nous le verrons plus loin, dans le cas du modèle local le plus simple, les dipôles de solvant ne se « voient » pas entre eux, et l'interaction solvant/solvant se réduit à celle de simples particules de Lennard-Jones.

Nous montrerons dans cette partie que ce modèle sur particules revient à un modèle de particules non polaires polarisables, dont la polarisabilité est liée aux grandeurs électrostatiques macroscopiques du milieu diélectrique continu. Les dipôles induits des pseudo-particules représentent à la fois la polarisation orientationnelle et électronique des molécules. Nous emploierons le terme de pseudo-particules car les particules de solvant sont considérées comme des particules pour les degrés de liberté de translation, mais leurs propriétés électrostatiques découlent d'équations macroscopiques à l'équilibre thermodynamique.

Nous présenterons d'abord le modèle appelé « non local », qui est le modèle électrostatique sur particules le plus fondé puisqu'il découle directement de la fonctionnelle macroscopique d'énergie libre de solvation. Ce modèle pose le problème de la détermination de la valeur de la polarisabilité  $\alpha$  des particules, grandeur microscopique, en fonction de la constante diélectrique  $\varepsilon$  du milieu continu, grandeur macroscopique. Nous verrons en effet que la relation connue de Clausius-Mosotti n'est plus vérifiée pour des fluides fortement polarisables. Ce modèle non local conduit à des équations auto-cohérentes qui nécessitent donc une résolution itérative, et cette résolution peut converger lentement. Des problèmes de convergence nous ont amenés à rajouter une dimension phénoménologique au modèle, en introduisant un effet de saturation diélectrique par l'intermédiaire d'une fonction de Langevin agissant sur les moments dipolaires ponctuels induits, d'une manière analogue à la méthode des dipôles de Langevin développée par Warshel *et al.* et vue au chapitre précédent.

Nous présenterons ensuite le modèle « local », qui découle d'une approximation sur la densité de polarisation. Cette approximation mène alors à des équations qui ne sont plus auto-cohérentes et les moments dipolaires ponctuels sont déterminés directement en fonction du champ créé par le soluté dans le vide. Le modèle local aura donc l'avantage

d'être très efficace numériquement, plus facile à implémenter et plus rapide que le modèle non local. Là encore nous introduirons des effets de saturation diélectrique par une fonction de Langevin. Remarquons que ces deux modèles de dipôles sur particules, non local et local, peuvent être mis en parallèle avec les méthodes itérative et non itérative des dipôles de Langevin sur grille.

Enfin, si notre fonctionnelle d'énergie libre est introduite de façon phénoménologique à partir d'une fonctionnelle macroscopique, notons qu'une approche de mécanique statistique des liquides basée sur la théorie de la fonctionnelle de densité (vue au chapitre précédent) appliquée à un solvant polaire microscopique aboutit à un modèle de solvant sur particules très proche de notre modèle non local de Langevin. Notre description macroscopique de l'électrostatique des pseudo-particules se trouve donc justifiée par une théorie microscopique des fluides dipolaires. Par souci de clarté, nous ne présenterons pas les développements théoriques de cette approche de mécanique statistique, mais les détails de la théorie peuvent être consultés dans l'article de référence [Ha-Duong *et al.*, 2002].

### 3.1.2 Modèle non local

Après avoir décrit la théorie du modèle non local, nous en présenterons quelques tests réalisés sur des systèmes simples. Dans le paragraphe suivant, nous nous intéresserons plus précisément à la relation liant la constante diélectrique macroscopique du milieu à la polarisabilité des particules de notre fluide.

#### Formulation du modèle non local sur particules

On considère un soluté présentant une distribution de charges  $\rho_0(\mathbf{r})$ , créant un champ électrique dans le vide  $\mathbf{E}_0(\mathbf{r})$ , immergé dans un milieu de constante diélectrique  $\varepsilon$ . Le système peut être décrit par une constante diélectrique dépendant de la position  $\varepsilon(\mathbf{r})$ , qui vaut  $\varepsilon_i$  à l'intérieur du soluté, et  $\varepsilon$  en-dehors des frontières du soluté. Le problème électrostatique peut être traité par une fonctionnelle non locale d'énergie libre dépendant de la densité de polarisation  $\mathbf{P}(\mathbf{r})$ , en accord avec la description macroscopique de Marcus [Marcus, 1956] :

$$\begin{aligned} \mathcal{F}_{NLP}[\mathbf{P}(\mathbf{r})] &= \frac{1}{2} \int \frac{\mathbf{P}(\mathbf{r})^2}{\varepsilon_0 \chi(\mathbf{r})} d\mathbf{r} - \int \mathbf{P}(\mathbf{r}) \cdot \mathbf{E}_0(\mathbf{r}) d\mathbf{r} \\ &\quad - \frac{1}{2} \int \int \mathbf{P}(\mathbf{r}) \cdot [\mathbf{T}(\mathbf{r} - \mathbf{r}') \cdot \mathbf{P}(\mathbf{r}')] d\mathbf{r} d\mathbf{r}', \end{aligned} \quad (3.1)$$

où  $\varepsilon_0$  est la constante diélectrique du vide,  $\chi(\mathbf{r}) = \varepsilon(\mathbf{r}) - 1$  est la susceptibilité électrique du milieu, et  $\mathbf{T}(\mathbf{r}) = -r^{-3}\mathbf{I} + 3r^{-5}\mathbf{r}\mathbf{r}$  est le tenseur dipolaire, avec  $\mathbf{I}$  la matrice identité. Notons que le tenseur dipolaire s'exprime aussi comme un double gradient :  $\nabla_{\mathbf{r}'} \cdot \nabla_{\mathbf{r}} |\mathbf{r} - \mathbf{r}'|^{-1}$ . La première intégrale de la fonctionnelle (3.1) représente l'énergie nécessaire pour créer le champ de polarisation, le deuxième terme représente l'interaction entre le milieu polarisable et le soluté, et le troisième terme l'interaction du milieu polarisable avec lui-même. Notons bien que la fonctionnelle d'énergie libre de polarisation ne représente que la partie électrostatique de l'énergie libre de solvation du soluté.

À l'équilibre thermodynamique, la fonctionnelle est minimale par rapport à la polarisation  $\mathbf{P}(\mathbf{r})$  :  $\partial\mathcal{F}_{NLP}[\mathbf{P}(\mathbf{r})]/\partial\mathbf{P}(\mathbf{r}) = 0$ . La résolution de cette dernière équation aboutit à l'expression de la densité de polarisation :

$$\mathbf{P}(\mathbf{r}) = \varepsilon_0\chi(\mathbf{r})\mathbf{E}(\mathbf{r}), \quad (3.2)$$

où  $\mathbf{E}(\mathbf{r})$  est le champ électrique de Maxwell du système qui s'écrit :

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}_0(\mathbf{r}) + \mathbf{E}_p(\mathbf{r}), \quad (3.3)$$

avec  $\mathbf{E}_p(\mathbf{r})$  le champ électrique de polarisation défini par :

$$\mathbf{E}_p(\mathbf{r}) = \int \mathbf{T}(\mathbf{r} - \mathbf{r}') \cdot \mathbf{P}(\mathbf{r}') d\mathbf{r}'. \quad (3.4)$$

Notons que Marcus [1956] a démontré que les équations découlant de la fonctionnelle à l'équilibre thermodynamique étaient équivalentes à l'équation de Poisson à laquelle obéit le potentiel électrostatique  $\phi(\mathbf{r})$  tel que  $\mathbf{E}(\mathbf{r}) = -\nabla\phi(\mathbf{r})$ , ou à l'équation de Maxwell,  $\nabla \cdot \mathbf{D}(\mathbf{r}) = \rho_0(\mathbf{r})$ , après avoir défini le déplacement diélectrique à l'équilibre par :

$$\mathbf{D}(\mathbf{r}) = \varepsilon_0\mathbf{E}(\mathbf{r}) + \mathbf{P}(\mathbf{r}). \quad (3.5)$$

L'expression du champ électrique de polarisation donnée par la formule (3.4) est en fait une intégrale impropre puisqu'elle n'est pas définie en  $\mathbf{r} = \mathbf{r}'$ . Le champ électrique de polarisation  $\mathbf{E}_p$  d'un milieu diélectrique sphérique soumis à un champ extérieur étant connu et égal à  $-\mathbf{P}/(3\varepsilon_0)$  [Ramshaw, 1971], on peut réécrire  $\mathbf{E}_p(\mathbf{r})$  sous la forme :

$$\mathbf{E}_p(\mathbf{r}) = -\frac{\mathbf{P}(\mathbf{r})}{3\varepsilon_0} + \int_{|\mathbf{r}-\mathbf{r}'|>\delta} \mathbf{T}(\mathbf{r} - \mathbf{r}') \cdot \mathbf{P}(\mathbf{r}') d\mathbf{r}', \quad (3.6)$$

où  $\delta$  est un rayon de coupure inférieur aux longueurs caractéristiques macroscopiques, et supérieur aux longueurs caractéristiques microscopiques.

En injectant l'expression du champ de polarisation donnée par (3.6) dans la fonctionnelle (3.1), la fonctionnelle non locale s'écrit de la façon suivante :

$$\begin{aligned} \mathcal{F}_{NLP}[\mathbf{P}(\mathbf{r})] &= \frac{1}{2} \int \left( \frac{3 + \chi(\mathbf{r})}{3\chi(\mathbf{r})} \right) \frac{\mathbf{P}(\mathbf{r})^2}{\varepsilon_0} d\mathbf{r} - \int \mathbf{P}(\mathbf{r}) \cdot \mathbf{E}_0(\mathbf{r}) d\mathbf{r} \\ &\quad - \frac{1}{2} \int \int_{|\mathbf{r}-\mathbf{r}'|>\delta} \mathbf{P}(\mathbf{r}) \cdot [\mathbf{T}(\mathbf{r}-\mathbf{r}') \cdot \mathbf{P}(\mathbf{r}')] d\mathbf{r} d\mathbf{r}'. \end{aligned} \quad (3.7)$$

Pour obtenir un modèle sur particules, nous allons discrétiser cette fonctionnelle macroscopique sur  $N$  particules de densité  $\rho$  et de paramètres de Lennard-Jones  $\sigma_s$  et  $\epsilon_s$  donnés. Chaque particule occupe un volume  $\Delta v = 1/\rho$ , et porte une densité de polarisation  $\mathbf{P}(\mathbf{r}_i) = \mathbf{p}_i/\Delta v$ . Dans la fonctionnelle d'énergie libre, les intégrales continues sur tout l'espace sont alors remplacées par des sommes discrètes sur les particules de solvant, car la polarisation est considérée nulle dans le soluté. Les différentielles  $d\mathbf{r}$  sont remplacées par  $\Delta v$ , et la fonction d'énergie libre s'écrit alors :

$$F_{NLP}(\{\mathbf{p}_i\}) = \frac{1}{2} \sum_i \left( \frac{3 + \chi}{3\chi} \right) \frac{\rho \mathbf{p}_i^2}{\varepsilon_0} - \sum_i \mathbf{p}_i \cdot \mathbf{E}_i^0 - \sum_{i<j} \mathbf{p}_i \cdot \mathbf{T}_{ij} \cdot \mathbf{p}_j, \quad (3.8)$$

où  $\mathbf{T}_{ij} = \mathbf{T}(\mathbf{r}_i - \mathbf{r}_j)$ , et  $\mathbf{E}_i^0$  est le champ électrique dans le vide créé par les  $M$  charges partielles du soluté  $q_k$  placées en  $\mathbf{R}_k$ , au point où se trouve la pseudo-particule  $i$  :

$$\mathbf{E}_i^0 = \mathbf{E}_0(\mathbf{r}_i; \{\mathbf{R}_k\}) = \frac{1}{4\pi\varepsilon_0} \sum_k q_k \frac{(\mathbf{r}_i - \mathbf{R}_k)}{|\mathbf{r}_i - \mathbf{R}_k|^3}, \quad (3.9)$$

Il est alors possible d'introduire la polarisabilité atomique  $\alpha$  des particules, qui s'exprime en fonction de la susceptibilité  $\chi$  du solvant, ou de la constante diélectrique  $\varepsilon$ , par les équations :

$$\alpha = \frac{3\varepsilon_0}{\rho} \left( \frac{\chi}{3 + \chi} \right) = \frac{3\varepsilon_0}{\rho} \left( \frac{\varepsilon - 1}{\varepsilon + 2} \right). \quad (3.10)$$

On reconnaît la relation de Clausius-Mosotti, que l'on sait vérifiée pour des fluides de faible polarisabilité, mais le raisonnement ci-dessus est approximatif et on ne sait pas si la relation est vérifiée pour des milieux fortement diélectriques tels que l'eau liquide. Il faudra trouver la polarisabilité  $\alpha$  des particules qui permet d'obtenir une constante diélectrique effective  $\varepsilon = 80$  correspondant à celle de l'eau. Nous reviendrons dans le paragraphe suivant de façon détaillée sur la relation entre  $\alpha$  et  $\varepsilon$ , qui est la clé du modèle. La fonctionnelle  $F_{NLP}$  peut alors s'écrire de façon plus simple :

$$F_{NLP}(\{\mathbf{p}_i\}) = \sum_i \frac{\mathbf{p}_i^2}{2\alpha} - \sum_i \mathbf{p}_i \cdot \mathbf{E}_i^0 - \sum_{i<j} \mathbf{p}_i \cdot \mathbf{T}_{ij} \cdot \mathbf{p}_j. \quad (3.11)$$

À l'équilibre thermodynamique,  $F_{NLP}$  est minimale par rapport aux  $\mathbf{p}_i$  pour une configuration donnée des particules de solvant et du soluté  $(\mathbf{r}_i, \mathbf{R}_k)$ , et on a donc :

$$\frac{\partial F_{NLP}}{\partial \mathbf{p}_i} = \mathbf{0}, \quad \forall i, \quad (3.12)$$

ce qui donne l'expression suivante pour les moments dipolaires induits  $\mathbf{p}_i$  des particules de solvant à l'équilibre thermodynamique :

$$\mathbf{p}_i = \alpha \mathbf{E}_i, \quad (3.13)$$

$$\mathbf{E}_i = \mathbf{E}_i^0 + \sum_{j \neq i} \mathbf{T}_{ij} \cdot \mathbf{p}_j. \quad (3.14)$$

En discrétisant la fonctionnelle macroscopique sur des particules, on obtient donc des pseudo-particules polarisables, dont le moment dipolaire induit est proportionnel au champ électrique au centre de la particule. Les équations obtenues sont auto-cohérentes, puisque le dipôle de la particule  $i$  dépend des dipôles de toutes les autres particules, qui dépendent eux-mêmes du dipôle de la particule  $i$ . La résolution de ces équations auto-cohérentes devra donc être faite par une méthode itérative. Lorsque les moments dipolaires induits des particules ont ainsi été calculés par les équations (3.13) et (3.14), l'énergie libre du système à l'équilibre prend alors l'expression simple :

$$F_{NLP}^{eq} = -\frac{1}{2} \sum_i \mathbf{p}_i \cdot \mathbf{E}_i^0. \quad (3.15)$$

Les forces électrostatiques (sans prendre en compte les interactions de Lennard-Jones) qui s'exercent entre les particules de solvant et les atomes du soluté peuvent être calculées simplement à partir de l'énergie libre de solvation électrostatique à l'équilibre thermodynamique. Sur une particule de solvant  $i$ , la force exercée par les autres particules de solvant et par les atomes du soluté s'exprime :

$$\mathbf{F}_i = -\frac{\partial F_{NLP}^{eq}}{\partial \mathbf{r}_i} = \mathbf{p}_i \cdot \left( \frac{\partial \mathbf{E}_i^0}{\partial \mathbf{r}_i} + \sum_{j \neq i} \frac{\partial \mathbf{T}_{ij}}{\partial \mathbf{r}_i} \cdot \mathbf{p}_j \right), \quad (3.16)$$

et sur un atome  $k$  du soluté, la force exercée par toutes les particules de solvant s'écrit :

$$\mathbf{F}_k = -\frac{\partial F_{NLP}^{eq}}{\partial \mathbf{R}_k} = \sum_i \mathbf{p}_i \cdot \frac{\partial \mathbf{E}_i^0}{\partial \mathbf{R}_k}. \quad (3.17)$$

Nous reviendrons de façon plus détaillée sur l'expression des forces et leur implémentation dans la partie sur la dynamique moléculaire.

### La saturation

Les équations auto-cohérentes donnant l'expression des dipôles induits sur les particules de solvant peuvent être résolues, pour chaque configuration du système, par une méthode itérative analogue à celle des dipôles de Langevin. Cependant, cette méthode itérative pose des problèmes de convergence numérique et l'équilibre thermodynamique est difficile à atteindre. Pour palier cette difficulté, nous ajoutons une dimension phénoménologique à notre modèle en introduisant un effet de saturation des dipôles. Nous représentons ces effets de saturation en modifiant les dipôles induits par une fonction de Langevin  $\mathcal{L}(x) = \coth(x) - 1/x$  (représentée sur la figure 3.2), de telle sorte que les dipôles sont proportionnels au champ total pour de faibles champs électrostatiques, mais pour de forts champs la valeur absolue du moment dipolaire sature à une valeur  $\mu$  donnée.

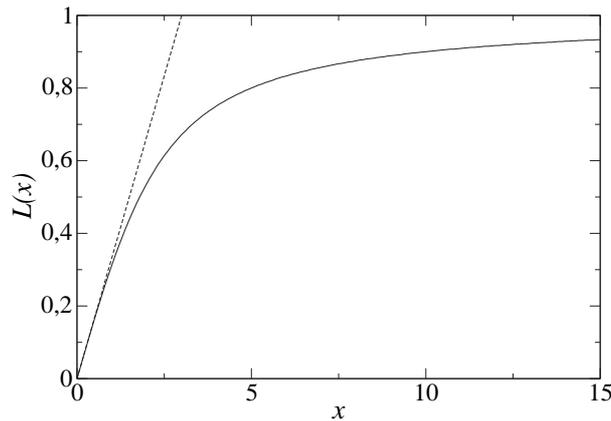


FIG. 3.2 : Représentation de la fonction de Langevin  $\mathcal{L}(x)$  en trait plein, et de  $x/3$  en trait pointillé : pour les  $x$  proches de 0, la fonction de Langevin est confondue avec  $x/3$ , et pour  $x \rightarrow \infty$  elle tend vers 1.

Les nouvelles équations s'écrivent alors à la place des équations (3.13) et (3.14) :

$$\mathbf{p}_i = \frac{\mu}{E_i} \mathcal{L}\left(\frac{3\alpha E_i}{\mu}\right) \mathbf{E}_i, \quad (3.18)$$

$$\mathbf{E}_i = \mathbf{E}_i^0 + \sum_{j \neq i} \mathbf{T}_{ij} \cdot \mathbf{p}_j, \quad (3.19)$$

où  $E_i = |\mathbf{E}_i|$  et  $\mu$  est le paramètre du dipôle de saturation à préciser.

La nouvelle fonctionnelle qui donne ces équations lorsqu'elle est minimisée par rapport aux  $\mathbf{p}_i$  doit donc être modifiée par rapport à la fonctionnelle de départ (3.11). Elle

peut s'écrire sous la forme :

$$F_{NLL}(\mathbf{p}_i) = \frac{\mu^2}{3\alpha} \sum_i \left( \mathcal{L}^{-1}(p_i/\mu) \frac{p_i}{\mu} - \ln \left[ \frac{\sinh[\mathcal{L}^{-1}(p_i/\mu)]}{\mathcal{L}^{-1}(p_i/\mu)} \right] \right) - \sum_{i \neq j} \mathbf{p}_i \cdot \mathbf{T}_{ij} \cdot \mathbf{p}_j - \sum_i \mathbf{p}_i \cdot \mathbf{E}_i^0. \quad (3.20)$$

Après avoir optimisé les dipôles à l'aide des équations auto-cohérentes (3.18) et (3.19), on obtient pour l'énergie libre de solvation à l'équilibre :

$$F_{NLL}^{eq} = -\frac{\mu^2}{3\alpha} \sum_i \ln \left[ \frac{\sinh(3\alpha E_i/\mu)}{3\alpha E_i/\mu} \right] + \sum_{i < j} \mathbf{p}_i \cdot \mathbf{T}_{ij} \cdot \mathbf{p}_j. \quad (3.21)$$

Notons que la reformulation de  $F_{NLL}^{eq}$  tenant compte de la saturation permet de ne pas modifier les expressions des forces données par les équations (3.16) et (3.17).

Le problème de la valeur du dipôle de saturation doit être examiné, car on ne dispose pas dans la littérature de valeur précise pour la saturation de nos pseudo-particules. Le dipôle effectif pour les modèles d'eau explicites de type TIP3 ou SPC est d'environ  $\mu = 2,3$  Debye, comme nous l'avons vu au chapitre précédent. La valeur  $\mu = 1,85$  Debye est la valeur expérimentale du dipôle de l'eau en phase gazeuse. En pratique, les tests du modèle non local ont été effectués avec des valeurs du dipôle de saturation de  $\mu = 1,85$  Debye, et de  $\mu = 5$  Debye, valeur qui n'a pas de justification théorique mais qui permet de se rapprocher d'un modèle sans saturation en restant en régime linéaire.

### Tests sur le modèle non local

Le modèle non local des pseudo-particules polarisables découle directement d'une théorie macroscopique exacte. Nous pensons donc pouvoir obtenir de bons résultats pour les propriétés électrostatiques. Il reste néanmoins quelques paramètres à ajuster : les paramètres de Lennard-Jones des particules, ainsi que la valeur de saturation des dipôles.

Nous avons implémenté le modèle non local dans un programme de dynamique moléculaire « maison » avec une résolution itérative des équations auto-cohérentes, le soluté étant un ensemble de sphères de Lennard-Jones fixes portant une charge ponctuelle. Nous avons ainsi testé avec le modèle non local le champ de polarisation autour d'un ion et le potentiel de force moyenne entre deux ions de charge opposée, pour différentes valeurs du dipôle de saturation, afin de valider le modèle en le comparant avec les résultats théoriques connus.

Nous avons choisi les paramètres de Lennard-Jones de nos particules afin de respecter au mieux les propriétés de l'eau liquide. Pour nos premiers tests, nous avons ainsi pris comme diamètre  $\sigma_s = 3,024 \text{ \AA}$ , qui correspond au diamètre d'une molécule d'eau. Afin de vérifier la condition sur la densité  $\rho^* = \rho\sigma_s^3 \simeq 0,8$  pour un liquide [Pollock *et al.*, 1980], nous avons donc une densité de  $\rho = 0,029 \text{ \AA}^3$ . Nous avons choisi  $\epsilon_s = 3,197 \text{ kJ/mol}$ , ce qui correspond de même à  $T^* = kT/\epsilon_s \simeq 0,8$  pour  $T = 300 \text{ K}$ . Pour tous les tests, nous avons pris  $N = 500$  particules de polarisabilité  $\alpha = 5,97 \text{ \AA}^3$  (qui correspond à  $\epsilon = 80$  par la relation d'Onsager pour  $x = 0,74$ , comme nous le verrons dans le paragraphe suivant). Nous avons effectué les dynamiques sur une durée d'environ 500 ps avec un pas d'intégration de  $\delta t = 10 \text{ fs}$ . Les conditions aux limites sont périodiques, et les interactions longue distance sont coupées à une distance de  $12 \text{ \AA}$  avec la convention de l'image minimum.

#### •La polarisation autour d'un ion :

Nous avons testé le modèle non local pour calculer la densité de polarisation, reliée au dipôle moyen  $\langle \mu(R) \rangle$ , autour d'un ion de charge  $q = +0,5e$ , pour différentes valeurs du dipôle de saturation  $\mu$ . Asymptotiquement à grande distance de l'ion, la théorie électrostatique prévoit que [Pollock *et al.*, 1980, Papazyan et Warshel, 1997] :

$$\lim_{R \rightarrow \infty} \langle \mu(R) \rangle = \lim_{R \rightarrow \infty} \frac{P(R)}{\rho} = \frac{\epsilon_0(\epsilon - 1)}{\rho\epsilon} E_0(R), \quad (3.22)$$

où  $\rho$  est la densité des particules de solvant,  $P(R)$  la polarisation du milieu, et  $E_0(R) = q^2/(4\pi\epsilon_0 R)$  est le champ créé par l'ion à la distance  $R$ .

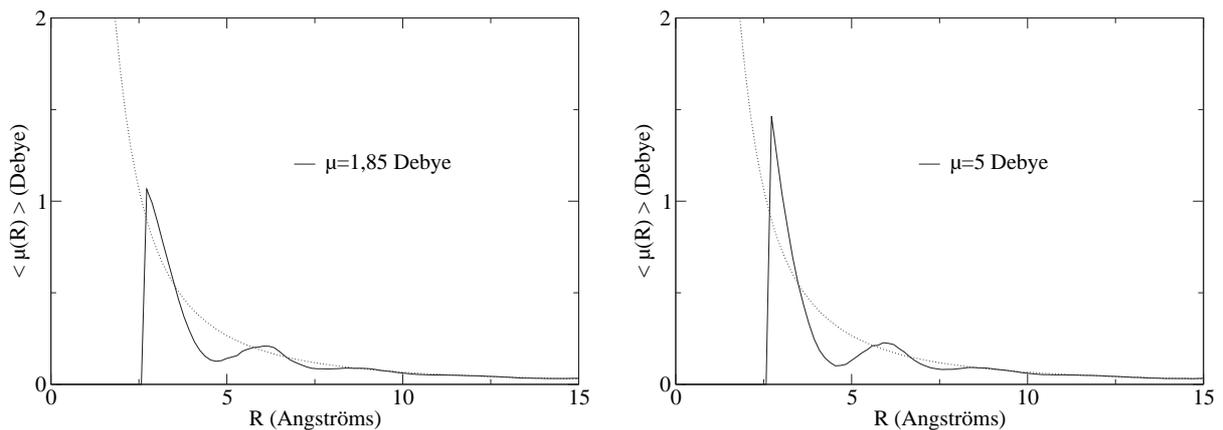


FIG. 3.3 : Dipôle moyen (en Debye) autour d'un ion avec le modèle non local, pour une valeur du dipôle de saturation  $\mu = 1,85 \text{ Debye}$  (à gauche) et  $\mu = 5 \text{ Debye}$  (à droite). La valeur théorique asymptotique est représentée en pointillés.

Le dipôle moyen en Debye est représenté sur la figure 3.3 pour une valeur du dipôle de saturation de  $\mu = 1,85 \text{ Debye}$  et  $\mu = 5 \text{ Debye}$ , ainsi que la courbe correspondant à la

valeur asymptotique théorique. On observe qu'à partir d'environ  $R = 7 \text{ \AA}$ , la courbe du modèle non local est confondue avec la courbe théorique asymptotique. Avant cette valeur de  $R$ , on observe des oscillations autour de la valeur asymptotique, dues aux interactions dipôle/dipôle. Quelle que soit la valeur du dipôle de saturation, notre modèle apparaît très proche de la théorie électrostatique.

**•Le potentiel de force moyenne entre deux ions :**

Nous avons calculé le potentiel de force moyenne (PMF) entre une paire d'ions de chlorure de sodium,  $\text{Na}^+$  et  $\text{Cl}^-$ , distants de  $R$ , avec le modèle non local des pseudo-particules polarisables. Pour cela, nous avons calculé la force moyenne induite par le solvant sur chaque ion, puis projeté la différence sur le vecteur unitaire entre les deux ions  $\mathbf{u}_{12}$ . Nous avons donc calculé donc  $F_s(R) = \frac{1}{2} \langle (\mathbf{F}_1 - \mathbf{F}_2) \cdot \mathbf{u}_{12} \rangle$  pour un certain nombre de distances  $R$  entre les deux ions, entre 2,5 et 8  $\text{Å}$ . La force totale entre les deux ions s'exprime par la somme de la force due au solvant et de la force dans le vide, due aux interactions de Lennard-Jones et de Coulomb entre les deux ions :  $F(R) = F_s(R) + F_{LJ}(R) + F_c(R)$ . Le PMF, que l'on note  $W(R)$ , est obtenu par intégration numérique de la force totale. Les études du PMF de la paire d'ions Na-Cl en solvant explicite [Berkowitz *et al.*, 1984, Dang *et al.*, 1990, Guardia *et al.*, 1991] ont montré que le potentiel présente deux minima à différentes distances :

- le premier minimum, appelé CIP - « Contact Ion Pair », correspond à la distance  $R$  pour laquelle les deux ions sont en contact ;
- le deuxième minimum, appelé SSIP - « Solvent Separated Ion Pair », correspond aux deux ions séparés par une particule de solvant.

La barrière de potentiel entre ces deux minima, ainsi que leurs valeurs relatives, varient selon le modèle d'eau utilisé et les conditions de la simulation. Ainsi, parfois le premier minimum est plus faible que le second [Berkowitz *et al.*, 1984, Dang *et al.*, 1990], et parfois c'est le minimum SSIP qui est plus faible [Guardia *et al.*, 1991]. Les barrières de dissociation et d'association sont en général autour d'une dizaine de kJ/mol.

Nous avons représenté sur la figure 3.4 le PMF calculé avec la méthode non locale sur particules pour deux valeurs différentes du dipôle de saturation. Par comparaison, nous avons tracé le PMF issu de la théorie électrostatique continue non locale sur grille (d'après les équations macroscopiques). On observe avec la méthode /sur particules une structure du PMF avec un premier creux qui correspond au « Contact Ion Pair » (CIP) des modèles explicites, puis une barrière de potentiel, moins importante pour une valeur de  $\mu$  plus élevée, et enfin un deuxième creux correspondant au minimum « Separated Ion Pair » (SSIP). Le minimum de potentiel de SSIP est plus important pour une valeur de saturation élevée.

Nous observons donc avec notre modèle non local sur particules un PMF de même

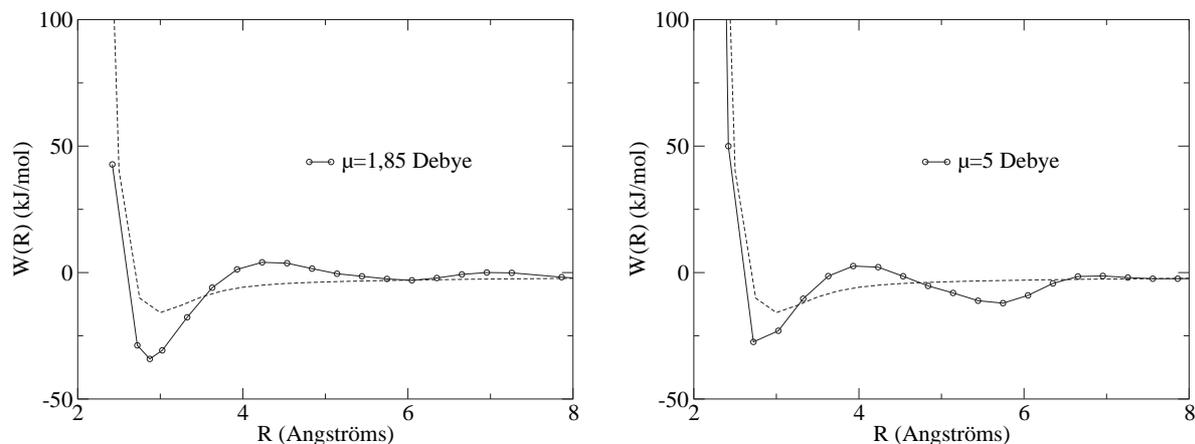


FIG. 3.4 : *PMF* entre deux ions de charges avec le modèle non local, pour une valeur du dipôle de saturation  $\mu = 1,85$  Debye (à gauche) et  $\mu = 5$  Debye (à droite). La valeur du *PMF* calculé avec la méthode non locale sur grille est représentée en pointillés.

type que celui des modèles explicites, avec les mêmes minima CIP et SSIP que ceux observés avec les modèles explicites pour une paire d'ions Na-Cl [Guardia *et al.*, 1991]. La valeur du dipôle de saturation influence la valeur de la barrière de potentiel entre les deux minima. C'est la prise en compte de l'aspect moléculaire du solvant qui permet de reproduire la forme du *PMF* des méthodes explicites avec ses deux minima, que les modèles continus sur grille ne peuvent reproduire.

Avec le modèle des pseudo-particules polarisables non local, une valeur du dipôle de saturation de  $\mu = 1,85$  Debye semble reproduire assez bien les propriétés électrostatiques d'un milieu diélectrique autour d'un ion, ainsi que le *PMF* entre deux ions des modèles explicites. De plus, pour une valeur de  $\mu$  plus faible, la méthode itérative convergera plus vite. L'introduction d'une saturation diélectrique semble donc adaptée pour reproduire l'électrostatique du solvant autour de systèmes simples. Dans le but d'utiliser ce modèle pour simuler des macromolécules biologiques, il faudrait effectuer un paramétrage plus complet, à l'image de celui effectué pour le modèle local dans la dernière partie de ce chapitre. Mais le modèle non local demeure assez coûteux en temps de calcul, et, pour cette raison, nous avons préféré utiliser le modèle local, plus performant numériquement, comme nous le verrons plus loin.

### 3.1.3 Étude de la relation entre $\varepsilon$ et $\alpha$ pour le modèle non local

Lors de la discrétisation de la fonctionnelle sur particules, nous avons introduit la polarisabilité  $\alpha$  des molécules, grandeur microscopique, reliée à la constante diélectrique

du milieu  $\varepsilon$ , grandeur macroscopique, par la relation (3.10), qui correspond à la relation connue de Clausius-Mosotti. Mais, si nous choisissons pour nos particules la polarisabilité  $\alpha$  correspondant à  $\varepsilon = 80$  par cette relation, nous observons dans les simulations que l'équilibre des dipôles n'était jamais atteint. Nous pouvons donc s'interroger sur la validité de cette relation.

Lorsque nous avons discrétisé la fonctionnelle, nous avons supposé que le champ électrique auquel était soumis la particule était le champ électrique créé par toutes les autres particules au centre de la particule. Nous avons alors fait une approximation, car la particule n'est pas un point sur une grille mais une cavité de volume  $\Delta v$ . Selon la théorie électrostatique utilisée pour calculer le champ dans cette cavité (Lorentz ou Onsager), on va obtenir soit la relation de Clausius-Mosotti, soit la relation d'Onsager, qui dépend d'un paramètre  $x = \rho\Delta v$ , pour exprimer  $\alpha$  en fonction de  $\varepsilon$ . Ces deux relations sont équivalentes pour des milieux faiblement diélectriques ou pour  $x = 1$ , mais différent pour des fluides fortement polarisables ou pour  $x < 1$ . Afin de savoir quelle est la relation vérifiée dans notre fluide polarisable, nous allons donc calculer la constante diélectrique effective  $\varepsilon$  obtenue pour une polarisabilité  $\alpha$  donnée des particules. Ainsi, nous pourrons par la suite utiliser la polarisabilité qui correspond à la constante diélectrique de 80 de l'eau. La relation entre  $\varepsilon$  et  $\alpha$  est donc la clé de ce modèle non local, car la connaissance de cette relation nous permettra de reproduire correctement les propriétés diélectriques de l'eau. En plus de l'information importante que cette étude apporte à notre modèle, il est intéressant d'étudier la validité des lois macroscopiques de l'électrostatique à l'échelle microscopique.

### Problématique de la relation entre $\varepsilon$ et $\alpha$

Les deux relations entre  $\varepsilon$  et  $\alpha$  théoriquement connues pour un fluide polarisable sont la relation de Clausius-Mosotti et la relation d'Onsager. Nous allons rappeler l'expression de ces deux relations et leur domaine de validité, puis montrer qu'elles sont équivalentes pour des milieux faiblement diélectriques en effectuant le développement limité de  $\varepsilon$  en fonction de  $\rho\alpha$  pour les deux relations.

#### •La relation de Clausius-Mosotti :

La relation de Clausius-Mosotti s'écrit :

$$\frac{\rho\alpha}{3\varepsilon_0} = \frac{\varepsilon - 1}{\varepsilon + 2}. \quad (3.23)$$

C'est la relation (3.10) qui est venue naturellement dans notre étude, en supposant que le champ électrique local dont dépend le moment dipolaire de la particule est le champ de

Lorentz au centre de la particule. Cette relation diverge à partir d'une certaine valeur de la polarisabilité,  $\rho\alpha_c = 3\varepsilon_0 \simeq 0,24$ , correspondant à l'asymptote de la courbe.

L'étude des propriétés électrostatiques des fluides non polaires polarisables effectuée par Alder *et al.* [Alder et Pollock, 1981, Pollock et Alder, 1978] s'est intéressée à la validité de l'équation de Clausius-Mosotti pour des systèmes de Lennard-Jones de densité caractéristique des liquides mais de faible polarisabilité. À partir de l'expression du tenseur dipolaire, ils calculent le développement limité du champ électrique autour d'un ion en fonction de  $\alpha$ , qui donne au premier ordre la relation de Clausius-Mosotti. Ils trouvent que les corrections à partir du second ordre sont négligeables pour les cas à faible polarisabilité, et ils confirment donc la validité de la relation de Clausius-Mosotti dans ce cas précis. Dans une autre étude [Pollock et Alder, 1977], ils observent une « catastrophe de polarisation », c'est-à-dire une polarisation spontanée du milieu, pour des valeurs plus élevées de la polarisabilité, à partir d'une valeur critique  $\alpha_c/\sigma^3 \simeq 0,2$ , soit, pour un liquide où  $\rho^* = \rho\sigma^3 = 0,8$ , à partir d'une valeur de  $\rho\alpha_c \simeq 0,16$ . Pourtant, la relation de Clausius-Mosotti diverge pour une valeur critique de la polarisabilité beaucoup plus élevée, pour  $\rho\alpha_c \simeq 0,24$ . La relation de Clausius-Mosotti ne semble donc pas être vérifiée dans le cas des fluides fortement polarisables.

#### •La relation d'Onsager :

La relation d'Onsager s'exprime en fonction d'un paramètre  $x = \rho\Delta v$  ( $0 < x \leq 1$ ) à déterminer pour le fluide considéré :

$$\frac{\rho\alpha}{3x\varepsilon_0} = \frac{\varepsilon - 1}{\varepsilon + 2 + \frac{9\varepsilon}{2\varepsilon+1}(x-1)}. \quad (3.24)$$

Pour  $x = 1$ , on retrouve exactement la relation de Clausius-Mosotti. L'asymptote de la courbe est en  $\rho\alpha_c = 3x\varepsilon_0$ , au lieu de  $\rho\alpha_c = 3\varepsilon_0$  pour Clausius-Mosotti. Pour  $x < 1$ , l'asymptote est plus proche de l'origine que celle de Clausius-Mosotti. La relation diverge donc à partir d'une valeur critique de la polarisabilité plus faible que la relation de Clausius-Mosotti. Cette valeur critique est déterminée par la valeur du paramètre  $x$ .

Dans un article de revue sur les traitements moléculaires des phénomènes diélectriques [Madden et Kivelson, 1984], le cas des particules non polaires polarisables de rayon  $r_0$ , donc de volume  $\Delta v = \frac{4}{3}\pi r_0^3$ , est étudié à partir d'une théorie microscopique. L'approximation qu'ils proposent consiste à supposer que la fonction de distribution des  $N$  particules  $g^N(\mathbf{r}_1, \dots, \mathbf{r}_N)$  s'écrit comme un produit de fonctions de distribution de paires simples de la forme  $1 - \theta(r_{ij} - 2r_0)$  où  $\theta$  est la fonction de Heavyside. Les auteurs expriment le champ électrique qu'ils appellent « local », c'est-à-dire le champ électrique utilisé pour calculer les moments dipolaires des particules par la formule (3.13). Ils retrouvent alors la théorie d'Onsager [Onsager, 1936] qui permet de calculer le champ électrique s'exerçant

sur chaque particule d'un fluide polarisable, en considérant la particule comme une cavité sphérique de volume  $\Delta v$  et de polarisabilité  $\alpha$  plongée dans un milieu diélectrique continu de constante diélectrique  $\varepsilon$ . L'expression de ce champ, à la place de l'expression du champ de Lorentz, permet de montrer que la relation vérifiée entre  $\varepsilon$  et  $\alpha$  est celle donnée par l'équation (3.24) et s'exprime en fonction du paramètre  $x = \rho\Delta v$ .

Afin de vérifier que les relations d'Onsager et de Clausius-Mosotti sont équivalentes à faibles polarisabilités, nous avons effectué le développement limité de  $\varepsilon$  en fonction de  $K = \rho\alpha/\varepsilon_0$ . Pour la relation de Clausius-Mosotti, le développement limité donne au voisinage de  $K = 0$  :

$$\varepsilon_{CM} = 1 + K + \frac{K^2}{3} + \frac{K^3}{9} + o(K^3), \quad (3.25)$$

alors que le développement limité de  $\varepsilon$  pour la relation d'Onsager donne :

$$\varepsilon_{Ons} = 1 + K + \frac{K^2}{3} + \frac{K^3}{9} \left( \frac{2}{x} - 1 \right) + o(K^3). \quad (3.26)$$

D'après ces deux relations, on voit que  $\varepsilon_{CM}$  et  $\varepsilon_{Ons}$  sont égaux jusqu'au troisième ordre. Cela confirme que dans l'étude d'Alder *et al.* qui s'intéressait au développement limité du champ électrique jusqu'au second ordre en  $\alpha$ , la relation de Clausius-Mosotti n'était alors pas mise en défaut. Pour des  $\alpha$  petits, les relations de Clausius-Mosotti et d'Onsager sont donc équivalentes, mais, dans le cas  $x \neq 1$ , elles diffèrent lorsque  $\rho\alpha$  se rapproche de la valeur critique  $\rho\alpha_c = 3x\varepsilon_0$ .

Les deux relations donnant  $\varepsilon$  en fonction de  $\rho\alpha$  sont représentées sur la courbe 3.5, la relation d'Onsager étant représentée pour une valeur de  $x < 1$  (ici  $x = 0,74$ ). On voit que jusqu'à  $\rho\alpha \simeq 0,12$  (correspondant à  $\alpha = 4,1 \text{ \AA}^3$  pour une densité de  $\rho = 0,029 \text{ \AA}^{-3}$ ) Clausius-Mosotti et Onsager sont confondues.

### Relation entre $\varepsilon$ et $\alpha$ obtenue avec le modèle non local

La relation entre  $\varepsilon$  et  $\alpha$  nécessite d'être étudiée à l'aide de notre modèle non local de pseudo-particules polarisables de Langevin. Nous allons donc calculer la constante diélectrique effective  $\varepsilon$  obtenue avec nos pseudo-particules ayant une polarisabilité  $\alpha$  donnée. Pour cela, nous avons utilisé le même programme de dynamique moléculaire « maison » que pour les tests présentés au paragraphe précédent ; ce programme fixe le soluté et laisse les particules de solvant libres, en résolvant de façon itérative les équations auto-cohérentes. Les paramètres de Lennard-Jones des particules de solvant seront les mêmes que précédemment,  $\sigma_s = 3,024 \text{ \AA}$  et  $\epsilon_s = 3,197 \text{ kJ/mol}$ , la température  $T = 300 \text{ K}$  et le

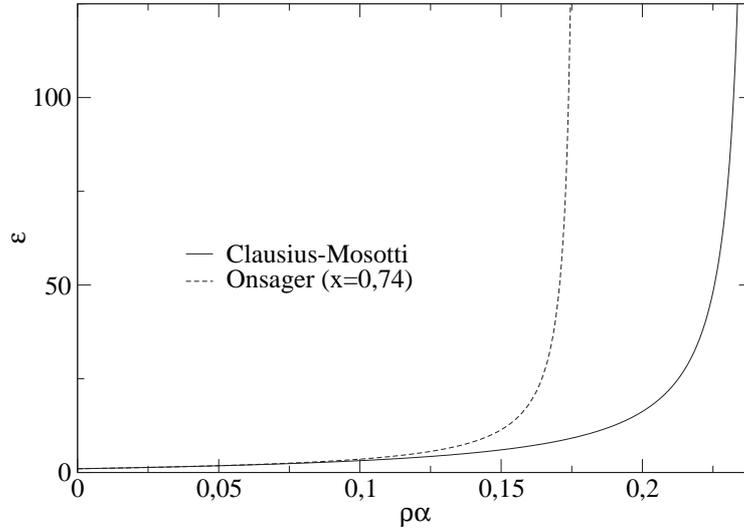


FIG. 3.5 : Représentation des deux relations différentes donnant  $\varepsilon(\alpha)$  : Clausius-Mosotti et Onsager sont confondues pour de faibles valeurs de  $\rho\alpha$ , mais l'asymptote de la relation de Clausius-Mosotti est en  $\rho\alpha = 3\varepsilon_0$  alors que celle d'Onsager est en  $\rho\alpha = 3x\varepsilon_0$  (avec ici  $x = 0,74$ ).

pas d'intégration de  $\delta t = 10$  fs. Le dipôle de saturation est fixé à  $\mu = 1,85$  Debye. Les conditions aux limites sont périodiques, et les interactions longue distance sont coupées à une distance de  $12 \text{ \AA}$  avec la convention de l'image minimum afin de limiter les effets de bords.

Nous considérons un fluide composé de particules polarisables de polarisabilité  $\alpha$  donnée et de densité  $\rho$  autour d'un ion, et nous cherchons à déterminer la constante diélectrique  $\varepsilon$  effective obtenue. Pour calculer cette constante diélectrique effective  $\varepsilon$ , on connaît la formule théorique donnant l'expression de la polarisation asymptotiquement autour d'un ion [Pollock *et al.*, 1980] en fonction du champ créé par l'ion dans le vide et de  $\varepsilon$ , déjà présentée au paragraphe précédent par l'équation (3.22) :

$$\lim_{R \rightarrow \infty} P(R) = \frac{\varepsilon_0(\varepsilon - 1)}{\varepsilon} E_0(R), \quad (3.27)$$

où  $E_0(R)$  est le champ à symétrie sphérique créé par l'ion dans le vide à la distance  $R$  de l'ion. Pour des particules discrètes de moment dipolaire  $\mathbf{p}_i$ , on relie le champ de polarisation  $P(R)$  aux moments dipolaires microscopiques par la formule :

$$P(R) = \frac{\langle \mu(R) \rangle}{\Delta v} = \rho \langle \mu(R) \rangle, \quad (3.28)$$

où  $\Delta v$  est le volume occupé par une particule et  $\rho$  est la densité des particules,  $\mu(R)$  est la moyenne sur la couche sphérique de rayon  $R$  et d'épaisseur  $\delta R$ , pour une configuration

donnée, des dipôles  $\mathbf{p}_i$  projetés selon la direction de  $\mathbf{E}_0(\mathbf{R})$  (car le problème électrostatique est à symétrie sphérique), et  $\langle \rangle$  représente la moyenne sur un certain nombre de configurations des particules.

Ainsi, pour calculer  $\varepsilon$  effectif, nous allons calculer un facteur  $S(R)$  défini de la façon suivante :

$$S(R) = \frac{\langle \mu(R) \rangle}{\varepsilon_0 E_0 / \rho}. \quad (3.29)$$

D'après la définition de  $S(R)$  et la formule théorique asymptotique (3.27) on a :

$$S_\infty = \lim_{R \rightarrow \infty} S(R) = \frac{\varepsilon - 1}{\varepsilon}, \quad (3.30)$$

ce qui nous permettra d'en déduire la constante diélectrique recherchée  $\varepsilon$  :

$$\varepsilon = \frac{1}{1 - S_\infty}. \quad (3.31)$$

Pour chaque simulation à  $\rho$  et  $\alpha$  donnés, nous allons donc tracer  $S(R)$ , déterminer  $S_\infty$  et en déduire  $\varepsilon$ . En faisant plusieurs simulations avec des  $\alpha$  différents, nous pourrions tracer  $\varepsilon(\alpha)$ , ou  $\varepsilon(\rho\alpha)$ . Nous allons comparer la courbe obtenue avec les courbes connues, pour déterminer quelle relation est vérifiée, et la valeur éventuelle du facteur  $x$  (on sait que  $x = \rho\Delta v$  mais on ne peut le calculer d'après les paramètres de nos particules car ce sont des particules de Lennard-Jones et non des sphères dures). Notons que dans le vide on aura  $S_\infty = 0$  car  $\varepsilon = 1$ .

Afin de tracer la relation donnant  $\varepsilon$  en fonction de  $\alpha$ , nous avons fait varier  $\alpha$  entre 1 et  $6 \text{ \AA}^3$ . La charge de l'ion est choisie à  $q = +0,5 e$  (afin que le champ électrique créé ne soit pas trop fort et que la méthode itérative converge plus vite), et nous avons considéré  $N = 864$  particules de solvant mélangées. Nous avons calculé  $S(R)$  comme une moyenne des configurations tous les 10 pas sur une dynamique d'au moins 500 ps. Plus  $\alpha$  est élevée, plus les oscillations sont importantes pour  $S(R)$  et plus il sera difficile de déterminer  $S_\infty$ , donc plus il faudra faire une dynamique longue pour avoir plus de statistiques et donc une courbe moyenne plus lisse. Malgré la convention de l'image minimum, on observe des effets de bords à une distance  $R$  élevée de l'ion. Pour déterminer  $S_\infty$ , nous avons coupé la courbe de  $S(R)$  avant les effets de bords et nous avons pris la valeur moyenne entre les deux extrema de la dernière oscillation, comme représenté sur la figure 3.6 pour le cas où  $\alpha = 4 \text{ \AA}^3$ . Nous avons vérifié que la valeur de  $S_\infty$  ne dépendait ni du nombre de particules de solvant, ni de la valeur du  $\mu$  de saturation, ni de la charge  $q$  de l'ion, en effectuant quelques simulations en changeant ces paramètres : pour  $\mu = 5$  Debye, avec  $N = 500$  et  $N = 1372$  particules, et pour  $q = +1 e$ . Ces résultats ne sont pas présentés ici.

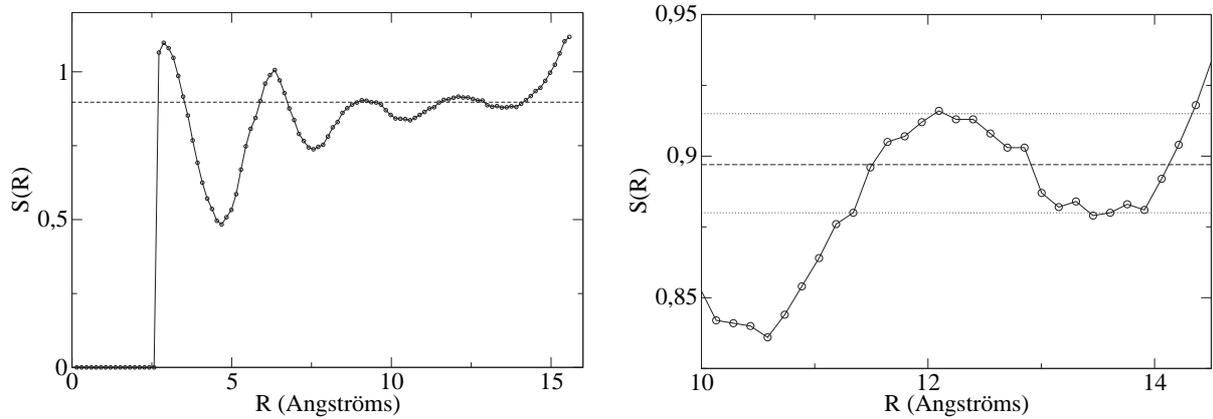


FIG. 3.6 : Valeur calculée du facteur  $S(R)$ , ici pour  $\alpha = 4 \text{ \AA}^3$ , à gauche pour  $R \in [0; 16 \text{ \AA}]$ , et à droite pour  $R \in [10; 15 \text{ \AA}]$ . La valeur de  $S_\infty$ , représentée par les traits pointillés, est la moyenne des deux valeurs extrêmes de la dernière oscillation, représentées par les pointillés sur la figure de droite.

Les points obtenus pour  $\varepsilon(\rho\alpha)$  avec le modèle non local des pseudo-particules polarisables sont représentés sur la figure 3.7. Nous avons tracé Clausius-Mosotti et Onsager pour différentes valeurs de  $x$  :  $x = 0,74$  et  $x = 0,75$ . On voit donc que la courbe obtenue par dynamique moléculaire colle extrêmement bien à la relation d'Onsager pour  $x = 0,74$ . Cependant, notons que la valeur exacte de  $x$  est difficile à déterminer précisément à cause de l'incertitude sur la mesure de  $S_\infty$ , donc sur  $\varepsilon$ , surtout à  $\varepsilon$  élevé, car l'incertitude absolue sur  $\varepsilon$  vaut  $\Delta\varepsilon = \varepsilon^2 \Delta S_\infty$ . Nous constatons donc que la valeur de la constante diélectrique effective d'un fluide polarisable ne vérifie la relation de Clausius-Mosotti que pour des valeurs de  $\rho\alpha$  faibles, et que la relation obtenue pour  $\varepsilon(\alpha)$  est la relation d'Onsager pour  $x < 1$ , qui diverge en  $\rho\alpha_c = 3x\varepsilon_0$ . Mais du fait de l'incertitude sur les valeurs de  $\varepsilon$  élevées, il n'est pas possible de déterminer très précisément l'asymptote de la courbe, et donc la valeur de  $x$ . Pour préciser cette valeur, nous pouvons la calculer d'une manière différente en étudiant la divergence qui s'opère dans le fluide à partir d'une certaine valeur de  $\alpha_c$ , comme nous allons le voir maintenant.

En effet, lorsque nous choisissons pour les particules une valeur de la polarisabilité située après l'asymptote de la courbe, nous observons alors une transition électrostatique dans le fluide, c'est-à-dire que les dipôles sont non nuls même sans la présence d'un champ extérieur (donc sans ion). C'est la catastrophe de polarisation, ou polarisation spontanée, observée par Pollock et Alder [1977]. L'étude de cette transition électrostatique aux alentours de la valeur de  $\rho\alpha_c$  critique de l'asymptote nous permettra de préciser la valeur exacte du paramètre  $x$ , ainsi que de comprendre la nature de la transition électrostatique.

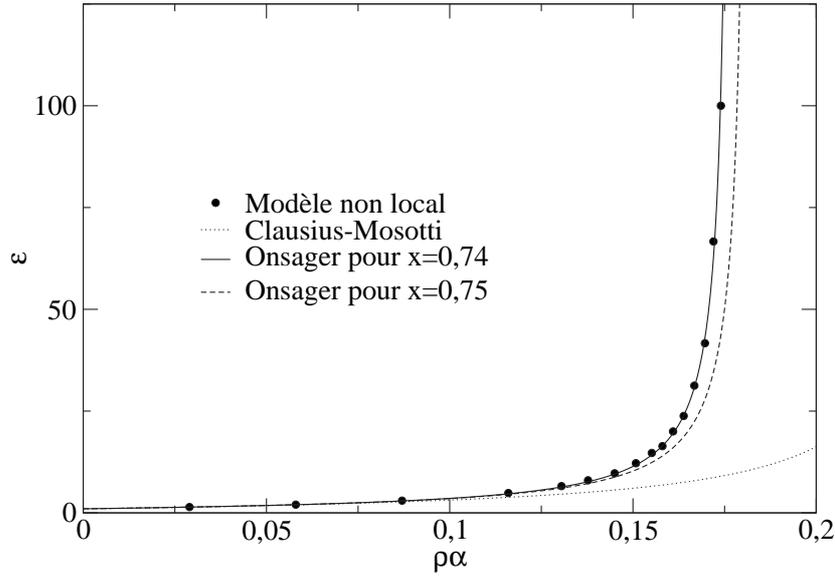


FIG. 3.7 : Valeurs de la constante diélectrique effective  $\varepsilon$  pour un fluide polarisable en fonction de  $\rho\alpha$ , comparées aux relations de Clausius-Mosotti et d'Onsager pour  $x = 0,74$  et  $x = 0,75$  : on voit que les valeurs calculées sont très proches de la courbe d'Onsager pour  $x = 0,74$ .

### Étude de la transition électrostatique dans notre fluide polarisable

Nous allons ainsi effectuer une dynamique avec nos pseudo-particules polarisables de Langevin, sans ion, à différents  $\alpha$  autour du  $\alpha_c$  critique présumé par l'étude de la relation  $\varepsilon(\alpha)$  précédente. La valeur de  $\alpha_c$  sera la valeur de la polarisabilité à partir de laquelle apparaîtront spontanément des dipôles induits non nuls.

Afin d'étudier la transition électrostatique dans le fluide, nous avons calculé la polarisation moyenne du système pour  $N$  particules de solvant de deux façons :

$$\langle P \rangle = \left\langle \frac{1}{N} \sum_i |\mathbf{p}_i| \right\rangle, \quad (3.32)$$

$$\langle P' \rangle = \left\langle \left| \frac{1}{N} \sum_i \mathbf{p}_i \right| \right\rangle. \quad (3.33)$$

En traçant  $\langle P \rangle$  et  $\langle P' \rangle$  en fonction de  $\alpha$ , on doit observer avant la transition  $\langle P \rangle = \langle P' \rangle = 0$ , et après la transition :

- $\langle P \rangle \neq 0$  et  $\langle P' \rangle \neq 0$  si la transition est ferroélectrique,
- ou  $\langle P \rangle \neq 0$  et  $\langle P' \rangle = 0$  si la transition est antiferroélectrique.

Nous avons étudié un système de  $N = 500$  particules de solvant avec des configurations des dipôles aléatoires au début de la simulation. Nous avons effectué des dynamiques

sur une durée d'environ 500 ps. Pour les valeurs de  $\alpha$  proches de la transition, comme un grand nombre d'itérations est nécessaire pour atteindre le critère de convergence, les  $\mathbf{p}_i$  à l'équilibre ne sont calculés par la méthode itérative que tous les 100 pas, et les moyennes  $\langle P \rangle$  et  $\langle P' \rangle$  sont calculées comme des moyennes de ces dipôles tous les 100 pas. Nous avons vérifié que les grandeurs  $\langle P \rangle$  et  $\langle P' \rangle$  dépendaient bien du produit  $\rho\alpha$  en effectuant des simulations à différents  $\rho$ , et la transition se produit pour la même valeur critique de  $\rho\alpha_c$ .

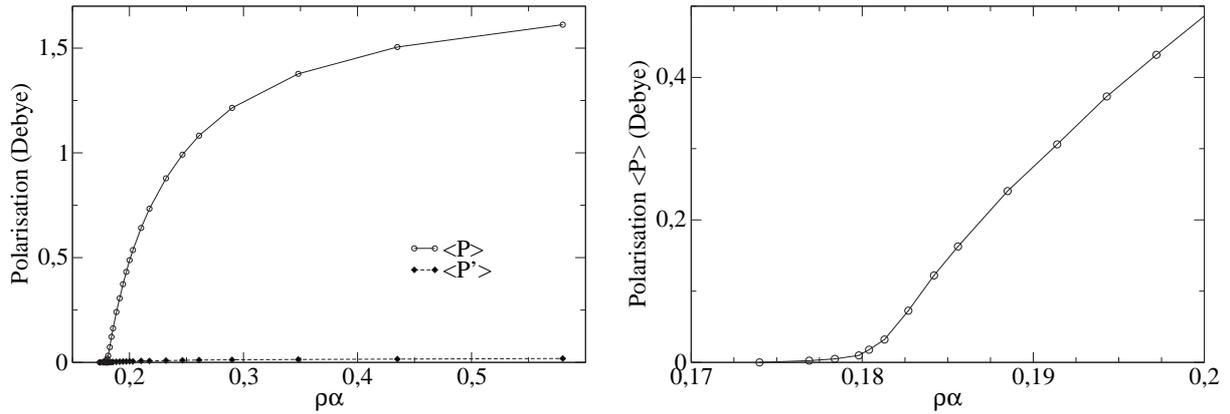


FIG. 3.8 : Polarisation moyenne des particules en fonction de la densité et de la polarisation, à gauche pour  $\rho\alpha \in [0; 0,6]$  et à droite pour  $\rho\alpha \in [0,17; 0,2]$  : on observe une transition antiferroélectrique vers  $\rho\alpha_c = 0,178$ .

La courbe obtenue, tracée sur la figure 3.8, donnant  $\langle P \rangle$  et  $\langle P' \rangle$  en fonction de  $\rho\alpha$  (tracée pour  $\rho = 0,029 \text{ \AA}^{-3}$ ) montre clairement que la transition est antiferroélectrique, et qu'elle se produit entre  $\rho\alpha = 0,177$  et  $0,18$ . À cause des effets de taille finie du système, la transition n'est pas franche, et il n'est possible de déterminer  $\alpha_c$  encore une fois que de manière peu précise. La valeur de  $\rho\alpha_c$  obtenue par cette étude nous donne néanmoins la valeur du paramètre  $x$  de la relation d'Onsager, puisque l'asymptote de la relation d'Onsager est  $\rho\alpha_c = 3x\varepsilon_0$ . Nous avons ainsi trouvé  $x \sim 0,74$ , puisque :

$$\begin{cases} x = 0,74 \Leftrightarrow \rho\alpha_c = 0,1767 \\ x = 0,75 \Leftrightarrow \rho\alpha_c = 0,1790 \end{cases}$$

L'étude de la relation entre  $\varepsilon$  et  $\alpha$  effectuée ici est valable pour des fluides non polaires polarisables. Nous avons donc montré que la constante diélectrique effective pour un fluide fortement polarisable ne vérifiait pas la relation de Clausius-Mosotti, mais celle d'Onsager, qui diverge à partir de  $\rho\alpha$  plus faibles que Clausius-Mosotti. Cependant, pour les milieux faiblement diélectriques, comme les deux relations sont équivalentes, il est possible d'utiliser la relation de Clausius-Mosotti. Au-delà de la valeur critique de  $\rho\alpha_c$  de

l'asymptote de la relation d'Onsager, nous avons observé une transition électrostatique dans le fluide dont nous avons montré qu'elle était de nature antiferroélectrique.

