

# Les méthodes numériques de la dynamique moléculaire

Christophe Chipot

*Equipe de chimie et & biochimie théoriques, Unité Mixte de Recherche CNRS/UHP 7565, Institut Nancéien de Chimie Moléculaire, Université Henri Poincaré, B.P. 239, 54506 Vandœuvre-lès-Nancy, France*  
*chipot@lctn.uhp-nancy.fr*

(Version: November 16, 2007)

## Table of contents

<b>1. Introduction</b>	<b>2</b>
1.1. Relier le microscopique au méso- et macroscopique . . . . .	2
1.2. De la légitimité des simulations de dynamique moléculaire . . . . .	4
<b>2. Les équations de la dynamique moléculaire</b>	<b>5</b>
2.1. La fonction énergie potentielle . . . . .	7
2.2. Les intégrateurs de la dynamique moléculaire . . . . .	12
2.3. Dynamique moléculaire à température constante . . . . .	14
2.4. Dynamique moléculaire à pression constante . . . . .	16
2.5. Interactions électrostatiques: Au-delà des approximations usuelles . . . . .	18
2.6. Quelques propriétés accessibles à partir des simulations . . . . .	22
<b>3. Conclusion</b>	<b>25</b>

## 1. Introduction

La connaissance à un niveau atomique des aspects structuraux et dynamiques des systèmes organisés est particulièrement importante pour mieux comprendre les fonctions de ces édifices moléculaires complexes. Dans de nombreux cas, l'obtention du détail microscopique par les techniques expérimentales conventionnelles s'avère impossible. Cependant, la véritable explosion des moyens informatiques amorcée depuis une dizaine d'années, et le développement d'algorithmes performants, rendent possible l'étude d'assemblages supramoléculaires de complexité croissante par les méthodes de la chimie théorique.

Ce cours a pour vocation d'examiner une facette de la chimie théorique, que constituent les simulations statistiques de la mécanique moléculaire. L'objectif de celles-ci est d'accéder au détail atomique de phases condensées au travers "d'expériences informatiques". Pour ce faire, de nombreuses techniques sont actuellement disponibles, parmi lesquelles la dynamique moléculaire, la dynamique stochastique et ses cas particuliers — par exemple, la dynamique brownienne ou la dynamique de Langevin — ou encore les simulations de Monte Carlo. Ces différentes approches théoriques peuvent être vues à bien des égards comme le pont reliant l'observation expérimentale macroscopique à son pendant microscopique. Dans ce qui suit, nous ne discuterons que de la dynamique moléculaire. [?]

### 1.1. Relier le microscopique au méso- et macroscopique

Est-il légitime de recourir aux simulations moléculaires pour modéliser des phases condensées ? En toute rigueur, l'étude complète d'un système aussi complexe qu'un liquide moléculaire nécessite la résolution de l'équation de Schrödinger dépendant du temps, pour un ensemble important {électrons + noyaux}. Une telle approche reste totalement illusoire, en dépit des progrès récents réalisés dans le domaine des calculs à croissance linéaire, nous contraignant, de fait, à nous limiter à une description classique de la dynamique du système. Même dans ce cadre, ne serait-ce que pour des raisons de temps de calcul, les simulations moléculaires sont généralement restreintes à un nombre de particules compris entre la centaine et quelques milliers.

Pour pouvoir corréler les propriétés du système microscopique avec celles de la phase macroscopique, il est impératif de s'affranchir des effets de bord. Dans la pratique, on a recours aux conditions de limites périodiques, qui consistent à répliquer l'ensemble fini de particules réparties dans une boîte

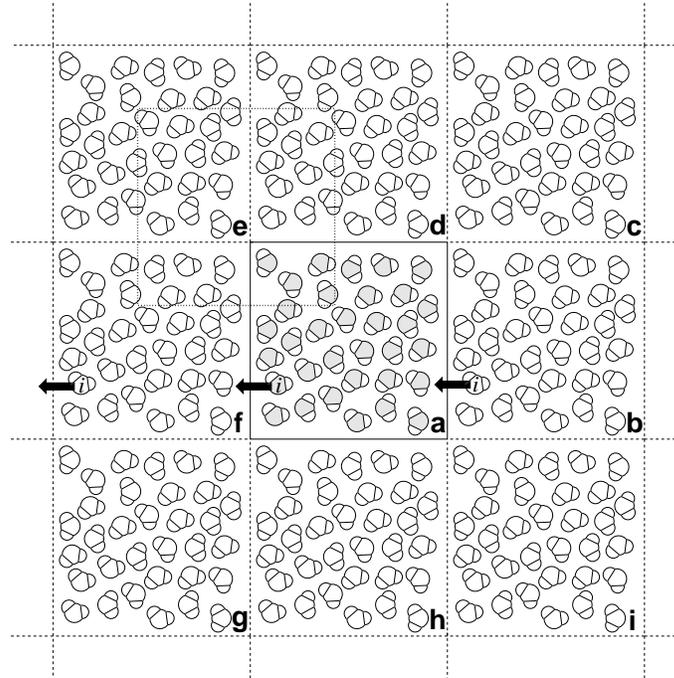


Figure 1: Vue bidimensionnelle d’une cellule de simulation répliquée dans les trois dimensions de l’espace. En utilisant les conditions de limites périodiques limites, lorsque la molécule  $i$  quitte la boîte centrale **a**, ses images dans les cellules fantômes voisines se déplacent de manière similaire. La boîte en pointillés, à cheval sur les cellules **a**, **d**, **e** et **f**, symbolise la convention de l’image minimale.

généralement parallélépipédique, selon les trois directions de l’espace (voir Figure 1). Il est apparu que la reproduction fiable de grandeurs thermodynamiques, à partir d’échantillons de tailles réduites, justifiait *a posteriori* de cette approche. Le caractère pseudo-infini du système ainsi généré nous contraint à effectuer certaines approximations quant au traitement des interactions entre molécules. [?] En particulier, l’approximation dite de “l’image minimale” suppose que chaque particule  $i$  de la cellule centrale interagit avec l’image la plus proche de toutes les autres particules  $j$ . Par ailleurs, l’introduction d’une sphère de troncature, ou *cut-off*, permet de supprimer les interactions au-delà d’une distance arbitraire, définie comme étant inférieure ou égale à la moitié du plus petit côté de la cellule de simulation (voir Figure 2). Il paraît absolument évident que ces approximations seront d’autant plus valables que la portée des interactions sera réduite. Si, le cas des interactions de dispersion ou de répulsion, de portée limitée, ne soulèvent généralement aucune difficulté, il n’en va pas de même pour certaines interactions électrostatiques. La taille du système évoluant globalement en  $r^3$ , on admet que le traitement des interactions en  $1/r^n$  — où  $n < 3$  — dans le contexte d’un modèle à distance de troncature, sera vraisemblablement erroné. Pour faire face à cette situation, il est recommandé d’employer une méthode de type somme de réseau, comme Ewald–Kornfeld ou Ladd, [?, ?] qui consiste à évaluer les interactions d’une molécule avec toutes les autres situées dans

la cellule centrale, ainsi que dans toutes les cellules images. Adopter une telle approche accroît, en revanche, de façon considérable le coût global du calcul, mais s'avère rigoureusement indispensable pour décrire correctement les interactions à très longue portée.

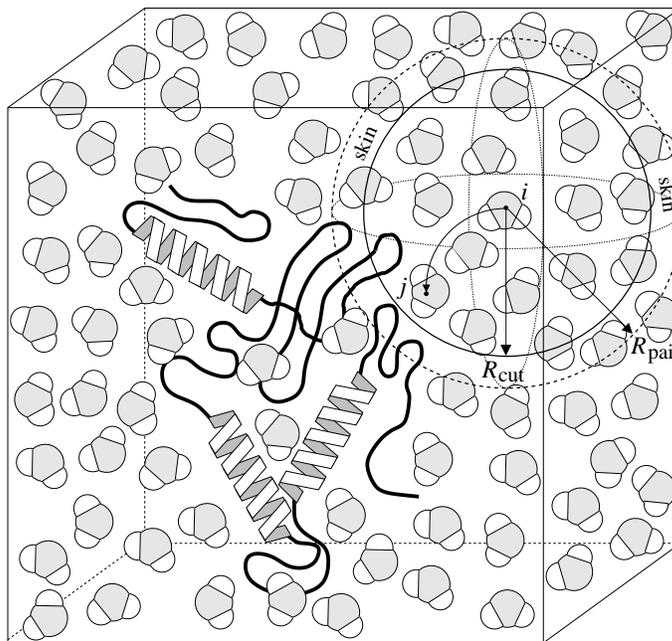


Figure 2: Utilisation d'une sphère de troncature de rayon  $R_{\text{cut}}$  pour limiter le calcul des interactions de la particule  $i$  avec ses voisins dans la convention de l'image minimale. Une sphère de rayon  $R_{\text{pair}}$ , supérieur à  $R_{\text{cut}}$ , est employée pour établir la liste de tous les voisins de  $i$ . La liste des paires de particules  $\{i, j\}$  est mise à jour périodiquement.

