



HAL
open science

Modélisation atomistique du calcium en solution appliquée à l'étude du fractionnement isotopique

Hugo Moison

► **To cite this version:**

Hugo Moison. Modélisation atomistique du calcium en solution appliquée à l'étude du fractionnement isotopique. Sciences de la Terre. Université de Toulouse, 2024. Français. NNT : 2024TLSES021 . tel-04797485

HAL Id: tel-04797485

<https://theses.hal.science/tel-04797485v1>

Submitted on 22 Nov 2024

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.

Doctorat de l'Université de Toulouse

préparé à l'Université Toulouse III - Paul Sabatier

Modélisation atomistique du calcium en solution appliquée à
l'étude du fractionnement isotopique

Thèse présentée et soutenue, le 11 janvier 2024 par

Hugo MOISON

École doctorale

SDU2E - Sciences de l'Univers, de l'Environnement et de l'Espace

Spécialité

Surfaces et interfaces continentales, Hydrologie

Unité de recherche

GET - Geosciences Environnement Toulouse

Thèse dirigée par

Merlin MEHEUT et Magali BENOIT

Composition du jury

M. Marc BLANCHARD, Président, CNRS / GET

Mme Sara BONELLA, Rapporteur, École Polytechnique Fédérale de Lausanne

Mme Marie-Pierre GAIGEOT, Rapporteur, Université d'Evry Val d'Essonne

M. Rodolphe VUILLEUMIER, Rapporteur, Sorbonne Université

M. Fabien BRIEUC, Examineur, CEA

M. Merlin MEHEUT, Directeur de thèse, Université Toulouse III - Paul Sabatier

Mme Magali BENOIT, Co-directrice de thèse, CNRS / CEMES

Résumé

La composition en isotopes stables du calcium de systèmes géochimiques permet d'obtenir des informations concernant les processus survenant dans le cycle biochimique du calcium de la zone critique. Ces processus sont encore mal connus et les simulations atomistiques permettent d'étudier le fractionnement isotopique entre deux systèmes, notamment entre le calcium en solution aqueuse, Ca^{2+} , et des minéraux. Ce fractionnement dépend de la structure autour de Ca^{2+} et en particulier de sa première sphère de coordination. Cependant la valeur de la coordinence est toujours sujette à débat expérimentalement et théoriquement.

Dans cette thèse, une première partie a donc été consacrée à l'étude de la coordinence de Ca^{2+} à l'aide de dynamiques moléculaires ab-initio (AIMD) et de la théorie de la fonctionnelle de la densité (DFT), le but étant de déterminer les paramètres numériques d'influence. Plusieurs fonctionnelles ont été testées confirmant la nécessité d'ajouter des termes de dispersion pour reproduire la structure de l'eau. La diffusion, ainsi que la variation du nombre de molécules d'eau dans la sphère de coordination, ont été étudiées pour des simulations effectuées à plusieurs températures ce qui a permis de mettre en évidence des problèmes d'équilibration à température ambiante. Cela nous a poussé à mettre en place des protocoles de validation de l'équilibration en utilisant des chaînes de Markov. Les résultats montrent aussi une légère dépendance en température de la coordinence de Ca^{2+} . Les propriétés du fractionnement isotopique ont ensuite été calculées dans le cadre de l'approximation harmonique.

Pour les liquides, l'approximation harmonique peut engendrer des erreurs. Pour s'en affranchir, une méthode consiste à réaliser des dynamiques moléculaires d'intégrales de chemin (PIMD). Les simulations PIMD ab-initio sont coûteuses en temps de calcul et l'utilisation de potentiels empiriques ne permet pas de reproduire correctement les propriétés du fractionnement isotopique de Ca^{2+} . Une seconde partie s'est donc concentrée sur l'utilisation de méthodes "machine learning" pour construire des potentiels interatomiques ayant une précision bien supérieure aux potentiels empiriques avec un temps de calcul réduit par rapport à la DFT. Les méthodes utilisant des réseaux de neurones ont donné de meilleurs résultats.

Finalement, des PIMD ont été réalisées avec ces potentiels. Le fractionnement isotopique a été calculé par intégration thermodynamique et comparé aux résultats obtenus en utilisant l'approximation harmonique.

Abstract

The stable calcium isotope composition of geochemical systems provides information about processes occurring in the biochemical calcium cycle of the critical zone. These processes are still poorly understood, and atomistic simulations make it possible to study the isotopic fractionation between two systems, in particular between calcium in aqueous solution, Ca^{2+} , and minerals. This fractionation depends on the structure around Ca^{2+} and in particular on its first coordination sphere. However, the value of the coordination sphere is still subject to debate, both experimentally and theoretically.

The first part of this thesis was therefore devoted to the study of Ca^{2+} coordination using ab-initio molecular dynamics (AIMD) and density functional theory (DFT), with the aim of determining the influencing numerical parameters. Several functionals were tested, confirming the need to add dispersion terms to reproduce the structure of water. Diffusion, as well as the variation in the number of water molecules in the coordination sphere, were studied for simulations carried out at several temperatures, revealing equilibration problems at room temperature. This led us to set up equilibration validation protocols using Markov chains. The results also show a slight temperature dependence of Ca^{2+} coordination.

Isotopic fractionation properties were then calculated using the harmonic approximation.

For liquids, harmonic approximation can lead to errors. To overcome this, one method is to perform path integral molecular dynamics (PIMD). Ab initio PIMD simulations are costly in terms of computation time, and the use of empirical potentials does not correctly reproduce the properties of Ca^{2+} isotope fractionation.

A second part therefore focused on the use of "machine learning" methods to construct interatomic potentials with a much higher accuracy than empirical potentials, with a reduced computation time compared to DFT. Methods using neural networks gave better results. Finally, PIMD was performed with these potentials. Isotopic fractionation was calculated by thermodynamic integration and compared with those obtained using the harmonic approximation.

Table des matières

Introduction	11
I Méthodes	15
I.1 Potentiels interatomiques	16
I.1.1 Théorie de la fonctionnelle de la densité	16
I.1.1.1 Le problème à plusieurs corps	16
I.1.1.2 Approximation de Born-Oppenheimer	17
I.1.1.3 Traitement classique des noyaux et forces atomiques	19
I.1.1.4 Etat fondamental et densité électronique	20
I.1.1.5 Théorèmes de Hohenberg-Kohn	20
I.1.1.6 Equations de Kohn-Sham	21
I.1.1.7 Fonctionnelles d'échange et corrélation	23
I.1.1.8 Solution numérique des équations de Kohn-Sham - Base d'ondes planes	24
I.1.1.9 Pseudopotentiel:	25
I.1.2 Potentiels empiriques	26
I.1.2.1 Potentiel coulombien	27
I.1.2.2 Potentiel Lennard-Jones	27
I.1.2.3 Potentiel (q)-SPC/fw - modèle de l'eau	27
I.1.3 Potentiels Machine Learning (ML)	29
I.1.3.1 Base de données	29
I.1.3.2 Descripteurs	30
I.1.3.3 Potentiel et ajustement	33
I.1.3.4 Méthodes et codes utilisés	38
I.2 Dynamique moléculaire	39
I.2.1 Algorithme de Verlet	39
I.2.2 Conditions aux limites périodiques	40
I.2.3 Conditions initiales	42
I.2.4 Ensemble canonique - Thermostats	42
I.2.5 Différents types de dynamiques moléculaires	44
I.3 Dynamique moléculaire d'intégrales de chemin	47
I.3.1 Fonction de partition canonique quantique	47
I.3.2 Echantillonnage et isomorphisme	48
I.3.3 Modes normaux	50
I.3.4 Thermostat de Langevin	52
II Fractionnement isotopique à l'équilibre	55
II.1 Origine quantique du fractionnement	56
II.2 Facteur de partage isotopique et constante d'échange	57
II.3 Calcul du fractionnement isotopique	58
II.3.1 Dynamique moléculaire ab-initio et approximation harmonique	59

II.3.2	Dynamique moléculaire d'intégrales de chemin et estimateurs . . .	61
III	Coordinance du calcium en solution aqueuse	65
III.1	Précédentes études	66
III.2	Système et protocole de simulation	68
III.3	Résultats	69
III.3.1	Propriétés de l'eau	69
III.3.2	Environnement du calcium	73
III.3.3	Etude en température	76
III.3.3.1	Structure de l'eau	76
III.3.3.2	Environnement du calcium	78
III.3.4	Evolution temporelle de la coordinance	79
III.3.5	Echanges et temps de résidence	82
III.3.6	Estimation de l'incertitude/critère de convergence	84
III.4	Conclusion et perspectives	88
IV	Génération d'un potentiel machine learning pour le calcium en solution aqueuse	91
IV.1	Echantillonnage canonique assisté par Machine Learning (MLACS) . . .	92
IV.1.1	Protocole	92
IV.1.2	Résultats	94
IV.1.2.1	Descripteurs SO(4)	94
IV.1.2.2	Descripteurs SO(3)	96
IV.2	Potentiel Machine learning et réseau de neurones	103
IV.2.1	Choix des paramètres	103
IV.2.1.1	Description des paramètres	105
IV.2.1.2	Résultats	107
IV.2.2	Entraînements et premiers résultats	115
IV.3	Coordinance du calcium	118
IV.3.1	Protocole	118
IV.3.2	Résultats	118
V	Calcul du fractionnement isotopique du calcium en solution aqueuse	137
V.1	Approximation harmonique	137
V.1.1	Protocole	138
V.1.2	Résultats	139
V.2	Dynamique moléculaire d'intégrales de chemin	144
V.2.1	Protocole	144
V.2.2	Résultats	145
	Conclusion générale	157
	Annexes	161
A	Outils d'analyse des trajectoires de dynamique moléculaire	161
B	Compléments du chapitre 3	163
B.1	Coordinance en fonction du temps	163
B.2	Fonctions de distribution radiale des simulations avec un contre- ion et pour des concentrations différentes	166
C	Compléments du chapitre 4	168

C.1	Equivalence entre les paramètres du réseau de neurones testés et les commandes N2P2	168
C.2	Descripteurs utilisés pour le réseau de neurones	168
C.3	Résultats complémentaires des dynamiques moléculaires avec le potentiel machine learning	173
D	Compléments du chapitre 5	181

Introduction

Les différents éléments chimiques rencontrés dans la nature sont, pour la plupart, présents sous différentes formes isotopiques. Les isotopes sont divisés en deux catégories, les isotopes radioactifs possédant des noyaux atomiques instables et se désintégrant, et les isotopes stables, qui, quant à eux, ne présentent pas de radioactivité décelable. Le travail de cette thèse porte sur l'étude des isotopes stables du calcium.

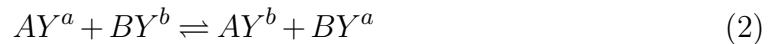
Il est possible de quantifier la proportion d'isotopes d'un élément sur Terre en mesurant son abondance relative exprimée en pourcentage. Par exemple, pour le calcium, les abondances relatives mesurées (entre parenthèses) donnent : $^{40}\text{Ca}(96,941)\%$, $^{42}\text{Ca}(0,647)\%$, $^{43}\text{Ca}(0,135)\%$, $^{44}\text{Ca}(2,086)\%$, $^{46}\text{Ca}(0,004)\%$ et $^{48}\text{Ca}(0,187)\%$.

Les isotopes ont une même configuration électronique et ont alors des propriétés chimiques semblables. Néanmoins, le rapport isotopique d'une phase, R , défini comme étant le rapport des densités atomiques n_{Y^a} et n_{Y^b} de deux isotopes Y^a et Y^b :

$$R = \frac{n_{Y^a}}{n_{Y^b}} \quad (1)$$

donne des valeurs différentes lorsqu'il est mesuré pour deux phases en contact. Par exemple entre le manteau terrestre et la croûte terrestre, entre deux minéraux d'une roche ou entre un minéral et une solution. Il existe alors un mécanisme d'enrichissement d'un isotope dans une phase par rapport à l'autre, dû à la différence de masse, que l'on appelle fractionnement isotopique.

On modélise le fractionnement isotopique à l'équilibre par une équation de réaction représentant un échange isotopique entre deux phases. Si l'on considère deux phases, A et B, composées de deux isotopes Y^a et Y^b , on a alors :



Lorsque les vitesses des deux sens de la réaction sont égales, l'équilibre thermodynamique est atteint et dépend des conditions thermodynamiques (pression et température). L'équilibre est caractérisé par une variation de l'enthalpie libre de réaction nulle. Par la suite, on s'intéressera uniquement au fractionnement isotopique à l'équilibre. A noter qu'il existe des mécanismes de fractionnement isotopique cinétique ou indépendants de la masse comme le fractionnement isotopique magnétique, l'effet de volume nucléaire ou bien provenant de la formation des éléments lors de la nucléosynthèse stellaire.

Les travaux de [118] sont le point de départ de la naissance de la théorie du fractionnement isotopique. La mesure des rapports isotopiques permet de comprendre les processus géologiques et géochimiques et de reconstituer les événements passés de

la Terre.

Les limites expérimentales ont, dans un premier temps, restreint l'étude du fractionnement isotopique aux isotopes stables dits traditionnels : l'hydrogène, le carbone, l'azote et l'oxygène. Le développement de nouveaux spectromètres de masse à couplage inductif à multicollecteur a permis l'étude d'autres éléments, comme le fer, le magnésium, le zinc ou encore le calcium.

Le calcium est le 5ème élément le plus abondant sur Terre et ses principaux réservoirs sont les roches carbonatées et les roches crustales (granit, basalt). Il est aussi présent sous forme dissoute dans les océans et les rivières ainsi que dans les plantes. Les applications du fractionnement du calcium sont multiples [51]. Il permet tout d'abord d'étudier la formation et la transformation de minéraux [55]. L'étude des os, dents ou des vestiges archéologiques peut fournir des informations sur les régimes alimentaires, les migrations humaines, les environnements paléoclimatiques [90, 91]. Par ailleurs, le calcium 40 est formé en permanence par la désintégration radioactive du potassium 40 ce qui permet d'envisager des applications géochronologiques [3]. Il permet aussi d'étudier le métabolisme osseux et a des applications en médecine [27, 25].

A la surface de la Terre, les variations des rapports isotopiques entre le calcium 40 et calcium 44 des différents réservoirs est de l'ordre de 4‰ (voir figure 1). Ainsi, l'étude du fractionnement isotopique du calcium permet de mieux comprendre son cycle bio-géochimique dans la zone critique (zone à l'interface entre l'atmosphère et la croûte terrestre continentale), qui est encore mal connu [37]. L'objectif de cette thèse est d'apporter des éléments de compréhension sur les processus opérant dans ce cycle.

Un premier processus important est l'altération des roches crustales et son lien avec le cycle du carbone. Lorsqu'une roche crustale silicatée est altérée par de l'eau, du CO_2 est capté à l'atmosphère et se retrouve dissout sous forme HCO_3^- . Après avoir été transporté jusqu'à la mer, ce HCO_3^- est capté par des organismes et se retrouve précipité sous forme de carbonate au fond des océans. Cette réaction nécessite des ions calcium. Le cycle du calcium est donc lié au cycle du carbone. Les processus de précipitation des carbonates de calcium peuvent être suivis en étudiant le fractionnement.

Finalement, de précédentes études [96] ont montré l'importance de l'étude du fractionnement isotopique du calcium à l'interface sol/plante (absorption par les racines, translocation, recyclage) afin d'étudier le rôle des plantes dans le cycle du calcium. Expérimentalement, les rapports isotopiques R varient peu, notamment pour les éléments de numéro atomique grand, on préfère alors mesurer la grandeur δ . Cette grandeur est définie comme étant l'écart relatif entre le rapport isotopique de la phase étudiée par rapport à un standard. Dans le cadre de l'étude des isotopes 40 et 44 du calcium, il a pour expression :

$$\delta^{44/40}\text{Ca}_{standard} = \left(\frac{\left(\frac{^{44}\text{Ca}}{^{40}\text{Ca}} \right)_{phase}}{\left(\frac{^{44}\text{Ca}}{^{40}\text{Ca}} \right)_{standard}} - 1 \right) \cdot 10^3 \quad (3)$$

Les résultats de précédentes études, avec le SRM915a comme matériel de référence

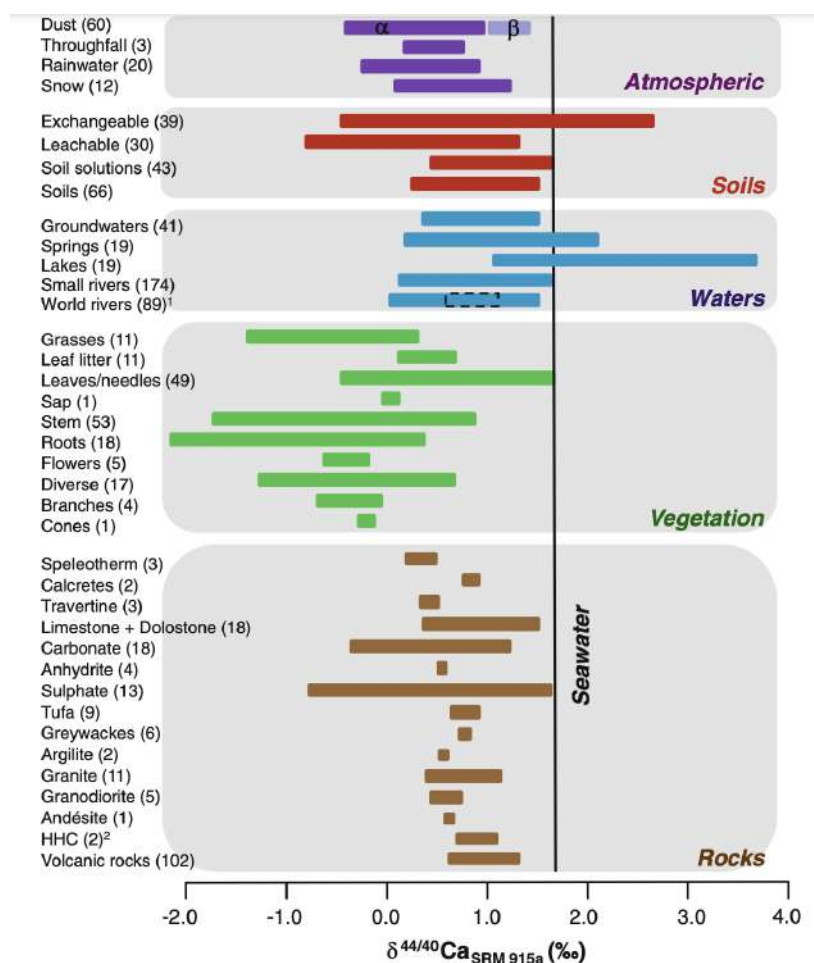


Figure 1: Valeurs expérimentales de $\delta^{44/40}\text{Ca}_{\text{SRM915a}}$ pour les différents reservoirs du calcium de la zone critique. La figure est extraite de [51]

standard, pour les différents réservoirs du calcium de la zone critique se trouvent figure 1.

L'utilisation de simulations numériques pour étudier le fractionnement isotopique du calcium apporte une approche complémentaire aux études expérimentales qui sont limitées par la difficulté à atteindre l'équilibre. Les simulations numériques permettent aussi d'explorer plus de systèmes, éventuellement dans des conditions difficilement reproductibles expérimentalement.

Lors des processus intervenant dans le cycle du calcium, le calcium est présent dans des minéraux mais aussi sous forme aqueuse Ca^{2+} . La forme aqueuse présente un intérêt particulier car c'est la forme mobile du calcium. De précédentes études numériques [28, 79], étudiant le fractionnement isotopique en présence d'une phase aqueuse, ont montré que le fractionnement isotopique entre des minéraux et une solution aqueuse dépend de la coordinence du calcium qui est encore mal connue à la fois expérimentalement et théoriquement. Ces études reposent sur la modélisation du calcium aqueux sous forme d'agrégat moléculaire, avec un atome calcium entouré de molécules d'eau. L'étude [28] obtient des valeurs de fractionnement isotopique en moyenne deux fois supérieures aux observations. D'après l'étude [79], les oxalates de

calcium devraient être enrichis en isotopes lourds du calcium, ce qui est en contradiction avec les observations réalisées dans les plantes [95]. Ces études soulignent la nécessité d'utiliser une autre approche pour étudier le fractionnement isotopique du calcium aqueux.

L'objectif de cette thèse est alors double. On souhaite, dans un premier temps, déterminer les paramètres numériques influençant la coordinence du calcium en solution aqueuse. Ces résultats nous permettront de déterminer le fractionnement isotopique du calcium aqueux à partir de schémas méthodologiques plus rigoureux. Une première méthode est basée sur une approche utilisée pour les solides employant l'approximation harmonique. Une seconde méthode, basée sur la dynamique moléculaire d'intégrales de chemin et l'intégration thermodynamique [35], permet de ne pas employer l'approximation harmonique qui n'est, a priori, pas justifiée pour les liquides.

Ce manuscrit est organisé de la façon suivante :

- Un premier chapitre présente l'ensemble des méthodes numériques utilisées dans ce travail : la dynamique moléculaire, la dynamique moléculaire d'intégrales de chemin et les différentes méthodes permettant de modéliser les interactions nécessaires à ces simulations.
- Un second chapitre présente plus en détail le fractionnement isotopique ainsi que son origine. Nous y présentons aussi la notion de facteur de partage isotopique et comment le calculer à partir des résultats des simulations numériques.
- Un troisième chapitre est consacré à l'étude de la coordinence du calcium en solution aqueuse à partir de simulations de dynamiques moléculaires ab-initio. Nous y étudions l'influence des différents paramètres de simulation sur la valeur de la coordinence et discutons des problèmes d'échantillonnage observés.
- Les simulations de dynamiques moléculaires d'intégrales de chemin étant coûteuses en temps de calcul, notamment lorsque sont utilisées des méthodes ab-initio, nous avons développé un potentiel de type "machine learning" qui est l'objet du chapitre 4.
- Finalement, un dernier chapitre présente les résultats du calcul du fractionnement isotopique du calcium en solution aqueuse obtenus à la fois dans l'approximation harmonique et par l'approche basée sur la dynamique moléculaire d'intégrale de chemin.

Méthodes

Dans ce chapitre est présenté l'ensemble des méthodes numériques utilisées au cours de cette thèse. Une partie (I.2) est consacrée à la dynamique moléculaire classique, une autre (I.3) à la dynamique moléculaire d'intégrales de chemin permettant de prendre en compte les effets quantiques des noyaux et de s'affranchir de l'approximation harmonique lors du calcul du fractionnement isotopique (voir chapitre II). La précision des résultats des simulations numériques dépend grandement des potentiels interatomiques utilisés pour décrire les interactions entre atomes, une première partie (I.1) leur est alors consacrée.

Table des matières

I.1 Potentiels interatomiques	16
I.1.1 Théorie de la fonctionnelle de la densité	16
I.1.2 Potentiels empiriques	26
I.1.3 Potentiels Machine Learning (ML)	29
I.2 Dynamique moléculaire	39
I.2.1 Algorithme de Verlet	39
I.2.2 Conditions aux limites périodiques	40
I.2.3 Conditions initiales	42
I.2.4 Ensemble canonique - Thermostats	42
I.2.5 Différents types de dynamiques moléculaires	44
I.3 Dynamique moléculaire d'intégrales de chemin	47
I.3.1 Fonction de partition canonique quantique	47
I.3.2 Echantillonnage et isomorphisme	48
I.3.3 Modes normaux	50
I.3.4 Thermostat de Langevin	52

I.1 Potentiels interatomiques

La description de la surface d'énergie potentielle (PES) d'un système atomique est un ingrédient nécessaire à toute simulation atomistique. Une fois connue, elle permet d'avoir accès à toutes les propriétés d'un système, ainsi que les forces agissant sur les atomes, nécessaires aux simulations de dynamique moléculaire.