Pour le cas qui nous intéresse ici, c'est-à-dire l'eau, la conclusion que nous tirons de cette étude est que pour reproduire correctement les propriétés diélectriques de l'eau avec notre modèle non local de particules polarisables, nous devons prendre comme valeur de la polarisabilité des particules le  $\alpha$  correspondant à  $\varepsilon = 80$  de l'eau par la relation d'Onsager pour  $x = 0,74$ , au lieu de la relation de Clausius-Mosotti, c'est-à-dire que nous aurons  $\alpha = 5,97 \text{ \AA}^3$  pour  $\rho = 0,029 \text{ \AA}^{-3}$ , alors qu'en utilisant Clausius-Mosotti pour les mêmes paramètres nous aurions  $\alpha = 7,95 \text{ \AA}^3$ .

### Conclusion sur le modèle non local

Le modèle non local des pseudo-particules polarisables, reproduit bien l'électrostatique, comme nous l'avons vu dans les quelques tests effectués, à condition de trouver la bonne relation reliant  $\varepsilon$  et  $\alpha$ . Il a cependant l'inconvénient d'être très coûteux en temps de calcul à cause de la résolution itérative des équations auto-cohérentes, et il serait donc peu intéressant pour des dynamiques de gros systèmes. À la place de la méthode itérative, une minimisation de la fonctionnelle par des méthodes de type Car-Parinello à l'image de la méthode sur grille développée au laboratoire de Modélisation des Systèmes Moléculaires Complexes [Marchi *et al.*, 2001] pour la fonctionnelle d'énergie libre de polarisation macroscopique (présentée dans le chapitre précédent) permettrait de gagner du temps de calcul. Une telle méthode n'a pas encore été développée mais constitue un axe de recherche futur intéressant. Par ailleurs, la coupure des interactions électrostatiques à une distance donnée provoque des effets de bords, et il faudrait envisager le calcul des interactions électrostatiques longue portée par une méthode de type Particle Mesh Ewald (PME), plus coûteuse que les méthodes de coupure. De ces différents constats, il paraît intéressant de développer une méthode plus performante en terme de temps de calcul, en faisant au besoin quelques approximations.

#### 3.1.4 Modèle local

Nous allons présenter dans ce paragraphe une simplification du modèle non local, appelé modèle local car l'approximation effectuée est locale dans l'espace. Après avoir effectué une approximation sur la polarisation, on obtient des équations pour les dipôles induits des particules qui ne sont plus auto-cohérentes, et on résout ainsi les problèmes de convergence du modèle non local.

### Les équations du modèle avec l'approximation locale

L'approximation dite « locale » consiste à supposer que la polarisation est longitudinale, c'est-à-dire qu'il existe une fonction  $f(\mathbf{r})$  telle que  $\mathbf{P}(\mathbf{r}) = -\nabla f(\mathbf{r})$  [Calef et Wolynes, 1983]. Il a été démontré que l'expression du champ électrique de polarisation, défini par l'équation (3.4), est alors la suivante [Calef et Wolynes, 1983, Kharkats *et al.*, 1976] :

$$\mathbf{E}_p(\mathbf{r}) = -\frac{\mathbf{P}(\mathbf{r})}{\varepsilon_0}. \quad (3.34)$$

La fonctionnelle non locale de Marcus donnée par (3.1) s'exprime alors de façon simplifiée. On remplace le champ électrique de polarisation par sa nouvelle expression et on obtient la fonctionnelle locale de Wolynes :

$$\mathcal{F}_{LP}[\mathbf{P}(\mathbf{r})] = \frac{1}{2} \int \left(1 + \frac{1}{\chi(\mathbf{r})}\right) \frac{\mathbf{P}(\mathbf{r})^2}{\varepsilon_0} d\mathbf{r} - \int \mathbf{P}(\mathbf{r}) \cdot \mathbf{E}_0(\mathbf{r}) d\mathbf{r}. \quad (3.35)$$

Dans cette fonctionnelle, le tenseur dipolaire n'intervient plus. Les interactions du milieu polarisable avec lui-même sont alors prises en compte dans la première intégrale. À l'équilibre thermodynamique, la fonctionnelle est minimale par rapport à  $\mathbf{P}(\mathbf{r})$  et on obtient l'expression suivante pour la densité de polarisation à l'équilibre :

$$\mathbf{P}(\mathbf{r}) = \left(\frac{\chi(\mathbf{r})}{\chi(\mathbf{r}) + 1}\right) \varepsilon_0 \mathbf{E}_0(\mathbf{r}). \quad (3.36)$$

D'après la définition du déplacement diélectrique à l'équilibre donnée par (3.5), il est facilement possible de montrer que la condition sur la polarisation (3.34) revient à faire l'approximation que le champ électrique dans le vide et le déplacement diélectrique sont identiques en tout point de l'espace :

$$\mathbf{D}(\mathbf{r}) = \varepsilon_0 \mathbf{E}_0(\mathbf{r}). \quad (3.37)$$

Notons que cette approximation (3.37) est la même que celle effectuée dans la méthode analytique de Born Généralisé vue au chapitre 2, par l'équation (2.13) [Schaefer et Karplus, 1996]. Sans approximation, il est connu que les divergences de ces deux vecteurs sont égales d'après l'équation de Maxwell :

$$\nabla \cdot \mathbf{D}(\mathbf{r}) = \nabla \cdot [\varepsilon_0 \mathbf{E}_0(\mathbf{r})], \quad (3.38)$$

ce qui veut dire qu'en réalité les deux vecteurs sont égaux à un rotationnel près. Cette approximation est exacte dans certains cas, soit pour un milieu diélectrique homogène, soit pour un soluté de géométrie sphérique [Kharkats *et al.*, 1976], c'est-à-dire lorsque  $\mathbf{E}_0(\mathbf{r}) \times \nabla \varepsilon(\mathbf{r}) = 0$ .

Pour développer le modèle sur particules, nous discrétisons alors la fonctionnelle macroscopique locale sur  $N$  particules de densité  $\rho$  comme pour le modèle non local (cf. (3.11)), avec  $\Delta v = 1/\rho$  et  $\mathbf{P}(\mathbf{r}_i) = \mathbf{p}_i/\Delta v$ . De même, nous négligeons la polarisabilité du soluté, et la somme discrète court sur toutes les particules de solvant. Nous obtenons une fonction locale de l'énergie libre :

$$F_{LP}(\{\mathbf{p}_i\}) = \sum_i \frac{\mathbf{p}_i^2}{2\alpha_e} - \sum_i \mathbf{p}_i \cdot \mathbf{E}_i^0, \quad (3.39)$$

où la polarisabilité atomique  $\alpha_e$  des particules est donnée par la formule :

$$\alpha_e = \frac{\varepsilon_0}{\rho} \left( \frac{\chi}{1 + \chi} \right) = \frac{\varepsilon_0(\varepsilon - 1)}{\rho\varepsilon}. \quad (3.40)$$

Notons que, pour le modèle local, il n'y a pas à étudier la relation entre  $\varepsilon$  et  $\alpha_e$  comme pour le modèle non local, car le champ électrique qui intervient ici n'est plus le champ de Maxwell mais simplement le champ électrique créé par le soluté dans le vide. Dans la fonctionnelle discrétisée, il n'y a donc plus de terme d'interaction dipôle/dipôle, car cette interaction est en fait prise en compte dans la première somme.

De la même manière que le modèle non local, la minimisation de la fonctionnelle par rapport aux  $\mathbf{p}_i$  donne pour l'expression des dipôles et de l'énergie libre correspondante à l'équilibre :

$$\mathbf{p}_i = \alpha_e \mathbf{E}_i^0, \quad (3.41)$$

$$F_{LP}^{eq} = -\frac{1}{2} \sum_i \mathbf{p}_i \cdot \mathbf{E}_i^0. \quad (3.42)$$

On voit donc que dans le modèle local, le moment dipolaire induit des particules est directement proportionnel au champ électrique créé par le soluté dans le vide, et les dipôles induits sont donc toujours orientés dans le sens de ce champ électrique. Les équations ne sont plus auto-cohérentes, et la résolution de ces équations sera donc beaucoup plus simple et plus rapide que pour le modèle non local.

Les forces électrostatiques dues au solvant s'expriment alors pour une particule  $i$  de solvant :

$$\mathbf{F}_i = -\frac{\partial F_{LP}^{eq}}{\partial \mathbf{r}_i} = \mathbf{p}_i \cdot \frac{\partial \mathbf{E}_i^0}{\partial \mathbf{r}_i}, \quad (3.43)$$

et pour un atome  $k$  du soluté :

$$\mathbf{F}_k = -\frac{\partial F_{LP}^{eq}}{\partial \mathbf{R}_k} = \sum_i \mathbf{p}_i \cdot \frac{\partial \mathbf{E}_i^0}{\partial \mathbf{R}_k}. \quad (3.44)$$

## La saturation

Pour introduire les effets de la saturation diélectrique de manière analogue au modèle non local, nous partons de la fonctionnelle non locale de Langevin (3.20), et nous faisons l'approximation locale (3.34) dans sa version discrétisée. Nous obtenons alors la fonctionnelle locale de Langevin sur particules qui s'exprime :

$$F_{LL}(\mathbf{p}_i) = \frac{\mu^2}{3\alpha_e} \sum_i \left( \mathcal{L}^{-1}(p_i/\mu) \frac{p_i}{\mu} - \ln \left[ \frac{\sinh[\mathcal{L}^{-1}(p_i/\mu)]}{\mathcal{L}^{-1}(p_i/\mu)} \right] \right) - \sum_i \mathbf{p}_i \cdot \mathbf{E}_i^0. \quad (3.45)$$

Après minimisation de la fonctionnelle par rapport aux  $\mathbf{p}_i$ , nous obtenons les équations analogues à celles du modèle non local de Langevin :

$$\mathbf{p}_i = \frac{\mu}{E_i^0} \mathcal{L} \left( \frac{3\alpha_e E_i^0}{\mu} \right) \mathbf{E}_i^0. \quad (3.46)$$

L'énergie libre du système s'écrit alors au minimum :

$$F_{LL}^{eq} = -\frac{\mu^2}{3\alpha_e} \sum_i \ln \left[ \frac{\sinh(3\alpha E_i^0/\mu)}{3\alpha E_i^0/\mu} \right]. \quad (3.47)$$

Nous verrons plus loin que l'introduction des effets de saturation diélectrique pour le modèle local est justifiée par les calculs du potentiel de force moyenne entre deux ions de charges opposées, qui sont fortement améliorés par l'introduction de la saturation.

De la même manière que pour le modèle non local, l'introduction de la saturation ne modifie pas l'expression des forces, données par les équations (3.43) et (3.44).

Afin de choisir la valeur du dipôle de saturation, nous savons que pour simuler des fluides de Stockmayer, constitués de dipôles permanents sur des particules de Lennard-Jones, afin de représenter un milieu diélectrique de constante  $\varepsilon$  élevée, une valeur de  $\mu = 1,7$  Debye pour le dipôle permanent a été utilisée [Pollock et Alder, 1980]. Par ailleurs, pour la méthode des dipôles de Langevin de Warshel sur une grille de  $3 \text{ \AA}$  (c'est-à-dire l'équivalent du diamètre des molécules d'eau), une valeur du dipôle de saturation de  $\mu = 1,25$  Debye permet de reproduire les énergies de solvation de nombreuses molécules [Florian et Warshel, 1997]. Pour notre modèle local, les dipôles sont à l'équilibre électrostatique et sont constamment orientés dans le sens du champ créé par le soluté, alors que les dipôles permanents des fluides de Stockmayer ne sont pas à l'équilibre et leur orientation fluctue. Nous en déduisons donc que le dipôle de saturation pour notre modèle polarisable local doit être plus faible que le dipôle permanent des fluides de Stockmayer, et donc situé entre 1,25 et 1,7 Debye. La valeur de  $\mu = 1,5$  Debye paraît donc un bon compromis.

## Conclusion sur la méthode locale

Le formalisme de la méthode locale aboutit à un ensemble de particules polarisables dont les moments dipolaires induits s'orientent en fonction du champ électrique créé par le soluté. Ainsi, contrairement à la méthode non locale, comme le moment dipolaire d'une particule ne dépend pas des moments dipolaires des autres particules, les particules de solvant ne se voient entre elles que comme des particules de Lennard-Jones. Cela rend la méthode encore plus performante en temps de calcul, et particulièrement simple à implémenter en dynamique moléculaire.

C'est pourquoi, dans la suite de l'étude, nous avons choisi d'utiliser la méthode locale, qui donne de bons résultats lorsqu'elle est correctement paramétrée. Nous reviendrons sur le paramétrage de la méthode locale dans la troisième partie de ce chapitre, mais nous allons auparavant détailler l'intégration des deux modèles de solvation dans l'algorithme de modélisation moléculaire que nous utiliserons pour la simulation des biomolécules.

## 3.2 Intégration du modèle en dynamique moléculaire

L'intérêt de notre modèle sur particules est qu'il peut être implémenté simplement dans un algorithme de dynamique moléculaire. En effet, nous avons une expression analytique des forces soluté/solvant et solvant/solvant, tant pour le point de vue particulaire (interactions de van der Waals) que pour le point de vue électrostatique (grâce à l'expression de la fonctionnelle). Ainsi, le modèle va pouvoir être intégré en terme d'un champ de force atomique dans l'hamiltonien du système. Avant de présenter comment le modèle de pseudo-particules polarisables est implémenté dans un programme de dynamique moléculaire, nous allons rappeler les principes fondamentaux de la dynamique moléculaire.

### 3.2.1 Le principe de la dynamique moléculaire

Pour modéliser un système moléculaire complexe constitué de nombreux atomes, il faut connaître toutes les interactions qui existent entre ces atomes. La méthode de dynamique moléculaire repose sur une expression de l'hamiltonien du système, c'est-à-dire son énergie totale, composée de l'énergie cinétique et de l'énergie potentielle. Cet

hamiltonien s'exprime pour  $N$  atomes de masse  $m_i$  à la position  $\mathbf{r}_i$  et de vitesse  $\dot{\mathbf{r}}_i$  :

$$H = \sum_{i=1}^N \frac{1}{2} m_i \dot{\mathbf{r}}_i^2 + V, \quad (3.48)$$

où  $V$  est le potentiel d'interaction entre les atomes. La donnée des paramètres de ce potentiel d'interaction s'appelle le champ de force du système, car les forces exercées sur les atomes du système dérivent de ce potentiel.

### Le champ de force

Le potentiel d'interaction entre les atomes peut être décomposé en deux parties distinctes : le potentiel d'interaction intermoléculaire entre les atomes non liés  $V_{inter}$  et le potentiel d'interaction intramoléculaire entre les atomes liés  $V_{intra}$ .

#### •Le potentiel intermoléculaire

Le potentiel  $V_{inter}$  comprend toutes les interactions entre les atomes non liés, d'une part les atomes de molécules différentes, et d'autre part les atomes d'une même molécule séparés par quatre liaisons et plus. Les interactions entre les atomes non liés sont de deux sortes : les interactions de van der Waals, qui sont le plus souvent modélisées par un potentiel de type Lennard-Jones, et les interactions électrostatiques dues aux charges ponctuelles des atomes, représentées par un potentiel de Coulomb. L'écriture du potentiel intermoléculaire est donc la suivante :

$$V_{inter} = V_{vdW} + V_{elec}, \quad (3.49)$$

avec :

$$V_{vdW} = \sum_{\text{atomes non liés}} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right), \quad (3.50)$$

$$V_{elec} = \sum_{\text{atomes non liés}} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}}, \quad (3.51)$$

où  $A_{ij}$  et  $B_{ij}$  sont les paramètres de Lennard-Jones pour le calcul des interactions de van der Waals entre les atomes  $i$  et  $j$  séparés de la distance  $r_{ij}$ , et  $q_i$  et  $q_j$  les charges partielles respectives des atomes  $i$  et  $j$ .

Pour les atomes séparés de moins de quatre liaisons, ces interactions de van der Waals et électrostatiques sont en fait prises en compte implicitement dans  $V_{intra}$ . Notons que souvent, au lieu des paramètres  $A_{ij}$  et  $B_{ij}$  pour le potentiel de Lennard-Jones, on donne la valeur absolue du potentiel au minimum  $\epsilon_{ij}$ , qui a donc la dimension d'une énergie, et

le diamètre des particules  $\sigma_{ij}$  qui est la valeur de  $r_{ij}$  pour laquelle le potentiel s'annule, ou  $2r_{ij}^{min}$ , qui est la valeur de  $r_{ij}$  pour laquelle le potentiel est minimal :

$$\begin{aligned} V_{vdW}(\sigma_{ij}) &= 0, \\ \frac{\partial V_{vdW}}{\partial r}(2r_{ij}^{min}) &= 0, \\ V_{vdW}(2r_{ij}^{min}) &= -\epsilon_{ij}. \end{aligned}$$

D'après l'expression du potentiel de Lennard-Jones (3.50), on relie simplement  $r_{ij}^{min}$  et  $\sigma_{ij}$  par la relation :  $2r_{ij}^{min} = 2^{1/6}\sigma_{ij}$ . Les deux paramètres sont donc équivalents. L'avantage de ces notations en  $\epsilon_{ij}$  et  $\sigma_{ij}$  est qu'elles ont une interprétation physique en terme d'énergie et de diamètre, et la donnée de ces deux paramètres permet de déterminer complètement le potentiel de la même façon que la donnée de  $A_{ij}$  et  $B_{ij}$ . D'une façon pratique, on donne la valeur de ces paramètres  $\epsilon_i$  et  $\sigma_i$  pour chaque atome  $i$ , et on détermine la valeur des paramètres du potentiel de Lennard-Jones entre deux atomes  $i$  et  $j$  par les formules  $\epsilon_{ij} = \sqrt{\epsilon_i\epsilon_j}$  et  $\sigma_{ij} = \frac{1}{2}(\sigma_i + \sigma_j)$ . Le potentiel de Lennard-Jones peut alors s'écrire en fonction de ces paramètres par la formule :

$$V_{vdW} = \sum_{i,j} 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$

### •Le potentiel intramoléculaire

Le potentiel intramoléculaire  $V_{intra}$  entre les atomes liés prend en compte toutes les interactions de liaisons : l'énergie d'élongation des liaisons covalentes, modélisée par un potentiel harmonique  $V_{liaison}$ , l'énergie de torsion des angles de valence  $V_{torsion}$ , représentée aussi par un potentiel harmonique, et l'énergie de torsion  $V_{dièdre}$  des angles dièdres, angles formés par quatre atomes successifs. Le potentiel intramoléculaire peut alors s'écrire sous la forme :

$$V_{intra} = V_{liaison} + V_{torsion} + V_{dièdre}, \quad (3.52)$$

avec :

$$V_{liaison} = \sum_{liaisons} K_d(d - d_0)^2, \quad (3.53)$$

$$V_{torsion} = \sum_{angles\ de\ valence} K_\theta(\theta - \theta_0)^2, \quad (3.54)$$

$$V_{dièdre} = \sum_{angles\ dièdres} K_\phi [1 + \cos(n\phi - \gamma)], \quad (3.55)$$

où  $K_d$  et  $K_\theta$  sont les constantes de force associées aux liaisons et aux angles de valence, et  $d_0$  et  $\theta_0$  leurs valeurs d'équilibre,  $2K_\phi$  est la valeur de la barrière de potentiel de l'angle dièdre  $\phi$ ,  $n$  la multiplicité de l'angle, et  $\gamma$  sa valeur à l'équilibre.

Ce potentiel ne dépend que de la position des noyaux des atomes et ne prend pas en compte les effets électroniques. Les constantes de liaison et valeurs à l'équilibre sont déterminées à partir de données expérimentales, comme des expériences de spectroscopie vibrationnelle. Quant aux charges partielles, elles sont souvent déterminées de façon à reproduire le champ et le potentiel électrostatique calculés par des méthodes *ab initio* sur des molécules simples. Notons qu'actuellement des champs de force prenant en compte la polarisabilité électronique des atomes de la molécule se développent pour simuler les molécules de façon plus réaliste [Cieplak *et al.*, 2001].

Dans le cas des molécules biologiques, les champs de force les plus couramment utilisés sont AMBER [Cornell *et al.*, 1995], CHARMM [MacKerell *et al.*, 1998] et GROMOS [van Gunsteren *et al.*, 1996]. Ces champs de force sont régulièrement actualisés, et plus ou moins bien adaptés aux simulations d'acides nucléiques ou de protéines.

La donnée du potentiel du système permet, à l'aide d'algorithmes de minimisation énergétique, d'atteindre des configurations énergétiquement favorables. Après la minimisation, on peut étudier la dynamique du système avec les algorithmes de dynamique moléculaire.

### L'algorithme de dynamique moléculaire

La technique de dynamique moléculaire permet de dépasser l'image statique de la molécule en lui donnant l'énergie cinétique nécessaire pour dépasser les barrières de potentiel entre les minima locaux de son hypersurface énergétique. Le principe de la dynamique moléculaire repose sur l'intégration des équations du mouvement classiques de Newton pour chaque atome, sachant que les forces qui s'exercent sur chaque atome dérivent à chaque instant du potentiel d'interaction  $V$  :

$$m_i \frac{\partial^2}{\partial t^2} \mathbf{r}_i(t) = \mathbf{F}_i = -\nabla_{\mathbf{r}_i} V, \quad (3.56)$$

où  $\mathbf{F}_i$  est la force exercée sur l'atome  $i$  de masse  $m_i$  et de position  $\mathbf{r}_i$  à l'instant  $t$ .

L'intégration de ces équations différentielles du mouvement est faite numériquement, à l'aide de différents algorithmes, dont le plus simple est celui d'Euler [Frenkel et Smit, 1996]. La vitesse et la position de l'atome  $i$  à l'instant  $t + \delta t$  s'expriment par rapport aux valeurs des vitesses, positions et forces à l'instant  $t$  par un développement de Taylor tronqué au second ordre :

$$\begin{cases} \mathbf{v}_i(t + \delta t) &= \mathbf{v}_i(t) + \frac{\partial \mathbf{v}_i(t)}{\partial t} \delta t = \mathbf{v}_i(t) + \frac{\mathbf{F}_i(t)}{m_i} \delta t \\ \mathbf{r}_i(t + \delta t) &= \mathbf{r}_i(t) + \mathbf{v}_i(t) \delta t \end{cases}$$

où  $\delta t$  est le pas de temps d'intégration. En pratique, l'algorithme d'Euler n'est jamais utilisé car il n'est ni réversible dans le temps ni conservatif (il ne conserve pas l'hamiltonien du système). On lui préférera l'algorithme de Verlet [Frenkel et Smit, 1996] qui exprime un développement de Taylor à l'ordre 3 de  $\mathbf{r}_i(t + \delta t)$  et de  $\mathbf{r}_i(t - \delta t)$  et qui permet d'obtenir une expression de la position valable à l'ordre 3 :

$$\mathbf{r}_i(t + \delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t - \delta t) + \frac{\mathbf{F}_i(t)}{m} \delta t^2$$

Cet algorithme est donc plus précis que celui d'Euler, et réversible dans le temps. Pour avoir une expression plus précise de la vitesse (qu'on obtient seulement au premier ordre avec l'algorithme de Verlet), on pourra utiliser d'autres algorithmes de même type que celui de Verlet.

Le choix de la valeur du pas d'intégration est un compromis entre la précision des calculs et le temps de calcul. Il faut impérativement que  $\delta t$  soit inférieur à la période des mouvements de plus haute fréquence du système pour éviter les erreurs d'intégration du mouvement. Pour une molécule, la plus petite période est celle de vibration des liaisons covalentes, qui est de l'ordre de  $10^{-14}$  s = 10 fs pour les liaisons impliquant des atomes d'hydrogène. On prendra donc en général  $\delta t \simeq 1$  fs. Notons qu'il est possible d'utiliser un pas d'intégration plus élevé si on contraint les mouvements de vibration des liaisons impliquant des atomes d'hydrogène, avec l'algorithme SHAKE [Rickaert *et al.*, 1977].

Les simulations de dynamique moléculaire sont le plus souvent effectuées à  $(N, V, E)$  constants (nombre de particules, volume et énergie totale du système) ou à  $(N, V, T)$  constants en utilisant un thermostat pour conserver la température. Le principe du thermostat de Nosé-Hoover [Nosé, 1984, Hoover, 1985] est de coupler le système à un réservoir externe de température, ce qui modifie en conséquence les vitesses des atomes par un facteur multiplicatif dépendant de la température. L'hamiltonien du système est alors modifié avec l'ajout d'une nouvelle variable dynamique liée au thermostat. Le thermostat d'Andersen couple le système par des collisions stochastiques avec un bain à la température voulue [Andersen, 1980]. Notons qu'afin de se rapprocher des conditions expérimentales, on peut aussi effectuer des simulations à  $(N, P, T)$  constants, dans l'ensemble isobare-isotherme, en couplant le système avec un barostat en plus du thermostat [Allen et Tidesley, 1987].

Par ailleurs, pour simuler des systèmes biologiques constitués de macromolécules entourées d'un très grand nombre de molécules d'eau, on ne peut considérer un système de taille finie qui impliquerait des effets de bords importants. C'est pourquoi des conditions aux bords périodiques [Allen et Tidesley, 1987] sont couramment utilisées pour diminuer les problèmes de discontinuité. Le principe de cette méthode est de placer la molécule dans une boîte de taille finie, le plus souvent parallélépipédique, entourée de tous les

côtés par un grand nombre d'images de la boîte initiale translattées. Il faut prendre une boîte de simulation assez grande pour que l'interaction de la molécule avec ses images soit négligeable, ou choisir une distance de coupure des interactions non liantes qui soit inférieure à la moitié de la plus petite arête de la boîte.

Nous utiliserons pour cette étude le programme de dynamique moléculaire ORAC développé au CEA par M. Marchi *et al.* [Procacci *et al.*, 1997]. L'un des avantages de ce programme est qu'il comprend toutes les options citées précédemment : thermostat, conditions aux bords périodiques,... De plus, ce programme peut être utilisé avec tous les champs de force, les fichiers de paramètres et de topologie étant modifiables. Nous pourrions donc tester indifféremment le champ de force de CHARMM ou celui d'AMBER. Enfin, ORAC possède la particularité d'utiliser des pas d'intégration multiples pour les différentes interactions courte et longue portée, et il permet d'utiliser l'algorithme de Particle Mesh Ewald pour les interactions électrostatiques longue portée. Nous allons présenter succinctement ces deux particularités.

### Le pas de temps multiple

Comme nous l'avons évoqué précédemment, le choix du pas d'intégration  $\delta t$  est dicté par la période la plus courte des mouvements caractéristiques du système. Cette contrainte limite donc le temps de simulation accessible. L'algorithme appelé r-RESPA pour « Reference System Propagation Algorithm » [Humphreys *et al.*, 1994] permet d'utiliser des pas d'intégration différents selon le temps caractéristique des mouvements considérés. Dans ORAC [Procacci et Marchi, 1998], il est possible d'utiliser jusqu'à cinq pas d'intégration différents, deux pour les mouvements dus aux interactions intramoléculaires, et trois pour les mouvements dus aux différentes composantes du potentiel  $V_{inter}$  à courte, moyenne et longue portée. Par exemple, on choisira le pas d'intégration le plus petit, entre 0,5 et 2 fs, pour les interactions covalentes et les interactions à courte portée (la partie répulsive du potentiel de Lennard-Jones), un pas d'intégration plus élevé pour les interactions à moyenne portée, puis le pas d'intégration le plus grand, entre 4 et 12 fs, pour les interactions longue portée (interactions coulombiennes et de Lennard-Jones à longue distance). Pendant la durée du pas d'intégration le plus court, on considère que les autres interactions ne modifient pas le mouvement. L'utilisation de ces différents pas d'intégration  $\delta t_1, \delta t_2, \dots$  repose donc sur le découpage du potentiel en différentes parties de temps caractéristiques distincts. Chaque pas d'intégration doit être un multiple du pas précédent dans l'ordre croissant. Les forces les plus « lentes » ne sont donc calculées qu'après que les forces les plus « rapides » ont été calculées un certain nombre de fois.

L'algorithme r-RESPA a la propriété d'être réversible et symplectique (il conserve la

mesure de l'espace des phases). D'un point de vue énergétique, il est possible d'observer en utilisant cet algorithme une dérive mineure de l'énergie totale du système [Procacci et Marchi, 1998]. Le gain de temps de calcul comparé à un pas de temps unique est d'un facteur 3 à 4 [Procacci et Marchi, 1998].

### Le traitement des interactions longue distance (PME)

Les interactions électrostatiques entre charges ponctuelles sont représentées par un potentiel coulombien en  $1/r_{ij}$ . Ces interactions sont donc à longue portée, c'est-à-dire qu'elles restent importantes même aux limites de la boîte de simulation. La méthode de coupure, ou « cutoff », consiste à négliger les interactions électrostatiques à partir d'une certaine distance de coupure  $R_c$ . Cependant, cette méthode néglige une partie importante des interactions électrostatiques, surtout dans le cas de molécules fortement chargées comme les acides nucléiques. Plusieurs simulations d'une boucle d'ARN de transfert en solvant explicite ont montré qu'avec un cutoff de 12 Å, la structure de la molécule se dénaturait très rapidement [Auffinger et Westhof, 1996, Capitaine, 2000].

La méthode des sommes d'Ewald [Frenkel et Smit, 1996] permet de calculer l'interaction électrostatique entre toutes les charges ponctuelles en séparant le potentiel en deux parties, une à courte portée, calculée dans l'espace direct, et l'autre à longue portée, calculée dans l'espace réciproque. Cela est possible après avoir introduit dans la première partie du potentiel des charges gaussiennes opposées aux charges ponctuelles, et dans la seconde des charges gaussiennes de même sens qui annulent les précédentes. Malheureusement, le calcul de ce potentiel est assez coûteux en temps de calcul, et varie en  $N^{3/2}$ , où  $N$  est le nombre de charges, donc plus coûteux qu'avec un cutoff. Pour des systèmes avec un grand nombre d'atomes, l'algorithme est donc souvent très prohibitif.

Ainsi, pour des systèmes importants avec des conditions aux bords périodiques, les méthodes de Particle Mesh Ewald [Darden *et al.*, 1997] ont été développées. Elles reposent sur la même décomposition du potentiel électrostatique que les sommes d'Ewald, mais permettent, par une interpolation tridimensionnelle des charges, de calculer efficacement le potentiel dans l'espace réciproque. Les charges gaussiennes considérées dans l'espace réciproque sont interpolées sur une grille par différentes fonctions selon la méthode utilisée. Par exemple, pour la méthode « Smooth Particle Mesh Ewald » ou SPME [Essmann *et al.*, 1995], des fonctions polynomiales « cardinal B-splines » sont utilisées. Puis, le potentiel électrostatique est calculé par des méthodes de transformées de Fourier rapides à trois dimensions. Les paramètres à choisir pour les méthodes PME sont la finesse de la grille et le facteur de convergence  $\alpha$  lié à la gaussienne des charges. L'algorithme de PME varie

en  $N \log(N)$ , ce qui le rend plus intéressant que les sommes d'Ewald à partir d'environ 900 atomes.

Il est intéressant, dans le cadre du modèle non local de pseudo-particules polarisables, de noter que des algorithmes des sommes d'Ewald et de PME existent non seulement pour des charges ponctuelles, mais aussi pour des multipôles, et en particulier des dipôles [Touk-maji et Sagui, 2000]. Nous pouvons même envisager de remplacer nos dipôles ponctuels par des dipôles gaussiens, pour ainsi n'avoir plus que la partie du potentiel électrostatique dans l'espace réciproque à calculer rapidement par des méthodes de FFT. Néanmoins, dans le développement de notre modèle, nous avons d'abord utilisé en première approximation la méthode plus simple de coupure des interactions longue distance.

### 3.2.2 L'intégration du modèle des pseudo-particules polarisables en dynamique moléculaire

#### Le nouvel hamiltonien et les forces associées

Pour introduire le modèle des pseudo-particules polarisables dans un programme de dynamique moléculaire, nous devons modifier l'hamiltonien présenté par l'équation (3.48), en distinguant le soluté, ensemble de charges ponctuelles dont les paramètres sont donnés par les champs de force classiques, et le solvant, particules polarisables sans charge ponctuelle, dont la masse et les paramètres de Lennard-Jones sont à préciser. Nous pouvons écrire simplement cet hamiltonien en considérant le solvant comme des particules de Lennard-Jones avec trois degrés de liberté de translation, et dont l'orientation des dipôles induits est celle à l'équilibre et dont le potentiel est donc donné par l'énergie libre de solvation à l'équilibre, à la place d'un potentiel électrostatique coulombien classique. Pour un système composé d'un soluté de  $M$  atomes de masses  $M_k$  situés à  $\mathbf{R}_k$  et de  $N$  particules de solvant de masse  $m_s$  situées à  $\mathbf{r}_i$ , l'hamiltonien s'écrit :

$$\begin{aligned}
 H = & \sum_{k=1}^M \frac{1}{2} M_k \dot{\mathbf{R}}_k^2 + V(\{\mathbf{R}_k\}) + \sum_{i=1}^N \frac{1}{2} m_s \dot{\mathbf{r}}_i^2 + \sum_{i,k} \Phi_k(|\mathbf{r}_i - \mathbf{R}_k|) \\
 & + \sum_{i < j} \Phi_s(|\mathbf{r}_i - \mathbf{r}_j|) + F(\{\mathbf{p}_i\}), \tag{3.57}
 \end{aligned}$$

où  $F(\{\mathbf{p}_i\})$  est l'énergie libre électrostatique de solvation  $F_{NLL}^{eq}$  donnée par (3.21) pour le modèle non local avec la saturation de Langevin, ou  $F_{LL}^{eq}$  donnée par (3.47) pour le modèle local avec la saturation de Langevin;  $V(\{\mathbf{R}_k\})$  est le potentiel du soluté comprenant la partie  $V_{intra}$  et  $V_{inter}$ , et  $\Phi_k$  et  $\Phi_s$  représentent respectivement le potentiel de

Lennard-Jones soluté/solvant et solvant/solvant.

Notons qu'en pratique nous prendrons naturellement comme masse des particules de solvant  $m_s$  la masse d'une molécule d'eau, c'est-à-dire  $m_s = 18,015$  g/mol.

Les forces dues à l'énergie de solvation exercées sur les atomes du soluté et sur les particules du solvant sont composées de deux parties :

- les forces dues au potentiel de Lennard-Jones :  $\Phi_k$  pour les atomes du soluté, et  $\Phi_s$  et  $\Phi_k$  pour les particules de solvant ;
- les forces électrostatiques dérivant de l'énergie libre électrostatique de solvation à l'équilibre : elles sont données par les équations (3.17) pour les atomes du soluté, et (3.16) pour les particules de solvant pour le modèle non local, et respectivement (3.44) et (3.43) pour le modèle local.

On comprend que nos particules de solvant sont des pseudo-particules, car si leurs coordonnées de position évoluent comme les atomes du soluté, le dipôle induit qu'elles portent est toujours à l'équilibre thermodynamique. Pour le modèle non local, il faudra à chaque pas calculer les moments dipolaires induits à l'équilibre par une méthode itérative, puis calculer les forces et l'énergie qui en résultent pour une configuration donnée du soluté et des particules de solvant, alors que pour le modèle local ces dipôles sont simplement proportionnels au champ électrique créé par le soluté dans le vide.

Il faut aussi souligner que par rapport aux méthodes de solvation implicites présentées dans le chapitre 2, l'avantage de notre méthode est qu'elle pourrait calculer la partie non polaire de l'énergie libre de solvation (dues aux interactions de van der Waals et de formation de la cavité) grâce au potentiel de Lennard-Jones soluté/solvant que nous pourrions moyenner sur les positions des particules de solvant, sans utiliser des méthodes de type SASA. Il faudrait alors optimiser les paramètres de Lennard-Jones de nos particules  $\epsilon_s$  et  $\sigma_s$  afin de reproduire les énergies libres de solvation de solutés hydrophobes (d'abord des sphères de différents rayons, puis des alcanes). Nous n'avons cependant pas développé cette approche, puisque nous nous sommes d'abord intéressés à reproduire la partie électrostatique de l'énergie libre de solvation.

## Le modèle de solvation et ORAC

D'un point de vue pratique, pour intégrer le modèle dans le programme ORAC, nous avons modifié le sous-programme qui calcule les forces en ajoutant les forces électrostatiques dues au solvant, et le sous-programme calculant l'hamiltonien en ajoutant à l'énergie totale l'énergie libre électrostatique de solvation. De plus, nous avons rajouté un sous-programme pour calculer les moments dipolaires induits des particules à l'équilibre à chaque pas d'intégration le plus faible. Les forces de Lennard-Jones soluté/solvant

et solvant/solvant sont calculées de la même façon que celles soluté/soluté, à condition d'avoir rajouté dans les fichiers de paramètres et de topologie un résidu (appelé DIP) composé d'un site neutre, et de paramètres  $m_s$ ,  $\epsilon_s$  et  $\sigma_s$  donnés.

Nous avons implémenté notre méthode avec une distance de coupure (par une fonction de coupure molle appelée fonction de « switch » pour éviter des problèmes de conservation d'énergie) pour les interactions non liantes. Nous sommes convaincus que la troncature des interactions non liées introduira seulement de faibles erreurs, d'une part parce que les interactions électrostatiques soluté/soluté sont calculées entre des groupements chimiques neutres (ou presque neutres), et d'autre part car le potentiel d'interaction charge/dipôle est en  $1/r^3$  et décroît donc plus vite que le potentiel d'interaction coulombien charge/charge habituel en  $1/r$ . Afin de corriger l'énergie libre électrostatique ainsi négligée, nous rajouterons à l'énergie de solvation électrostatique  $F$  tronquée à la distance  $R_c$  une correction du bulk (due au champ de réaction) calculée par la formule de Born pour une cavité de rayon  $R_c$  immergée dans un milieu diélectrique continu :  $\Delta G_{Bulk}(kcal/mol) = -166(Q^2/R_c)(1 - 1/\epsilon)$ , où  $Q$  est la charge totale du soluté. Pour un soluté neutre, nous rajouterons la correction du bulk calculée par la formule de Kirkwood pour un dipôle immergé dans une cavité mais en pratique cette contribution (de l'ordre de 1 kcal/mol) est négligeable par rapport à l'énergie de solvation tronquée.

### 3.2.3 Évaluation du temps de calcul

#### Le modèle non local

Pour tester la rapidité du modèle non local s'il était implémenté avec l'algorithme PME, nous avons utilisé un programme calculant l'énergie électrostatique de charges et dipôles ponctuels avec un algorithme de PME pour des charges et dipôles par des méthodes de FFT. Nous avons comparé le temps de calcul d'un pas de dynamique pour un système de  $3N$  charges ponctuelles sans dipôle (représentant un modèle d'eau de type TIP3), et pour un système de  $N$  dipôles sans charges ponctuelles (représentant notre modèle non local sur particules). Selon les paramètres de PME choisis, il est possible de gagner un facteur 8 avec des dipôles par rapport aux charges pour le calcul des interactions électrostatiques solvant/solvant.

#### Le modèle local

Le modèle local, avec une méthode de coupure pour les interactions longue distance, devrait être très intéressant d'un point de vue du temps de calcul par rapport aux mé-

thodes explicites. En effet, nos pseudo-particules de solvant n'ont qu'un seul site au lieu de 3, ce qui devrait faire gagner un facteur 9 pour le calcul des interactions de Lennard-Jones solvant/solvant. De plus, il n'y a pas d'interactions électrostatiques à calculer pour la partie solvant/solvant. Pour les interactions soluté/solvant, il devrait être possible de gagner un facteur 3.

L'implémentation du modèle de solvant local des pseudo-particules polarisables (PPP) dans le programme de dynamique moléculaire ORAC a été en partie optimisée, afin de comparer le temps de calcul obtenu pour une simulation avec notre modèle et celui obtenu avec un modèle de solvant explicite à trois sites TIP3. Ces tests ont été effectués avec ORAC et des pas d'intégration multiples : 4 fs pour les interactions intermoléculaires à longue portée coupées à 12 Å, 2 fs pour les interactions à moyenne portée coupées à 7,2 Å, et 1 fs pour les interactions à courte portée coupées à 4,5 Å et pour les interactions intramoléculaires. La liste des voisins est mise à jour toutes les 60 fs. Nous avons calculé le temps CPU, sur un PC avec un processeur à 2,6 GHz et 1 Go de mémoire de RAM, pour une dynamique d'une durée d'1 ps. Pour le modèle PPP, les interactions solvant/solvant, uniquement de type Lennard-Jones, sont coupées à 7,2 Å.

Les résultats sont présentés sur le tableau 3.1 pour un système composé du solvant pur, et pour des molécules de différentes tailles parmi celles étudiées dans la thèse, entourées de solvant : le peptide Rn24 (étudié au chapitre 4), l'ARNt et la double hélice d'ADN-B2 (étudiés au chapitre 5), et la protéine 1PGB (étudiée au chapitre 4), avec pour chaque système le même nombre de particules de solvant PPP ou TIP3. Pour les molécules étudiées, le temps de calcul est ramené à un pour une dynamique du soluté dans le vide.

<i>Molécule étudiée</i>	<i>Nombre de particules</i>	<i>Vide</i>	<i>PPP</i>	<i>TIP3</i>	<i>Rapport TIP3/PPP</i>
Solvant pur	0/686		1,0	12,5	<b>12,5</b>
Rn24	195/1930	1,0	12,6	76,7	<b>6,1</b>
ARNt	562/5248	1,0	15,9	85,9	<b>5,4</b>
ADN	782/5169	1,0	11,6	57,8	<b>5,0</b>
1PGB	855/5202	1,0	7,7	35,1	<b>4,5</b>

TAB. 3.1 : Temps de calcul en fonction du nombre de particules du système soluté/solvant, pour différentes molécules dans le vide, avec le solvant PPP et avec un solvant TIP3 (le temps de calcul est ramené à un pour des dynamiques des solutés dans le vide, sauf pour la simulation du solvant pur où il est ramené à un pour une dynamique avec PPP).

Le gain de temps de calcul du modèle PPP par rapport au modèle explicite TIP3 diminue quand le nombre d'atomes du soluté augmente, car les interactions soluté/soluté

coûtent aussi cher quel que soit le modèle de solvant utilisé. Ce rapport PPP/TIP3 est en moyenne autour de 5 pour des molécules de taille importante, ce qui rend notre modèle de solvant très efficace par rapport aux modèles explicites. De plus, le rapport du temps de calcul du modèle PPP par rapport au vide décroît avec l'augmentation du nombre d'atomes du soluté. Cela veut dire que plus le soluté est important, plus la rapidité de notre modèle de solvant devient comparable à celle d'une simulation dans le vide. Le rapport PPP/vide atteint 7,7 pour des protéines globulaires de presque un millier d'atomes comme 1PGB, ce qui rend alors notre modèle de solvant compétitif par rapport aux méthodes de solvation implicites de type Born Généralisé.

Notons que nous pourrions encore améliorer ces résultats en jouant sur la coupure des interactions solvant/solvant. En effet, la particularité du modèle local est qu'il n'y a pas d'interactions électrostatiques solvant/solvant à calculer, et les interactions entre les particules de solvant sont uniquement de type Lennard-Jones. En modifiant ce potentiel de Lennard-Jones solvant/solvant pour le rendre à encore plus courte portée, nous pourrions ainsi améliorer le gain de temps de calcul par rapport à un modèle de solvant explicite. De plus, nous pourrions choisir un pas d'intégration plus élevé pour calculer les interactions solvant/solvant, typiquement jusqu'à 10 fs pour un fluide de Lennard-Jones pur. Cela nécessiterait de modifier de façon plus complexe le programme ORAC, mais permettrait d'améliorer fortement la rapidité du modèle.

### 3.3 Le paramétrage du modèle local

Pour aboutir à la formulation locale du modèle de pseudo-particules polarisables, on a effectué un certain nombre d'approximations. Avant de l'appliquer à des systèmes complexes, nous allons donc devoir paramétrer notre modèle en comparant les résultats obtenus avec notre modèle à la théorie électrostatique exacte ou à des modèles explicites. Cette comparaison portera sur quelques systèmes simples : un ion, deux ions, puis de petites molécules et des résidus d'acides nucléiques et d'acides aminés.

#### 3.3.1 Étude d'un ion

##### Le dipôle moyen autour d'un ion

Pour cette étude, le système comporte  $N = 500$  particules de paramètres de Lennard-Jones  $\sigma_s = 3,024 \text{ \AA}$  et  $\epsilon_s = 3,197 \text{ kJ/mol}$ , à la température  $T = 300 \text{ K}$ . La polarisabilité des particules est  $\alpha = 2,71 \text{ \AA}^3$ , correspondant à  $\epsilon = 80$  d'après la relation (3.40).

De la même façon que pour le modèle non local, nous avons étudié la polarisation moyenne autour d'un ion de charge  $q = 1e$ , représentée par la valeur du dipôle moyen  $\langle \mu(R) \rangle$ , comparé à la valeur théorique asymptotique  $\frac{\epsilon_0(\epsilon-1)}{\rho\epsilon} E_0(R)$ . Pour le modèle local sur particules sans la saturation de Langevin, on a, d'après l'expression des moments dipolaires induits donnée par (3.41),  $\mathbf{p}_i = \alpha_e \mathbf{E}_i^0$  avec  $\alpha_e = \frac{\epsilon_0(\epsilon-1)}{\rho\epsilon}$  d'après l'équation (3.40). On voit d'après ces équations que l'on retrouve exactement l'équation théorique asymptotique. En effet, l'approximation locale est exacte pour un soluté de symétrie sphérique [Kharkats *et al.*, 1976]. Avec l'introduction de la saturation, le dipôle moyen s'éloigne de l'électrostatique pour de faibles valeurs de  $R$ , mais s'en approche dès qu'on est en régime linéaire, c'est-à-dire dès que le champ électrique devient plus faible, donc quand  $R$  augmente, comme on le voit sur la figure 3.9 pour deux valeurs différentes du dipôle de saturation.

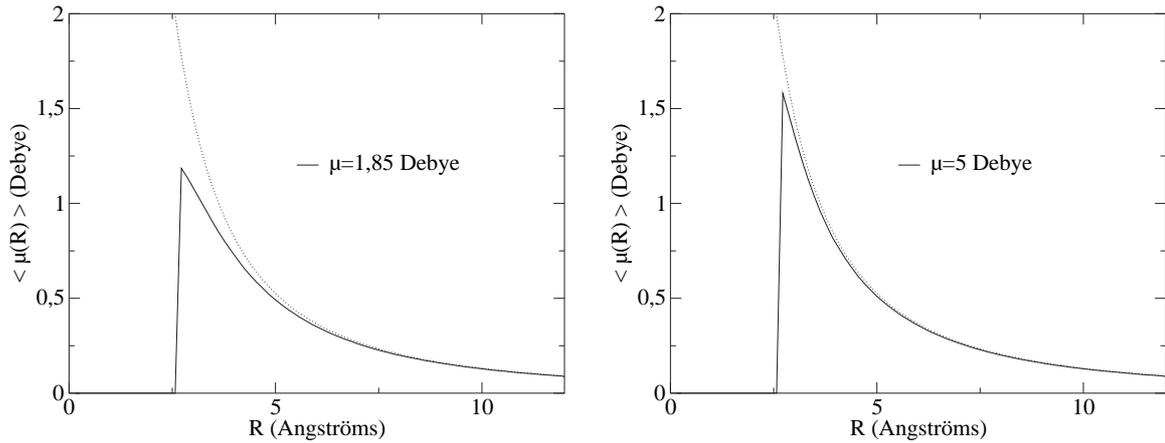


FIG. 3.9 : Dipôle moyen autour d'un ion en Debye avec le modèle local, pour une valeur du dipôle de saturation  $\mu = 1,85$  Debye (à gauche) et  $\mu = 5$  Debye (à droite). La valeur théorique asymptotique est représentée en pointillés.

On voit que pour une valeur élevée de la saturation, le moment dipolaire induit avec la fonction de Langevin donné par l'équation (3.46) est presque linéaire, puisque la courbe obtenue avec le modèle local est presque entièrement confondue avec la courbe asymptotique. Plus la saturation est faible, plus le dipôle moyen s'éloigne de l'électrostatique à de faibles distances, mais dans tous les cas il tend asymptotiquement à grande distance vers la valeur théorique. Pour le modèle local, on n'observe pas d'oscillations comme pour le modèle non local, car les interactions dipôles/dipôles ne sont pas calculées explicitement.

### L'énergie de solvation d'un ion

Nous avons testé la validité du modèle non local pour le calcul de l'énergie libre de solvation électrostatique d'un ion de charge  $q(e)$  et de rayon  $R(\text{Å})$ . Nous avons comparé l'énergie  $F_{LL}^{eq}$  obtenue avec les formules du modèle local à l'énergie libre de solvation de Born, valable pour un ion, en kJ/mol :  $\Delta G_{Born} = -685 \frac{q^2}{R} (1 - \frac{1}{\epsilon})$ . Nous avons observé que, sans aucun paramétrage, notre modèle ne reproduit pas bien l'énergie de solvation de Born, surtout pour des ions de faibles rayons, et nous allons donc tenter de corriger cette erreur.

Selon la charge et le rayon de l'ion considéré, nous attribuons à l'ion un rayon effectif  $R_{eff} = R \times \Delta G_{Born} / F_{LL}^{eq}$  pour corriger l'erreur du modèle local. Nous obtenons ainsi un facteur de rayon (option « slt-radius » de notre programme) qui modifie le rayon de l'atome par une fonction quadratique dépendant du rayon initial et de la charge de l'ion. Cette fonction quadratique est obtenue par régression quadratique sur le calcul de  $R_{eff}$  pour différentes charges et différents rayons. La fonction quadratique doit être calculée pour chaque valeur du dipôle de saturation.

On a tracé sur la figure 3.10 l'énergie de Born et l'énergie libre électrostatique de solvation calculée avec et sans le facteur de rayon, pour une valeur du dipôle de saturation de  $\mu = 1,5$  Debye, et pour des charges ioniques entre 0,25 et 1,5  $e$ , avec comme paramètres de Lennard-Jones pour les particules de solvant  $\sigma_s = 2,88 \text{ Å}$  et  $\epsilon_s = 3,197$  kJ/mol, à la température  $T = 300$  K. Grâce au facteur de rayon, nous arrivons à réduire l'erreur relative, qui était en moyenne de 10 à 20 % sans le facteur de rayon, à une erreur de 5 à 10 %. On remarque que les erreurs relatives sont plus importantes pour les atomes de faibles rayons (de moins de 1 Å). On peut l'expliquer par la saturation du dipôle, car, comme nous l'avons vu pour le calcul du dipôle moyen autour d'un ion, la saturation a essentiellement une influence à faible distance de l'ion.

L'intérêt du facteur de rayon est qu'il est paramétré sous la forme d'une fonction quadratique qui renormalise de façon automatique tous les rayons des différents types d'atomes selon leurs charges et leurs rayons initiaux. Nous n'aurons donc pas besoin de calculer le rayon de chaque type d'atome séparément et nous pourrions utiliser comme paramètres de notre programme l'ensemble des rayons des champs de force classique.

### La fonction de distribution radiale autour d'un ion

Comme notre modèle est sur particules, il est intéressant d'étudier la fonction de distribution radiale de nos particules autour d'un ion par rapport à un modèle d'eau explicite.

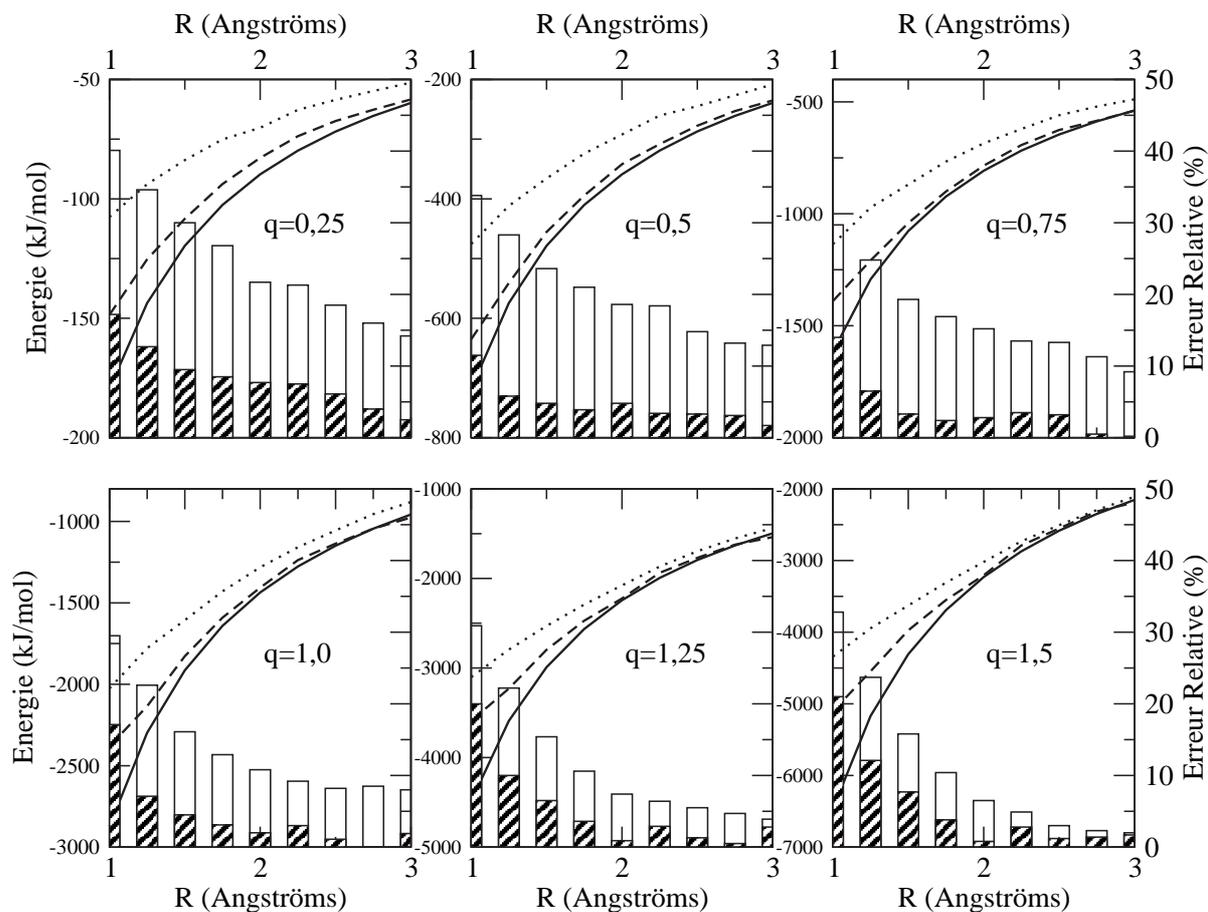


FIG. 3.10 : Énergie de solvation d'un ion en fonction de son rayon  $R$  pour différentes charges  $q$  de l'ion. L'énergie de Born de référence est représentée en trait plein, l'énergie de solvation avec le modèle local sans renormalisation des rayons en pointillés, et avec la renormalisation des rayons en traits pointillés. Les erreurs relatives par rapport à l'énergie de Born sont représentées avec le facteur de rayon en barres hachurées, et sans en barres blanches.

En effet, comme avec un modèle explicite, nous voulons pouvoir étudier la structure des couches d'hydratation autour des molécules avec notre modèle. La fonction de distribution radiale autour d'un ion, notée  $g(r)$ , représente la probabilité de trouver une particule de solvant à la distance  $r$  de l'ion. Nous avons étudié l'influence du dipôle de saturation sur  $g(r)$ , ainsi que l'influence du facteur de rayon introduit précédemment, et l'influence du diamètre de Lennard-Jones  $\sigma_s$  des particules de solvant. Nous avons comparé le  $g(r)$  moyenné sur environ 500 ps pour le modèle local au  $g(r)$  calculé avec l'eau explicite TIP3 (en considérant le centre de masse de chaque molécule d'eau) sur une dynamique de 900 ps.

Les simulations ont été effectuées avec ORAC, et le calcul de  $g(r)$  avec le module d'analyse d'ORAC modifié pour notre étude. Les paramètres de Lennard-Jones des par-

ticules sont  $\sigma_s = 3,024 \text{ \AA}$ , ou  $\sigma_s = 2,88 \text{ \AA}$ , et  $\epsilon_s = 3,197 \text{ kJ/mol}$ . Pour les simulations avec  $\sigma_s = 3,024 \text{ \AA}$ , la densité est  $\rho = 0,029 \text{ \AA}^{-3}$ , la boîte carrée de côté  $a = 28,7 \text{ \AA}$  et la polarisabilité des particules de  $\alpha = 2,71 \text{ \AA}^3$ . Pour celles avec  $\sigma_s = 2,88 \text{ \AA}$ , la densité est  $\rho = 0,033 \text{ \AA}^{-3}$ , la boîte de côté  $a = 27,3 \text{ \AA}$  et la polarisabilité  $\alpha = 2,33 \text{ \AA}^3$ . Les dynamiques sont effectuées avec un thermostat de Nosé-Hoover à  $T = 300 \text{ K}$ . Nous avons utilisé un cutoff pour les interactions à longue portée avec une fonction de switch entre 12 et 13  $\text{\AA}$ . Nous avons utilisé l'algorithme de pas d'intégration multiple, avec comme plus grand pas 6 fs pour les interactions à longue portée (entre 9 et 12  $\text{\AA}$ ), puis 3 fs pour les interactions à moyenne portée (entre 6 et 9  $\text{\AA}$ ), et 1 fs pour toutes les autres interactions à courte portée (à moins de 6  $\text{\AA}$ ) et internes. L'ion est un ion Sodium chargé positivement,  $q = +1 e$ , et de paramètres de Lennard-Jones  $\epsilon_{sod} = 0,4189 \text{ kJ/mol}$  et  $\sigma_{sod} = 2,863 \text{ \AA}$ .

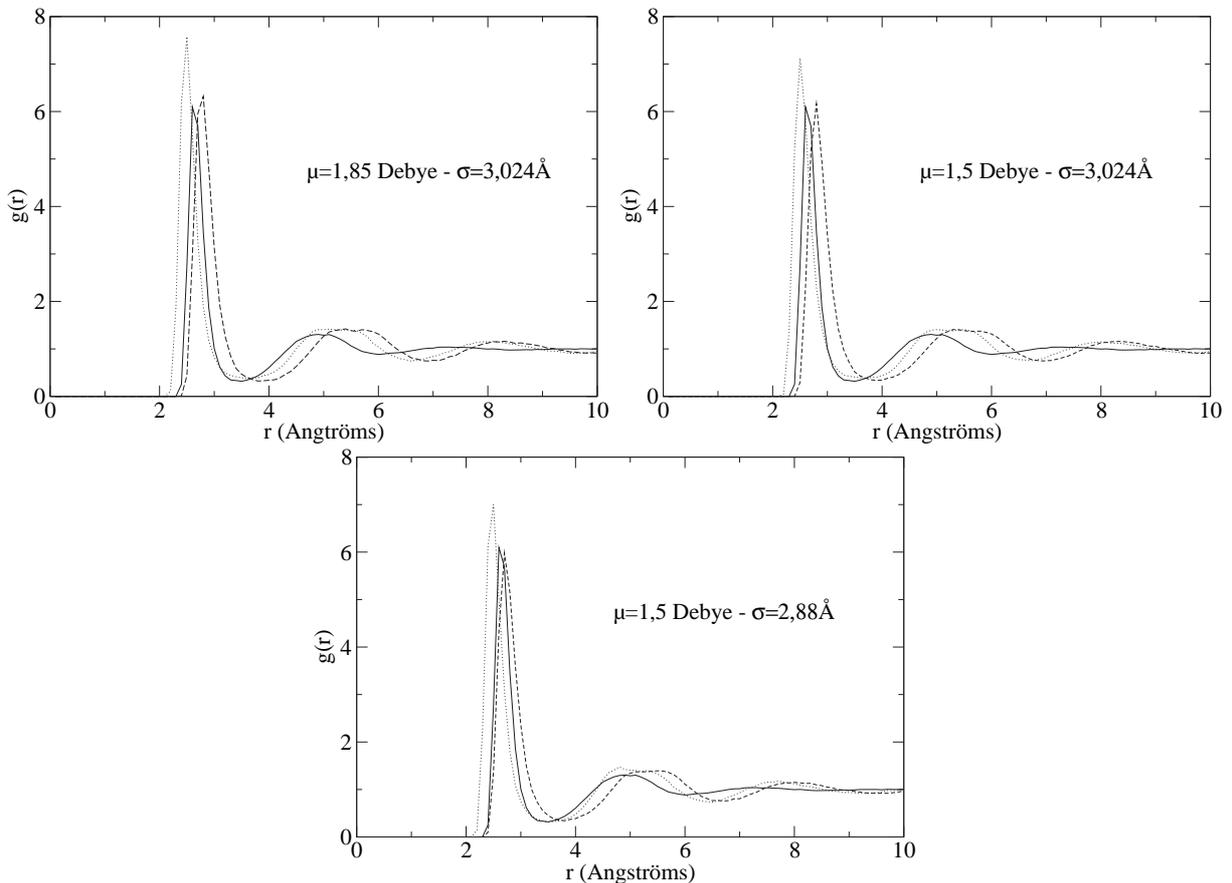


FIG. 3.11 : Fonction de distribution radiale autour d'un ion, avec le modèle non local, avec  $\sigma_s = 3,024 \text{ \AA}$  et  $\sigma_s = 2,88 \text{ \AA}$  pour une valeur du dipôle de saturation  $\mu = 1,85 \text{ Debye}$  et  $\mu = 1,5 \text{ Debye}$ . En traits pointillés sans la renormalisation des rayons, et en pointillés avec le facteur de rayon. La valeur de  $g(r)$  avec TIP3 est représentée en trait plein.