### 1.2. De la légitimité des simulations de dynamique moléculaire

Le principe de la dynamique moléculaire, particulièrement simple, consiste à générer les trajectoires d'un ensemble fini de particules en intégrant de façon numérique les équations classiques du mouvement. Cette approche, *a priori* discutable, se trouve justifiée par deux faits remarquables: (i) compte tenu de l'approximation de Born–Oppenheimer, il est possible de dissocier le mouvement des électrons de celui des noyaux, et, (ii) du fait que, dans la majorité des cas, la longueur d'onde de de Broglie associée à une particule soit nettement inférieure à la distance intermoléculaire, les effets quantiques sont globalement négligeables. Les trajectoires ainsi déterminées sont utilisées pour évaluer les propriétés statiques et dynamiques par des moyennes temporelles, qui coïncident avec les moyennes statistiques pour des systèmes ergodiques:

$$\lim_{t \rightarrow \infty} \overline{\mathcal{A}}_t = \langle \mathcal{A} \rangle \quad (1)$$

Ici,  $\mathcal{A}$  désigne une propriété quelconque, observable,  $\overline{\mathcal{A}}_t$ , sa moyenne temporelle, et  $\langle \mathcal{A} \rangle$ , sa moyenne d'ensemble statistique. Dans la pratique, on constate que le postulat d'ergodicité est vérifié, tout au moins pour des liquides simples.

## 2. Les équations de la dynamique moléculaire

En dynamique moléculaire classique, [?, ?, ?] la trajectoire des différentes composantes du système moléculaire est générée par intégration des équations de Newton du mouvement, qui pour chaque particule  $i$ , s'écrivent:

$$\begin{cases} m_i \frac{d^2 \mathbf{r}_i(t)}{dt^2} = \mathbf{f}_i(t) \\ \mathbf{f}_i(t) = -\frac{\partial \mathcal{V}(\mathbf{r}^N)}{\partial \mathbf{r}_i(t)} \end{cases} \quad (2)$$

$\mathcal{V}(\mathbf{r}^N)$  est la fonction énergie potentielle du système à  $N$  particules, qui ne dépend que des coordonnées cartésiennes  $\{\mathbf{r}_i\}$ . Les équations (2) sont intégrées numériquement en utilisant un pas,  $\delta t$ , infinitésimal, garantissant la conservation de l'énergie du système — typiquement 1–2 fs.

Il est, cependant, parfaitement illusoire d'espérer générer une trajectoire *exacte* aux temps longs, dès lors que les équations newtonniennes du mouvement sont résolues numériquement, avec un pas d'intégration fini. L'exactitude de la solution des équations (2) n'est, cependant, pas aussi crucial qu'il n'y paraît au premier abord. Ce qui importe, en revanche, est que le comportement statistique de la trajectoire soit, quant à lui, correct, pour garantir la reproduction des propriétés dynamiques et thermodynamiques du système avec suffisamment de précision. Cette condition n'est remplie que si l'intégrateur utilisé pour propager le mouvement possède la propriété de symplecticité. [?] Un propagateur dit *symplectique* conserve la métrique invariante de l'espace phase. Il en résulte que l'erreur associée à ce propagateur est nécessairement bornée:

$$\lim_{n_{\text{pas}} \rightarrow \infty} \left( \frac{1}{n_{\text{pas}}} \right) \sum_{k=1}^{n_{\text{pas}}} \left| \frac{\mathcal{E}(k\delta t) - \mathcal{E}(0)}{\mathcal{E}(0)} \right| \leq \varepsilon_{\text{MD}} \quad (3)$$

Ici,  $n_{\text{pas}}$  désigne le nombre de pas de la simulation,  $\mathcal{E}(0) \equiv \mathcal{H}(\mathbf{r}^N, \mathbf{p}^N; 0)$ , l'énergie totale initiale du système équilibré, et  $\varepsilon_{\text{MD}}$ , la borne supérieure de conservation de l'énergie — *viz.*  $10^{-4}$  constitue une valeur acceptable. En supposant que le pas d'intégration est limité, l'intégration des équations du mouvement ne donne pas lieu à une croissance erratique de l'erreur de conservation de l'énergie, qui pourrait affecter significativement le comportement statistique de la dynamique moléculaire aux temps longs. Il est intéressant de noter que pour un système hamiltonien, la propriété de symplecticité entraîne que le jacobien:

$$\mathbf{J}(\mathbf{\Gamma}_{\delta t}, \mathbf{\Gamma}_0) = \frac{\partial(\mathbf{\Gamma}_{\delta t}^1, \dots, \mathbf{\Gamma}_{\delta t}^N)}{\partial(\mathbf{\Gamma}_0^1, \dots, \mathbf{\Gamma}_0^N)} \quad (4)$$

soit unitaire.  $\mathbf{\Gamma}_0$  représente le vecteur initial de l'espace des phases à  $N$  dimensions, contenant toutes les variables de position,  $\mathbf{r}$ , et d'impulsion,  $\mathbf{p}$ , décrivant le système.

Comme nous l'avons souligné précédemment, la nature à longue portée des interactions charge–dipôle, *i.e.*  $1/r^2$ , et, *a fortiori*, des interactions charge–charge, *i.e.*  $1/r$ , impose l'utilisation d'algorithmes adaptés pour la prise en compte de telles contributions, pouvant augmenter le coût de la simulation de façon significative. En écrivant les équations (2) plus formellement:

$$\mathbf{\Gamma}_t = e^{i\mathcal{L}t} \mathbf{\Gamma}_0 \quad (5)$$

où  $\mathcal{L}$  désigne l'opérateur de Liouville permettant de générer une distribution  $\varrho(\mathbf{\Gamma}, t)$  pour un ensemble thermodynamique donné, selon:

$$\frac{\partial \varrho(\mathbf{\Gamma}, t)}{\partial t} = -i\mathcal{L}\varrho(\mathbf{\Gamma}, t) \quad (6)$$

et en appliquant la formule de Trotter:

$$e^{i\mathcal{L}\Delta t} = e^{i\mathcal{L}_1 \frac{\Delta t}{2}} e^{i\mathcal{L}_2 \Delta t} e^{i\mathcal{L}_1 \frac{\Delta t}{2}} + \mathcal{O}(\Delta t^3) \quad (7)$$

dans laquelle  $i\mathcal{L} = i\mathcal{L}_1 + i\mathcal{L}_2$ , la déconvolution des contributions à courte et à longue portée devient évidente. Ainsi, des pas d'intégration de longueurs différentes peuvent être utilisés en fonction de la nature de l'interaction considérée. En découplant, par exemple, l'hamiltonien total,

$\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N)$ , régissant le comportement du système en une contribution cinétique,  $\mathcal{T}(\mathbf{p}^N)$ , de valence,  $\mathcal{V}_{\text{valence}}(\mathbf{r}^N)$ , et électrostatique à courte,  $\mathcal{V}_{\text{short}}(\mathbf{r}^N)$ , et à longue,  $\mathcal{V}_{\text{long}}(\mathbf{r}^N)$ , portées, il s'ensuit que:

$$e^{i\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N)\Delta t} = e^{i\mathcal{V}_{\text{long}}(\mathbf{r}^N)\frac{\Delta t}{2}} \left\{ e^{i\mathcal{V}_{\text{short}}(\mathbf{r}^N)\frac{\Delta t}{2n}} \left[ e^{i\mathcal{V}_{\text{valence}}(\mathbf{r}^N)\frac{\Delta t}{2pn}} e^{i\mathcal{T}(\mathbf{p}^N)\frac{\Delta t}{pn}} \right. \right. \quad (8)$$

$$\left. \left. \times e^{i\mathcal{V}_{\text{valence}}(\mathbf{r}^N)\frac{\Delta t}{2pn}} \right]^p e^{i\mathcal{V}_{\text{short}}(\mathbf{r}^N)\frac{\Delta t}{2n}} \right\}^n e^{i\mathcal{V}_{\text{long}}(\mathbf{r}^N)\frac{\Delta t}{2}}$$

Ce partitionnement des diverses contributions de  $\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N)$  constitue l'idée centrale des approches dites à pas d'intégration multiples, comme RESPA. [?] Il met clairement en évidence l'utilisation de pas distincts pour la mise à jour de ces contributions, réduisant de fait l'investissement informatique de la simulation de manière appréciable.

## 2.1. La fonction énergie potentielle

Cette fonction constitue la clé de voûte des calculs moléculaires, car son rôle est de reproduire les interactions intra- et intermoléculaires du système aussi fidèlement que possible. En principe, cette fonctionnelle s'écrit comme une somme de  $N$  termes:

$$\mathcal{V}(\mathbf{r}^N) = \sum_i v_1(\mathbf{r}_i) + \sum_i \sum_{j>i} v_2(\mathbf{r}_i, \mathbf{r}_j) + \sum_i \sum_{j>i} \sum_{k>j>i} v_3(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) + \dots \quad (9)$$

dans laquelle  $v_1(\mathbf{r}_i)$ ,  $v_2(\mathbf{r}_i, \mathbf{r}_j)$ , ... représentent le potentiel intramoléculaire, le potentiel d'interaction de paires, ...  $\mathcal{V}(\mathbf{r}^N)$  caractérise donc un problème à  $N$  corps, même si l'on peut arguer du fait que  $v_2(\mathbf{r}_i, \mathbf{r}_j)$  constitue sans doute le terme le plus significatif de la contribution intermoléculaire. [?] Ce point de vue est d'ailleurs à l'origine de l'approximation de paires, dans laquelle les effets d'ordre supérieur sont partiellement inclus dans un potentiel effectif:

$$\mathcal{V}(\mathbf{r}^N) \simeq \sum_i v_1(\mathbf{r}_i) + \sum_i \sum_{j>i} v_2^{\text{effectif}}(r_{ij}) \quad (10)$$