Pour déterminer la PES, on se doit de résoudre l'équation de Schrödinger pour un système de coordonnées atomiques données. Dans le cas général, il n'existe pas de solution à cette équation. Plusieurs méthodes reposant sur des approximations sont alors utilisées, l'une d'elles étant la théorie de la fonctionnelle de la densité (partie I.1.1). Même avec ces méthodes, le temps de calcul numérique reste important, notamment lorsqu'elles sont utilisées dans les simulations de dynamique moléculaire où l'évaluation des forces doit être faite un grand nombre de fois.

Il est possible d'utiliser des formes analytiques, avec des paramètres ajustables, pour décrire la PES. Une première approche consiste à déterminer une fonction en se basant sur des considérations physiques et d'ajuster les paramètres sur des propriétés connues du système (partie I.1.2). Ces fonctions ont une forme simple et sont rapides à estimer au détriment d'une précision moindre. Une autre approche, non physique, consiste à utiliser une forme analytique complexe avec un grand nombre de paramètres à ajuster sur une base de données. Les potentiels de type machine learning (partie I.1.3) font partie de cette dernière catégorie.

I.1.1 Théorie de la fonctionnelle de la densité

La théorie de la fonctionnelle de la densité (DFT) [60] est une théorie quantique utilisée en physique, dans le cadre de l'étude de la matière, et en chimie quantique. Elle permet d'étudier les propriétés électroniques d'un système et est largement utilisée dans les simulations numériques.

Dans cette partie, on commence par exposer le problème à plusieurs corps et les approximations usuellement utilisées en théorie de la structure électronique permettant de simplifier la résolution de l'équation de Schrödinger. On présente ensuite les concepts centraux de la DFT (densité électronique, théorème de Hohenberg-Kohn, équation de Kohn-Sham et fonctionnelles). Finalement, les aspects pratiques de résolution numérique des équations, notamment l'utilisation de base d'ondes planes et de pseudo-potentiels, sont discutés.

I.1.1.1 Le problème à plusieurs corps

L'équation de Schrödinger que l'on souhaite résoudre est celle d'un système composé de N électrons et M noyaux. La fonction d'onde ψ qui décrit ce système dépend des N positions, $\{\vec{r}_i\}_{i=1,\dots,N}$, des électrons et M positions, $\{\vec{R}_I\}_{I=1,\dots,M}$, des noyaux (le spin n'est pas pris en compte) :

$$\psi = \psi(\vec{r}_1, \dots, \vec{r}_N, \vec{R}_1, \dots, \vec{R}_M) \quad (\text{I.1})$$

L'équation de Schrödinger est alors donnée par :

$$H\psi = E_{tot}\psi, \quad (\text{I.2})$$

où E_{tot} est l'énergie totale du système et H son hamiltonien.

L'hamiltonien se décompose en une somme de termes d'énergie cinétique et d'énergie potentielle. Les énergies cinétiques sont celles des électrons, T_e , et des noyaux, T_n . Les énergies potentielles intervenant dans notre description sont de type coulombien. On note V_{ee} l'énergie potentielle coulombienne entre électrons, V_{nn} celle entre noyaux et V_{ne} celle entre noyaux et électrons.

On obtient alors :

$$\begin{aligned}
 & \left[\begin{array}{c} \xrightarrow{T_n} \\ -\sum_{I=1}^M \frac{\hbar^2}{2M_I} \vec{\nabla}_I^2 \\ \xrightarrow{T_e} \\ -\sum_{i=1}^N \frac{\hbar^2}{2m_e} \vec{\nabla}_i^2 \\ \xrightarrow{V_{ee}} \\ +\frac{1}{2} \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N \frac{e^2}{4\pi\epsilon_0} \frac{1}{|\vec{r}_i - \vec{r}_j|} \\ \xrightarrow{V_{nn}} \\ +\frac{1}{2} \sum_{I=1}^M \sum_{\substack{J=1 \\ J \neq I}}^M \frac{e^2}{4\pi\epsilon_0} \frac{Z_I Z_J}{|\vec{R}_I - \vec{R}_J|} \\ \xrightarrow{V_{ne}} \\ -\sum_{i=1}^N \sum_{I=1}^M \frac{e^2}{4\pi\epsilon_0} \frac{Z_I}{|\vec{r}_i - \vec{R}_I|} \end{array} \right] \psi = E_{tot} \psi \quad (I.3)
 \end{aligned}$$

où m_e est la masse de l'électron, les M_I celles des noyaux et les Z_I sont les numéros atomiques.

On s'attend à avoir des énergies de l'ordre de l'énergie coulombienne entre un proton et un électron. On divise alors chaque terme par l'énergie d'Hartree et on travaille dans le système d'unités atomiques ($\hbar = m_e = \frac{e^2}{4\pi\epsilon_0} = 1$). Les distances sont alors exprimées en Bohr, les masses en multiple de la masse de l'électron et les énergies en Hartree. L'équation de Schrödinger devient :

$$\begin{aligned}
 & \left[-\sum_{i=1}^N \frac{1}{2} \vec{\nabla}_i^2 - \sum_{I=1}^M \frac{1}{2M_I} \vec{\nabla}_I^2 + \frac{1}{2} \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N \frac{1}{|\vec{r}_i - \vec{r}_j|} \right. \\
 & \left. + \frac{1}{2} \sum_{I=1}^M \sum_{\substack{J=1 \\ J \neq I}}^M \frac{Z_I Z_J}{|\vec{R}_I - \vec{R}_J|} - \sum_{i=1}^N \sum_{I=1}^M \frac{Z_I}{|\vec{r}_i - \vec{R}_I|} \right] \psi = E_{tot} \psi \quad (I.4)
 \end{aligned}$$

Dans la quasi-totalité des cas, cette équation n'est pas solvable analytiquement et impossible numériquement. La difficulté vient des interactions coulombiennes et il est nécessaire d'avoir recours à des approximations. La première, qui est à la base de l'étude des structures électroniques, est l'approximation de Born-Oppenheimer [17].

I.1.1.2 Approximation de Born-Oppenheimer

Commençons par réécrire l'équation I.4 précédente :

$$[T_n + H_0] \psi = E_{tot} \psi \quad \text{avec} \quad H_0 = T_e + V_{ee} + V_{nn} + V_{ne} \quad (I.5)$$

L'approximation de Born-Oppenheimer se base tout d'abord sur une première constatation. La masse d'un noyau étant très supérieure à celle d'un électron ($M_I \gg m_e$), leur mouvement est alors beaucoup plus lent que celui des électrons.

La première étape de l'approximation consiste à considérer que le mouvement des

noyaux est tellement lent que les électrons les voient fixes. Le terme d'énergie cinétique des noyaux est alors nul ($T_n = 0$) et on s'intéresse à la résolution de l'équation pour les électrons décrits par la fonction d'onde ϕ :

$$H_0\phi(\{\vec{r}_i\};\{\vec{R}_I\}) = E_{tot}\phi(\{\vec{r}_i\};\{\vec{R}_I\}) \quad (\text{I.6})$$

où la notation ";" indique que les positions atomiques $\{\vec{R}_I\}$ sont maintenant des paramètres.

L'application de conditions aux limites induit une quantification :

$$H_0\phi_m(\{\vec{r}_i\};\{\vec{R}_I\}) = E_m\phi_m(\{\vec{r}_i\};\{\vec{R}_I\}) \quad (\text{I.7})$$

les ϕ_m sont les fonctions d'ondes électroniques, orthogonales et normalisées formant une base.

$$\int \prod_i d\vec{r}_i \phi_m^* \phi_{m'} = \delta_{mm'} \quad (\text{I.8})$$

La deuxième étape de l'approximation est alors de supposer que l'on a trouvé les solutions de l'équation I.7 et de développer la fonction d'onde du système :

$$\psi(\{\vec{r}_i\},\{\vec{R}_I\}) = \sum_m \chi_m(\{\vec{R}_I\})\phi_m(\{\vec{r}_i\};\{\vec{R}_I\}) \quad (\text{I.9})$$

Pour déterminer complètement la fonction d'onde ψ , il nous reste à trouver l'expression des χ_m . Pour cela, on introduit l'expression de ψ dans l'équation I.5:

$$\begin{aligned} [T_n + H_0] \sum_m \chi_m \phi_m &= E_{tot} \sum_m \chi_m \phi_m \\ \Leftrightarrow \sum_m T_n (\chi_m \phi_m) + \chi_m H_0 \phi_m &= E_{tot} \sum_m \chi_m \phi_m \\ \Leftrightarrow \sum_m T_n (\chi_m \phi_m) + E_m \chi_m \phi_m &= E_{tot} \sum_m \chi_m \phi_m \end{aligned} \quad (\text{I.10})$$

On développe ensuite le premier terme :

$$\begin{aligned} \Leftrightarrow \sum_m (T_n \chi_m) \phi_m + \chi_m (T_n \phi_m) + \sum_{I=1}^M \frac{1}{M_I} (\vec{\nabla}_I \chi_m) (\vec{\nabla}_I \phi_m) + E_m \chi_m \phi_m &= E_{tot} \sum_m \chi_m \phi_m \\ \Leftrightarrow \sum_m (T_n \chi_m) \phi_m + E_m \chi_m \phi_m = E_{tot} \sum_m \chi_m \phi_m - \sum_{I=1}^M \frac{1}{M_I} (\vec{\nabla}_I \chi_m) (\vec{\nabla}_I \phi_m) - \chi_m (T_n \phi_m) \end{aligned} \quad (\text{I.11})$$

En multipliant par $\phi_{m'}^*$, intégrant sur les \vec{r}_i et en utilisant I.8 on obtient :

$$[T_n + E_m] \chi_m = E_{tot} \chi_m - \sum_m \Lambda_{m'm} \chi_m \quad (\text{I.12})$$

avec :

$$\Lambda_{m'm} = \int \prod_i d\vec{r}_i \sum_m \phi_{m'}^* \left[T_n \phi_m + \sum_{I=1}^M \frac{1}{M_I} (\vec{\nabla}_I \phi_m) (\vec{\nabla}_I) \right] \quad (\text{I.13})$$

Les $\Lambda_{m'm}$ sont des opérateurs de couplage.

Pour finir, on effectue une dernière approximation dite adiabatique qui consiste à négliger les $\Lambda_{m'm}$ et à découpler le mouvement des électrons, décrit par les ϕ_m , et des noyaux, décrit par les χ_m . En effet, on a alors le jeu d'équations suivant :

$$\psi_m(\{\vec{r}_i\}, \{\vec{R}_I\}) = \chi_m(\{\vec{R}_I\})\phi_m(\{\vec{r}_i\}; \{\vec{R}_I\}) \quad (\text{I.14})$$

$$H_{elec}\phi_m(\{\vec{r}_i\}; \{\vec{R}_I\}) = E_{elec,m}\phi_m(\{\vec{r}_i\}; \{\vec{R}_I\}) \quad (\text{I.15})$$

$$[T_n + E_m]\chi_m = E_{tot}\chi_m \quad (\text{I.16})$$

L'équation I.15 est l'équation fondamentale de la théorie des structures électroniques. V_{nn} étant constant pour les électrons, on a posé $H_{elec} = T_e + V_{ee} + V_{ne}$ et $E_{elec,m} = E_m - V_{nn}$.

Finalement, dans l'approximation de Bohr-Oppenheimer, les électrons, ayant un mouvement plus rapide que les noyaux, s'adaptent instantanément au mouvement lent des noyaux. Les noyaux quant à eux voient le mouvement moyen des électrons. L'approximation de Bohr-Oppenheimer est d'autant plus vérifiée que l'on se trouve proche de l'état fondamental. Dans la suite, c'est cet état fondamental que l'on cherchera à déterminer et on omettra l'indice m .

I.1.1.3 Traitement classique des noyaux et forces atomiques

Le mouvement des noyaux étant décrit par l'équation I.16, on peut définir l'hamiltonien nucléaire à l'équilibre H_n :

$$\begin{aligned} H_n = T_n + E_{elec} + V_{nn} &= -\sum_{I=1}^M \frac{1}{2M_I} \vec{\nabla}_I^2 + E_{elec} + \frac{1}{2} \sum_{I=1}^M \sum_{\substack{J=1 \\ J \neq I}}^M \frac{e^2}{4\pi\epsilon_0} \frac{Z_I Z_J}{|\vec{R}_I - \vec{R}_J|} \\ &= -\sum_{I=1}^M \frac{1}{2M_I} \vec{\nabla}_I^2 + U \end{aligned} \quad (\text{I.17})$$

Faisons maintenant une seconde approximation, le mouvement des noyaux peut être traité classiquement. Concrètement, cela consiste à considérer que la fonction d'onde des noyaux a une extension spatiale tellement faible qu'ils peuvent être traités comme des particules. Plus les noyaux sont lourds, plus l'approximation est vérifiée.

Dans ce cas, l'hamiltonien classique des noyaux devient, en utilisant $p_I^2 \leftarrow -\vec{\nabla}_I^2$:

$$\mathcal{H} = \sum_{I=1}^M \frac{p_I^2}{2M_I} + U \quad (\text{I.18})$$

Les forces agissant sur les noyaux I sont alors :

$$\vec{F}_I = -\vec{\nabla}_I U \quad (\text{I.19})$$

Ces forces, nécessaires aux simulations de dynamique moléculaire, peuvent donc être obtenues à partir de la structure électronique en déterminant E_{elec} . L'enjeu est alors de résoudre l'équation de Schrödinger électronique I.15 et dans la suite on omettra l'indice $elec$.

I.1.1.4 Etat fondamental et densité électronique

A partir de maintenant, on souhaite résoudre l'équation de Schrödinger électronique et obtenir l'énergie de l'état fondamental E . Le principe variationnel nous permet d'écrire :

$$E = \min_{\phi} \langle \phi | H | \phi \rangle = \min_{\phi} E[\phi] \quad (\text{I.20})$$

L'énergie est alors une fonctionnelle de la fonction d'onde et H est :

$$H = T_e + V_{ee} + V_{ne} \quad \text{avec} \quad V_{ne} = - \sum_{i=1}^N \sum_{I=1}^M \frac{Z_I}{|\vec{r}_i - \vec{R}_I|} = \sum_{i=1}^N v_{ne}(\vec{r}_i) \quad (\text{I.21})$$

Les noyaux étant fixes, v_{ne} joue le rôle d'un potentiel extérieur. La détermination de la fonction d'onde nous permet de calculer la densité électronique :

$$n(\vec{r}) = N \int d\vec{r}_2 \dots d\vec{r}_n |\phi(\vec{r}, \vec{r}_2, \dots, \vec{r}_N)|^2 \quad (\text{I.22})$$

Il existe donc une relation entre le potentiel extérieur et la densité électronique :

$$v_{ne} \longrightarrow n(\vec{r}) \quad (\text{I.23})$$

La théorie de la fonctionnelle de la densité est une théorie de l'état fondamental et repose sur le fait que la relation inverse existe.

I.1.1.5 Théorèmes de Hohenberg-Kohn

Hohenberg et Kohn montrent en 1964 [60] que pour un système électronique soumis à un potentiel extérieur, dans notre cas v_{ne} , la densité électronique de l'état fondamental détermine, à une constante près, ce potentiel. En d'autres termes, il n'existe pas deux potentiels extérieurs, qui diffèrent de plus d'une constante, donnant la même densité électronique de l'état fondamental. Il existe donc une relation :

$$n(\vec{r}) \longrightarrow v_{ne} \quad (\text{I.24})$$

Preuve pour état non dégénéré :

La preuve de ce théorème repose sur une démonstration par l'absurde. Considérons deux potentiels extérieurs, $v_{ne}^{(1)}$ et $v_{ne}^{(2)}$, donnant la même densité électronique $n(\vec{r})$. Les fonctions d'onde de l'état fondamental pour ces deux potentiels sont $\phi^{(1)}$ et $\phi^{(2)}$ et leur hamiltonien $H^{(1)}$ et $H^{(2)}$. On a alors :

$$E^{(1)} = \langle \phi^{(1)} | H^{(1)} | \phi^{(1)} \rangle = \langle \phi^{(1)} | T_e + V_{ee} + \sum_{i=1}^N v_{ne}^{(1)} | \phi^{(1)} \rangle \quad (\text{I.25})$$

$$= \langle \phi^{(1)} | T_e + V_{ee} | \phi^{(1)} \rangle + \int d\vec{r} v_{ne}^{(1)} n(\vec{r}) \quad (\text{I.26})$$

$$E^{(2)} = \langle \phi^{(2)} | H^{(2)} | \phi^{(2)} \rangle = \langle \phi^{(2)} | T_e + V_{ee} + \sum_{i=1}^N v_{ne}^{(2)} | \phi^{(2)} \rangle \quad (\text{I.27})$$

$$= \langle \phi^{(2)} | T_e + V_{ee} | \phi^{(2)} \rangle + \int d\vec{r} v_{ne}^{(2)} n(\vec{r}) \quad (\text{I.28})$$

où les égalités I.26 et I.28 sont obtenues en utilisant la symétrie de la fonction d'onde électronique et la définition de la densité électronique.

De plus, d'après le principe variationnel :

$$E^{(1)} < \langle \phi^{(2)} | H^{(1)} | \phi^{(2)} \rangle = \langle \phi^{(2)} | T_e + V_{ee} | \phi^{(2)} \rangle + \int d\vec{r} v_{ne}^{(1)} n(\vec{r}) \quad (\text{I.29})$$

$$= E^{(2)} + \int d\vec{r} (v_{ne}^{(1)} - v_{ne}^{(2)}) n(\vec{r}) \quad (\text{I.30})$$

et :

$$E^{(2)} < \langle \phi^{(1)} | H^{(2)} | \phi^{(1)} \rangle = \langle \phi^{(1)} | T_e + V_{ee} | \phi^{(1)} \rangle + \int d\vec{r} v_{ne}^{(2)} n(\vec{r}) \quad (\text{I.31})$$

$$= E^{(1)} + \int d\vec{r} (v_{ne}^{(2)} - v_{ne}^{(1)}) n(\vec{r}) \quad (\text{I.32})$$

En sommant les deux équations on obtient la contradiction suivante :

$$E^{(1)} + E^{(2)} < E^{(1)} + E^{(2)} \quad (\text{I.33})$$

Il ne peut donc pas exister deux potentiels extérieurs, qui diffèrent de plus d'une constante, donnant la même densité électronique.

Finalement, la densité électronique de l'état fondamental détermine le potentiel extérieur, et donc l'hamiltonien et tout ce qu'on a besoin de connaître. L'énergie de l'état fondamental est maintenant une fonctionnelle de la densité électronique $E = E[n(\vec{r})]$ et il en va de même pour la fonction d'onde $\phi = \phi[n(\vec{r})]$.

Considérons un système avec une densité électronique de l'état fondamental $n(\vec{r})^{(1)}$, associé à $\phi^{(1)}$ et $H^{(1)}$, l'énergie est donc :

$$E^{(1)}[n(\vec{r})^{(1)}] = \langle \phi^{(1)} | H^{(1)} | \phi^{(1)} \rangle \quad (\text{I.34})$$

Si l'on considère maintenant une densité électronique différente, $n(\vec{r})^{(2)}$, associée à $\phi^{(2)}$ nécessairement différente de $\phi^{(1)}$, par le principe variationnel on a :

$$E^{(1)}[n(\vec{r})^{(1)}] = \langle \phi^{(1)} | H^{(1)} | \phi^{(1)} \rangle < \langle \phi^{(2)} | H^{(1)} | \phi^{(2)} \rangle = E^{(2)}[n(\vec{r})^{(2)}] \quad (\text{I.35})$$

Ainsi, minimiser l'énergie par rapport à la densité électronique donne l'énergie de l'état fondamental. Une fonctionnelle universelle de l'énergie $E[n(\vec{r})]$ peut être définie en fonction de la densité électronique. L'énergie de l'état fondamental est obtenue en minimisant cette fonctionnelle et la densité électronique associée est celle de l'état fondamental. Cela constitue le deuxième théorème de Hohenberg-Kohn.

I.1.1.6 Equations de Kohn-Sham

Déterminons l'énergie électronique de l'état fondamental, fonctionnelle de la densité électronique

$$E[n] = T[n] + V_{ee}[n] + \int d\vec{r} v_{ne} n(\vec{r}) \quad (\text{I.36})$$

en appliquant le principe variationnel.

Les seules densités électroniques valides sont celles qui conservent le nombre d'électrons

: $\int d\vec{r}n(\vec{r}) = N$. On définit alors la fonctionnelle $A[n] = E[n] - \mu \int d\vec{r}n(\vec{r})$, avec μ le multiplicateur de Lagrange associé à la conservation du nombre d'électrons (c'est le potentiel chimique). Le principe variationnel donne alors :

$$\frac{\delta A[n]}{\delta n} = 0 \Leftrightarrow \frac{\delta E[n]}{\delta n} = \mu \quad (\text{I.37})$$

L'expression de $T[n]$ est compliquée, en revanche, on la connaît pour un système d'électrons qui ne sont pas en interaction. L'idée de Kohn-Sham [67] est alors la suivante. On considère un système d'électrons indépendants qui a la même densité électronique de l'état fondamental que le vrai système. On note $T_s[n]$ et v_s , l'énergie cinétique et le potentiel extérieur de ce système.

On réécrit alors :

- $T[n] = T_s[n] + T_c[n]$
- $V_{ee}[n] = V_H[n] + \Delta V[n]$ avec le potentiel de Hartree $V_H = \frac{1}{2} \int \int d\vec{r}d\vec{r}' \frac{n(\vec{r})n(\vec{r}')}{|\vec{r}-\vec{r}'|}$
- $E_{xc}[n] = T_c[n] + \Delta V[n]$

$E_{xc}[n]$ est l'énergie d'échange-correlation, dans laquelle se trouvent les termes d'échange et de corrélation électronique ainsi que la correction à l'énergie cinétique $T_s[n]$.

L'application du principe variationnel donne alors :

$$\frac{\delta E[n]}{\delta n} = \frac{\delta T_s[n]}{\delta n} + v_{ne} + \int d\vec{r}' \frac{n(\vec{r}')}{|\vec{r}-\vec{r}'|} + \frac{\delta E_{xc}[n]}{\delta n} = \mu \quad (\text{I.38})$$

or, si l'on revient à notre système d'électrons indépendants, d'énergie cinétique $T_s[n]$, soumis à un potentiel extérieur v_s , on a :

$$\frac{\delta E[n]}{\delta n} = \frac{\delta T_s[n]}{\delta n} + v_s = \mu \quad (\text{I.39})$$

et on sait que pour un système d'électrons indépendants, en définissant ϕ_i la fonction d'onde d'un électron, l'équation de Schrödinger est :

$$\left[\frac{1}{2} \vec{\nabla}^2 + v_s \right] \phi_i = \epsilon_i \phi_i \quad i = 1, \dots, N \quad (\text{I.40})$$

et

$$n(\vec{r}) = \sum_i^N |\phi_i|^2 \quad (\text{I.41})$$

En identifiant à l'aide des équations I.38 et I.39 l'expression de v_s on obtient les équations de Kohn-Sham :

$$\left[\frac{1}{2} \vec{\nabla}^2 + v_{ne} + V_H + V_{xc} \right] \phi_i = \epsilon_i \phi_i \quad \text{avec} \quad V_H = \frac{\delta E_H}{\delta n} \quad \text{et} \quad V_{xc} = \frac{\delta E_{xc}}{\delta n} \quad (\text{I.42})$$

Supposons que l'on connaisse l'expression de E_{xc} , on se retrouve alors avec un problème aux valeurs propres qu'il est possible de résoudre. Mais, pour obtenir les fonctions propres ϕ_i , il faut d'abord connaître V_H et V_{xc} , qui dépendent de la densité électronique $n(\vec{r})$ et donc des ϕ_i . On se retrouve donc avec un problème auto-cohérent.