La figure 3.11 représente les fonctions de distribution radiale (FDR) du modèle local

avec des particules de solvant de diamètre  $\sigma_s = 3,024 \text{ \AA}$ , pour deux valeurs du dipôle de saturation (1,85 et 1,5 Debye), avec et sans le facteur de rayon, ainsi que la FDR avec des particules de solvant de diamètre  $\sigma_s = 2,88 \text{ \AA}$  et  $\mu = 1,5$  Debye avec et sans le facteur de rayon, par rapport à la fonction de distribution radiale de l'eau explicite TIP3. On voit que sans le facteur de rayon, le premier pic de la FDR de l'eau est relativement bien reproduit, mais le premier creux et le deuxième pic sont fortement décalés. L'utilisation du facteur de rayon change peu le premier pic, mais rapproche le premier creux et le deuxième pic. On voit aussi qu'avec le facteur de rayon, la hauteur du premier pic augmente, et que pour  $\mu = 1,85$  Debye cette valeur devient un peu plus élevée. Même avec le facteur de rayon, pour  $\sigma_s = 3,024 \text{ \AA}$ , le deuxième pic de la FDR est plus large que celui de l'eau. Pour les simulations avec  $\sigma_s = 2,88 \text{ \AA}$  et avec le facteur de rayon, la courbe de FDR obtenue est très proche de celle de l'eau explicite. Le premier pic est légèrement plus proche de l'ion et un peu plus élevé, mais le premier creux correspond tout à fait à celui de l'eau, et le deuxième pic se rapproche fortement de celui de l'eau par rapport aux simulations avec  $\sigma_s = 3,024 \text{ \AA}$ . Le troisième creux est un peu plus loin de l'ion, et les courbes ne correspondent plus vraiment au-delà, indiquant que les deuxièmes et troisièmes couches d'hydratation seraient moins bien reproduites. Mais le but du modèle de pseudo-particules polarisables est de reproduire les premières couches d'hydratation des molécules tout en apportant des informations sur les grandeurs thermodynamiques moyennes. Nous considérons donc que reproduire la FDR de l'eau jusqu'au second pic nous permettra de reproduire correctement les premières couches de la structure d'hydratation des molécules.

### 3.3.2 Le potentiel de force moyenne entre deux ions

De la même façon que pour le modèle non local, nous avons étudié le potentiel de force moyenne (PMF)  $W(R)$  entre une paire d'ions de charges opposées,  $\text{Na}^+$  et  $\text{Cl}^-$  du chlorure de sodium, distants de  $R$ .

Le PMF et les différentes composantes de la force entre les deux ions sont représentés sur la figure 3.12, pour une valeur de la saturation de  $\mu = 1,85$  Debye, avec et sans le facteur de rayon, pour un diamètre des particules de  $\sigma_s = 3,024 \text{ \AA}$ . On distingue sur cette figure les deux composantes de la force totale qui se compensent à longue distance : la force des deux ions dans le vide et la force due au solvant, qui a un effet d'écrantage des interactions électrostatiques dans le vide.

Le PMF pour trois valeurs différentes de la saturation est représenté sur la figure 3.13, avec et sans le facteur de rayon, avec comme diamètre des particules  $\sigma_s = 3,024 \text{ \AA}$ , et avec  $\sigma_s = 2,88 \text{ \AA}$  pour une valeur du dipôle de saturation de  $\mu = 1,5$  D.

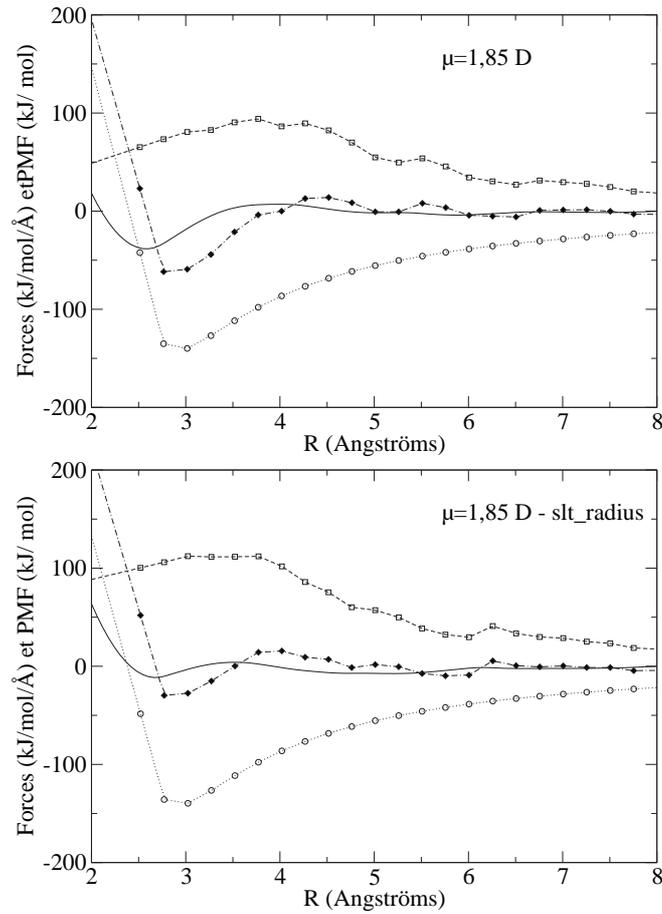


FIG. 3.12 : Forces et PMF entre  $\text{Na}^+$  et  $\text{Cl}^-$ , avec le modèle local, pour une valeur du dipôle de saturation de  $\mu = 1,85$  Debye, avec et sans le facteur de rayon. la force entre les deux ions dans le vide (en  $\text{kJ/mol}/\text{\AA}$ ) est représentée en pointillés (et cercles), la force due au solvant en traits pointillés (et carrés), et la force totale en traits mixtes (et losanges pleins).

On voit que sans le facteur de rayon, pour une valeur du dipôle de saturation élevé ( $\mu = 2,35$  D) le PMF présente un minimum de potentiel de CIP (Contact Ion Pair) assez élevé, et une barrière de potentiel beaucoup trop importante entre CIP et SSIP (plus de  $30$   $\text{kJ/mol}$ ). Une valeur plus faible du dipôle de saturation pour le modèle local améliore considérablement le PMF, elle abaisse le minimum CIP, même si la barrière de potentiel reste relativement trop élevée entre les deux minima par rapport aux méthodes explicites [Guardia *et al.*, 1991, Berkowitz *et al.*, 1984, Dang *et al.*, 1990]. On comprend donc que, comme nous l'avions annoncé précédemment, l'introduction de la saturation pour le modèle local améliore le PMF entre deux ions de charges opposées. L'introduction du facteur de rayon diminue fortement la valeur de la barrière de potentiel entre les deux minima. Pour une valeur du dipôle de saturation de  $\mu = 2,35$  D, on observe mal le minimum SSIP,

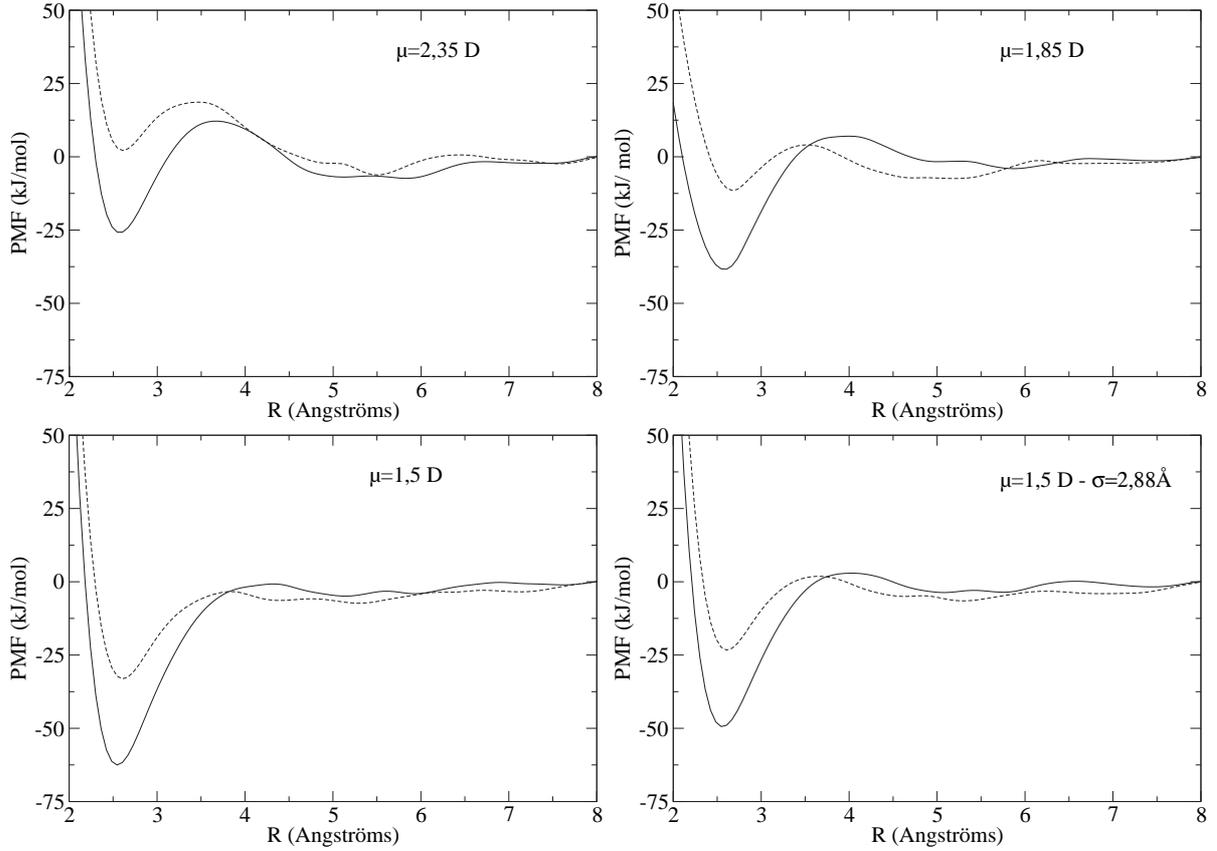


FIG. 3.13 : PMF entre  $Na^+$  et  $Cl^-$ , avec le modèle local, pour des valeurs de différentes valeurs du dipôle de saturation, avec  $\sigma_s = 3,024 \text{ \AA}$  sauf pour la dernière courbe où  $\sigma_s = 2,88 \text{ \AA}$ , en trait plein sans le facteur de rayon, et en trait pointillé avec le facteur de rayon.

alors que pour  $\mu = 1,85 \text{ D}$  et  $\mu = 1,5 \text{ D}$ , on observe bien les deux minima à des distances comparables à celles des précédentes simulations en solvant explicite (CIP pour environ  $R \simeq 2,6 \text{ \AA}$ , et SSIP pour  $R \simeq 4,5\text{--}5 \text{ \AA}$ ). Cependant, pour  $\mu = 1,5 \text{ D}$ , la barrière de potentiel paraît un peu trop élevée, et le deuxième minimum est moins prononcé que pour les valeurs plus élevées de  $\mu$ . La diminution du diamètre des particules à  $\sigma_s = 2,88 \text{ \AA}$ , induite par l'étude de la fonction de distribution radiale, avec une valeur du dipôle de saturation de  $\mu = 1,5 \text{ D}$ , améliore encore les caractéristiques du PMF. Le premier minimum est moins profond, et avec le facteur de rayon, la barrière de potentiel entre CIP et SSIP est moins élevée.

Les deux PMF reproduisant au mieux les barrières de potentiel entre CIP et SSIP des résultats en solvant explicite sont toujours avec le facteur de rayon, et avec  $\sigma_s = 3,024 \text{ \AA}$  et  $\mu = 1,85 \text{ D}$ , ou avec  $\sigma_s = 2,88 \text{ \AA}$  et  $\mu = 1,5 \text{ D}$ . Cependant, il n'existe pas de consensus sur les caractéristiques exactes que doit avoir le PMF de la paire d'ions Na-Cl,

et il est donc difficile de choisir entre les deux jeux de paramètres avec ce seul critère. L'étude de la fonction de distribution radiale nous a montré que la valeur du dipôle de saturation  $\mu = 1,5$  D associée à un diamètre des particules de  $\sigma_s = 2,88$  Å donnait des résultats plus proches du solvant explicite TIP3. De plus, des simulations de dynamique moléculaire sur des macromolécules biologiques avec un dipôle de saturation de  $\mu = 1,85$  D ont été effectuées (résultats non présentés ici) et n'ont jamais donné de résultats satisfaisants. Nous utiliserons donc par la suite  $\mu = 1,5$  D et  $\sigma_s = 2,88$  Å, toujours avec le facteur de rayon puisqu'il améliore à la fois l'énergie de solvation d'un ion, la fonction de distribution radiale autour d'un ion et le potentiel de force moyenne entre deux ions. Ce sont ces derniers paramètres que nous utiliserons aussi dans les chapitres 4 et 5 pour l'étude des macromolécules biologiques.

### 3.3.3 L'énergie de solvation de petites molécules

L'intérêt du modèle sur particules est non seulement de reproduire par son aspect particulaire les propriétés d'un solvant explicite, mais aussi de reproduire l'électrostatique, en particulier les calculs d'énergies libres de solvation. Comme nous l'avons expliqué plus haut, il faut moyenner sur un certain nombre de configurations de particules de solvant pour obtenir l'énergie libre de solvation d'un soluté, comme si on moyennait sur plusieurs grilles assez grossières. Pour pouvoir strictement comparer à une méthode électrostatique non dynamique (de type résolution de l'équation de Poisson-Boltzmann), il faut donc effectuer une dynamique moléculaire avec le soluté fixé, et moyenner l'énergie libre électrostatique de solvation sur une durée de simulation significative.

Nous avons choisi une dizaine de petites molécules organiques chargées (acétate, méthoxyde,...), une dizaine de neutres (méthanol, acide acétique,...), ainsi que des résidus isolés d'acides nucléiques et d'acides aminés (bases, sucres, et squelette phosphodiester pour les acides nucléiques, et tous les acides aminés des protéines). La liste exacte de toutes les molécules étudiées est présentée en Annexe A.3 de ce chapitre.

Avec le modèle local des pseudo-particules polarisables (que nous noterons PPP), nous avons utilisé les paramètres suivants :

- pour les conditions de simulation : nous avons utilisé les mêmes pas d'intégration que pour le calcul de  $g(r)$ , un thermostat de Nosé à  $T = 300$  K, une boîte de simulation cubique de côté  $27,3$  Å, pour un temps total de simulation de 60 ps dont 42 ps d'équilibration (où nous avons augmenté progressivement la température du système de 20 à 300 K), puis 18 ps sur lesquelles nous avons calculé la moyenne de l'énergie libre de solvation (on voit sur la figure 3.14 que cette énergie de solvation reste constante en moyenne sur ces 18 ps) ;

- pour les 686 particules de solvant PPP mobiles : nous avons utilisé les paramètres de Lennard-Jones  $\epsilon_s = 3,197$  kJ/mol,  $\sigma_s = 2,88$  Å, le dipôle de saturation  $\mu = 1,5$  Debye, le facteur de rayon, et une polarisabilité des particules de  $\alpha = 2,33$  Å<sup>3</sup> correspondant à  $\epsilon = 80$  ;
- pour le soluté immobile : pour les petites molécules organiques, nous avons utilisé les paramètres et topologies du champ de force CHARMM « all 22 » [MacKerell *et al.*, 1998], et pour les résidus des acides nucléiques et protéines les paramètres et topologies du champ de force de seconde génération AMBER94 [Cornell *et al.*, 1995].

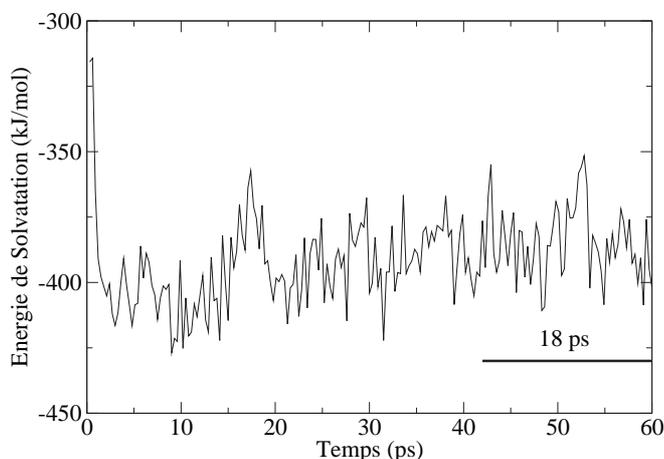


FIG. 3.14 : Évolution de l'énergie libre de solvation électrostatique pendant la simulation avec le solvant PPP (ici pour le méthoxyde) : nous avons calculé la moyenne sur les dernières 18 ps de simulation.

Nous avons ainsi comparé la valeur moyenne obtenue pour  $F_{LL}^{eq}$  sur les 18 dernières picosecondes de la simulation avec la valeur obtenue avec les mêmes paramètres de charges et de rayons par le programme DelPhi [Nicholls et Honig, 1991], dont on sait qu'il reproduit correctement l'électrostatique. Pour les paramètres de DelPhi, nous avons utilisé une constante diélectrique de  $\epsilon_i = 1$  à l'intérieur du soluté, et  $\epsilon = 80$  à l'extérieur. L'espacement de la grille est fixée à  $10$  Å<sup>-1</sup>, et le soluté occupe 75 % de la boîte. La taille de la grille est alors calculée automatiquement en fonction de la vitesse du processeur disponible. La frontière entre le soluté et le solvant est définie par la surface du soluté calculée par la méthode des sphères de Connolly. Cette méthode consiste à faire passer une sphère de rayon 1,4 Å, représentant le solvant, à la surface des volumes de van der Waals des atomes du soluté, et à calculer les portions convexes et concaves où les sphères de solvant peuvent s'introduire.

### Introduction du facteur de charge

Sans aucun autre paramétrage, nous observons une mauvaise corrélation entre les énergies libres de solvation obtenues avec le modèle PPP et celles calculées avec DelPhi, surtout pour les molécules neutres. Or, il a été montré [Lévy, 2002] qu'en utilisant la fonctionnelle macroscopique de notre modèle, l'approximation locale introduit une erreur systématique pour l'énergie libre de solvation d'un dipôle ponctuel  $\mu$  dans une cavité de rayon  $R$  (avec des conditions de sphère dure) immergée dans un milieu diélectrique continu de constante diélectrique  $\varepsilon$ . En effet, on obtient avec la formulation locale une énergie libre électrostatique de solvation pour le dipôle de :

$$\Delta F_{Local} = -\frac{1}{2} \left(\frac{2}{3}\right) \frac{(\varepsilon - 1) \mu^2}{\varepsilon R^3}, \quad (3.58)$$

alors que la formule exacte de l'énergie libre de solvation d'un dipôle dans une telle cavité a été donnée par Kirkwood [Kirkwood, 1934] :

$$\Delta F_{Kirkwood} = -\frac{1}{2} \frac{(\varepsilon - 1) \mu^2}{\varepsilon + \frac{1}{2}} \frac{1}{R^3}. \quad (3.59)$$

On voit que pour  $\varepsilon \rightarrow \infty$ , c'est-à-dire pour des valeurs de la constante diélectrique élevée comme pour l'eau, il y a une erreur systématique de 2/3 sur l'énergie libre électrostatique de solvation du dipôle avec le modèle local. Afin de corriger cette erreur, il est possible de modifier le dipôle dans le modèle local en utilisant un dipôle effectif  $\mu_{eff} = \sqrt{3/2}\mu$  pour le calcul de l'énergie libre de solvation.

Cela revient, pour un résidu de  $N$  charges ponctuelles  $q_i$  (qui possédera donc un moment dipolaire permanent  $\mu$ ), à renormaliser les charges par la formule :

$$q_i^{eff} = \lambda_q q_i + (1 - \lambda_q) \frac{Q}{N}, \quad (3.60)$$

où  $Q = \sum_{i=1}^N q_i$  est la charge totale du résidu. On voit donc que pour un résidu neutre ( $Q = 0$ ), le facteur de charge est simplement un facteur multiplicatif identique pour toutes charges. Pour le cas d'un ion ( $N = 1$ ), la charge de l'ion n'est pas modifiée, ce qui paraît normal puisqu'il n'y a pas d'erreur systématique sur l'énergie libre de solvation d'un ion. En pratique, nous renormalisons les charges pour chaque groupe d'atomes, neutre ou presque, défini dans ORAC.

Afin d'obtenir le facteur  $\sqrt{3/2}$  pour le moment dipolaire, il faut prendre un facteur de charge  $\lambda_q = \sqrt{3/2} \simeq 1,225$ . Cependant, notons que cette valeur du facteur de charge correspond à des conditions de sphères dures, alors que dans notre modèle nous utilisons

un potentiel de Lennard-Jones « mou ». La valeur du facteur de charge doit donc être adapté à notre étude autour de la valeur de  $\sqrt{3/2}$ .

D'un point de vue purement théorique, ce facteur  $\lambda_q$  renormalisant les charges des atomes du soluté ne devrait être pris en compte que pour le calcul de l'énergie libre de solvation du soluté, donc uniquement pour le calcul des interactions soluté/solvant. Cependant, en pratique, une telle renormalisation ne permet pas d'obtenir des trajectoires de dynamiques stables (résultats non montrés ici), et nous avons donc utilisé le facteur de charge pour renormaliser les charges du soluté de façon cohérente pour le calcul de toutes les interactions. Nous pouvons justifier ce choix par le fait que les charges issues des champs de force classiques sont adaptées pour convenir aux modèles de solvant explicites, et pas à notre modèle de solvant.

### Résultats sur les molécules chargées

Nous présentons d'abord les résultats obtenus sur les 18 molécules et résidus chargés, avec différentes valeurs du facteur de charge (option « slt-chrg » de notre programme) et sans le facteur de charge (ou slt-chrg=1.0). On a tracé sur la figure 3.15 en abscisse la valeur de l'énergie libre de solvation obtenue avec DelPhi (en kJ/mol) et en ordonnée la valeur moyenne obtenue avec le modèle local des pseudo-particules polarisables (PPP). Le coefficient de corrélation, la pente et l'erreur relative moyenne pour chaque facteur de charge sont présentés dans le tableau 3.2. Le tableau de valeurs des énergies et des erreurs relatives pour toutes les molécules est présenté dans l'Annexe A.3.

<i>Facteur de Charge</i>	<i>Coef. de corrélation</i>	<i>Pente</i>	<i>Erreur relative</i>
<i>1,0 (sans facteur)</i>	0,925	1,251	9,96 %
<i>1,185</i>	0,937	1,270	5,94 %
<i>1,2</i>	0,940	1,282	5,85 %
<i>1,225</i>	0,936	1,265	6,51 %

TAB. 3.2 : Coefficients de corrélation, pente et erreur relative moyenne entre les énergies libres de solvation calculées avec le modèle PPP et avec DelPhi, pour les molécules et résidus chargés.

Le facteur de charge doit modifier les résultats des énergies libres de solvation pour un dipôle, et peu pour les molécules chargées. On voit en effet sur les courbes et le tableau que le facteur de charge change peu les coefficients de corrélation et les pentes mais diminue l'erreur relative par rapport à la valeur de DelPhi de presque un facteur

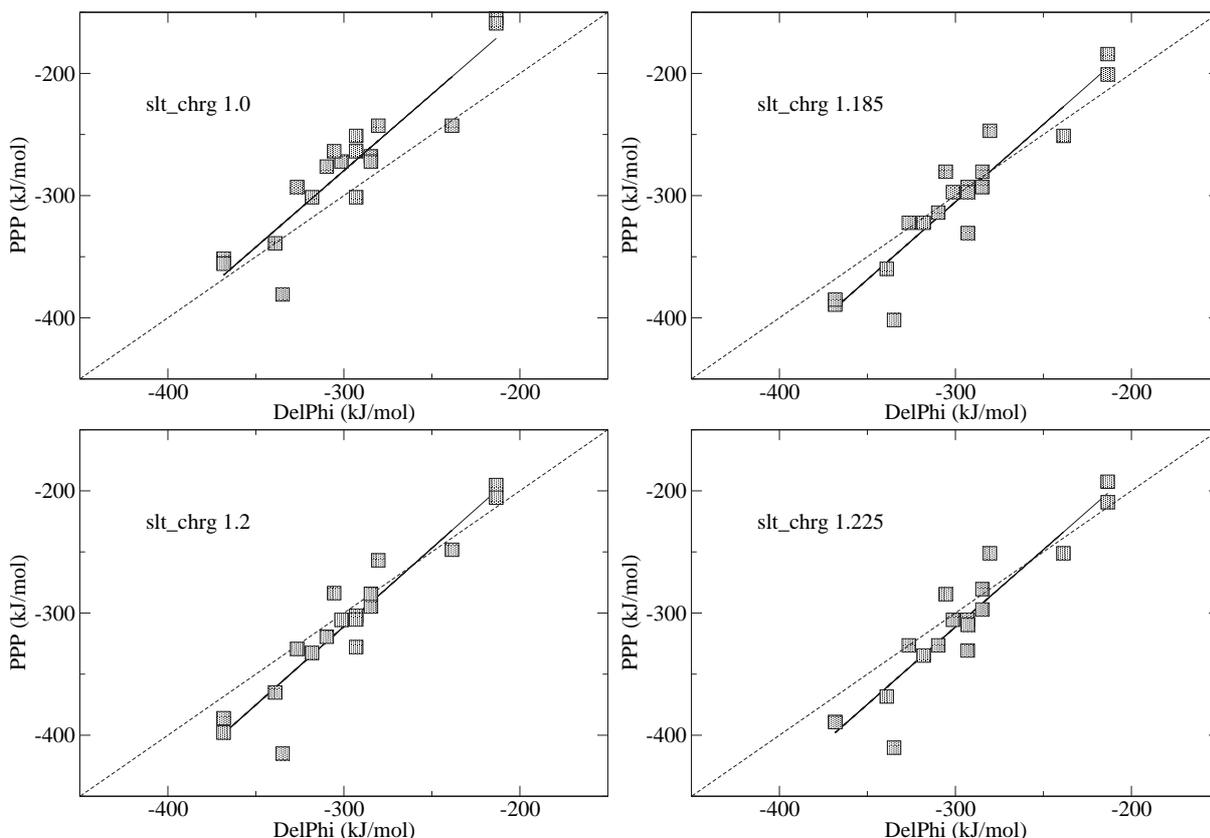


FIG. 3.15 : Corrélations entre les énergies libres électrostatiques de solvation des molécules et résidus chargés calculées avec le modèle PPP et avec DelPhi pour différents facteurs de charge.

2. Avec un facteur de charge de  $\lambda_q = \sqrt{3/2} = 1,225$ , l'erreur relative est un peu plus élevée (6,51 %) qu'avec les deux autres facteurs. Avec ou sans facteur de charge, les pentes des droites de corrélation sont légèrement supérieures à un. Les résultats sur les molécules chargées ne permettent pas de déterminer le meilleur facteur de charge. On voit sur les courbes qu'un point, correspondant au méthylammonium, est très en-dessous de la droite dans tous les cas, et il augmente fortement la pente de la droite et fait diminuer les coefficients de corrélation. Cette molécule contient plusieurs atomes d'hydrogène de faibles rayons, et on sait que notre modèle est moins adapté pour calculer l'énergie libre de solvation pour des atomes de petits rayons.

### Résultats sur les molécules neutres

Les mêmes calculs effectués sur 38 molécules et résidus neutres sont présentés sur la figure 3.16 pour les droites de corrélation, et sur le tableau 3.3 pour les valeurs des

coefficients de corrélation, pentes, et erreurs relatives moyennes. Le tableau donnant les valeurs des énergies et erreurs relatives pour toutes les molécules étudiées est présenté en Annexe A.3.

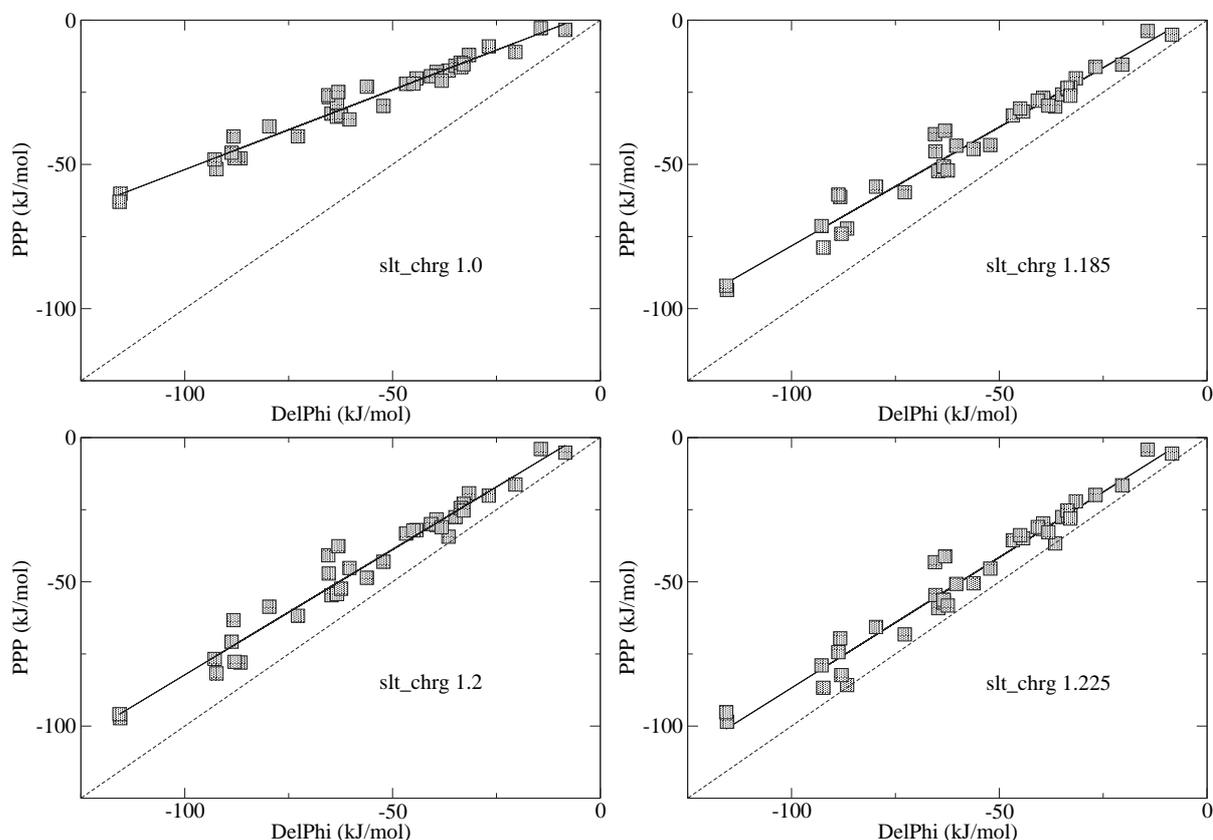


FIG. 3.16 : Corrélations entre les énergies libres électrostatiques de solvation des molécules et résidus neutres calculées avec le modèle PPP et avec DelPhi pour différents facteurs de charge.

Facteur de Charge	Coef. de corrélation	Pente	Erreur relative
1,0 (sans facteur)	0,980	0,552	52,58 %
1,185	0,983	0,824	27,33 %
1,2	0,981	0,869	23,86 %
1,225	0,980	0,905	18,74 %

TAB. 3.3 : Coefficients de corrélation, pente et erreur relative moyenne entre les énergies libres de solvation calculées avec le modèle PPP et avec DelPhi, pour les molécules et résidus neutres.

Là encore, le facteur de charge améliore significativement l'erreur relative, en la di-

minuant d'au moins un facteur 2. Cependant, l'erreur relative reste toujours assez élevée, de l'ordre de 20 à 25 %. On observe donc un léger décalage d'une dizaine de kJ/mol, qui, pour de faibles valeurs des énergies de solvatation, va entraîner une erreur relative élevée. Pour ces molécules et résidus neutres, l'introduction du facteur de charge améliore nettement la pente de la droite de corrélation, qui est très faible sans ce facteur (0,552). Les pentes des droites restent légèrement en-dessous de un, contrairement à celles des molécules chargées qui étaient au-dessus de un. Les coefficients de corrélation sont meilleurs pour les molécules neutres que pour les molécules chargées, ils sont autour de 0,98 quelque soit le facteur de charge. Le meilleur coefficient de corrélation (0,983) est obtenu avec un facteur de charge de  $\lambda_q = 1,185$ , mais la meilleure pente (0,905) et la meilleure erreur relative (18,74 %) sont obtenues pour un facteur de charge de  $\lambda_q = 1,225$ . L'énergie de solvatation obtenue est correcte à une faible constante près, ce qui importera peu si nous nous intéressons à des différences d'énergies libres de solvatation.

### Conclusion sur les énergies libres de solvatation des petites molécules

Nous avons obtenu pour les molécules neutres et chargées des coefficients de corrélation de l'ordre de 0,95 en moyenne, ce qui est tout à fait encourageant, et les pentes des droites de corrélation sont relativement proches de un. Le facteur de charge améliore considérablement la précision de notre modèle pour les coefficients de corrélation, les pentes des droites et les erreurs relatives. On peut penser qu'une valeur de  $\lambda_q = 1,185$  ou 1,2 constituerait un bon compromis pour les molécules chargées et non chargées, d'autant plus que le facteur de charge de  $\lambda_q = 1,225$  ne permet pas de reproduire la structure de petits peptides, comme nous le verrons au chapitre suivant.

#### 3.3.4 Conclusion relative au paramétrage du modèle local

Le paramétrage du modèle local de pseudo-particules polarisables a consisté à reproduire les propriétés électrostatiques de l'eau ainsi que les propriétés structurales des premières couches d'hydratation. Nous avons introduit un facteur de rayon qui renormalise les rayons des atomes pour obtenir des énergies libres de solvatation plus précises. Ce facteur de rayon permet aussi d'obtenir une fonction de distribution radiale du solvant autour d'un ion plus en accord avec un solvant explicite, ainsi qu'un meilleur potentiel de force moyenne entre deux ions de charges opposées. Pour obtenir une énergie libre de solvatation correcte pour les molécules neutres, nous avons aussi introduit un facteur de charge, qui permet d'obtenir de meilleurs résultats pour les énergies libres de solvatation de petites molécules et résidus. Ces deux facteurs transforment de façon systématique

les charges et rayons issus des champs de force classique pour qu'ils conviennent à notre modèle. Cela nous permettra dans la partie suivante de ne pas avoir à chercher de nouveaux paramètres pour chaque nouvelle molécule étudiée, contrairement aux méthodes habituelles (par exemple Born Généralisé).

Nous avons trouvé que les paramètres de Lennard-Jones qui convenaient le mieux pour les particules de solvant étaient  $\epsilon_s = 3,197$  kJ/mol et  $\sigma_s = 2,88$  Å, qui vérifient à  $T = 300$  K et  $\rho = 0,033$  Å<sup>-3</sup> les conditions  $T^* = kT/\epsilon_s = 0,8$  et  $\rho^* = \rho\sigma_s^3 = 0,8$ . La polarisabilité des particules est alors  $\alpha = 2,33$  Å<sup>3</sup>. Nous avons vu que pour avoir une bonne fonction de distribution radiale autour d'un ion et un bon potentiel de force moyenne entre deux charges opposées, la valeur du dipôle de saturation de  $\mu = 1,5$  Debye était relativement bien adaptée. En pratique, quelques simulations de dynamique moléculaire sur de gros systèmes biologiques entourés d'ions ont été effectuées avec des valeurs de  $\mu$  plus élevées, et nous avons observé une déstructuration très rapide des molécules. Nous ne présenterons pas ces simulations dans les chapitres suivants, où nous aurons toujours une valeur de  $\mu = 1,5$  Debye.

Nous avons vu que l'énergie libre de solvation que nous calculons avec notre modèle n'est pas aussi précise qu'avec les méthodes de résolution de l'équation de Poisson. Cependant, notre méthode possède l'avantage d'être sur particules et de pouvoir reproduire aussi les propriétés structurales d'un solvant explicite. Notre modèle de solvant apparaît donc comme un bon compromis entre des propriétés électrostatiques et structurales de l'eau. Une fois implémenté dans le programme ORAC et correctement paramétré, nous pouvons maintenant l'appliquer à la dynamique moléculaire de systèmes biologiques. Nous allons présenter ces applications dans les deux chapitres suivants.

## Annexe A.3

### Énergies libres de solvation des petites molécules avec le solvant PPP

Sur les tableaux 3.4 et 3.5, sont représentées les valeurs des énergies libres électrostatiques de solvation respectivement des molécules chargées et neutres, calculées avec DelPhi et avec le modèle de solvant local des pseudo-particules polarisables (PPP) pour quatre différents facteurs de charge  $\lambda_q$ , ainsi que les erreurs relatives.

Molécule	DelPhi	$\lambda_q = 1,0$	$\lambda_q = 1,185$	$\lambda_q = 1,2$	$\lambda_q = 1,225$
Acétate	-318,1	-301,4 (5%)	-322,3 (1%)	-332,6 (5%)	-334,8 (5%)
Éthoxyde	-368,3	-351,6 (5%)	-389,3 (6%)	-397,9 (8%)	-389,3 (6%)
Méthylammonium	-334,8	-380,9 (14%)	-401,8 (20%)	-415,0 (24%)	-410,2 (22%)
Méthoxyde	-368,3	-355,8 (3%)	-385,1 (5%)	-386,0 (5%)	-389,3 (6%)
Guandinium	-326,5	-293,0 (10%)	-322,3 (1%)	-329,4 (1%)	-326,5 (0%)
Diméthylphosphate	-301,4	-272,1 (10%)	-297,2 (1%)	-305,6 (1%)	-305,5 (1%)
Éthylthiolate	-284,6	-267,9 (6%)	-293,0 (3%)	-294,7 (4%)	-297,2 (4%)
Imidazolium	-293,0	-301,4 (3%)	-330,7 (13%)	-327,9 (12%)	-330,7 (13%)
Méthylthiolate	-284,6	-272,1 (4%)	-280,4 (1%)	-284,2 (0%)	-280,4 (1%)
Tétraméthylammonium	-238,6	-242,8 (2%)	-251,1 (5%)	-248,2 (4%)	-251,1 (5%)
Arginine	-309,7	-276,2 (11%)	-313,9 (1%)	-319,4 (3%)	-326,5 (5%)
Acide Aspartique	-280,4	-242,8 (13%)	-246,9 (12%)	-256,9 (8%)	-251,1 (10%)
Acide Glutamique	-305,5	-263,7 (14%)	-280,4 (8%)	-283,7 (7%)	-284,6 (7%)
Lysine	-339,0	-339,0 (0%)	-360,0 (6%)	-365,0 (8%)	-368,3 (9%)
Désoxyribose-Phosphate	-293,0	-263,7 (10%)	-293,0 (0%)	-302,2 (3%)	-305,5 (4%)
Ribose-Phosphate	-293,0	-251,1 (14%)	-297,2 (1%)	-305,3 (4%)	-309,7 (6%)
D-Phosphate 3'	-213,5	-154,9 (27%)	-184,2 (14%)	-195,3 (9%)	-192,5 (10%)
R-Phosphate 3'	-213,5	-159,0 (25%)	-200,9 (6%)	-205,6 (4%)	-209,3 (2%)

TAB. 3.4 : Énergies libres électrostatiques de solvation des molécules chargées en  $\text{kJ/mol}$  : les valeurs de référence calculées avec DelPhi sont dans la deuxième colonne, puis les valeurs calculées avec PPP pour quatre facteurs de charge différents. Les erreurs relatives de PPP par rapport à DelPhi sont indiquées entre parenthèses. « D-Phosphate 3' » et « R-Phosphate 3' » représentent respectivement le désoxyribose-phosphate et le ribose-phosphate de l'extrémité 3' du brin.

Molécule	DelPhi	$\lambda_q = 1, 0$	$\lambda_q = 1, 185$	$\lambda_q = 1, 2$	$\lambda_q = 1, 225$
Méthanol	-31,58	-12,10 (62%)	-20,13 (36%)	-19,30 (39%)	-22,10 (30%)
Acide Acétique	-36,56	-17,45 (52%)	-29,97 (18%)	-34,36 (6%)	-36,71 (0%)
Éthanol	-26,85	-9,08 (66%)	-16,20 (40%)	-20,13 (25%)	-19,88 (26%)
Acétamide	-39,38	-17,96 (54%)	-27,79 (29%)	-30,30 (23%)	-31,68 (20%)
N-méthylacétate	-39,44	-19,17 (51%)	-27,00 (32%)	-28,34 (28%)	-29,72 (25%)
Méthylacétate	-20,43	-11,05 (46%)	-15,44 (24%)	-16,28 (20%)	-16,53 (19%)
Formamide	-46,76	-22,10 (53%)	-33,02 (29%)	-33,27 (29%)	-35,58 (24%)
Méthylamine	-14,31	-2,76 (81%)	-3,77 (74%)	-3,98 (72%)	-4,19 (71%)
Methanethiol	-8,43	-3,39 (60%)	-5,06 (40%)	-5,23 (38%)	-5,57 (34%)
Alaninedipeptide	-52,16	-29,72 (43%)	-43,28 (17%)	-43,03 (18%)	-45,41 (13%)
Alanine	-40,74	-19,42 (52%)	-27,92 (31%)	-29,97 (26%)	-31,01 (24%)
Asparagine	-79,65	-36,79 (54%)	-57,72 (28%)	-58,64 (26%)	-65,63 (18%)
Cystéine	-34,87	-15,82 (55%)	-25,74 (26%)	-27,58 (21%)	-27,54 (21%)
Glutamine	-88,28	-40,31 (54%)	-61,28 (31%)	-63,28 (28%)	-69,60 (21%)
Glycine	-44,26	-20,17 (54%)	-31,68 (28%)	-32,02 (28%)	-34,91 (21%)
Histidine	-86,64	-47,92 (45%)	-72,24 (17%)	-77,98 (10%)	-85,80 (1%)
Isoleucine	-32,89	-15,32 (53%)	-23,31 (29%)	-22,98 (30%)	-27,04 (18%)
Leucine	-33,47	-16,28 (51%)	-23,98 (28%)	-24,53 (27%)	-27,08 (19%)
Méthionine	-44,99	-22,02 (51%)	-30,60 (32%)	-32,19 (28%)	-33,82 (25%)
Phénylalanine	-33,61	-14,73 (56%)	-23,52 (30%)	-24,44 (27%)	-25,28 (25%)
Proline	-38,21	-21,01 (45%)	-29,63 (22%)	-31,06 (19%)	-32,81 (14%)
Sérine	-65,37	-26,66 (59%)	-45,50 (30%)	-47,13 (28%)	-54,58 (17%)
Thréonine	-56,17	-23,06 (59%)	-44,66 (20%)	-48,59 (13%)	-50,52 (10%)
Tryptophane	-64,69	-32,44 (50%)	-52,32 (19%)	-54,66 (16%)	-59,14 (9%)
Tyrosine	-63,37	-28,46 (55%)	-50,81 (20%)	-53,78 (15%)	-57,22 (10%)
Valine	-32,94	-15,32 (53%)	-26,20 (20%)	-25,28 (23%)	-28,08 (15%)
Adénine	-65,48	-25,99 (60%)	-39,51 (40%)	-40,81 (38%)	-43,19 (34%)
Cytosine	-92,37	-51,61 (44%)	-78,81 (15%)	-81,74 (12%)	-86,72 (6%)
Guanine	-115,53	-60,15 (48%)	-93,55 (19%)	-97,27 (16%)	-98,48 (15%)
Thymine	-63,37	-33,48 (47%)	-50,64 (20%)	-54,24 (14%)	-56,21 (11%)
Uracile	-62,38	-33,15 (47%)	-52,11 (16%)	-52,40 (16%)	-58,22 (7%)
1-méthyl-guanine	-88,73	-45,96 (48%)	-60,48 (32%)	-70,73 (20%)	-74,38 (16%)
Pseudo-uracile	-72,76	-40,26 (45%)	-59,64 (18%)	-61,78 (15%)	-68,22 (6%)
D-Phosphate 5'	-60,36	-34,40 (43%)	-43,53 (28%)	-45,20 (25%)	-50,73 (16%)
R-Phosphate 5'	-92,79	-48,30 (48%)	-71,36 (23%)	-76,72 (17%)	-78,98 (15%)

TAB. 3.5 : Énergies libres électrostatiques de solvation des molécules neutres en kJ/mol. « D-Phosphate 5' » et « R-Phosphate 3' » représentent respectivement le désoxyribose-phosphate et le ribose-phosphate de l'extrémité 5' du brin.



## Chapitre 4

# Application aux peptides et protéines



# Application aux peptides et protéines

Le modèle de solvant local des pseudo-particules polarisables (noté PPP) présenté au chapitre précédent a été implémenté dans le programme ORAC et paramétré sur des systèmes simples pour reproduire le mieux possible à la fois les propriétés électrostatiques et structurales des premières couches de solvation. Dans ce chapitre, nous allons tester la validité de notre modèle sur la stabilité et la dynamique de chaînes polypeptidiques. Le but de notre démarche est d'étudier avec notre modèle la dynamique moléculaire de systèmes biologiques connus et déjà étudiés expérimentalement ou en dynamique moléculaire avec d'autres modèles de solvant (solvant explicite, Born Généralisé,...).

L'étude de la dynamique moléculaire des systèmes biologiques avec notre modèle présentée dans ce chapitre comporte plusieurs applications, allant de peptides courts à des systèmes plus complexes. Tout d'abord, nous avons étudié des peptides simples, peu ou pas chargés, composés d'une dizaine d'acides aminés, et pour lesquels on dispose de données expérimentales et théoriques précises. Nous étudierons ainsi plus particulièrement la stabilité des hélices protéiques en solution. Puis, parmi les protéines dont la structure cristallographique est résolue, nous avons choisi d'étudier des protéines largement étudiées en dynamique moléculaire et pour lesquelles on dispose de données expérimentales : la protéine BPTI (« Bovine Pancreatic Trypsine Inhibitor » - Inhibiteur de la trypsine pancréatique bovine), et le domaine de liaison B1 à l'immunoglobuline de la protéine G du streptocoque.

## 4.1 Application aux peptides

Avant de tester notre modèle sur des systèmes biologiques importants comme les protéines et les acides nucléiques, nous avons effectué des dynamiques moléculaires sur des peptides hélicoïdaux courts. Nous voulons ainsi étudier la stabilité des hélices protéiques

avec notre modèle, et comparer nos résultats avec des résultats expérimentaux ou issus d'autres dynamiques.

Nous avons d'abord étudié un peptide « artificiel », l'octa-alanine, en partant d'une forme en hélice  $\alpha$  canonique. Puis nous avons étudié un peptide issu d'une protéine biologique, la ribonucléase A, peptide dont on sait que seulement une partie est en hélice  $\alpha$  en solution. Pour ces deux peptides, nous présenterons d'abord leurs structures en solution et les résultats expérimentaux et théoriques connus, puis les paramètres de simulation que nous avons utilisés, et les paramètres que nous avons choisis d'analyser pendant ces dynamiques. Enfin, nous présenterons les résultats des analyses structurales et énergétiques obtenus avec le modèle de solvant PPP, et nous les comparerons aux résultats expérimentaux ou issus d'autres simulations.

### 4.1.1 L'octa-alanine

L'octa-alanine que nous avons étudiée est composée de huit résidus alanine et globalement neutre. Elle compte au total 83 atomes, et les deux extrémités N et C-terminales sont chargées ( $-1e$  et  $+1e$ ). Nous avons pris comme conformation de départ une hélice  $\alpha$  canonique, présentée sur la figure 4.1. Il est connu qu'en solution cette hélice comporte trop peu d'acides aminés pour être stable [Chakrabarty et Baldwin, 1995]. Une simulation en solvant explicite [Weber *et al.*, 2000] a montré que l'hélice de l'octa-alanine se dépliait en moins d'une nanoseconde de dynamique pour adopter une conformation en « fer à cheval » circulaire, où les deux extrémités se rapprochent, telle que celle présentée sur la figure 4.1.

#### Paramètres de simulation utilisés

Les simulations ont été effectuées à une température  $T = 300$  K avec un thermostat de Nosé-Hoover. Les interactions de Lennard-Jones et électrostatiques sont coupées par une fonction de « switch » à une distance de coupure de  $R_c = 12$  Å. Nous avons utilisé un pas d'intégration multiple : 4 fs pour les interactions longue portée (entre 8 et 12 Å), 2 fs pour les interactions moyenne portée (entre 5 et 8 Å) et 1 fs pour les interactions courte portée (à moins de 5 Å) et intramoléculaires. L'algorithme SHAKE [Rickaert *et al.*, 1977] est utilisé pour contraindre les liaisons impliquant des atomes d'hydrogène à leur longueur d'équilibre.

Nous avons pris une boîte de simulation cubique de 39,1 Å de côté, contenant 1967 particules de solvant (nous avons généré 2000 particules de solvant régulièrement réparties par un réseau cubique centré, et celles qui recouvrent le soluté, ou qui sont à moins

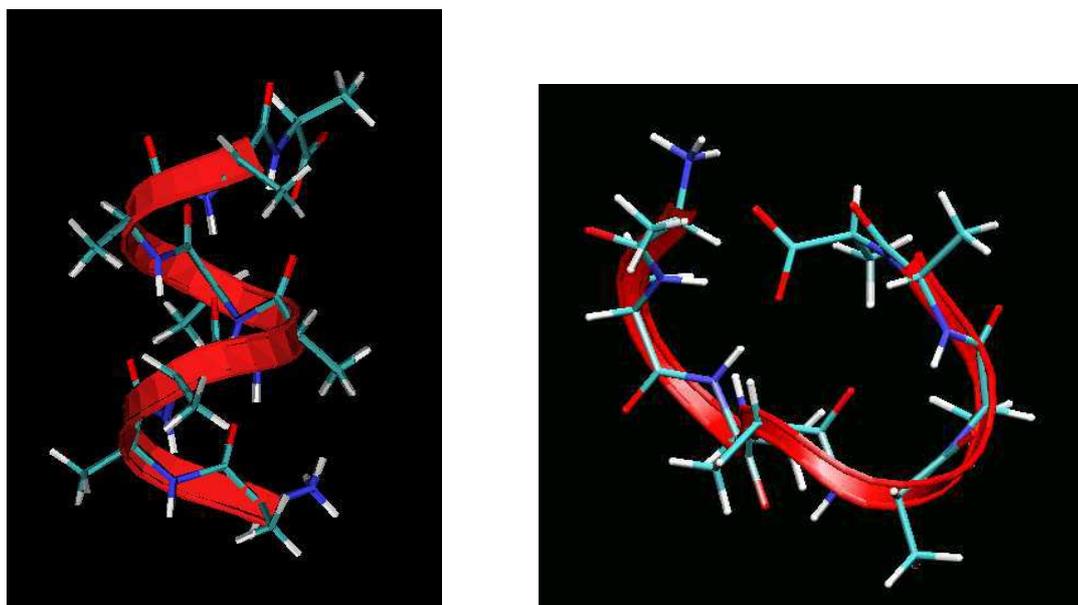


FIG. 4.1 : Structure de l'octa-alanine en hélice  $\alpha$  au début de la simulation (à droite), et en « fer à cheval » à la fin de la simulation (structure observée après 1 ns de dynamique avec le solvant PPP et  $\lambda_q=1,185$ ).

de 0,7 Å du soluté, sont retirées) de paramètres de Lennard-Jones  $\sigma_s = 2,88$  Å, et  $\epsilon_s = 3,197$  kJ/mol. La polarisabilité des particules est  $\alpha = 2,33$  Å<sup>3</sup>, correspondant à la constante diélectrique  $\epsilon = 80$  de l'eau par la formule (3.40), et la valeur du dipôle de saturation de  $\mu = 1,5$  Debye. Pour le soluté, nous avons pris les paramètres du champ de force d'Amber-95 [Cornell *et al.*, 1995] pour les protéines. Nous avons utilisé le facteur de rayon introduit au chapitre 3. Nous avons effectué plusieurs simulations avec différentes valeurs du facteur de charge  $\lambda_q$ .

Après avoir minimisé la structure du soluté dans le vide par la méthode de gradients conjugués (avec un critère de convergence de  $10^{-3}$ ), nous avons équilibré pendant 12 ps les particules de solvant à 300 K autour de la structure minimisée et rigide du soluté. Ensuite, nous avons augmenté progressivement la température du système sans contrainte de 20 à 300 K en plusieurs paliers, puis laissé le système s'équilibrer à 300 K avec le thermostat, pendant 48 ps au total. Pendant ces étapes de chauffage et d'équilibration, les vitesses des atomes sont réassignées à chaque fois que l'énergie ou la température sort de la fourchette demandée. À chaque étape de montée en température, les vitesses des atomes sont attribuées de façon aléatoire selon la distribution de Boltzmann correspondant à la température demandée. La production de dynamique a ensuite été effectuée, sans aucune contrainte, sur une durée d'une nanoseconde pour chaque simulation. Les résultats présentés ensuite sont calculés et analysés sur cette nanoseconde de dynamique.

## Paramètres analysés

### •Analyse structurale

Il est possible de chiffrer la déformation structurale de la molécule à l'aide de plusieurs indices, comme l'écart quadratique moyen des configurations tridimensionnelles de la molécule par rapport à une structure donnée, et la distance entre certains résidus.

Nous avons étudié pour chaque simulation l'évolution dans le temps du RMS, c'est-à-dire l'écart quadratique moyen (« Root Mean Square ») des structures de la molécule pendant la dynamique par rapport à la structure initiale. Le RMS, exprimé en Å, est défini entre deux conformations A et B d'une molécule de  $N$  atomes par la formule :

$$\text{RMS}_{A-B} = \sqrt{\frac{1}{N} \sum_{i=1}^N |\mathbf{r}_i^A - \mathbf{r}_i^B|^2}. \quad (4.1)$$

Étudié au cours du temps par rapport à la structure initiale, le RMS nous permet d'avoir un indice de déformation globale de la molécule. Il est en général calculé sur tous les atomes sauf les hydrogène (les atomes lourds), ou sur les atomes lourds de la chaîne principale ou squelette (donc sans les chaînes latérales des acides aminés), ou encore seulement sur les atomes de carbone  $C\alpha$  de la chaîne principale. Si la simulation aboutit à une structure stable, les variations du RMS doivent se stabiliser autour d'une valeur moyenne. Mais le RMS ne nous renseigne que sur une déformation globale de la molécule, et non sur les déformations précises et locales de la molécule.

Pour l'octa-alanine, afin d'étudier précisément le rapprochement des deux extrémités observé en solvant explicite, nous nous sommes intéressés à l'évolution dans le temps de la distance entre les deux résidus terminaux. Pour cela, nous avons donc étudié la distance entre l'atome d'azote de l'extrémité N-terminale et l'atome de carbone de l'extrémité C-terminale. Si la structure prend une conformation en fer à cheval, les deux extrémités se rapprochent et cette distance devrait diminuer.

### •Analyse énergétique

Dans cette étude, deux énergies nous intéressent particulièrement : l'énergie potentielle du soluté, qui varie avec les déformations du soluté, et l'énergie libre électrostatique de solvatation, calculée grâce à notre modèle par la formule (3.42).

Afin de vérifier que l'énergie libre électrostatique de solvatation du soluté calculée avec notre modèle est correcte, nous la comparerons avec celle issue de méthodes de résolution de l'équation de Poisson-Boltzmann. Nous avons utilisé le programme DelPhi [Honig *et al.*, 1993, Honig et Nicholls, 1995] pour calculer les énergies électrostatiques

de référence. Notre modèle donne l'énergie libre de solvation en faisant une moyenne sur un certain nombre de configurations des particules de solvant. Or, si les particules de solvant sont mobiles pendant la simulation, le soluté est mobile lui aussi. Nous ne pouvons donc comparer rigoureusement les énergies moyennes sur une dynamique du solvant et du soluté à des énergies calculées sur des structures fixes du soluté. C'est pourquoi nous avons comparé l'énergie de solvation moyenne avec notre modèle sur 12 ps de dynamique à l'énergie de solvation calculée avec DelPhi sur une structure moyenne du soluté sur les mêmes 12 ps. Dans un cas nous calculons donc la moyenne d'une énergie sur une structure mobile, et dans l'autre l'énergie d'une structure moyenne figée. On comprend que cette comparaison n'est pas exacte (contrairement au calcul des énergies de solvation des petites molécules dans le chapitre 3, où les particules de solvant étaient mobiles autour d'un soluté rigide), mais elle nous permet de vérifier si l'ordre de grandeur de l'énergie libre électrostatique de solvation calculée avec notre modèle est correcte et varie bien avec les déformations de la molécule. Les paramètres de DelPhi utilisés sont les suivants : l'espacement de la grille est de  $5 \text{ \AA}^{-1}$ , le pourcentage d'occupation du soluté est de 75 %, et les autres paramètres sont les mêmes que pour les petites molécules étudiées au chapitre 3.

## Résultats

Nous avons effectué trois simulations d'une nanoseconde sur l'octa-alanine avec les paramètres donnés plus haut, et avec trois facteurs de charge différents :  $\lambda_q = 1,185$  (simulation 1),  $\lambda_q = 1,2$  (sim. 2) et  $\lambda_q = 1,225$  (sim. 3). Pour chaque simulation, nous avons étudié l'évolution dans le temps du RMS des atomes lourds (sans les hydrogène), de la distance entre les extrémités de la molécule, de l'énergie potentielle du soluté, et de l'énergie libre électrostatique de solvation, ainsi que la corrélation entre l'énergie libre électrostatique de solvation moyenne sur chacune des 83 étapes de 12 ps de la dynamique et celle calculée avec DelPhi sur les 83 structures moyennes.

Les évolutions du RMS et de la distance entre les deux extrémités pour les trois simulations sont présentées sur la figure 4.2. L'analyse du RMS montre que, dans tous les cas, le RMS par rapport à la structure initiale augmente fortement après 200 à 300 ps de simulation, puis se stabilise. Avec  $\lambda_q = 1,185$  et  $1,2$ , le RMS augmente après 200 ps de simulation et se stabilise autour de  $4 \text{ \AA}$  pour la simulation 1, et autour de  $3,5 \text{ \AA}$  pour la simulation 2. Pour la simulation 3 avec  $\lambda_q = 1,225$ , la transition de structure s'opère un peu plus tard, vers 300 ps, et le RMS se stabilise vers  $4 \text{ \AA}$ . L'analyse de la distance entre les deux extrémités de la molécule représentée également sur la figure 4.2 montre cependant que la structure atteinte après ces transitions n'est pas toujours la structure attendue en fer à cheval. En effet, pour la simulation 3, la distance entre les deux extrémités

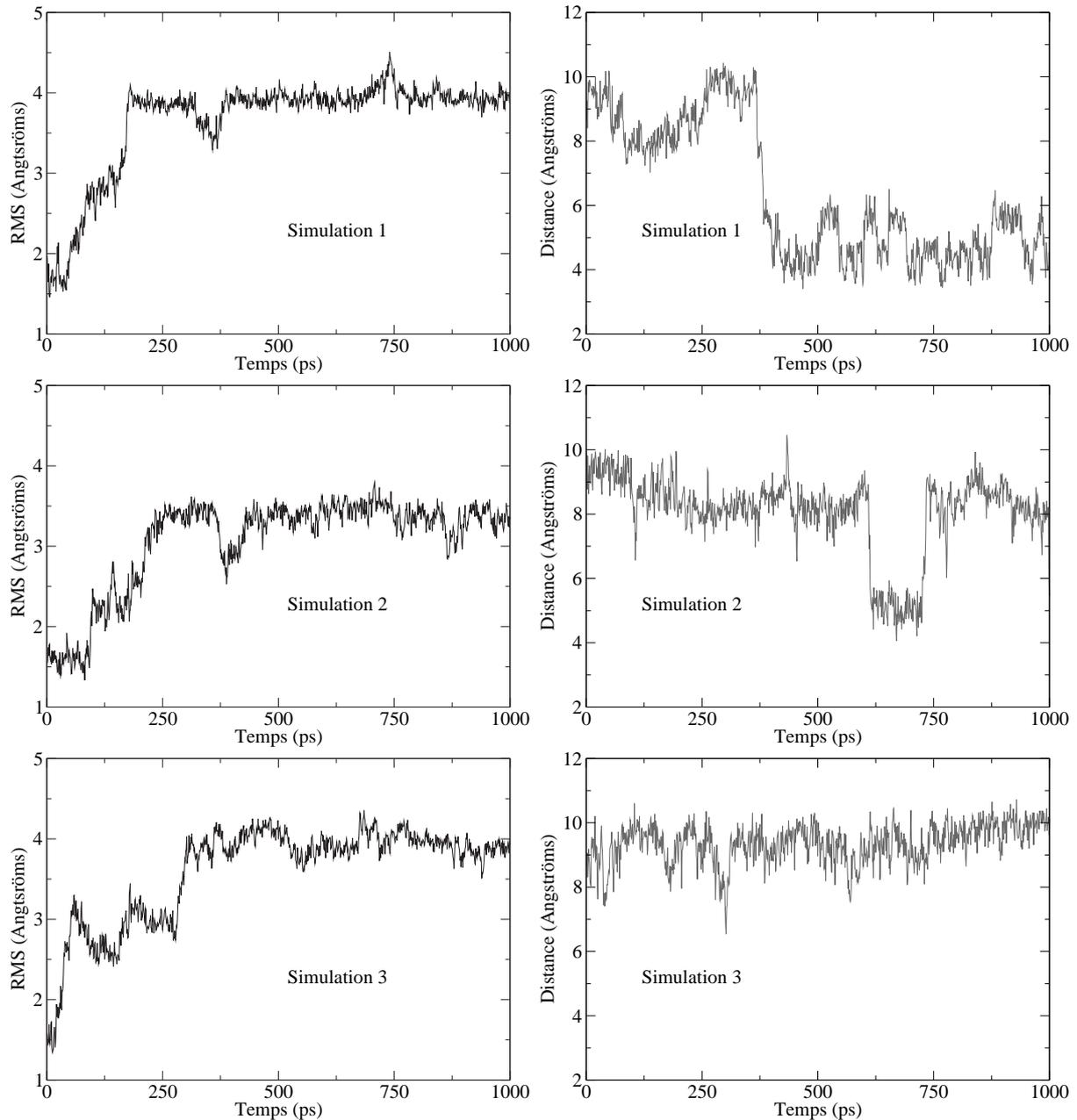


FIG. 4.2 : Évolution du RMS des atomes lourds de l'octa-alanine par rapport à la structure initiale (à gauche), et de la distance entre les deux extrémités (à droite) pendant les trois simulations : simulation 1 avec  $\lambda_q = 1,185$ , simulation 2 avec  $\lambda_q = 1,2$  et simulation 3 avec  $\lambda_q = 1,225$ .

reste globalement constante, entre 8 et 10 Å, ce qui montre que les deux extrémités ne se rapprochent pas, même si l'hélice s'est dépliée. Pour la simulation 2, la distance moyenne est autour de 8 Å, avec une chute temporaire vers 5 Å, pendant environ 100 ps, au bout de 600 ps de simulation. Pour la simulation 1, la distance décroît brusquement

au bout d'environ 350 ps de 10 à 4 Å, indiquant que la conformation adoptée est bien celle circulaire en fer à cheval présentée sur la figure 4.1. Si les trois simulations montrent des déformations structurales importantes par rapport à la structure hélicoïdale initiale, seule la simulation 1 avec  $\lambda_q = 1,185$  montre une transition structurale comparable à celle observée dans l'eau explicite SPC [Weber *et al.*, 2000], la simulation avec  $\lambda = 1,2$  ne montrant qu'une transition passagère vers cette structure et celle avec  $\lambda_q = 1,225$  n'adoptant jamais cette conformation.