Cette approximation est utilisée dans la plupart des champs de force commerciaux, en particulier ceux destinés à l'étude de systèmes macromoléculaires, pour lesquels le temps de calcul est intimement relié à la complexité de  $\mathcal{V}(\mathbf{r}^N)$ . Parmi ces fonctions énergie potentielle, nous avons souvent eu

recours à celle de la suite de programmes AMBER: [?, ?]

$$\begin{aligned}
\mathcal{V}(\mathbf{r}) = & \sum_{\text{liaisons}} k_r (r - r_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 \\
& + \sum_{\text{dièdres}} \sum_n \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] \\
& + \frac{1}{k_{\text{vdW}}^{1-4}} \sum_{\substack{i < j \\ \{i,j\} \in 1-4}} \varepsilon_{ij} \left[ \left( \frac{R_{ij}^*}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{ij}^*}{r_{ij}} \right)^6 \right] + \frac{1}{k_{\text{Coulomb}}^{1-4}} \sum_{\substack{i < j \\ \{i,j\} \in 1-4}} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1 r_{ij}} \\
& + \sum_{\substack{i < j \\ \{i,j\} > 1-4}} \varepsilon_{ij} \left[ \left( \frac{R_{ij}^*}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{ij}^*}{r_{ij}} \right)^6 \right] + \sum_{\substack{i < j \\ \{i,j\} > 1-4}} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1 r_{ij}}
\end{aligned} \tag{11}$$

dans laquelle  $k_r$  et  $r_0$  représentent, respectivement, la constante de force de liaison et la longueur de liaison à l'équilibre,  $k_\theta$  et  $\theta_0$ , la constante de force angulaire et l'angle de valence à l'équilibre,  $V_n/2$ ,  $n$  et  $\gamma$ , la barrière de torsion, sa périodicité et sa phase.  $\epsilon_0$  et  $\epsilon_1$  sont respectivement la permittivité du vide et la permittivité relative.  $q_i$  est la charge partielle portée par l'atome  $i$ .  $R_{ij}^*$  et  $\varepsilon_{ij}$  correspondent aux paramètres de van der Waals pour la paire atomique  $\{ij\}$ , obtenus à partir des règles de combinaison de Lorentz–Berthelot:

$$\begin{cases} \varepsilon_{ij} & = \sqrt{\varepsilon_i \varepsilon_j} \\ R_{ij}^* & = R_i^* + R_j^* \end{cases} \tag{12}$$

Compte tenu du fait que, lors de leur paramétrisation, généralement à partir de calculs quantochimiques sophistiqués, les dièdres contiennent d'emblée une composante électrostatique et de van der Waals, la plupart des champs de force commerciaux font une distinction entre les interactions d'atomes séparés par exactement trois liaisons chimiques (termes dits "1-4"), et toutes les autres, dès lors que  $\{i, j\}$  ne sont pas séparés par une ou deux liaisons (voir Figure 3). Les contributions "1-4" sont pondérées par les facteurs  $1/k_{\text{vdW}}^{1-4}$  et  $1/k_{\text{Coulomb}}^{1-4}$  apparaissant dans l'expression (11).

La description des angles dièdres constitue un aspect particulièrement crucial de la mise au point de fonctions énergie potentielle. Le comportement du potentiel de torsion,  $\mathcal{V}(\phi)$ , est souvent trop complexe pour qu'un seul terme de la série de Fourier dans l'équation (11) soit suffisant pour en garantir une restitution fidèle. Le cas des phospholipides est une bonne illustration de cette difficulté. En effet, la clé d'une reproduction précise lors d'une simulation statistique des paramètres d'ordre,

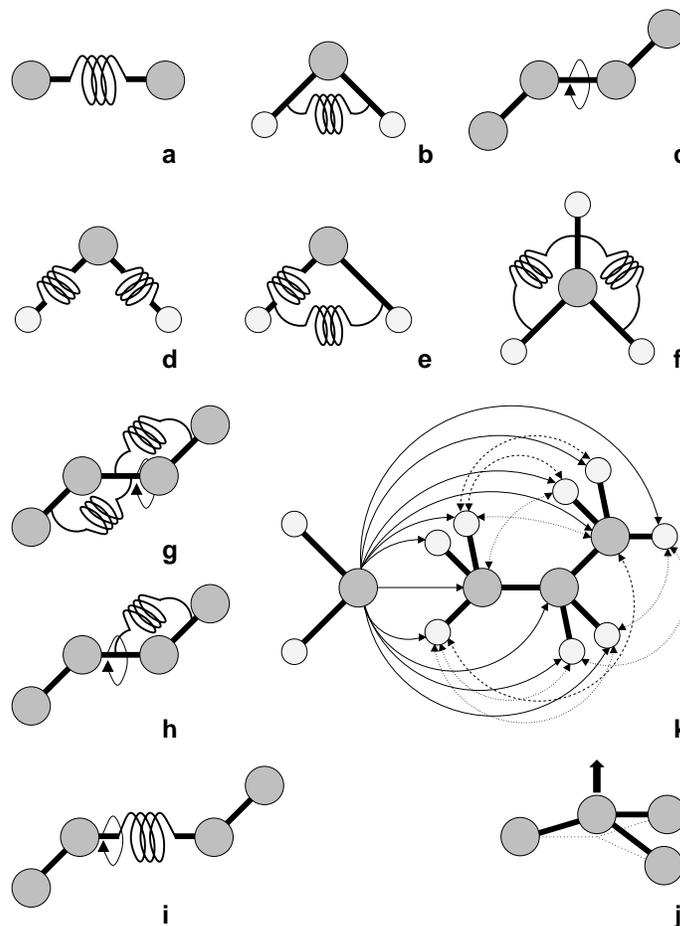


Figure 3: Illustration des différents termes d'une fonction énergie potentielle empirique. Les termes **a-j** représentent le champ de force de valence, parmi lesquels **d-j** sont des termes dits "croisés". **j** est un terme dit de *out-of-plane*, garantissant que l'atome central reste dans le plan des trois voisins auxquels il est lié chimiquement. **k** caractérise les interactions coulombiennes et de van der Waals d'atomes non liés chimiquement: Interactions intermoléculaires (trait plein), interactions intramoléculaires dites "1-4" (pointillés), et interactions intramoléculaires > 1-4 (tirets).

$S_{CD}$ , des chaînes aliphatiques réside dans la description des dièdres de celles-ci par un potentiel approprié. Seul un potentiel à plusieurs termes sera capable de rendre compte du subtil équilibre *trans-gauche* le long des caténaire. Le potentiel de Ryckaert et Belleman, [?]

$$\mathcal{V}(\phi) = \sum_{i=1}^6 a_i \cos^i \phi \quad (13)$$

dans lequel les coefficients  $a_i$  ont été optimisés sur la base de la rotation interne du *n*-butane, est employé le plus souvent pour la simulation de bicouches phospholipidiques.

Tout aussi subtil est l'équilibre entre les différents termes du champ de force. La composante torsionnelle ne représente qu'une facette de cet équilibre. Le choix des paramètres de Lennard-Jones,

$R_{ij}^*$  et  $\varepsilon_{ij}$ , et des charges partielles,  $q_i$ , n'est pas moins critique, dans la mesure où la justesse de celui-ci conditionne la précision des quantités thermodynamiques calculées. Une fonction énergie potentielle est un édifice dont les éléments constitutifs ont été calibrés pour la reproduction globale de propriétés physico-chimiques clé, sans qu'à ces éléments ne soit nécessairement attachée une signification physique ou chimique.

Une des approches en vogue pour la mise au point de modèles de charges ponctuelles consiste en un ajustement sur le potentiel électrostatique, véritable empreinte digitale de la molécule. En se limitant à un développement monopolaire, le jeu optimal de  $N_{\text{atomes}}$  charges nettes atomiques,  $\{q_k\}$ , est obtenu en minimisant la fonctionnelle:

$$f(\{q_k\}) = \sum_{i=1}^{N_{\text{points}}} \left[ V^{\text{référence}}(\mathbf{r}_i) - \sum_{j=1}^{N_{\text{atomes}}} \frac{q_j}{r_{ij}} \right]^2 \quad (14)$$

où  $V^{\text{référence}}(\mathbf{r}_i)$  est le potentiel électrostatique évalué au point  $\mathbf{r}_i$  d'une grille de  $N_{\text{points}}$  entourant la molécule.  $V^{\text{référence}}(\mathbf{r}_i)$  est obtenu par les calculs de la chimie quantique, *i.e.*  $\langle \Psi | 1/r_i | \Psi \rangle$ , généralement à un niveau de sophistication avancé.

La détermination des paramètres de Lennard-Jones se révèle souvent plus délicate. Une voie possible consiste en un ajustement des contributions de répulsion et de dispersion à partir d'un nombre élevé de calculs quanto-chimiques très fins, réalisés sur différentes configurations. Cette approche est généralement applicable à des systèmes de taille réduite, comme par exemple l'hétéro-dimère formamide-eau, pour lequel on souhaite définir le potentiel d'interaction de van der Waals atome-atome. Une approche plus heuristique, alternative aux calculs de la chimie quantique, repose sur les simulations statistiques. Partant d'un jeu de paramètres de Lennard-Jones,  $\{R_{ij}^*, \varepsilon_{ij}\}$ , donné, peut-on reproduire quantitativement les grandeurs thermodynamiques fondamentales d'un liquide moléculaire, comme sa densité,  $\rho$ , son enthalpie de vaporisation,  $\Delta H_{\text{vap}}$ , voire son coefficient de diffusion,  $D$  ?