En pratique, on démarre des positions nucléaires et on calcule le potentiel v_{ne} . On

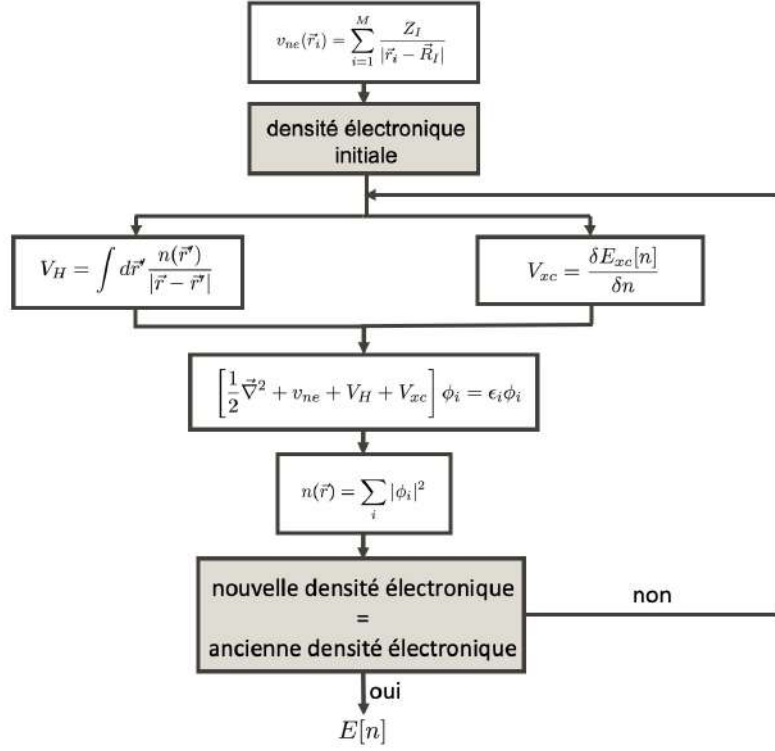


Figure I.1: Représentation schématique du calcul auto-cohérent

choisit ensuite une densité électronique de départ pour déterminer V_H et V_{xc} . On résout numériquement les équations de Kohn-Sham, on obtient alors de nouvelles fonctions d'onde ϕ_i , et une nouvelle densité électronique $n(\vec{r})$. On répète la procédure jusqu'à obtenir la convergence de la densité électronique et enfin obtenir l'énergie de l'état fondamental à l'aide de l'équation :

$$E[n] = T_s[n] + V_{ne} + E_H + E_{xc} \quad (\text{I.43})$$

La procédure est résumée figure I.1. Il est évident qu'un élément essentiel est l'expression de l'énergie d'échange-corrélation E_{xc} qui est inconnue. Le problème est alors d'en trouver une expression approchée.

I.1.1.7 Fonctionnelles d'échange et corrélation

Différentes fonctionnelles E_{xc} ont été développées. La plus simple, proposée par Kohn et Sham [67], est la "Local-Density-Approximation (LDA)" :

$$E_{xc}^{LDA}[n] = \int d\vec{r} n(\vec{r}) e_{xc}[n(\vec{r})] \quad (\text{I.44})$$

où $e_{xc}[n(\vec{r})]$ est l'énergie d'échange-corrélation par électron d'un gaz homogène d'électrons de densité constante qu'il est possible de calculer formellement.

Dans un système réel, la distribution électronique n'est pas homogène. Pour prendre en compte les inhomogénéités, ont été développées des fonctionnelles de type "Generalized-gradient approximation (GGA)" :

$$E_{xc}^{GGA}[n] = \int d\vec{r} n(\vec{r}) e_{xc}[n(\vec{r}), \vec{\nabla} n(\vec{r})] \quad (\text{I.45})$$

où maintenant e_{xc} dépend du gradient de la densité électronique. Plusieurs fonctionnelles GGA ont été proposées. Dans cette thèse nous avons utilisé les fonctionnelles PBE [86], BLYP [11, 69], PBEsol [87].

Pour les systèmes où les interactions de Van der Waals sont présentes, l'utilisation de fonctionnelles locales n'est pas en mesure de les modéliser correctement. En effet, les fonctionnelles GGA permettent de décrire localement les variations de densité alors que les interactions de Van der Waals impliquent des corrélations électroniques à longue distance. Pour prendre en compte explicitement les termes de dispersion de type Van der Waals, il est possible d'utiliser deux méthodes :

- Correction de dispersion semi-empirique: on ajoute à la fonctionnelle un terme semi-empirique de la forme :

$$E_{disp} = -s \sum_{I < J} f(R_{IJ}) \frac{C_{IJ}}{R_{IJ}^6} \quad (\text{I.46})$$

où s est un paramètre, C_{IJ} des coefficients de dispersion et f une fonction d'amortissement.

Cette approche est nommée "DFT-D" par Grimme [49].

- Fonctionnelle de Van der Waals non locale: on ajoute à la fonctionnelle, une autre fonctionnelle, non-locale, de la forme :

$$E_{nl}[n] = \frac{1}{2} \int \int d\vec{r} d\vec{r}' n(\vec{r}) n(\vec{r}') k(\vec{r}, \vec{r}') \quad (\text{I.47})$$

où k est une fonction de corrélation dépendant de $|\vec{r} - \vec{r}'| f(\vec{r})$ et $|\vec{r} - \vec{r}'| f(\vec{r}')$
où f est une fonction de la densité électronique et de son gradient [71].

I.1.1.8 Solution numérique des équations de Kohn-Sham - Base d'ondes planes

Nous avons maintenant tous les ingrédients nécessaires pour déterminer l'énergie. On souhaite alors trouver une stratégie pour calculer les ϕ_i à partir de l'équation de Kohn-Sham. La première chose à faire est de déterminer un domaine spatial pour notre système. On définit alors une cellule dans laquelle se trouve le système, définie par trois vecteurs du réseau $\vec{a}_1, \vec{a}_2, \vec{a}_3$, le volume de la cellule est alors $\Omega = \vec{a}_1 \cdot (\vec{a}_2 \times \vec{a}_3)$. A ce stade, on peut définir les vecteurs du réseau réciproque, $\vec{b}_1, \vec{b}_2, \vec{b}_3$ tels que :

$$\vec{a}_i \cdot \vec{b}_j = 2\pi \delta_{ij} \quad i, j = 1, 2, 3 \quad (\text{I.48})$$

En physique de la matière, on utilise la régularité des structures pour utiliser les conditions aux limites périodiques :

$$\phi_i(\vec{r} + \vec{a}_1) = \phi_i(\vec{r}) \quad \text{et} \quad \vec{\nabla} \phi_i(\vec{r} + \vec{a}_1) = \vec{\nabla} \phi_i(\vec{r}) \quad (\text{I.49})$$

Le fait que le potentiel effectif soit maintenant périodique permet l'apparition du nombre quantique \vec{k} , le vecteur d'onde du réseau réciproque. On décompose les fonctions d'onde ϕ_i sur une base d'ondes planes. D'après le théorème de Bloch :

$$\phi_{i,\vec{k}}(\vec{r}) = \sum_{\vec{G}} c_{i,\vec{k}}(\vec{G}) e^{i(\vec{k} + \vec{G}) \cdot \vec{r}} \quad (\text{I.50})$$

où $\vec{G} = m_1\vec{b}_1 + m_2\vec{b}_2 + m_3\vec{b}_3$, avec m_1, m_2, m_3 des entiers. La somme se fait alors sur un ensemble infini de vecteurs \vec{G} définis tels que $\vec{G} \cdot \vec{a}_i = 2\pi m_i$.

Pour des raisons pratiques, on définit une énergie de cutoff E_{cut} telle que :

$$\frac{|\vec{G} + \vec{k}|^2}{2} \leq E_{cut} \quad (\text{I.51})$$

On a alors :

$$\phi_{i,\vec{k}}(\vec{r}) = \sum_{\vec{G}, \frac{|\vec{G} + \vec{k}|^2}{2} \leq E_{cut}} c_{i,\vec{k}}(\vec{G}) e^{i(\vec{k} + \vec{G}) \cdot \vec{r}} \quad (\text{I.52})$$

Plus l'énergie de cutoff est élevée, plus il est possible de décrire de rapides variations de la fonction d'onde. L'équation de Kohn-Sham devient alors :

$$\left[\frac{1}{2} \vec{\nabla}^2 + v_{ne} + V_H + V_{xc} \right] \phi_{i,\vec{k}} = \epsilon_i \phi_{i,\vec{k}} \quad (\text{I.53})$$

et la densité électronique :

$$n(\vec{r}) = \sum_i^N \sum_{\vec{k}} w_{\vec{k}} |\phi_{i,\vec{k}}|^2 \quad (\text{I.54})$$

où les $w_{\vec{k}}$ sont des poids associés au vecteur d'onde \vec{k} . Comme il apparaît dans l'expression la fonction d'onde au carré, on utilise une énergie de cutoff pour la densité électronique typiquement prise égale à quatre fois celle des fonctions d'ondes (sauf dans le cas des pseudopotentiels ultrasoft où on la choisit plus grande, voir partie suivante).

L'avantage d'utiliser ce formalisme est qu'il permet l'utilisation de "Fast Fourier Transform" et de travailler dans l'espace réciproque. En revanche, une bonne résolution peut demander un grand nombre d'ondes planes et donc augmente considérablement le temps de calcul. On a dans ce cas recours aux pseudopotentiels.

I.1.1.9 Pseudopotentiel:

On peut diviser les électrons en deux catégories, les électrons de coeur et les électrons de valence. Les fonctions d'onde associées à ces deux types d'électrons comportent deux types de problèmes. Pour les électrons de coeur, les fonctions d'onde sont très piquées près du noyau, pour les électrons de valence, elles comportent des oscillations près du noyau. Pour pouvoir les décrire, il est alors nécessaire d'avoir un E_{cut} grand. Le but étant de réduire le temps de calcul, il est possible de résoudre ce problème en plusieurs étapes.

On remarque tout d'abord que les électrons de coeur sont peu sensibles au changement d'environnement chimique, autrement dit, il n'est pas nécessaire de les prendre en compte dans notre description. On prend leurs fonctions d'ondes identiques aux fonctions d'onde de l'atome isolé. Les électrons de coeur sont dit gelés et on cherche à résoudre les équations de Kohn-Sham seulement pour les électrons de valence.

Concernant les électrons de valence, l'idée est de remplacer la partie de la fonction d'onde oscillante ϕ , proche du noyau, définie par une distance r_c , par une pseudo

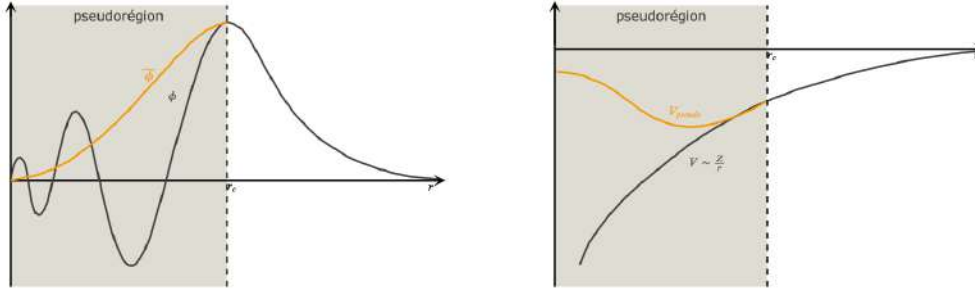


Figure I.2: Dans la pseudo-région définie par r_c , la fonction d'onde et le potentiel sont remplacés par une pseudo fonction d'onde et un pseudo potentiel.

fonction d'onde $\tilde{\phi}$ plus lisse choisie pour donner la même densité électronique que la vraie fonction d'onde entre $0 \leq r \leq r_c$, et ayant la même valeur que la vraie fonction d'onde en $r = r_c$.

On construit ensuite un pseudopotentiel tel que, pour $r \leq r_c$, la solution de l'équation de Kohn-Sham donne $\tilde{\phi}$ (voir Fig. I.2). Il existe différentes méthodes pour générer de tels pseudopotentiels. Au cours de cette thèse, nous avons utilisé des pseudopotentiels ultrasoft [119] présents dans la librairie du code Quantum Espresso [46, 45]. La construction de ces pseudopotentiels permet d'avoir un r_c plus grand rendant le potentiel plus "lisse" et diminuant ainsi l'énergie de cutoff E_{cut} .

L'utilisation de la DFT comme méthode de résolution de l'équation de Schrödinger a connue de nombreux succès et cette méthode est encore largement utilisée en physique de la matière. Néanmoins, elle demande de longs temps de calcul (scaling en $\mathcal{O}(N^3)$) rendant difficile son utilisation pour de grands systèmes, de longues simulations de dynamique moléculaire et pour la dynamique moléculaire d'intégrale de chemin. Une alternative est d'utiliser des potentiels avec une forme analytique connue comme les potentiels empiriques ou les potentiels machine learning.

I.1.2 Potentiels empiriques

On souhaite construire une fonction du potentiel interatomique. Pour cela, pour prendre en compte les différents types d'interaction, on décompose un potentiel V en termes d'interaction à un, deux, trois, jusqu'à M corps, M étant le nombre d'atomes du système :

$$V(\{\vec{R}\}) = \sum_{I=1}^M V_1(\vec{R}_I) + \frac{1}{2} \sum_{I=1}^M \sum_{\substack{J=1 \\ J \neq I}}^M V_2(\vec{R}_I, \vec{R}_J) + \frac{1}{6} \sum_{I=1}^M \sum_{\substack{J=1 \\ J \neq I}}^M \sum_{\substack{K=1 \\ K \neq I \\ K \neq J}}^M V_3(\vec{R}_I, \vec{R}_J, \vec{R}_K) + \dots \quad (\text{I.55})$$

où $\{\vec{R}\} = \vec{R}_1, \dots, \vec{R}_M$ et les \vec{R}_I sont les positions atomiques.

Dans cette thèse, ne seront utilisés que des potentiels à deux corps (Coulomb, Lennard-Jones et harmonique) et à trois corps (harmonique). Une fois le potentiel déterminé, on obtient les forces agissant sur un atome en prenant la dérivée du potentiel :

$$\vec{F}_I(\{\vec{R}\}) = -\vec{\nabla}_I V(\{\vec{R}\}) \quad (\text{I.56})$$

I.1.2.1 Potentiel coulombien

L'interaction électrostatique entre deux atomes est modélisée par le potentiel coulombien :

$$V(R) = \frac{K q_1 q_2}{\epsilon_r R} \quad (\text{I.57})$$

où R est la distance interatomique, q_1 et q_2 les charges électriques des atomes, K une constante dépendant du système d'unité utilisé et ϵ_r la constante diélectrique. Il est possible d'ajuster les charges ainsi que la constante diélectrique pour rendre compte des observations.

I.1.2.2 Potentiel Lennard-Jones

Le potentiel de type Lennard-Jones fait partie des potentiels à deux corps les plus couramment utilisés lors de simulations numériques. Il est composé de deux termes :

- un terme attractif rendant compte des interactions de dispersion de London
- un terme répulsif, à courte distance, empirique, conséquence du principe de Pauli pour les électrons

Il a pour expression :

$$V(R) = 4\epsilon \left[\left(\frac{\sigma}{R} \right)^{12} - \left(\frac{\sigma}{R} \right)^6 \right] \quad (\text{I.58})$$

où R est la distance interatomique et ϵ et σ sont des paramètres à ajuster et correspondent au minimum du potentiel et à la distance où le potentiel change de signe. L'ajustement des paramètres se fait sur des données expérimentales.

Au cours de cette thèse, nous avons utilisé des potentiels Lennard-Jones pour décrire l'interaction entre un calcium et un oxygène [89] (Table I.1), ainsi qu'entre deux oxygènes pour le modèle de l'eau (q)-SPC/fw (partie I.1.2.3).

I.1.2.3 Potentiel (q)-SPC/fw - modèle de l'eau

Plusieurs modèles ont été développés pour décrire l'eau liquide. Nous utiliserons le modèle flexible Single Point Charge (SPC/fw) [125]. Les interactions entre atomes d'oxygène et d'hydrogène sont décomposées en un terme intramoléculaire et un terme intermoléculaire:

$$V = V_{inter} + V_{intra} \quad (\text{I.59})$$

V_{inter} est composé de potentiels coulombiens et Lennard-Jones :

$$V_{inter} = \sum_{I,J} 4\epsilon_{ij} \left[\left(\frac{\sigma_{IJ}}{R_{IJ}} \right)^{12} - \left(\frac{\sigma_{IJ}}{R_{IJ}} \right)^6 \right] + K \frac{q_I q_J}{R_{IJ}} \quad (\text{I.60})$$

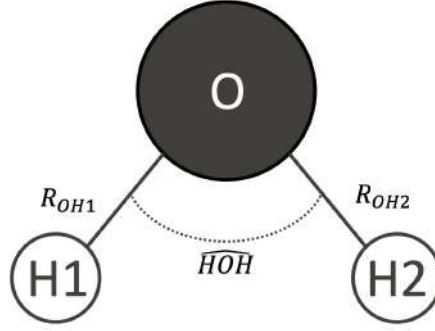


Figure I.3: Schéma de la molécule d'eau

où les R_{IJ} sont les distances entre atome I et atome J , ϵ_{IJ} et σ_{IJ} les paramètres des potentiels de Lennard-Jones et q_I et q_J leurs charges. La constante diélectrique est prise égale à un.

Pour la partie intramoléculaire, la molécule d'eau est définie par trois sites correspondant à l'atome d'oxygène et aux deux atomes d'hydrogène. On a alors trois degrés de liberté, les distances des deux liaisons O-H, R_{OH1} et R_{OH2} ainsi que l'angle \widehat{HOH} (voir Figure I.3). Des potentiels harmoniques décrivent les interactions à deux et trois corps :

$$V_{intra} = \frac{k_b}{2} \left[(R_{OH1} - R_{OH}^0)^2 + (R_{OH2} - R_{OH}^0)^2 \right] + \frac{k_a}{2} \left(\widehat{HOH} - \widehat{HOH}^0 \right)^2 \quad (\text{I.61})$$

avec k_b et k_a deux paramètres harmoniques et, R_{OH}^0 et \widehat{HOH}^0 , la longueur de la liaison O-H et l'angle \widehat{HOH} à l'équilibre. Les valeurs des paramètres du modèle se trouvent Table I.1. S'y trouvent aussi les valeurs pour le modèle q-SPC/fw [84], qui est une reparamétrisation du modèle SPC/fw, effectuée à partir de simulations quantiques. Ce modèle est plus adapté, et donne une meilleure description de l'eau, pour les simulations quantiques, par exemple les dynamiques moléculaires d'intégrales de chemin.

	Potentiel	Paramètres
Modèle (q)-SPC/fwref	Coulomb	$q(O) = -0.82(-0.84)$, $q(H) = 0.41(0.42)$
	Lennard-Jones	$\epsilon_{OO} = 0.006739(0.006739)$, $\sigma_{OO} = 3.165492(3.165492)$
	Harmonique O-H	$k_b = 1059.162(1059.162)$, $R_{OH}^0 = 1.012(1.000)$
	Harmonique H-O-H	$k_a = 75.90(75.90)$, $\widehat{HOH}^0 = 113.24(112.00)$
Calcium-Oxygène ref	Lennard-Jones	$\epsilon_{CaO} = 0.00095$, $\sigma_{CaO} = 3.35$

Table I.1: Paramètres des potentiels empiriques utilisés. Pour la modélisation de l'eau, les valeurs en noir correspondent au modèle SPC/fw et en orange au modèle q-SPC/fw. Les charges sont en e , ϵ en eV , σ et r_{OH}^0 en Å et \widehat{HOH}^0 en degrés.

I.1.3 Potentiels Machine Learning (ML)

L'utilisation de potentiels empiriques réduit grandement le temps de calcul. En revanche, du fait de leur simplicité, leur précision est grandement réduite comparée aux méthodes ab-initio. Dans certains cas, ils ne permettent pas de rendre compte correctement des propriétés d'un système.

Les potentiels machine learning permettent de combler le manque entre les potentiels empiriques et les méthodes ab initio et d'avoir une précision proche de l'ab initio avec un temps de calcul grandement réduit.

L'idée est, comme pour les potentiels empiriques, de construire une fonction de l'énergie à partir de laquelle on peut déterminer les forces. En revanche, la forme de cette fonction ne repose pas sur des considérations physiques, on la choisit aussi complexe que l'on souhaite. Le principe général est le suivant :

- On construit une base de données constituées de configurations atomiques, $\{\vec{R}_I\}$, avec l'énergie, E , les forces, $\{\vec{F}_I\}$ et/ou les contraintes, qui sont calculées à partir de méthodes ab-initio. Ces données représentent des points de la PES que l'on souhaite reproduire avec le potentiel machine learning (voir Figure I.4).
- On construit un nouvel ensemble de coordonnées, appelées descripteurs, $\{G_J\}$, à partir des configurations atomiques $\{\vec{R}_I\}$ (voir partie I.1.3.2). Nous utilisons des descripteurs dit locaux, représentant l'environnement d'un atome et on suppose ensuite que l'énergie de notre système est la somme des énergies atomiques de chaque atome E_I :

$$E(\{G_J\}) = \sum_{I=1}^M E_I \quad (\text{I.62})$$

La PES que l'on souhaite reproduire dépend maintenant des descripteurs (voir Figure I.4).

- On détermine une forme du potentiel machine learning, comportant des paramètres, que l'on ajuste par regression à l'aide des données ab-initio (voir partie I.1.3.3 et Figure I.4).

Le potentiel machine learning ainsi généré peut être utilisé dans des simulations atomistiques. La composante k d'une force atomique ($k = x, y, z$) peut être calculée :

$$F_k = -\frac{\partial E}{\partial k} = -\sum_{I=1}^M \frac{\partial E_I}{\partial k} = -\sum_{I=1}^M \sum_{J=1}^{N_J} \frac{\partial E_I}{\partial G_{IJ}} \frac{\partial G_{IJ}}{\partial k} \quad (\text{I.63})$$

où G_{IJ} est le descripteur J de l'atome I , et N_J le nombre de descripteurs de l'atome I .

Ces différentes étapes sont décrites en détail dans les parties suivantes.