L'analyse énergétique des trois simulations présentée sur la figure 4.3 correspond bien à l'analyse structurale. On remarque que les variations de l'énergie potentielle du soluté accompagnent les variations de la distance entre les deux extrémités, plutôt que celles du RMS. On voit ainsi que le rapprochement des extrémités correspond à un minimum de l'énergie potentielle du soluté en solution, qu'on observe bien pour la simulation 1 et de façon transitoire pour la simulation 2. L'énergie potentielle du soluté pour la simulation 3 ne montre pas de transition, mais une légère baisse continue. Quant à l'énergie électrostatique de solvation calculée avec notre modèle, on voit que son évolution est symétriquement opposée à celle de l'énergie potentielle du soluté. L'évolution de l'énergie de solvation compense donc celle de l'énergie potentielle du soluté.

<i>Sim.</i>	$\lambda_q$	<i>Coef. de corrélation</i>	<i>Pente</i>	<i>Erreur absolue</i>	<i>Erreur relative</i>
1	1,185	0,980	1,106	93,7 kJ/mol	22,7 %
2	1,2	0,987	0,934	87,6 kJ/mol	18,7 %
3	1,225	0,876	0,593	104,3 kJ/mol	18,4 %

TAB. 4.1 : Coefficients de corrélation entre les énergies libres de solvation moyennes avec le modèle PPP et les énergies des structures moyennes avec DelPhi, pente de la droite de corrélation et erreur moyenne absolue et relative, pour l'octa-alanine.

Les droites de corrélation entre les énergies libres de solvation électrostatiques calculées avec le solvant PPP et celles calculées avec DelPhi sont présentées sur la figure 4.4. Les valeurs des coefficients de corrélation, des pentes des droites et des erreurs absolues et relatives sont précisées dans le tableau 4.1. On voit que l'énergie libre électrostatique de solvation calculée avec notre modèle est toujours 20 % en-dessous de la valeur de DelPhi, contrairement aux résultats obtenus au chapitre précédent sur les molécules neutres. On remarque ici que pour les simulations 1 et 2 avec  $\lambda_q = 1,185$  et 1,2, le coefficient de corrélation et la pente de la droite sont très proches de un, contrairement à la simulation 3 avec  $\lambda_q = 1,225$  pour laquelle le coefficient de corrélation et la pente en sont beaucoup moins proches. Cela confirme que le facteur  $\lambda_q = 1,225$  est mal adapté à la simulation de peptides. Pour les deux autres facteurs de charge, les variations de l'énergie libre de

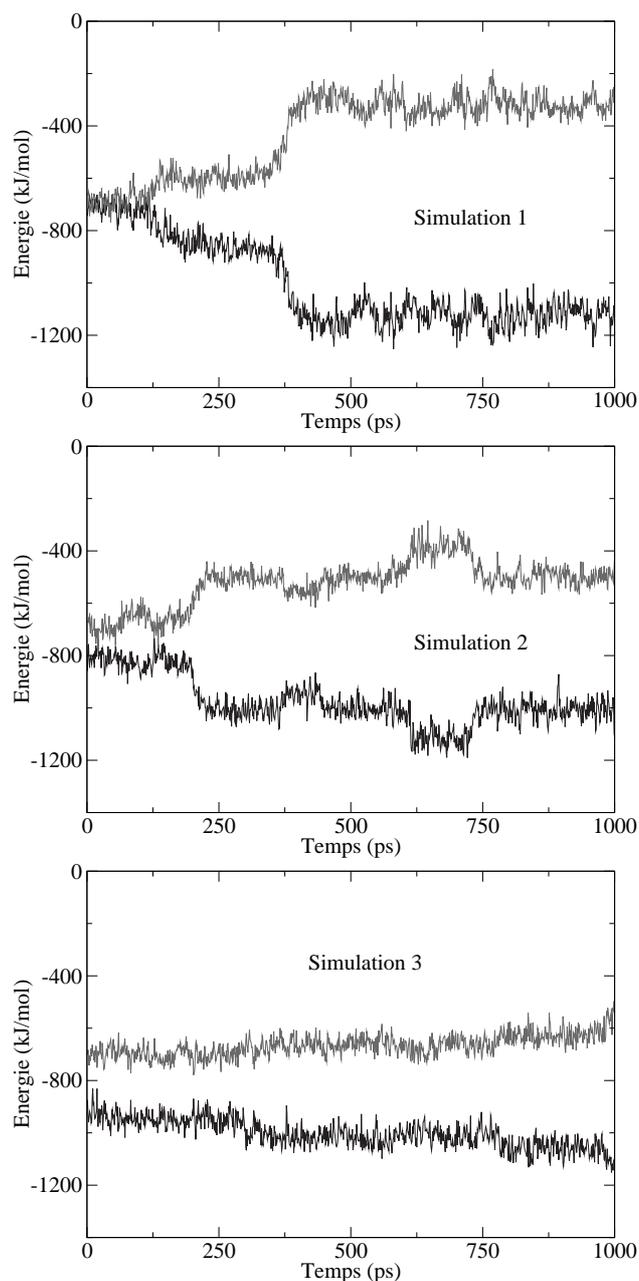


FIG. 4.3 : Évolution de l'énergie potentielle du soluté (en noir) et de l'énergie libre électrostatique de solvation (en gris) pendant les trois simulations : simulation 1 avec  $\lambda_q = 1,185$ , simulation 2 avec  $\lambda_q = 1,2$  et simulation 3 avec  $\lambda_q = 1,225$ .

solvation dues aux changements conformationnels de la molécule suivent bien les valeurs de Poisson-Boltzmann.

Pour les simulations de l'octa-alanine, seul le facteur de charge  $\lambda_q = 1,185$  montre très rapidement la même transition structurale que celle observée en solvant explicite, le facteur  $\lambda_q = 1,2$  ne montrant la transition que pendant une courte période transitoire.

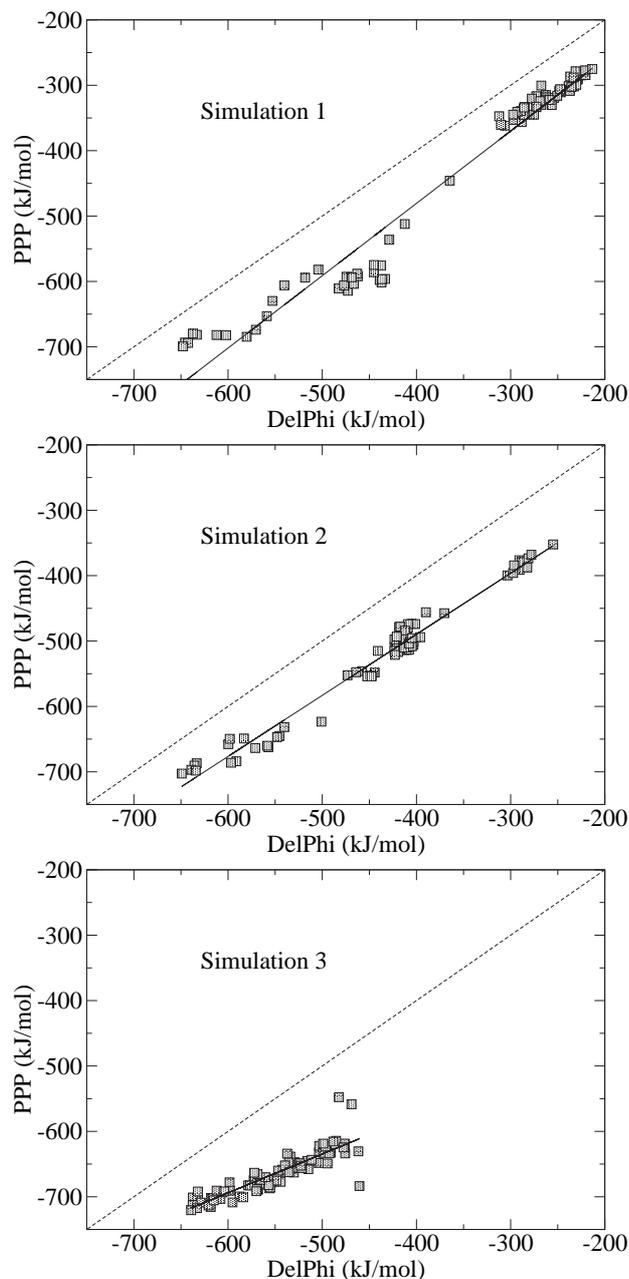


FIG. 4.4 : Corrélations entre les énergies libres électrostatiques de solvation moyennes avec le modèle PPP et les énergies des structures moyennes avec DelPhi, pour les trois simulations de l'octa-alanine.

Pour ces deux facteurs de charge, l'énergie libre de solvation est en bonne corrélation avec celle calculée par résolution de Poisson-Boltzmann. Notre modèle permet donc de retrouver à la fois les changements structuraux observés en solvant explicite, et approximativement les valeurs d'énergies libres électrostatiques des méthodes continues. Le facteur de charge de  $\lambda_q = 1,225$  ne semble pas adapté à cette étude, puisque les déformations

structurales observées ne correspondent pas à celles qui s'opèrent en solvant explicite, et les énergies libres de solvation calculées ne sont pas en bon accord avec celles calculées avec DelPhi. Dans la suite, on n'utilisera plus comme facteur de charge que  $\lambda_q = 1,185$  ou  $\lambda_q = 1,2$ .

### 4.1.2 L'hélice N-terminale du ribonucléase A

Pour compléter l'étude des peptides, nous avons étudié une partie d'une protéine biologique dont la structure cristallographique est résolue. Nous avons choisi un peptide analogue de l'hélice N-terminale du ribonucléase A (ou peptide C, appelé Rn-24). La séquence des 13 acides aminés de ce peptide, de charge totale  $+1e$ , est la suivante :  $\text{NH}_2\text{-Ala-Glu}^{-1}\text{-Thr-Ala-Ala-Ala-Lys}^{+1}\text{-Phe-Leu-Arg}^{+1}\text{-Ala-His}^{+1}\text{-Ala-CO}_2^{-1}$ . Le principal intérêt de ce peptide est que l'on sait qu'il peut former une courte hélice en solution, à basse température, et que sa structure est assez stable, même coupée du reste de la protéine. On a représenté sur la figure 4.5 la structure tridimensionnelle du peptide.

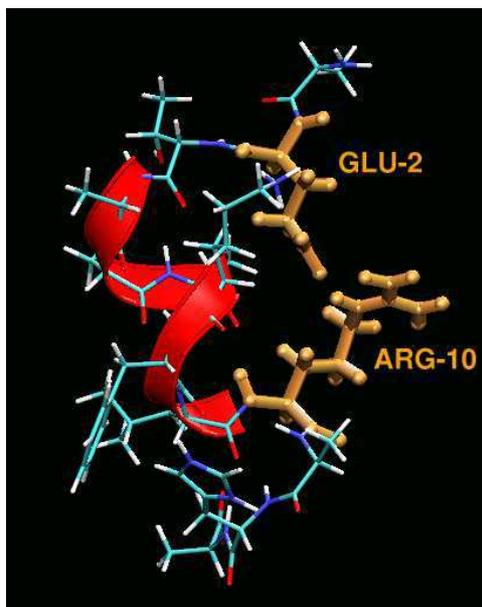


FIG. 4.5 : Structure tridimensionnelle du peptide analogue de Rn-24 : l'hélice  $\alpha$  est représentée en rouge, et les résidus chargés Glu-2 et Arg-10 sont mis en évidence en orange. Image réalisée avec VMD [Humphrey et al., 1996].

De nombreuses données expérimentales et théoriques sont disponibles pour ce peptide ou des peptides analogues<sup>1</sup>. Des expériences de dichroïsme circulaire, permettant de

<sup>1</sup>les peptides analogues sont des peptides dont les acides aminés principaux sont les mêmes que le peptide Rn-24, mais pour lesquels certains acides aminés sont différents.

distinguer les structures secondaires, ont montré que ce peptide en solution contenait 50 à 60 % de ses résidus impliqués dans une hélice  $\alpha$  [Schoemaker *et al.*, 1987]. Des expériences de RMN-2D [Osterhout *et al.*, 1989] ont montré que les trois résidus de l'extrémité N-terminale étaient moins hélicoïdaux que les autres résidus du peptide. Ces mêmes expériences de RMN ainsi que d'autres expériences de dichroïsme circulaire [Fairman *et al.*, 1990] sur ce peptide et des mutants<sup>2</sup> concordent pour mettre en évidence la formation d'un pont salin entre les deux résidus chargés Glu-2 et Arg-10, qui stabiliserait la conformation du peptide en hélice partielle. Si un de ces deux acides aminés est muté, la structure n'est plus hélicoïdale [Fairman *et al.*, 1990]. Cette propriété structurale de pont salin a été confirmée par des études théoriques de simulations de dynamique moléculaire en solvant explicite [Tirado-Rives et Jorgensen, 1991], et de simulations de Monte-Carlo avec une représentation implicite du solvant sur des peptides analogues [Hansmann et Okamoto, 1998].

### Paramètres de simulation

Pour la structure initiale du peptide étudié, nous sommes partis de la structure cristallographique du Ribonucléase A déposée dans la Protein Data Bank avec le code d'entrée 1AFU [Leonidas *et al.*, 1997], coupée aux 13 premiers acides aminés. Puis nous avons substitué certains acides aminés (la Lys-1 en Ala-1, le Glu-9 en Leu-9, la Gln-11 en Ala-11 et la Met-13 en Ala<sup>-1</sup>-13) afin que la séquence corresponde exactement à celle du peptide étudié par RMN [Osterhout *et al.*, 1989]. Puis nous avons minimisé par la méthode des gradients conjugués (avec un critère de convergence de  $10^{-3}$ ) la structure obtenue après ces substitutions. Les paramètres et la topologie utilisés pour le peptide sont ceux d'Amber-95.

Les paramètres de simulation utilisés ici pour notre modèle de solvant ainsi que pour les pas d'intégration et la distance de coupure sont les mêmes que pour l'octa-alanine. Nous avons effectué deux simulations à  $T = 275$  K avec un facteur de charge de  $\lambda_q = 1,185$  (simulation 1) et  $\lambda_q = 1,2$  (sim. 2), et une simulation à  $T = 355$  K avec un facteur de charge de  $\lambda_q = 1,185$  (sim. 3), pour vérifier que l'hélice du peptide se dénaturait à haute température comme l'ont observé d'autres études expérimentales et théoriques [Osterhout *et al.*, 1989, Fairman *et al.*, 1990]. La boîte de simulation est la même que l'octa-alanine, sauf qu'elle contient 1933 particules de solvant après avoir introduit le soluté. L'équilibration du système en température a été effectuée plus progressivement que pour l'octa-alanine, sur un total de 130 ps, car la structure du peptide est plus sensible

---

<sup>2</sup>les mutants sont des peptides dans lesquels certains des acides aminés principaux ont été substitués.

à la température. Les trois simulations ont été produites sur deux nanosecondes, en 166 étapes de 12 ps.

### Paramètres analysés

Pour cette étude, nous avons analysé l'évolution dans le temps du RMS des atomes lourds par rapport à la structure minimisée de départ. Nous nous sommes aussi intéressés à l'évolution temporelle de la distance entre les résidus chargés Glu-2 et Arg-10, car les études expérimentales et théoriques ont montré la formation d'un pont salin entre les deux chaînes latérales de ces résidus. Plus précisément, nous avons analysé la distance entre le carbone du groupe  $\text{CO}_2^-$  à l'extrémité de la chaîne latérale du résidu Glu-2, et le carbone du groupe  $\text{C}(\text{NH}_2)_2^+$  à l'extrémité de la chaîne latérale du résidu Arg-10.

Pour affiner l'étude structurale, nous nous sommes intéressés à l'évolution du pourcentage d'hélicité globale du peptide, et au pourcentage d'hélicité par résidu, moyenné sur toute la simulation. Pour cela, nous avons calculé les angles  $\phi$  et  $\psi$  (présentés au chapitre 1) pour chaque résidu, et nous avons utilisé le critère suivant pour considérer que le résidu est hélicoïdal :  $\phi \in [-85^\circ, -25^\circ]$  et  $\psi \in [-75^\circ, 0^\circ]$ . Le critère est assez large pour pouvoir prendre en compte à la fois les hélices  $\alpha$  et les hélices  $3_{10}$ . Nous avons fait ce calcul pour chacune des 166 structures PDB issues de notre simulation, et nous en avons déduit le pourcentage d'hélicité pour chaque résidu. En faisant la moyenne de l'hélicité sur tous les résidus pour chaque structure PDB, on obtient le pourcentage d'hélicité globale du peptide en fonction du temps, et en faisant la moyenne de l'hélicité de chaque résidu sur toutes les structures PDB, on obtient le pourcentage d'hélicité moyen par résidu sur la dynamique.

Enfin, pour l'étude énergétique, nous avons calculé les corrélations entre les énergies libres électrostatiques de solvation moyennes avec notre modèle et les calculs électrostatiques de Poisson-Boltzmann (avec DelPhi) sur les structures moyennes de la même façon que pour l'octa-alanine. Dans le chapitre 3, pour l'étude de l'énergie libre de solvation des molécules chargées, nous avons rajouté la contribution du bulk  $\Delta G_{bulk}$ , calculée par la formule de Born pour une cavité sphérique de rayon  $R_c$  et de la charge de la molécule. Cependant, pour des molécules de taille plus importante, ce  $\Delta G_{bulk}$  n'a plus vraiment de sens car on ne peut pas considérer que la cavité est sphérique (puisque la cavité est formée par un rayon  $R_c$  autour de la molécule que l'on ne peut pas considérer comme étant sphérique). Nous avons donc calculé l'énergie électrostatique de solvation directement avec la formule 3.47. Notons que dans le cas de l'octa-alanine, la molécule était neutre et on avait donc  $\Delta G_{bulk} = 0$ . Les paramètres de DelPhi utilisés pour calculer les énergies de ce peptide sont les mêmes que pour l'octa-alanine.

## Résultats

L'évolution du RMS des atomes lourds par rapport à la structure minimisée initiale pour les trois simulations est représentée sur la figure 4.6. Pour les deux simulations à  $T = 275$  K, le RMS reste d'abord autour de 2 Å, puis augmente vers 500 ps de simulation pour se stabiliser autour de 3 Å. Le RMS de la simulation 2 passe légèrement au-dessus de celui de la simulation 1 après 1,3 ns de simulation. Quant à la simulation 3, son RMS passe par plusieurs paliers, un à 3 Å, puis à 4 Å, et enfin, après un peu plus d'une nanoseconde de simulation, il semble se stabiliser à plus de 5 Å, indiquant comme attendu que la structure hélicoïdale du peptide se dénature à haute température.

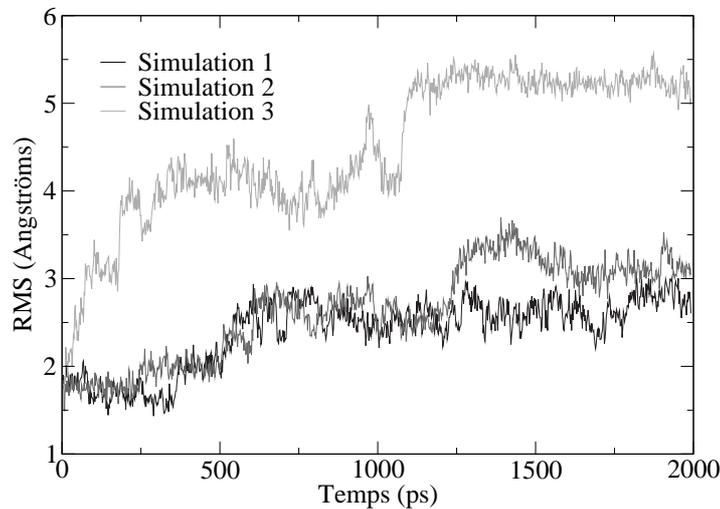


FIG. 4.6 : Évolution du RMS des atomes lourds du peptide *Rn-24* par rapport à la structure initiale pendant les trois simulations : simulations 1 (avec  $\lambda_q = 1,185$ , en noir) et 2 ( $\lambda_q = 1,2$ , en gris foncé) à  $T = 275$  K, et simulation 3 à  $T = 355$  K ( $\lambda_q = 1,185$ , en gris clair).

Afin de mieux voir l'influence de la température sur la structure, nous avons calculé l'hélicité globale moyenne de la molécule pendant les trois simulations. Pour les simulations 1 et 2, le pourcentage d'hélicité globale varie autour d'une valeur moyenne de respectivement 55 % et 52 %, alors que pour la simulation 3 l'hélicité globale varie autour de 30 %, confirmant qu'une température élevée dénature la structure hélicoïdale du peptide. Afin de voir plus précisément les résidus du peptide qui sont ou non hélicoïdaux, on a représenté sur la figure 4.7 le pourcentage d'hélicité moyen par résidu pour les trois simulations. On voit clairement que pour les simulations 1 et 2, les trois premiers résidus ne participent pas à l'hélice, et que ce sont les résidus 4 à 11 qui sont les plus hélicoïdaux, confirmant les résultats expérimentaux et théoriques cités plus haut. Pendant la

simulation 3, on voit que l'hélice n'est conservée que partiellement entre les résidus 4 et 7, confirmant que l'hélice  $\alpha$  du peptide n'est stable qu'à basse température.

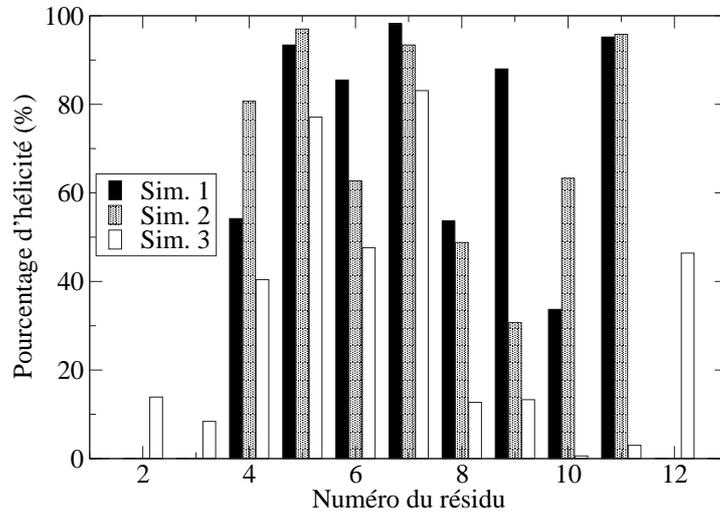


FIG. 4.7 : Hélicité moyenne sur toute la simulation de chaque résidu du peptide *Rn-24* pour les trois simulations.

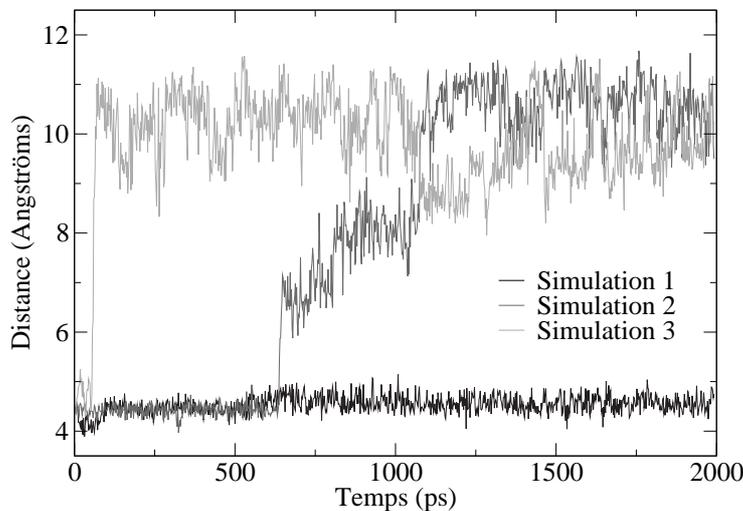


FIG. 4.8 : Évolution de la distance entre *Glu-2* et *Arg-10* du peptide *Rn-24* pendant les trois simulations.

Nous nous sommes aussi intéressés à la distance entre les groupements chargés des chaînes latérales de *Glu-2* et *Arg-10*, dont l'évolution est représentée pour les trois simulations sur la figure 4.8. Pour la simulation 1, la distance reste constante autour de 4,5 Å pendant toute la durée de la simulation. Pour la simulation 2, cette distance reste aussi autour de 4,5 Å jusqu'à environ 600 ps de simulation, puis augmente en passant par deux paliers et atteint une distance d'un peu plus de 10 Å, comparable à celle la

simulation 3, qui augmente dès le début de la simulation. Notons que pour la simulation 2, la brusque augmentation de la distance entre les deux résidus ne s'accompagne pas d'une augmentation importante du RMS (qui reste comparable à celui de la simulation 1). Pour la simulation 1, on remarque que les variations de la distance sont beaucoup plus faibles que pour les deux autres simulations, indiquant que la structure est très stable. Nous retrouvons donc avec un facteur de charge de  $\lambda_q = 1,185$  le résultat expérimental et théorique du pont salin entre les chaînes latérales de ces deux résidus, qui n'a lieu qu'à 275 K [Osterhout *et al.*, 1989, Fairman *et al.*, 1990]. Avec  $\lambda_q = 1,2$ , on n'observe ce pont salin que pendant les 600 premières picosecondes de la simulation.

Enfin, l'étude énergétique des corrélations entre les énergies libres de solvation moyennes avec le modèle PPP et les énergies calculées avec DelPhi sur les structures moyennes est représentée sur la figure 4.9. La valeur de l'énergie de solvation avec le solvant PPP est toujours en-dessous de celle de Poisson-Boltzmann. Les coefficients de corrélation, les pentes et les erreurs relatives pour les trois simulations sont précisés dans le tableau 4.2. Les corrélations ne sont pas aussi bonnes que pour l'octa-alanine, et on peut l'expliquer par le fait que notre modèle donne de moins bonnes corrélations de l'énergie libre de solvation pour les molécules chargées, comme nous l'avons vu au chapitre précédent (ici, la charge totale du peptide est  $+1 e$ ). De plus, comme les molécules de solvant bougent en même temps que les atomes du soluté pendant la dynamique, l'énergie calculée avec notre modèle n'est jamais vraiment une énergie libre à l'équilibre, contrairement à la théorie de Poisson-Boltzmann. Néanmoins, on observe qu'avec  $\lambda_q = 1,185$ , les pentes des droites de corrélation sont plus proches de un que pour  $\lambda_q = 1,2$ . C'est aussi pour la simulation 2, avec  $\lambda_q = 1,2$ , que l'erreur relative est la plus élevée. Le facteur de charge  $\lambda_q = 1,185$  semble une nouvelle fois mieux adapté à l'étude des petits peptides.

<i>sim.</i>	$\lambda_q$	$T$	<i>Coef. de corrélation</i>	<i>Pente</i>	<i>Erreur relative</i>
1	1,185	275 K	0,751	1,027	8,23 %
2	1,2	275 K	0,806	1,929	18,13 %
3	1,185	355 K	0,883	0,814	14,25 %

TAB. 4.2 : Coefficients de corrélation entre les énergies libres de solvation moyennes avec le modèle PPP et les énergies calculée avec DelPhi sur les structures moyennes, pente de la droite de corrélation, et erreur relative moyenne pour le peptide Rn24.

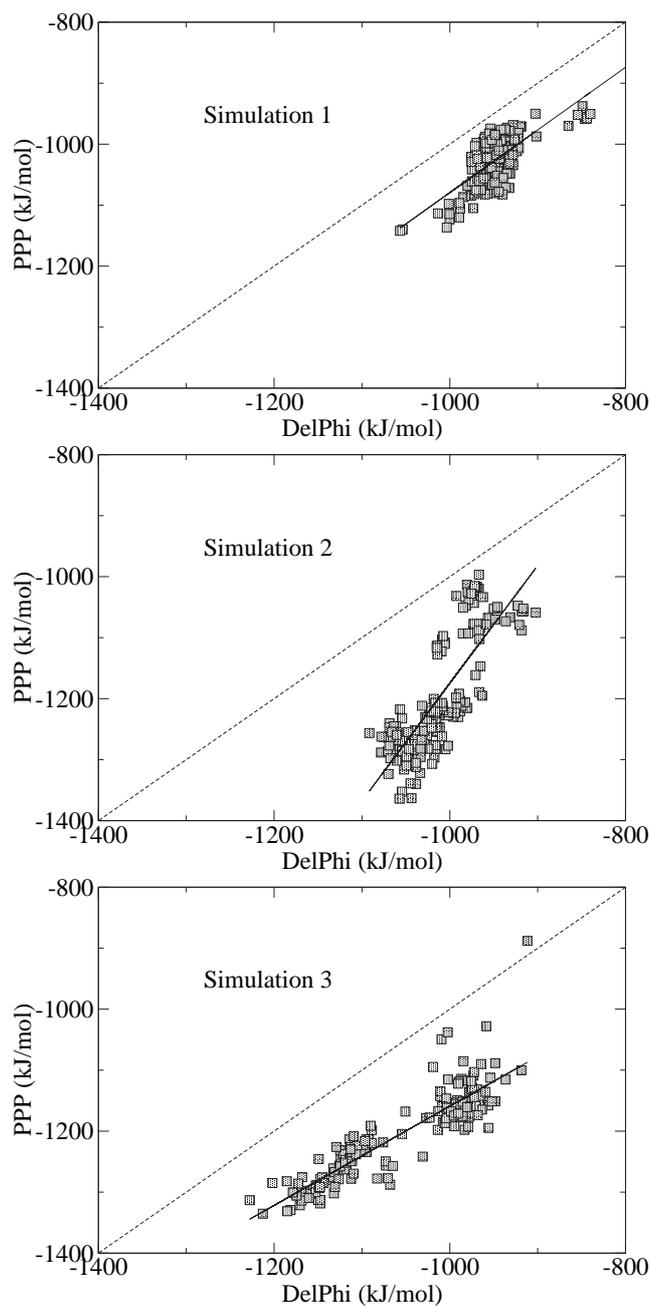


FIG. 4.9 : Corrélations entre les énergies libres électrostatiques de solvation moyennes avec PPP et les énergies des structures moyennes avec DelPhi, pour les trois simulations du peptide Rn24.

Notre modèle, s'il ne donne pas une valeur très précise de l'énergie libre de solvation à chaque instant, donne néanmoins une évolution de l'énergie libre de solvation qui est globalement en accord avec la théorie de Poisson-Boltzmann.

### 4.1.3 Conclusion sur l'étude des peptides

Ces deux simulations de peptides courts sont très encourageantes, car elles réussissent à reproduire avec notre modèle de solvant PPP des résultats structuraux en accord avec les résultats expérimentaux ou théoriques. Même si l'énergie libre électrostatique de solvatation calculée avec notre modèle n'est pas à chaque instant exactement celle calculée avec les méthodes basées sur la résolution de Poisson-Boltzmann, nous avons vu que ces énergies libres corrélaient relativement bien avec les changements structuraux du soluté. Un facteur de charge de  $\lambda_q = 1,185$  donne les meilleurs résultats pour ces deux études, aussi bien d'un point de vue structural qu'énergétique.

## 4.2 Application aux protéines

Les deux protéines que nous avons choisi d'étudier sont des protéines globulaires de taille moyenne qui ont fait l'objet de nombreuses études expérimentales et théoriques, et qui sont donc de bons systèmes biologiques pour tester notre modèle. Il s'agit de l'inhibiteur de la trypsine pancréatique bovine (BPTI), et le domaine B1 de liaison à l'immunoglobuline de la protéine G du streptocoque, comportant respectivement 58 et 56 résidus. Nous allons tout d'abord présenter les principales données expérimentales et théoriques disponibles pour les deux protéines. Puis, nous décrirons la méthode utilisée pour la simulation et l'analyse des dynamiques de ces deux protéines. Enfin, nous présenterons les résultats obtenus sur les dynamiques des deux protéines avec notre modèle de solvant PPP.

### 4.2.1 Les protéines étudiées

#### L'inhibiteur de la trypsine pancréatique bovine (BPTI)

La protéine BPTI est une protéine du boeuf très abondante qui régule, en l'inhibant, l'action enzymatique de la trypsine, enzyme protéase du suc pancréatique permettant la digestion. Elle est composée de 58 résidus (892 atomes), et comporte trois ponts disulfures (entre les cystéines 5 et 55, 14 et 38, et 30 et 51), qui stabilisent sa structure tertiaire, ainsi que quatre prolines. Les études expérimentales et théoriques sont nombreuses sur BPTI, car elle constitue un bon modèle pour étudier le dépliement des protéines comportant des ponts disulfures [Kazmirski et Daggett, 1998]. Plusieurs études comparant les résultats issus d'expériences de RMN et ceux issus de simulations de dynamique moléculaire se

sont intéressées à la dynamique interne de BPTI en solution [Smith *et al.*, 1995, Schneider *et al.*, 1999]. Des simulations de dynamique moléculaire dans l'eau explicite et dans le vide ont été effectuées afin d'étudier l'effet du solvant sur les mouvements collectifs de la protéine [Hayward *et al.*, 1993]. Enfin, des simulations de dynamique moléculaire en solvant explicite à haute température ont été effectuées pour étudier le dépliement de BPTI [Kazmirski et Daggett, 1998].

La structure cristallographique de BPTI est disponible dans la base de données des protéines (PDB). Nous avons pris comme structure initiale des simulations la structure issue de la cristallographie par diffraction des rayons X, de code d'entrée 1BPI [Parkin *et al.*, 1996] dans la PDB, dont la charge totale est de  $+6e$ . La structure de la protéine comporte deux feuillets  $\beta$  anti-parallèles :

- $\beta_1$ , du résidu 18 au résidu 24,
- $\beta_2$ , du résidu 29 à 35.

Elle comporte aussi deux courtes hélices  $\alpha$  :

- $\alpha_1$ , du résidu 3 à 6,
- $\alpha_2$ , du résidu 48 à 55.

Sa structure tridimensionnelle avec la mise en évidence des structures secondaires est représentée sur la figure 4.10.

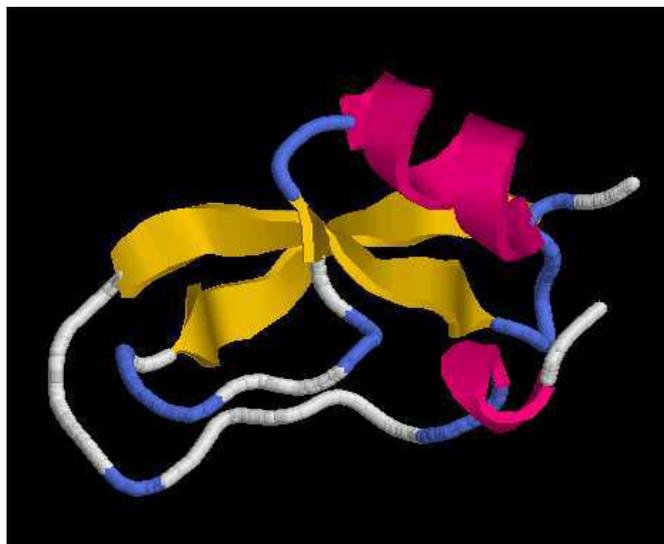


FIG. 4.10 : Structure tridimensionnelle de la protéine de l'inhibiteur de la trypsine pancréatique bovine (BPTI) : seule la chaîne principale est représentée, avec la mise en évidence des structures secondaires (les hélices  $\alpha$  sont en rose et les feuillets  $\beta$  sont représentés par des flèches jaunes). Image réalisée avec Rasmol [Sayle et Milner-White, 1995].

### Le domaine B1 de la protéine G (1PGB)

La protéine que nous noterons 1PGB est le domaine B1 de liaison à l'immunoglobuline de la protéine G du streptocoque. La protéine G est une protéine membranaire participant à la transduction de signaux entre des récepteurs et effecteurs membranaires. La protéine G de la paroi cellulaire de plusieurs espèces de streptocoques (bactéries) contient plusieurs domaines de liaison à l'immunoglobuline, un anticorps, de forte homologie entre eux, et dont la structure est stable en solution séparée du reste de la protéine [Sheinerman et Brooks, 1997]. Le domaine B1 est constituée de 56 résidus (855 atomes). Cette protéine est intéressante à étudier car une très grande proportion de ses résidus sont impliqués dans des structures secondaires. De plus, elle ne comporte aucun pont disulfure ni aucune proline (la proline pose des problèmes d'isomérisation cis/trans qui compliquent les processus de repliement). Tout cela en fait un bon modèle pour étudier le repliement protéique simple et la stabilité des structures secondaires. Les simulations de dynamique moléculaire sur 1PGB sont nombreuses, certaines en solvant explicite avec un modèle d'eau TIP3 [Sheinerman et Brooks, 1997], d'autres avec la méthode de Born Généralisé, pour tester différents paramètres des volumes atomiques [Calimet *et al.*, 2001], et enfin, plus récemment, avec une méthode de diélectrique dépendant de la distance [Hassan *et al.*, 2003] (méthode des « Screened Coulomb Potentials » décrite au chapitre 2). Ces différentes simulations concordent pour conclure que la structure de 1PGB est très stable en solution, en particulier au niveau de ses structures secondaires.

La structure initiale de nos simulations est la structure cristallographique disponible dans la PDB par le code d'entrée 1PGB [Gallagher *et al.*, 1994], dont la charge totale est de  $-4e$ . Notons qu'il existe aussi une structure tridimensionnelle issue d'expériences de RMN pour cette protéine [Gronenborn *et al.*, 1991], de code d'entrée 2GB1 dans la PDB. Les deux structures diffèrent légèrement par l'orientation de l'hélice  $\alpha$  par rapport aux feuillets  $\beta$ , et pour la conformation d'une des boucles. La simulation en solvant explicite [Sheinerman et Brooks, 1997] a montré que cette protéine en solution adoptait une structure plus proche de la structure cristallographique que de la structure issue de la RMN. La structure secondaire de la protéine 1PGB comporte 4 feuillets  $\beta$  anti-parallèles deux à deux :

- $\beta_1$ , du résidu 2 à 8,
- $\beta_2$ , anti-parallèle à  $\beta_1$ , du résidu 13 à 19,
- $\beta_3$ , du résidu 42 à 46,
- $\beta_4$ , anti-parallèle à  $\beta_3$  et parallèle à  $\beta_1$ , du résidu 51 à 56.

Elle comporte aussi une longue hélice  $\alpha$  du résidu 23 à 36, de longueur 21,5 Å. La structure tridimensionnelle de la protéine 1PGB est représentée avec Rasmol sur la figure 4.11, en mettant en évidence les structures secondaires.

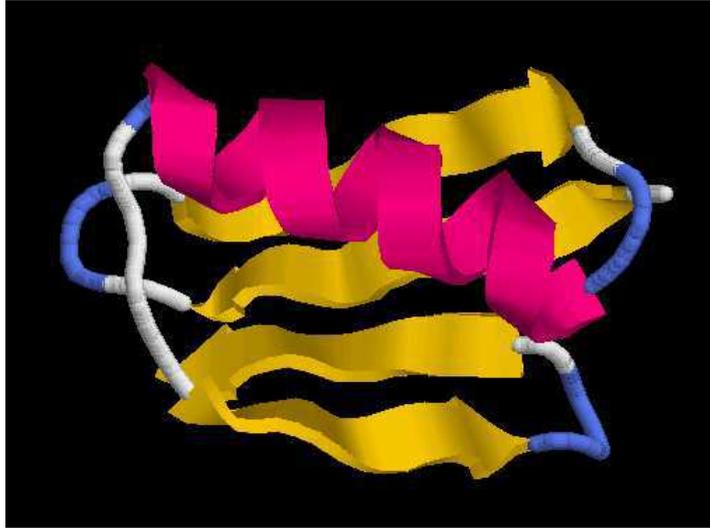


FIG. 4.11 : Structure tridimensionnelle du domaine B1 de la protéine G : seule la chaîne principale est représentée, l'hélice  $\alpha$  est en rose et les quatre feuillets  $\beta$  sont représentés par des flèches jaunes.

## 4.2.2 Paramètres de simulation et d'analyse

### Paramètres de simulation

Nous avons utilisé ici une boîte de simulation rectangulaire de dimensions :  $48 \text{ \AA} \times 48 \text{ \AA} \times 70,7 \text{ \AA}$ , contenant 5196 particules de solvant pour BPTI, et 5209 pour 1PGB (5488 particules de solvant sont générées sur un réseau cubique, puis les particules qui recouvrent le soluté, ou qui sont à moins de  $0,7 \text{ \AA}$  du soluté, sont retirées), de mêmes paramètres de Lennard-Jones et de saturation dipolaire que les simulations précédentes :  $\sigma_s = 2,88 \text{ \AA}$ ,  $\epsilon_s = 3,197 \text{ kJ/mol}$ , et  $\mu = 1,5 \text{ Debye}$ . La température finale de simulation est maintenue à  $300 \text{ K}$  par un thermostat de Nosé-Hoover. Les interactions longue distance sont coupées à une distance  $R_c = 12 \text{ \AA}$ . Les pas d'intégration multiples sont toujours de  $4 \text{ fs}$  pour les interactions longue portée,  $2 \text{ fs}$  pour les moyenne portée, et  $1 \text{ fs}$  pour les interactions courte portée et intramoléculaires. L'algorithme SHAKE est utilisé pour les liaisons impliquant des atomes d'hydrogène. Les paramètres et topologie d'Amber-95 sont utilisés pour les deux protéines.

Nous avons d'abord minimisé la structure du soluté dans le vide par la méthode des gradients conjugués, avec un critère de convergence de  $10^{-2}$ , puis nous avons équilibré les particules de solvant autour du soluté rigide pendant  $24 \text{ ps}$ . Ensuite, pour équilibrer le système en température, nous avons progressivement augmenter la température en plusieurs paliers de  $20$  à  $300 \text{ K}$  sur un total de  $96 \text{ ps}$ . Pendant cette période de chauffage

du système, les vitesses des atomes sont réassignées dès que l'énergie ou la température sort de la fourchette désirée. Au début de chaque étape de 8 ps de l'équilibration, les vitesses des atomes sont attribuées de façon aléatoire selon une distribution de Boltzmann. Enfin, la production des dynamiques des deux protéines a été effectuée pour différents facteurs de charge,  $\lambda_q = 1,185$  et  $\lambda_q = 1,2$ , sur deux nanosecondes pour chaque simulation (166 étapes de 12 ps). Nous avons dénommé les simulations de la façon suivante :

- simulation A1 : simulation de BPTI avec  $\lambda_q = 1,185$  ;
- simulation A2 : BPTI avec  $\lambda_q = 1,2$  ;
- simulation B1 : 1PGB avec  $\lambda_q = 1,185$  ;
- simulation B2 : 1PGB avec  $\lambda_q = 1,2$ .

### Paramètres analysés

Pour chaque simulation, nous avons effectué une analyse structurale, une analyse énergétique et une analyse de l'hydratation des protéines.

#### •Analyse structurale

Notre analyse structurale des protéines va reposer sur l'étude de plusieurs paramètres : l'évolution temporelle du RMS de la protéine par rapport à la structure initiale, le RMS moyen par résidu sur la simulation, le pourcentage d'hélicité et de feuillet  $\beta$  par résidu et le facteur d'ordre  $S^2$ , que nous allons détailler ci-dessous. La définition du RMS a été donnée plus haut. Nous le calculerons ici pour les atomes lourds de la chaîne principale.

Pour le RMS moyen par résidu, nous calculons la moyenne par résidu des fluctuations atomiques pendant la dynamique (ou une partie de la dynamique) des atomes qui le composent. La fluctuation atomique d'un atome  $i$  pendant une dynamique, par rapport à une structure de référence notée avec l'indice 0, est définie par la formule :

$$\delta r(i) = \sqrt{\frac{1}{N_c} \sum_{k=1}^{N_c} |\mathbf{r}_i(k) - \mathbf{r}_i^0|^2}, \quad (4.2)$$

où  $N_c$  est le nombre total de configurations analysées. La moyenne des fluctuations atomiques sur les atomes du résidu donne ce que nous appellerons le RMS moyen par résidu. Comme pour le RMS, on peut calculer le RMS par résidu pour les atomes lourds, ou pour les atomes lourds de la chaîne principale (on ne prend alors pas en compte les fluctuations des atomes des chaînes latérales). On peut calculer ces fluctuations atomiques sur la durée entière de la simulation, ou sur une partie de la simulation pour laquelle le RMS global s'est stabilisé, afin de voir précisément les résidus les plus flexibles d'une structure globale stable. Enfin, on calcule le RMS par résidu par rapport à une structure de référence : par

exemple la structure initiale (la structure cristallographique minimisée dans le vide) ou la structure moyenne de la simulation (ou d'une partie stable de la simulation), et il nous renseigne alors sur la rigidité relative des différents résidus.

Le pourcentage d'hélicité de la protéine est calculé de la même façon que pour le peptide Rn-24, sur les fichiers PDB de la simulation, avec les critères suivants sur les angles  $(\phi, \psi)$  d'un résidu pour considérer qu'il est hélicoïdal :  $(\phi, \psi) \in ([-85^\circ; -25^\circ], [-75^\circ; 0^\circ])$ . Nous ne calculerons ici que le pourcentage d'hélicité moyen par résidu, nous voulons en effet vérifier que les hélices  $\alpha$  des protéines sont conservées par rapport à la structure initiale. De plus, afin d'étudier la stabilité des feuillets  $\beta$  avec notre modèle de solvant, nous calculerons aussi le pourcentage de feuillet  $\beta$  moyen sur la simulation, pour chaque résidu. Le critère choisi sur les angles  $\phi$  et  $\psi$  pour qu'un résidu appartienne à un feuillet  $\beta$  est :  $(\phi, \psi) \in ([-165^\circ; -95^\circ], [90^\circ; 160^\circ])$ .

Enfin, un autre paramètre structural permet d'évaluer la dynamique interne des structures secondaires pendant la simulation, et peut être directement comparé à l'expérience : le facteur d'ordre des liaisons N–H de la chaîne principale. Ce facteur d'ordre, noté  $S^2$ , calculé pour chaque résidu, reflète la mobilité interne angulaire des liaisons amides de la chaîne principale. Il est défini par la fonction suivante [Smith *et al.*, 1995] :

$$S^2 = \frac{1}{T^2} \int_0^T \int_0^T \left[ \frac{3}{2} (\mu(t) \cdot \mu(t + \tau))^2 - \frac{1}{2} \right] d\tau dt, \quad (4.3)$$

où  $\mu(t)$  est le vecteur unitaire porté par la liaison N–H. Il a été montré [Henry et Szabo, 1985] que le facteur d'ordre pouvait être calculé de la façon suivante :

$$S^2 = \frac{3}{2} \sum_{\alpha=1}^3 \sum_{\beta=1}^3 \langle \mu_\alpha \mu_\beta \rangle^2 - \frac{1}{2}, \quad (4.4)$$

où  $\mu_\alpha$  pour  $\alpha = 1, 2$  et  $3$  sont les composantes en  $x, y$  et  $z$  du vecteur unitaire porté par la liaison N–H du résidu étudié, et  $\langle \rangle$  représente une moyenne sur les configurations de la molécule pendant la simulation. Comme le calcul du facteur d'ordre n'est pas implémenté dans ORAC, nous le calculerons à partir des fichiers PDB que nous écrivons toutes les 12 ps (166 fichiers pour 2 ns). Le nombre de configurations utilisées pour le calcul de  $S^2$  est donc assez faible, et les résultats que nous obtiendrons devront être interprétés avec précaution. Afin de ne prendre en compte que la mobilité interne de la protéine, toutes les structures d'une même dynamique sont superposées par une méthode des moindres carrés sur une structure de référence. Contrairement au RMS par résidu qui sera plus élevé pour les résidus les plus mobiles, le facteur d'ordre sera plus élevé pour les résidus les plus rigides.

### •Analyse énergétique

Pour l'analyse énergétique, nous étudierons d'abord l'évolution temporelle de certaines composantes énergétiques importantes du système : l'énergie totale, l'énergie potentielle du soluté, et l'énergie libre électrostatique de solvation, afin de les comparer entre les différentes simulations d'une même molécule, et de vérifier si elles suivent les évolutions conformationnelles de la molécule.

De la même façon que pour les peptides, nous étudierons les droites de corrélation entre les énergies électrostatiques de solvation moyennes avec le modèle PPP et les énergies de solvation calculées par résolution de l'équation de Poisson-Boltzmann avec DelPhi sur les structures moyennes. Vu la taille importante des molécules étudiées, les calculs électrostatiques avec DelPhi sont assez coûteux, et nous avons donc fait l'étude des corrélations sur seulement une structure sur deux issue des dynamiques (83 structures au lieu de 166). Les paramètres de DelPhi sont les mêmes que pour les peptides étudiés dans la première partie de ce chapitre.

### •Analyse de l'hydratation

Enfin, nous nous sommes intéressés à l'hydratation de ces deux protéines par nos pseudo-particules de solvant. En effet, l'un des principaux intérêts de notre modèle de solvant est de tenir compte de la nature moléculaire de l'eau et nous pouvons directement étudier les sites préférentiels d'hydratation des protéines étudiées. Pour cela, nous avons procédé de la manière suivante : nous avons regardé pour chaque résidu les particules de solvant qui étaient dans la première couche d'hydratation<sup>3</sup> de certains groupements donneurs et accepteurs de liaisons hydrogène :

- à moins de 4 Å de l'atome d'oxygène du groupement carboxyle de la chaîne principale,
- à moins de 3 Å de l'atome d'hydrogène du groupement amide de la chaîne principale (car le rayon de l'atome d'hydrogène est plus petit que celui des autres atomes et la première couche d'hydratation est donc plus proche du centre de l'atome),
- à moins de 3 ou 4 Å des atomes polaires (N, O ou H) de la chaîne latérale des acides aminés polaires.

Nous avons identifié ces particules de solvant dans les 166 fichiers PDB, et nous avons calculé pour chaque résidu le pourcentage du temps total de la dynamique pendant lequel est restée la particule de solvant qui est restée le plus longtemps proche de ces atomes. Nous pourrions ainsi visualiser les résidus les plus hydratés, et nous pourrions ensuite comparer ces résultats avec les molécules d'eau identifiées expérimentalement, soit dans la structure cristallographique, soit dans les études de RMN.

---

<sup>3</sup>la taille de la première couche d'hydratation est définie grâce la fonction de distribution radiale (FDR) présentée au chapitre 3 : c'est la valeur de la distance au premier minimum de la FDR.

### 4.2.3 Résultats

#### Analyse structurale

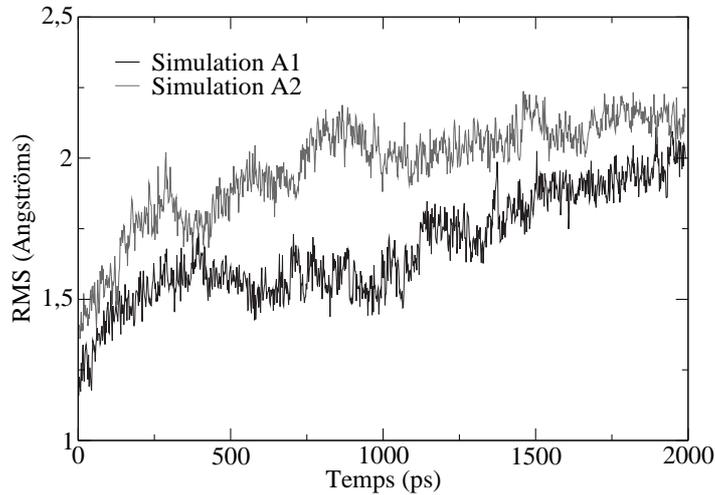


FIG. 4.12 : Évolution du RMS des atomes de la chaîne principale de BPTI par rapport à la structure initiale pendant les deux simulations : simulation A1 avec un facteur de charge de  $\lambda_q = 1,185$  (en noir) et simulation A2 avec  $\lambda_q = 1,2$  (en gris).

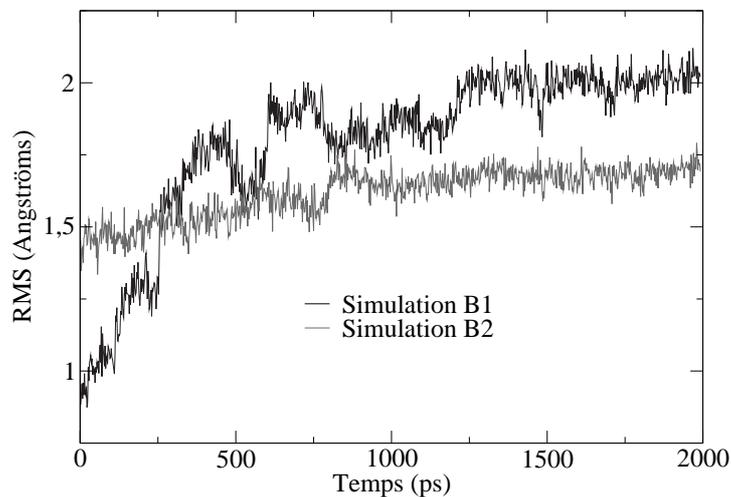


FIG. 4.13 : Évolution du RMS des atomes de la chaîne principale de 1PGB par rapport à la structure initiale pendant les deux simulations : simulation B1 avec un facteur de charge de  $\lambda_q = 1,185$  (en noir) et simulation B2 avec  $\lambda_q = 1,2$  (en gris).

Les évolutions temporelles du RMS des atomes lourds de la chaîne principale des deux protéines étudiées par rapport à leur structure initiale, pour les différents facteurs de charge envisagés, sont présentées sur la figure 4.12 pour BPTI, et sur la figure 4.13

pour 1PGB. Pour les simulations A1, A2 et B1, le RMS augmente fortement pendant les premières 250 ps de dynamique, et augmente plus légèrement pour la simulation B2. Cela montre que le système, malgré la phase d'équilibration, n'est pas encore complètement stabilisé au début de la dynamique. Pour BPTI, le RMS de la simulation A2 reste toujours au-dessus de celui de la simulation A1. Mais il semble se stabiliser après 800 ps de dynamique autour d'une valeur d'environ 2,1 Å, alors que celui de la simulation A1, qui semblait s'être stabilisé entre 500 ps et 1 ns de dynamique vers 1,6 Å, continue d'augmenter pendant la dernière nanoseconde et tend vers 2 Å. Pour 1PGB, les RMS des deux simulations se stabilisent après environ 700 ps de dynamique. Le RMS de la simulation B2, qui est au départ plus élevé que celui de la simulation B1, passe en-dessous au bout d'environ 300 ps, et est beaucoup plus stable tout au long de la simulation. Il se stabilise autour d'une valeur moyenne de 1,7 Å, alors que celui de la simulation B1 se stabilise autour d'une valeur de 2,0 Å. Globalement, à part pour la simulation A1, le squelette des protéines semble se stabiliser autour d'une conformation stable dans la première nanoseconde de simulation.

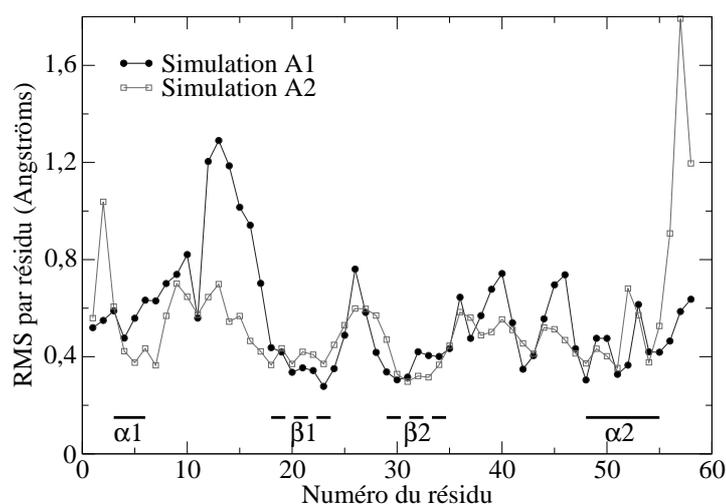


FIG. 4.14 : RMS par résidu des atomes lourds de la chaîne principale de BPTI par rapport à la structure moyenne sur 1,75 ns des simulations A1 et A2. On a représenté les structures secondaires de la structure initiale.

Nous avons calculé le RMS moyen par résidu pour les atomes lourds de la chaîne principale sur 1,75 ns de dynamique (car nous avons enlevé les premières 250 ps des productions pendant lesquelles les structures n'étaient pas stables), par rapport à la structure moyenne calculée sur la même partie de la dynamique. Ce RMS moyen par résidu, qui nous renseigne donc sur la flexibilité ou la rigidité des résidus, est présenté sur la figure 4.14 pour les deux simulations de BPTI, et sur la figure 4.15 pour les deux simulations de 1PGB. Dans les deux cas, on observe que le RMS par résidu est minimum pour les résidus

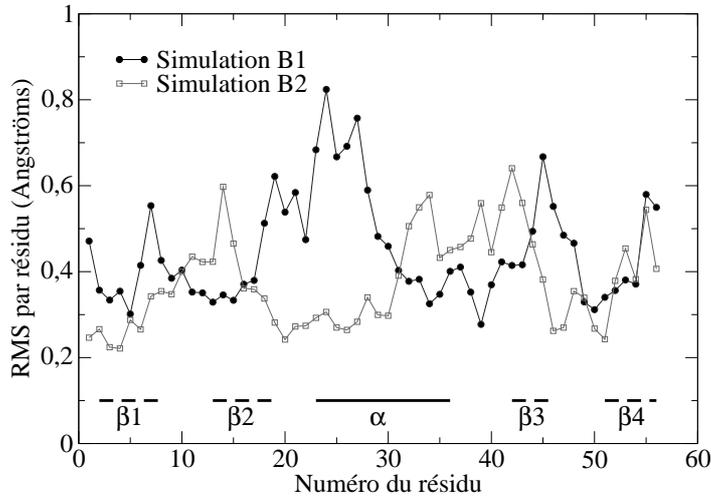


FIG. 4.15 : RMS par résidu des atomes de la chaîne principale de 1PGB par rapport à la structure moyenne sur 1,75 ns des simulations B1 et B2.

impliqués dans des structures secondaires, indiquant comme attendu que la chaîne principale est plus flexible au niveau des boucles qu'au niveau des structures secondaires. Pour 1PGB, le RMS par résidu est plus faible que pour BPTI, comme le RMS global, montrant que la protéine 1PGB est plus rigide que BPTI en solution. Cela s'explique assez bien par le fait que plus de résidus de 1PGB sont impliqués dans des structures secondaires que dans BPTI. La simulation B2 semble en moyenne un peu plus rigide que la simulation B1, surtout au niveau des boucles qui paraissent plus flexibles pendant la simulation B1, en particulier au début de l'hélice  $\alpha$ .