Dans de nombreux cas, la description minimaliste imposée par la fonctionnelle (11) peut s'avérer insuffisante. Comme c'est le cas de la plupart des champs de force commerciaux, cette fonctionnelle a un but polyvalent, même si elle a été, au départ, conçue pour l'étude de biopolymères et, plus particulièrement, de protéines et d'acides nucléiques. L'investigation plus fine de petites molécules

organiques, souvent nécessaire dans le domaine émergent du *drug design de novo*, requiert un niveau de calcul dépassant les hypothèses simplistes de l'expression (11). Parmi celles-ci, on peut citer le remplacement du terme harmonique d'élongation de liaison (voir Figure 3 **a**) par un potentiel dissociatif de Morse:

$$\mathcal{V}(r) = D_0 \left[ e^{-\alpha(r-r_0)} - 1 \right]^2 \quad (15)$$

dans lequel  $D_0$  représente précisément l'énergie de dissociation et  $r_0$ , la longueur de liaison d'équilibre. La description harmonique, elle-même, peut apparaître inadaptée, dès lors que les effets d'anharmonicité ne sont plus négligeables. Aussi bien pour l'élongation de liaisons que pour la déformation d'angles de valence, des corrections cubiques et quartiques peuvent être nécessaires:

$$\mathcal{V}(r) = k_r (r - r_0)^2 \left[ 1 - k'_r (r - r_0) + k''_r (r - r_0)^2 \right] \quad (16)$$

L'ouverture d'un angle de valence a pour résultat de réduire les longueurs de liaison — effet clairement absent dans l'équation (11). Pour corriger ce défaut, on peut enrichir la fonction énergie potentielle de termes dits de couplage, comme par exemple:

$$\mathcal{V}(r-\theta) = k_{r\theta} (r - r_0) (\theta - \theta_0) \quad (17)$$

Outre le *stretch-bend* caractérisé par l'expression ci-dessus, il est possible d'introduire d'autres termes de couplage dans  $\mathcal{V}(\mathbf{r}^N)$ , comme le montre la Figure 3, **e-i**. Hormis l'alourdissement du calcul de la fonction énergie potentielle à chaque pas,  $\delta t$ , d'intégration des équations (2), l'introduction de termes de couplage implique également un effort supplémentaire de paramétrisation.

Le traitement de l'électrostatique est sans doute le plus problématique, ce, à bien des égards. Tout d'abord, dans la mesure où l'approximation monopolaire de la relation (14) n'est pas nécessairement suffisante pour n'importe quelle molécule. Il peut alors être désirable d'inclure des dipôles permanents au simple modèle de charges centrées sur les atomes. Plus cruciaux sont les effets d'induction, évidemment absents dans l'approximation additive de paires (10). Pour pallier cette difficulté, le nombre de solutions disponibles reste limité. L'approche la moins onéreuse, encore très largement utilisée dans les simulations de systèmes macromoléculaires, consiste à accroître artificiellement les

charges nettes atomiques, de façon à ce que ces dernières reproduisent un moment dipolaire permanent caractéristique, non pas de la phase gazeuse, mais d'un environnement polaire. Il s'agit là d'une représentation des effets de polarisation au sens "moyen" du terme. Plus rigoureuse, mais également bien plus coûteuse, la solution de choix consiste à introduire explicitement des contributions de polarisabilité dans le champ de force. Dans ce cas, la contribution électrostatique totale s'écrit comme:

$$\mathcal{V}_{\text{elec}}(\mathbf{r}^N) = \frac{1}{2} \sum_i q_i V_i \quad (18)$$

Ici, le potentiel électrostatique,  $V_i$ , naît de deux sources — d'une part, des charges ponctuelles du système; d'autre part, des moments multipolaire induit au site  $i$ . En nous limitant ici au moment dipolaire induit,  $\boldsymbol{\mu}_i$ , lié linéairement au champ électrique,  $\mathbf{E}_i$ , créé en ce point par tous les autres sites polarisables  $j \neq i$ , à savoir,  $\boldsymbol{\mu}_i = \alpha_i \mathbf{E}_i$ , le potentiel s'exprime sous la forme:

$$V_i = \sum_{j \neq i} \left[ \frac{q_j}{4\pi\epsilon_0 r_{ij}} + \frac{\mathbf{r}_{ij} \cdot \boldsymbol{\mu}_j}{4\pi\epsilon_0 r_{ij}^3} \right] \quad (19)$$

Même en utilisant comme point de départ au temps  $t + \delta t$ , les moments induits,  $\boldsymbol{\mu}_i$ , obtenus au temps  $t$ , la convergence de ces derniers au cours d'une simulation de dynamique moléculaire d'un liquide polarisable accroît le temps de calcul d'un facteur  $\approx 3-4$  par rapport à une simulation dans l'approximation additive de paires.

## 2.2. Les intégrateurs de la dynamique moléculaire

Il existe plusieurs approches pour intégrer numériquement les équations de Newton du mouvement (2). Parmi celles-ci, nous en distinguerons ici trois. Sans doute la plus simple, l'algorithme de Verlet est basé sur l'utilisation du triplet  $\{\mathbf{r}_i(t), \mathbf{r}_i(t - \delta t), \mathbf{a}_i(t)\}$ , où  $\mathbf{a}_i(t) = \ddot{\mathbf{r}}_i(t) = d^2\mathbf{r}_i(t)/dt^2 = \mathbf{f}_i(t)/m_i$  désigne l'accélération de la particule  $i$ . [?] La modification des positions des particules est obtenue à partir d'un développement de série de Taylor de la position en  $t - \delta t$  et en  $t + \delta t$ , conduisant à:

$$\mathbf{r}_i(t + \delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t - \delta t) + \mathbf{a}_i(t) \delta t^2 \quad (20)$$

qui implique des possibilités d'erreurs en  $\mathcal{O}(\delta t^4)$ . Il est à noter que les vitesses,  $\mathbf{v}_i(t) = \dot{\mathbf{r}}_i(t) =$

$dr_i(t)/dt$ , n'interviennent pas directement dans cette formule. Elles ont été éliminées lors du développement en série de Taylor de  $\mathbf{r}_i(t + \delta t)$  et de  $\mathbf{r}_i(t - \delta t)$ . Bien que non nécessaires à la description des trajectoires, leur calcul est obligatoire pour évaluer l'énergie cinétique,  $\mathcal{T}(\mathbf{p})$ , qui ne dépend que des variables d'impulsion  $\mathbf{p}$ , et, par conséquent, l'énergie totale,  $\mathcal{E} \equiv \mathcal{H}(\mathbf{r}^N, \mathbf{p}^N)$ , du système, selon:

$$\mathbf{v}_i(t) = \frac{\mathbf{r}_i(t + \delta t) - \mathbf{r}_i(t - \delta t)}{2 \delta t} \quad (21)$$

conduisant, à chaque pas d'intégration, à des erreurs en  $\mathcal{O}(\delta t^2)$ .

L'algorithme *leap-frog*, dérivé du précédent, utilise le triplet  $\{\mathbf{r}_i(t), \mathbf{v}_i(t - \delta t/2), \mathbf{a}_i(t)\}$ . L'origine de son nom apparaît clairement dans l'écriture de l'algorithme:

$$\begin{cases} \mathbf{r}_i(t + \delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t + \frac{\delta t}{2}) \delta t \\ \mathbf{v}_i(t + \frac{\delta t}{2}) = \mathbf{v}_i(t - \frac{\delta t}{2}) + \mathbf{a}_i(t) \delta t \end{cases} \quad (22)$$

Dans la pratique, la première étape est le calcul de  $\mathbf{v}_i(t + \delta t/2)$ , à partir duquel on déduit  $\mathbf{v}_i(t)$ , nécessaire à l'évaluation du terme cinétique  $\mathcal{T}(\mathbf{p})$ , selon:

$$\mathbf{v}_i(t) = \frac{\mathbf{v}_i(t + \frac{\delta t}{2}) + \mathbf{v}_i(t - \frac{\delta t}{2})}{2} \quad (23)$$

Enfin, l'algorithme Verlet–vitesses corrige le défaut principal de Verlet classique et *leap-frog*, à savoir la définition des vitesses — dont l'erreur associée est en  $\mathcal{O}(\delta t^2)$ . L'incorporation explicite des vitesses dans l'algorithme de Verlet peut s'écrire:

$$\begin{cases} \mathbf{r}_i(t + \delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t) \delta t + \frac{1}{2} \mathbf{a}_i(t) \delta t^2 \\ \mathbf{v}_i(t + \delta t) = \mathbf{v}_i(t) + \frac{\mathbf{a}_i(t) + \mathbf{a}_i(t + \delta t)}{2} \delta t \end{cases} \quad (24)$$

Ce schéma implique les deux étapes suivantes:

$$\mathbf{v}_i(t + \frac{\delta t}{2}) = \mathbf{v}_i(t) + \frac{1}{2} \mathbf{a}_i(t) \delta t \quad (25)$$

qui permet, alors, d'évaluer les forces thermodynamiques  $\mathbf{f}_i$ , ainsi que les accélérations  $\mathbf{a}_i$ , au temps  $t + \delta t$ . A la suite de quoi:

$$\mathbf{v}_i(t + \delta t) = \mathbf{v}_i(t + \frac{\delta t}{2}) + \frac{1}{2} \mathbf{a}_i(t + \delta t) \delta t \quad (26)$$

dont on peut déduire l'énergie cinétique à l'instant  $t + \delta t$ , alors que le l'énergie potentielle,  $\mathcal{V}(\mathbf{r}^N)$ , au même moment, est calculée dans la boucle des forces.