I.1.3.1 Base de données

La base de données représente l'ensemble des points de la PES que l'on souhaite reproduire. Elle est utilisée pour ajuster les paramètres du potentiel machine learning. A chaque étape de la regression, on évalue la différence entre les valeurs données par

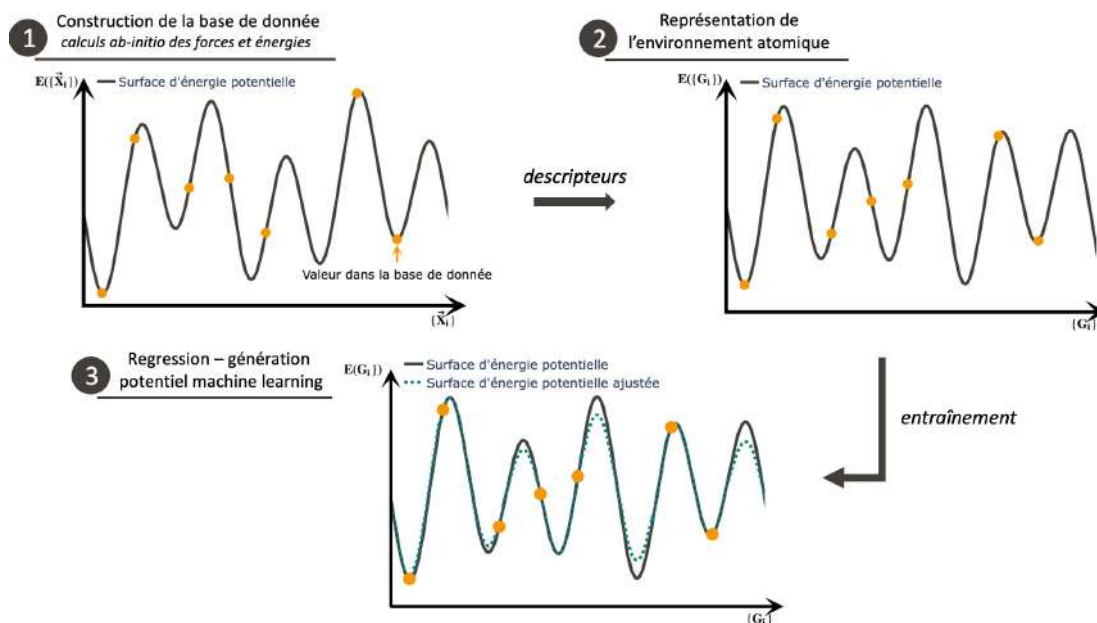


Figure I.4: Etapes de la construction d'un potentiel machine learning. Inspiré de [31].

le potentiel machine learning et les valeurs de la base de données. On souhaite alors avoir une base de données la plus petite possible, pour que la régression ne soit pas trop gourmande en temps de calcul, et échantillonnant l'ensemble de la PES, et pas seulement les minima locaux, pour qu'il soit possible d'interpoler correctement entre les points de la base de données. On utilise des méthodes ab-initio pour construire cette base de données. Une première approche est alors d'effectuer des dynamiques moléculaires ab initio et de récupérer des configurations à partir de cette simulation. Comme l'on souhaite avoir un bon échantillonnage de la PES, il peut être nécessaire d'avoir recours à de longues simulations, rendant cette méthode peu efficace. D'autres méthodes ont été proposées permettant une sélection plus efficace des configurations de la base de données (voir [13] pour un aperçu). Ces dernières années, des méthodes d'apprentissage actif ont notamment été développées et leur principe est expliqué à la fin de cette partie.

La base de données est alors composée des positions en coordonnées cartésiennes des différentes configurations qui la composent ainsi que les énergies et forces associées. Néanmoins, la génération d'un potentiel machine learning nécessite l'utilisation de coordonnées différentes, les descripteurs.

I.1.3.2 Descripteurs

L'algorithme de régression permettant la génération du potentiel machine learning dépend d'un vecteur d'entrée. Si ce vecteur d'entrée est composé des coordonnées cartésiennes des configurations de la base de données, il n'est alors plus possible d'utiliser le potentiel pour des systèmes ayant un nombre d'atomes différents à cause de la taille fixe du vecteur. De plus, une simple rotation, translation ou permutation de deux éléments identiques va changer les valeurs des coordonnées cartésiennes,

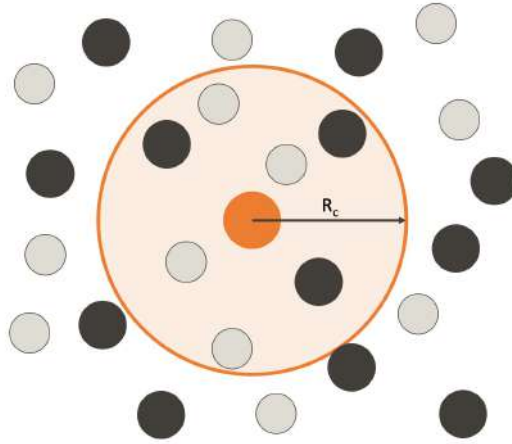


Figure I.5: Environnement atomique d'un atome (en orange) défini par la zone sphérique de rayon R_c

et donc le vecteur d'entrée, alors que l'énergie totale du système est inchangée. On définit alors un autre jeu de coordonnées, appelées descripteurs, dépendant des positions cartésiennes des atomes, mais donnant la même valeur pour toutes représentations énergétiquement équivalentes d'une molécule.

Les descripteurs que nous utilisons permettent la description de l'environnement atomique, défini par un rayon de cutoff R_C , d'un atome (voir Fig. I.5). Pour s'assurer que les descripteurs et leurs dérivées décroissent lentement vers zéro au rayon de cutoff, on utilise généralement une fonction de cutoff. Une fonction de cutoff régulièrement utilisée est :

$$f_c(R_{IJ}) = \begin{cases} \frac{1}{2} \left[\cos\left(\pi \frac{R_{IJ}}{R_c}\right) + 1 \right] & \text{pour } R_{IJ} \leq R_c \\ 0 & \text{pour } R_{IJ} > R_c \end{cases} \quad (\text{I.64})$$

où R_{IJ} est la distance interatomique. D'autres fonctions, tangente hyperbolique, exponentielle, polynomiales, peuvent aussi être utilisées. On décompose ensuite généralement les descripteurs en une partie radiale et une partie angulaire.

Descripteurs atome-centré (G)

Behler [12] propose des descripteurs à deux corps :

$$G_I = \sum_{J \neq I} e^{-\eta(R_{IJ}-R_s)^2} \cdot f_c(R_{IJ}) \quad (\text{I.65})$$

et à trois corps :

$$G_I = 2^{1-\zeta} \sum_{\substack{J, K \neq I \\ J < K}} (1 + \lambda \cos \theta_{IJK})^\zeta e^{-\eta((R_{IJ}-R_s)^2 + (R_{IK}-R_s)^2 + (R_{JK}-R_s)^2)} \cdot f_c(R_{IJ}) \cdot f_c(R_{IK}) \cdot f_c(R_{JK}) \quad (\text{I.66})$$

avec $\cos(\theta_{IJK}) = \vec{R}_{IJ} \cdot \vec{R}_{IK} / (R_{IJ}R_{IK})$. Les paramètres η , R_s , λ et ζ ne sont pas optimisés pendant l'entraînement du potentiel et sont appelés hyperparamètres. En

général, quelques dizaines de descripteurs sont utilisés pour chaque élément chimique.

Descripteurs smooth-SO3 (p)

D'autres types de descripteurs, basés sur une description harmonique de la densité du voisinage atomique, ont été développés. On commence par définir la fonction de la densité du voisinage atomique, ρ , d'un atome, définie par le rayon de cutoff R_C :

$$\rho(R) = \sum_{I, \vec{R}_I \leq R_C} \delta(\vec{R} - \vec{R}_I) \quad (\text{I.67})$$

Il est ensuite possible de décomposer cette fonction en somme d'harmoniques sphériques Y_{lm} :

$$\rho(\hat{\vec{R}}) = \sum_{l=0}^{+\infty} \sum_{m=-l}^l c_{lm} Y_{lm}(\hat{\vec{R}}) \quad (\text{I.68})$$

où $\hat{\vec{R}}$ est le vecteur donnant le point sur la sphère unité donné par la direction de \vec{R} . Les coefficients de l'expansion sont donnés par :

$$c_{lm} = \langle \rho | Y_{lm} \rangle = \sum_J Y_{lm}(\hat{\vec{R}}_{IJ}) \quad (\text{I.69})$$

à partir desquels on construit des descripteurs, p_l , appelés descripteurs SO3, satisfaisant les propriétés d'invariance :

$$p_l = \sum_{m=-l}^l c_{lm} c_{lm}^* \quad (\text{I.70})$$

Le problème rencontré avec les descripteurs SO3 est qu'ils ne contiennent pas d'information radiale. Les auteurs de [10] proposent une modification. Tout d'abord, pour avoir une fonction de mesure de similarité des environnements atomiques plus lisse, ils définissent la fonction de densité du voisinage atomique comme la somme de termes gaussiens :

$$\rho(\vec{R}) = \sum_{I, \vec{R}_I \leq R_C} e^{-\alpha |\vec{R} - \vec{R}_I|^2} \quad (\text{I.71})$$

qu'ils décomposent ensuite en utilisant les harmoniques sphériques et une base radiale g_n :

$$\rho(\vec{R}) = \sum_{l=0}^{+\infty} \sum_{m=-l}^l c_{nlm} g_n(R) Y_{lm}(\hat{\vec{R}}) \quad (\text{I.72})$$

avec :

$$c_{nlm} = \langle Y_{lm} g_n | \rho \rangle \quad \text{et} \quad g_n(R) = \sum_{k=1}^{n_{max}} S_{nk}^{1/2} \phi_k(R) \quad (\text{I.73})$$

où n_{max} est un hyperparamètre et :

$$\phi_k(R) = \frac{(R_C - R)^{k+2}}{N_k}, N_k = \left(\frac{2R_C^{2k+7}}{(2k+5)(2k+6)(2k+7)} \right)^{\frac{1}{2}}, S_{pq} = \int_0^{R_C} R^2 \phi_p \phi_q dr \quad (\text{I.74})$$

Finalement, de la même façon que pour les descripteurs SO3, on définit des descripteurs "smooth" SO3, $p_{n_1 n_2 l}$, par :

$$p_{n_1 n_2 l} = \sum_{m=-l}^l c_{n_1 l m} c_{n_2 l m}^* \quad (\text{I.75})$$

Descripteur SO4 (B)

Une approche alternative pour prendre en compte la composante radiale est de représenter la fonction de densité du voisinage sur une hyper-sphère, de rayon R_0 , en quatre dimensions. Les coordonnées polaires sont données par :

$$\begin{aligned} \theta &= \arccos\left(\frac{z}{R}\right), \\ \phi &= \arctan\left(\frac{y}{x}\right), \\ \omega &= \frac{\pi R}{R_0} \end{aligned} \quad (\text{I.76})$$

Pour éviter les problèmes de discontinuité, on introduit une fonction de cutoff dans la définition de la densité du voisinage :

$$\rho(\vec{R}) = \delta(\vec{R}) + \sum_J f_c(\vec{R}) \delta(\vec{R} - \vec{R}_J) \quad (\text{I.77})$$

que l'on décompose à l'aide des fonctions harmoniques d'hyper-sphère $U_{m'm}^j$:

$$\rho = \sum_{j=0}^{+\infty} \sum_{m', m=-j}^j c_{m'm}^j U_{m'm}^j(\omega, \theta, \phi) \quad \text{avec} \quad c_{m'm}^j = \langle U_{m'm}^j | \rho \rangle \quad (\text{I.78})$$

Finalement, on construit les descripteurs SO4 $B_{jj_1 j_2}$:

$$B_{jj_1 j_2} = \sum_{m', m=-j}^j c_{m'm}^{*j} \sum_{m'_1, m_1=-j_1}^{j_1} c_{m'_1 m_1}^{j_1} \sum_{m'_2, m_2=-j_2}^{j_2} c_{m'_2 m_2}^{j_2} C_{m m_1 m_2}^{j j_1 j_2} C_{m' m'_1 m'_2}^{j j_1 j_2} \quad (\text{I.79})$$

où les $C_{\alpha\beta\gamma}^{ijk}$ sont les coefficients de Clebsh-Gordan.

Pour une description complète des différents descripteurs utilisés pour la construction de potentiel machine learning, ainsi que leurs implémentations, le lecteur est invité à consulter [10, 126, 127, 110].

I.1.3.3 Potentiel et ajustement

Le dernier élément pour la construction du potentiel machine learning consiste à déterminer la forme de la fonction de l'énergie $E(\{G_I\})$ et de procéder à l'ajustement. Trois approches sont couramment utilisées dans la littérature: les méthodes à noyaux,

les regressions linéaires et les réseaux de neurones. Nous développons ici les deux dernières.

Regression linéaire

Ce type de regression consiste à développer, sous forme de combinaisons linéaires des descripteurs, l'énergie atomique d'un élément chimique α_I [111] :

$$E_I = w_0^{\alpha_I} + \vec{w}^{\alpha_I} \cdot \vec{G}_I \quad (\text{I.80})$$

où $w_0^{\alpha_I}$ et \vec{w}^{α_I} sont des poids à ajuster et \vec{G}_I le vecteur des descripteurs de l'atome I . On somme ensuite les énergies atomiques pour obtenir l'énergie totale du système.

Regression par réseau de neurones

L'élément de base de l'utilisation des réseaux de neurones pour la construction d'un potentiel sont les réseaux de neurones à propagation avant schématisés Fig. I.6. Il est composé de différentes couches. La première couche, dite d'entrée, est composée des descripteurs et la dernière couche, dite de sortie, est l'énergie. Entre les deux se trouvent des couches "cachées", composées de neurones (voir Fig. I.7), qui servent à construire la fonction de l'énergie. Entre deux couches, les neurones sont connectés et un poids $w_{ki}^{j-1,j}$ est appliqué à chaque entrée de neurone. La valeur y_i^j du neurone i dans la couche j est alors donnée par la combinaison linéaire des valeurs des neurones de la couche $j-1$ et de leurs poids. Un biais b_i^j est ensuite ajouté et une fonction d'activation f_i^j appliquée donnant alors :

$$y_i^j = f_i^j \left(b_i^j + \sum_{k=1}^{N_{j-1}} w_{ki}^{j-1,j} y_k^{j-1} \right) \quad (\text{I.81})$$

où N_{j-1} est le nombre de neurones dans la couche $j-1$.

La fonction d'activation est une fonction non linéaire sauf pour le neurone de sortie où la fonction $y = x$ est utilisée. La valeur de sortie est alors un ensemble de fonctions imbriquées. Dans l'exemple de la figure I.6, elle est donnée par :

$$E = f_1^3 \left(b_1^3 + \sum_{l=1}^{N_2} w_{kl}^{23} f_k^2 \left(b_k^2 + \sum_{k=1}^{N_1} w_{jk}^{12} f_j^1 \left(b_j^1 + \sum_{i=1}^{N_G} w_{ij}^{01} G_i \right) \right) \right) \quad (\text{I.82})$$

où N_1 et N_2 sont les nombres de neurones des couches cachées 1 et 2 et N_G le nombre de descripteurs d'entrée.

Comme mentionné dans la partie I.1.3.2, on utilise des descripteurs locaux pour s'affranchir des problèmes d'invariance de l'énergie et de taille du système et on suppose que l'énergie peut s'écrire comme la somme des énergies atomiques. On utilise alors des réseaux de neurones à haute dimension permettant de rendre compte de cette description. On construit un réseau de neurones à propagation avant pour chaque élément chimique du système qui donne en sortie l'énergie atomique. On peut ensuite sommer l'ensemble de ces énergies. Le principe est schématisé Fig.I.8 pour un système composé de deux éléments chimiques différents.

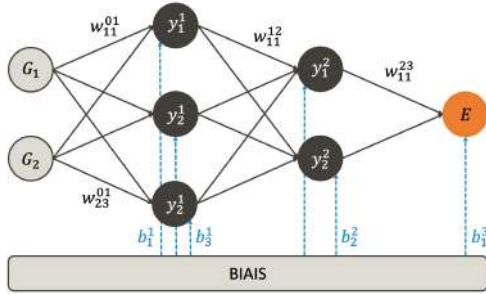


Figure I.6: Représentation schématique d'un réseau de neurones à propagation avant composé de deux neurones d'entrée, de deux couches cachées chacune composée de trois et deux neurones et d'un neurone de sortie. Inspiré de [14]

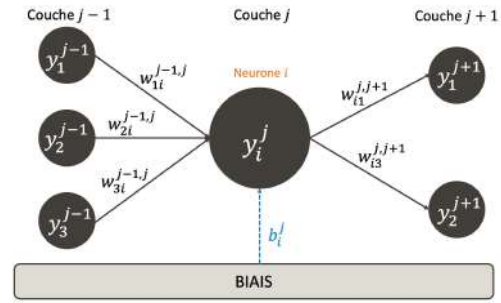


Figure I.7: Représentation schématique d'un neurone i dans la couche j . La valeur y_i^j est la combinaison linéaire des valeurs y_i^{j-1} de neurone de la couche $j-1$ et des poids à laquelle est ajouté un biais b_i^j . La valeur y_i^j est ensuite transmise aux neurone de la couche $j+1$. Inspiré de [14]

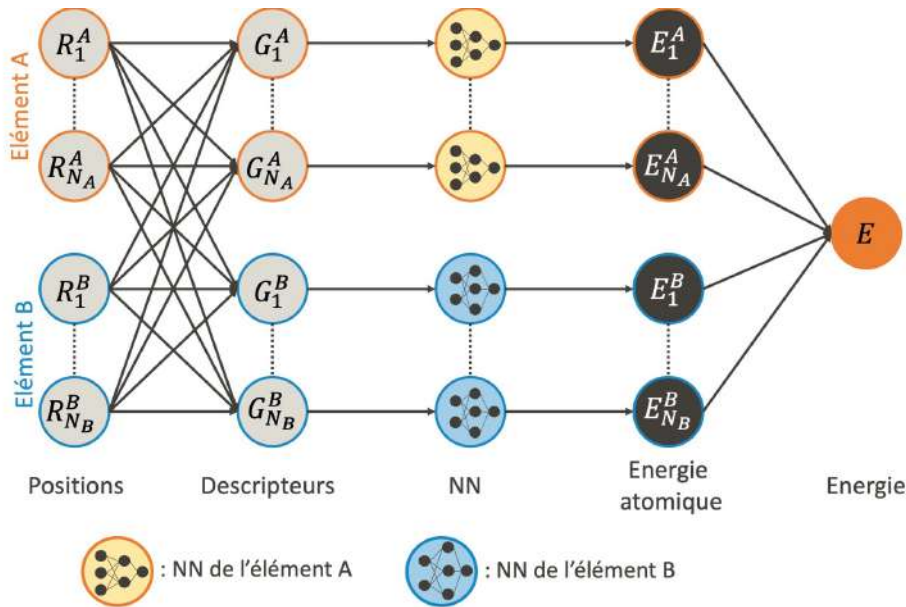


Figure I.8: Représentation schématique d'un réseau de neurones à haute dimension pour deux éléments chimiques A et B. Les positions atomiques sont utilisées pour construire les vecteurs des descripteurs de chaque atome qui sont ensuite donnés en entrée du réseau de neurones à propagation avant (NN) de l'élément. Chaque réseau de neurones à propagation avant donne en sortie les énergies atomiques qui sont ensuite sommées. Inspiré de [14]

Optimisation

L'optimisation des poids w et des biais b , que l'on appelle entraînement, se fait par minimisation d'une fonction de coût Γ de la forme :

$$\Gamma = \sum_{i=1}^{N_{data}} (E_i - E_i^{data})^2 + \beta^2 \sum_{i=1}^{N_{data}} \sum_{j=1}^{M_j} (\vec{F}_{ij} - \vec{F}_{ij}^{data})^2 \quad (\text{I.83})$$

où N_{data} est le nombre de configurations dans la base de données, l'exposant $data$ indique l'énergie et les forces de la base de données et M_j le nombre d'atomes dans la configuration j . Si les contraintes sont aussi utilisées pour l'ajustement du potentiel, un terme supplémentaire est ajouté. Le paramètre β contrôle l'importance relative des forces par rapport à l'énergie.

Les poids sont initialisés selon une procédure choisie, et l'entraînement se fait ensuite itérativement. Une itération correspond à une "époque". Plusieurs algorithmes d'optimisation peuvent être utilisés, comme la rétropropagation du gradient [93] ou la méthode du filtre de Kalman [100].

Pour suivre l'entraînement et juger de la précision du potentiel, il est possible de calculer la racine de l'erreur quadratique moyenne ($RMSE$) de l'énergie et des forces :

$$RMSE(E) = \sqrt{\frac{1}{N_{data}} \sum_{i=1}^{N_{data}} (E_i - E_i^{data})^2} \quad (\text{I.84})$$

$$RMSE(F) = \sqrt{\frac{1}{N_{data}} \sum_{i=1}^{N_{data}} \sum_{j=1}^{M_j} (\vec{F}_{ij} - \vec{F}_{ij}^{data})^2} \quad (\text{I.85})$$

Lors de la procédure, il peut arriver que l'on observe un "surentraînement". Le potentiel généré donne de très bon résultats pour les configurations de la base de données mais donne de mauvaises prédictions en dehors. Pour détecter ce phénomène, on sépare la base de données en deux. Une partie servant à l'apprentissage, une autre au test (voir Figure I.9).

Apprentissage actif

La précision du potentiel machine learning, en particulier dans son interpolation, dépend grandement du choix des configurations de la base de données. Pour éviter de nombreux calculs ab-initio, des méthodes d'apprentissage actif, ou "on-the-fly", ont été développées. L'idée générale (voir Figure I.10) repose sur une procédure auto-cohérente. On commence par sélectionner un nombre N_{data} de configurations atomiques. Pour ces N_{data} configurations, un calcul ab-initio est effectué donnant une première base de données. On construit ensuite un premier potentiel ML. On teste alors un critère de convergence, par exemple la convergence des RMSEs ou d'une propriété du système. Si le critère est validé, on sort de la procédure, sinon, on génère une dynamique moléculaire (voir partie I.2). A la fin de cette dynamique, un nombre N_2 de configurations est sélectionné. Elles peuvent être prises par exemple à intervalle régulier le long de la trajectoire ou choisies, à l'aide du calcul d'une fonction de similitude, comme étant les moins similaires aux N_{data} déjà présentes dans la base de données. Les N_2 configurations sont ajoutées à la base de données et on répète la procédure.

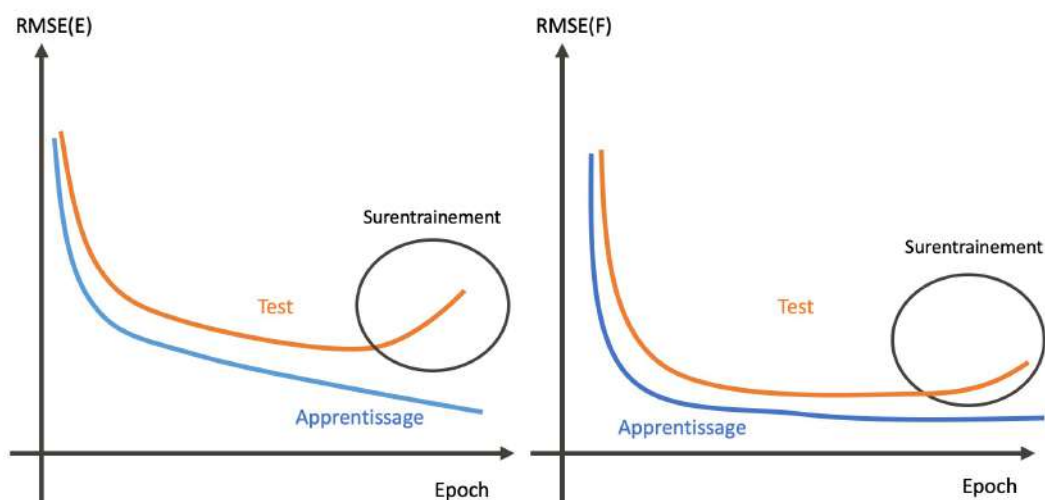


Figure I.9: Erreur quadratique de l'énergie et des forces sur la base de données servant à l'apprentissage et celle servant de test en fonction de l'époque. Le surentraînement est caractérisé par l'augmentation de l'erreur calculé sur la base de données de test alors que l'erreur sur la base de données de l'apprentissage continue à diminuer.