Nous devons compléter ces remarques sur la flexibilité des structures secondaires par l'étude du pourcentage d'hélicité et de feuillet  $\beta$  par résidu pour chaque simulation, sur les dernières 1,75 ns. La figure 4.16 présente les résultats pour les simulations de BPTI. Pour les deux simulations, on voit que l'hélice  $\alpha_2$  (résidus 48–55) et le feuillet  $\beta_1$  (résidus 18–24) sont assez bien conservés pendant la dynamique. Le feuillet  $\beta_2$  (résidus 29–35) et l'hélice  $\alpha_1$  (résidus 3–6) paraissent moins stables, ils apparaissent plus courts pendant la dynamique que dans la structure initiale. On voit apparaître pour les deux simulations A1 et A2 une courte hélice dans la boucle entre  $\alpha_1$  et  $\beta_1$ , au niveau des résidus 10–16, ainsi qu'entre les deux feuillets  $\beta_1$  et  $\beta_2$ , au niveau des résidus 25–27. Cette dernière observation, ainsi que la stabilité relative des structures secondaires  $\alpha_1$  et  $\beta_2$ , confirment des résultats de RMN [Pan *et al.*, 1995, Carulla *et al.*, 2000] et de dynamique moléculaire [Schneider *et al.*, 1999] observés sur cette même protéine. Les résultats analogues du pourcentage d'hélicité pour 1PGB sont présentés sur la figure 4.17. On retrouve assez clairement pour les deux simulations les structures secondaires de départ, en particulier l'hélice  $\alpha$  (résidus 23–36) qui semble extrêmement stable en solution. La simulation B2 conserve un peu moins

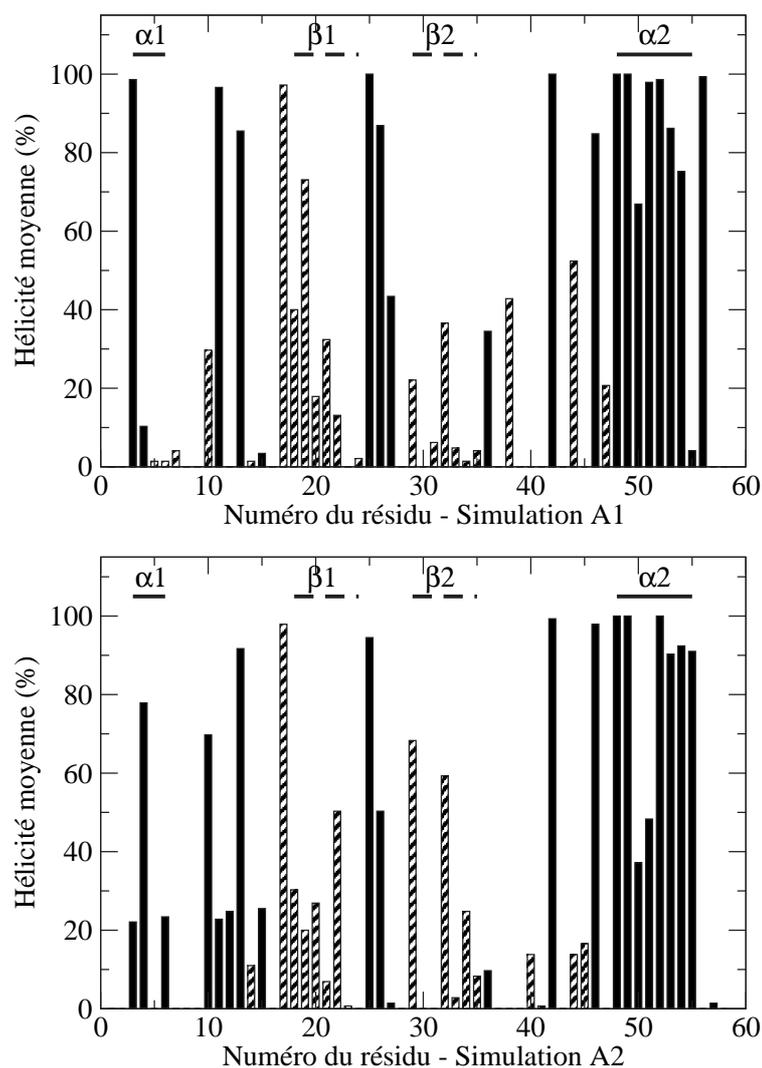


FIG. 4.16 : Pourcentage d'hélicité et de feuillet  $\beta$  par résidu pour les deux simulations A1 et A2 avec BPTI : les hélices  $\alpha$  sont en noir et les feuillets  $\beta$  en hachuré.

bien cette hélice (au niveau des résidus 29 et 30). Là encore, on observe l'apparition d'une courte hélice au niveau de la boucle entre les feuillets  $\beta_3$  et  $\beta_4$  (résidus 47–49). Cette observation est aussi en accord avec différentes études de RMN [Frank *et al.*, 1995, Kobayashi *et al.*, 2000] et théoriques [Sheinerman et Brooks, 1997] qui avaient souligné que la boucle au niveau des résidus 46–51 adoptait une conformation particulièrement rigide, due aux interactions entre les acides aminés Asp-46, Ala-48, Thr-49 et Thr-51.

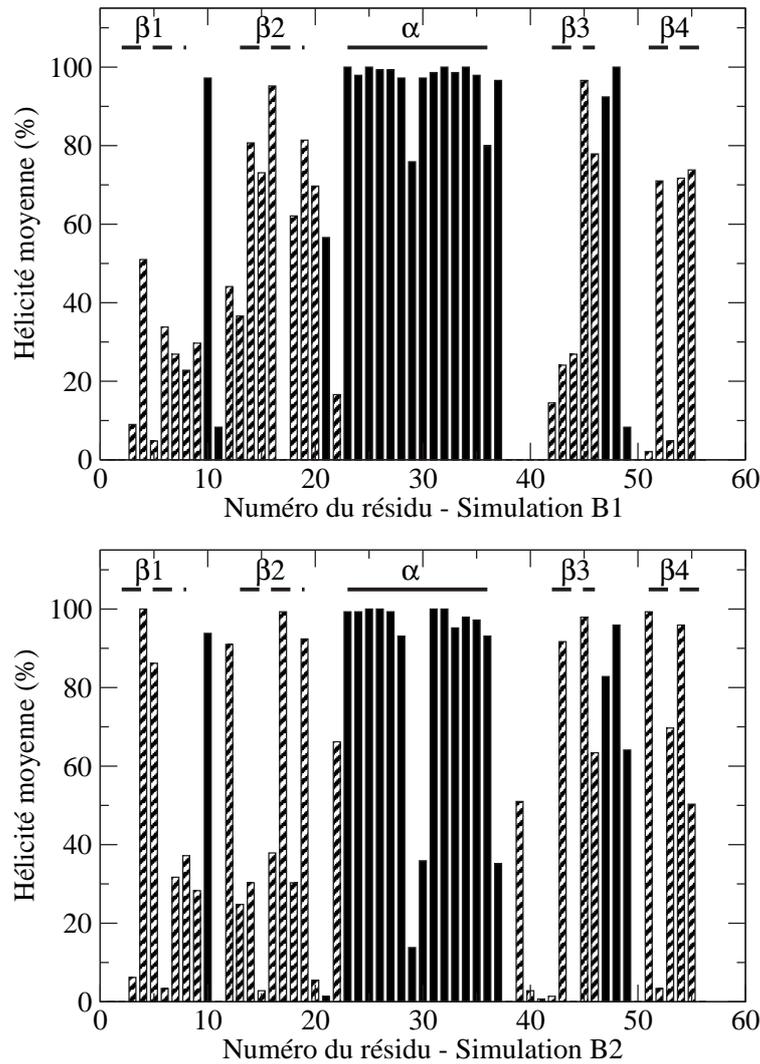


FIG. 4.17 : Pourcentage d'hélicité et de feuillet  $\beta$  par résidu pour les deux simulations B1 et B2 avec 1PGB : les hélices  $\alpha$  sont en noir et les feuillets  $\beta$  en hachuré.

La flexibilité relative des structures secondaires des protéines peut être étudiée en calculant le facteur d'ordre des liaisons N-H de la chaîne principale. On a tracé ce facteur  $S^2$  pour les deux simulations de BPTI sur la figure 4.18, et pour les deux simulations de 1PGB sur la figure 4.19, toujours sur les dernières 1,75 ns de la dynamique. On observe comme attendu que le facteur d'ordre est plus élevé dans les régions où il y a des structures secondaires, confirmant que les régions en boucle sont les plus flexibles. Si la forme des courbes de  $S^2$  est globalement en accord avec les résultats expérimentaux et théoriques pour BPTI [Smith *et al.*, 1995, Schneider *et al.*, 1999] et 1PGB [Sheinerman et Brooks, 1997, Calimet *et al.*, 2001], il faut noter que les valeurs obtenues avec notre modèle sont plus élevées que les valeurs des autres études, qui sont plutôt en-dessous de 0,8 pour BPTI [Smith *et al.*, 1995] et autour de 0,8 en moyenne pour 1PGB [Sheinerman et Brooks, 1997].

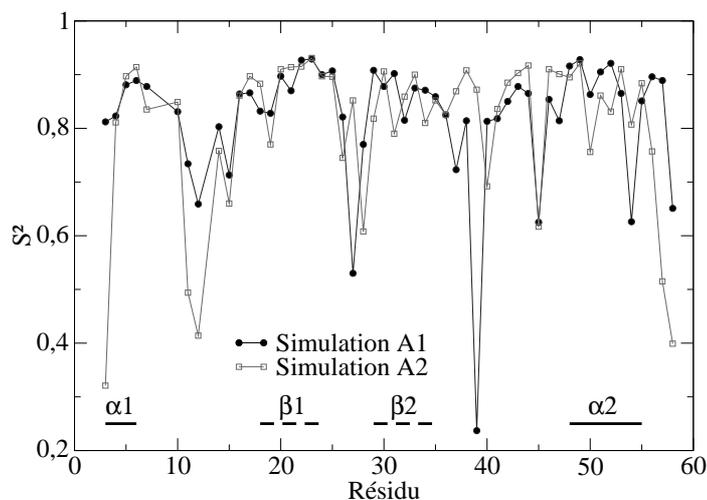


FIG. 4.18 : Facteur d'ordre  $S^2$  sur les liaisons N-H des résidus de la chaîne principale de BPTI, sur 1,75 ns des deux simulations A1 et A2.

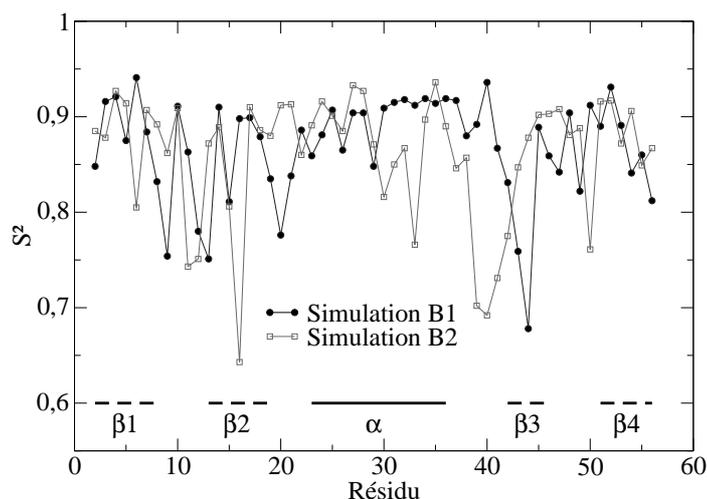


FIG. 4.19 : Facteur d'ordre  $S^2$  par résidu de 1PGB, sur 1,75 ns des deux simulations B1 et B2.

On observe également que les boucles les plus flexibles dans nos simulations ne sont pas les mêmes que dans d'autres simulations. Ces différences peuvent provenir de la méthode utilisée pour calculer le facteur d'ordre. En effet, nous calculons le facteur d'ordre sur des configurations prélevées toutes les 12 ps de la trajectoire, et sur une simulation de seulement 2 ns, ce qui représente un échantillonnage assez faible. Nous ne pouvons donc pas comparer de façon stricte nos résultats à ceux des autres études et en particulier celles expérimentales. Nous pouvons cependant comparer les simulations entre elles. Pour BPTI, les résultats obtenus sur les deux simulations sont comparables au niveau des structures secondaires, mais différent au niveau des boucles. On remarque pour 1PGB, l'hélice  $\alpha$  de

la simulation B2 semble plus flexible que celle de la simulation B1, confirmant les résultats sur l'hélicité moyenne par résidu. On remarque aussi que la boucle entre  $\beta_3$  et  $\beta_4$  est plus rigide que les autres boucles, confirmant encore les résultats sur l'hélicité. Globalement, nos simulations reproduisent assez bien la relative rigidité des structures secondaires, mais la flexibilité relative des boucles n'est pas tout à fait claire.

Les simulations sont différentes pour les deux facteurs de charge au niveau de la stabilité de certaines structures, mais les résultats obtenus sont assez mitigés et ne nous permettent pas vraiment de trancher entre les deux valeurs de  $\lambda_q = 1,185$  et  $1,2$ . En moyenne, les structures obtenues avec  $\lambda_q = 1,2$  (simulations A2 et B2) paraissent légèrement moins flexibles qu'avec  $\lambda_q = 1,185$  (A1 et B1). Pourtant, les structures secondaires paraissent légèrement moins bien conservées avec  $\lambda_q = 1,2$ , et globalement le facteur  $\lambda_q = 1,185$  produit des résultats un peu plus satisfaisants.

### Analyse énergétique

Les évolutions temporelles des énergies totale, potentielle du soluté, et électrostatique de solvation des différentes simulations sont représentées sur la figure 4.20 pour BPTI et sur la figure 4.21 pour 1PGB. Pour toutes les simulations, on observe une légère dérive de l'énergie totale. Nous avons observé qu'en diminuant le pas d'intégration maximum, nous diminuons l'importance de cette dérive (les résultats ne sont pas montrés ici), et on sait aussi que l'algorithme de pas d'intégration multiple avec ORAC ne conserve pas strictement l'énergie totale du système [Procacci *et al.*, 1997]. Enfin, comme on a représenté toutes les énergies du système à la même échelle, on voit que cette dérive est négligeable par rapport aux fluctuations des énergies potentielles, et négligeable aussi par rapport à la valeur absolue de l'énergie totale (en moyenne 0,5 %). On remarque aussi que les fluctuations de l'énergie totale sont très faibles par rapport aux fluctuations des énergies potentielles, indiquant que les simulations de dynamique moléculaire avec notre modèle de solvant sont globalement conservatives. Pour BPTI, les courbes des énergies des simulations A1 et A2 évoluent de façon très similaire. On remarque que la compensation des variations de l'énergie potentielle du soluté par l'énergie électrostatique de solvation est moins visible que pour l'étude des petits peptides, même si les allures des deux courbes sont assez symétriques. Pour 1PGB, les courbes des deux simulations B1 et B2 sont assez différentes, et on remarque une nouvelle fois qu'une transition dans l'énergie potentielle du soluté de la simulation B2 est accompagnée d'une transition inverse dans l'énergie électrostatique de solvation.

Les droites de corrélation entre les énergies libres électrostatiques de solvation calculées avec notre modèle PPP et celles calculées avec DelPhi sur les structures moyennes

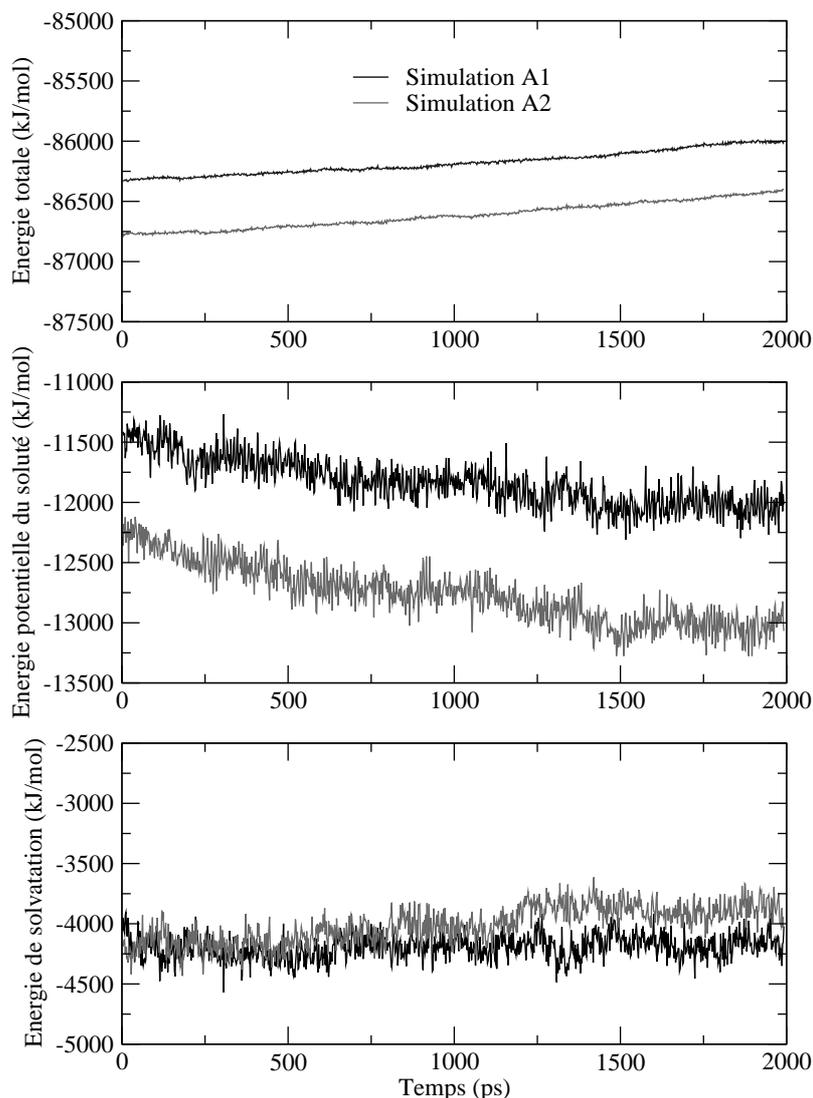


FIG. 4.20 : Évolution des énergies totale, potentielle du soluté et de solvation pendant les deux simulations de BPTI : simulation A1 avec un facteur de charge de  $\lambda_q = 1,185$  (en noir) et simulation A2 avec  $\lambda_q = 1,2$  (en gris).

sont présentées sur la figure 4.22. Le tableau 4.3 présente les valeurs des coefficients de corrélation, des pentes des droites et des erreurs relatives moyennes. Nous avons aussi représenté sur la figure 4.23 les évolutions comparées de ces deux énergies libres de solvation. On voit qu'avec le facteur de charge de  $\lambda_q = 1,2$ , nous avons les meilleurs coefficients de corrélation (autour de 0,9). Cependant, pour ces deux simulations A2 et B2, les pentes des droites sont très supérieures à un. Pour la simulation B1, on remarque que la pente est très proche de un, alors que pour la simulation A1, elle est en-dessous de un. Cependant, nous avons vu sur la figure 4.12 montrant l'évolution du RMS pour BPTI que la structure issue de la dynamique A1 n'est pas encore bien stabilisée après 2 ns, et le

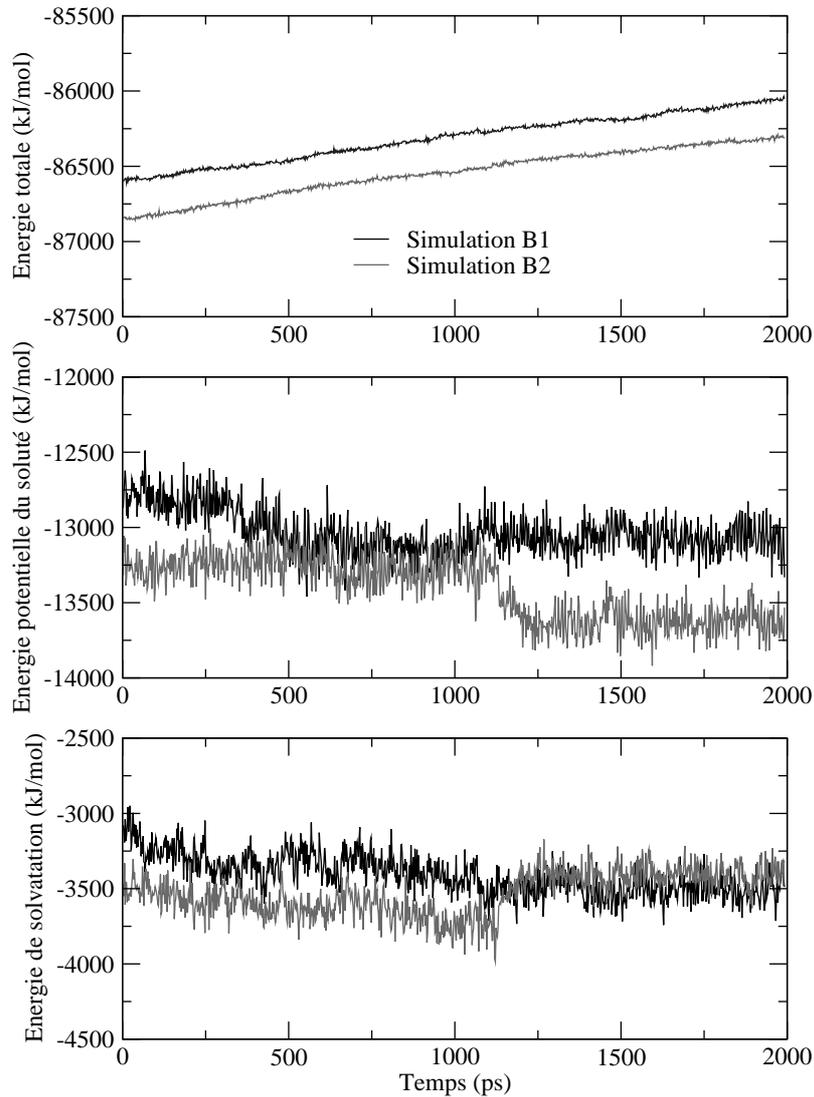


FIG. 4.21 : Évolution des énergies totale, potentielle du soluté et de solvation pendant les deux simulations de 1PGB : simulation B1 avec un facteur de charge de  $\lambda_q = 1,185$  (en noir) et simulation B2 avec  $\lambda_q = 1,2$  (en gris).

solvant autour de la protéine n'est donc peut-être pas bien équilibré. Les erreurs relatives moyennes sont de l'ordre de 10 % pour BPTI, et de 2 % pour 1PGB, quel que soit le facteur de charge. Nous avons vu au chapitre 3 avec les tests sur les petites molécules que notre modèle donne de moins bons résultats sur l'énergie libre électrostatique de solvation des molécules chargées. Or, la protéine BPTI est globalement chargée  $+6e$  alors que 1PGB est chargée  $-4e$ . On comprend ainsi que les résultats soient un peu meilleurs pour la molécule la moins chargée. On remarque sur les courbes d'évolution des énergies libres de solvation de la figure 4.23 que l'énergie de solvation moyenne calculée avec le solvant PPP suit globalement les variations de l'énergie libre de solvation des structures

moyennes calculées avec DelPhi, sauf pour les premières 250 ps environ, montrant qu'il est important de bien équilibrer les degrés de translation des particules de solvant autour du soluté pour obtenir une bonne énergie libre de solvatation.

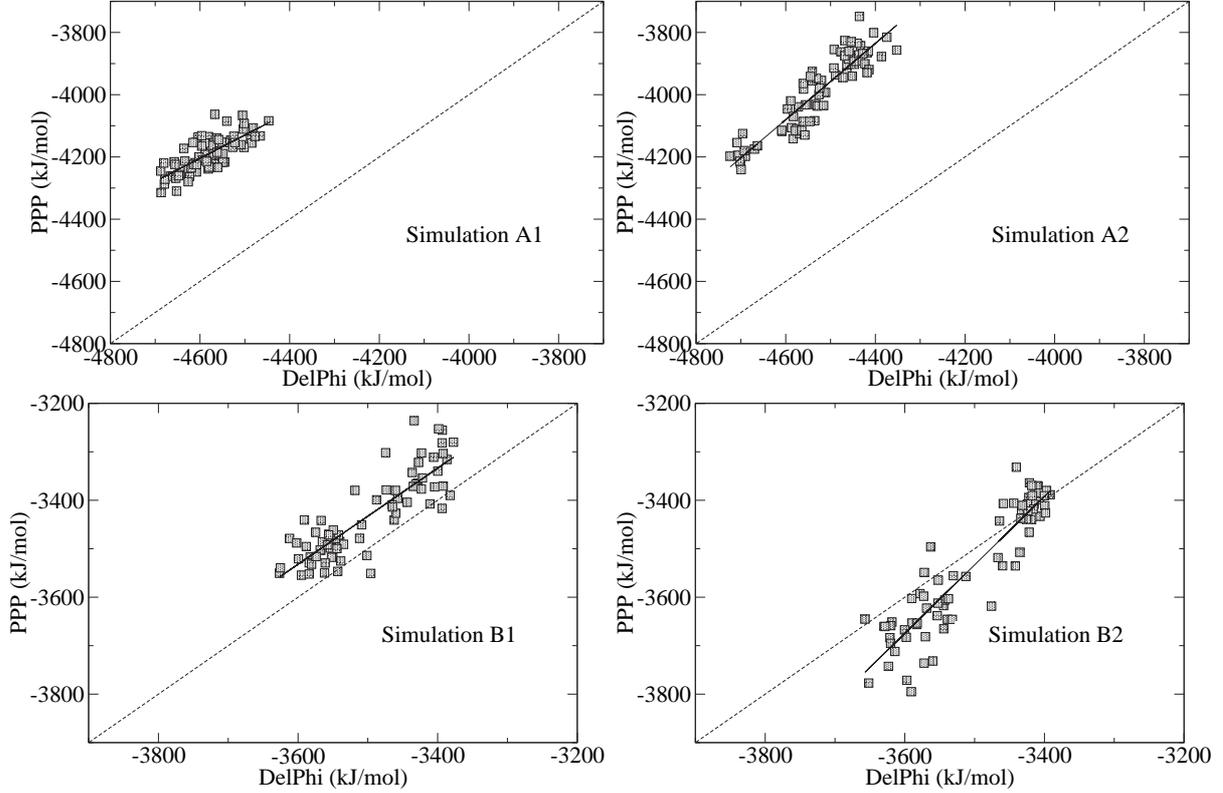


FIG. 4.22 : Corrélations entre les énergies électrostatiques de solvatation moyennes avec PPP et les énergies des structures moyennes avec DelPhi, pour les deux simulations A1 et A2 sur BPTI et B1 et B2 sur 1PGB.

Sim.	Protéine	$\lambda_q$	Coef. de corrélation	Pente	Erreur relative
A1	BPTI	1,185	0,756	0,743	8,54 %
A2	BPTI	1,2	0,916	1,225	11,86 %
B1	1PGB	1,185	0,856	0,991	2,00 %
B2	1PGB	1,2	0,915	1,411	1,55 %

TAB. 4.3 : Coefficients de corrélation entre les énergies de solvatation moyennes avec le modèle PPP et les énergies des structures moyennes avec DelPhi, pente de la droite de corrélation et erreur relative moyenne pour les simulations des protéines BPTI et 1PGB.

D'un point de vue énergétique, comme pour les simulations des peptides, on remarque donc que notre modèle ne donne pas une valeur précise de l'énergie libre de

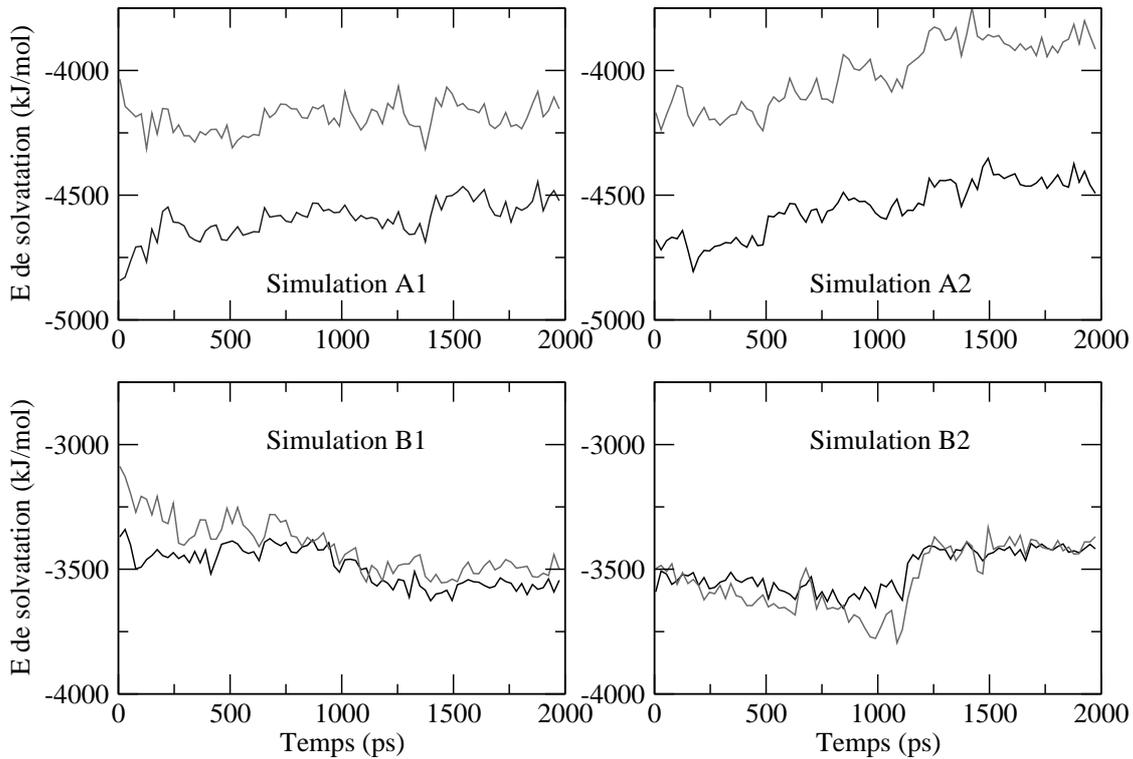


FIG. 4.23 : Évolutions des énergies libres électrostatiques de solvation des structures moyennes avec DelPhi (en noir) et des énergies de solvation moyennes avec PPP (en gris), pour les deux simulations A1 et A2 sur BPTI et B1 et B2 sur 1PGB.

solvation à chaque instant. Cependant, là encore, les évolutions de l'énergie libre de solvation calculée avec notre modèle suivent, parfois avec un peu de retard, celles calculées avec DelPhi. Encore une fois, le facteur de charge  $\lambda_q = 1,185$  semble être un meilleur facteur que  $\lambda_q = 1,2$ , car les pentes des droites de corrélation obtenues avec  $\lambda_q = 1,185$  sont plus proches de un.

### Analyse de l'hydratation

Nous avons étudié les sites d'hydratation préférentiels des deux protéines pour les différentes simulations par la méthode exposée plus haut. L'intérêt de notre modèle semi-explicite est que nous pouvons identifier des molécules de solvant aussi simplement qu'en solvant explicite. Pour chaque simulation, nous avons tracé donc le pourcentage maximum de temps pendant lequel une molécule d'eau est restée proche des groupements NH et CO des chaînes principales, et des chaînes latérales des résidus polaires. Nous avons fait cette étude sur les dernières 1,75 ns des dynamiques. Comme nous l'avons vu au chapitre 2, les études expérimentales mettant en évidence les molécules d'eau fortement

liées aux protéines ont montré que les molécules d'eau intervenaient dans des réseaux de liaisons hydrogène avec les groupes d'atomes donneurs et accepteurs de la protéine. Celles qui sont emprisonnées dans les cavités à l'intérieur de la molécule (les molécules d'eau internes) complètent les liaisons hydrogène des groupements donneurs et accepteurs de la protéine laissés libres en formant jusqu'à quatre liaisons hydrogène, et elles stabilisent ainsi fortement la structure tridimensionnelle repliée de la protéine.

Les différentes structures cristallographiques de la protéine BPTI contiennent chacune une soixantaine de molécules d'eau localisées, mais qui ne correspondent pas toujours entre les différentes structures. Les études de RMN ont confirmé que seules quatre molécules d'eau internes, et une partiellement exposée au solvant, étaient aussi présentes en solution [Wlodawer *et al.*, 1987, Tüchsen *et al.*, 1987, Otting et Wüthrich, 1989, Denisov et Halle, 1995]. Les quatre molécules d'eau internes forment des liaisons hydrogène avec les groupements de la chaîne principale suivants : Pro-8 (CO), Tyr-10 (CO), Tyr-10 (NH), Thr-11 (CO), Cys-14 (NH), Cys-38 (CO), Cys-38 (NH), Lys-41 (NH) et Asn-43 (CO), ainsi qu'avec les groupements des chaînes latérales de Glu-7 et Asn-44. La molécule d'eau partiellement exposée au solvant (bulk) forme des liaisons hydrogène avec les résidus Tyr-23 (CO) et Ala-25 (NH).

La figure 4.24 représente les résultats obtenus sur l'étude de l'hydratation avec notre modèle pour les deux simulations de BPTI. On retrouve plus ou moins bien parmi les sites fortement hydratés ceux observés expérimentalement. En effet, les sites expérimentaux correspondent dans nos simulations à des pourcentages d'occupation souvent au-dessus de 50 % (sur des dynamiques de 1,75 ns au total). Néanmoins, on ne retrouve pas dans la simulation A2 l'hydratation du groupement amide de Cys-38. Nous trouvons plusieurs sites d'hydratation supplémentaires par rapport aux résultats expérimentaux, indiquant les sites où le potentiel électrostatique est favorable à la présence d'une molécule d'eau mais qui ne correspondent pas forcément à un site préférentiel car nous ne tenons pas compte des liaisons hydrogène qui sont géométriquement contraignantes.

Les études expérimentales par RMN de l'hydratation de 1PGB ont mis en évidence principalement deux molécules d'eau liées à la chaîne principale de la protéine [Clare et Gronenborn, 1992]. Ce ne sont pas des molécules d'eau internes comme dans le cas de BPTI. Ces molécules d'eau participent à deux liaisons hydrogène « bifurquées » : elles acceptent une liaison hydrogène d'un groupement NH d'un résidu, et donnent une liaison hydrogène à un groupement CO d'un autre résidu, qui forme lui-même déjà une liaison hydrogène avec le groupement NH du premier résidu. La première molécule d'eau se trouve entre les résidus Met-1 (CO) et Ala-20 (NH), et la seconde entre Val-29 (CO) et Tyr-33 (NH). La seconde molécule d'eau se trouve au niveau de l'hélice  $\alpha$  de 1PGB et elle est

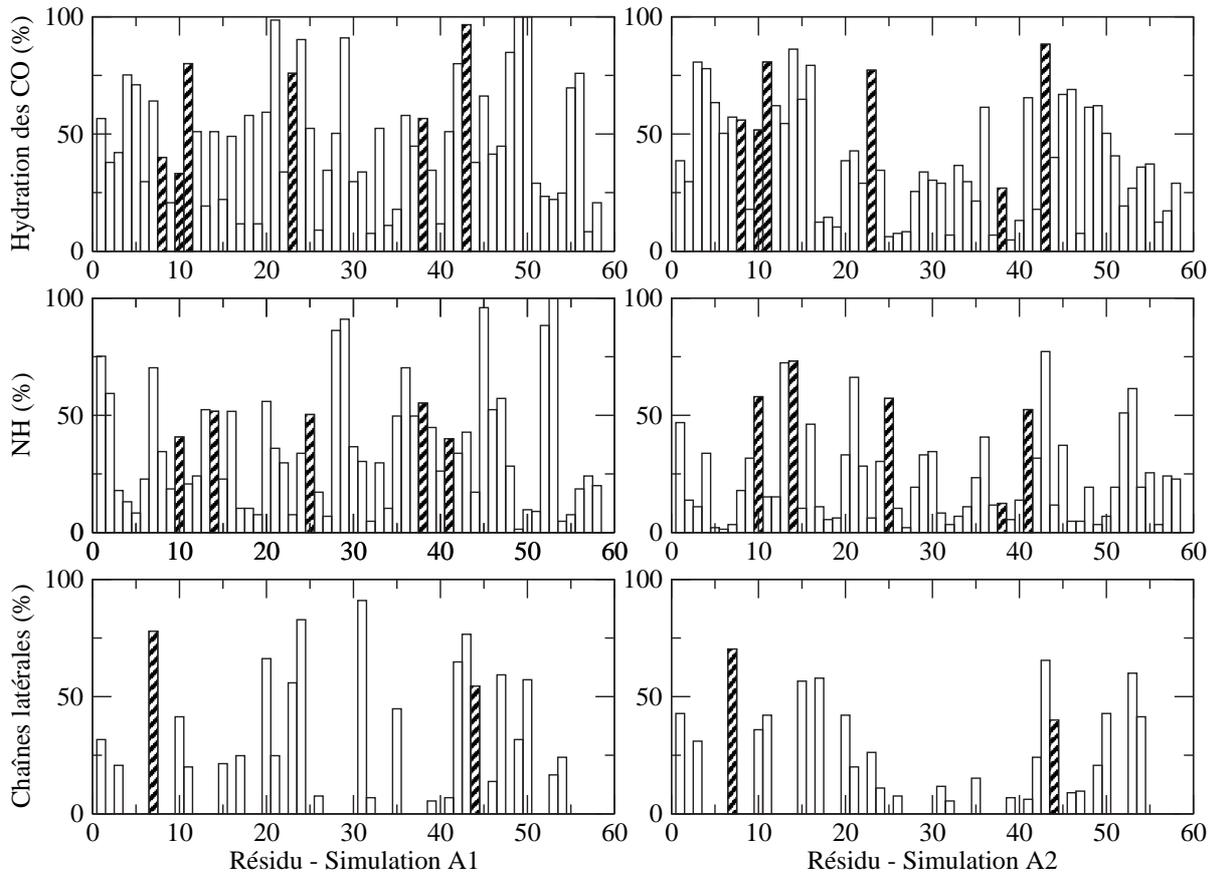


FIG. 4.24 : Représentation du pourcentage maximum d'hydratation des résidus de BPTI pour les simulations A1 et A2, pour les groupements CO de NH de la chaîne principale et les groupements polaires des chaînes latérales. Les molécules d'eau identifiées expérimentalement sont mises en évidence par des hachures.

très accessible au solvant. Notons que cette deuxième molécule d'eau n'a pas été observée dans la simulation de 1PGB en solvant explicite [Sheinerman et Brooks, 1997]. Une autre étude de RMN plus récente [Khare *et al.*, 1999] a identifié une troisième molécule d'eau entre les groupements amides NH des résidus Thr-11 et Leu-12 et le carboxylate de la chaîne latérale de Glu-56.

La figure 4.25 représente les sites d'hydratation obtenus avec notre modèle sur les deux simulations de 1PGB. Parmi tous les sites fortement hydratés, on observe bien que les groupements carboxyles de Met-1 et Val-29 sont relativement bien hydratés, mais pas les groupements amides de Ala-20 et Tyr-33, comme la simulation en solvant explicite l'a aussi constaté [Sheinerman et Brooks, 1997]. Par contre, notre modèle permet d'identifier la présence d'une autre molécule d'eau dans l'hélice  $\alpha$ , entre les liaisons hydrogène reliant les résidus Lys-28 (CO) et Gln-32 (NH) pour la simulation B1, et entre Glu-27 (CO) et Lys-

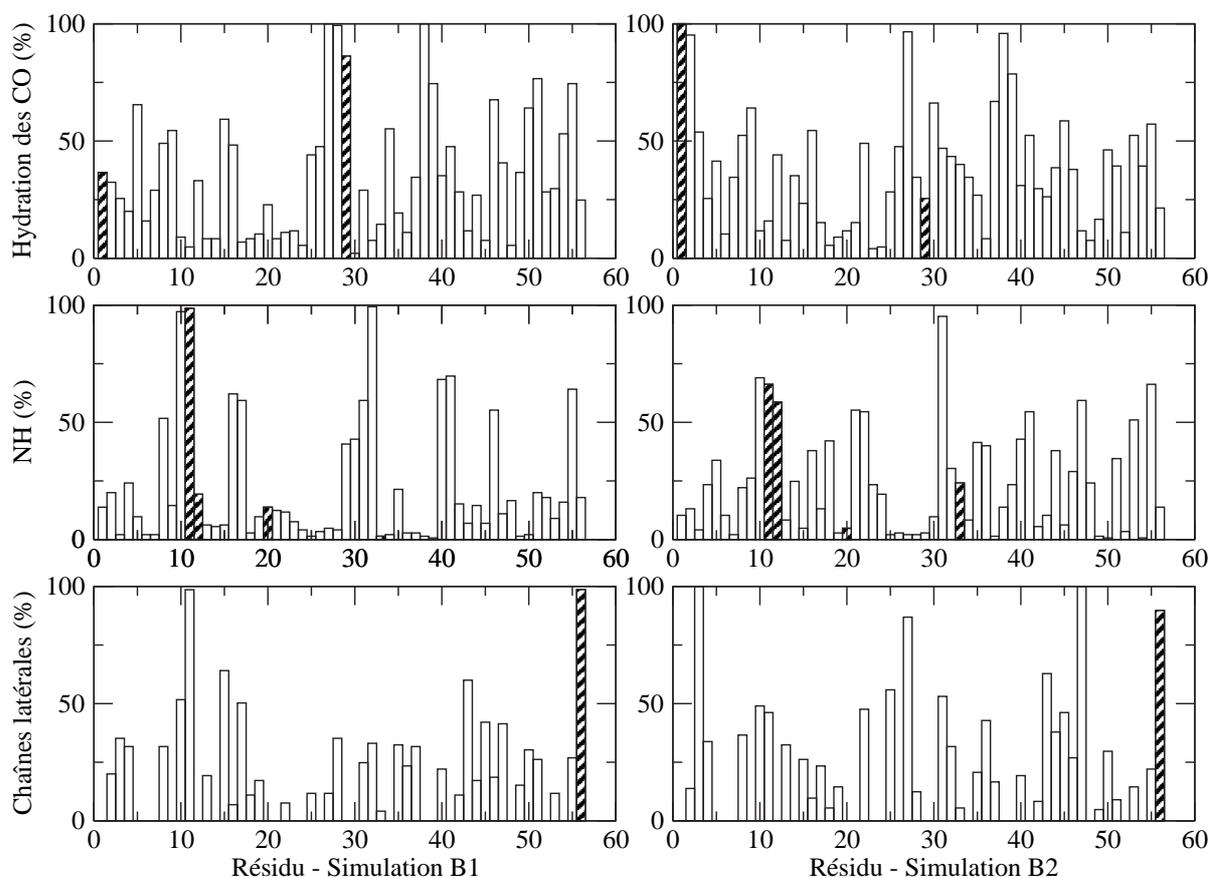


FIG. 4.25 : Représentation du pourcentage maximum d'hydratation des résidus de 1PGB pour les simulations B1 et B2, pour les groupements CO de NH de la chaîne principale et les groupements polaires des chaînes latérales. Les molécules d'eau identifiées expérimentalement sont mises en évidence par des hachures.

31 (NH) pour la simulation B2. Cette molécule d'eau serait donc un ou deux résidus plus haut dans l'hélice que celle observée par RMN, confirmant que cette région de l'hélice est préférentiellement hydratée. Enfin, on retrouve assez bien avec notre modèle la troisième molécule d'eau identifiée expérimentalement, surtout pour la chaîne latérale de Glu-56 qui apparaît très fortement liée à une même pseudo-particule de solvant. Nous pouvons expliquer les différences observées entre les sites d'hydratation identifiés avec notre modèle de solvant et ceux observés expérimentalement par le fait que certaines molécules d'eau sont liées par des réseaux de liaisons hydrogène complexes que seul un modèle de solvant explicite triatomique pourrait reproduire.

Les résultats de l'étude de l'hydratation avec le modèle PPP révèlent donc des ressemblances intéressantes avec les résultats expérimentaux, mais aussi quelques différences, quelque soit le facteur de charge. Notre modèle ne peut reproduire les liaisons hydrogène

de l'eau comme un modèle explicite, et on comprend que l'on ne puisse pas retrouver exactement les résultats expérimentaux. Cependant, nous arrivons à retrouver globalement la plupart des sites d'hydratation expérimentaux, ou des sites voisins, suggérant que les phénomènes d'hydratation des protéines seraient d'abord gouvernés par les interactions électrostatiques, et dans une moindre mesure par les liaisons hydrogène.

#### 4.2.4 Conclusion sur l'étude des protéines

L'étude de la dynamique des deux protéines effectuée dans cette partie a montré que le modèle PPP conservait assez bien les éléments de structures secondaires des molécules sur une dynamique de deux nanosecondes, que l'énergie libre électrostatique de solvation ainsi calculée suivait globalement l'énergie de solvation calculée par des méthodes implicites, et que nous arrivions aussi à reproduire certaines propriétés d'hydratation observées expérimentalement ou en solvant explicite. Ces résultats sont très encourageants, et répondent bien aux objectifs que nous nous étions fixés de développer un modèle similaire à la fois aux méthodes implicites pour les énergies électrostatiques et aux méthodes explicites pour les résultats structuraux ou d'hydratation [Basdevant *et al.*, 2003a].

Apparemment, le facteur de charge  $\lambda_q = 1,185$  semble être le plus adapté pour l'étude des protéines et des peptides. En effet, c'est pour cette valeur que nous arrivons à reproduire le mieux les structures secondaires, ainsi que les énergies libres électrostatiques de solvation qui suivent le mieux les évolutions de l'énergie de solvation calculée par les méthodes de Poisson-Boltzmann. Il semblerait qu'après une déformation structurale importante de la molécule, le solvant PPP mette un peu de temps à s'équilibrer autour du soluté, et il faut bien attendre la stabilisation du solvant avant d'extraire l'énergie libre de solvation. Les résultats encourageants obtenus sur l'hydratation nous incitent à continuer notre étude sur des molécules fortement chargées, et pour lesquelles l'hydratation joue un rôle primordial, telles que les acides nucléiques.

# Chapitre 5

## Application aux acides nucléiques



# Application aux acides nucléiques

Nous nous intéresserons dans ce chapitre à la dynamique moléculaire des acides nucléiques avec le modèle de solvant des pseudo-particules polarisables (PPP). Les acides nucléiques sont des systèmes très chargés qui interagissent fortement avec l'eau et les ions qui les entourent, et donc pour lesquels les interactions électrostatiques jouent un rôle majeur dans la stabilisation de leurs structures tridimensionnelles. La représentation du solvant dans les simulations d'acides nucléiques est donc un problème particulièrement important.

Nous étudierons d'abord une boucle d'ARN de transfert (ARNt), connue pour être sensible à la représentation du solvant, et dont l'étude de la stabilité pendant la dynamique avec le solvant PPP constituera donc un bon test. Nous étudierons ensuite les dynamiques de plusieurs double hélices d'ADN de forme A et de forme B, afin de vérifier en particulier la stabilité des double hélices d'ADN en solution avec notre modèle de solvant.

## 5.1 Présentation des molécules étudiées et des simulations

Nous présenterons tout d'abord les différentes molécules étudiées, leurs particularités, leurs structures tridimensionnelles, et les principales données expérimentales et théoriques disponibles les concernant. Puis, nous décrirons les paramètres de simulation utilisés dans cette étude, ainsi que les paramètres des dynamiques que nous avons choisis d'analyser. Enfin, nous présenterons et analyserons les résultats structuraux, énergétiques et d'hydratation de ces simulations, d'abord pour l'ARNt, puis pour les double hélices d'ADN-A et B.

### 5.1.1 Les molécules étudiées

#### La boucle anticodon de l'ARN de transfert

Nous avons choisi d'étudier la molécule d'ARNt de l'acide aspartique, et plus précisément la partie contenant la boucle de l'anticodon. Les structures cristallographiques de plusieurs ARNt ont été résolues et affinées, le plus souvent pour la levure ou pour la bactérie *Escherichia Coli*. Toutes ces études cristallographiques ont déterminé la structure tridimensionnelle de l'ARNt en forme de L, comme nous l'avons vu au chapitre 1 (cf. fig. 1.22). Cependant, cette structure en forme de L ne serait pas parfaitement identique pour tous les ARNt, puisque l'angle entre les deux branches du L pourrait varier d'un ARNt à l'autre. Une étude cristallographique [Moras *et al.*, 1985] a ainsi montré que l'ARNt de l'acide aspartique (ARNt-Asp) avait un angle plus ouvert que l'ARNt de la phénylalanine (ARNt-Phe). Une hypothèse proposée par les auteurs de cette étude pour expliquer cette différence est que la structure cristallographique de l'ARNt-Asp correspondrait à l'ARNt lié à l'ARN messenger (ARNm), alors que l'ARNt-Phe représenterait un ARNt libre en solution [Moras *et al.*, 1985]. En effet, la nature de la séquence GUC de l'anticodon de l'ARNt-Asp entraîne la formation de dimères dans le cristal par appariement des deux anticodons via deux paires stables G-C et une paire U-U, mimant la reconnaissance entre l'anticodon et le codon complémentaire GAC de l'ARNm. Cette explication concorde avec des études de RMN qui ont montré que l'appariement de l'ARNt à l'ARNm aboutirait à l'ouverture de l'angle du L [Gassen, 1980]. L'ARNt-Asp apparaît donc comme un bon modèle pour étudier des phénomènes de reconnaissance entre l'ARNt et l'ARNm lors de la traduction génétique.

La partie de la molécule de l'ARNt-Asp contenant la boucle de l'anticodon que nous avons choisie d'étudier a déjà été simulée en dynamique moléculaire avec un solvant explicite [Auffinger et Westhof, 1996]. Cette étude a montré que, même coupée du reste de la molécule, cette boucle était stable en solution, mais seulement avec un traitement des interactions longue distance de type Particle Mesh Ewald (PME). Nous avons confirmé ces observations en simulant cette boucle avec ORAC en solvant explicite TIP3P [Capitaine, 2000] : avec une coupure des interactions non liantes à 12 Å, la structure se dénature très vite, rendant nécessaire l'utilisation de l'algorithme PME. Cependant, des simulations similaires ont été effectuées sur la boucle anticodon de l'ARNt-Phe avec une coupure des interactions longue distance à 12 Å, et ont montré que l'on pouvait obtenir des trajectoires stables [Lahiri et Nillson, 2000]. S'il n'y a donc pas de consensus sur la méthode de traitement des interactions électrostatiques à longue portée, ces différentes études s'accordent néanmoins sur le fait que les interactions électrostatiques et la représentation du solvant sont primordiales pour la stabilité de cette boucle qui se dénature extrêmement

vite dans le vide, et même avec un modèle de constante diélectrique dépendant de la distance [Capitaine, 2000], et cela en fait un bon système pour tester notre modèle de solvant PPP.

Le point de départ de nos simulations est la structure cristallographique de l'ARN de transfert de l'acide aspartique de la levure disponible dans la base de données des acides nucléiques (NDB) par l'entrée « trna05 » [Comarmond *et al.*, 1986]. Nous avons coupé la molécule aux 17 nucléotides de la boucle anticodon (résidus 27 à 43), et la partie de la molécule étudiée comporte cinq paires de bases en double hélice et une boucle monocaténaire de sept résidus reliant les deux brins de l'hélice. La séquence contient deux bases modifiées dans la partie en boucle, la 1-méthyl-guanine (notée 1mg ou m<sup>1</sup>G) et la pseudo-uracile (notée Ψ), dont les formules chimiques sont données dans le premier chapitre (cf. fig. 1.20). La courte double hélice comporte quatre paires G–C de grande stabilité, et une paire de bases non standard G–U, appelée paire « wobble », dont la stabilité est moins importante. Le triplet GUC de l'anticodon est complémentaire du codon GAC de l'ARN messager, qui code pour l'acide aspartique. Les structures secondaire et tridimensionnelle de cette boucle sont représentées sur la figure 5.1.

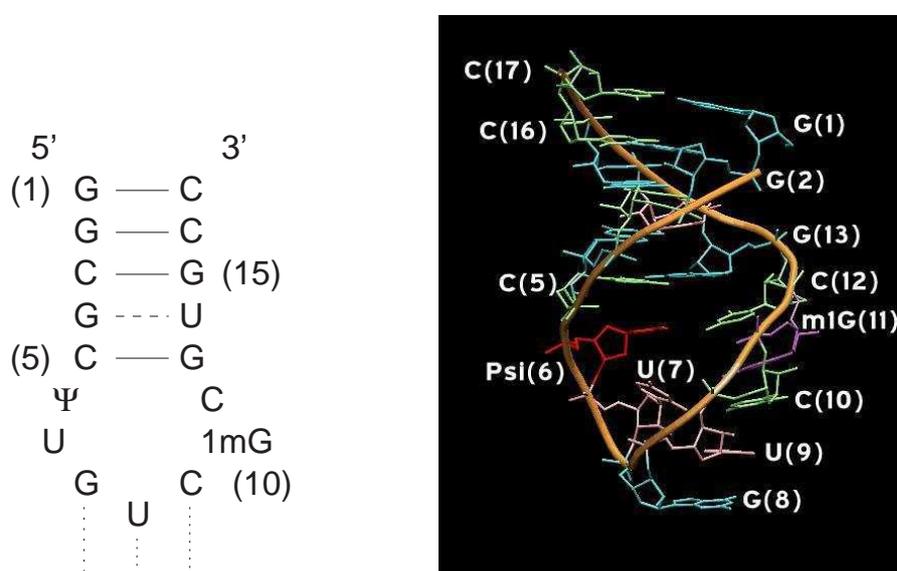


FIG. 5.1 : Structure secondaire (à gauche) et tridimensionnelle (à droite) de la boucle anticodon de l'ARN de transfert de l'acide aspartique.

### Les double hélices d'ADN A et B

Les structures cristallographiques de nombreuses double hélices d'ADN-B et A sont résolues et disponibles dans la NDB, comme on l'a vu au chapitre 1. La forme B de

la double hélice d'ADN est reconnue par les méthodes expérimentales comme étant la forme dominante en solution [Franklin et Goslin, 1953], la forme A étant favorisée par une faible activité de l'eau ou par une séquence de bases riche en paires CG [Wang *et al.*, 1987]. Différentes études de dynamique moléculaire s'intéressant aux transitions entre les formes A et B de double hélices d'ADN ont confirmé ces résultats expérimentaux, en montrant qu'on observait le plus souvent lors de simulations en solvant explicite des double hélices de forme B, ou des formes hybrides entre A et B pour le cas des séquences riches en oligo-(dG), même en partant parfois d'une conformation A [Cheatham *et al.*, 1998, Sprous *et al.*, 1998, Steff *et al.*, 2001]. Cependant, la grande stabilité des double hélices d'ADN-B au détriment des formes A lors des simulations doit être interprétée avec précaution, car le champ de force utilisé pourrait favoriser une hélice de forme B dans le cas d'AMBER, le plus couramment utilisé pour les simulations de double hélices d'ADN, alors que CHARMM favoriserait plutôt les formes A [Feig et Pettitt, 1998].

Par ailleurs, plusieurs revues des études théoriques de dynamique moléculaire sur les acides nucléiques ont mis en évidence que la stabilité des double hélices d'ADN dépendait fortement des méthodes utilisées pour représenter le solvant et pour prendre en compte les interactions électrostatiques à longue distance [Auffinger et Westhof, 1998b, Beveridge et McConnell, 2000, Giudice et Lavery, 2002]. Aujourd'hui, la plupart des dynamiques moléculaires d'acides nucléiques sont effectuées en solvant explicite, et avec l'algorithme PME [Young *et al.*, 1997, Beveridge et McConnell, 2000]. Quelques simulations d'ADN ont été effectuées avec des méthodes de solvation implicites de résolution de Poisson-Boltzmann pour le calcul de l'énergie libre électrostatique de solvation, mais avec un solvant explicite pour le calcul des trajectoires de la dynamique [Srinivasan *et al.*, 1998]. Plus récemment, des simulations de dynamique moléculaire sur des double hélices d'ADN-A et B ont été effectuées avec la méthode de Born Généralisé soigneusement paramétrée, et les structures issues de ces trajectoires sont en bon accord avec les simulations en solvant explicite [Tsui et Case, 2000]. Mais les simulations de double hélices d'ADN avec Born Généralisé sont encore peu nombreuses, et ne permettraient pas de produire des trajectoires stables pour certaines séquences de bases [Giudice et Lavery, 2002].

Nous avons choisi de simuler avec le solvant PPP plusieurs acides nucléiques de séquences différentes, en double hélice A ou B, et dont la structure cristallographique est disponible dans la NDB :

- le décimère double brin de forme B de séquence d(CCGCCGGCGG)<sub>2</sub>, que nous noterons ADN-B1, de code d'entrée « bd0015 » dans la NDB [Timsit et Moras, 1994] ;
- le décimère double brin de même séquence d(CCGCCGGCGG)<sub>2</sub>, mais en forme

- A, noté ADN-A1, de code d'entrée « adj0109 » dans la NDB [Mayer-Jung *et al.*, 1998] ;
- le dodécamère double brin de forme B de séquence  $d(\text{CGCGAATTCGCG})_2$ , que nous appellerons ADN-B2, de code d'entrée « bdl001 » dans la NDB [Drew *et al.*, 1981].

Les structures tridimensionnelles de ces trois molécules sont représentées sur la figure 5.2.

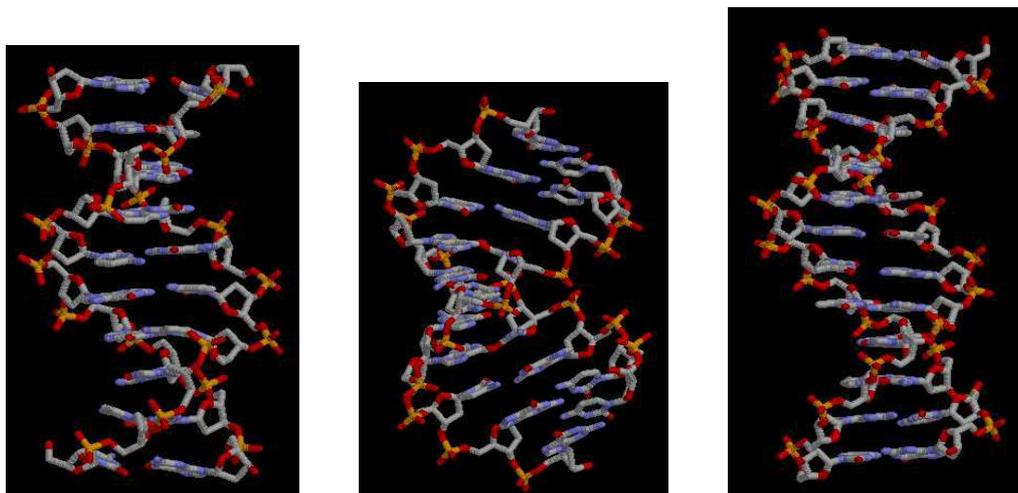


FIG. 5.2 : Structures tridimensionnelles des trois double hélices d'ADN étudiées. De gauche à droite : l'ADN-B1, l'ADN-A1 et l'ADN-B2.

Le décimère de séquence  $d(\text{CCGCCGGCGG})_2$  a d'abord été cristallisé sous la forme B [Timsit et Moras, 1994], puis sous la forme A [Mayer-Jung *et al.*, 1998]. Le fait que la séquence soit riche en paires C–G nous incite à penser que cette molécule d'ADN en solution pourrait potentiellement adopter une forme A, mais il n'y a pas à ce jour de résultats connus sur cette séquence précise en solution permettant de le confirmer. Cependant, nous pourrions comparer les résultats de nos simulations à des résultats plus généraux sur la structure et l'hydratation des ADN-A et B en solution et des ADN de séquence riche en paires C–G. Plusieurs études expérimentales et théoriques des transitions A→B ou B→A ont montré que, plutôt que l'abondance de paires C–G, ce serait les « G-tracts » (séquences dans le sens 5'–3' d'au moins quatre bases successives ne contenant que des d(GG) ou d(GC), mais pas de d(CG)) qui favoriseraient la forme A en solution [Steff *et al.*, 2001, Hud et Plavec, 2003]. La séquence de l'ADN étudié ici ne contient pas de G-tract, mais au mieux une succession de trois bases GGC, et nous ne pouvons donc pas affirmer qu'elle va adopter la forme A en solution. De plus, si des formes d'ADN-A sont restées stables sous certaines conditions en dynamique moléculaire [Sprous *et al.*,

1998], la transition d'ADN-B en ADN-A est difficile à observer pendant une dynamique, même en diminuant l'activité de l'eau (dans un solvant constitué d'un mélange d'eau et d'éthanol) donc en favorisant la forme A déshydratée [Sproun *et al.*, 1998], sans doute à cause des temps de simulation limités et des champs de force utilisés qui favorisent le plus souvent la forme B [Feig et Pettitt, 1998]. Nous regarderons donc précisément les paramètres hélicoïdaux des structures issues de chacune des dynamiques de l'ADN-B1 et de l'ADN-A1 afin de voir si nous observons une transition structurale pendant une de ces dynamiques, ou si obtient une forme hybride entre A et B. Notons que la seule donnée de la phase des sucres n'est pas toujours un critère suffisant pour distinguer les formes A des formes B, à cause de l'existence des formes hybrides [Steffl *et al.*, 2001, Ng et Dickerson, 2002], et nous étudierons donc plusieurs autres paramètres hélicoïdaux et d'hydratation.

La double hélice d'ADN-B2, aussi communément appelée séquence de Drew et Dickerson, est la première molécule d'ADN de forme B à avoir été cristallisée [Dickerson et Drew, 1981, Drew et Dickerson, 1981]. L'ADN-B2 contient dans ses 6 paires de bases centrales le site de reconnaissance de l'enzyme de restriction endonucléase EcoRI, dont le complexe a été le premier complexe ADN-protéine à avoir été cristallisé [Frederick *et al.*, 1984]. Cette séquence constitue le système de référence pour les ADN de forme B, et a fait l'objet de nombreuses études expérimentales et théoriques dont nous ne présenterons ici que les principaux résultats. La structure cristallographique [Drew *et al.*, 1981] montre quelques particularités structurales : la structure tridimensionnelle n'est pas tout à fait symétrique malgré la symétrie de la séquence, et le petit sillon est plus étroit au niveau de la région AATT. Cette structure cristallographique a également mis en évidence la présence d'une épine d'hydratation dans le petit sillon de l'hélice. Des études de RMN ont montré que cette molécule était stable en forme B en solution, et ont également confirmé l'existence d'une épine d'hydratation dans le petit sillon [Lane *et al.*, 1991, Liepinsh *et al.*, 1994]. Young *et al.* [1997] ont effectué une simulation de cette molécule avec AMBER et PME en solvant explicite sur une durée de 5 ns, qui a ensuite été poursuivie de façon stable sur 14 ns et comparée avec succès à des résultats de RMN [Arthanari *et al.*, 2003]. Nous espérons pouvoir observer sur cette molécule d'ADN bien connue des effets de séquence sur les paramètres étudiés pendant la simulation.

### 5.1.2 Paramètres des simulations

#### Le solvant, le soluté et le protocole de simulation

Les paramètres utilisés pour le solvant PPP dans les simulations de ces acides nucléiques sont les mêmes que ceux utilisés pour les protéines au chapitre précédent. La boîte

de simulation est rectangulaire de dimensions  $48 \text{ \AA} \times 48 \text{ \AA} \times 70,7 \text{ \AA}$ , contenant 5264 particules de solvant pour l'ARNt, 5241 pour l'ADN-B1, 5232 pour l'ADN-A1, et 5174 pour l'ADN-B2. Les paramètres de Lennard-Jones des particules de solvant sont toujours de  $\sigma_s = 2,88 \text{ \AA}$  et  $\epsilon_s = 3,197 \text{ kJ/mol}$ , la valeur du dipôle de saturation est de  $\mu = 1,5 \text{ Debye}$  et la polarisabilité des particules de  $\alpha = 2,33 \text{ \AA}^3$ . Le champ de force d'AMBER (parm94 [Cornell *et al.*, 1995]) est utilisé pour les acides nucléiques. Pour les bases modifiées de l'ARNt, les paramètres sont calculés à partir de ceux des bases standard dont la formule chimique est proche.