Lors de l'intégration simultanée des équations de Newton du mouvement, l'énergie totale du système est conservée, et, pour peu que le volume soit maintenu constant, la simulation générera un ensemble statistique micro-canonique, c'est à dire,  $(N, V, \mathcal{E})$ . Une telle situation n'est, cependant, pas toujours satisfaisante et il peut être désirable de réaliser des simulations dans lesquelles la température, ou la pression, sont des grandeurs indépendantes, et non des propriétés dérivées.

### 2.3. Dynamique moléculaire à température constante

Plusieurs méthodes, plus ou moins sophistiquées, permettant la réalisation de simulations de dynamique moléculaire isothermes ont été proposées. Sans doute la plus simple, consiste à recalibrer périodiquement les vitesses d'un facteur  $\sqrt{T/T_{\mathcal{J}}}$ , où  $T_{\mathcal{J}}$  désigne la température cinétique instantanée — soit  $2\mathcal{T}(\mathbf{p})/3Nk_B$  — et  $T$ , la température désirée. Cependant, l'application de ce facteur à chaque pas d'intégration ne conduit pas à une dynamique moléculaire newtonienne, à proprement parler. La mécanique newtonienne implique que soient conservées l'énergie et la quantité de mouvement. Une dynamique moléculaire à température cinétique constante nécessite que soient résolues les équations du mouvement contraintes: [?]

$$\begin{cases} \dot{\mathbf{r}}_i = \frac{\mathbf{p}_i}{m_i} \\ \dot{\mathbf{p}}_i = \mathbf{f}_i - \xi(\mathbf{r}; \mathbf{p}) \mathbf{p} \end{cases} \quad (27)$$

dans lesquelles  $\xi(\mathbf{r}; \mathbf{p})$  peut être assimilé à un coefficient de friction garantissant que  $\dot{T}_{\mathcal{J}} = 0$ . Cette contrainte est choisie de manière à perturber le moins possible la trajectoire newtonienne:

$$\xi(\mathbf{r}; \mathbf{p}) = \frac{\sum_i \mathbf{p}_i \cdot \mathbf{f}_i}{\sum_i |\mathbf{p}_i|^2} \quad (28)$$

Une seconde approche, plus rigoureuse, consiste à introduire dans les équations du mouvement un degré de liberté,  $s$ , supplémentaire. La vitesse de la particule  $i$  s'écrit alors  $\mathbf{v}_i = s \dot{\mathbf{r}}_i = \mathbf{p}_i/m_i s$ . Au degré de liberté  $s$ , assimilable au thermostat du système, sont associés les termes potentiel et cinétique:

$$\begin{cases} \mathcal{V}_s = (f + 1) k_B T \ln s \\ \mathcal{T}_s = \frac{1}{2} \mathcal{Q} \dot{s}^2 \end{cases} \quad (29)$$

où  $\mathcal{Q}$  désigne le paramètre d'inertie thermique, qui régule les fluctuations de la température, et  $f$ , le nombre de degrés de liberté du système —  $3N - 3$  si l'impulsion totale  $\mathbf{p}$  est constante. Une telle approche est connue sous le vocable de lagrangien étendu, dans la mesure où celui-ci prend, à présent, la forme  $\mathcal{L}_s(\mathbf{r}; \mathbf{p}) = \mathcal{T}(\mathbf{p}) + \mathcal{T}_s(\mathbf{p}) - \mathcal{V}(\mathbf{r}) - \mathcal{V}_s(\mathbf{r})$ . Les équations du mouvement peuvent alors être réétablies comme:

$$\begin{cases} \ddot{\mathbf{r}}_i = \frac{\mathbf{f}_i}{m_i s^2} - 2 \frac{\dot{s} \dot{\mathbf{r}}_i}{s} \\ \mathcal{Q} \ddot{s} = \sum_i m_i \dot{\mathbf{r}}_i^2 s - \frac{(f + 1) k_B T}{s} \end{cases} \quad (30)$$

Ce formalisme, proposé par Nosé, [?] a été revisité par Hoover, [?, ?] qui y a supprimé le paramètre  $s$  dépendant du temps. Dans les équations du mouvement contraintes (27), le terme de friction est maintenant donné par une équation différentielle du premier ordre:

$$\dot{\xi} = \frac{f}{\mathcal{Q}} k_B (T_{\mathcal{T}} - T) \quad (31)$$

La quantité conservée ici est l'hamiltonien  $\mathcal{H}_s(\mathbf{r}; \mathbf{p}) = \mathcal{T}(\mathbf{p}) + \mathcal{T}_s(\mathbf{p}) + \mathcal{V}(\mathbf{r}) + \mathcal{V}_s(\mathbf{r})$ .

La dernière approche, dite *weak coupling*, [?] consiste à relaxer la température cinétique instantanée  $T_{\mathcal{T}}(t)$  vers une valeur de référence  $T$ , selon:

$$\frac{dT_{\mathcal{F}}(t)}{dt} = \frac{T - T_{\mathcal{F}}(t)}{\tau_T} \quad (32)$$

où  $\tau_T$  représente précisément le temps de relaxation associé aux fluctuations de température. L'énergie cinétique est modifiée d'une quantité  $\Delta\mathcal{F}$  donnée par:

$$\Delta\mathcal{F} = \frac{1}{2} (\chi^2 - 1) N k_B T_{\mathcal{F}}(t) \quad (33)$$

durant un pas d'intégration  $\delta t$ , en procédant au réajustement des vitesses d'un facteur  $\chi$ , tel que:

$$\chi = \left[ 1 + \frac{\delta t}{\tau_T} \left( \frac{T}{T_{\mathcal{F}}(t)} - 1 \right) \right]^{1/2} \quad (34)$$

Ce couplage aperiodique à un "réservoir de chaleur", grâce à un processus du premier ordre, ne conduit pas à des réponses oscillantes aux changements de température. En revanche, il a été prouvé que cet algorithme ne conduisait pas rigoureusement à une distribution canonique, à la différence de l'approche de Nosé et Hoover.

#### 2.4. *Dynamique moléculaire à pression constante*

Là encore, plusieurs méthodes, plus ou moins sophistiquées, permettent de maintenir la cellule de simulation à une pression constante au cours du temps. Il peut, en effet, être désirable, dans certaines situations, de générer des trajectoires dans l'ensemble isobare–isotherme,  $(N, P, T)$ .

Tout comme pour le maintien de la température à une valeur constante, le formalisme du lagrangien étendu est applicable à la pression. Initialement proposée par Andersen, [?] cette méthode implique le couplage du système à une variable externe,  $V$ , caractérisant le volume de la boîte de simulation. Ce couplage symbolise l'action qu'exercerait un piston sur le système, à laquelle on associe les termes cinétique et potentiel:

$$\begin{cases} \mathcal{V}_V = \frac{1}{2} m_P \dot{V}^2 \\ \mathcal{F}_V = P V \end{cases} \quad (35)$$

où  $m_P$  est assimilable à la masse du piston, et  $P$  représente la pression désirée. La mise à l'échelle des

variables de position,  $\mathbf{r}$ , et de vitesse,  $\mathbf{v}$ , sous la forme  $\mathbf{s} = \mathbf{r}/V^{1/3}$  et  $\dot{\mathbf{s}} = \mathbf{v}/V^{1/3}$ , permet d'écrire les énergies cinétique et potentielle comme:  $\mathcal{V}(\mathbf{r}) \equiv \mathcal{V}(V^{1/3}\mathbf{s})$  et  $\mathcal{T}(\mathbf{p}) = \frac{1}{2} m V^{2/3} \sum_i \dot{s}_i^2$ . Il s'ensuit qu'à partir du lagrangien,  $\mathcal{L}_V(\mathbf{r}; \mathbf{p}) = \mathcal{T}(\mathbf{p}) + \mathcal{T}_V(\mathbf{p}) - \mathcal{V}(\mathbf{r}) - \mathcal{V}_V(\mathbf{r})$ , il est possible d'établir les nouvelles équations du mouvement:

$$\begin{cases} \ddot{\mathbf{s}}_i = \frac{\mathbf{f}_i}{m_i V^{1/3}} - \frac{2}{3} \frac{\dot{\mathbf{s}}_i \dot{V}}{V} \\ \ddot{V} = \frac{P_{\mathcal{P}} - P}{m_P} \end{cases} \quad (36)$$

dans lesquelles la force,  $\mathbf{f}_i$  et la pression instantanée,  $P_{\mathcal{P}}$  — dérivée du viriel,  $P_{\mathcal{P}} = \frac{1}{V} (Nk_B T - \frac{1}{2} \sum_i \mathbf{r}_i \cdot \mathbf{f}_i)$  — sont évaluées à partir des coordonnées et des impulsions non remises à l'échelle. Ici, la quantité conservée au cours de la dynamique moléculaire est l'hamiltonien du système étendu,  $\mathcal{H}_V(\mathbf{r}; \mathbf{p}) = \mathcal{T}(\mathbf{p}) + \mathcal{T}_V(\mathbf{p}) + \mathcal{V}(\mathbf{r}) + \mathcal{V}_V(\mathbf{r})$ , c'est à dire son enthalpie, à laquelle s'ajoute une contribution cinétique de  $\frac{1}{2}k_B T$  due à la fluctuation du volume de la boîte. Il convient de souligner que, formellement, cet algorithme génère une distribution isobare–enthalpique,  $(N, P, H)$ . Son couplage à un thermostat, tel que celui régi par l'équation (30), permet d'accéder à l'ensemble isobare–isotherme.