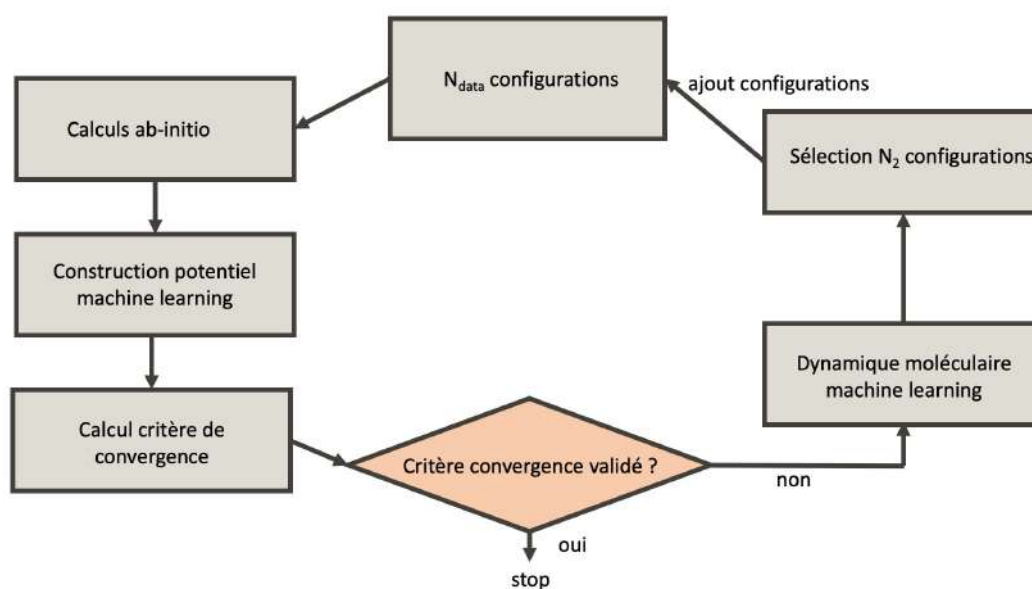


Figure I.10: Principe schématique de l'apprentissage actif.

I.1.3.4 Méthodes et codes utilisés

Lors de cette thèse, deux méthodes ont été utilisées pour générer un potentiel machine learning pour un système composé d'un calcium et de 62 molécules d'eau (voir chapitre IV).

Un premier potentiel machine learning à réseau de neurones a été construit à l'aide du code N2P2 [101]. La base de données est composée des configurations issues de dynamiques moléculaires, ab-initio (DFT) à différentes températures provenant de l'étude de la coordinence du calcium en solution aqueuse (voir chapitre III).

Un second potentiel a lui été construit à l'aide du code MLACS [21] et les calculs ab-initio ont été réalisés par DFT. C'est un code utilisant l'apprentissage actif. Dans notre cas, nous avons commencé la procédure avec une seule configuration de départ. Le potentiel machine learning ainsi lors de la première étape donne de mauvais résultats et il est possible que les dynamiques moléculaires faites avec celui-ci explosent. Nous avons donc utilisé une option du code MLACS qui permet de définir un potentiel de référence. Le potentiel machine learning ajuste alors, par régression linéaire, l'écart entre les valeurs du potentiel de référence et celles de la DFT. Pour notre système, nous avons utilisé le modèle SPC/fw pour décrire l'eau et un potentiel de type Lennard-Jones pour l'interaction calcium-oxygène (voir partie I.1.2). A chaque dynamique moléculaire de la boucle auto-cohérente, c'est la dernière configuration qui a été sélectionnée et ajoutée à la base de données.

Ces deux méthodes sont interfacées avec le code Lammmps [109, 32] avec lequel nous avons effectué des dynamiques moléculaires.

Pour conclure, on notera que dû à leur caractère "boîte noire", les potentiels machine learning peuvent avoir un comportement non-intuitif. Par exemple, bien que nous contrôlons la convergence de l'apprentissage en calculant les erreurs quadratiques de l'énergie et des forces, ce n'est pas forcément le potentiel de l'époque où l'erreur est la plus faible qui donnera de meilleurs résultats. Il peut aussi être nécessaire de tester les différents descripteurs ainsi que la taille de la base de données. La construction du potentiel demande donc de nombreux tests.

I.2 Dynamique moléculaire

Les simulations atomistiques permettent d'avoir accès aux grandeurs thermodynamiques ainsi qu'aux propriétés d'équilibre d'un système. Dans l'ensemble microcanonique, la moyenne d'ensemble d'une grandeur a est donnée par :

$$\langle a \rangle = \frac{\int dx a(x) \delta(\mathcal{H}(x) - E)}{\int \delta(\mathcal{H}(x) - E)} \quad (\text{I.86})$$

où $\mathcal{H}(x)$ est l'hamiltonien du système.

Le principe ergodique permet de remplacer la moyenne d'ensemble par une moyenne temporelle :

$$\langle a \rangle = \bar{a} = \lim_{T \rightarrow +\infty} \frac{1}{T} \int_0^T dt a(x(t)) \quad (\text{I.87})$$

La dynamique moléculaire repose sur ce principe. Etant donné un système (=ensemble d'atomes), on génère une trajectoire ergodique, telle que $\mathcal{H}(x) = E$, en résolvant les équations du mouvement numériquement et en discrétisant le temps avec un pas Δt , pour chaque atome I :

$$\dot{\vec{q}}_I = \frac{\partial \mathcal{H}}{\partial \vec{p}_I}, \quad \dot{\vec{p}}_I = -\frac{\partial \mathcal{H}}{\partial \vec{q}_I} \quad (\text{I.88})$$

où $\{\vec{q}_I\}$ sont les coordonnées généralisées et $\{\vec{p}_I\}$ les moments conjugués.

On récupère le long de la trajectoire les données nécessaires au calcul de la grandeur d'intérêt, et la moyenne est alors donnée par :

$$\bar{a} = \frac{1}{N_{steps}} \sum_{n=1}^{N_{steps}} a(x(n\Delta t)) \quad (\text{I.89})$$

avec N_{steps} le nombre de pas de la dynamique.

A noter que la dynamique moléculaire permet aussi d'avoir accès aux grandeurs dynamiques du système.

Pour les simulations de dynamique moléculaire, on a alors besoin des positions et vitesses initiales du système d'atomes et des interactions interatomiques régissant cette dynamique. Il est possible d'ajouter des contraintes extérieures pour travailler dans d'autres ensembles statistiques, on verra partie I.2.4 le cas de la température et de l'ensemble canonique (NVT).

La première étape est donc de construire un algorithme faisant évoluer les positions des atomes. On verra le cas d'algorithmes symplectiques et réversibles. Ces derniers ont la propriété d'avoir une faible dérive de l'énergie du système sur des temps de simulation longs. Pour une construction générale, le lecteur est renvoyé à [116]. On présente le cas de l'algorithme le plus couramment utilisé, l'algorithme de Verlet [122].

I.2.1 Algorithme de Verlet

Le temps étant discrétisé, on cherche la position de l'atome I au temps $t + \Delta t$ connaissant sa position au temps t . L'algorithme de Verlet repose sur un développement

de Taylor :

$$\vec{R}_I(t + \Delta t) = \vec{R}_I(t) + \frac{d\vec{R}_I}{dt}(t)\Delta t + \frac{1}{2} \frac{d^2\vec{R}_I}{dt^2}(t)\Delta t^2 + \frac{1}{6} \frac{d^3\vec{R}_I}{dt^3}(t)\Delta t^3 + \mathcal{O}(\Delta t^4) \quad (\text{I.90})$$

En faisant un autre développement de Taylor au temps $t - \Delta t$:

$$\vec{R}_I(t - \Delta t) = \vec{R}_I(t) - \frac{d\vec{R}_I}{dt}(t)\Delta t + \frac{1}{2} \frac{d^2\vec{R}_I}{dt^2}(t)\Delta t^2 - \frac{1}{6} \frac{d^3\vec{R}_I}{dt^3}(t)\Delta t^3 + \mathcal{O}(\Delta t^4) \quad (\text{I.91})$$

et en sommant les équations I.90 et I.91, on obtient :

$$\vec{R}_I(t + \Delta t) + \vec{R}_I(t - \Delta t) = 2\vec{R}_I(t) + \frac{d^2\vec{R}_I}{dt^2}(t)\Delta t^2 + \mathcal{O}(\Delta t^4) \quad (\text{I.92})$$

$$\Leftrightarrow \vec{R}_I(t + \Delta t) = 2\vec{R}_I(t) - \vec{R}_I(t - \Delta t) + \frac{d^2\vec{R}_I}{dt^2}(t)\Delta t^2 + \mathcal{O}(\Delta t^4) \quad (\text{I.93})$$

En utilisant la deuxième loi de Newton et la notation $\vec{F}_I(t) = \vec{F}_I(\vec{R}_1(t), \dots, \vec{R}_M(t))$ on a finalement :

$$\vec{R}_I(t + \Delta t) = 2\vec{R}_I(t) - \vec{R}_I(t - \Delta t) + \frac{\vec{F}_I(t)}{M_I}\Delta t^2 + \mathcal{O}(\Delta t^4) \quad (\text{I.94})$$

L'équation I.94 correspond à l'algorithme de Verlet. En pratique, on démarre d'un ensemble de positions initiales $\vec{R}_1(0), \dots, \vec{R}_M(0)$ et on utilise l'équation I.90 à l'ordre 2 pour déterminer les positions à l'instant Δt , $\vec{R}_1(\Delta t), \dots, \vec{R}_M(\Delta t)$. La trajectoire est ensuite générée en utilisant l'équation I.94. Si besoin, les vitesses sont calculées à partir de :

$$\frac{d\vec{R}_I}{dt}(t) = \vec{v}_I(t) = \frac{\vec{R}_I(t + \Delta t) - \vec{R}_I(t - \Delta t)}{2\Delta t} \quad (\text{I.95})$$

L'algorithme de Verlet ne prend pas en compte explicitement les vitesses, il existe une alternative appelée "velocity Verlet" [107]. Cette fois-ci un développement de Taylor est effectué pour les positions et les vitesses et permet d'obtenir :

$$\vec{R}_I(t + \Delta t) = \vec{R}_I(t) + \vec{v}_I(t)\Delta t + \frac{\vec{F}_I}{2M_I}(t)\Delta t^2 + \mathcal{O}(\Delta t^3) \quad (\text{I.96})$$

$$\vec{v}_I(t + \Delta t) = \vec{v}_I(t) + \frac{\Delta t}{2m_I} \left(\vec{F}_I(t) + \vec{F}_I(t + \Delta t) \right) \quad (\text{I.97})$$

Il faut alors aussi fournir les vitesses initiales. On commence par évaluer les forces au temps t , ce qui permet en utilisant I.96 de déterminer les positions au pas $t + \Delta t$. On peut alors calculer les forces au temps $t + \Delta t$ et finalement les vitesses, en utilisant l'équation I.97, au temps $t + \Delta t$.

I.2.2 Conditions aux limites périodiques

Lors d'une dynamique moléculaire, un système de N atomes dans une cellule est étudié. Malgré les progrès de performance, le calcul des interactions interatomiques et de l'évolution temporelle du système ne permet pas de simuler des systèmes réels

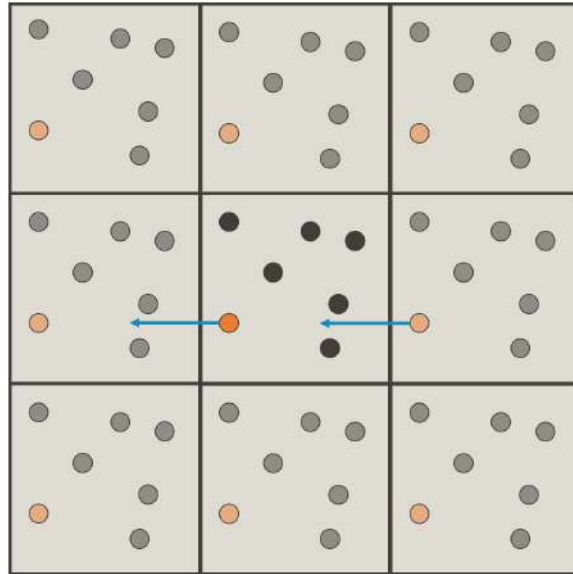


Figure I.11: Conditions aux limites périodiques pour un système 2D. La boîte centrale est répétée dans toutes les directions de l'espace. Si un atome sort de la boîte centrale (en orange), celui se trouvant dans la réplique opposée rentre dans la boîte centrale.

($N \sim 10^{23}$). Les conditions aux limites périodiques permettent de simuler un système de taille infinie à partir de répétitions de la cellule de départ. Cela permet d'obtenir des propriétés macroscopiques du système tout en considérant un nombre réduit d'atomes.

On commence par déterminer une boîte de simulation, celle-ci est alors répliquée dans toutes les directions de l'espace. Si, lors de la simulation, un atome de la boîte de départ sort de celle-ci et se retrouve dans une réplique, l'atome correspondant de la réplique opposée rentre dans la boîte de départ, le principe est schématisé figure I.11 L'utilisation des conditions aux limites à elles seules ne permet pas de résoudre le problème du nombre d'interactions interatomiques à calculer.

Pour les interactions à courte portée, on adopte alors la convention d'image minimum. C'est à dire qu'on introduit un rayon de cutoff R_{cut} , au delà duquel une interaction n'est plus évaluée. Pour s'assurer qu'un atome n'interagisse pas avec l'une de ses copies, le rayon de cutoff ne doit pas excéder la moitié de la taille de la cellule (dans le cas d'une cellule cubique, voir Fig.I.12).

Dans le cas des interactions à longue portée, comme l'interaction coulombienne, on utilise la sommation d'Ewald [36] ou des méthodes dérivées comme la "Particle-Particle-Particle-Mesh" [58]. La sommation des contributions des différentes interactions coulombiennes entre atomes est alors effectuée dans l'espace réel pour les atomes se trouvant à une distance inférieure à un certain rayon de cutoff $R_{cut,coul}$. Pour les atomes se situant à une distance plus grande, le calcul est fait dans l'espace réciproque accélérant grandement la vitesse de calcul et permettant de résoudre les problèmes de convergence.

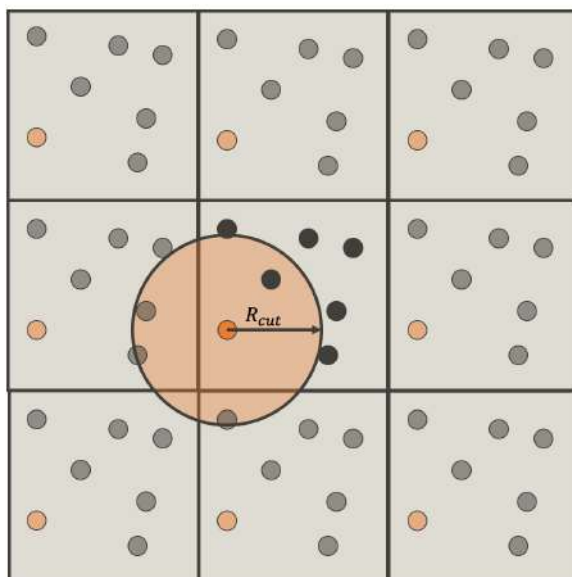


Figure I.12: Convention d'image minimum pour l'atome orange définie par le cercle de rayon R_{cut} . Les interactions entre l'atome orange et les atomes en dehors du cercle ne sont pas calculées.

I.2.3 Conditions initiales

Une dynamique moléculaire démarre d'une configurations initiale. Il est nécessaire de déterminer les positions et vitesses de cette configuration. Pour les positions, la méthode utilisée dépend de la complexité du système. Pour les solides, il est d'usage d'utiliser les données expérimentales. Pour les liquides, nous avons utilisé dans cette thèse le code PACKMOL [77] qui génère des positions aléatoires mais telles que les distances entre atomes ne soient pas trop faibles, empêchant d'avoir de fortes répulsions.

Pour les vitesses, elles sont choisies pour que leur distribution corresponde à une distribution de Maxwell-Boltzmann correspondant à la température désirée.

I.2.4 Ensemble canonique - Thermostats

Jusqu'à maintenant, l'ensemble thermodynamique considéré était l'ensemble micro-canonique. En général, les expériences ne sont pas réalisées à énergie constante mais à température (ou pression) constante. Pour réaliser une dynamique moléculaire dans l'ensemble canonique, il est alors nécessaire de simuler l'effet d'un thermostat. Le but étant de générer une trajectoire dont les microétats sont en accord avec une distribution canonique, c'est à dire d'échantillonner l'espace des phases canonique.

Thermostat de Langevin

Une première méthode repose sur l'équation du mouvement de Langevin décrivant l'interaction d'une particule I avec un solvant à l'aide d'une force de friction $\gamma\vec{p}_I$ et d'une force aléatoire $\sigma\vec{\xi}_I$, de distribution gaussienne tel que $\langle \xi_I \rangle = 0$ et

$\langle \xi_I(0)\xi_I(t) \rangle = \delta(t)$, modélisant les collisions entre les particules et le solvant :

$$\frac{d\vec{p}_I}{dt} = -\nabla_I V(\vec{r}_I) - \gamma\vec{p}_I + \sigma\vec{\xi}_I(t) \quad \text{avec : } \sigma = 2M_I kT\gamma \quad (\text{I.98})$$

où k est la constante de Boltzmann et T la température.

Le solvant joue alors le rôle d'un thermostat, les contributions des forces de friction et aléatoires permettent de simuler l'échange d'énergie avec un bain thermique environnant.

Chaînes de Nose-Hoover

Les thermostats dit "Chaînes de Nose-Hoover" reposent sur les thermostats de Nose-Hoover [83, 61]. On ajoute au système des variables supplémentaires, η et p_η , servant à simuler un bain de chaleur. L'Hamiltonien \mathcal{H}_{NH} associé est alors :

$$\mathcal{H}(\vec{R}, \eta, \vec{p}, p_\eta)_{NH} = \mathcal{H}(\vec{R}, \vec{p}) + \frac{p_\eta^2}{2Q} + dNkT\eta \quad (\text{I.99})$$

où \mathcal{H} est l'hamiltonien du système, d la dimension de l'espace, N le nombre d'atomes, k la constante de Boltzmann, T la température et Q un paramètre. Les équations du mouvement sont alors [76] :

$$\dot{\vec{R}}_I = \frac{\vec{p}_I}{M_I} \quad (\text{I.100})$$

$$\dot{\vec{p}}_I = \vec{F}_I - \frac{p_\eta}{Q}\vec{p}_I \quad (\text{I.101})$$

$$\dot{\eta} = \frac{p_\eta}{Q} \quad (\text{I.102})$$

$$\dot{p}_\eta = \sum_{I=1}^M \frac{\vec{p}_I}{M_I} - dNkT \quad (\text{I.103})$$

On voit apparaître un terme de friction dans l'équation d'évolution des moments conjugués servant à réguler la température. Dans certains cas, la dynamique définie par le système d'équations I.103 ne permet pas d'obtenir la distribution canonique recherchée. On ajoute alors d'autres variables sous forme de chaîne, c'est à dire qu'un thermostat de Nose-Hoover est attaché à p_η . On a alors deux nouvelles variables η_2 et p_{η_2} et on peut attacher un nouveau thermostat à p_{η_2} et ainsi de suite jusqu'à obtenir N_t thermostats de Nose-Hoover. De plus, en pratique, pour obtenir une équilibration rapide, on attache à chaque atome I une chaîne de Nose-Hoover. Finalement, on

obtient les équations du mouvement [76] :

$$\dot{\vec{R}}_I = \frac{\vec{p}_I}{M_I} \quad (\text{I.104})$$

$$\dot{\vec{P}}_I = \vec{F}_I - \frac{\vec{p}_{\eta 1, I}}{Q_1} \vec{p}_I \quad (\text{I.105})$$

$$\dot{\eta}_{j, I} = \frac{p_{\eta j, I}}{Q_j} \quad j = 1, \dots, N_t \quad (\text{I.106})$$

$$\dot{p}_{\eta 1, I} = \left[\frac{\vec{p}_I^2}{M_I} - dkT \right] - \frac{p_{\eta 2, I}}{Q_2} p_{\eta 1, I} \quad (\text{I.107})$$

$$\dot{p}_{\eta j, I} = \left[\frac{p_{\eta j-1, I}^2}{Q_{j-1}} - kT \right] - \frac{p_{\eta j+1, I}}{Q_{j+1}} p_{\eta j, I} \quad j = 2, \dots, N_t - 1 \quad (\text{I.108})$$

$$\dot{p}_{\eta N_t, I} = \left[\frac{p_{\eta N_t-1, I}^2}{Q_{N_t-1}} - kT \right] \quad (\text{I.109})$$

Typiquement, dans une simulation, on utilise un nombre de thermostats de Nose-Hoover dans une chaîne de l'ordre de quelques unités. Pour les paramètres Q_1, \dots, Q_{N_t} [76] propose de les choisir tel que :

$$Q_1 = dNkT\tau^2 \quad (\text{I.110})$$

$$Q_j = kT\tau^2 \quad j \neq 1 \quad (\text{I.111})$$

où τ est un temps caractéristique du système, que l'on choisit en général de l'ordre de quelques dizaines de pas de temps Δt .

I.2.5 Différents types de dynamiques moléculaires

Un élément important de la dynamique moléculaire est le calcul des forces agissant sur les atomes qui régissent leur mouvement. Selon la méthode utilisée pour la détermination des potentiels interatomiques, et donc des forces, on distingue différents types de dynamique moléculaire.

Dynamique moléculaire classique et machine learning

Dans le cas de la dynamique moléculaire classique ou la dynamique moléculaire de type machine learning (MLMD), le principe est le même. A chaque pas de temps de la dynamique, l'énergie est déterminée à partir de la fonction de l'énergie, correspondant aux potentiels empiriques ou machine learning. Les forces sont ensuite calculées et les positions et vitesses mises à jour.