Les pas d'intégration multiples sont les mêmes que précédemment : 4 fs pour les interactions longue portée, 2 fs pour les moyenne portée et 1 fs pour les interactions courte portée et intramoléculaires. Les interactions entre atomes non liés sont coupées par une fonction de « switch » entre 12 et 13  $\text{\AA}$ . L'algorithme SHAKE est utilisé pour les liaisons impliquant des atomes d'hydrogène. Nous utilisons un thermostat de Nosé-Hoover pendant la dynamique, afin de maintenir la température à  $T = 300 \text{ K}$ .

Avant les dynamiques, les structures des solutés issues de la cristallographie sont d'abord minimisées avec ORAC dans le vide, sans les contre-ions, par la méthode des gradients conjugués avec un critère de convergence de  $5 \cdot 10^{-3}$ . Puis, nous avons rajouté les ions et le solvant et nous avons équilibré les particules de solvant autour du soluté et des ions figés à 300 K pendant 24 ps, et pendant 48 ps avec le soluté figé et les ions mobiles. Ensuite, nous avons augmenté la température par paliers sans aucune contrainte entre 20 et 300 K pendant une durée totale de 156 ps. Au début de chacune des étapes de chauffage et d'équilibration de 12 ps, les vitesses des atomes sont attribuées de façon aléatoire selon une distribution de Boltzmann. Enfin, pour les différentes simulations, la production de dynamique est effectuée sur 1080 ps.

Nous avons utilisé pour toutes les simulations d'acides nucléiques un facteur de charge de  $\lambda_q = 1,185$ , non seulement d'après les conclusions tirées des simulations des protéines présentées au chapitre précédent, mais aussi parce que nous avons observé que l'utilisation d'un facteur de charge de  $\lambda_q = 1,2$  dénaturait rapidement la structure des acides nucléiques fortement chargés (résultats non présentés dans cette thèse).

## Les contre-ions

Les acides nucléiques sont chargés au niveau des groupements phosphates  $\text{PO}_4^-$  de la chaîne polynucléotidique. Afin d'étudier un système neutre, on peut donc introduire des contre-ions chargés positivement. Dans le cytoplasme de la cellule, on trouve différentes sortes d'ions chargés positivement ou négativement, des ions chlorure  $\text{Cl}^-$ , sodium  $\text{Na}^+$ ,

potassium  $K^+$ , ou des ions magnésium  $Mg^{2+}$  [Watson *et al.*, 1994]. Dans le cas de l'ARNt, des études de RMN ont montré l'importance des ions  $Mg^{2+}$  dans la stabilité de la structure de l'ARNt en solution, surtout au niveau du bras accepteur [Choi et Redfield, 1995, Friederich *et al.*, 1998]. Une étude de dynamique moléculaire sur l'ARNt-Asp entier en solvant explicite (SPC/E) avec des ions ammonium  $NH_4^+$  [Auffinger *et al.*, 1999] a montré que, si les ions  $Mg^{2+}$  participaient sans doute à la stabilisation de la structure générale de l'ARNt, la partie contenant la boucle de l'anticodon était peu sensible à la présence d'ions divalents. Dans le cas des ADN, le rôle des ions dans l'environnement des double hélices a été mis en évidence par les méthodes expérimentales et théoriques. Des ions sont parfois présents dans les structures cristallographiques, mais ils peuvent être difficiles à distinguer des molécules d'eau [Soler-Lopez *et al.*, 2000]. Des simulations de dynamique moléculaire ont montré que les contre-ions faisaient partie intégrante des premières couches d'hydratation des acides nucléiques, en prenant parfois la place de certaines molécules d'eau [Young *et al.*, 1997, Feig et Pettitt, 1999].

Le placement des contre-ions autour du soluté au début des simulations d'acides nucléiques est très variable selon les simulations dans la littérature, mais l'étude de la dynamique d'une double hélice d'ADN [Young *et al.*, 1997] a montré que ce placement initial (sur la bissectrice des phosphates, à des minima du potentiel électrostatique, par des méthodes de type Monte-Carlo,...) avait très peu d'influence sur les simulations qui suivaient, les ions étant très mobiles pendant la dynamique.

Nous avons choisi de placer au début de nos simulations des ions sodium à proximité des groupements phosphates des acides nucléiques, sur la bissectrice de l'angle formé par les deux liaisons P-O chargées, à une distance de 6 Å (pour l'ARNt et l'ADN-B2) ou 3 Å (pour l'ADN-A1 et B1) de l'atome de phosphore. Comme le brin phosphodiester de l'extrémité 5' n'est pas chargé, nous avons rajouté pour l'ARNt 16 ions  $Na^+$ , pour l'ADN-B2 22 ions, et pour les ADN-B1 et A1 18 ions. Les paramètres de Lennard-Jones des ions sodium que nous avons utilisés sont :  $\sigma_{sod} = 2,922$  Å et  $\epsilon_{sod} = 0,419$  kJ/mol.

### 5.1.3 Paramètres analysés

#### Analyse structurale

Pour toutes les simulations d'acides nucléiques, nous étudierons l'évolution du RMS des atomes lourds des molécules pendant la simulation, par rapport à la structure minimisée initiale. Notons que pour le cas des ADN double brin, nous calculerons le RMS sur les deux brins comme une seule molécule. Pour le cas particulier de la boucle anticodon

de l'ARNt, afin d'étudier précisément les fluctuations de certains nucléotides, nous calculerons également le RMS moyen par résidu par rapport à la structure moyenne sur une partie stable de la simulation.

Afin de caractériser sur le plan structural les conformations des acides nucléiques échantillonnées toutes les 12 ps pendant la dynamique, nous calculerons les paramètres hélicoïdaux (présentés au chapitre 1) de ces structures en utilisant le programme CURVES 51 [Lavery et Sklenar, 1996]. L'algorithme de CURVES repose sur le calcul du meilleur axe global non linéaire pour la double hélice [Stofer et Lavery, 1994], et permet donc de calculer de façon précise, même pour les hélices courbées, les paramètres hélicoïdaux globaux pour les paires de bases, les inter-paires de bases, et les bases seules. Ce programme permet aussi de calculer la phase des sucres de tous les nucléotides. Nous présenterons ces résultats sous forme d'histogrammes représentant la répartition des différents paramètres calculés sur les 90 structures issues d'une même simulation, prélevées toutes les 12 ps. Nous comparerons ces répartitions aux valeurs de référence des ADN-A et B (cf. tableau 1.2), ainsi que, pour l'ARNt dont la structure est plus inhabituelle, aux valeurs des paramètres de la structure initiale.

### Analyse énergétique

Comme pour les protéines, nous étudierons d'abord l'évolution temporelle des différentes composantes énergétiques des systèmes étudiés : l'énergie totale, l'énergie potentielle du soluté et l'énergie libre électrostatique de solvation. Nous calculerons également les corrélations entre les énergies libres électrostatiques moyennes avec le solvant PPP et celles calculées sur les structures moyennes avec DelPhi, pour seulement une structure sur deux issue de la dynamique. Notons que pour que le système étudié avec les deux méthodes soit neutre, les énergies sont calculées avec DelPhi pour le système composé du soluté et des contre-ions traités explicitement. Les paramètres de DelPhi sont les mêmes qu'au chapitre précédent : l'espacement de la grille est de  $5 \text{ \AA}^{-1}$  et le pourcentage d'occupation du soluté dans la boîte est de 75 %.

### Analyse de l'hydratation

D'une manière analogue à l'étude des protéines, nous avons étudié les sites préférentiels d'hydratation des acides nucléiques par nos pseudo-particules de solvant, en particulier au niveau des groupements polaires des nucléotides, et plus précisément :

- autour des atomes d'oxygène des groupements phosphates du brin phosphodiester ;

- autour des différents groupements polaires donneurs ou accepteurs de liaisons hydrogène (N, O ou H) des bases, du côté du grand sillon ;
- autour des groupements polaires des bases du côté du petit sillon.

Nous calculerons un pourcentage maximum d'hydratation pour chacun de ces groupements pour tous les nucléotides en procédant de la même façon que pour l'étude des protéines. Nous pourrions ainsi identifier les sites préférentiels d'hydratation des acides nucléiques selon leur séquence, et préciser si ces sites se trouvent au niveau des phosphates, du petit ou du grand sillon de chaque nucléotide. Nous comparerons ces résultats avec ceux expérimentaux ou théoriques disponibles dans la littérature.

## 5.2 Résultats obtenus sur l'ARN de transfert

### 5.2.1 Analyse dynamique et structurale

#### Analyse du RMS

L'évolution du RMS des atomes lourds de la boucle anticodon de l'ARNt-Asp (que nous noterons simplement ARNt par la suite) pendant la dynamique est représentée sur la figure 5.3. Le RMS stagne autour de 1,35 Å pendant les premières 250 ps de la simulation, puis augmente et se stabilise après 300 ps de dynamique autour d'environ 1,45 Å. La structure semble s'être globalement stabilisée à la fin de la dynamique, même si on remarque encore de faibles variations du RMS, indiquant que la structure n'est pas complètement figée.

Le RMS moyen par résidu de l'ARNt, calculé par rapport à la structure moyenne sur la période stable de la dynamique d'après l'évolution du RMS, c'est-à-dire sur les dernières 720 ps de la simulation, est représenté sur la figure 5.4. On remarque qu'en moyenne, le RMS par résidu est plus élevé dans la partie en boucle (nucléotides 6–12) que dans la partie en double hélice de l'ARNt. En effet, nous avons vu au chapitre 1 que les nucléotides participant à une double hélice étaient rigidifiés non seulement par les interactions d'empilement entre bases d'un même brin mais aussi par les liaisons hydrogène entre bases complémentaires. Dans la partie en double hélice, la paire G1–C17, qui est à l'extrémité et n'est pas contrainte pendant la simulation, est plus mobile que les autres. La paire « wobble » G4–U14 paraît aussi rigide, voir plus, que les paires Watson-Crick standard, et cela avait déjà été observé dans la simulation du même ARNt en solvant explicite [Auffinger et Westhof, 1996]. Au sein de la boucle, le triplet de l'anticodon est très mobile au niveau des bases G8 et U9, alors que la troisième base C10 est plus figée.

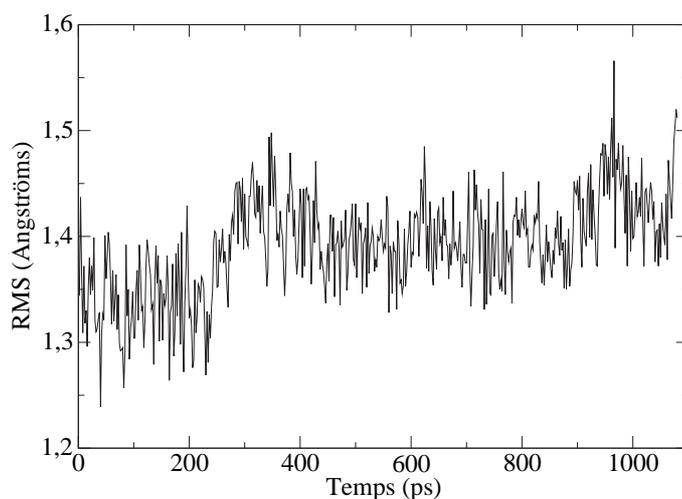


FIG. 5.3 : Évolution du RMS des atomes lourds de l'ARNt par rapport à la structure initiale pendant la simulation.

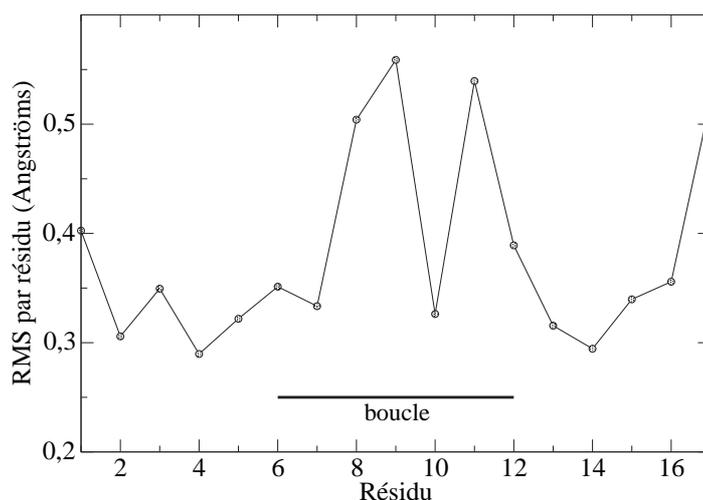


FIG. 5.4 : RMS moyen par résidu des atomes lourds de l'ARNt par rapport à la structure moyenne sur 720 ps de dynamique.

La forte mobilité de la première base de l'anticodon (G8) avait aussi été remarquée dans la simulation de cette même molécule en solvant explicite [Auffinger et Westhof, 1996], ainsi que dans la simulation de la boucle équivalente de l'ARNt-Phe [Lahiri et Nilsson, 2000]. En effet, cette base G8 se trouve à l'extrémité de la boucle, comme on le voit sur la structure tridimensionnelle de la figure 5.1, et elle est donc plus accessible au solvant que les autres et moins contrainte par les interactions d'empilement. Une étude des particularités structurales des boucles anticodon [Auffinger et Westhof, 2001] a suggéré que le nucléotide en position 10 formait des liaisons hydrogène stabilisant la boucle entre son groupement phosphate et la base en 7 (ici U7), et cela pourrait expliquer pourquoi le

nucléotide C10 reste très figé pendant notre simulation. Enfin, la base modifiée 1mG11 est elle aussi très mobile, et nous pensons que ce serait probablement parce qu'elle ne peut former aucune liaison hydrogène avec les autres bases de la boucle, ni avec  $\Psi 6$  qui est plus ou moins apparié à C12, ni avec U7 qui interagit avec C10.

### Analyse des paramètres hélicoïdaux avec Curves

Nous avons effectué une étude structurale plus fine de l'ARNt pendant toute la durée de la simulation avec CURVES, sur les 90 structures PDB de l'ARNt prélevées toutes les 12 ps pendant la dynamique. Nous avons d'abord analysé la conformation des sucres des 17 nucléotides, puis les paramètres hélicoïdaux de la partie en double hélice constituée de cinq paires de bases (nucléotides 1–5 et 13–17), et enfin les paramètres d'une partie du simple brin constituée des cinq nucléotides empilés après le tournant abrupt de la boucle (nucléotides 8–12).

#### •La phase des sucres du squelette :

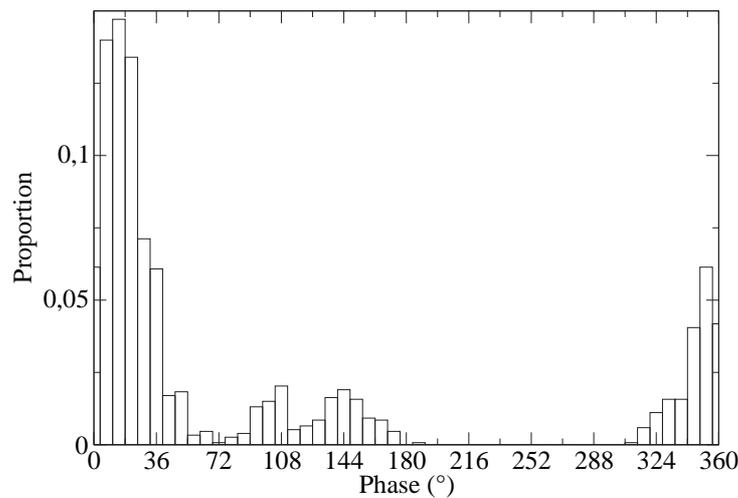


FIG. 5.5 : Répartition des nucléotides en fonction de la phase de leur sucre en degrés pendant toute la simulation de l'ARNt.

La figure 5.5 représente la répartition de l'ensemble des nucléotides de l'ARNt pendant toute la simulation en fonction de la phase de leur sucre. On voit que les phases des sucres sont très majoritairement entre 0 et 36°, c'est-à-dire que les sucres sont principalement en conformation C3'-endo ou Nord, conformation préférentielle pour les double hélices d'ADN de forme A ou d'ARN. Quelques sucres ont transité en conformation C2'-endo ou Sud (autour de 144°), conformation majoritaire pour les double hélices d'ADN-B. Le tableau 5.1 donne pour chaque nucléotide la valeur de la phase de la structure initiale,

et la valeur moyenne de la phase, l'écart-type et la conformation moyenne pendant la simulation. Pour la partie de l'ARNt en double hélice, les sucres sont presque tous en moyenne en conformation C3'-endo, sauf pour le nucléotide C17 dont la phase varie beaucoup, et qui est à l'extrémité de l'hélice donc plus mobile. Dans la partie en boucle (nucléotides 6–12), on remarque que les nucléotides 8 à 12, fortement empilés les uns au-dessus des autres en simple brin, ont aussi un sucre moyen autour de C3'-endo. Les sucres des nucléotides  $\Phi$ 6 et U7 ne sont pas du tout en conformation C3'-endo, ces deux bases étant les moins empilées et se trouvant très proches du tournant abrupt de la boucle. Ces deux nucléotides  $\Phi$ 6 et U7 sont aussi ceux pour lesquels la phase pendant la simulation est la plus éloignée de la phase de la structure initiale, qui n'était pas non plus en C3'-endo.

<i>nuc.</i>	$P_0(^{\circ})$	$P(^{\circ})$	<i>conformation</i>	<i>nuc.</i>	$P_0(^{\circ})$	$P(^{\circ})$	<i>conformation</i>
<b>G1</b>	2	11 (11)	C3'-endo	<b>C17</b>	15	94 (74)	O1'-endo
<b>G2</b>	14	1 (16)	C3'-endo	<b>C16</b>	5	21 (16)	C3'-endo
<b>C3</b>	24	20 (19)	C3'-endo	<b>G15</b>	23	23 (22)	C3'-endo
<b>G4</b>	21	16 (9)	C3'-endo	<b>U14</b>	20	16 (10)	C3'-endo
<b>C5</b>	25	-1 (8)	C2'-exo	<b>G13</b>	18	22 (15)	C3'-endo
$\Psi$ 6	56	101 (18)	O1'-endo	<b>C12</b>	20	17 (13)	C3'-endo
<b>U7</b>	37	135 (17)	C1'-exo	<b>m<sup>1</sup>G11</b>	3	3 (10)	C3'-endo
<b>G8</b>	-14	-24 (14)	C2'-exo	<b>C10</b>	6	13 (10)	C3'-endo
<b>U9</b>	30	20 (21)	C3'-endo				

TAB. 5.1 : Phase des sucres de la structure initiale ( $P_0$ ) et phase moyenne des sucres de tous les nucléotides pendant la simulation de l'ARNt,  $P \in [-180^{\circ}; +180^{\circ}]$ . L'écart-type est indiqué entre parenthèses.

#### •Les paramètres de la partie double hélice :

Pour la partie en double hélice (1–5 et 13–17), nous avons étudié les paramètres hélicoïdaux pour chaque paire de bases ou inter-paires de bases.

Pour les paramètres axe-paires de bases, les répartitions des deux translations  $X_{\text{disp}}$  et  $Y_{\text{disp}}$  sont représentées sur la figure 5.6 pour chacune des cinq paires de bases pendant toute la simulation. Nous avons vu au chapitre 1 que, pour une double hélice d'ADN-A, le  $X_{\text{disp}}$  est en moyenne autour de  $-5,3 \text{ \AA}$  et le  $Y_{\text{disp}}$  autour de  $0 \text{ \AA}$ . On observe que toutes les paires de bases de la double hélice de l'ARNt ont des  $X_{\text{disp}}$  et  $Y_{\text{disp}}$  autour de  $-2 \text{ \AA}$ . Les paires de bases seraient donc régulièrement translatées par rapport à l'axe comparé à un ADN-A classique. Les paires de bases dans la conformation initiale avaient un  $X_{\text{disp}}$  aux alentours de  $-2,5 \text{ \AA}$ , et un  $Y_{\text{disp}}$  légèrement positif. Les paires de bases

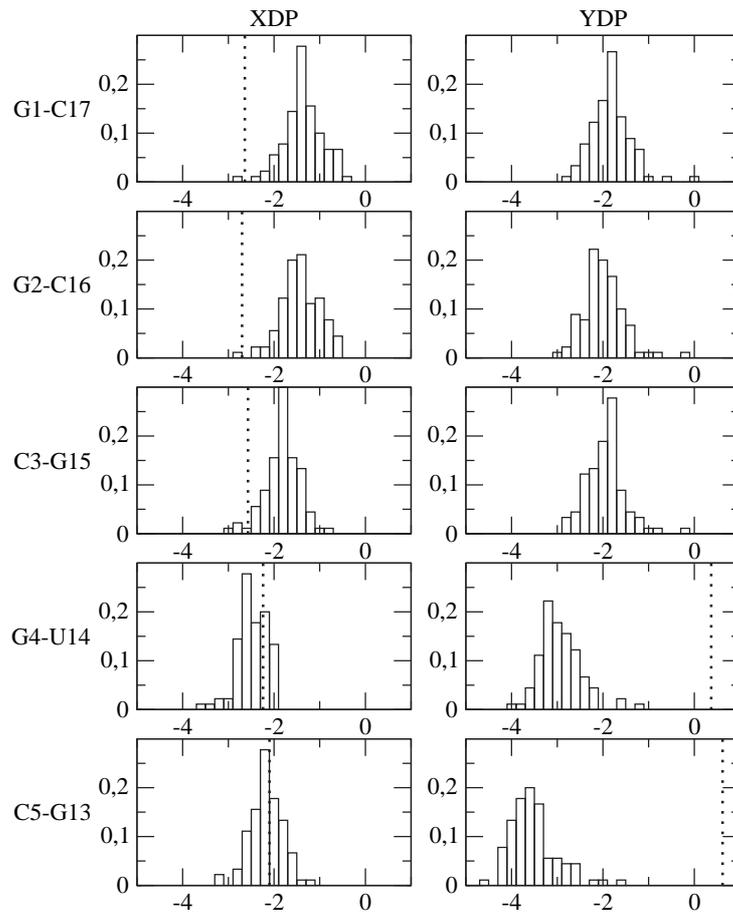


FIG. 5.6 : Répartition des paramètres axe-paires de bases :  $X_{disp}$  ( $XDP$ ) et  $Y_{disp}$  ( $YDP$ ) en Å, pour les différentes paires de bases de la partie en double hélice de l'ARNt. Les valeurs des paramètres de la structure initiale sont représentés en pointillés.

se sont donc déplacées de l'axe pendant la simulation par rapport à la structure initiale. Les paramètres de la paire « wobble » G4-U14 se distinguent de ceux des autres paires, son  $X_{disp}$  et son  $Y_{disp}$  étant tous deux plutôt aux alentours de  $-3$  Å, et cette paire est donc plus éloignée de l'axe que les autres pendant la simulation. On observe également que les paramètres de la dernière paire de la double hélice C5-G13 sont en léger décalage par rapport aux autres (avec un  $Y_{disp}$  autour de  $-3,5$  Å en moyenne), et la paire est donc un plus éloignée de l'axe que les autres, sans doute parce qu'elle est à la jonction avec la partie en boucle.

Pour les paramètres intra-paire de bases, les répartitions de la translation Stretch, et des trois angles Buckle, Propeller Twist et Opening pour chacune des cinq paires de bases pendant toute la simulation sont représentées sur la figure 5.7. On remarque le Stretch de la paire wobble est autour de  $-1,2$  Å, alors que tous les autres Stretches sont autour de  $0$  Å. Afin de s'apparier avec G, il est connu que la base U est tournée vers le grand sillon

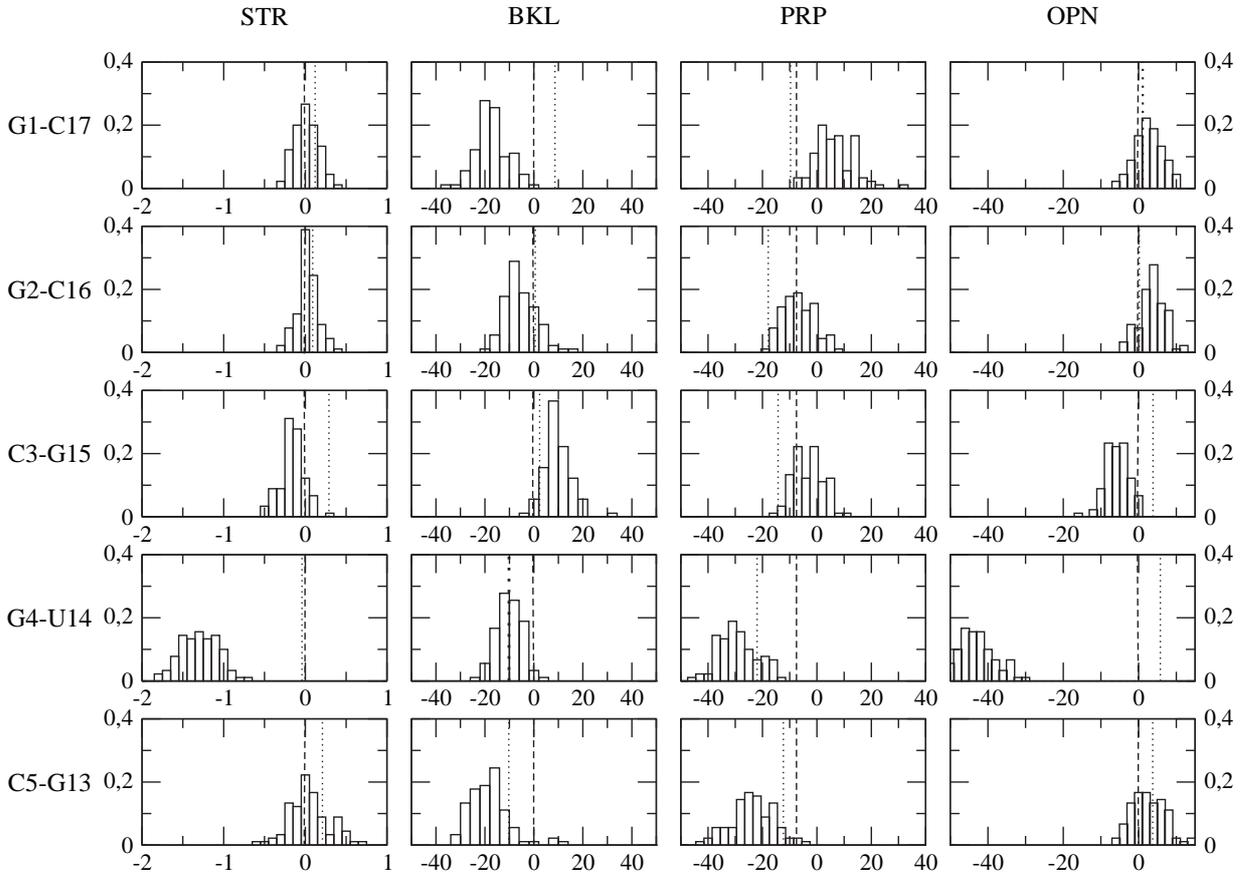


FIG. 5.7 : Répartition des paramètres intra-paire de bases : Stretch (*STR*) en Å, Buckle (*BKL*), Propeller Twist (*PRP*) et Opening (*OPN*) en degrés, pour les différentes paires de bases de l'ARNt. Les paramètres moyens des ADN-A sont représentés en traits pointillés, et les paramètres de la structure initiale en pointillés.

par rapport à la base C canonique [Masquida et Westhof, 2000]. Pour cette même paire wobble, le Propeller Twist et l'Opening sont beaucoup plus négatifs (respectivement  $-30^\circ$  et  $-45^\circ$  en moyenne) que pour les autres paires. Alors que le Propeller Twist de cette paire dans la structure initiale était déjà plus négatif que les autres ( $-22^\circ$ ), l'Opening était au contraire légèrement positif ( $6^\circ$ ). Pour les autres paires de bases, les valeurs des paramètres intra-paires sont en général autour de la valeur moyenne des ADN-A et des valeurs de la structure de départ. Pour les paires des deux extrémités de l'hélice G1-C17 et C5-G13, ces paramètres s'éloignent un peu plus des valeurs de référence, en particulier pour les angles Buckle et Propeller Twist. Ces deux paires étant aux extrémités de l'hélice, on comprend qu'elles puissent s'éloigner plus aisément de la structure initiale pendant la simulation.

Pour les paramètres inter-paires de bases, les répartitions de la translation Rise, et

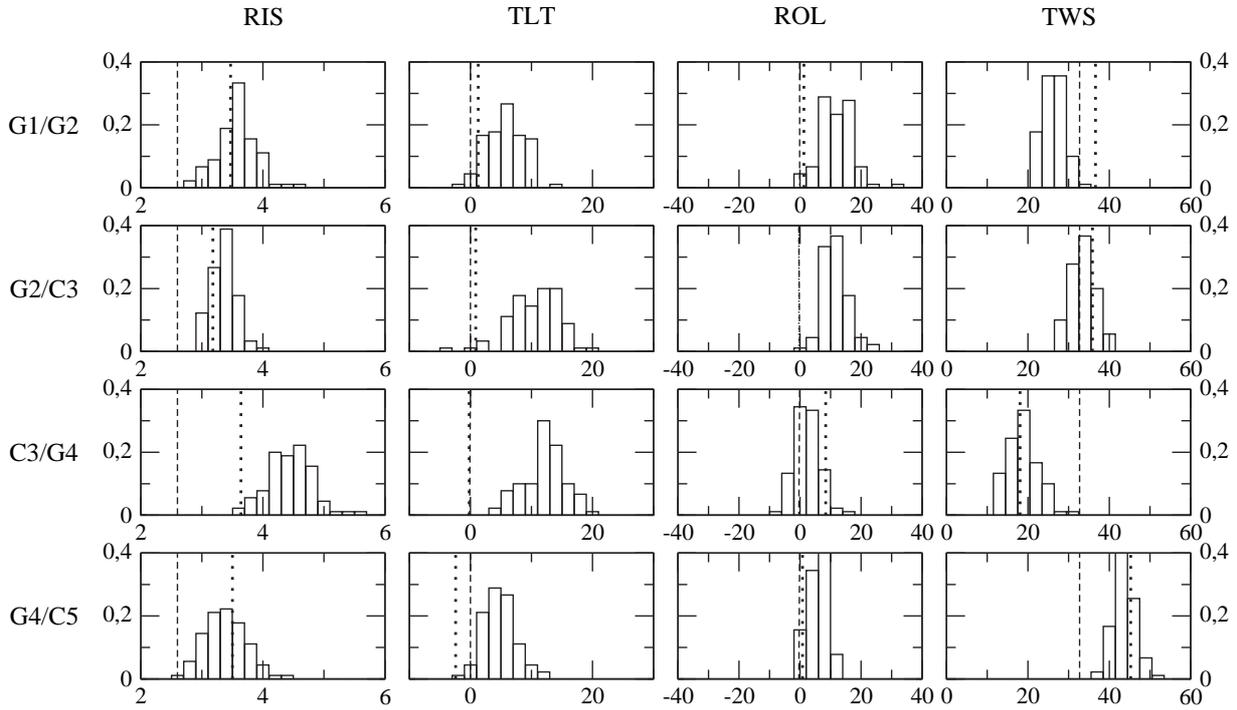


FIG. 5.8 : Répartition des paramètres inter-paires de bases de l'ARNt : Rise (RIS) en Å, Tilt (TLT), Roll (ROL) et Twist (TWS) en degrés pour les différents inter-paires de bases de l'ARNt. Les paramètres moyens des ADN-A sont représentés en traits pointillés et les paramètres de la structure initiale en pointillés.

des trois angles Tilt, Roll et Twist sont représentées sur la figure 5.8 pour chacun des quatre inter-paires de bases pendant la simulation. On remarque une fois encore que la paire wobble se démarque des autres paires de bases, en particulier pour le Rise qui est plus élevé que les autres et que sa valeur initiale (autour de 4,4 Å au lieu de 3,6 Å au départ pour le pas C3/G4). La paire wobble est de plus sous-enroulée, car le Twist est plus faible que les autres pour le pas C3/G4 (autour de 20° au lieu d'environ 35° pour les autres pas). Ce sous-enroulement semble être ensuite compensé par un sur-enroulement pour le pas G4/C5 (le Twist est autour de 45°), ces deux paramètres de Twist restant autour de leurs valeurs initiales. Ces résultats sur le Twist de la paire wobble concordent avec ceux observés lors de la simulation de cette même molécule en solvant explicite [Rubin-Carrez, 1992], et avec les résultats connus sur la structure des paires wobble [Masquida et Westhof, 2000]. Pour le reste de la double hélice, le Rise est en moyenne plus élevé que pour un ADN-A classique, comme lors des simulations en solvant explicite [Rubin-Carrez, 1992], et reste plutôt autour de la valeur initiale, assez proche de la valeur de 3,4 Å caractéristique des ADN-B. Le Tilt des différents inter-paires de bases est aussi plus élevé que la valeur de référence et que les valeurs initiales, montrant que les paires de bases ne sont plus tout

à fait parallèles entre elles. Le Roll est aussi légèrement positif pour toutes les paires de bases, autour d'une dizaine de degrés en moyenne, un peu plus élevé que dans la structure initiale, mais un peu plus faible que pour les simulations en solvant explicite pendant lesquelles il était plutôt autour de  $25^\circ$  [Rubin-Carrez, 1992], ce qui suggère que l'hélice serait légèrement courbée.

La structure hélicoïdale de la partie en double hélice de la boucle anticodon de l'ARNt-Asp est donc globalement bien conservée mais montre pendant la simulation certaines différences par rapport aux valeurs de référence des double hélices d'ADN-A et par rapport à la structure initiale, surtout au niveau de la paire wobble G4-U14, connue pour avoir une structure différente des paires standard [Masquida et Westhof, 2000]. Les paramètres inter-paires de bases sont un peu éloignés des valeurs de référence pour toute la double hélice, montrant que cette hélice n'est pas tout à fait rectiligne, et qu'elle adopterait une conformation légèrement plus « globulaire », comme cela a été observé dans des simulations de la même boucle en solvant explicite [Rubin-Carrez, 1992].

#### • Les paramètres de la boucle en simple brin :

Pour la partie de l'ARNt en simple brin, et plus précisément pour les bases 8–12 qui sont empilées les unes au-dessus des autres après le tournant abrupt de la boucle, nous avons étudié plusieurs paramètres hélicoïdaux pour chaque base et inter-base.

Parmi les paramètres axe-bases, nous avons calculé les translations  $X_{\text{disp}}$  et  $Y_{\text{disp}}$  par rapport à l'axe pour les cinq bases, et leurs répartitions sont représentées sur la figure 5.9. On voit qu'en moyenne le  $X_{\text{disp}}$  est nul alors qu'il valait un peu moins que  $-2 \text{ \AA}$  pour la structure initiale, alors que le  $Y_{\text{disp}}$  vaut en moyenne  $-2 \text{ \AA}$  au lieu d'un peu moins de  $0 \text{ \AA}$  pour la structure initiale. Cela suggère que les bases sont placées régulièrement par rapport à l'axe global, mais en translation uniforme par rapport à la structure initiale. On remarque aussi que les valeurs de ces paramètres sont assez étalées par rapport aux valeurs équivalentes dans la double hélice où on observait des pics de répartition plus fins. Même si ces bases sont fortement empilées, le fait qu'elles ne forment pas de double hélice, et donc pas de liaisons hydrogène stabilisatrices, les rend beaucoup plus mobiles.

On a représenté sur la figure 5.10 les répartitions de plusieurs paramètres inter-bases, la translation Rise, et les trois angles Tilt, Roll et Twist pour les quatre inter-bases. Le Rise entre les deux bases de l'anticodon U9/C10 apparaît plus faible que les autres. Ces deux bases seraient donc fortement empilées, ce qui est en accord avec l'étude du RMS par résidu qui avait mis en évidence la faible mobilité de la base C10 (cf. figure 5.4). Le Rise est au contraire plus élevé que sa valeur initiale pour le pas G8/U9, suggérant que

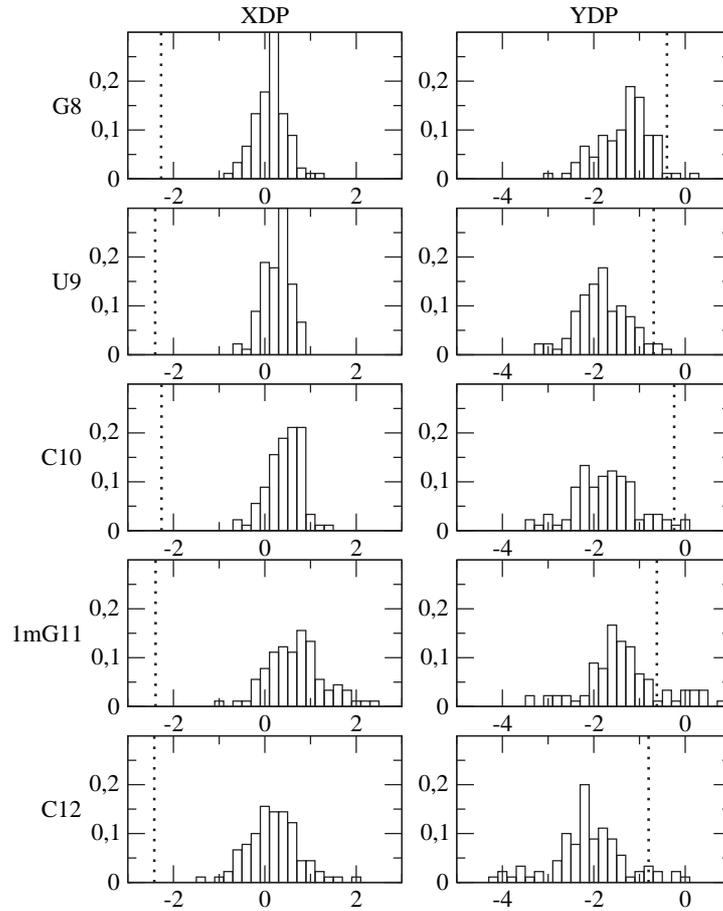


FIG. 5.9 : Répartition des paramètres axe-bases :  $X_{disp}$  ( $XDP$ ) et  $Y_{disp}$  ( $YDP$ ), en Å pour différentes bases de la partie en boucle de l'ARNt. Les paramètres de la structure initiale sont représentés en pointillés.

les interactions entre ces deux bases sont plutôt faibles, ce qui les rendrait plus mobiles, comme l'a montré le RMS par résidu. Le Roll est pour tous les pas nucléotidiques autour de  $20^\circ$ , indiquant que ces bases ne sont pas parallèles entre elles et que le brin est relativement courbé. Le Twist baisse régulièrement entre G8 et C12, et le brin se déroule donc de plus en plus entre G8 et C12.

L'étude de cette partie de la boucle de l'ARNt a malheureusement peu de points de comparaison dans la littérature. Une étude analogue sur les bases de la boucle de l'ARNt-Phe [Lahiri et Nilsson, 2000] a montré également une importante variabilité des différents paramètres par rapport à ceux initiaux, mais nous ne pouvons les comparer directement à nos résultats car la séquence de bases est ici très différente. Nous pouvons simplement déduire de cette étude que la structure de cette boucle est relativement flexible, beaucoup moins régulière qu'une structure hélicoïdale classique, et qu'elle s'éloigne de façon conséquente de la structure initiale.

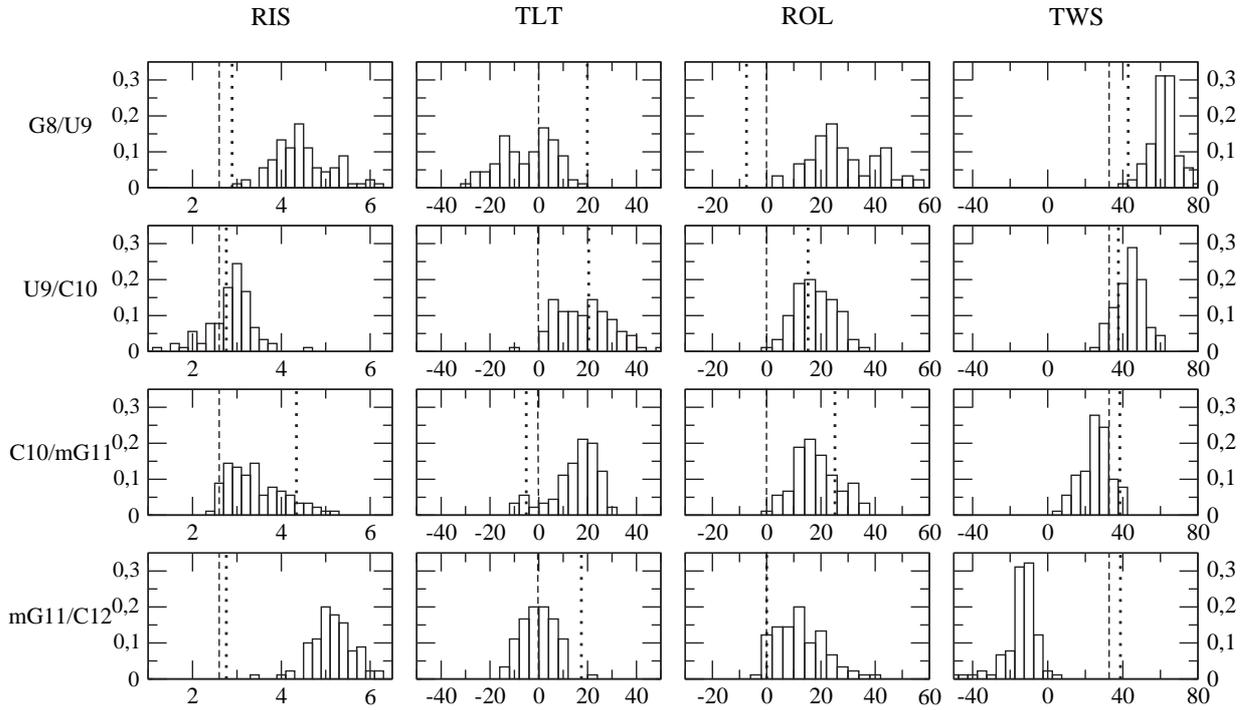


FIG. 5.10 : Répartition des paramètres inter-paires de bases : Rise (RIS) en Å, Tilt (TLT), Roll (ROL) et Twist (TWS) en degrés pour les différents inter-bases de l'ARNt. Les paramètres moyens des ADN-A sont représentés en traits pointillés et ceux de la structure initiale en pointillés.

L'analyse structurale de l'ARNt a montré que la structure de l'ARNt se stabilisait pendant la simulation autour d'une structure assez différente de la structure cristallographique. Les paramètres des structures obtenus pour la double hélice sont plutôt réguliers, et assez proches d'une hélice canonique de forme A, à part pour la paire wobble, qui montre des différences structurales notables (surtout pour le Twist) par rapport aux autres paires. Pour la partie en simple brin, la boucle proprement dite, les paramètres sont plus variables, même si les sucres des nucléotides qui la compose restent majoritairement en conformation C3'-endo. Les conclusions de cette analyse structurale de l'ARNt concordent globalement avec les simulations de la même molécule effectuées en solvant explicite [Auffinger et Westhof, 1996], en particulier pour les résultats sur la paire wobble et pour le léger repliement de la boucle sur elle-même qui se manifeste par un Roll positif. Le modèle de solvant PPP permettrait donc d'obtenir pour les ARN en boucle des simulations équivalentes d'un point de vue structural à celles en solvant explicite.

## 5.2.2 Analyse énergétique

Les évolutions de l'énergie totale, de l'énergie potentielle du soluté et de l'énergie libre de solvation sont représentées sur la figure 5.11. On remarque une fois encore une légère dérive de l'énergie totale, mais toujours négligeable par rapport aux fluctuations des autres énergies et par rapport à sa valeur moyenne. On remarque aussi que les évolutions de l'énergie potentielle du soluté et de l'énergie libre électrostatique de solvation sont encore une fois symétriques. On observe une importante variation de ces deux énergies vers 300 ps de dynamique qui correspond à la transition structurale observée dans l'évolution du RMS (cf. fig. 5.3).

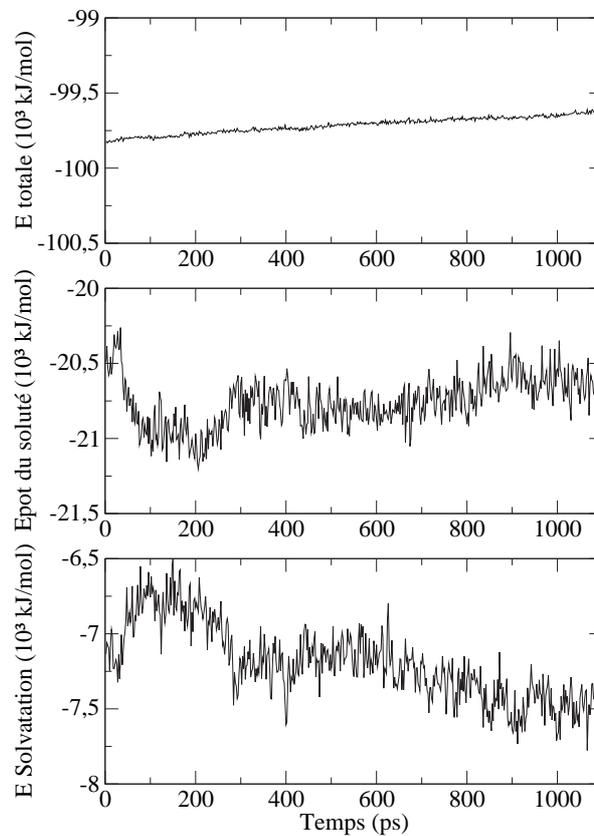


FIG. 5.11 : Évolution des énergies totale, potentielle du soluté et de solvation pendant la simulation de l'ARNt (en  $10^3$  kJ/mol).

La corrélation entre les énergies libres électrostatiques de solvation moyennes calculées avec le solvant PPP et les énergies de solvation des structures moyennes calculées avec DelPhi, ainsi que les évolutions simultanées de ces deux énergies sont représentées sur la figure 5.12. Le coefficient de corrélation est très proche de un (0,975), et il est meilleur que pour les simulations des protéines du chapitre précédent (au mieux 0,916 pour la

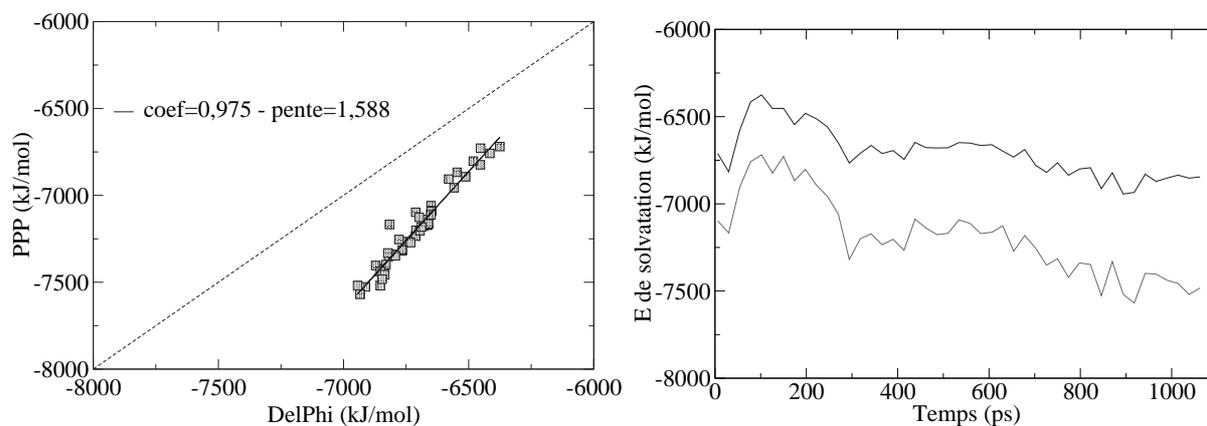


FIG. 5.12 : Corrélation entre l'énergie libre de solvation moyenne de l'ARNt calculée avec PPP et l'énergie de la structure moyenne avec DelPhi (à gauche) et évolution de ces deux énergies (à droite) en gris pour PPP et en noir pour DelPhi.

simulation de BPTI A1). L'ensemble de la boucle de l'ARNt-Asp et des ions sodium est électriquement neutre, alors que les protéines étaient légèrement chargées, et nous avons vu à plusieurs reprises que le modèle PPP donne de meilleures corrélations des énergies libres de solvation pour les molécules neutres. Cependant, la pente de la droite de corrélation est très au-dessus de un (1,588), et on observe toujours une erreur systématique (de 7 % en moyenne). On remarque sur le graphe des évolutions des énergies (cf. fig. 5.12) que l'évolution des deux énergies, malgré une différence systématique, est très similaire pendant toute la durée de la simulation. Les variations de l'énergie calculée avec le solvant PPP paraissent plus importantes que celles de l'énergie calculée avec DelPhi. De nouveau, on peut expliquer cela par le fait que notre modèle de solvant PPP nécessite de bien équilibrer les degrés de translation des particules de solvant avant d'extraire l'énergie libre de solvation, surtout après d'importants changements de conformation du soluté.

### 5.2.3 Analyse de l'hydratation

Peu de résultats expérimentaux sont disponibles pour l'hydratation de cette molécule particulière d'ARNt, mais plusieurs études cristallographiques [Eisenstein et Shakked, 1995, Schneider *et al.*, 1995] se sont intéressées à l'hydratation des double hélices d'ADN-A, dont la structure est proche de la partie en double hélice de l'ARNt. Ces études expérimentales ont mis en évidence que le grand sillon des ADN-A était particulièrement hydraté alors que le petit sillon était plus pauvrement hydraté, et que les groupements phosphates étaient fortement hydratés, mais de façon moins localisée que dans le grand sillon [Eisenstein et Shakked, 1995, Schneider *et al.*, 1995]. L'hydratation de la molécule

d'ARNt a été étudiée de façon théorique par dynamique moléculaire en solvant explicite par Auffinger et Westhof [1997], en terme de pourcentage de liaisons hydrogène et de temps de résidence des molécules d'eau. Cette étude a montré que les groupements phosphates étaient très accessibles au solvant, et que les bases G étaient particulièrement accessibles au solvant dans le grand sillon alors que les bases C étaient accessibles dans le petit sillon. Les auteurs de cette étude ont mis en évidence que la paire wobble était fortement hydratée, car les interactions entre les deux bases sont plus faibles et les groupements donneurs et accepteurs de liaisons hydrogène plus accessibles au solvant. Au niveau des bases de la boucle, cette étude a souligné la forte hydratation des bases modifiées  $\Psi$ 6 et  $m^1G11$  par de longs temps de résidence de molécules d'eau. La base U7 a également été remarquée fortement hydratée, au niveau de son groupement phosphate et du grand sillon. Le triplet de l'anticodon (8–10) ne montre pas particulièrement de longs temps de résidence de molécules d'eau [Auffinger et Westhof, 1997].

Le paramètre qu'on a calculé pour mesurer l'hydratation lors de nos simulations n'est ni un temps de résidence ni un pourcentage de liaison hydrogène, mais il nous renseigne sur les sites préférentiels d'hydratation autour de la molécule d'un point de vue électrostatique. Nous avons représenté sur la figure 5.13 le pourcentage d'hydratation maximum obtenu sur toute la simulation avec les pseudo-particules de solvant PPP, pour les différents nucléotides de l'ARNt, pour les groupements phosphates et les bases dans le grand et le petit sillon. La double hélice paraît en moyenne plus fortement hydratée que la partie en boucle, surtout dans le grand et le petit sillon. On remarque que le grand sillon de la double hélice est très fortement hydraté, ce qui est en accord avec les résultats expérimentaux sur les ADN-A [Eisenstein et Shakked, 1995], et en particulier au niveau des bases G, comme dans la simulation de la même molécule en solvant explicite [Auffinger et Westhof, 1997]. Cependant, on ne retrouve pas avec notre modèle de solvant de site préférentiel d'hydratation dans le petit sillon des bases C. La paire wobble G4–U14 apparaît très fortement hydratée comme dans la simulation en solvant explicite [Auffinger et Westhof, 1997], à la fois dans le grand et le petit sillon. Une étude de l'hydratation des différentes paires de bases présentes dans les double hélices d'ARN [Auffinger et Westhof, 1998a] avait également souligné la forte hydratation des paires G–U en général, et l'avait expliquée par la présence de cavités à la surface des sillons au niveau de cette paire inhabituelle, formant des poches d'hydratation. Nous avons en effet vu que les paramètres hélicoïdaux de cette paire pendant notre simulation étaient très différents des autres bases, ce qui pourrait la rendre particulièrement accessible aux particules de solvant. Les bases 8 à 12 de la partie en boucle, fortement empilées, paraissent peu hydratées dans le petit sillon, et légèrement plus dans le grand sillon, à part pour l'anticodon GUC qui est peu hydraté dans les deux sillons, comme observé dans la simulation en solvant explicite [Auffinger et Westhof,

1997]. Ce triplet de bases est pourtant particulièrement accessible au solvant, mais aucune molécule d'eau spécifique ne se lierait fortement à une de ces bases. Les groupements phosphates de l'anticodon sont au contraire particulièrement hydratés, ce qui n'avait pas été clairement observé lors de la simulation en solvant explicite. L'hydratation de la base m<sup>1</sup>G11 est très similaire à celle des bases G et ne se distingue pas particulièrement, mais on retrouve pendant notre simulation que les bases Ψ6 et U7 sont fortement hydratées, comme en solvant explicite [Auffinger et Westhof, 1997], surtout au niveau du petit et du grand sillon, et dans une moindre mesure au niveau de leur groupement phosphate.

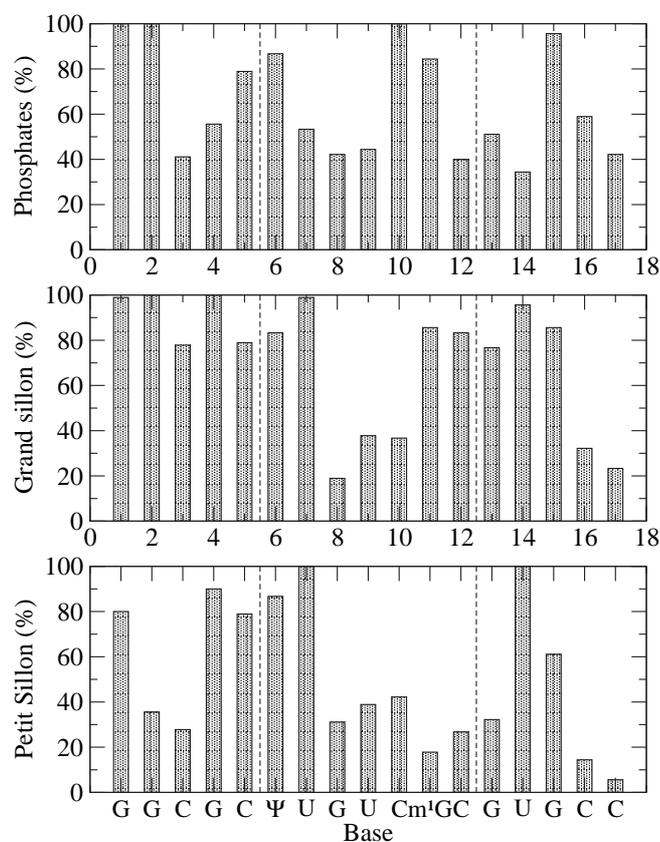


FIG. 5.13 : Pourcentage d'hydratation de l'ARNt pour les groupements phosphates, les atomes donneurs et accepteurs de liaisons hydrogène des bases du grand sillon, et ceux du petit sillon.

Le manque de résultats connus, surtout expérimentaux, sur les sites préférentiels d'hydratation de cette molécule d'ARNt dont la séquence est très particulière nous empêche de valider ou de rejeter certains des sites fortement hydratés observés avec nos pseudo-particules de solvant. Cependant, il est encourageant de voir que l'on retrouve certains des sites d'hydratation observés en solvant explicite, et, pour la partie de la molécule en double hélice, que nous trouvons des sites d'hydratation en accord avec

les études cristallographiques plus générales sur les ADN-A et sur la paire wobble des ARN. Ainsi, même sans pouvoir reproduire la géométrie des liaisons hydrogène entre le soluté et le solvant, nous pouvons encore une fois observer avec notre modèle de solvant des sites préférentiels d'hydratation globalement en accord avec les résultats expérimentaux et théoriques.

#### 5.2.4 Conclusion de la simulation de l'ARNt

L'étude de la dynamique de la boucle anticodon de l'ARNt avec le solvant PPP a montré que nous obtenions non seulement une trajectoire stable sur environ une nanoseconde, mais aussi des paramètres structuraux proches de ceux des simulations en solvant explicite. L'énergie libre électrostatique de solvatation calculée avec notre modèle suit assez bien celle calculée par les méthodes implicites, avec une très bonne corrélation, et nous obtenons également des propriétés d'hydratation souvent en accord avec celles observées en solvant explicite ou par les méthodes expérimentales.

Ces résultats sont très encourageants, car on sait que cette molécule est particulièrement sensible au solvant et aux interactions électrostatiques, et nous avons obtenu une structure stable lors de la dynamique avec le solvant PPP avec pourtant une coupure des interactions électrostatiques et de Lennard-Jones. Cela suggère que l'utilisation d'une distance de coupure pour ces interactions n'entraîne pas d'erreurs importantes avec notre modèle. En ce qui concerne les détails structuraux et d'hydratation obtenus avec le solvant PPP, le peu de résultats théoriques et surtout expérimentaux disponibles pour cette molécule nous empêche d'être complètement satisfaits. L'étude de double hélices d'ADN, pour lesquelles plus de résultats sont établis, nous permettra de mieux comparer les différents paramètres analysés.

### 5.3 Résultats obtenus sur les double hélices d'ADN

#### 5.3.1 Analyse structurale

##### Analyse du RMS

Les évolutions des RMS des atomes lourds des ADN-B1 et A1 par rapport à leurs structures initiales respectives sont représentées sur la figure 5.14. Le RMS de l'ADN-B1 par rapport à sa structure initiale augmente fortement pendant les premières 150 ps de

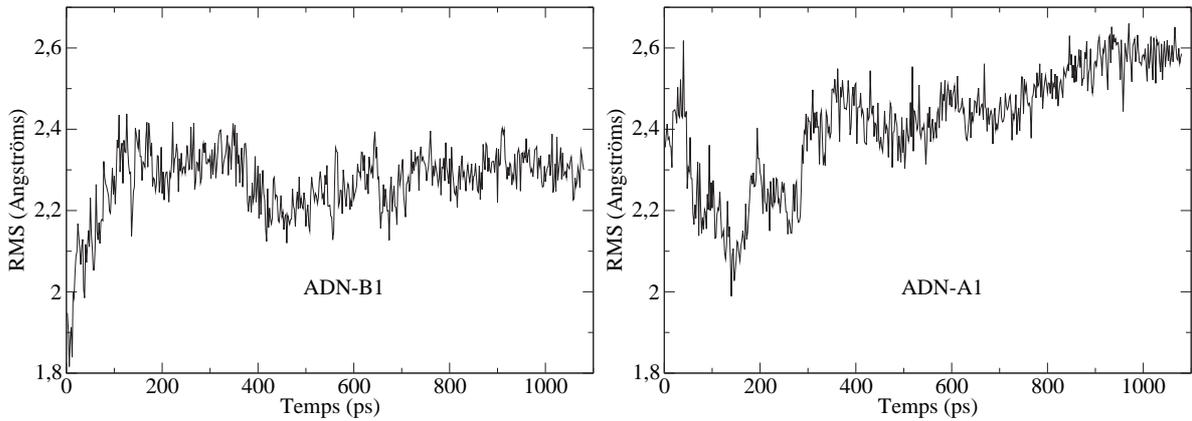


FIG. 5.14 : Évolution du RMS des atomes lourds des ADN-B1 (à gauche) et A1 (à droite) par rapport à leur structure initiale pendant la simulation.