Il a été observé que cette approche conduisait à des oscillations de  $P_{\mathcal{P}}$ , dépendant de la masse du piston,  $m_P$ . Feller *et al.* ont proposé un schéma supprimant cet effet indésirable en amortissant le degré de liberté du piston au travers de l'équation de Langevin. En reprenant l'équation (36), il vient:

$$\begin{cases} \ddot{\mathbf{s}}_i = \frac{\mathbf{f}_i}{m_i V^{1/3}} - \frac{2}{3} \frac{\dot{\mathbf{s}}_i \dot{V}}{V} \\ \ddot{V} = \frac{P_{\mathcal{P}} - P}{m_P} - \gamma \dot{V} + R(t) \end{cases} \quad (37)$$

où  $\gamma$  désigne la fréquence de collision, et  $R(t)$  est une force aléatoire, tirée d'une distribution gaussienne de moyenne et de variance nulles. Il est intéressant de noter que  $R(t)$  satisfait à la relation de fluctuation–dissipation, *i.e.*  $\langle R(t_1)R(t_2) \rangle = \frac{1}{m_P} 2k_B T \kappa(t_1 - t_2)$ , dans laquelle  $\kappa(t)$  représente un facteur d'amortissement.

Une seconde approche, proposée par Berendsen *et al.*, [?] est une extension à la pression de la méthode *weak coupling*, décrite précédemment. Tout comme pour l'algorithme à température constante, les

équations du mouvement sont modifiées suite à la relaxation de la pression instantanée  $P_{\mathcal{P}}(t)$  vers une valeur de référence  $P$ , selon:

$$\frac{dP_{\mathcal{P}}(t)}{dt} = \frac{P - P_{\mathcal{P}}(t)}{\tau_P} \quad (38)$$

où  $\tau_P$  est le temps de relaxation associé aux fluctuations de pression. En réajustant les coordonnées atomiques et la taille de la cellule périodique par un facteur  $\varsigma$ , le volume total se voit modifié de  $\Delta V = (\varsigma^3 - 1) V$ , ce qui, naturellement, entraîne une variation de pression, s'exprimant alors sous la forme:

$$\Delta P = \frac{\Delta V}{\beta_{\mathcal{P}} V} \quad (39)$$

dans laquelle  $\beta_{\mathcal{P}}$  désigne la compressibilité isotherme. En résolvant les équations (38) et (39), pour une valeur de  $\varsigma$  donnée, il vient:

$$\varsigma = \left[ 1 - \beta_{\mathcal{P}} \delta t \frac{P - P_{\mathcal{P}}(t)}{\tau_P} \right]^{1/3} \quad (40)$$

Tout comme pour la température, cet algorithme ne conduit pas à un ensemble thermodynamique clairement défini.

## 2.5. Interactions électrostatiques: Au-delà des approximations usuelles

Un aspect particulièrement critique des simulations de la dynamique moléculaire concerne le traitement approprié des interactions électrostatiques. Pour d'évidentes raisons de coût, la troncature sphérique, dont nous avons parlé précédemment, reste encore très largement utilisée, surtout lorsque les échelles de temps explorées dépassent le domaine de la ns, ce qui est le cas, par exemple, de la simulation de bicouches phospholipidiques. Si la nature à longue portée des interactions dipôle-dipôle en  $1/r^3$  est suffisamment limitée pour garantir une reproduction satisfaisante des propriétés structurales et des moyennes d'ensemble statistique du système, il n'en reste pas moins que l'influence de ces interactions sur des processus complexes comme le repliement de protéines mériterait un examen plus détaillé. La présence d'espèces ioniques est nettement plus problématique dans la mesure où l'utilisation d'un *cut-off* sphérique induit de nombreux artéfacts faussant clairement les simu-

lations réalisées. Il est d'ailleurs nécessaire de souligner, que même en se limitant à des espèces dipolaires, l'emploi d'une troncature sphérique entraîne une singularité des dérivées de l'énergie potentielle à la frontière du *cut-off*. Les effets délétères de cette troncature sphérique peuvent être atténués par l'adjonction d'une fonction dite de *switching*, [?] remplaçant le comportement abrupt de type Heavyside par une décroissance plus douce au voisinage de la sphère. Cette approche ne résoud, toutefois, en rien les difficultés introduites par la présence d'ions dans le système. Une méthode pour traiter plus rigoureusement les interactions charge-charge et charge-dipôle, *a priori* peu onéreuse, repose sur la théorie de Debye-Hückel et la résolution de l'équation de Poisson-Boltzmann linéarisée. [?] Comme l'illustre la Figure 5, cette méthode, connue sous le vocable de champ de réaction généralisé, n'élimine cependant pas totalement les artéfacts lorsque  $r \rightarrow R_{\text{cut-off}}$ , le rayon de troncature. L'approche sans doute la plus rigoureuse est celle proposée par Ewald. Partant du fait que la somme coulombienne,

$$\mathcal{V}_{\text{Coulomb}}(\mathbf{r}) = \sum_{i < j} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1 r_{ij}} \quad (41)$$

étendue à la boîte centrale ainsi qu'à toutes ses voisines, ne converge pas formellement, l'idée directrice de la méthode dite des sommes d'Ewald consiste à décomposer la relation (41) en une somme dans l'espace direct et une somme dans l'espace réciproque:

$$\sum_{\mathbf{n}} \frac{1}{|\mathbf{n}|} \mathcal{F}(\mathbf{n}) + \sum_{\mathbf{m}} \frac{1}{|\mathbf{m}|} [1 - \mathcal{F}(\mathbf{n})] \quad (42)$$

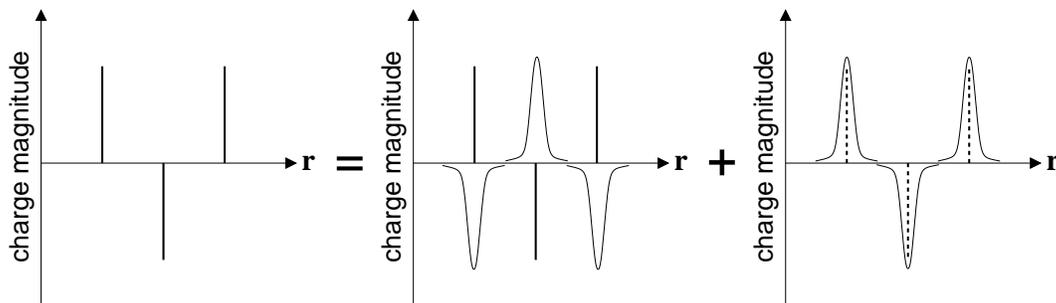


Figure 4: Composantes d'une somme d'Ewald dans un système unidimensionnel de charges ponctuelles. Dans l'espace direct, chaque charge est entourée d'une distribution de charge gaussienne,  $\varrho_i(\mathbf{r})$ , d'amplitude égale, mais de signe opposé. Cette contribution est contre-balançée dans l'espace réciproque par une distribution gaussienne  $\varrho_j(\mathbf{r})$  de signe contraire.

En entourant chaque charge ponctuelle du système d'une distribution de charge gaussienne:

$$\varrho_i(\mathbf{r}) = q_i \alpha^3 \frac{\exp(-\alpha^2 r^2)}{\sqrt{\pi^3}} \quad (43)$$

où  $\alpha$  est un paramètre positif caractérisant la largeur de la distribution gaussienne, la première somme converge rapidement quand  $\mathbf{n} \rightarrow \infty$ , car  $\mathcal{F}(\mathbf{n})$  décroît très rapidement. La contribution dans l'espace direct est de courte portée — voir Figure (4). La seconde somme évaluée dans l'espace réciproque utilise une transformée de Fourier pour résoudre l'équation de Poisson, *i.e.*  $\nabla^2 \mathcal{V}_i(\mathbf{r}) = -4\pi \varrho_i(\mathbf{r})$ . La transformée décroît rapidement, et la somme converge de la même manière. [?]

L'implantation des sommes de réseaux selon le schéma proposé par Ewald dans un code de dynamique moléculaire utilisant un champ de force macromoléculaire comme celui décrit par l'équation (11), peut se résumer par:

$$\begin{aligned} \mathcal{V}_{\text{Ewald}}(\mathbf{r}) &= \frac{1}{2V\epsilon_0} \sum_{\mathbf{k} \neq \mathbf{0}} \frac{\exp(-k^2/4\alpha^2)}{k^2} \left[ \sum_j q_j \exp(-i\mathbf{k} \cdot \mathbf{r}_j) \right] \left[ \sum_j q_j \exp(i\mathbf{k} \cdot \mathbf{r}_j) \right] \\ &+ \frac{1}{4\pi\epsilon_0} \sum_i \sum_{j>i} \frac{q_i q_j}{r_{ij}} \operatorname{erfc}(\alpha r_{ij}) - \frac{\alpha}{4\pi^{3/2}\epsilon_0} \sum_i q_i^2 \\ &+ \frac{1}{4\pi\epsilon_0} \sum_i \sum_{\substack{j \text{ lié à } i \\ j>i}} \frac{q_i q_j}{r_{ij}} \end{aligned} \quad (44)$$

Le premier terme correspond à une somme dans l'espace réciproque sur les vecteurs  $\mathbf{k}$ ;  $\alpha$  est un paramètre positif caractérisant la largeur de la distribution de charge gaussienne entourant chaque charge ponctuelle du système. Le second terme correspond à une somme dans l'espace direct;  $\operatorname{erfc}(x)$  est la fonction erreur complémentaire, *i.e.*  $1 - \operatorname{erf}(x)$ . Les troisième et quatrième contributions sont correctives, compte tenu du fait que la somme dans l'espace réciproque porte sur toutes les paires atomiques  $\{i, j\}$ , incluant, par conséquent, les termes *self*, 1–2 et 1–3.