Dynamique moléculaire ab initio (AIMD) :

Les dynamiques moléculaires ab initio correspondent à l'ensemble des dynamiques moléculaires dont le calcul des forces se fait à partir de la résolution de l'équation de Schrödinger. Pour cela, on utilise le théorème de Hellmann-Feynman [39, 54]. Soit un hamiltonien H_λ dépendant d'un paramètre continu λ , ϕ_λ une fonction d'onde propre normée qui dépend implicitement de λ et $E_\lambda = \langle \phi_\lambda | H_\lambda | \phi_\lambda \rangle$ l'énergie propre,

alors :

$$\frac{dE_\lambda}{d\lambda} = \frac{d}{d\lambda} (\langle \phi_\lambda | H_\lambda | \phi_\lambda \rangle) \quad (\text{I.112})$$

$$= \left\langle \frac{d\phi_\lambda}{d\lambda} \middle| H_\lambda | \phi_\lambda \right\rangle + \langle \phi_\lambda | H_\lambda \left| \frac{d\phi_\lambda}{d\lambda} \right\rangle + \langle \phi_\lambda | \frac{dH_\lambda}{d\lambda} | \phi_\lambda \rangle \quad (\text{I.113})$$

$$= E_\lambda \left\langle \frac{d\phi_\lambda}{d\lambda} \middle| \phi_\lambda \right\rangle + E_\lambda \left\langle \phi_\lambda \middle| \frac{d\phi_\lambda}{d\lambda} \right\rangle + \langle \phi_\lambda | \frac{dH_\lambda}{d\lambda} | \phi_\lambda \rangle \quad (\text{I.114})$$

$$= E_\lambda \frac{d}{d\lambda} \langle \phi_\lambda | \phi_\lambda \rangle + \langle \phi_\lambda | \frac{dH_\lambda}{d\lambda} | \phi_\lambda \rangle \quad (\text{I.115})$$

$$= \langle \phi_\lambda | \frac{dH_\lambda}{d\lambda} | \phi_\lambda \rangle \quad (\text{I.116})$$

Si maintenant λ est une position d'un noyau I , \vec{R}_I , et que l'on reprend les notations de la partie I.1.1.3 et I.1.1.4 :

$$\begin{aligned} \vec{F}_I &= -\vec{\nabla}_I U = -\vec{\nabla}_I \langle \phi | (T_e + V_{ee} + V_{ne}) | \phi \rangle - \vec{\nabla}_I \left(\frac{1}{2} \sum_{I=1}^M \sum_{\substack{J=1 \\ J \neq I}}^M \frac{e^2}{4\pi\epsilon_0} \frac{Z_I Z_J}{|\vec{R}_I - \vec{R}_J|} \right) \quad (\text{I.117}) \\ &= -\langle \phi | \vec{\nabla}_I (T_e + V_{ee} + V_{ne}) | \phi \rangle - \vec{\nabla}_I \left(\frac{1}{2} \sum_{I=1}^M \sum_{\substack{J=1 \\ J \neq I}}^M \frac{e^2}{4\pi\epsilon_0} \frac{Z_I Z_J}{|\vec{R}_I - \vec{R}_J|} \right) \end{aligned}$$

or T_e et V_{ee} ne dépendent pas des positions atomiques. On a alors :

$$\vec{F}_I = -\langle \phi | \vec{\nabla}_I V_{ne} | \phi \rangle - \vec{\nabla}_I \left(\frac{1}{2} \sum_{I=1}^M \sum_{\substack{J=1 \\ J \neq I}}^M \frac{e^2}{4\pi\epsilon_0} \frac{Z_I Z_J}{|\vec{R}_I - \vec{R}_J|} \right) \quad (\text{I.118})$$

$$= -\int n(\vec{r}) \vec{\nabla}_I v_{ne} d\vec{r} - \vec{\nabla}_I \left(\frac{1}{2} \sum_{I=1}^M \sum_{\substack{J=1 \\ J \neq I}}^M \frac{e^2}{4\pi\epsilon_0} \frac{Z_I Z_J}{|\vec{R}_I - \vec{R}_J|} \right) \quad (\text{I.119})$$

Ainsi, les forces peuvent être calculées connaissant la densité électronique. En pratique, la base choisie n'est pas nécessairement composée des fonctions propres de l'hamiltonien du système et donc les deux premiers termes de l'équation I.114 ne se simplifient pas. Une méthode consiste à choisir une base dont les fonctions ne dépendent pas des positions des noyaux (ou dont les dérivées par rapport aux positions des noyaux font elles mêmes parties de la base). C'est le cas des bases d'ondes planes et l'équation I.119 peut être utilisée.

- AIMD Bohr-Oppenheimer (BOMD) :

Lors d'une dynamique moléculaire ab initio dite "Bohr-Oppenheimer", à chaque pas de temps, un calcul auto cohérent est effectué à partir duquel on obtient l'énergie de l'état fondamental du système. Les forces sont ensuite calculées et les équations du mouvement résolues et seule la dynamique des noyaux est prise en compte. Cette méthode est très répandue mais a le désavantage d'être lourde en temps de calcul puisque le calcul auto cohérent doit être effectué un grand nombre de fois.

- AIMD Car-Parinello (CPMD) :

Pour répondre à ce problème, Car et Parinello [20], ont développé une méthode qui consiste à prendre en compte à la fois la dynamique des noyaux et des électrons. Seule la dynamique des noyaux a un sens physique et les forces agissant sur ceux-ci sont calculées de la même manière que pour la BOMD. Les électrons évoluent selon une dynamique électronique 'fictive' adiabatique de telle sorte que les orbitales électroniques restent proches de l'état fondamental. Car et Parinello proposent alors le Lagrangien suivant :

$$\mathcal{L}_{CPMD} = \sum_{I=1}^M \frac{M_I \vec{R}_I^2}{2} + \sum_i \frac{\mu_i \langle \dot{\phi}_i | \dot{\phi}_i \rangle}{2} - E[n, \{\vec{R}_I\}] + \sum_{ij} \Lambda_{ij} (\langle \phi_i | \phi_j \rangle - \delta_{ij}) \quad (\text{I.120})$$

où $E[n, \{\vec{R}_I\}] = E[n] + V_{nn}$ définis dans les équations I.3 et I.43, les ϕ_i sont les orbitales électroniques contraintes orthogonales à l'aide du dernier terme de l'équation et les μ_i les "masses fictives" associées aux orbitales. Les équations de Lagrange nous donnent les équations du mouvement :

$$\frac{d}{dt} \frac{\partial \mathcal{L}_{CPMD}}{\partial \dot{\vec{R}}_I} = \frac{\partial \mathcal{L}_{CPMD}}{\partial \vec{R}_I} \iff M_i \ddot{\vec{R}}_I = -\nabla_I E[n, \{\vec{R}_I\}] \quad (\text{I.121})$$

$$\frac{d}{dt} \frac{\partial \mathcal{L}_{CPMD}}{\partial \dot{\phi}_i^*} = \frac{\partial \mathcal{L}_{CPMD}}{\partial \phi_i^*} \iff \mu_i \ddot{\phi}_i = -\frac{\delta E[n, \{\vec{R}_I\}]}{\delta \phi_i^*} + \sum_j \Lambda_{ij} \phi_j \quad (\text{I.122})$$

Les masses fictives μ_i sont des paramètres de la simulation choisies pour que la dynamique reste adiabatique, les électrons doivent rester proche de l'état fondamental et ne doivent pas échanger d'énergie avec les ions. Pour cela, les fréquences associées aux ions et aux électrons ne doivent pas se superposer. Un ordre de grandeur de la valeur de la fréquence électronique la plus basse est donnée par [85] :

$$\omega_e^{min} \approx \left(\frac{\Delta E_g}{\mu_i} \right)^{\frac{1}{2}} \quad (\text{I.123})$$

où ΔE_g est l'énergie de gap. On doit alors respecter la condition $\omega_e^{min} \gg \omega_n^{max}$ avec ω_n^{max} la fréquence maximale associée aux noyaux. Le moyen le plus simple est alors de prendre une masse fictive faible, mais en contrepartie, cela implique que la fréquence maximale associée aux électrons augmente. Elle est donnée, dans le cas de l'utilisation d'une base d'ondes planes :

$$\omega_e^{max} \approx \left(\frac{E_{cut}}{\mu_i} \right)^{\frac{1}{2}} \quad (\text{I.124})$$

Il est alors nécessaire d'utiliser un pas de temps de simulation Δt plus faible permettant de décrire le mouvement des électrons :

$$\Delta t < \frac{1}{\omega_e^{max}} = \left(\frac{\mu_i}{E_{cut}} \right)^{\frac{1}{2}} \quad (\text{I.125})$$

Il faut donc choisir un jeu de paramètres Δt et μ_i de sorte à conserver l'adiabaticité mais avec un pas de temps le plus grand possible pour réduire le temps de calcul.

I.3 Dynamique moléculaire d'intégrales de chemin

Les simulations de dynamique moléculaire reposent sur un traitement classique des noyaux justifié par leur masse très grande par rapport à celle des électrons. Néanmoins, cette description ne permet pas de rendre compte d'effets quantiques des noyaux apparaissant pour des noyaux légers comme l'effet tunnel et la liaison hydrogène, la délocalisation, la fluctuation du point zero, et pour le fractionnement isotopique.

Nous développons ici la méthode de dynamique moléculaire d'intégrales de chemin permettant de prendre en compte les effets quantiques des noyaux, nous verrons aussi partie II.3 qu'elle permet de s'affranchir de l'approximation harmonique lors du calcul du fractionnement isotopique.

I.3.1 Fonction de partition canonique quantique

Oublions pour le moment les simulations numériques et la dynamique moléculaire. Nous souhaitons avoir accès aux grandeurs thermodynamiques d'un système quantique composé de M atomes, de positions atomiques $\{\vec{R}_I\}_{I=1,\dots,M}$. On veut alors connaître la fonction de partition Z du système à partir de laquelle nous pourrions calculer ces grandeurs. On se place dans l'ensemble canonique, pour un système quantique d'hamiltonien H , la fonction de partition est alors donnée par :

$$Z = \text{Tr} \left(e^{-\beta H} \right), \quad \text{avec } \beta = \frac{1}{kT} \quad \text{et } H = T + V = \frac{p^2}{2M} + V(\vec{R}) \quad (\text{I.126})$$

ce qui donne dans l'espace des positions :

$$Z = \int d\vec{R} \langle \vec{R} | e^{-\beta H} | \vec{R} \rangle \quad (\text{I.127})$$

$$= \int d\vec{R} \langle \vec{R} | e^{-\beta(T+V)} | \vec{R} \rangle \quad (\text{I.128})$$

Les opérateurs T et V ne commutent pas, on utilise alors la formule de Trotter [114] :

$$e^{-\beta(T+V)} = \lim_{P \rightarrow +\infty} \left(e^{-\frac{\beta}{P}V} e^{-\frac{\beta}{P}T} \right)^P \quad (\text{I.129})$$

On obtient alors :

$$Z = \lim_{P \rightarrow +\infty} \int d\vec{R} \langle \vec{R} | \left(e^{-\frac{\beta}{P}V} e^{-\frac{\beta}{P}T} \right)^P | \vec{R} \rangle \quad (\text{I.130})$$

$$= \lim_{P \rightarrow +\infty} \int d\vec{R} \langle \vec{R} | \underbrace{\left(e^{-\frac{\beta}{P}V} e^{-\frac{\beta}{P}T} \right) \dots \left(e^{-\frac{\beta}{P}V} e^{-\frac{\beta}{P}T} \right)}_{P \text{ fois}} | \vec{R} \rangle \quad (\text{I.131})$$

En introduisant $P - 1$ relation de fermeture $I_d = \int d\vec{R}^{(i)} |\vec{R}^{(i)}\rangle \langle \vec{R}^{(i)}|$ pour $i = 2, \dots, P$ entre chaque terme entre parenthèses et en notant $\vec{R} = \vec{R}^{(1)}$, on obtient :

$$Z = \lim_{P \rightarrow +\infty} \int d\vec{R}^{(1)} \dots d\vec{R}^{(P)} \langle \vec{R}^{(1)} | \left(e^{-\frac{\beta}{P}V} e^{-\frac{\beta}{P}T} \right) | \vec{R}^{(1)} \rangle \dots \langle \vec{R}^{(P)} | \left(e^{-\frac{\beta}{P}V} e^{-\frac{\beta}{P}T} \right) | \vec{R}^{(P)} \rangle \quad (\text{I.132})$$

$$= \lim_{P \rightarrow +\infty} \int \prod_{i=1}^P d\vec{R}^{(i)} \left[\prod_{i=1}^P d\vec{R}^{(i)} \langle \vec{R}^{(i)} | \left(e^{-\frac{\beta}{P}V} e^{-\frac{\beta}{P}T} \right) | \vec{R}^{(i)} \rangle \right]_{\vec{R}^{(1)} = \vec{R}^{(P+1)}} \quad (\text{I.133})$$

On cherche maintenant à évaluer le terme :

$$\langle \vec{R}^{(i)} | \left(e^{-\frac{\beta}{P}V} e^{-\frac{\beta}{P}T} \right) | \vec{R}^{(i)} \rangle = e^{-\frac{\beta}{P}V(\vec{R}^{(i)})} \langle \vec{R}^{(i)} | e^{-\frac{\beta}{P}T} | \vec{R}^{(i)} \rangle \quad (\text{I.134})$$

$$= e^{-\frac{\beta}{P}V(\vec{R}^{(i)})} \langle \vec{R}^{(i)} | e^{-\frac{\beta p^2}{2MP}} | \vec{R}^{(i)} \rangle \quad (\text{I.135})$$

On introduit ensuite la relation de fermeture $I_d = \int d\vec{p} |\vec{p}\rangle \langle \vec{p}|$ et on obtient (pour le calcul se référer à [116]) :

$$\langle \vec{R}^{(i)} | \left(e^{-\frac{\beta}{P}V} e^{-\frac{\beta}{P}T} \right) | \vec{R}^{(i)} \rangle = \left(\frac{MP}{2\pi\beta\hbar^2} \right)^{\frac{3}{2}} e^{-\frac{\beta}{P}V(\vec{R}^{(i)})} e^{-\frac{MP}{2\beta\hbar^2} (\vec{R}^{(i+1)} - \vec{R}^{(i)})^2} \quad (\text{I.136})$$

En utilisant la relation précédente dans I.133 on obtient finalement :

$$Z = \lim_{P \rightarrow +\infty} \left(\frac{MP}{2\pi\beta\hbar^2} \right)^{\frac{3P}{2}} \int \prod_{i=1}^P d\vec{R}^{(i)} e^{\left[-\sum_{i=1}^P \frac{MP}{2\beta\hbar^2} (\vec{R}^{(i+1)} - \vec{R}^{(i)})^2 + \frac{\beta}{P}V(\vec{R}^{(i)}) \right]_{\vec{R}^{(1)} = \vec{R}^{(P+1)}}} \quad (\text{I.137})$$

Pour un système de M particules, pour des températures où la statistique de Boltzmann est vérifiée, les particules sont discernables et la fonction de partition du système est le produit des fonctions de partition à une particule. Dans ce cas, on obtient :

$$Z = \lim_{P \rightarrow +\infty} \prod_{I=1}^M \left(\frac{M_I P}{2\pi\beta\hbar^2} \right)^{\frac{3P}{2}} \int \prod_{I=1}^M \prod_{i=1}^P d\vec{R}_I^{(i)} e^{\left[-\sum_{i=1}^P \left(\sum_{I=1}^M \frac{M_I P}{2\beta\hbar^2} (\vec{R}_I^{(i+1)} - \vec{R}_I^{(i)})^2 + \frac{\beta}{P}V(\{\vec{R}_I^{(i)}\}) \right) \right]_{\vec{R}_I^{(1)} = \vec{R}_I^{(P+1)}}} \quad (\text{I.138})$$

Pour une discussion sur le lien avec la formulation en terme d'intégrale de chemin en mécanique quantique, le lecteur pourra se référer à [116].

Maintenant que nous avons l'expression de la fonction de partition, nous souhaitons trouver un moyen permettant de l'échantillonner à l'aide de simulations numériques.

I.3.2 Echantillonnage et isomorphisme

Nous allons voir qu'il est possible d'échantillonner la fonction de partitions d'un système quantique grâce à des simulations numériques en faisant une analogie entre l'expression I.138 de la fonction de partition avec celle d'un système classique. Il est possible d'utiliser des méthodes de type Monte-Carlo ou de dynamiques moléculaires.

Nous développons ici la deuxième méthode. Pour rendre la discussion plus claire, commençons par remplacer les MP termes du préfacteur de l'expression de la fonction de partition I.138 par des exponentielles gaussiennes tel que :

$$\left(\frac{M_I P}{2\pi\beta\hbar^2}\right)^{\frac{3}{2}} = \int d\vec{p}_I^{(i)} e^{-\frac{\beta}{P} \frac{\vec{p}_I^{(i)2}}{2M_I^{(i)'}}} \quad (\text{I.139})$$

La fonction de partition se réécrit :

$$Z = \lim_{P \rightarrow +\infty} \int \prod_{I=1}^M \prod_{i=1}^P d\vec{R}_I^{(i)} d\vec{p}_I^{(i)} e^{-\beta_P \left[\sum_{i=1}^P \left(\sum_{I=1}^M \frac{\vec{p}_I^{(i)2}}{2M_I^{(i)'}} + \frac{1}{2} M_i \omega_P^2 \left(\vec{R}_I^{(i+1)} - \vec{R}_I^{(i)} \right)^2 + V(\{\vec{R}_I^{(i)}\}) \right) \right]}_{\vec{R}_I^{(1)} = \vec{R}_I^{(P+1)}} \quad (\text{I.140})$$

avec $\omega_P = \frac{P}{\beta\hbar}$ et où on a introduit $\beta_P = \frac{\beta}{P}$. En toute rigueur, $M_I^{(i)'} = \frac{M_I}{(2\pi\hbar)^2}$, mais sa valeur change seulement le préfacteur de la fonction de partition, on peut donc choisir d'autres valeurs de $M_I^{(i)'}$ sans changer les propriétés thermodynamiques dérivées à partir de la fonction de partition.

Exprimée comme ceci, la fonction de partition du système quantique est celle d'un système classique d'hamiltonien :

$$\mathcal{H}_{cl} = \sum_{i=1}^P \left(\sum_{I=1}^M \frac{\vec{p}_I^{(i)2}}{2M_I^{(i)'}} + \frac{1}{2} M_i \omega_P^2 \left(\vec{R}_I^{(i+1)} - \vec{R}_I^{(i)} \right)^2 + V(\{\vec{R}_I^{(i)}\}) \right) \Big|_{\vec{R}_I^{(1)} = \vec{R}_I^{(P+1)}} \quad (\text{I.141})$$

où chaque particule du système quantique a été dupliquée en P "beads", formant une chaîne cyclique. Chaque bead est soumis à un potentiel harmonique avec ses beads voisins, et au potentiel V entre lui et les atomes de mêmes numéro (i) de bead. Il y a donc un isomorphisme entre le système quantique et un système classique de chaînes de polymères cycliques [24], voir Fig. I.13. Echantillonner la fonction de partition du système classique, à la température $P \times T$, nous permet alors aussi, lorsque $P \rightarrow +\infty$, d'obtenir la fonction de partition du système quantique. Il est possible d'utiliser les méthodes de dynamique moléculaire en résolvant les équations du mouvement :

$$\dot{\vec{R}}_I^{(i)} = \frac{\partial \mathcal{H}_{cl}}{\partial \vec{p}_I^{(i)}}, \quad \dot{\vec{p}}_I^{(i)} = -\frac{\partial \mathcal{H}_{cl}}{\partial \vec{R}_I^{(i)}} \quad (\text{I.142})$$

et d'utiliser des thermostats pour travailler dans l'ensemble canonique. On parle alors de dynamique moléculaire d'intégrales de chemin. Il est important de noter qu'en l'état, la dynamique moléculaire d'intégrales de chemin n'est utilisée que pour échantillonner la fonction de partition du système. Pour pouvoir extraire des propriétés dynamiques, il faut avoir recours aux fonctions de corrélation [52].

Lors de nos simulations, il n'est pas possible d'avoir un nombre infini de beads. On utilise alors un développement du théorème de Trotter I.129. A l'ordre deux, il est donné par :

$$Tr(e^{-\beta(T+V)}) = Tr\left(\left(e^{-\frac{\beta}{P}T} e^{-\frac{\beta}{P}V}\right)^P\right) + \mathcal{O}\left(\frac{\beta^3}{P^2}\right) \quad (\text{I.143})$$

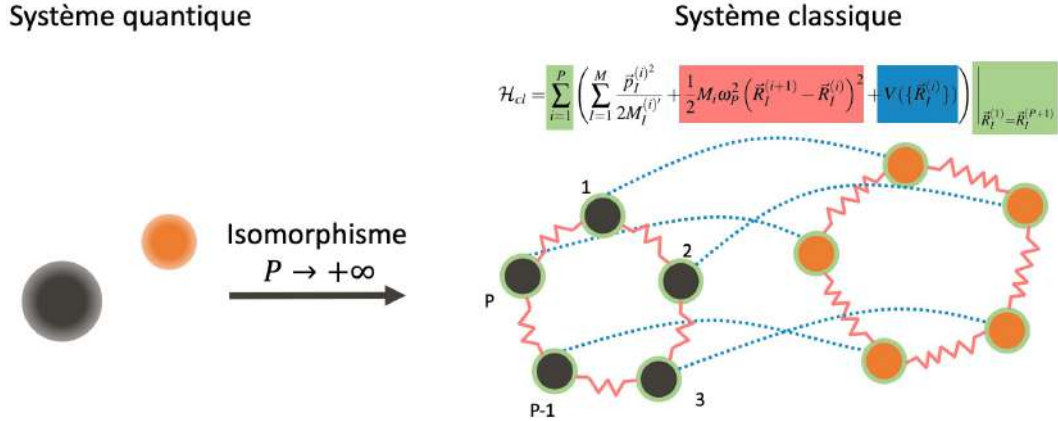


Figure I.13: Isomorphisme entre un système quantique et un système classique de chaînes cycliques de polymère pour deux atomes. Chaque atome est dupliqué en P beads formant une chaîne de polymère représentant la délocalisation spatiale d'une particule quantique.

Les étapes de calcul pour obtenir la fonction de partition, Z_P , à partir de ce développement sont les mêmes qu'à la partie I.3.1, seule la $\lim_{P \rightarrow +\infty}$ disparaît. Et on obtient :

$$Z_P = \prod_{I=1}^M \left(\frac{M_I P}{2\pi\beta\hbar^2} \right)^{\frac{3P}{2}} \int \prod_{I=1}^M \prod_{i=1}^P d\vec{R}_I^{(i)} e^{\left[-\sum_{i=1}^P \left(\sum_{l=1}^M \frac{M_l P}{2\beta\hbar^2} \left(\vec{R}_l^{(i+1)} - \vec{R}_l^{(i)} \right)^2 + \frac{\beta}{P} V(\{\vec{R}_l^{(i)}\}) \right) \right]_{\vec{R}_l^{(1)} = \vec{R}_l^{(P+1)}}} \quad (\text{I.144})$$

et donc :

$$Z = \lim_{P \rightarrow +\infty} Z_P \quad (\text{I.145})$$

En pratique, on effectue des dynamiques moléculaires d'intégrales de chemin en augmentant progressivement le nombre de beads. Lorsque la grandeur d'intérêt ne varie plus, le nombre de beads est suffisant pour caractériser les effets quantiques.

I.3.3 Modes normaux

L'utilisation des équations du mouvement I.142 pour générer un ensemble de configurations de l'espace des phases entraîne des problèmes d'ergodicité [53]. En effet,

réécrivons l'hamiltonien classique I.141 sous la forme d'une somme de deux termes :

$$\mathcal{H}_{cl} = H_{cl}^0 + V_P \quad \text{avec :} \quad \mathcal{H}_{cl}^0 = \sum_{i=1}^P \left(\sum_{I=1}^M \frac{\vec{p}_I^{(i)2}}{2M_I^{(i)'}} + \frac{1}{2} M_i \omega_P^2 \left(\vec{R}_I^{(i+1)} - \vec{R}_I^{(i)} \right)^2 \right) \quad (\text{I.146})$$

$$V_P = \sum_{i=1}^P V(\{\vec{R}_I^{(i)}\}) \quad (\text{I.147})$$

\mathcal{H}_{cl}^0 correspond à l'hamiltonien d'un ensemble d'oscillateurs couplés. On peut réécrire cet hamiltonien dans la représentation des modes normaux [23] :

$$\vec{p}_I^{(i)} = \sum_{j=1}^P C_{ji} \vec{p}_I^{(j)} \quad \text{et} \quad \vec{R}_I^{(i)} = \sum_{j=1}^P C_{ji} \vec{R}_I^{(j)} \quad (\text{I.148})$$

avec, pour P pair :

$$C_{ji} = \begin{cases} \sqrt{1/P} & i = 0 \\ \sqrt{2/P} \cos(2\pi i j / P) & 1 \leq k \leq P/2 - 1 \\ \sqrt{1/P} (-1)^j & k = P/2 \\ \sqrt{2/P} \sin(2\pi i j / P) & P/2 + 1 \leq k \leq P - 1 \end{cases} \quad (\text{I.149})$$

\mathcal{H}_{cl}^0 devient :

$$\sum_{i=0}^{P-1} \left(\sum_{I=1}^M \frac{\vec{p}_I^{(i)2}}{2M_I^{(i)'}} + \frac{1}{2} M_i \omega_i^2 \vec{R}_I^{(i)2} \right) \quad \text{avec} \quad \omega_i = 2\omega_P \sin\left(\frac{i\pi}{P}\right) \quad (\text{I.150})$$

La dynamique comporte donc des modes de vibrations d'autant plus élevés que P augmente ($\omega_P = P/\beta\hbar$) nécessitant alors un pas de temps Δt petit. Il faut donc des simulations très longues pour échantillonner correctement les modes de vibrations de fréquences plus faibles.