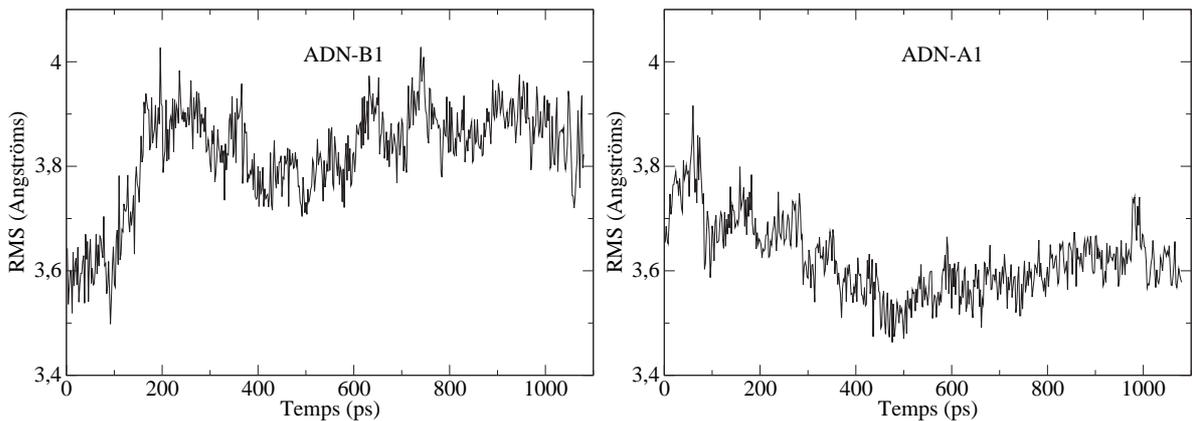


FIG. 5.15 : Évolution du RMS des atomes lourds de l'ADN-B1 (à gauche) par rapport à la structure initiale de l'ADN-A1 et inversement pour l'ADN-A1 (à droite).

simulation, puis se stabilise vers  $2,35 \text{ \AA}$ , redescend légèrement après 350 ps de simulation, et se stabilise autour de  $2,3 \text{ \AA}$  jusqu'à la fin de la simulation. La structure de l'ADN-B1 semble donc très stable pendant toute la simulation après les premières 150 ps. Le RMS de l'ADN-A1 par rapport à sa structure initiale est plus élevé que celui de l'ADN-B1. Il diminue pendant les premières 150 ps de dynamique, puis augmente par paliers pour se stabiliser autour de  $2,45 \text{ \AA}$  après 350 ps. Enfin, il augmente encore légèrement et se stabilise pendant les dernières 200 ps de simulation autour de  $2,55 \text{ \AA}$ . L'ADN-A1 semble donc adopter plusieurs structures assez différentes pendant la simulation, plus ou moins éloignées de la structure initiale. Nous avons aussi représenté sur la figure 5.15 les évolutions des RMS des atomes lourds de l'ADN-B1 par rapport à la structure initiale de l'ADN-A1, et inversement de l'ADN-A1 par rapport à la structure initiale de l'ADN-B1, afin de voir si les double hélices d'ADN transitaient vers des formes B ou A pendant

les simulations. L'évolution du RMS de l'ADN-B1 par rapport à la structure initiale de l'ADN-A1 est assez similaire à celle par rapport à sa structure initiale. Le RMS se stabilise dès 200 ps autour de 3,85 Å. La structure de l'ADN-B1 reste donc plus proche de l'ADN-B que de l'ADN-A. Au contraire, le RMS de l'ADN-A1 par rapport à la structure initiale de l'ADN-B1 diminue pendant 500 ps, et semble se stabiliser autour de 3,6 Å jusqu'à la fin de la simulation. L'ADN-A1, même si sa structure pendant la dynamique reste assez proche de sa structure initiale, semble adopter une structure qui se rapproche légèrement de celle de l'ADN-B. L'étude des paramètres hélicoïdaux nous permettra de confirmer ou non que nous obtenons bien une structure hybride entre la forme A et la forme B.

Nous avons représenté sur la figure 5.16 les structures tridimensionnelles des ADN-B1 et A1 à la fin de la dynamique. On remarque que la structure de l'ADN-B1 est assez proche d'une double hélice de forme B, même si certaines paires de bases ne sont plus très parallèles. La structure de l'ADN-A1 semble plus étirée qu'une forme A classique, l'hélice semble plus fine, avec un Rise plus élevé, même si les paires de bases restent inclinées comme pour une forme A.

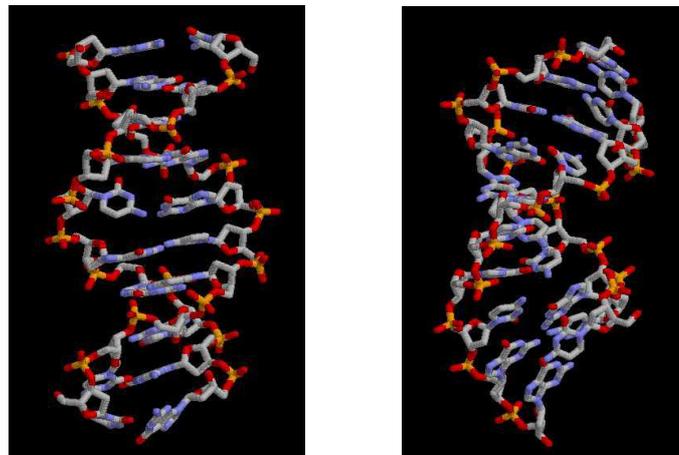


FIG. 5.16 : Structures tridimensionnelles après 1080 ps de simulation de l'ADN-B1 (à gauche) et de l'ADN-A1 (à droite).

L'évolution du RMS des atomes lourds de l'ADN-B2 par rapport à la structure initiale pendant la dynamique est représentée sur la figure 5.17. Le RMS, d'abord autour de 2,1 Å, augmente après 200 ps de dynamique pour se stabiliser autour de 2,3 Å pendant environ 500 ps. Puis il augmente à nouveau après 750 ps de simulation, et fluctue légèrement avant de se stabiliser vers 2,55 Å jusqu'à la fin de la simulation. La structure n'est donc pas figée pendant toute la durée de la simulation, et le RMS passe par différents paliers pour lesquels la structure paraît stable.

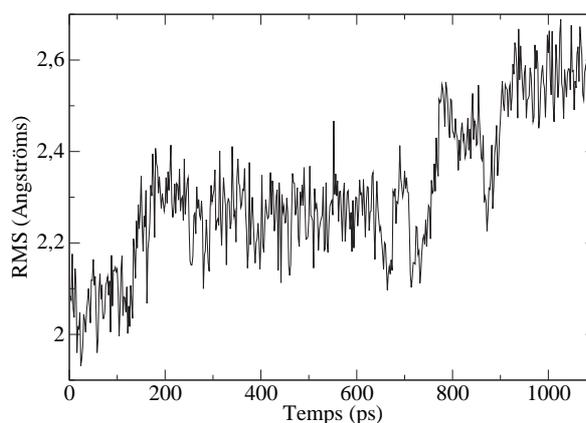


FIG. 5.17 : Évolution du RMS des atomes lourds de l'ADN-B2 par rapport à la structure initiale pendant la simulation.

### Analyse des paramètres hélicoïdaux avec Curves

#### •Les ADN-B1 et A1

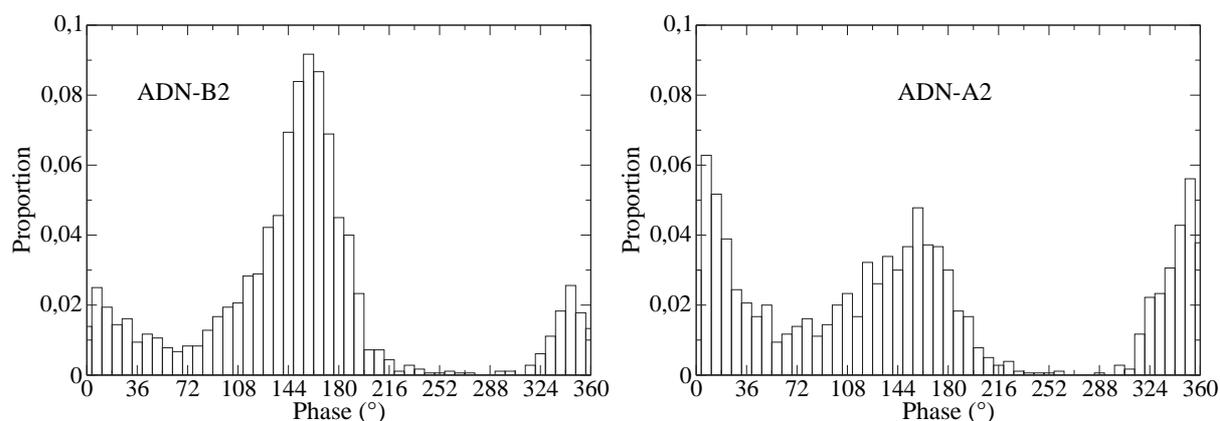


FIG. 5.18 : Répartition des nucléotides en fonction de la phase de leur sucre en degrés pendant les simulations des ADN-B1 et A1.

Les répartitions des phases des sucres des nucléotides pendant les simulations des ADN-B1 et A1 sont représentées sur la figure 5.18. On remarque que, pour l'ADN-B1, les sucres sont en majorité en conformation C2'-endo (Sud), conformation préférentielle des ADN-B, même si le pic de répartition est assez large. Le pic de répartition en C3'-endo, caractéristique des ADN-A, est non négligeable. Pour l'ADN-A1, les deux pics de répartition des sucres au niveau des deux conformations Sud et Nord sont presque équivalents, avec une légère préférence pour la conformation C3'-endo (Nord). Les phases de la structure initiale, et les phases des sucres des nucléotides pendant les deux simulations sont précisées sur le tableau 5.2. On remarque de fortes similitudes de la conformation

majoritaire des sucres de plusieurs nucléotides entre l'ADN-B1 et l'ADN-A1, suggérant que les deux double hélices se rapprochent localement. Ainsi, les quatre premières paires de bases, de C1–G20 à C4–G17, ainsi que la paire C8–G13, ont des sucres dans une conformation proche dans les deux simulations B1 et A1. La conformation des sucres pendant la simulation de l'ADN-A1, répartis entre C3'-endo et C2'-endo, tendrait à nous faire penser une nouvelle fois que la structure obtenue serait hybride entre la forme A et la forme B, confirmant l'observation des RMS et des structures tridimensionnelles.

## ADN-B1

<i>nuc.</i>	$P_0(^{\circ})$	$P (^{\circ})$	<i>conf.</i>	<i>nuc.</i>	$P_0(^{\circ})$	$P (^{\circ})$	<i>conf.</i>
<b>C1</b>	138	14	C3'-endo	<b>G20</b>	233	166	C2'-endo
<b>C2</b>	217	166	C2'-endo	<b>G19</b>	145	151	C2'-endo
<b>G3</b>	175	346	C2'-exo	<b>C18</b>	118	151	C2'-endo
<b>C4</b>	182	166	C2'-endo	<b>G17</b>	160	137	C1'-exo
<b>C5</b>	98	7	C3'-endo	<b>G16</b>	107	158	C2'-endo
<b>G6</b>	158	187	C3'-exo	<b>C15</b>	152	86	O1'-endo
<b>G7</b>	177	173	C2'-endo	<b>C14</b>	147	166	C2'-endo
<b>C8</b>	132	166	C2'-endo	<b>G13</b>	89	338	C2'-exo
<b>G9</b>	160	140	C1'-exo	<b>C12</b>	79	130	C1'-exo
<b>G10</b>	212	115	C1'-exo	<b>C11</b>	189	29	C3'-endo

## ADN-A1

<i>nuc.</i>	$P_0(^{\circ})$	$P (^{\circ})$	<i>sucré</i>	<i>nuc.</i>	$P_0(^{\circ})$	$P (^{\circ})$	<i>sucré</i>
<b>C1</b>	5	43	C4'-exo	<b>G20</b>	162	122	C1'-exo
<b>C2</b>	13	158	C2'-endo	<b>G19</b>	355	137	C1'-exo
<b>G3</b>	8	349	C2'-exo	<b>C18</b>	16	158	C2'-endo
<b>C4</b>	11	173	C2'-endo	<b>G17</b>	6	194	C3'-exo
<b>C5</b>	16	7	C3'-endo	<b>G16</b>	358	353	C2'-exo
<b>G6</b>	358	7	C3'-endo	<b>C15</b>	16	338	C2'-exo
<b>G7</b>	6	349	C2'-exo	<b>C14</b>	11	14	C3'-endo
<b>C8</b>	16	173	C2'-endo	<b>G13</b>	8	324	C2'-exo
<b>G9</b>	355	122	C1'-exo	<b>C12</b>	13	22	C3'-endo
<b>G10</b>	161	79	O1'-endo	<b>C11</b>	6	122	C1'-exo

TAB. 5.2 : Phase initiale ( $P_0$ ), phase ( $P$ ) en degrés et conformation majoritaire des sucres des nucléotides pendant les simulations des ADN-B1 et A1.

Les répartitions de quelques-uns des paramètres axe-paire de bases pendant la simulation des ADN-B1 et A1 sont représentées respectivement sur les figures 5.19 et 5.20. Le  $X_{disp}$  de l'ADN-B1 reste très proche de l'origine, alors que le  $Y_{disp}$  est plus étalé, surtout pour la seconde moitié de la séquence de G6–C15 à G10–C11. Le  $X_{disp}$  de l'ADN-A1 est moins négatif que celui d'un ADN-A canonique ( $-5, 3 \text{ \AA}$ ) ainsi que du  $X_{disp}$  moyen de la structure initiale ( $-4, 54 \text{ \AA}$ ). Le  $Y_{disp}$  est quant à lui autour de 0 pour la première moitié de l'ADN, de C1–G20 à C4–G17, et plus étalé et légèrement négatif pour la seconde moitié. Pour les deux formes d'ADN, on observe une asymétrie de la structure, la première moitié semblant plus figée que la seconde. On observe globalement que l'ADN-A1 a un paramètre  $X_{disp}$  entre ceux des ADN-B et A canoniques.

Les répartitions des paramètres intra-paire de bases pendant les simulations sont représentées sur la figure 5.21 pour l'ADN-B1 et sur la figure 5.22 pour l'ADN-A1. Pour l'ADN-B1, le *Stretch* est en moyenne nul, ainsi que le *Buckle* et l'*Opening*, à part pour les paires extrêmes et pour la paire C5–G16 qui montre un *Stretch* légèrement positif ( $0,5 \text{ \AA}$ ) et un *Opening* légèrement négatif ( $-10^\circ$ ). Notons que cette paire C5–G16 est à la jonction d'un pas CpC et GpG. Le *Propeller Twist* est plus variable, le pic de répartition est assez étalé, et seules les paires C3–G18 et C4–G17 ont leur maximum autour de la valeur canonique des ADN-B de  $-13,3^\circ$ . Les autres paires sont donc assez mobiles en rotation autour de l'axe  $y$ . Dans le cas de l'ADN-A1, les pics de répartition sont en moyenne beaucoup plus étalés. Seul l'*Opening* reste en majorité autour de  $0^\circ$ , alors qu'il était légèrement positif ( $4,35^\circ$ ) pour la structure initiale. Le *Stretch* est assez étalé mais reste autour de l'origine. Quant au *Buckle*, il est positif pour la première moitié de la séquence (de C1–G20 à C4–G17), et négatif pour la seconde moitié (de G6–C15 à G9–C12), alors que cette tendance n'avait été observée que pour les extrémités de l'ADN-B1. Le *Propeller Twist* est aussi très variable pour l'ADN-A1, et montre quelques pics de répartition majoritaires à la valeur des ADN-B pour les paires C1–G20, G7–C14 et C9–G12. Notons que la structure initiale de l'ADN-A1 avait un *Propeller Twist* moyen de  $-19,98^\circ$ , alors qu'il était de  $-18,44^\circ$  pour l'ADN-B1, et dans les deux cas il était très variable le long de la séquence. Cette séquence riche en pas GpG (ou CpC) montre donc une grande variabilité des paramètres intra-paire de bases, ce qui confirmerait la plus grande flexibilité des pas GpG observée expérimentalement par Ng et Dickerson [2002].

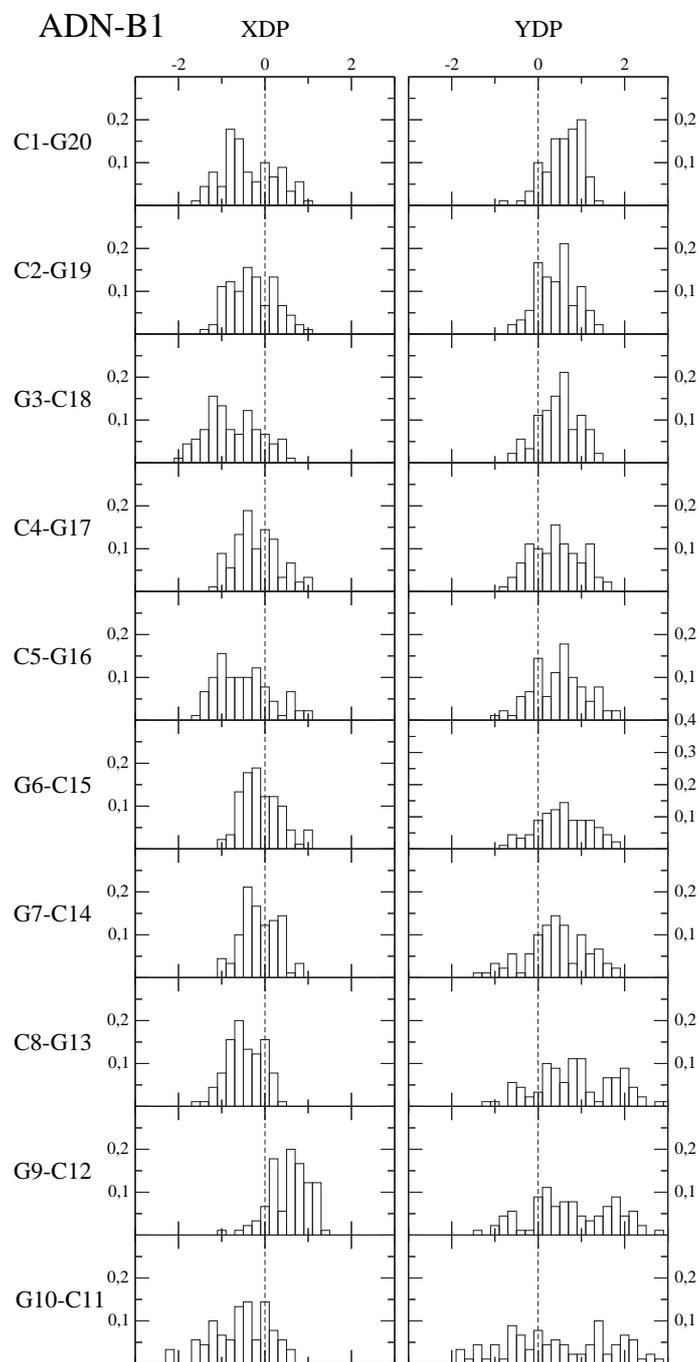


FIG. 5.19 : Répartition des paramètres axe-paires de bases :  $Xdisp$  ( $XDP$ ) et  $Ydisp$  ( $YDP$ ), en Å pour les différentes paires de bases de l'ADN-B1. Les valeurs de référence des ADN-B sont représentées en traits pointillés.

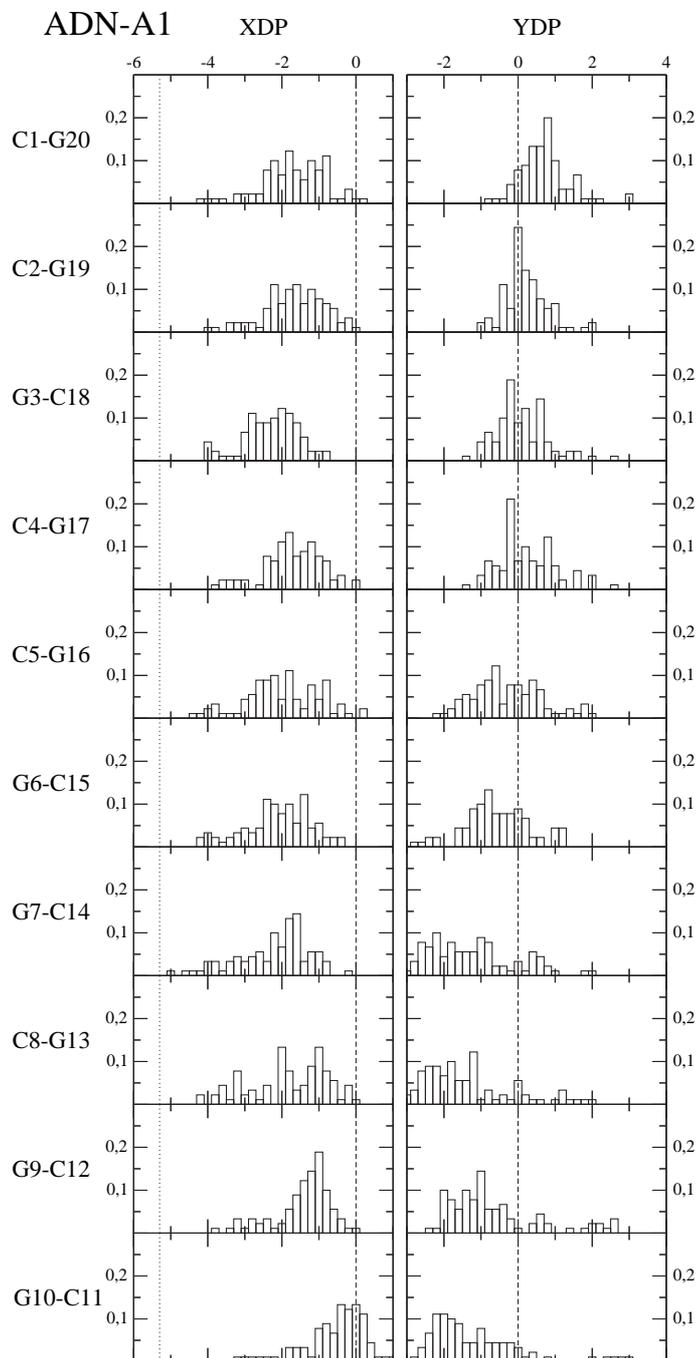


FIG. 5.20 : Répartition des paramètres axe-paires de bases :  $Xdisp$  ( $XDP$ ) et  $Ydisp$  ( $YDP$ ) en Å, pour les différentes paires de bases de l'ADN-A1. Les valeurs de référence des ADN-A et B sont représentées respectivement en pointillés et en traits pointillés.

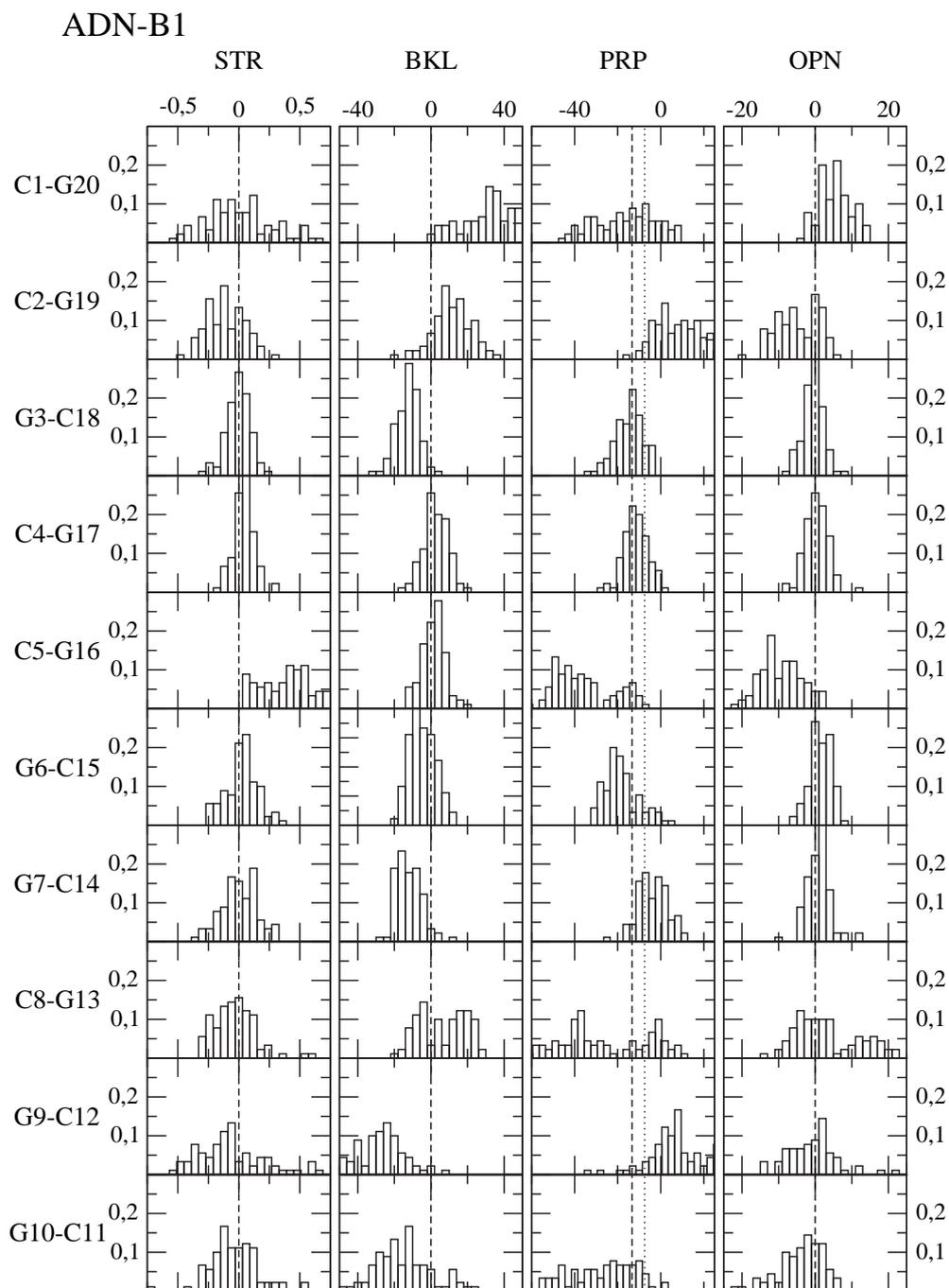


FIG. 5.21 : Répartition des paramètres intra-paire de bases : Stretch (STR) en Å, Buckle (BKL), Propeller Twist (PRP) et Opening (OPN) en degrés pour les différentes paires de bases de l'ADN-B1. Les valeurs de référence des ADN-A et B sont représentées respectivement en pointillés et en traits pointillés.

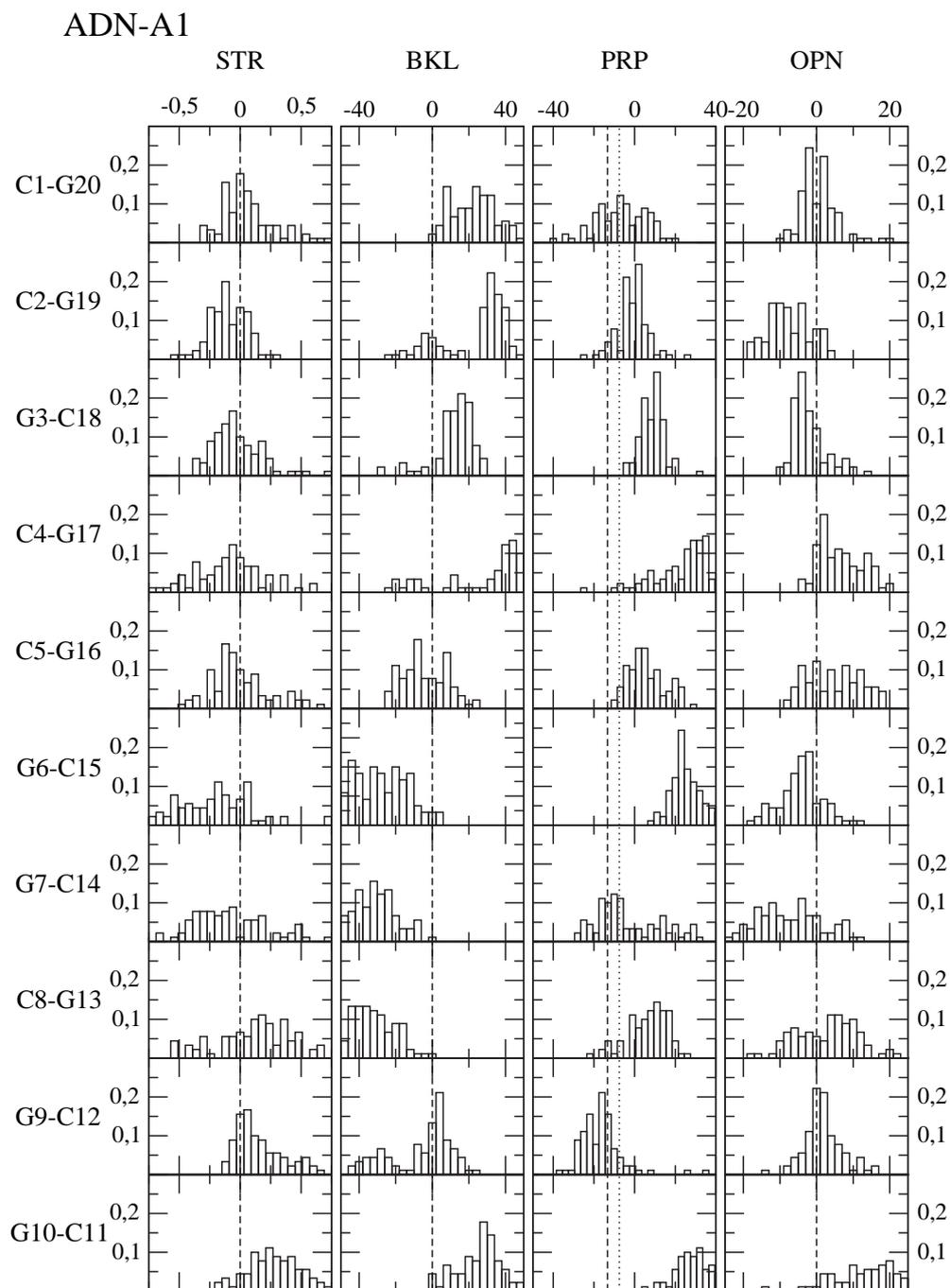


FIG. 5.22 : Répartition des paramètres intra-paire de bases : Stretch (STR) en Å, Buckle (BKL), Propeller Twist (PRP) et Opening (OPN) en degrés pour les différentes paires de bases de l'ADN-A1. Les valeurs de référence des ADN-A et B sont représentées respectivement en pointillés et en traits pointillés.

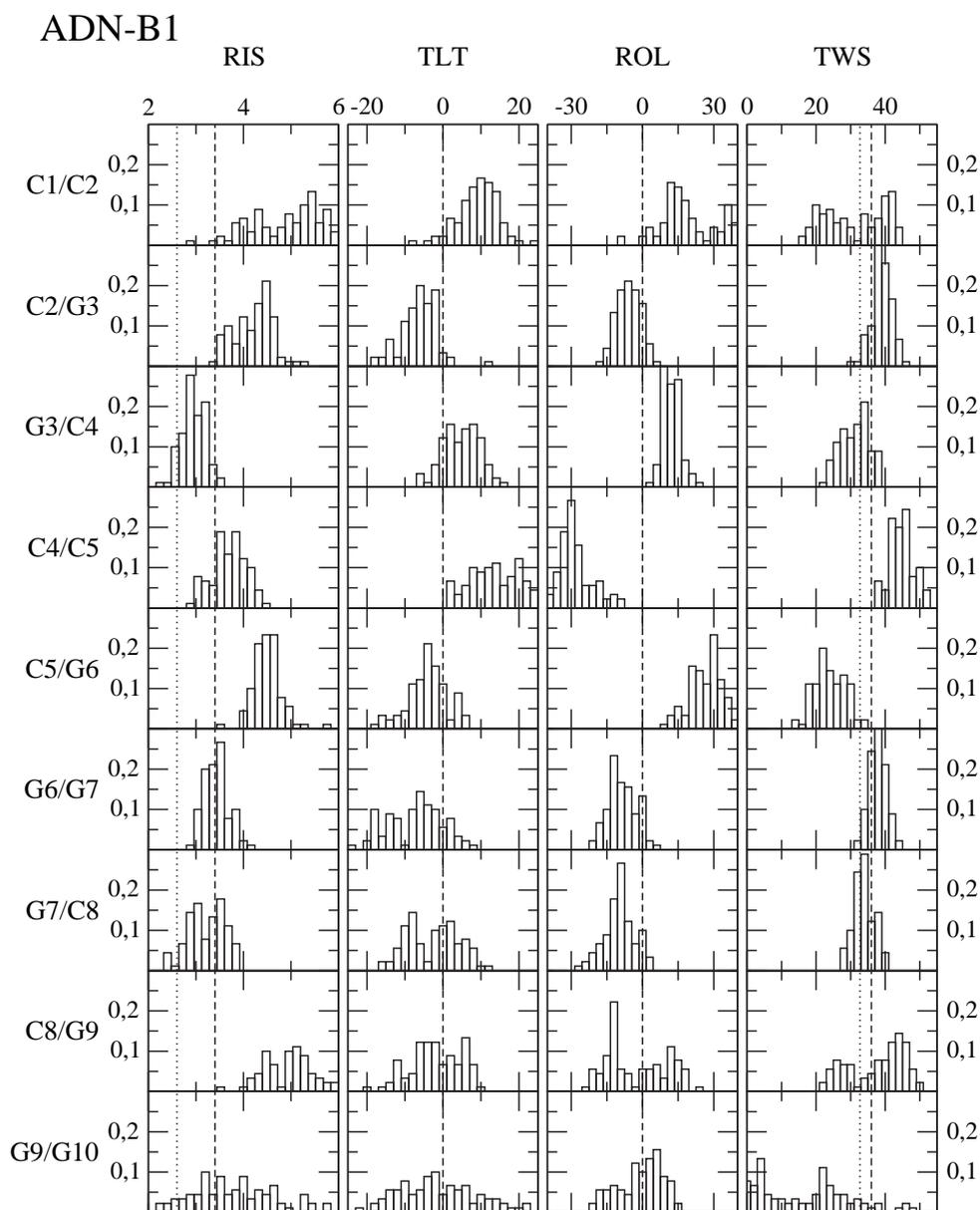


FIG. 5.23 : Répartition des paramètres inter-paires de bases : Rise (RIS) en Å, Tilt (TLT), Roll (ROL) et Twist (TWS) en degrés pour les différentes inter-paires de bases de l'ADN-B1. Les valeurs de référence des ADN-A et B sont représentées respectivement en pointillés et en traits pointillés.

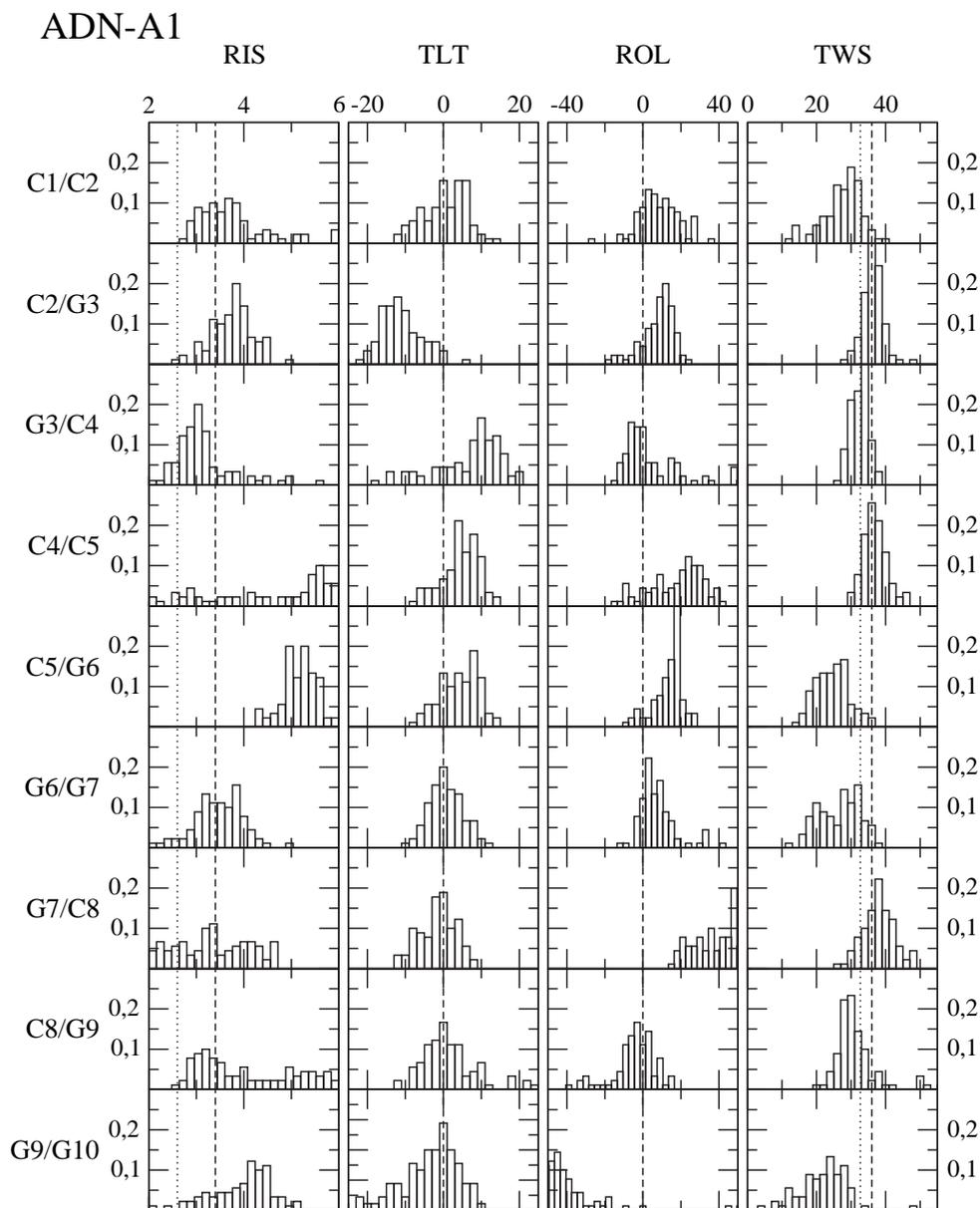


FIG. 5.24 : Répartition des paramètres inter-paires de bases : Rise (RIS) en Å, Tilt (TLT), Roll (ROL) et Twist (TWS) en degrés pour les différentes inter-paires de bases de l'ADN-A1. Les valeurs de référence des ADN-A et B sont représentées respectivement en pointillés et en traits pointillés.

Les répartitions des paramètres inter-paires de bases pendant les simulations des ADN-B1 et A1 sont représentées respectivement sur les figures 5.23 et 5.24. On observe certaines similarités, mais aussi certaines différences, entre les deux simulations. Ainsi, le pas CpG semble dans les deux cas augmenter le Rise et diminuer le Twist (surtout pour le pas nucléotidique C5/G6). Dans les deux simulations, le pas CpC semble plutôt augmenter le Twist, et le pas GpC diminue le Rise. Le pas GpG montre une répartition élargie à la fois pour l'ADN-B1 et l'ADN-A1, ce qui confirmerait encore une fois que c'est le pas le plus flexible [Ng et Dickerson, 2002]. Le Roll varie fortement pour les deux formes d'ADN entre  $-40^\circ$  et  $+40^\circ$ , mais on n'observe pas de similitude entre les deux simulations B1 et A1. Pour l'ADN-B1, le Roll s'éloigne fortement de la valeur canonique au niveau des pas C4/C5 ( $-30^\circ$ ) et C5/G6 ( $+30^\circ$ ), alors que pour l'ADN-A1 c'est au niveau des pas G7/C8 ( $+45^\circ$ ) et G9/G10 ( $-40^\circ$ ) que le Roll s'éloigne le plus de la valeur de référence. Le Rise de l'ADN-A1, qui était à l'origine en moyenne de 2,66 Å, augmente pendant la simulation et est en moyenne beaucoup plus proche d'une valeur caractéristique d'ADN-B de 3,4 Å.

L'étude des paramètres hélicoïdaux des ADN-B1 et A1 confirme l'étude des RMS, c'est-à-dire que la simulation de cette séquence en solution semble produire des structures plus proches d'un ADN-B que d'un ADN-A. Alors que la structure de l'ADN-B1 reste stable autour d'une forme B pendant la nanoseconde de dynamique, la structure de l'ADN-A1 paraît se rapprocher d'une structure B d'après le RMS par rapport à la structure initiale de l'ADN-B1, la conformation des sucres de ses nucléotides, et la répartition du Xdisp et du Rise pendant la simulation. Rappelons que des formes de double hélices hybrides entre A et B ont déjà été observées dans des simulations en solvant explicite [Sproun *et al.*, 1998], ainsi que par des méthodes expérimentales [Steff *et al.*, 2001, Ng et Dickerson, 2002]. De plus, nous retrouvons avec notre modèle de solvant des résultats expérimentaux sur la plus grande flexibilité des pas GpG [Ng et Dickerson, 2002].

### •L'ADN-B2

Nous avons représenté sur la figure 5.25 la répartition des phases des sucres des nucléotides de l'ADN-B2 pendant toute la simulation. On remarque que les sucres sont majoritairement en conformation C2'-endo ( $144^\circ$ – $180^\circ$ ) ou Sud, conformation préférentielle pour les ADN-B. La courbe de répartition est assez étendue et dépasse assez largement vers la conformation C1'-exo ( $108^\circ$ – $144^\circ$ ). Cette observation est en accord avec la simulation de la même molécule en solvant explicite [Young *et al.*, 1997] ainsi qu'avec l'étude de RMN de cette double hélice en solution [Lane *et al.*, 1991]. On remarque aussi un léger pic pour les conformations C3'-endo ( $0$ – $36^\circ$ ), conformation des ADN-A, mais qui reste très minoritaire. Le tableau 5.3 précise la phase de la structure initiale, et la valeur moyenne et

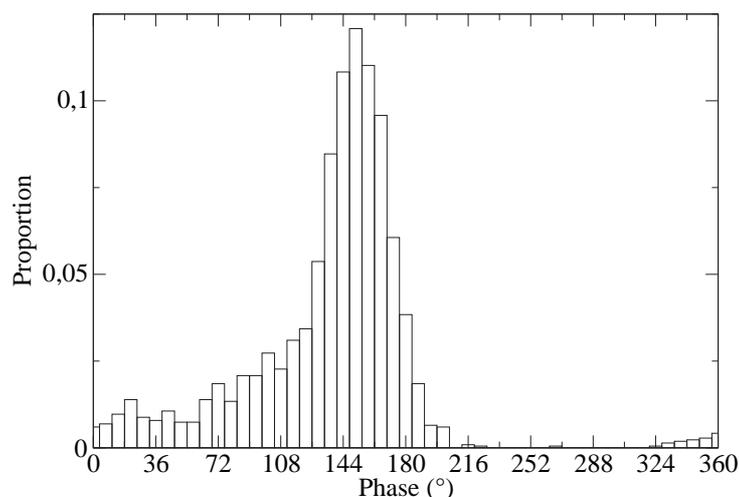


FIG. 5.25 : Répartition des nucléotides en fonction de la phase de leur sucre en degrés pendant la simulation de l'ADN-B2.

<i>nuc.</i>	$P_0(^{\circ})$	$P(^{\circ})$	<i>conformation</i>	<i>nuc.</i>	$P_0(^{\circ})$	$P(^{\circ})$	<i>conformation</i>
<b>C1</b>	141	126 (23)	C1'-exo	<b>G24</b>	31	93 (26)	O1'-endo
<b>G2</b>	156	181 (83)	C3'-exo	<b>C23</b>	157	166 (13)	C2'-endo
<b>C3</b>	133	162 (18)	C2'-endo	<b>G22</b>	156	159 (14)	C2'-endo
<b>G4</b>	154	104 (22)	O1'-endo	<b>C21</b>	104	126 (20)	C1'-exo
<b>A5</b>	160	143 (12)	C1'-exo	<b>T20</b>	136	49 (92)	C4'-exo
<b>A6</b>	147	140 (11)	C1'-exo	<b>T19</b>	115	163 (14)	C2'-endo
<b>T7</b>	130	78 (34)	O1'-endo	<b>A18</b>	164	111 (19)	C1'-exo
<b>T8</b>	132	170 (9)	C2'-endo	<b>A17</b>	159	159 (8)	C2'-endo
<b>C9</b>	161	130 (12)	C1'-exo	<b>G16</b>	160	156 (14)	C2'-endo
<b>G10</b>	152	176 (15)	C2'-endo	<b>C15</b>	90	152 (12)	C2'-endo
<b>C11</b>	168	153 (17)	C2'-endo	<b>G14</b>	164	149 (10)	C2'-endo
<b>G12</b>	148	79 (77)	O1'-endo	<b>C13</b>	146	146 (24)	C2'-endo

TAB. 5.3 : Phase de la structure initiale ( $P_0$ ), phase ( $P$ ) et conformation moyenne des sucres des nucléotides pendant la simulation de l'ADN-B2,  $P \in [0^{\circ}; 360^{\circ}]$ . L'écart-type est indiqué entre parenthèses.

l'écart-type de la phase, ainsi que la conformation du sucre correspondante, pour chaque nucléotide pendant la simulation. La plupart des nucléotides ont un sucre en moyenne autour de C2'-endo ou en C1'-exo, confirmant l'observation précédente. Les nucléotides ont en général une phase proche de celle de la structure initiale, pour laquelle presque tous les nucléotides (à part G24) étaient en conformation C2'-endo. Les nucléotides extrêmes

G12 et G24 ont une phase un peu plus éloignée de la conformation Sud, ainsi que certains autres nucléotides (G4 et T20) se trouvant à la jonction entre les séquences CGCG et AATT, cette jonction étant connue expérimentalement pour entraîner des irrégularités de structure [Drew *et al.*, 1981, Lane *et al.*, 1991].

Quelques-uns des paramètres axe-paires de bases sont représentés sur la figure 5.26. En moyenne, le Xdisp des paires de bases est légèrement négatif entre 0 et  $-1 \text{ \AA}$ , alors que le Xdisp d'un ADN-B canonique est nul. La simulation en solvant explicite de la même molécule avait déjà mis en évidence ce phénomène, et le Xdisp moyen était alors de  $-0,92 \text{ \AA}$  [Young *et al.*, 1997]. Le Ydisp fluctue aussi légèrement vers les valeurs négatives, sans que l'on trouve un lien évident avec la séquence. Les valeurs de ces paramètres axe-paire de bases montrent que la double hélice de l'ADN-B2 pendant la simulation n'est ni tout à fait rectiligne, ni tout à fait symétrique, mais ces irrégularités sont en accord avec les précédentes simulations.

Les répartitions de certains des paramètres intra-paire de bases sont représentées sur la figure 5.27. Le Stretch des paires C–G montre des pics de répartition autour de l'origine beaucoup plus fins que ceux des paires A–T, pour lesquelles le Stretch varie fortement et est parfois légèrement positif en moyenne (mais reste en-dessous de  $0,5 \text{ \AA}$  en moyenne). Les paires C–G sont liées par trois liaisons hydrogène, alors que les paires A–T ne sont liées que par deux liaisons hydrogène, et on comprend que les paires G–C soient plus rigides et que les bases de ces paires soient plus proches que celles des paires A–T pendant la simulation. Le Buckle varie légèrement autour de la valeur canonique de  $0^\circ$ , sauf pour la paire extrême C1–G24 (le Buckle vaut alors environ  $10^\circ$  en moyenne), et les paires A6–T19 (on observe un pic maximum à  $-20^\circ$ ) et C9–G16 ( $-20^\circ$  en moyenne), qui seraient donc légèrement pliées. Le Propeller Twist des différentes paires est très variable selon la séquence, et sa répartition assez étalée. On remarque qu'il est légèrement plus élevé que la valeur canonique des ADN-B ( $-13,3^\circ$ ) en moyenne pour les paires C–G, et plus négatif d'une dizaine de degrés pour les paires A–T, ce qui avait déjà été observé lors de la simulation de l'ADN-B2 en solvant explicite [Young *et al.*, 1997]. Quant à l'Opening, il est majoritairement nul et montre des pics de répartition très fins pour les paires C–G, alors qu'il est beaucoup plus étalé pendant la simulation pour les paires A–T. Ce paramètre a une valeur moyenne de  $-20^\circ$  pour la paire T8–A17, et cette paire serait donc assez ouverte, et presque séparée d'après son Stretch qui atteint  $0,5 \text{ \AA}$  en moyenne. D'après l'étude de ces paramètres intra-paire de bases, on remarque clairement des différences selon la séquence, les paires C–G étant en moyenne beaucoup plus rigides et régulières que les paires A–T.

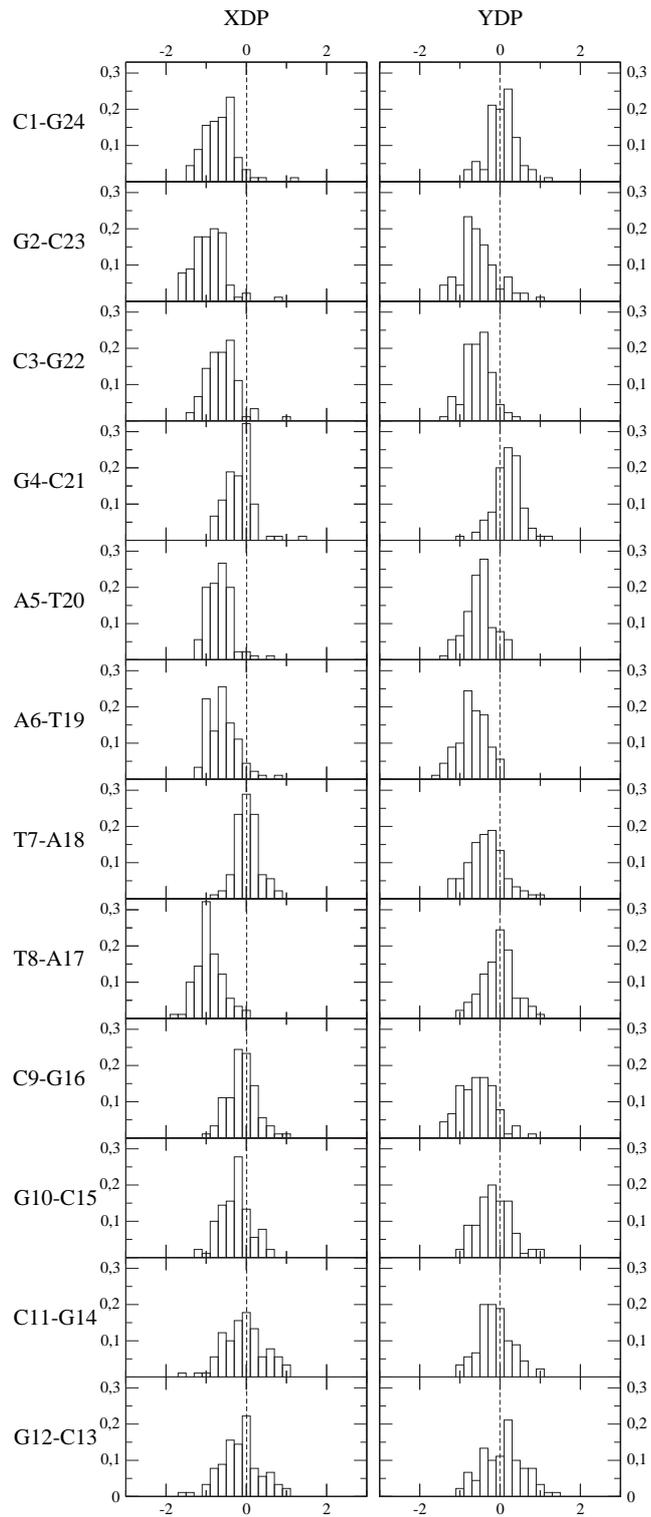


FIG. 5.26 : Répartition des paramètres axe-paires de bases :  $X_{disp}$  ( $XDP$ ) et  $Y_{disp}$  ( $YDP$ ) en Å, pour les différentes paires de bases de l'ADN-B2. Les valeurs de référence des ADN-B sont représentées en traits pointillés.

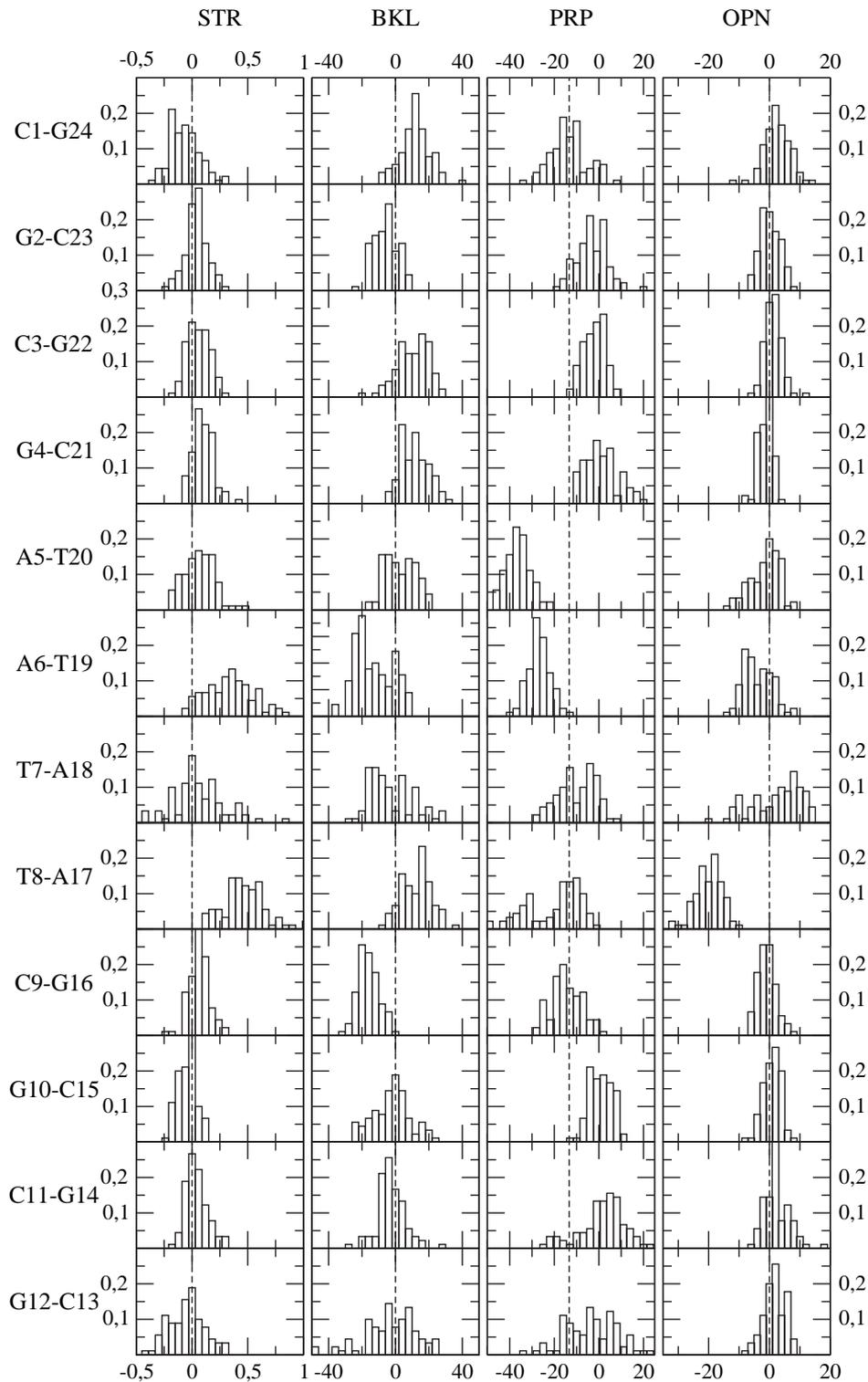


FIG. 5.27 : Répartition des paramètres intra-paire de bases : Stretch ( $STR$ ) en Å, Buckle ( $BKL$ ), Propeller Twist ( $PRP$ ) et Opening ( $OPN$ ) en degrés pour les différentes paires de bases de l'ADN-B2. Les valeurs de référence des ADN-B sont représentées en traits pointillés.

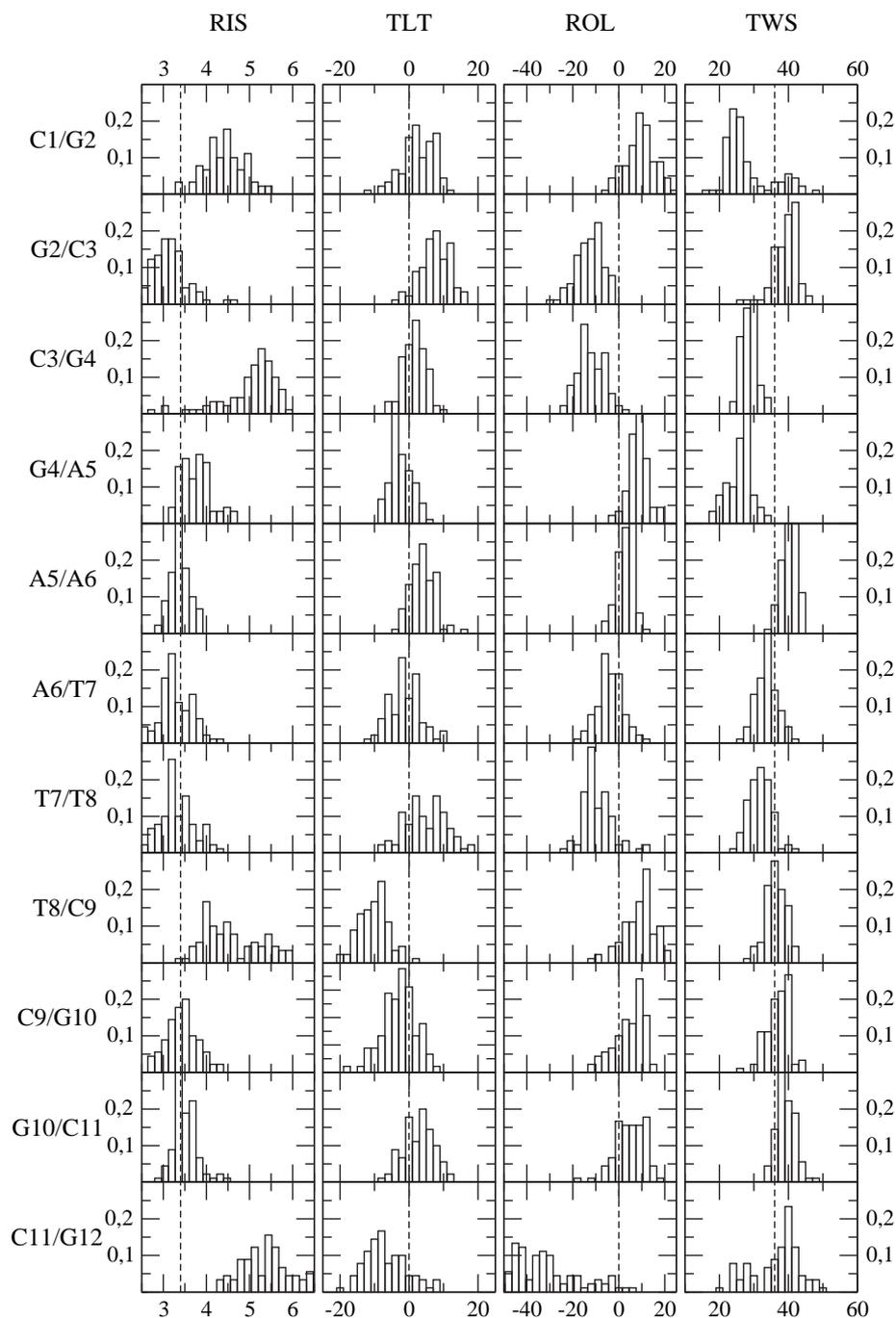


FIG. 5.28 : Répartition des paramètres inter-paires de bases : Rise (RIS) en Å, Tilt (TLT), Roll (ROL) et Twist (TWS) en degrés pour les différentes inter-paires de bases de l'ADN-B2. Les valeurs de référence des ADN-B sont représentées en traits pointillés.

La répartition de paramètres inter-paires de bases de l'ADN-B2 représentée sur la figure 5.28 montre des Rises parfois très élevés, allant jusqu'à 4,5 Å pour le pas nucléotidique C1/G2, et jusqu'à plus de 5 Å en moyenne pour les pas C3/G4 et C11/G12. Pour les paires placées aux extrémités, on comprend que le Rise puisse augmenter pendant la simulation puisque l'on n'a appliqué aucune contrainte. Pour le pas C3/G4, nous avons observé qu'un ion sodium s'est partiellement intercalé entre les deux paires de bases dans le petit sillon, entraînant une augmentation du Rise. Le Tilt reste en moyenne très proche de la valeur de 0° de l'ADN-B canonique, à part à l'extrémité C11/G12, et pour le pas T8/C9 à la jonction entre les séquences AATT et CGCG. On observe que le Roll montre des déplacements par rapport à sa valeur canonique au niveau de la jonction entre CGCG et AATT (le Roll devient positif à chacune des deux jonctions pendant 2 ou 3 pas, à condition de ne pas prendre en compte les paires de bases extrêmes), comme cela avait été observé dans la simulation en solvant explicite de la même molécule [Young *et al.*, 1997]. Le Twist est en moyenne autour de la valeur canonique de 36°, en accord avec les résultats de RMN, mais on n'observe pas de « untwist » global pour toute la séquence comme dans la simulation en solvant explicite où le Twist moyen observé était autour de 34,6° [Arthanari *et al.*, 2003]. Cependant, on remarque dans notre simulation quelques disparités des valeurs du Twist selon la séquence, en particulier pour le premier pas nucléotidique C1/G2 qui semble légèrement déroulé, puis pour les deux pas C3/G4 et G4/A5 qui sont aussi un peu déroulés, contrairement au pas A5/A6 qui est légèrement plus enroulé.

L'étude des paramètres hélicoïdaux de l'ADN-B2 a donc mis en évidence que la double hélice montrait quelques variations pendant la simulation par rapport à une hélice B canonique, mais restait globalement en conformation B. On observe certains effets de la séquence sur les paramètres hélicoïdaux, en particulier à la jonction entre les deux régions CGCG et AATT, qui semble plus flexible. La plupart des variations observées sont en accord avec les simulations en solvant explicite [Young *et al.*, 1997] et les résultats d'expériences de RMN [Lane *et al.*, 1991].