Formellement, l'investissement informatique impliqué dans l'approche classique des sommes de réseau d'Ewald est  $\mathcal{O}(N^2)$ , ou  $N$  est le nombre de particules formant le système. Ce coût, comme l'ont montré d'une part Perram *et al.*, et Fincham, d'autre part, peut être réduit à  $\mathcal{O}(N^{3/2})$  grâce à un choix judicieux de la largeur des distributions gaussiennes,  $\alpha$ , du nombre de vecteurs  $\mathbf{k}$ , et de la troncature des interactions de paires dans l'espace direct. D'une manière générale, il est nécessaire que soient équilibrés les temps CPU investis dans l'évaluation des sommes de l'espace direct et de

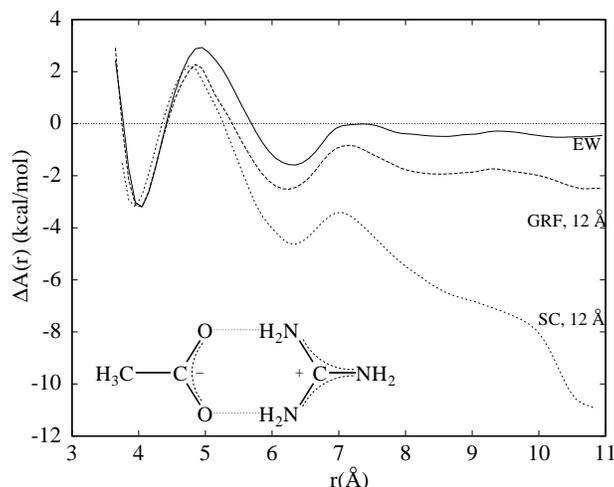


Figure 5: Profil d'énergie libre caractérisant l'approche d'un cation guanidinium d'un anion acétate dans la géométrie  $\mathcal{C}_{2v}$ , en phase aqueuse. [?] Le profil en trait plein correspond à une simulation en somme d'Ewald (EW) —  $\alpha = 0.3 \text{ \AA}^{-1}$ ; celui en tirets, à une simulation avec champ de réaction généralisé (GRF) —  $R_{\text{cut-off}} = 12 \text{ \AA}$ ; celui en pointillés, à une simulation avec troncature sphérique (SC) —  $R_{\text{cut-off}} = 12 \text{ \AA}$ .

l'espace réciproque pour atteindre une croissance en  $N^{3/2}$ .

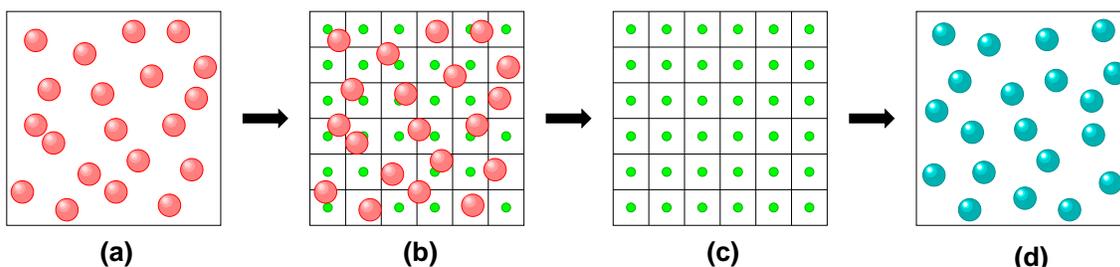


Figure 6: Approche de type *particle-mesh* sur un réseau bidimensionnel. **(a)** Système de particules chargées. **(b)** Les charges sont interpolées sur une grille bidimensionnelle. **(c)** En utilisant une approche de transformée de Fourier rapide (FFT), le potentiel et les forces sont évalués en chaque point de la grille. **(d)** Les forces sont interpolées en retour vers les particules, dont la position est ensuite mise à jour.

Des alternatives moins onéreuses à l'approche d'Ewald classique reposent sur une évaluation de la somme dans l'espace réciproque grâce à une transformée de Fourier rapide (FFT). Une grille tridimensionnelle remplissant l'espace dans lequel est réalisée la simulation de dynamique moléculaire est construite. Les charges portées par les particules constituant le système sont interpolées sur cette grille, et la distribution de charge correspondante,  $\rho(\mathbf{r})$ , est calculée. En recourant à la méthode FFT, la transformée de la distribution de charge,  $\hat{\rho}(\mathbf{k})$ , est déterminée sur la base des vecteurs  $\mathbf{k}$  de l'espace réciproque. La contribution à longue portée du potentiel électrostatique est alors évaluée selon  $\hat{V}_{\text{long}}(\mathbf{k}) = \hat{\mathcal{G}}(\mathbf{k})\hat{\rho}(\mathbf{k})$ , où  $\hat{\mathcal{G}}(\mathbf{k})$  est la fonction d'influence définie par  $\hat{\mathcal{G}}(\mathbf{k}) = \hat{\lambda}(\mathbf{k})/\epsilon_0 k^2$ , dans laquelle  $\lambda(\mathbf{r})$  est une distribution dépendant des seules caractéristiques géométriques de la boîte de

simulation. Par transformée inverse, la contribution  $V_{\text{long}}(\mathbf{r})$  est estimée aux différents points de la grille tridimensionnelle. Les forces d'origine électrostatique sont alors déterminées par dérivation numérique du potentiel. Enfin, le champ électrique et le potentiel sont interpolés de la grille vers la position des particules — voir Figure (6). Ce schéma constitue l'idée centrale des algorithmes de type *particle-mesh*, [?, ?] comme *particle-mesh Ewald* (PME) ou *particle-particle-particle-mesh* [?] ( $P^3M$ ), dont le coût est formellement en  $\mathcal{O}(N \ln N)$ .

## 2.6. Quelques propriétés accessibles à partir des simulations

Parmi les propriétés statiques qu'il est possible d'extraire des trajectoires générées au cours de simulations de dynamique moléculaire, les grandeurs structurales sont particulièrement intéressantes, dans la mesure où elles rendent compte de l'ordre local du système moléculaire. Les fonctions de distribution radiale,  $g(r)$ , indiquent la probabilité de trouver une paire d'atomes séparés par une distance  $r$ , par rapport à la probabilité attendue pour une distribution totalement aléatoire à la même densité. [?, ?, ?] La définition de  $g(r)$  passe par l'intégration de la fonction de distribution configurationnelle sur toutes les positions atomiques, à l'exception de celles de deux atomes marqués:

$$g(\mathbf{r}_1; \mathbf{r}_2) = \frac{N(N-1)}{\rho^2 \int e^{-\mathcal{V}(\mathbf{r}_1, \dots, \mathbf{r}_N)/k_B T} d\mathbf{r}_1 \dots d\mathbf{r}_N} \int e^{-\mathcal{V}(\mathbf{r}_1, \dots, \mathbf{r}_N)/k_B T} d\mathbf{r}_3 \dots d\mathbf{r}_N \quad (45)$$

où  $\rho$  désigne la densité du liquide. Pour un système dans lequel les atomes seraient tous identiques, la fonction de distribution radiale se simplifie comme une moyenne d'ensemble sur des paires d'atomes:

$$g(r) = \frac{V}{N^2} \left\langle \sum_i \sum_{j \neq i} \delta(\mathbf{r} - \mathbf{r}_{ij}) \right\rangle \quad (46)$$

Pour une paire d'atomes  $\{i, j\}$ , la fonction de distribution radiale peut être facilement évaluée par la relation:

$$g_{ij}(r) = \frac{\langle n_j(r + \delta r) \rangle}{4\pi \rho_j \int r^2 dr} \quad (47)$$

dans laquelle  $\langle n_j(r + \delta r) \rangle$  correspond au nombre moyen de sites  $j$ , dont la distance à  $i$  est comprise entre  $r$  et  $r + \delta r$ ;  $\rho_j$  est la densité moyenne de sites  $j$  dans l'échantillon considéré.

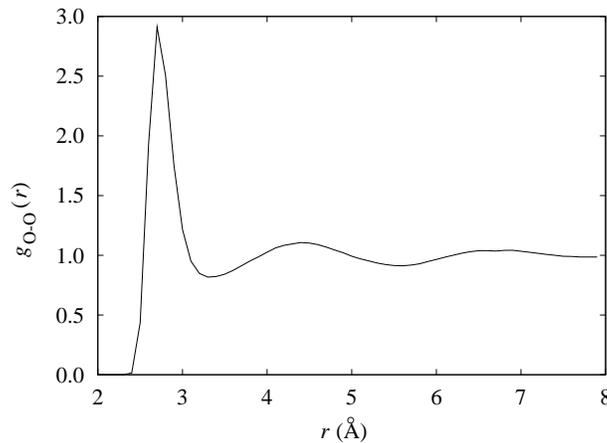


Figure 7: Fonction de distribution radiale oxygène–oxygène de l’eau TIP4P, obtenue à partir d’une simulation de dynamique moléculaire à 300 K.

Une seconde quantité tout aussi pertinente pour apprécier l’ordre du liquide moléculaire est le facteur de Kirkwood,  $G_K(R)$ , dépendant de la distance. [?] Cette grandeur caractérise la corrélation entre le moment dipolaire,  $\mu_i$  d’une molécule  $i$  de référence et celui porté par les molécules voisines  $j$  contenues dans une sphère de rayon  $R$  centrée en  $i$ :

$$G_K(R) = \frac{\left\langle \sum_{i,j;r_{ij}<R} \mu_i \cdot \mu_j \right\rangle}{N\mu^2} \quad (48)$$

Le facteur de Kirkwood est particulièrement sensible au traitement de l’électrostatique au cours de la simulation. Une troncature sphérique est à l’origine de nombreux artéfacts détériorant la corrélation dipôle–dipôle au bord de la sphère.