Plusieurs méthodes existent pour palier à ce problème, comme l'utilisation de variables "staging" [117]. On développe ici la méthode des modes normaux retenue par le code i-PI [22] et utilisée dans cette thèse.

La réécriture de l'hamiltonien en somme des deux termes H_{cl}^0 et V_P permet de décomposer l'intégration des équations du mouvement en plusieurs étapes [23], l'intégration des termes harmoniques se faisant directement dans la représentation en modes normaux :

1. évolution de $\vec{p}_I^{(i)}$ pour un intervalle de temps $\Delta t/2$ soumis au potentiel V_P :

$$\vec{p}_I^{(i)} = \vec{p}_I^{(i)} - \frac{\Delta t}{2} \frac{\partial V(\{\vec{R}_I^{(i)}\})}{\partial \vec{R}_I^{(i)}} \quad (\text{I.151})$$

2. passage en représentation des modes normaux par transformée de Fourier :

$$\vec{p}_I^{(j)} = \sum_{i=1}^P C_{ji} \vec{p}_I^{(i)}, \quad \vec{R}_I^{(j)} = \sum_{i=1}^P C_{ji} \vec{R}_I^{(i)} \quad (\text{I.152})$$

3. évolution des quantités de mouvement et positions pour l'hamiltonien \mathcal{H}_{cl}^0 :

$$\begin{pmatrix} \tilde{p}_I^{(j)} \\ \tilde{R}_I^{(j)} \end{pmatrix} = \begin{pmatrix} \cos(\omega_j \Delta t) & -M'_I \omega_j \sin(\omega_j \Delta t) \\ (1/M'_I \omega_j) \sin(\omega_j \Delta t) & \cos(\omega_j \Delta t) \end{pmatrix} \begin{pmatrix} \tilde{p}_I^{(j)} \\ \tilde{R}_I^{(j)} \end{pmatrix} \quad (\text{I.153})$$

4. retour à la représentation originale par transformée de Fourier inverse :

$$\tilde{p}_I^{(i)} = \sum_{j=0}^{P-1} C_{ji} \tilde{p}_I^{(j)}, \quad \tilde{R}_I^{(i)} = \sum_{j=0}^{P-1} C_{ji} \tilde{R}_I^{(j)} \quad (\text{I.154})$$

5. évolution de $\tilde{p}_I^{(i)}$ pour un intervalle de temps $\Delta t/2$ soumis au potentiel V_P :

$$\tilde{p}_I^{(i)} = \tilde{p}_I^{(i)} - \frac{\Delta t}{2} \frac{\partial V(\{\tilde{R}_I^{(i)}\})}{\partial \tilde{R}_I^{(i)}} \quad (\text{I.155})$$

I.3.4 Thermostat de Langevin

Pour échantillonner la fonction de partition dans l'ensemble canonique, il est possible d'utiliser des thermostats de type chaîne de Nose-Hoover. Leur implémentation est assez coûteuse en temps de calcul et les auteurs de [23] ont montré que l'utilisation de thermostats de type Langevin permet d'obtenir un échantillonnage aussi efficace à moindre coup.

Cette méthode repose sur l'équation du mouvement de Langevin (voir partie I.2.4). Dans le cas de la PIMD, il est possible, de la même manière que dans la partie précédente, de décomposer l'intégration des équations du mouvement et d'ajouter avant l'étape 1 et après l'étape 5 la dynamique de Langevin dans la représentation des modes normaux :

$$\tilde{p}_I^{(j)} = \sum_{i=1}^P C_{ji} \tilde{p}_I^{(i)} \quad (\text{I.156})$$

$$\tilde{p}_I^{(i)} = e^{\frac{\Delta t}{2} \gamma^{(i)}} + \sqrt{\frac{M'_I}{\beta_P} (1 - e^{\Delta t \gamma^{(i)}})} \xi_I^{(i)} \quad (\text{I.157})$$

$$\tilde{p}_I^{(i)} = \sum_{j=0}^{P-1} C_{ji} \tilde{p}_I^{(j)} \quad (\text{I.158})$$

où ξ_I^i est de distribution gaussienne tel que $\langle \xi_I^i \rangle = 0$ et $\langle \xi_I^i(0) \xi_I^i(t) \rangle = \delta(t)$. L'avantage de travailler dans la représentation des modes normaux est qu'il est possible d'avoir un coefficient de friction $\gamma^{(i)}$ différent pour chaque mode. Les auteurs de [23] proposent de les choisir tel que $\gamma^{(i)} = 2\omega_i$.

Le problème est que pour le mode $j = 0$, on a alors $\gamma^{(0)} = 0$. Il faut donc fixer sa valeur à l'aide d'un paramètre de simulation à ajuster τ_0 tel que $\gamma^{(0)} = \frac{1}{\tau_0}$. Ce type de thermostat est appelé PILE_L.

L'inconvénient de ce thermostat est qu'il peut être trop "agressif" pour obtenir des propriétés sensibles au mouvement lent et global des atomes. Pour résoudre ce

problème, il est possible d'employer un algorithme "rescaling velocity" [23, 19] pour les modes 0 :

$$\tilde{p}_I^{(0)} = \alpha \tilde{p}_I^{(0)} \quad (\text{I.159})$$

avec :

$$\alpha = c + \frac{(1+c)((\xi_1^{(0)})^2 + \sum_{I=2}^M (\xi_I^{(0)})^2)}{2\beta_P K} + 2\xi_1^{(0)} \sqrt{\frac{c(1-c)}{2\beta_P K}} \quad (\text{I.160})$$

et :

$$c = e^{-\frac{\Delta t}{\tau_0}}, \quad K = \sum_{I=1}^M \frac{(\tilde{p}_I^{(0)})^2}{2M_I} \quad (\text{I.161})$$

Ce type de thermostat est appelé PILE_G.

Le choix du paramètre τ_0 est primordial pour obtenir un échantillonnage correct de l'espace de phase et est à tester. Une valeur trop petite a tendance à donner des valeurs de grandeurs thermodynamiques convergées fausses, bien que l'échantillonnage soit rapide. En revanche, pour une valeur trop grande, l'échantillonnage est long.

Les outils d'analyse, fonction de distribution radiale, fonction de déplacement quadratique moyen et fonction d'autocorrélation des vitesses, des trajectoires de dynamiques moléculaires et dynamiques moléculaires d'intégrales de chemin, utilisés dans les chapitres suivant sont présentés en annexe A.

Fractionnement isotopique à l'équilibre

On présente dans ce chapitre plus en détail la notion de fractionnement isotopique à l'équilibre en commençant par justifier son origine quantique (partie II.1). La partie suivante (II.2) présente le facteur de partage isotopique, grandeur quantifiant le fractionnement et mesurable expérimentalement. On y présente aussi le lien entre ce facteur et la variation d'énergie libre de la réaction associée au fractionnement en y introduisant le facteur β dépendant d'un seul système. C'est ce facteur que l'on cherchera à déterminer par la suite à l'aide de simulations numériques (voir partie II.3).

Table des matières

II.1 Origine quantique du fractionnement	56
II.2 Facteur de partage isotopique et constante d'échange	57
II.3 Calcul du fractionnement isotopique	58
II.3.1 Dynamique moléculaire ab-initio et approximation harmonique	59
II.3.2 Dynamique moléculaire d'intégrales de chemin et estimateurs	61

Commençons par considérer deux phases, A et B, par exemple un solide et un liquide, en contact. Ces deux phases sont composées d'éléments chimiques ayant un isotope léger Y^a et un isotope lourd Y^b (voir Figure II.1). Deux isotopes peuvent s'échanger au sein d'une phase ou entre les deux phases à leur interface. Ces échanges s'effectuent jusqu'à un équilibre thermodynamique que l'on peut modéliser par une équation de réaction à laquelle on associe une constante d'équilibre K_{eq} .



Il est alors possible de mesurer le rapport isotopique, c'est à dire le rapport des densités atomiques, R_A et R_B , de l'isotope Y^a , n^a , et de l'isotope Y^b , n^b , dans une même phase.

$$R_A = \left. \frac{n^b}{n^a} \right|_A, \quad R_B = \left. \frac{n^b}{n^a} \right|_B \tag{II.2}$$

Bien que les isotopes présentent les mêmes propriétés chimiques de part leurs configurations électroniques identiques, l'expérience montre que les rapports isotopiques R_A et R_B , peuvent avoir des valeurs différentes. Il existe donc un mécanisme

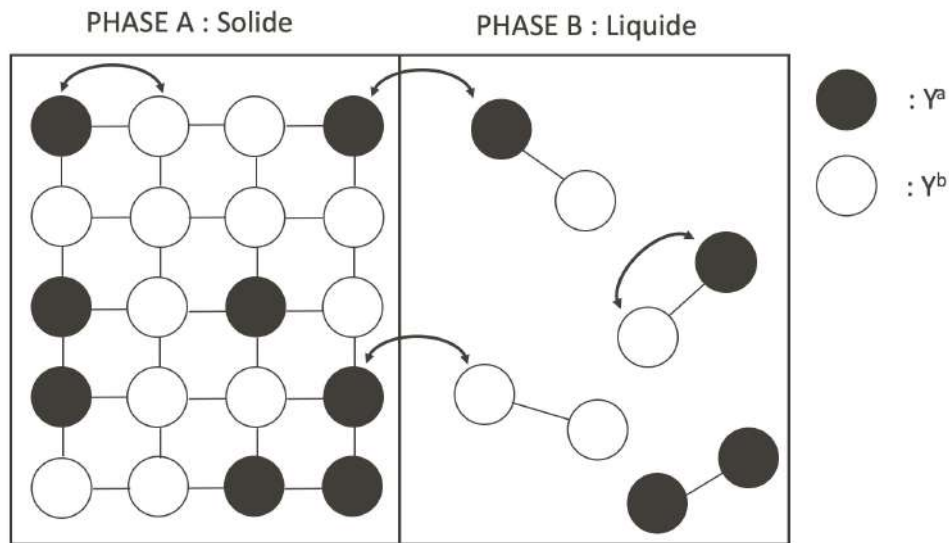


Figure II.1: Exemple d'un solide et d'un liquide en contact composé de deux isotopes Y^a et Y^b . Les échanges d'isotopes sont représentés par une double flèche noire et peuvent se produire au sein d'une même phase ou à l'interface des deux phases.

d'enrichissement d'un isotope dans une phase par rapport à une autre que l'on appelle fractionnement isotopique.

II.1 Origine quantique du fractionnement

Il est possible de comprendre l'origine du fractionnement isotopique à partir d'une modélisation simple. Considérons que la liaison entre un atome d'une phase et un isotope peut être modélisée par un oscillateur harmonique. L'une des deux phases aura une liaison plus rigide que l'autre (voir Figure II.2).

Si l'on effectue une analyse classique, à l'équilibre, pour les deux isotopes, le système est au minimum du puits de potentiel des oscillateurs harmoniques. L'échange de deux isotopes n'entraîne alors pas de variation d'énergie et nous ne devrions pas observer de fractionnement isotopique.

Si nous faisons maintenant une analyse quantique, à l'équilibre, le minimum de l'énergie est défini par :

$$E_0 = \frac{1}{2} \hbar \omega(k, M^i) \quad (\text{II.3})$$

où \hbar est la constante de Planck réduite, ω la pulsation qui dépend de la raideur de la liaison k et de la masse de l'isotope M^i , $i = a, b$. Le minimum de l'énergie dépend alors maintenant de la masse, et un échange isotopique entraîne une variation de l'énergie. L'origine du fractionnement isotopique est donc quantique. Les liaisons plus rigides favorisent les isotopes lourds (voir Figure II.2). Ce dernier point est important, en effet, cela implique donc que les isotopes lourds seront préférentiellement dans un système où la liaison est courte et donc où la coordinence est faible. Cela a été observé expérimentalement et théoriquement, notamment pour le calcium [28, 79], et

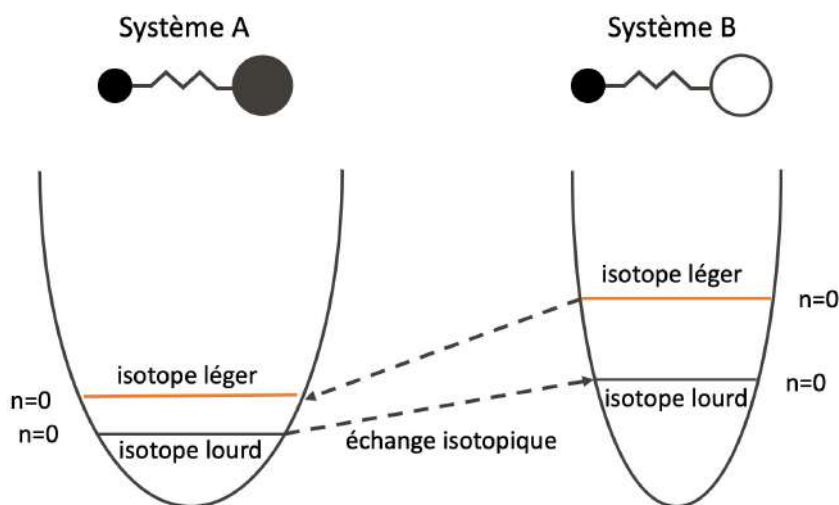


Figure II.2: Echange isotopique entre deux phases dont les interactions sont modélisées par des oscillateurs harmoniques. Le minimum de l'énergie étant différent pour les deux isotopes dans le cadre de la mécanique quantique, les liaisons rigides favorisent les isotopes lourds.

est la raison de l'étude de la coordinence du calcium en solution aqueuse du chapitre 3.

II.2 Facteur de partage isotopique et constante d'échange

Pour quantifier le fractionnement isotopique on définit le facteur de partage isotopique α comme le rapport des rapports isotopiques définis eq. II.2 :

$$\alpha(A, B) = \frac{R_A}{R_B} = \frac{\frac{n^b}{n^a} \Big|_A}{\frac{n^b}{n^a} \Big|_B} \quad (\text{II.4})$$

On peut réécrire le facteur de partage isotopique en fonction des concentrations des espèces isotopiques, et, si l'on considère que les conditions pour lesquelles les activités peuvent être prises égales à la concentration des espèces sont remplies, le facteur de partage isotopique est alors égal à la constante d'équilibre de la réaction II.1 :

$$\alpha(A, B) = \frac{[AY^b][BY^a]}{[AY^a][BY^b]} = K_{eq} \quad (\text{II.5})$$

L'énergie libre de Gibbs de la réaction II.1, ΔG , est reliée à la constante d'équilibre par :

$$\Delta G = -kT \ln K_{eq} \quad (\text{II.6})$$

avec k la constante de Boltzman et T la température du système.

Les réactions d'échanges isotopiques n'ont pas, en général, d'influence sur le volume molaire, de plus, le nombre d'atomes reste inchangé. Finalement, la réaction s'effectue à volume constant et la variation de l'énergie libre de Gibbs est la même que la variation de l'énergie libre ΔF [94]. Cela nous permettra dans la suite de travailler dans l'ensemble canonique. On obtient donc :

$$\alpha(A, B) = e^{-\frac{\Delta F}{kT}} \quad (\text{II.7})$$

Tout l'enjeu est alors de déterminer la variation de l'énergie libre associée à la réaction d'échange isotopique. Ce facteur de partage dépend de deux phases, on préfère alors travailler avec le facteur beta, \mathcal{B} , défini comme étant le facteur de partage isotopique entre une phase et un gaz monoatomique parfait constitué des isotopes Y^a et Y^b dans les proportions naturelles. Pour une phase A, il est donné par :

$$\mathcal{B}(A) = \frac{\left. \frac{n^b}{n^a} \right|_A}{\left. \frac{n^b}{n^a} \right|_g} \quad (\text{II.8})$$

où l'indice g dénote le gaz parfait monoatomique. Le facteur \mathcal{B} est associé à la réaction, de variation d'énergie libre ΔF_A :

$$AY^a + Y_{(g)}^b \rightleftharpoons AY^b + Y_{(g)}^a, \quad K_{A,eq} = \mathcal{B}(A) = e^{-\frac{\Delta F_A}{kT}} \quad (\text{II.9})$$

En faisant de même pour une seconde phase B :

$$BY^a + Y_{(g)}^b \rightleftharpoons BY^b + Y_{(g)}^a, \quad K_{B,eq} = \mathcal{B}(B) = e^{-\frac{\Delta F_B}{kT}} \quad (\text{II.10})$$

On peut déterminer le facteur de partage isotopique α à partir des facteurs \mathcal{B} :

$$\alpha(A, B) = \frac{\mathcal{B}(A)}{\mathcal{B}(B)} \quad (\text{II.11})$$

Au lieu de déterminer le facteur de partage isotopique α dépendant de deux phases, on se retrouve à déterminer deux facteurs \mathcal{B} dépendent cette fois ci d'une seule phase, les termes apparaissant dans la variation de l'énergie libre provenant du gaz parfait étant calculable analytiquement. Par la suite, les simulations atomistiques réalisées seront donc celle d'une phase. On détaille dans la partie suivante comment obtenir le facteur \mathcal{B} à partir de ces simulations.

II.3 Calcul du fractionnement isotopique

Dans cette partie nous exposons les deux méthodes utilisées pour déterminer le facteur de fractionnement \mathcal{B} relié à la variation d'énergie libre associée à l'équation de réaction :



où (S) désigne le système d'étude, (g) le gaz parfait monoatomique et les exposants a et b les isotopes légers et lourds.

Le facteur beta est donné par :

$$\mathcal{B} = e^{-\beta\Delta F} \quad (\text{II.13})$$

L'énergie libre d'un gaz parfait monoatomique pour un atome est :

$$F_{(g)}(Y^i) = -kT \ln \left(\left(\frac{2\pi M^i kT}{h^2} \right)^{\frac{3}{2}} V \right) \quad (\text{II.14})$$

où M^i est la masse de l'isotope, V le volume, k la constante de Boltzmann, T la température et h la constante de Planck.

ΔF se met alors sous la forme :

$$\Delta F = F_{(S)}(Y^a) - F_{(S)}(Y^b) + F_{(g)}(Y^b) - F_{(g)}(Y^a) = \Delta F_{(S)} + \frac{3}{2}kT \ln \left(\frac{M^a}{M^b} \right) \quad (\text{II.15})$$

L'enjeu est alors de déterminer $\Delta F_{(S)}$.

II.3.1 Dynamique moléculaire ab-initio et approximation harmonique

L'énergie libre d'un système est reliée à sa fonction de partition par :

$$F_{(S)}(Y^i) = -kT \ln(Z^{(i)}) \quad (\text{II.16})$$

La fonction de partition peut se décomposer, pour une molécule, en produit de trois contributions [92]. Une contribution correspondant aux rotations, une aux translations et une aux vibrations. Si on considère un solide cristallin, seules les vibrations interviennent dans le calcul du fractionnement isotopique [18]. La fonction de partition est alors la fonction de partition vibrationnelle.

Lorsqu'un atome est à sa position d'équilibre, il est possible d'approximer le potentiel par un oscillateur harmonique autour de cette position (voir Figure. II.3). Par la suite, on considère un système périodique et l'approximation harmonique justifiée. On considère aussi de grands systèmes, leur zone de Brillouin est alors petite. Dans ce cas on peut prendre seulement le point Γ pour échantillonner la zone et la fonction de partition vibrationnelle s'écrit [118] :

$$Z^{(i)} = \prod_{I=1}^{3M} \frac{e^{-\frac{h\nu_I^{(i)}}{2kT}}}{1 - e^{-\frac{h\nu_I^{(i)}}{2kT}}} \quad (\text{II.17})$$

avec $\nu_I^{(i)}$ les modes de vibrations du système.

La variation d'énergie libre du système S est alors :

$$\Delta F_{(S)} = -kT \ln(Z^{(a)}) + kT \ln(Z^{(b)}) = kT \prod_{I=1}^{3M} \frac{e^{-\frac{h\nu_I^b}{2kT}}}{1 - e^{-\frac{h\nu_I^b}{2kT}}} \times \frac{1 - e^{-\frac{h\nu_I^a}{2kT}}}{e^{-\frac{h\nu_I^a}{2kT}}} \quad (\text{II.18})$$

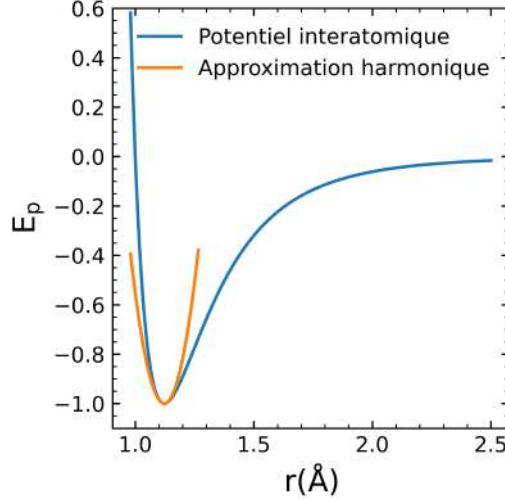


Figure II.3: Approximation du potentiel interatomique E_p autour de la position d'équilibre par un potentiel harmonique.

En utilisant les équations II.13 et II.15 on obtient alors :

$$\mathcal{B} = \left(\prod_{I=1}^{3M} \frac{e^{-\frac{h\nu_I^b}{2kT}}}{1 - e^{-\frac{h\nu_I^b}{2kT}}} \times \frac{1 - e^{-\frac{h\nu_I^a}{2kT}}}{e^{-\frac{h\nu_I^a}{2kT}}} \right) \left(\frac{M^a}{M^b} \right)^{\frac{3}{2}} \quad (\text{II.19})$$

A haute température, $\frac{h\nu_I^{(i)}}{kT} \ll 1$, le système devient classique. Le fractionnement isotopique étant un phénomène quantique, le facteur \mathcal{B} doit tendre vers 1 [118]. Dans ce cas :

$$1 = \left(\prod_{I=1}^{3M} \frac{\nu_I^b}{\nu_I^a} \right) \left(\frac{M^a}{M^b} \right)^{\frac{3}{2}} \quad (\text{II.20})$$

En imposant cette règle du produit haute température, on peut remplacer le rapport des masses dans II.19. Finalement :

$$\mathcal{B} = \left(\prod_{I=1}^{3M} \frac{\nu_I^a}{\nu_I^b} \frac{e^{-\frac{h\nu_I^b}{2kT}}}{1 - e^{-\frac{h\nu_I^b}{2kT}}} \times \frac{1 - e^{-\frac{h\nu_I^a}{2kT}}}{e^{-\frac{h\nu_I^a}{2kT}}} \right) \quad (\text{II.21})$$

Ainsi, pour un solide cristallin, dans le cadre de l'approximation harmonique, le facteur \mathcal{B} peut être déterminé par la seule connaissance des modes de vibrations.