### 5.3.2 Analyse énergétique

La figure 5.29 représente les évolutions de l'énergie totale, de l'énergie potentielle et de l'énergie libre électrostatique de solvation pendant les simulations des ADN-B1 et A1. Alors que l'énergie totale est en moyenne très stable et que sa dérive apparaît toujours négligeable devant sa valeur moyenne, l'énergie potentielle du soluté diminue dans les deux cas. Pour l'ADN-B1, l'énergie libre électrostatique de solvation suit assez symétriquement l'évolution de l'énergie potentielle du soluté, alors que pour l'ADN-A1 l'évolution de l'énergie de solvation ne semble compenser celle de l'énergie potentielle

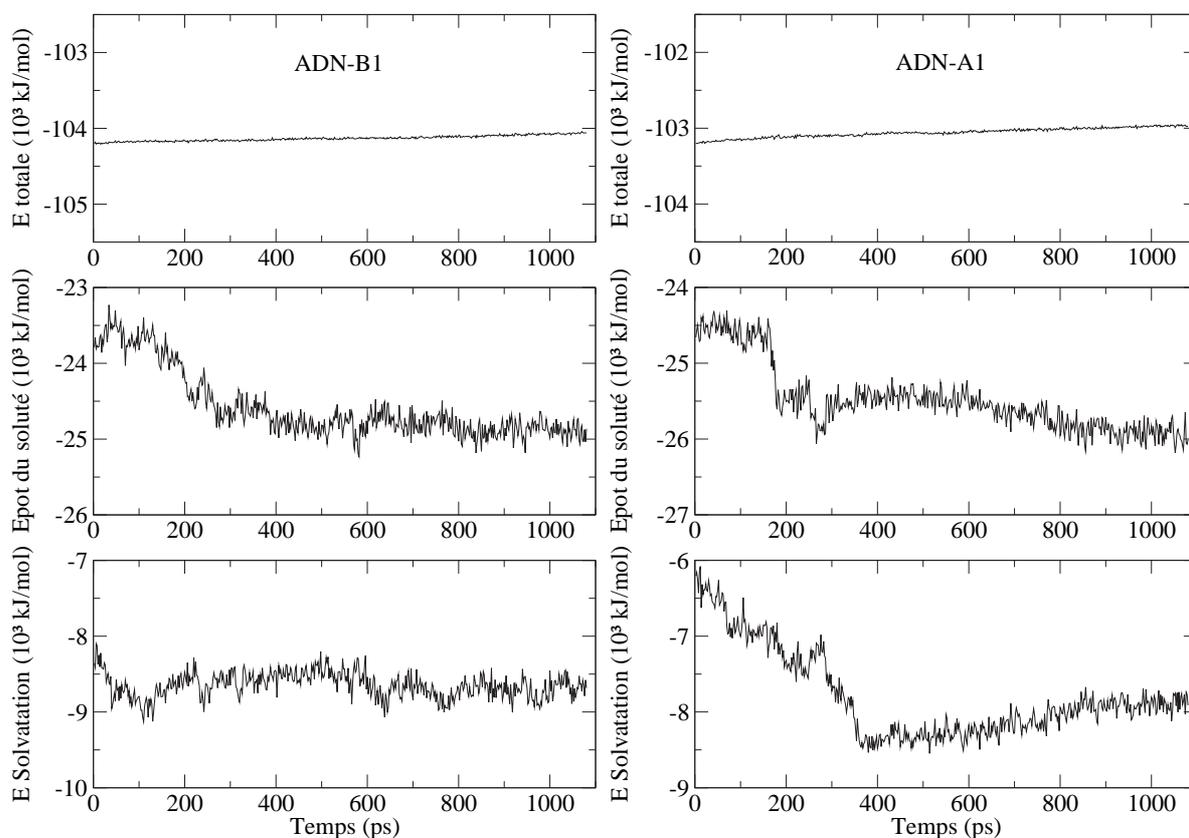


FIG. 5.29 : Évolution des énergies totale, potentielle du soluté et de solvation (en  $10^3$  kJ/mol) pendant les simulations des ADN-B1 (à gauche) et A1 (à droite).

du soluté que vers la fin de la dynamique. Au début de la simulation de l'ADN-A1, une diminution de l'énergie libre de solvation accompagne celle de l'énergie potentielle du soluté. Contrairement aux précédentes simulations des protéines et des autres acides nucléiques avec le solvant PPP, les changements importants de conformation de l'ADN-A semblent entraîner un changement radical des couches d'hydratation, et l'énergie libre de solvation ne compenserait plus du tout les variations de l'énergie potentielle du soluté.

Les évolutions de l'énergie totale, de l'énergie potentielle et de l'énergie libre électrostatique de solvation pendant la simulation de l'ADN-B2 sont représentées sur la figure 5.30. La dérive de l'énergie totale est toujours négligeable, et l'énergie libre de solvation reste plutôt stable malgré les légères variations de l'énergie potentielle du soluté. Les changements de conformation du soluté (composé de l'ADN et des ions sodium) observés d'après l'évolution de l'énergie potentielle n'entraînent apparemment que des changements d'hydratation mineurs, puisque son énergie libre de solvation reste globalement constante. Nous pensons que si les deux courbes d'énergie libre de solvation et d'énergie potentielle du soluté sont moins symétriques dans cette étude que lors des

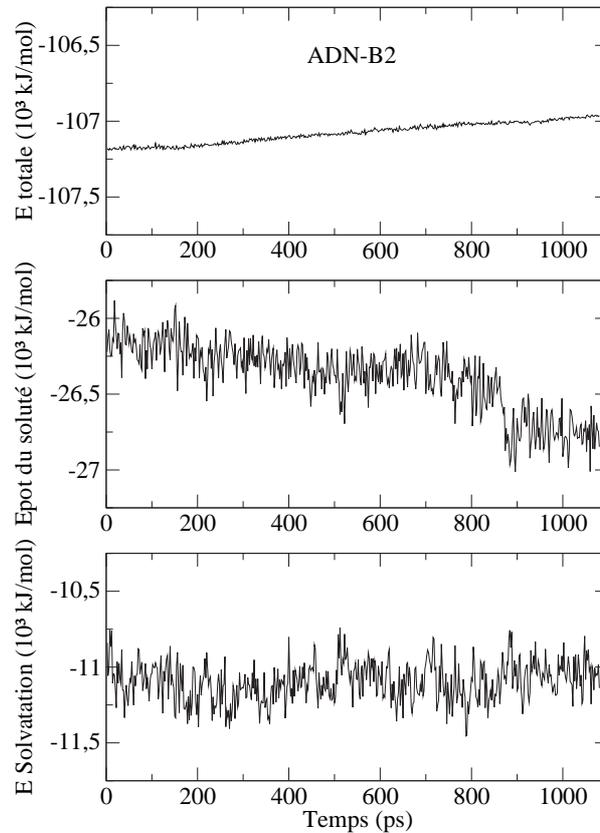


FIG. 5.30 : Évolution des énergies totale, potentielle du soluté et de solvation pendant la simulation de l'ADN-B2 (en  $10^3$  kJ/mol).

simulations de peptides et protéines, c'est peut-être parce que nous prenons ici en compte dans le soluté les ions explicites très mobiles, et que le déplacement d'un ion lié à l'ADN peut entraîner un changement d'énergie potentielle électrostatique du soluté, sans influencer de façon significative sur l'énergie libre de solvation.

La droite de corrélation entre les énergies libres électrostatiques de solvation calculées avec PPP et avec DelPhi pour la simulation de l'ADN-B2 ainsi que les évolutions des deux énergies sont représentées sur la figure 5.31. Comme nous l'avons remarqué sur la figure 5.30, l'énergie de solvation varie peu pendant la simulation, et le solvant PPP ne permet alors pas d'obtenir de très bonnes corrélations avec DelPhi. Le coefficient de corrélation n'est donc pas très proche de un (0,758), ainsi que la pente (0,771), mais l'erreur systématique observée est faible, de 4,4 % en moyenne. Pour le système neutre composé de l'ADN-B2 et des contre-ions, le modèle de solvant PPP ne permet donc pas d'observer des variations de l'énergie libre de solvation aussi précises que celles que nous avons observées pour l'ARNt. Cependant, les deux courbes d'évolution des énergies libres de solvation ont encore une fois des formes assez similaires.

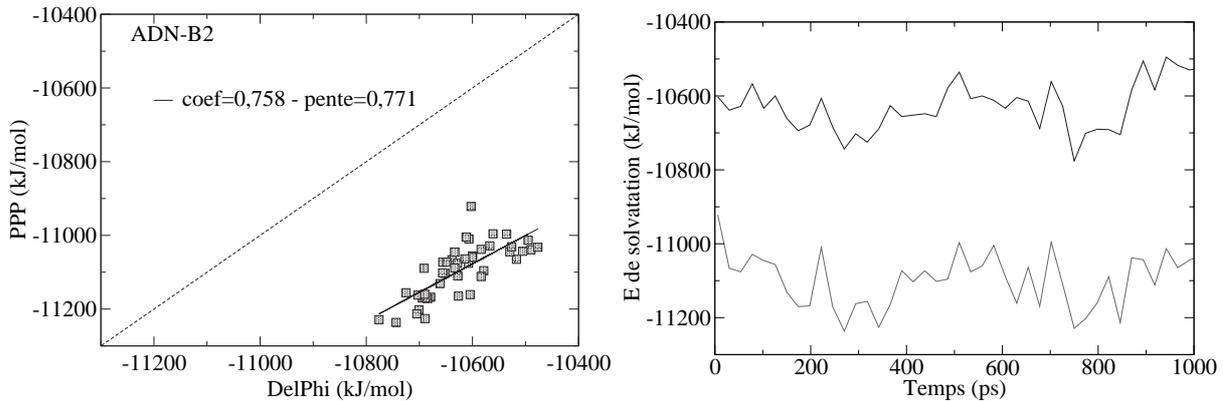


FIG. 5.31 : Corrélation entre l'énergie libre de solvation moyenne de l'ADN-B2 calculée avec PPP et l'énergie de la structure moyenne avec DelPhi (à gauche) et évolution des ces deux énergies (à droite) en gris pour PPP et en noir pour DelPhi.

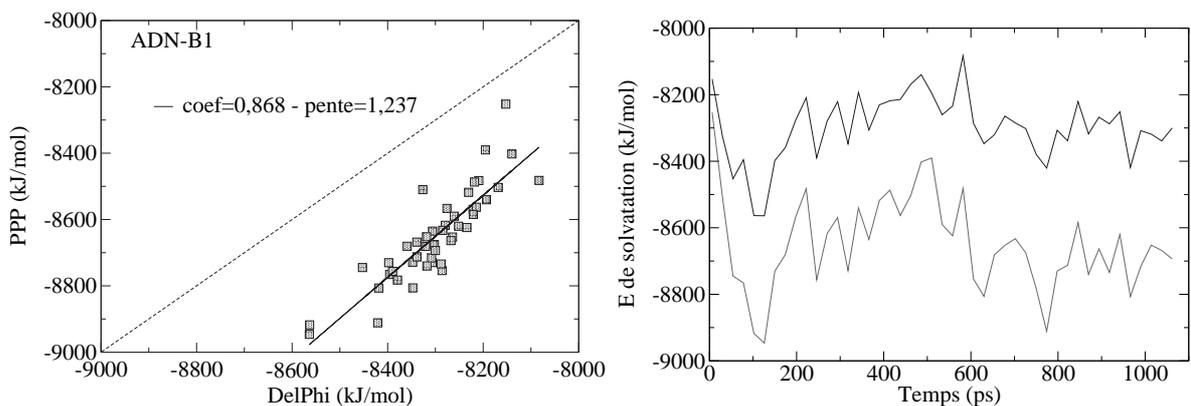


FIG. 5.32 : Corrélation entre l'énergie libre de solvation moyenne de l'ADN-B1 calculée avec PPP et l'énergie de la structure moyenne avec DelPhi (à gauche) et évolution des ces deux énergies (à droite) en gris pour PPP et en noir pour DelPhi.

La droite de corrélation entre les énergies libres électrostatiques de solvation calculées avec PPP et avec DelPhi pour la simulation de l'ADN-B1 est représentée sur la figure 5.32, ainsi que les évolutions de ces deux énergies. L'énergie de solvation varie plus pendant la simulation de l'ADN-B1 que pendant celle de l'ADN-B2, et le coefficient de corrélation (0,868) est meilleur que pour l'ADN-B2, alors que la pente (1,237) est supérieure à un. L'erreur relative est encore une fois faible, de 4,2 % en moyenne. La courbe des évolutions montre que les deux énergies suivent une évolution très parallèle.

La droite de corrélation des énergies libres électrostatiques de solvation pour la simulation de l'ADN-A1 est représentée sur la figure 5.33. Dans ce cas, l'énergie de solvation varie beaucoup plus que pour les deux autres molécules d'ADN pendant la dynamique. Le coefficient de corrélation est meilleur que pour l'ADN-B1, et même meilleur

que pour l'ARNt, il est extrêmement proche de un (0,997), alors que la pente est encore une fois supérieure à un (1,319). L'erreur relative est encore plus faible que pour les autres simulations, de 3,2 % en moyenne. La courbe des évolutions des énergies montre bien que le modèle PPP permet de suivre correctement les variations de l'énergie libre électrostatique de solvation calculée avec DelPhi tout au long de la dynamique.

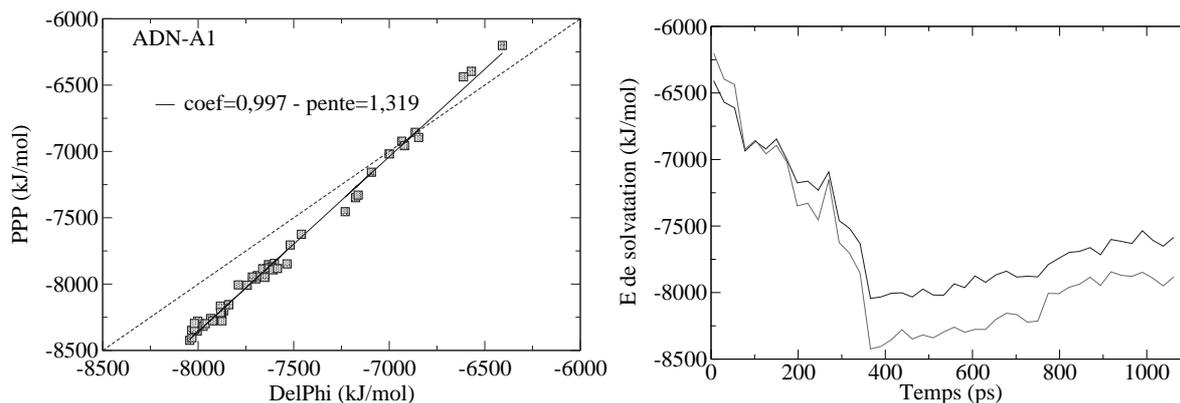


FIG. 5.33 : Corrélation entre l'énergie libre de solvation moyenne de l'ADN-A1 calculée avec PPP et l'énergie de la structure moyenne avec DelPhi (à gauche) et évolution des ces deux énergies (à droite) en gris pour PPP et en noir pour DelPhi.

Pour les trois simulations, l'énergie libre électrostatique de solvation calculée avec le modèle PPP suit les variations de l'énergie calculée avec DelPhi, avec une erreur relative systématique assez faible, entre 3 et 4 %. On remarque que pour l'ADN en forme A, comme pour l'ARNt, le modèle PPP permet d'évaluer plus précisément l'énergie libre de solvation que pour les ADN-B, en montrant de meilleures corrélations avec les méthodes de résolution de Poisson-Boltzmann. Dans toutes les simulations, notre modèle de solvant, s'il ne permet pas de calculer une valeur de l'énergie libre électrostatique de solvation très précise du système à chaque instant, donne cependant des variations très correctes de cette énergie tout au long des simulations.

### 5.3.3 Analyse de l'hydratation

Nous avons évoqué au chapitre 2 les structures d'hydratation mises en évidence par les méthodes expérimentales pour les double hélices d'ADN en forme B et en forme A. Pour les ADN-B, une épine d'hydratation serait observée dans le petit sillon de la double hélice [Drew *et al.*, 1981, Liepinsh *et al.*, 1994], avec une préférence pour les séquences riches en A-T [Kochoyan et Leroy, 1995] et pour les séquences répétées et non alternées

A–T ou G–C [Chalikian *et al.*, 1994b]. Pour les ADN-A, c'est le grand sillon qui est particulièrement hydraté [Eisenstein et Shakked, 1995, Mayer-Jung *et al.*, 1998]. De plus, une étude cristallographique de nombreux cristaux d'ADN-A et B s'est intéressée à l'hydratation des groupements phosphates et a montré que les groupements phosphates des ADN-A partageraient entre eux leur couche d'hydratation, alors que celles des ADN-B n'étaient pas en contact les unes avec les autres [Schneider *et al.*, 1995]. Nous nous attendons donc à ce que les phosphates des ADN-A apparaissent plus fortement hydratés que ceux des ADN-B.

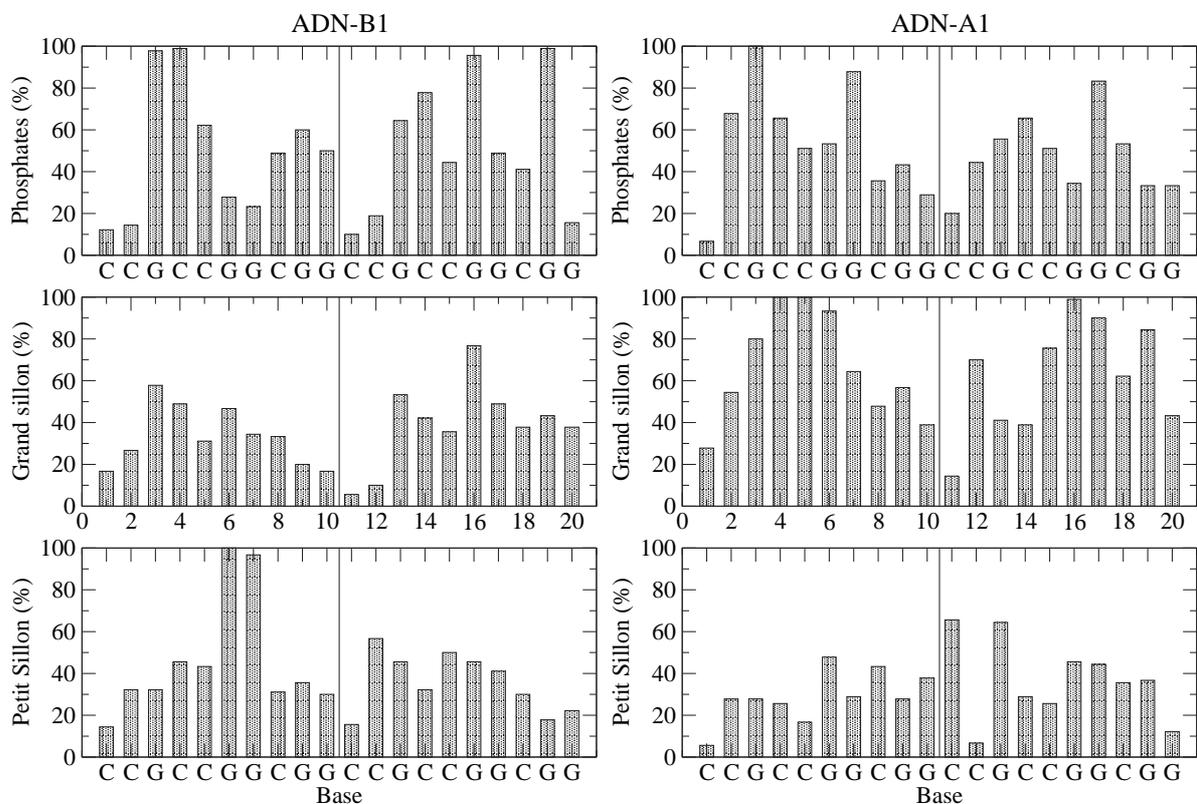


FIG. 5.34 : Pourcentage d'hydratation des double hélices d'ADN-B1 (à gauche) et A1 (à droite) pour les groupements phosphates, les atomes donneurs ou accepteurs de liaisons hydrogène des bases du grand sillon, et ceux du petit sillon.

La figure 5.34 représente le pourcentage d'hydratation maximum selon la séquence pour les simulations des ADN-B1 et A1. Même si la structure de l'ADN-A1 s'éloigne d'un ADN-A classique pendant la simulation, son hydratation présente les principales caractéristiques d'un ADN-A, c'est-à-dire des groupements phosphates assez fortement hydratés, le petit sillon pauvrement hydraté et le grand sillon fortement hydraté. Le schéma d'hydratation de l'ADN-B1 est plus mitigé : alors que les phosphates sont presque autant hydratés que pour l'ADN-A1, le petit et le grand sillon sont hydratés de façon équivalente, ne mon-

trant pas d'épine d'hydratation dans le petit sillon comme nous l'observerons un peu plus loin pour l'ADN-B2. Seules les bases G6 et G7 sont très hydratées dans le petit sillon. Ces résultats suggèrent que l'épine d'hydratation ne serait pas favorisée par les séquences riches en paires C–G. L'ADN-B1 ne possède pas l'hydratation typique d'un ADN-B, alors que l'ADN-A1 présente une structure d'hydratation assez proche d'un ADN-A classique. Nous pouvons même avancer que le schéma d'hydratation de l'ADN-B1 semble être situé entre celui d'un ADN-A et d'un ADN-B classique.

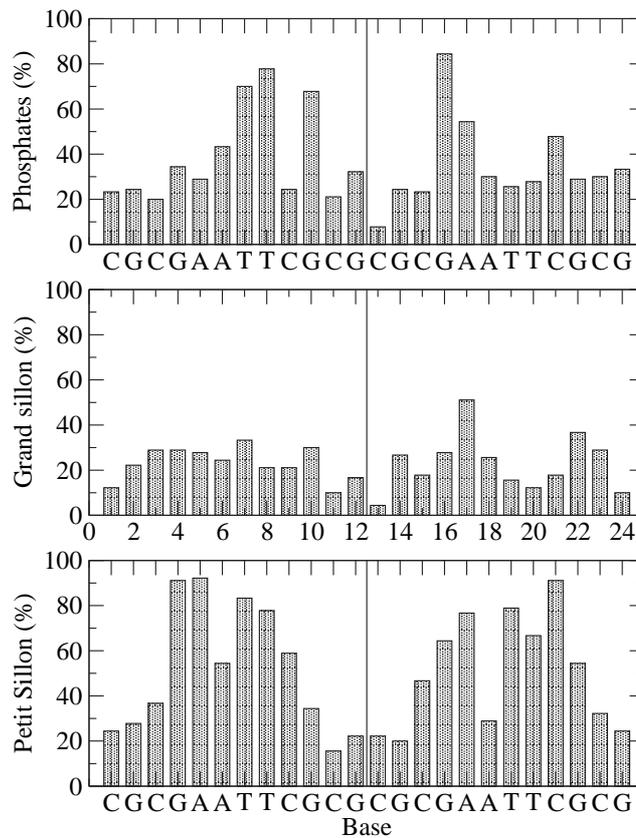


FIG. 5.35 : Pourcentage d'hydratation de la double hélice d'ADN-B2 pour les groupements phosphates, les atomes donneurs ou accepteurs de liaisons hydrogène des bases du grand sillon, et ceux du petit sillon.

La figure 5.35 représente le pourcentage d'hydratation maximum selon la séquence pour la simulation de l'ADN-B2. Le grand sillon apparaît clairement peu hydraté, en particulier aux extrémités de la molécule. Comme attendu, le petit sillon est au contraire plus fortement hydraté en moyenne, et particulièrement au niveau des bases centrales de la molécule, donc pour la région AATT. La comparaison avec le schéma d'hydratation de l'ADN-B1 suggère que l'épine d'hydratation dans le petit sillon de la double hélice est favorisée par des séquences riches en paires A–T. Nous retrouvons donc dans nos simulations avec le solvant PPP les résultats expérimentaux concernant l'épine d'hydratation

dans le petit sillon. Les groupements phosphates ne sont pas particulièrement hydratés, à part pour quelques nucléotides (T7, T8, G10 pour le premier brin, et G16 pour le second).

#### 5.3.4 Conclusion sur les simulations des double hélices d'ADN

Ces trois simulations de double hélices d'ADN avec le solvant PPP nous ont apporté des informations différentes et complémentaires. Dans tous les cas, les simulations aboutissent à des structures relativement stables, et les énergies libres électrostatiques de solvation calculées avec notre modèle suivent assez bien les énergies calculées par résolution de Poisson-Boltzmann avec DelPhi. Les simulations des ADN-A1 et B1 ont des résultats plus mitigés que l'ADN-B2, peut-être à cause de la longueur assez réduite de la séquence (seulement dix paires de bases au lieu de douze pour l'ADN-B2, soit pas tout à fait un tour complet d'hélice). On observe cependant quelques indices structuraux et d'hydratation nous permettant de conclure que l'ADN-A1 se rapproche en solution d'une structure hybride entre la forme A et la forme B. Cette structure hybride est relativement proche de la forme B au niveau des paramètres hélicoïdaux et des conformations des sucres, et très proche de la forme A pour l'hydratation. L'ADN-B1 reste quant à lui globalement en forme B, avec un schéma d'hydratation mitigé. La simulation de l'ADN-B2 nous a permis quant à elle de retrouver des effets de séquence sur les paramètres hélicoïdaux, ainsi que sur les sites d'hydratation avec nos pseudo-particules de solvant, tout à fait en accord avec ceux observés expérimentalement et théoriquement.

Les simulations des acides nucléiques avec le solvant PPP sont très encourageantes et répondent aux différents objectifs que nous nous étions fixés [Basdevant *et al.*, 2003b]. En effet, elles sont relativement stables sur presque une nanoseconde pour une boucle d'ARNt et sur plusieurs double hélices d'ADN, alors que les acides nucléiques sont particulièrement sensibles aux interactions électrostatiques et donc à la représentation du solvant. De plus, elles nous ont permis de retrouver des résultats structuraux en solution en accord avec ceux expérimentaux ou théoriques connus, mais aussi de calculer une énergie libre électrostatique de solvation dont l'ordre de grandeur et les variations sont globalement satisfaisants. Enfin, ces simulations permettent de caractériser des sites d'hydratation très proches de ceux connus pour les acides nucléiques, selon la séquence et selon la forme de l'hélice. Il nous faudrait cependant continuer notre étude sur d'autres séquences, par exemple les « A-tracts », ou plus généralement les séquences riches en paires A-T, afin de valider complètement notre modèle de solvant pour l'étude des acides nucléiques.



# Conclusion et Perspectives



## Conclusion et Perspectives

La représentation du solvant, dont le rôle est très important dans la stabilisation et la dynamique des macromolécules biologiques, constitue un enjeu majeur de la modélisation moléculaire. Le nouveau modèle de solvant proposé dans cette thèse est semi-implicite, puisqu'il représente le solvant comme un ensemble de particules microscopiques dont les propriétés diélectriques sont calculées à partir des lois macroscopiques de l'électrostatique.

En partant d'une fonctionnelle macroscopique d'énergie libre électrostatique de solvation discrétisée sur particules pour décrire le solvant, nous avons obtenu un fluide de particules de Lennard-Jones polarisables à l'équilibre thermodynamique, dont le moment dipolaire induit dépend du champ électrique créé par la macromolécule. La prise en compte des effets de saturation diélectrique nous a amenés à introduire une fonction de Langevin dans l'expression des moments dipolaires induits des particules. Notre méthodologie permet d'obtenir simplement une expression de l'énergie libre électrostatique de solvation de la macromolécule étudiée en moyennant sur les positions des particules de solvant. Les deux versions du modèle, non locale et locale (qui consiste à faire une approximation sur la densité de polarisation par rapport à la version non locale), ont été présentées au chapitre 3. La version non locale découle d'une théorie macroscopique exacte et utilise une méthode itérative pour calculer les moments dipolaires des particules. Elle nous a permis d'étudier de façon plus générale les propriétés diélectriques d'un fluide fortement polarisable, et en particulier la relation entre la constante diélectrique du milieu et la polarisabilité des particules. La version locale du modèle, pour laquelle les moments dipolaires des particules sont directement proportionnels au champ créé par le soluté, est plus simple à implémenter et plus rapide en temps de calcul. Elle a donc été développée de façon plus poussée, et a été paramétrée de la manière la plus simple possible, afin de reproduire le mieux possible à la fois les propriétés structurales de l'eau explicite pour les premières couches d'hydratation et les propriétés électrostatiques des modèles de solvation implicites. Après avoir fixé les paramètres de Lennard-Jones de nos particules de solvant et avoir trouvé le meilleur paramètre pour le dipôle de saturation de Langevin, nous avons introduit un facteur de rayon et un facteur de charge, qui renormalisent les

rayons atomiques et les charges du soluté par rapport aux paramètres des champs de force classiques, adaptés pour les simulations en solvant explicite.

La version locale du modèle a ensuite été appliquée avec succès à différentes macromolécules biologiques, acides nucléiques et protéines. Les trajectoires des dynamiques moléculaires obtenues sont stables sur une à deux nanosecondes, et les structures des molécules issues de nos simulations sont le plus souvent en accord d'un point de vue structural avec les simulations en solvant explicite et les résultats expérimentaux, et ce même pour des molécules fortement chargées telles que les acides nucléiques. De plus, les énergies libres électrostatiques de solvation des molécules étudiées calculées avec notre modèle corrént relativement bien avec celles calculées par des méthodes de résolution de l'équation de Poisson-Boltzmann, et leurs variations suivent ainsi assez bien les déformations du soluté. Enfin, l'étude de l'hydratation des acides nucléiques et protéines avec nos pseudo-particules de solvant a montré que, même sans prendre en compte explicitement les liaisons hydrogène des molécules d'eau, on pouvait retrouver les sites d'hydratation privilégiés observés en solvant explicite ou par les méthodes expérimentales. Notre modèle de solvant permet donc d'obtenir des résultats de simulations très encourageants et qui répondent bien aux objectifs de notre projet.

Nous pourrions envisager, dans la continuité de notre travail, d'appliquer la version locale de notre modèle, qui a maintenant fait ses preuves sur différentes biomolécules, à des complexes macromoléculaires, afin d'étudier les phénomènes de reconnaissance moléculaire avec nos pseudo-particules de solvant. En parallèle, afin d'obtenir des résultats plus précis sur les énergies libres de solvation calculées avec le modèle local, nous pourrions envisager de réaliser un paramétrage plus fin et moins systématique que celui effectué jusqu'ici. À l'image de la méthode des charges variationnelles utilisée pour améliorer la méthode implicite locale sur grille développée par Borgis *et al.* [2003], nous pourrions calculer un facteur de charge adapté à chaque type d'atome ou groupement d'atomes. Enfin, afin de mieux modéliser les systèmes fortement chargés (en particulier certaines séquences d'acides nucléiques riches en paires A-T, qui se sont montrées assez fragiles dans les premières tentatives de simulation avec notre modèle de solvant, et dont nous n'avons pas présenté les résultats ici), nous pourrions tenter de réfléchir à une représentation plus efficace des contre-ions par des méthodes implicites inspirées par exemple de la théorie de la fonctionnelle de la densité [Löwen *et al.*, 1993]. Ainsi, en partant d'une fonctionnelle d'énergie libre de solvation équivalente à Poisson-Boltzmann au lieu de Poisson, nous n'aurions plus à introduire les contre-ions explicitement dans le système.

Une autre perspective intéressante serait de développer le modèle non local dans une version non itérative, avec une méthode de type Car-Parinello [Car et Parinello, 1985], qui

permettrait aux moments dipolaires des particules de solvant de suivre adiabatiquement à chaque pas de dynamique les déformations de la molécule tout en restant à l'équilibre thermodynamique. Couplée à un traitement de type Particle Mesh Ewald des interactions électrostatiques à longue distance entre les charges du soluté et les dipôles induits des particules de solvant et entre les dipôles, cette version non locale pourrait ainsi être appliquée efficacement à des macromolécules biologiques.

Enfin, comme notre modèle de solvant représente chaque molécule d'eau triatomique comme une seule particule, nous pouvons considérer que notre représentation du solvant fait partie des méthodes de type « coarse-graining » développées récemment [Minichino et Voth, 1997, Shelley *et al.*, 2001, Grooth et Rabone, 2001, Müller *et al.*, 2003], où les molécules modélisées ne sont plus tout à fait représentées au niveau atomique. Le principe de ces méthodes de « coarse-graining » est en effet de considérer un ensemble constitué d'un petit nombre d'atomes comme une seule particule. Cela permet, après avoir redéfini des potentiels d'interaction simplifiés entre les particules artificielles du soluté, de gagner un temps de calcul très important pour simuler des systèmes encore plus grands sur des temps plus longs. Dans cette optique, la taille de nos particules de solvant devient un paramètre arbitraire, et nous pourrions donc simplifier encore plus notre modèle de solvant en représentant les particules qui se trouvent loin du soluté d'une taille plus importante que celles qui sont proches du soluté. Une telle méthodologie pour représenter le solvant et le soluté permettrait de simuler efficacement des systèmes biologiques de taille très importante, comme par exemple des systèmes membranaires, ou des complexes biologiques faisant intervenir trois ou quatre macromolécules.



# Bibliographie

- B. J. Alder et E. L. Pollock. *Ann. Rev. Phys. Chem.*, 32 :311–329, 1981.
- M. P. Allen et D. J. Tidesley. In *Computer Simulations of Liquids*. Oxford University Press, New-York, 1987.
- H. C. Andersen. *J. Chem. Phys.*, 72 :2384–2393, 1980.
- H. Arthanari, K. J. McConnell, R. Beger, M. A. Young, D. L. Beveridge, et P. H. Bolton. *Biopolymers*, 68 :3–15, 2003.
- P. Auffinger, S. Louise-May, et E. Westhof. *Biophys. J.*, 76 :50–64, 1999.
- P. Auffinger et E. Westhof. *Biophys. J.*, 71 :940–954, 1996.
- P. Auffinger et E. Westhof. *J. Mol. Biol.*, 269 :326–341, 1997.
- P. Auffinger et E. Westhof. *J. Biomol. Struct. Dyn*, 16 :739–1102, 1998a.
- P. Auffinger et E. Westhof. *Curr. Opin. Struct. Biol.*, 8 :227–236, 1998b.
- P. Auffinger et E. Westhof. *RNA*, 7 :334–341, 2001.
- N. Ban, P. Nissen, J. Hansen, P. B. Moore, et T. A. Steitz. *Science*, 289 :905–920, 2000.
- N. Basdevant, D. Borgis, et T. Ha-Duong. *Biophys. J.*, 2003a. (soumis).
- N. Basdevant, D. Borgis, et T. Ha-Duong. 2003b. (en préparation).
- D. Beglov et B. Roux. *J. Phys. Chem. B*, 101 :7821–7826, 1997.
- H. J. C. Berendsen, J. R. Grigera, et T. P. Straatsma. *J. Phys. Chem.*, 91 :6269–6271, 1987.
- M. Berkowitz, O. A. Karim, J. A. McCammon, et P. J. Rossky. *Chem. Phys. Lett.*, 105 : 577–580, 1984.

- H. M. Berman. *Curr. Opin. Struct. Biol.*, 4 :345–350, 1994.
- H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, et P. E. Bourne. *Nucleic Acid Res.*, 28 :235–242, 2000.
- D. L. Beveridge et K. J. McConnell. *Curr. Opin. Struct. Biol.*, 10 :182–196, 2000.
- A. R. Bizzarri et S. Cannistraro. *J. Phys. Chem.*, 106 :6617–6633, 2002.
- D. Borgis, N. Lévy, et M. Marchi. *J. Chem. Phys.*, 119 :3516–3528, 2003.
- C. J. F. Bottcher. In *Theory of Electric Polarization*. Elsevier Press, Amsterdam, 1973.
- D. F. Calef et P. G. Wolynes. *J. Phys. Chem.*, 87 :3387–3399, 1983.
- N. Calimet, M. Schaefer, et T. Simonson. *Proteins*, 45 :144–158, 2001.
- N. Capitaine. *Simulation par dynamique moléculaire de la boucle anticodon d'un ARN de transfert*. Rapport de stage de 3ème année, École Centrale Paris, 2000.
- R. Car et M. Parinello. *Phys. Rev. Lett.*, 55 :2471–2474, 1985.
- N. Carulla, C. Woodward, et G. Barany. *Biochemistry*, 39 :7927–7937, 2000.
- A. Chakrabarty et R. L. Baldwin. *Adv. Protein Chem.*, 46 :141–176, 1995.
- T. V. Chalikian et K. J. Breslauer. *Curr. Opin. Struct. Biol.*, 8 :657–664, 1998.
- T. V. Chalikian, G. E. Plum, A. P. Sarvazyan, et K. J. Breslauer. *Biochemistry*, 33 :8629–8640, 1994a.
- T. V. Chalikian, A. P. Sarvazyan, G. E. Plum, et K. J. Breslauer. *Biochemistry*, 33 :2394–2401, 1994b.
- D. Chandler, J. D. McCoy, et S. J. Singer. *J. Chem. Phys.*, 85 :5971–5982, 1986.
- T. E. Cheatham, III, J. Srinivasan, et P. A. Kollman. *J. Biomol. Struct. Dyn.*, 2 :265–280, 1998.
- J. W. Cheng, S. H. Chou, M. Salazar, et B. R. Reid. *J. Mol. Biol.*, 228 :118–137, 1992.
- B. S. Choi et A. G. Redfield. *J. Biochem.*, 117 :515–520, 1995.
- P. Cieplak, J. Caldwell, et P. Kollman. *J. Comput. Chem.*, 22 :1048–1057, 2001.
- G. M. Clore et A. M. Gronenborn. *J. Mol. Biol.*, 223 :853–856, 1992.

- M. B. Comarmond, R. Giégé, J. C. Thierry, D. Moras, et J. Ficher. *Acta Crystallograph. B*, 42 :272–280, 1986.
- W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, et P. A. Kollman. *J. Am. Chem. Soc.*, 117 : 5179–5197, 1995.
- L. X. Dang, J. E. Rice, et P. A. Kollman. *J. Chem. Phys.*, 93 :7528–7529, 1990.
- T. A. Darden, A. Toukmaji, et L. G. Pedersen. *J. Chim. Phys.*, 94 :1346–1364, 1997.
- V. P. . Denisov et B. Halle. *Faraday Discuss.*, 103 :227–244, 1996.
- V. P. Denisov et B. Halle. *J. Mol. Biol.*, 245 :682–697, 1995.
- R. E. Dickerson, M. Bansal, C. R. Calladine, S. Dieckmann, W. N. Hunter, O. Kennard, R. Lavery, H. C. M. Nelson, W. K. Olson, W. Saenger, Z. Shakked, H. Sklenar, D. M. Soumpasis, C. S. Tung, E. V. Kitzing, A. H. J. Wang, et V. D. Zhurkin. *J. Mol. Biol.*, 205 :787–791, 1989.
- R. E. Dickerson et H. R. Drew. *J. Mol. Biol.*, 149 :761–786, 1981.
- B. N. Dominy et C. L. Brooks, III. *J. Phys. Chem. B*, 103 :3765–3773, 1999.
- H. R. Drew et R. E. Dickerson. *J. Mol. Biol.*, 151 :535–556, 1981.
- H. R. Drew, R. M. Ming, T. Takano, C. Broka, S. Tanaka, K. Itakura, et R. E. Dickerson. *Proc. Nat. Acad. Sci. USA*, 78 :2179–2183, 1981.
- S. Ehrenson. *J. Comput. Chem.*, 10 :77–93, 1989.
- M. Eisenstein et Z. Shakked. *J. Mol. Biol.*, 248 :662–678, 1995.
- U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, et L. G. Pedersen. *J. Chem. Phys.*, 103 :8577–8592, 1995.
- R. Fairman, K. R. Schoemaker, E. J. York, J. M. Stewart, et R. L. Baldwin. *Biophys. Chem.*, 37 :107–119, 1990.
- M. Feig et B. M. Pettitt. *Biophys. J.*, 75 :134–149, 1998.
- M. Feig et B. M. Pettitt. *Biophys. J.*, 77 :1769–1781, 1999.
- B. A. Fields, F. A. Goldbaum, W. Dall’Acqua, E. L. Malchiodi, A. Cauerhff, F. P. Schwarz, X. Ysern, R. J. Poljak, et R. A. Mariuzza. *Biochemistry*, 35 :15494–15503, 1996.

- J. Florian et A. Warshel. *J. Phys. Chem.*, 101 :5583–5595, 1997.
- F. Fologari, G. Esposito, P. Viglino, et H. Molinari. *J. Comput. Chem.*, 22 :1830–1842, 2001.
- M. K. Frank, G. M. Glore, et A. M. Gronenborn. *Protein Sci.*, 4 :2605–2615, 1995.
- R. E. Franklin et R. G. Goslin. *Nature*, 171 :740–741, 1953.
- C. A. Frederick, J. Grable, M. Melia, C. Samudzi, L. Jen-Jacobson, B.-C. Wang, P. Greene, H. W. Boyer, et J. M. Rosenberg. *Nature*, 309 :327–330, 1984.
- D. Freifelder. In Masson, editor, *Biologie Moléculaire*. 1990.
- D. Frenkel et B. Smit. In *Understanding Molecular Simulation*. Academic Press, 1996.
- M. W. Friederich, E. Vacano, et P. J. Hagerman. *Proc. Natl. Acad. Sci.*, 95 :3572–3577, 1998.
- P. Frodl et S. Dietrich. *Phys. Rev. A*, 45 :7330–7354, 1992.
- T. Gallagher, P. Alexander, P. Bryan, et G. L. Gilliland. *Biochem.*, 33 :4721–4729, 1994.
- M. M. Garner et D. C. Rau. *EMBO J.*, 14 :1257–1263, 1995.
- H. G. Gassen. *Prog. Nucleic Acid Res. Mol. Biol.*, 24 :57–86, 1980.
- D. T. Gewirth et P. B. Sigler. *Nat. Struct. Biol.*, 2 :386–394, 1995.
- M. K. Gilson, J. A. McCammon, et J. D. Madura. *J. Comput. Chem.*, 16 :1081–1095, 1995.
- E. Giudice et R. Lavery. *Acc. Chem. Res.*, 35 :350–357, 2002.
- F. A. Goldbaum, F. P. Schwarz, E. Einsenstein, A. Cauerhff, R. A. Mariuzza, et R. J. Poljak. *J. Mol. Recogn.*, 9 :6–12, 1996.
- A. M. Gronenborn, D. R. Filpula, N. Z. Essig, A. Achari, M. Whitlow, P. T. Wingfield, et G. M. Clore. *Science*, 253 :657–661, 1991.
- R. D. Groot et K. L. Rabone. *Biophys. J.*, 81 :725–736, 2001.
- E. Guardia, R. Rey, et J. A. Padro. *Chem. Phys.*, 155 :187–195, 1991.
- T. Ha-Duong, S. Phan, M. Marchi, et S. Phan. *J. Chem. Phys.*, 117 :541–556, 2002.

- J.-P. Hansen et I. R. McDonald. Academic Press, 2nd edition, 1986.
- U. H. E. Hansmann et Y. Okamoto. *J. Phys. Chem. B*, 102 :653–656, 1998.
- B. Hartmann et R. Lavery. *Quart. Rev. Biophys.*, 29 :309–368, 1996.
- S. C. Harvey. *Proteins*, 5 :78–92, 1989.
- S. A. Hassan, F. Guarnieri, et E. L. Mehler. *J. Phys. Chem. B*, 104 :6478–6489, 2000.
- S. H. Hassan, E. L. Mehler, D. Zhang, et H. Weinstein. *Proteins*, 51 :109–125, 2003.
- S. Hayward, A. Kitao, F. Hirata, et N. Go. *J. Mol. Biol.*, 234 :1207–1217, 1993.
- E. R. Henry et A. Szabo. *J. Chem. Phys.*, 82 :4753–4761, 1985.
- B. Hingerty, R. H. Richie, T. L. Ferre, et T. E. Turner. *Biopolymers*, 24 :427–439, 1985.
- B. Honig, K. Sharp, et A. Yang. *J. Phys. Chem.*, 97 :1101–1109, 1993.
- B. Honig et A. Nicholls. *Science*, 268 :1144–1149, 1995.
- W. G. Hoover. *Phys. Rev. A*, 31 :1695–1697, 1985.
- N. V. Hud et J. Plavec. *Biopolymers*, 69 :144–159, 2003.
- W. Humphrey, A. Dalke, et K. Schulten. *J. Molec. Graphics*, 14.1 :33–38, 1996.
- D. D. Humphreys, R. A. Friesner, et B. J. Berne. *J. Phys. Chem.*, 98 :6885–6892, 1994.
- D. E. Hyre et L. D. Spicer. *Biochemistry*, 34 :3212–3221, 1995.
- W. Im, D. Beglov, et B. Roux. *Comp. Phys. Comm.*, 111 :59, 1998.
- J. D. Jackson. In Wiley, editor, *Classical Electrodynamics*. 3rd edition, New-York, 1999.
- W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, et M. L. Klein. *J. Chem. Phys.*, 79 :926–935, 1983.
- S. L. Kazmirski et V. Daggett. *J. Mol. Biol.*, 277 :487–506, 1998.
- D. Khare, P. Alexander, et J. Orban. *Biochemistry*, 38 :3918–3925, 1999.
- Y. I. Kharkats, A. A. Kornyshev, et M. A. Vorotynstev. *J. Chem. Soc. Faraday Trans. II*, 72 :361–371, 1976.
- J. E. Kilpatrick, K. S. Pitzer, et R. Spitzer. *J. Am. Chem. Soc.*, 69 :2483–2488, 1947.

- J. G. Kirkwood. *J. Chem. Phys.*, 2 :351–361, 1934.
- N. Kobayashi, S. Honda, H. Yoshii, et E. Munekata. *Biochemistry*, 39 :6564–6571, 2000.
- M. Kochoyan et J.-L. Leroy. *Curr. Opin. Struct. Biol.*, 5 :329–333, 1995.
- P. A. Kollman. *Chem. Rev.*, 93 :2395–2417, 1993.
- D. Kosztin, T. C. Bishop, et K. Schulten. *Biophys. J.*, 73 :557–570, 1997.
- H. Kovacs, A. E. Mark, et W. E. van Gunsteren. *Proteins*, 27 :395–404, 1997.
- L. A. Labeots et M. A. Weiss. *J. Mol. Biol.*, 269 :113–128, 1997.
- J. E. Ladbury, J. G. Wright, J. M. Strurtevant, et P. B. Sigler. *J. Mol. Biol.*, 238 :669–681, 1998.
- A. Lahiri et L. Nilsson. *Biophys. J.*, 79 :2276–2289, 2000.
- A. N. Lane, T. C. Jenkins, T. Brown, et S. Neidle. *Biochemistry*, 30 :1372–1385, 1991.
- R. Lavery. In Olson, Sarma, et Sundaralingam, editors, *Structure and Expression. DNA bending and curvature*, volume 3, pages 191–211. Adenine Press, New-York, 1988.
- R. Lavery et H. Sklenar. *Curves 51 - Helical Analysis of irregular nucleic acids*. CNRS URA 77, Institut de Biologie Physico-Chimique, 1996.
- R. Lavery, H. Sklenar, K. Zakrzewska, et B. Pullman. *J. Biomol. Struct. Dyn.*, 3 :989–1014, 1986.
- R. Lavery et K. Zakrewzka. In Neidle, editor, *Oxford Handbook of Nucleic Acide Structure*. Oxford University Press, 1997.
- A. R. Leach. In Longman, editor, *Molecular Modelling. Principles and Applications*. Singapore, 1996.
- D. D. Leonidas, R. Shapiro, L. I. Irons, N. Russo, et K. R. Acharya. *Biochemistry*, 36 : 5578–5588, 1997.
- M. Levitt et A. Warshel. *J. Am. Chem. Soc.*, 100 :2607–2613, 1978.
- E. Liepinsh, W. Leupin, et G. Otting. *Nucleic Acid Res.*, 22 :2249–2254, 1994.
- T. Lündback et T. Härd. *Proc. Natl. Acad. Sci. USA*, 93 :4754–4759, 1996.
- V. Lounnas et B. M. Pettitt. *Proteins*, 18 :133–147, 1994.

- N. Lévy. *Solvatation des macromolécules complexes : une approche fonctionnelle de l'équation de Poisson-Boltzmann*. Thèse de Doctorat, Université d'Evry Val d'Essonne, 2002.
- H. Löwen, J.-P. Hansen, et P. A. Madden. *J. Chem. Phys.*, 98 :3275–3289, 1993.
- T. W. Lynch et S. G. Sligar. *Biochim. Biophys. Acta*, 1595 :277–282, 2002.
- A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, et M. Karplus. *J. Phys. Chem. B*, 102 :3586–3616, 1998.
- P. A. Madden et D. Kivelson. *Adv. Chem. Phys.*, 56 :467–566, 1984.
- M. W. Mahoney et W. L. Jorgensen. *J. Chem. Phys.*, 112 :8910–8922, 2000.
- V. A. Makarov, B. K. Andrews, P. E. Smith, et B. M. Pettitt. *Biophys. J.*, 79 :2966–2974, 2000.
- V. A. Makarov, B. M. Pettitt, et M. Feig. *Acc. Chem. Res.*, 35 :376–384, 2002.
- M. Marchi, D. Borgis, N. Levy, et P. Ballone. *J. Chem. Phys.*, 114 :4377–4385, 2001.
- M. Marchi, F. Sterpone, et M. Ceccarelli. *J. Am. Chem. Soc.*, 124 :6787–6791, 2002.
- R. A. Marcus. *J. Chem. Phys.*, 24 :979–989, 1956.
- B. Masquida et E. Westhof. *RNA*, 6 :9–15, 2000.
- C. Mayer-Jung, D. Moras, et Y. Timsit. *EMBO J.*, 17 :2709–2718, 1998.
- C. Minichino et G. A. Voth. *J. Phys. Chem. B*, 101 :4544–4552, 1997.
- M. Müller, K. Katsov, et M. Schick. *J. Polym. Sci. B*, 41 :1441–1450, 2003.
- D. Moras, A. C. Dock, P. Dumas, E. Westhof, P. Romby, J. P. Ebel, et R. Giégé. *J. Biomol. Struct. Dyn.*, 3 :479–493, 1985.
- L. R. Murphy, N. Matubayasi, V. A. Payne, et R. M. Levy. *Fold Des.*, 3 :105–118, 1998.
- N. Nassar, G. Horn, C. Herrmann, C. Block, R. Janknecht, et A. Wittinghofer. *Nat. Struct. Biol.*, 3 :723–729, 1996.
- H.-L. Ng et R. E. Dickerson. *Nucleic Acids Res.*, 30 :4061–4067, 2002.

- A. Nicholls et B. Honig. *J. Comput. Chem.*, 12 :435–445, 1991.
- S. Nosé. *J. Chem. Phys.*, 81 :511–519, 1984.
- L. Onsager. *J. A. C. S.*, 58 :1486–1493, 1936.
- J. J. Osterhout, Jr., R. L. Baldwin, E. J. York, J. M. Stewart, H. J. Dyson, et P. E. Wright. *Biochemistry*, 28 :7059–7064, 1989.
- G. Otting, E. Liepinsh, et K. Wüthrich. *Science*, 254 :974–980, 1991.
- G. Otting et K. Wüthrich. *J. Am. Chem. Soc.*, 111 :1871–1875, 1989.
- H. Pan, E. Barbar, G. Barany, et C. Woodward. *Biochemistry*, 34 :13974–13981, 1995.
- A. Papazyan et A. Warshel. *J. Phys. Chem. B*, 101 :11254–11264, 1997.
- S. Parkin, B. Rupp, et H. Hope. *Acta Crystallogr. D*, 52 :18–29, 1996.
- B. M. Pettitt et M. Karplus. *J. Phys. Chem.*, 90 :6335–6345, 1986.
- E. L. Pollock et B. J. Alder. *Phys. Rev. Lett.*, 39 :299–302, 1977.
- E. L. Pollock et B. J. Alder. *Phys. Rev. Lett.*, 41 :903–906, 1978.
- E. L. Pollock et B. J. Alder. *Physica A*, 102 :1–21, 1980.
- E. L. Pollock, B. J. Alder, et L. R. Pratt. *Proc. Natl. Acad. Sci. USA*, 77 :49–51, 1980.
- P. Procacci, T. A. Darden, E. Paci, et M. Marchi. *J. Comput. Chem.*, 18 :1848–1862, 1997.
- P. Procacci et M. Marchi. In C. Zannoni et P. Pasini, editors, *Advances in Computer Simulations of Liquid Crystals*. NATO-ASI School, 1998.
- D. Qiu, P. S. Shenkin, F. P. Hollinger, et W. C. Still. *J. Phys. Chem.*, 101 :3005–3014, 1997.
- R. Ramirez, R. Gebauer, M. Mareschal, et D. Borgis. *Phys. Rev. E*, 66 :031206, 2002.
- J. D. Ramshaw. *J. Chem. Phys.*, 55 :1763, 1971.
- S. N. Rao et P. Kollman. *J. Am. Chem. Soc.*, 107 :1611–1617, 1985.
- J.-P. Rickaert, G. Ciccotti, et H. J. C. Berendsen. *J. Comput. Phys.*, 23 :327–341, 1977.
- C. R. Robinson et S. G. Sligar. *J. Mol. Biol.*, 234 :302–306, 1993.

- C. R. Robinson et S. G. Sligar. *Proc. Natl. Acad. Sci. USA*, 95 :2186–2195, 1998.
- B. Roux et T. Simonson. *Biophys. Chem.*, 78 :1–20, 1999.
- C. Rubin-Carrez. *Modélisation moléculaire des propriétés conformationnelles et dynamiques d'acides ribo-nucléiques : représentation électrostatique et influence de l'environnement*. Thèse de Doctorat, Université Louis Pasteur de Strasbourg, 1992.
- R. Sayle et E. J. Milner-White. *Trends in Biochem. Sciences*, 20 :374, 1995.
- M. Schaefer et M. Karplus. *J. Phys. Chem. B*, 100 :1578–1599, 1996.
- B. Schneider, K. Patel, et H. M. Berman. *Biophys. J.*, 75 :2422–2434, 1995.
- T. R. Schneider, A. T. Brünger, et M. Nilges. *J. Mol. Biol.*, 285 :727–740, 1999.
- K. R. Schoemaker, P. S. Kim, E. J. York, J. M. Stewart, et R. L. Baldwin. *Nature*, 326 :563–567, 1987.
- J. W. R. Schwabe. *Curr. Opin. Struct. Biol.*, 7 :126–134, 1997.
- F. B. Sheinerman et C. L. Brooks, III. *Proteins*, 29 :193–202, 1997.
- J. C. Shelley, M. Y. Shelley, R. C. Reeder, et S. B. M. L. Klein. *J. Phys. Chem. B*, 105 :4464–4470, 2001.
- T. Simonson. *Curr. Opin. Struct. Biol.*, 11 :243–253, 2001.
- T. Simonson. *Rep. Prog. Phys.*, 66 :737–787, 2003.
- P. E. Smith, R. C. van Schaik, T. Szyperski, K. Wüthrich, et W. F. van Gunsteren. *J. Mol. Biol.*, 246 :356–365, 1995.
- M. Soler-Lopez, L. Malinina, et J. A. Subirina. *J. Biol. Chem.*, 275 :23034–23044, 2000.
- D. Sprous, M. A. Young, et D. L. Beveridge. *J. Phys. Chem. B*, 102 :4658, 1998.
- J. Srinivasan, T. E. Cheatham, III, P. Cieplak, P. A. Kollman, et D. A. Case. *J. Am. Chem. Soc.*, 120 :9401–9409, 1998.
- R. Steff, L. Trantirek, M. Volickova, J. Koca, V. Sklenart, et J. Kypr. *J. Mol. Biol.*, 307 :513–524, 2001.
- W. C. Still, A. Tempczyk, R. C. Hawley, et T. Hendrickson. *J. Am. Chem. Soc.*, 112 :6127–6129, 1990.

- E. Stofer et R. Lavery. *Biopolymers*, 34 :337–346, 1994.
- M. Sundaralingam et B. Pan. *Biophys. Chem.*, 95 :273–282, 2002.
- M. Sunnerhagen, V. P. Denisov, K. Venu, A. M. J. J. Bonvin, J. Carey, B. Halle, et G. Otting. *J. Mol. Biol.*, 282 :847–858, 1998.
- E. Tüchsen, J. M. Hayes, S. Ramaprasad, V. Copie, et C. Woodward. *Biochemistry*, 26 : 5163–5172, 1987.
- P. I. Teixeira et M. M. T. da Gama. *J. Phys. : Condens. Matter*, 3 :111–125, 1991.
- Y. Timsit et D. Moras. *EMBO J.*, 13 :2737–2746, 1994.
- J. Tirado-Rives et W. L. Jorgensen. *Biochemistry*, 30 :3864–3871, 1991.
- B. Tooze. In *Introduction à la Structure des Protéines*. De Boeck Université, 1996.
- A. Toukmaji et C. Sagui. *J. Chem. Phys.*, 113 :10913–10927, 2000.
- V. Tsui et D. Case. *J. Am. Chem. Soc.*, 122 :2489–2498, 2000.
- V. Tsui et D. A. Case. *Biopolymers*, 56 :275–291, 2001.
- N. B. Ulyanov, U. Schmitz, A. Kumar, et T. L. James. *Biophys. J.*, 78 :5937–5939, 1995.
- D. van der Spoel, P. J. van Maaren, et H. J. C. Berdendsen. *J. Chem. Phys.*, 108 : 10220–10230, 1998.
- W. S. van Gasteren, S. R. Billeter, A. A. Eising, P. Hünenberger, P. Krüger, A. Mark, W. Scott, et I. Tironi. In V. H. A. an der ETH Zürich, editor, *Biomolecular Simulation : The GROMOS 96 Manual and User Guide*. Zürich, Switzerland, 1996.
- Y. Wang, G. A. Thomas, et W. L. Peticolas. *J. Biomol. Struct. Dyn.*, 5 :249–274, 1987.
- Y.-X. Wang, D. I. Freedberg, S. Grzesiek, D. A. Torchia, P. T. Wingfield, J. D. Kaufman, S. J. S. anf C.-H. Chang, et C. N. Hodge. *Biochemistry*, 35 :12694–12704, 1996.
- A. Warshel et M. Levitt. *J. Mol. Biol.*, 103 :227–249, 1976.
- A. Warshel et A. Papazyan. *Curr. Opin. Struct. Biol.*, 8 :211–217, 1998.
- A. Warshel et S. T. Russel. *Quart. Rev. Biophys.*, 17 :283–422, 1984.
- J. D. Watson et F. H. C. Crick. *Nature*, 171 :964–967, 1953.

- J. D. Watson, M. Gilman, J. Witoswski, et M. Zoller. In *ADN Recombinant*. De Boeck Université, 1994.
- W. Weber, P. H. Hünenberger, et J. A. McCammon. *J. Phys. Chem. B*, 104 :3668–3675, 2000.
- M. H. F. Wilkins, A. R. Stockes, et H. R. Wilson. *Nature*, 171 :738–740, 1953.
- A. Wlodawer, J. Deisenhofer, et R. Huber. *J. Mol. Biol.*, 193 :145–156, 1987.
- M. A. Young, G. Ravishanker, et D. L. Beveridge. *Biophys. J.*, 73 :2313–2336, 1997.
- X.-J. Zhang et B. W. Matthews. *Prot. Sci.*, 3 :1031–1039, 1994.

## Résumé :

Dans la cellule des organismes vivants, le solvant (l'eau) joue un rôle très important dans la stabilisation des structures tridimensionnelles des macromolécules biologiques et lors de leurs interactions. Les méthodes théoriques de simulations de modélisation moléculaire permettent de compléter les informations partielles sur l'hydratation des biomolécules obtenues par les méthodes expérimentales. Nous avons développé un nouveau modèle de solvation semi-implicite pour représenter le solvant en modélisation moléculaire. Ce modèle décrit le solvant comme des particules microscopiques dont les propriétés diélectriques découlent des lois macroscopiques de l'électrostatique. Nous obtenons ainsi à l'équilibre électrostatique un fluide de particules de Lennard-Jones non polaires, polarisables par le champ électrique créé par le soluté. Ce modèle a l'intérêt de prendre en compte la structure moléculaire du solvant tout en calculant efficacement l'énergie libre électrostatique de solvation du système. De plus, il est d'un faible coût numérique comparé aux méthodes explicites. Après avoir implémenté notre modèle dans un programme de dynamique moléculaire et l'avoir paramétré de façon simple, nous l'avons appliqué à plusieurs peptides, protéines et acides nucléiques (ADN et ARN de transfert). Les trajectoires de ces simulations sont stables sur une à deux nanosecondes, et les structures obtenues sont tout à fait en accord avec les méthodes expérimentales et les méthodes théoriques de solvation explicites. Notre modèle permet également de retrouver les sites préférentiels d'hydratation des molécules étudiées identifiés expérimentalement ou théoriquement, malgré l'absence de liaisons hydrogène dans notre solvant. De plus, nous observons de bonnes corrélations entre les énergies libres électrostatiques de solvation calculées avec notre modèle et celles calculées avec les méthodes de résolution de l'équation de Poisson-Boltzmann, et ces résultats paraissent très encourageants.

**Mots-clés :** Solvation, Modélisation Moléculaire, Dynamique Moléculaire, Protéine, ADN, ARN de transfert, Énergie libre électrostatique, Sites d'hydratation.

## Abstract :

In the cell of the living organisms, the solvent (water) plays an important part in the stabilization of the tridimensional structures of biological macromolecules and in their interactions. Theoretical methods like molecular modeling simulations can complete the partial informations on biomolecules hydration provided by experimental methods. We have developed a new semi-implicit solvation model to represent solvent in molecular modeling of biological macromolecules. Our model considers the solvent as microscopic particles with dielectric properties based on macroscopic electrostatic equations. The resulting solvent is a fluid composed, at electrostatic equilibrium, by non polar Lennard-Jones particles, polarizable by the electric field created by the solute. This model can describe the molecular details of solvent and can also easily calculate the electrostatic solvation free energy of the system. It is also numerically very efficient compared with explicit models. We have implemented our model in a molecular dynamics code, parameterized it in a simple way and applied it to the dynamics of several peptides, proteins and nucleic acids (DNA and transfer RNA). The trajectories of these dynamics are stable over one or two nanoseconds, and their structural properties are in very good agreement with experimental results, and with simulations using explicit methods. Our model allows also to find the preferential hydration sites of the molecules identified by experimental or theoretical studies, in spite of the absence of hydrogen bonds in our solvent. Moreover, correlations between the electrostatic solvation free energies calculated with our method and those from the Poisson-Boltzmann resolution methods are quite good, and thus provide very encouraging results.

**Key words :** Solvation, Molecular Modeling, Molecular Dynamics, Protein, DNA, Transfer RNA, Electrostatic Free Energy, Hydration sites.