La dynamique moléculaire, comme son nom l’indique, permet également d’accéder à des propriétés dynamiques du système étudié. Parmi celles-ci, les fonctions de corrélation temporelle sont très informatives sur les temps de relaxation des différents degrés de liberté de ce système. [?, ?] D’une façon très générale, la corrélation entre deux quantités  $\mathcal{B}$  et  $\mathcal{B}'$  est donnée par le coefficient:

$$c_{\mathcal{B}\mathcal{B}'} = \frac{\langle \delta\mathcal{B} \delta\mathcal{B}' \rangle}{\sigma(\mathcal{B}) \sigma(\mathcal{B}')} \quad (49)$$

où  $\delta\mathcal{B} = \mathcal{B} - \langle \mathcal{B} \rangle$ , avec  $\langle \mathcal{B} \rangle$ , la moyenne d’ensemble statistique sur la grandeur  $\mathcal{B}$ ;  $\sigma(\mathcal{B}) = \sqrt{\langle \mathcal{B}^2 \rangle - \langle \mathcal{B} \rangle^2}$ .  $c_{\mathcal{B}\mathcal{B}'}$  varie entre 0 et 1, 0 correspondant à l’absence de corrélation. Cette formulation peut être généralisée au cas où  $\mathcal{B}$  et  $\mathcal{B}'$  sont évaluées à des instants distincts. Il s’ensuit que la fonction

de corrélation s'écrit:

$$c_{\mathcal{B}\mathcal{B}'}(t) = \frac{\langle \delta\mathcal{B}(t) \delta\mathcal{B}'(0) \rangle}{\sigma(\mathcal{B}) \sigma(\mathcal{B}')} \quad (50)$$

La moyenne temporelle au numérateur est effectuée sur toutes les origines des temps. Dans l'éventualité où  $\mathcal{B}' \equiv \mathcal{B}$ , la fonction calculée est une fonction dite d'auto-corrélation, définie par:

$$c_{\mathcal{B}\mathcal{B}}(t) = \frac{\langle \delta\mathcal{B}(t) \delta\mathcal{B}(0) \rangle}{\langle \delta\mathcal{B}(0) \delta\mathcal{B}(0) \rangle} \quad (51)$$

L'intégration de la fonction de corrélation, entre  $t = 0$  et  $t = \infty$ , conduit au temps de corrélation. A titre d'illustration, considérons le temps de corrélation réorientationnel du benzène dans l'eau. Il s'agit d'une simulation nécessairement plus longue que pour un liquide pur, étant donné que la moyenne statistique ne porte que sur une seule molécule.

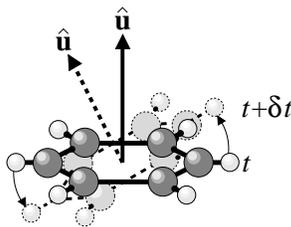


Figure 8: Réorientation d'une molécule de benzène au cours du temps.  $\hat{\mathbf{u}}$  est le vecteur unitaire porté par l'axe  $\mathcal{C}_6$  de la molécule.

La quantité évaluée ici est le coefficient d'auto-corrélation  $c_{\hat{\mathbf{u}}\hat{\mathbf{u}}}(t) = \langle \hat{\mathbf{u}}(t) \hat{\mathbf{u}}(0) \rangle$  où  $\hat{\mathbf{u}}$  est le vecteur unitaire porté par l'axe  $\mathcal{C}_6$  ou l'axe  $\mathcal{C}'_2$  du benzène. Le tenseur d'inertie de la molécule laisse, toutefois, supposer que le mouvement de rotation autour de  $\mathcal{C}'_2$  sera plus rapidement décorrélié que celui autour de  $\mathcal{C}_6$ .

A partir de la connaissance des fonctions de corrélation, non pas sur la grandeur  $\mathcal{B}$ , mais sur sa dérivée,  $\dot{\mathcal{B}}$ , les simulations numériques permettent d'accéder aux coefficients de transport. La dynamique moléculaire à l'équilibre offre, par exemple, la possibilité d'estimer le coefficient de diffusion,  $D$ , via l'intégrale: [?]

$$D = \frac{1}{3} \int_0^\infty \langle \dot{\mathbf{r}}_i(t) \cdot \dot{\mathbf{r}}_i(0) \rangle dt \quad (52)$$

dans laquelle  $\dot{\mathbf{r}}_i(t) \equiv \mathbf{v}_i(t)$  est la vitesse du centre de masse de la molécule  $i$ . Pour des temps suffisamment longs,  $D$  peut être obtenu en recourant à la relation d'Einstein:

$$D = \frac{1}{3} \frac{\langle |\mathbf{r}_i(t) - \mathbf{r}_i(0)|^2 \rangle}{2t} \quad (53)$$

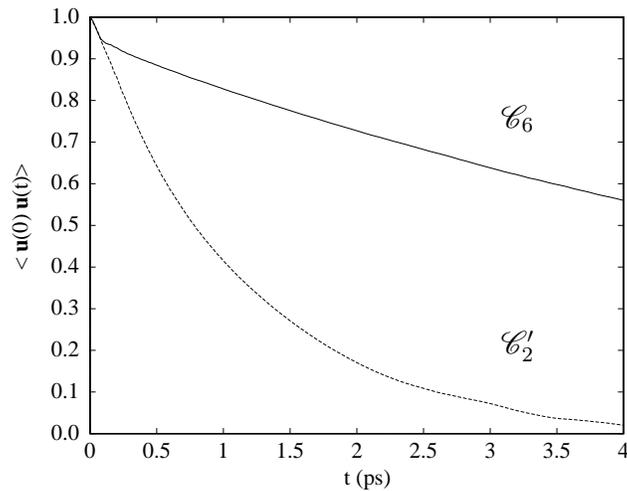


Figure 9: Fonctions d'auto-corrélation réorientationnelles du benzène dans l'eau liquide à 300 K. L'intégration des profils conduit aux temps de corrélation  $\tau(\mathcal{C}_6) = 2.4$  ps et  $\tau(\mathcal{C}'_2) = 0.5$  ps.

### 3. Conclusion

Les quelques exemples présentés dans ce texte donnent un aperçu du champ d'investigation accessible aux simulations numériques pour traiter des problèmes auxquels peut être confronté quotidiennement le modélisateur. Ce champ d'investigation va en s'élargissant au fur et à mesure que le rapport prix/performance des ordinateurs de calcul intensif diminue. En outre, l'accès aux architectures massivement parallèles, encore cependant inégal de par le monde, ouvre la voie à des applications particulièrement complexes pour la dynamique moléculaire. L'étude de systèmes macromoléculaires d'intérêt physico-chimique ou biologique, sur des échelles de temps significatives, en est l'illustration. Néanmoins, les progrès réalisés durant ces dernières années suggèrent deux réflexions. En premier lieu, la qualité du calcul entrepris est fortement conditionnée par le protocole de simulation choisi. En fonction du problème, certaines approches sont à proscrire et il devient

nécessaire de se tourner vers des méthodes plus rigoureuses, et souvent plus onéreuses. C'est le cas du traitement de l'électrostatique, limité dans sa forme la moins chère à une approximation additive de paires et à une troncature sphérique des forces à longue portée, mais pouvant, dans une forme plus sophistiquée, inclure une somme de réseaux, voire des modèles de polarisabilités distribuées. Si la méthodologie est à présent bien établie et robuste, tout ne peut pas encore être simulé par la dynamique moléculaire, attendu que l'intégration numérique des équations de Newton du mouvement est réalisée par pas infinitésimaux de 1 à 2 fs, qu'il faut comparer aux échelles de temps explorées, ainsi qu'à la taille du système considéré. Ce peut être là un obstacle rédhibitoire à une utilisation routinière de ces calculs. Par ailleurs, la nature du système à étudier requiert la plus grande prudence dans la paramétrisation de la fonction énergie potentielle. Si, fruits de nombreuses années de développement, les champs de force commerciaux ont atteint la maturité nécessaire pour aborder avec confiance des assemblages macromoléculaires variés, il n'en reste pas moins que de nombreuses applications demandent d'importants efforts supplémentaires de mise au point. C'est, nous l'avons vu, le cas des membranes lipidiques, pour lesquelles l'emploi de fonctionnelles multi-usage ne constitue pas une solution satisfaisante. Enfin, si avec l'apparition de l'outil *modélisation moléculaire*, généralisée à toutes les facettes de la chimie, il est tentant de recourir de plus en plus aux simulations numériques, dans la majorité des situations, les codes utilisés ne doivent pas être considérés comme des "boîtes noires". En tout état de cause, seul un examen attentif des résultats obtenus, une analyse détaillée des trajectoires générées, et un contrôle approfondi des propriétés thermodynamiques dérivées confirmeront la justesse et le sens physico-chimique des calculs entrepris. [?]

## Remerciements

Je tiens à remercier le Centre Charles Hermite pour l'accès aux machines parallèles SGI Origin 2000 et Power Challenge Array. Merci également à François Dehez pour avoir accepté de relire ce texte et me faire part de ses commentaires, ainsi qu'à Mounir Tarek pour ses nombreuses suggestions pour en améliorer le contenu.

## References