L'approximation harmonique se traduit par un développement de Taylor à l'ordre 2 du potentiel U s'exerçant sur les noyaux (voir equation I.19). Ce développement se fait autour de la position d'équilibre \vec{R}_I^0 . En notant \vec{u}_I le déplacement par rapport à la position d'équilibre, on obtient :

$$U(\{\vec{R}_I\}) = U(\{\vec{R}_I^0\}) + \left. \frac{\partial U}{\partial R_{I\alpha}} \right|_{\{\vec{R}_I\}=\{\vec{R}_I^0\}} u_{I\alpha} + \frac{1}{2} \left. \frac{\partial^2 U}{\partial R_{I\alpha} \partial R_{J\beta}} \right|_{\{\vec{R}_I\}=\{\vec{R}_I^0\}} u_{I\alpha} u_{J\beta} \quad (\text{II.22})$$

$$\text{avec } \vec{R}_I = \vec{R}_I^0 + \vec{u}_I$$

Où on utilise la sommation d'Einstein et $I\alpha$ est la $\alpha = x, y, z$ coordonnées de l'atome I .

Les dérivées sont prises aux positions d'équilibre, ainsi, le second terme du développement, correspondant aux forces, est nul. On définit la matrice de constante des forces K dont les éléments sont donnés par :

$$K_{I\alpha, J\beta} = \frac{\partial^2 U}{\partial R_{I\alpha} \partial R_{J\beta}} \Big|_{\{\vec{R}_I\}=\{\vec{R}_I^0\}} \quad (\text{II.23})$$

Cette matrice est symétrique et on peut alors écrire l'équation du mouvement pour les noyaux dans l'approximation harmonique :

$$M_I \ddot{u}_{I\alpha} = -K_{I\alpha, J\beta} u_{J\beta} \quad (\text{II.24})$$

$$\iff M_I^{\frac{1}{2}} \ddot{u}_{I\alpha} = -\frac{K_{I\alpha, J\beta}}{(M_I M_J)^{\frac{1}{2}}} M_J^{\frac{1}{2}} u_{J\beta} \quad (\text{II.25})$$

$$\iff \ddot{v}_{I\alpha} = -D_{I\alpha, J\beta} v_{J\beta} \quad \text{avec} \quad v_{I\alpha} = M_I^{\frac{1}{2}} u_{I\alpha} \quad \text{et} \quad D_{I\alpha, J\beta} = \frac{K_{I\alpha, J\beta}}{(M_I M_J)^{\frac{1}{2}}} \quad (\text{II.26})$$

La matrice D est appelée matrice dynamique. L'équation du mouvement décrit alors le mouvement de M oscillateurs couplés, et la diagonalisation de la matrice dynamique permet d'obtenir les valeurs propres qui sont les pulsations des modes de vibrations du système et ainsi de calculer le facteur \mathcal{B} à l'aide de la formule II.21. En pratique, le calcul de la matrice dynamique peut se faire à l'aide de la théorie de la fonctionnelle de densité perturbative [9].

Dans le cas où plusieurs points de la zone de Brillouin sont utilisés, il est possible de dériver les mêmes équations avec une matrice dynamique associée à chaque vecteur d'onde de l'espace réciproque. Dans cette thèse, tous les calculs ont été faits au point Γ mais le lecteur intéressé peut se référer à [48, 81].

Dans le cas où le système est anharmonique, c'est à dire qu'il n'est pas possible d'approximer le potentiel autour de la position d'équilibre par un oscillateur harmonique, il est nécessaire d'avoir recours à d'autres méthodes. Les simulations de dynamique moléculaire d'intégrales de chemin permettent de calculer la variation de l'énergie libre par intégration thermodynamique.

II.3.2 Dynamique moléculaire d'intégrales de chemin et estimateurs

Dans le cadre de la dynamique moléculaire d'intégrales de chemin, pour calculer la variation d'énergie libre $\Delta F_{(S)}$ on utilise l'intégration thermodynamique.

Intégration thermodynamique

$$\Delta F_{(S)} = \int_0^1 \frac{dF}{d\lambda} d\lambda \quad (\text{II.27})$$

où λ est un paramètre faisant varier le système entre un état initial et final. Dans notre cas, c'est un paramètre faisant évoluer la masse d'une particule entre un état

initial avec l'isotope léger Y^a à un état final avec l'isotope lourd Y^b .

Pour calculer des grandeurs thermodynamiques on est donc amené à calculer des grandeurs de type $\frac{dF}{d\lambda} = -\frac{kT}{Z} \frac{dZ}{d\lambda}$. En reprenant les notations de la partie I.3, la fonction de partition pour une particule de masse $M(\lambda)$, dans le cadre de la dynamique moléculaire d'intégrales de chemin, est donnée par :

$$Z = \left(\frac{M(\lambda)P}{2\pi\beta\hbar^2} \right)^{\frac{3P}{2}} \int \prod_{i=1}^P d\vec{R}^{(i)} e^{-\beta V_{eff}^{(2)}}, \quad \text{avec} \quad V_{eff}^{(2)} = \frac{M(\lambda)P}{2\beta^2\hbar^2} \sum_{I=1}^P \left(\vec{R}^{(i+1)} - \vec{R}^{(i)} \right)^2 + \sum_{I=1}^P \frac{1}{P} V \quad (\text{II.28})$$

où l'exposant (2) indique que l'on considère le développement de la formule de Trotter à l'ordre 2 (I.143). A partir de cette expression, on peut construire des estimateurs, à l'ordre 2, de grandeurs thermodynamiques.

Estimateur primitif à l'ordre 2 $A^{(2)}$

On cherche :

$$A^{(2)} = \frac{dF}{d\lambda} = -\frac{1}{\beta Z} \frac{dZ}{d\lambda} \quad (\text{II.29})$$

$$= -\frac{1}{\beta Z} \left[\frac{3P}{2} \frac{1}{M(\lambda)} \frac{dM(\lambda)}{d\lambda} Z + \left(\frac{M(\lambda)P}{2\pi\beta\hbar^2} \right)^{\frac{3P}{2}} \int \prod_{i=1}^P d\vec{R}^{(i)} \left(-\beta \frac{\partial V_{eff}^{(2)}}{\partial \lambda} \right) e^{-\beta V_{eff}^{(2)}} \right] \quad (\text{II.30})$$

$$= -\frac{3P}{2\beta} \frac{1}{M(\lambda)} \frac{dM(\lambda)}{d\lambda} + \left\langle \frac{\partial V_{eff}^{(2)}}{\partial \lambda} \right\rangle \quad (\text{II.31})$$

or :

$$\frac{\partial V_{eff}^{(2)}}{\partial \lambda} = \frac{dM(\lambda)}{d\lambda} \frac{P}{2\beta^2\hbar^2} \sum_{i=1}^P \left(\vec{R}^{(i+1)} - \vec{R}^{(i)} \right)^2 \quad (\text{II.32})$$

Finalement :

$$A^{(2)} = -\frac{1}{M(\lambda)} \frac{dM(\lambda)}{d\lambda} \left\langle \frac{3P}{2\beta} - \frac{M(\lambda)P}{2\beta^2\hbar^2} \sum_{i=1}^P \left(\vec{R}^{(i+1)} - \vec{R}^{(i)} \right)^2 \right\rangle \quad (\text{II.33})$$

Estimateur de viriel à l'ordre 2 $A_{vir}^{(2)}$

On peut montrer que la moyenne d'ensemble apparaissant dans l'équation II.33 est l'estimateur à l'ordre 2 de l'énergie cinétique $T^{(2)}$ [116]. L'énergie cinétique croit donc linéairement avec P . Ainsi, lorsque P augmente, les termes harmoniques peuvent avoir des fluctuations rapides rendant la convergence de l'estimateur primitif compliquée. On utilise alors le théorème du Viriel et montre que [56] :

$$T^{(2)} = \left\langle \frac{3P}{2\beta} - \frac{M(\lambda)P}{2\beta^2\hbar^2} \sum_{i=1}^P \left(R^{(i+1)} - R^{(i)} \right)^2 \right\rangle = \left\langle \frac{1}{2P} \sum_{i=1}^P \vec{R}^{(i)} \cdot \frac{\partial V}{\partial \vec{R}^{(i)}} \right\rangle \quad (\text{II.34})$$

Ainsi, l'estimateur de viriel à l'ordre 2 est :

$$A_{vir}^{(2)} = -\frac{1}{M(\lambda)} \frac{dM(\lambda)}{d\lambda} \left\langle \frac{1}{2P} \sum_{i=1}^P \vec{R}^{(i)} \cdot \frac{\partial V}{\partial \vec{R}^{(i)}} \right\rangle \quad (\text{II.35})$$

Ecrit de cette manière, l'estimateur n'est pas invariant par translation, on peut le réécrire :

$$A_{vir}^{(2)} = -\frac{1}{M(\lambda)} \frac{dM(\lambda)}{d\lambda} \left\langle \frac{1}{2P} \sum_{i=1}^P (\vec{R}^{(i)} - \vec{R}_c) \cdot \frac{\partial V}{\partial \vec{R}^{(i)}} + \frac{1}{2P} \sum_{i=1}^P \vec{R}_c \cdot \frac{\partial V}{\partial \vec{R}^{(i)}} \right\rangle \quad (\text{II.36})$$

$$\text{avec : } \vec{R}_c = \frac{1}{P} \sum_{i=1}^P \vec{R}^{(i)} \text{ la position du centroïde} \quad (\text{II.37})$$

Le dernier terme correspond à l'énergie cinétique du centroïde, égale à $\frac{3}{2\beta}$, finalement :

$$A_{vir}^{(2)} = -\frac{1}{M(\lambda)} \frac{dM(\lambda)}{d\lambda} \left\langle \frac{1}{2P} \sum_{i=1}^P (\vec{R}^{(i)} - \vec{R}_c) \cdot \frac{\partial V}{\partial \vec{R}^{(i)}} + \frac{3}{2\beta} \right\rangle \quad (\text{II.38})$$

$$= -\frac{1}{M(\lambda)} \frac{dM(\lambda)}{d\lambda} \left\langle \frac{1}{2P} \sum_{i=1}^P (\vec{R}^{(i)} - \vec{R}_c) \cdot \vec{F}^{(i)} + \frac{3}{2\beta} \right\rangle \quad (\text{II.39})$$

La dynamique moléculaire d'intégrales de chemin permet de générer des configurations à partir desquelles on récupère positions et forces permettant de calculer les moyennes intervenant dans le calcul des estimateurs. Des simulations effectuées pour plusieurs valeurs de λ , séparées d'une valeur $\Delta\lambda$, permettent alors de déterminer la variation d'énergie libre du système :

$$\Delta F_{(S)} = \sum_{\{\lambda\}} A_{vir}^{(2)}(\lambda) \Delta\lambda \quad (\text{II.40})$$

On obtient ensuite le facteur \mathcal{B} à l'aide des équations II.13 et II.15.

Choix des masses et paramètres λ

Pour le calcul de la variation de l'énergie libre, on se doit de choisir une paramétrisation telle que λ varie entre 0 et 1 avec $M(\lambda=0) = M^a$ et $M(\lambda=1) = M^b$. Il est alors possible de choisir une paramétrisation linéaire:

$$M(\lambda) = M^a(1 - \lambda) + \lambda M^b \quad (\text{II.41})$$

Néanmoins, on souhaite effectuer le moins de simulations possible. Connaissant l'énergie libre, F_{harm} , d'un oscillateur harmonique de pulsation ω :

$$F_{harm} = \frac{\hbar\omega}{2} + \frac{1}{\beta} \ln(1 - e^{-\beta\hbar\omega}) \quad (\text{II.42})$$

et dans le cas où le système est proche de son état fondamental :

$$F_{harm} \approx \frac{\hbar\omega}{2} \quad (\text{II.43})$$

sachant que $\omega \propto 1/\sqrt{M}$, [75] propose de choisir une paramétrisation de la forme :

$$\frac{1}{\sqrt{M(\lambda)}} = \frac{1 - \lambda}{\sqrt{M^a}} + \frac{\lambda}{\sqrt{M^b}} \quad (\text{II.44})$$

Dans ce cas, pour un système proche d'un système harmonique, la dérivée de l'énergie libre, c'est à dire l'estimateur, devrait varier (quasi-)linéairement avec λ . Il est alors possible de d'utiliser cette linéarité pour calculer la variation d'énergie libre à partir d'une seule simulation avec $\lambda = 1/2$ au lieu de l'intégration II.27. Ainsi :

$$\Delta F_{(S)} = A_{vir}^{(2)}(\lambda = 1/2) \quad (\text{II.45})$$

Il reste nécessaire de vérifier cette méthode. Elle a toutefois été vérifiée par [35] pour une solution ionique $\text{Li}_{(\text{aq})}^+$.

Coordinnence du calcium en solution aqueuse

Comme nous l'avons vu au chapitre II, la coordinnence est directement liée au fractionnement isotopique. Ce chapitre est donc consacré à l'étude de la coordinnence du calcium en solution aqueuse $\text{Ca}_{(\text{aq})}^{2+}$, c'est-à dire au nombre de molécules d'eau présentes dans la première sphère de coordinnation du calcium, à partir de dynamiques moléculaires ab-initio. Il s'agit de trouver les paramètres de simulation permettant une bonne description de l'environnement du calcium, ce qui nous permettra ensuite de calculer le facteur de partage isotopique.

La détermination de la coordinnence de $\text{Ca}_{(\text{aq})}^{2+}$ a déjà fait l'objet de nombreuses études, à la fois expérimentales et numériques. On présente alors dans une première partie (III.1) ces résultats et on verra que la valeur de la coordinnence est toujours sujette à débat, motivant notre étude. Après avoir décrit, partie III.2, les paramètres des simulations, une troisième partie III.3 est consacrée à l'influence du choix de la fonctionnelle utilisée lors des dynamiques moléculaires. On verra notamment que nous avons pu mettre en évidence des problèmes d'échantillonnage. C'est pourquoi nous avons utilisé des méthodes d'analyse statistique se basant sur les chaînes de Markov afin d'étudier l'influence de la température sur la coordinnence, ainsi que la convergence de sa valeur.

Table des matières

III.1 Précédentes études	66
III.2 Système et protocole de simulation	68
III.3 Résultats	69
III.3.1 Propriétés de l'eau	69
III.3.2 Environnement du calcium	73
III.3.3 Etude en température	76
III.3.4 Evolution temporelle de la coordinnence	79
III.3.5 Echanges et temps de résidence	82
III.3.6 Estimation de l'incertitude/critère de convergence	84
III.4 Conclusion et perspectives	88

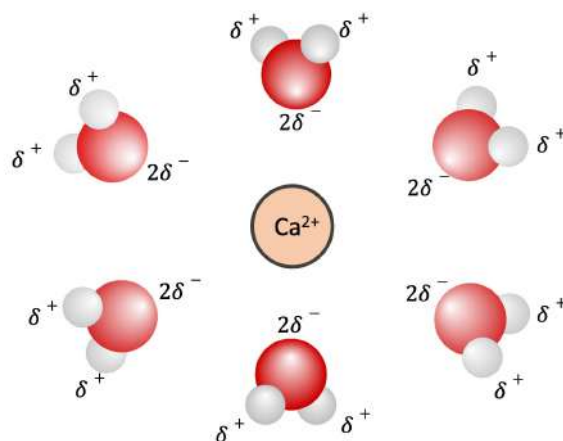


Figure III.1: Représentation schématique du complexe $[\text{Ca}(\text{H}_2\text{O})_6]^{2+}$. Les molécules d'eau étant polaires, elles s'orientent de telle sorte que les atomes d'oxygène soient au plus proche du calcium qui est chargé positivement.

III.1 Précédentes études

Le calcium en solution aqueuse est présent sous forme ionique de charge +2. Les molécules d'eau présentes dans la solution vont s'arranger autour de l'ion formant une sphère. Par exemple, une coordinence égale à 6 signifie qu'il a comme plus proches voisins 6 molécules d'eau (voir Figure III.1). Et l'ensemble forme un complexe aqua que l'on note $[\text{Ca}(\text{H}_2\text{O})_6]^{2+}$.

De par son importance dans la compréhension de processus (bio)physico-chimiques, la détermination de la valeur de la coordinence de $\text{Ca}_{(\text{aq})}^{2+}$ a été l'objet de nombreuses études.

Expérimentalement, il est possible d'utiliser des techniques de diffraction, de rayon X ou par neutron, ou de spectroscopie de rayon X. En fonction de la méthode utilisée, des valeurs de coordinence allant de 5,1 à 10 ont été trouvées [40, 88, 78, 1, 72, 30, 4, 106, 65, 104, 57] (voir Tableau III.1).

Des simulations numériques, de dynamique moléculaire ab-initio [16, 73, 33, 7, 82, 26, 112, 124] ou avec potentiel empirique [88, 65], de dynamiques ab-initio (DFT) contraintes [5, 63], de type Monte-Carlo [15], ou des simulations hybrides ab-initio/empirique [99, 98] ont aussi été effectuées dans le but de mieux comprendre les processus se passant à l'échelle atomique et d'avoir accès à des informations inaccessibles expérimentalement. Pour l'ensemble de ces simulations, des valeurs de coordinence variant aussi de 5.1 à 10 ont été obtenues. On reporte tableau III.2 les valeurs de la coordinence obtenues, allant de 5.9 à 8, pour les simulations de dynamique moléculaire ab-intio (DFT), qui nous serviront de comparaison par la suite.

L'ensemble de ces résultats montre à la fois que la méthode utilisée influe sur la détermination de la coordinence et que le calcium en solution aqueuse est présent sous différentes formes de complexes aqua. Les études de dynamiques ab-initio contraintes, permettant d'avoir accès à l'énergie libre en fonction de la coordinence montrent que

Référence	sel	Méthode*	Coordinnence
[40]	CaCl ₂	XAFS	7.2 ± 1.2
[88]	CaCl ₂	XRD	6.9
[78]	CaCl ₂	XRS	5.1 – 8.0
[1]	CaCl ₂	XRD	8.0 – 8.2
[72]	CaCl ₂	XRD	6
[30]	CaCl ₂	ND	5.5 ± 0.3
[4]	CaCl ₂	ND	6.95 – 7.3
[106]	CaCl ₂	EXAFS	8
[65]	CaCl ₂ /CaI ₂ /CaBr ₂ /Ca(ClO ₄) ₂	LAXS/EXAFS	8
[104]	Ca(NO ₃) ₂	XRD	6
[57]	CaCl ₂	XRD/ND	10 – 6.4

Table III.1: Résultats expérimentaux de la coordinnence du calcium en solution aqueuse de précédentes études.

*XRD : diffraction par rayons X, ND : diffraction de neutron, (E)XAFS : spectroscopie d'absorption par rayon X (étendue), XRS : diffusion par rayon X, LAXS : diffusion par rayon X grand angle.

Référence	Méthode	Système	Fonctionnelle	Température(K)	t _{simu} (ps)	Coordinnence
[16]	NVT/CPMD	1Ca ²⁺ +64H ₂ O	PBE96	300	33	6.67
[73]	NVT/CPMD	1Ca ²⁺ +60H ₂ O	PBE	285.1	10	6.2
				289.1	6	6.2
				291.0	40	7
[33]	NVT/CPMD	1Ca ²⁺ +53D ₂ O	PBE	400	15 – 20	5.9
		1CaHCO ₃ ⁺ +52D ₂ O				6.4
		1CaCO ₃ +52D ₂ O				6.0
[7]	NVE/CPMD	1Ca ²⁺ +54D ₂ O	BLYP	306	7	6
[82]	NVT/CPMD	1Ca ²⁺ +31D ₂ O	BLYP	300	14.2	7 – 8
[26]	NVT/CPMD	1Ca ²⁺ +48H ₂ O	BLYP	300	50	6
[112]	NVT/CPMD	1CaCl ₂ +58H ₂ O	BLYP	320	7.2	6 – 7
		2CaCl ₂ +58H ₂ O				6
		2Ca ²⁺ +62H ₂ O				6
[124]	NVT/CPMD	Ca ²⁺ +63H ₂ O	PBE	300	50	6.0
			PBE-D3		50	6.0
					100	6.0
					500	6.0
		Ca ²⁺ +124H ₂ O	PBE-D3	300	200	6.5
		Ca ²⁺ +64H ₂ O	revPBE	300	50	6.6
		Ca ²⁺ +64H ₂ O	revPBE-D3	300	50	6.1
		Ca ²⁺ +64H ₂ O	BLYP	300	50	6.3
		Ca ²⁺ +64H ₂ O	BLYP-D3	300	500	6.9

Table III.2: Résultats de précédentes études du calcium en solution aqueuse déterminés à partir de dynamique moléculaire ab-initio (DFT).

le profil énergétique du calcium comporte effectivement plusieurs minima accessibles à température ambiante.

La détermination de la coordinence est donc difficile, motivant notre étude de l'influence des paramètres de simulations. Nous avons, dans un premier temps, effectué des dynamiques moléculaires ab-initio dans le but de tester l'influence du choix de la fonctionnelle. Nous verrons que cela nous a amené par la suite à étudier l'influence de la température et à développer des méthodes permettant de juger du bon échantillonnage de nos simulations.

D'autres simulations, dans le but d'étudier l'influence d'un contre-ion et de la taille de la boîte de simulation ont été effectuées et sont présentées en annexe B.2.

III.2 Système et protocole de simulation

Des simulations de dynamique moléculaire ab-initio (DFT) ont été effectuées avec le code Quantum Espresso [46]. Nous avons étudié les fonctionnelles, de type GGA, BLYP, PBE et PBEsol (voir partie I.1.1.7). Des corrections "Grimme D2" [49] ont été ajoutées aux fonctionnelles PBE et BLYP, que l'on nommera PBE-D2 et BLYP-D2. Nous avons aussi utilisé la fonctionnelle non locale vdW-DF2 [71].

Pour l'ensemble des fonctionnelles, des pseudopotentiels de type Trouiller-Martins [115] ont été utilisés pour l'oxygène et l'hydrogène. Pour le calcium, des pseudopotentiels ultrasoft provenant de la librairie "PSLibrary" ont été utilisés pour les fonctionnelles BLYP et PBEsol. Pour les fonctionnelles PBE et vdW-DF2, nous avons utilisé des potentiels de type RRKJ générés à partir du code "OPIUM" (opium.sourceforge.net).

Pour l'ensemble des simulations, une énergie de cutoff pour les fonctions d'onde de $80Ry$ a été appliquée et une énergie de cutoff de $320Ry$ pour la densité électronique. Les simulations ont été effectuées dans l'ensemble statistique NVT, avec des thermostats de type Nose-Hoover composés de chaînes de longueur 4 pour chaque atome. Le pas de temps a été fixé à $\Delta t = 2.5a.u.$ et des conditions aux limites périodiques ont été appliquées.

Le système est composé d'un ion calcium $Ca_{(aq)}^{2+}$ et de 62 molécules d'eau dans une boîte cubique de côté égale à 12.41\AA donnant une densité de l'eau de $0.9697g.cm^{-3}$. Un fond homogène chargé a été ajouté pour assurer la neutralité électrique du système.

Les dynamiques moléculaires de l'eau présentant des problèmes d'ergodicité à température ambiante [120], nous avons choisi d'effectuer un première ensemble de simulations à 400K avant d'effectuer des simulations à 300K. Pour compléter notre étude, des simulations à 300K, 350K, 450K, 500K, 550K et 600K ont ensuite été conduites, uniquement avec la fonctionnelle BLYP-D2.

Pour les simulations à 400K, après relaxation du système, nous avons commencé dans l'ensemble NVE. Nous avons ensuite ajouté le thermostat pour progressivement arriver à la température désirée. Pour les simulations aux autres températures, nous avons récupéré la dernière configuration des simulations à 400K et ensuite baissé ou

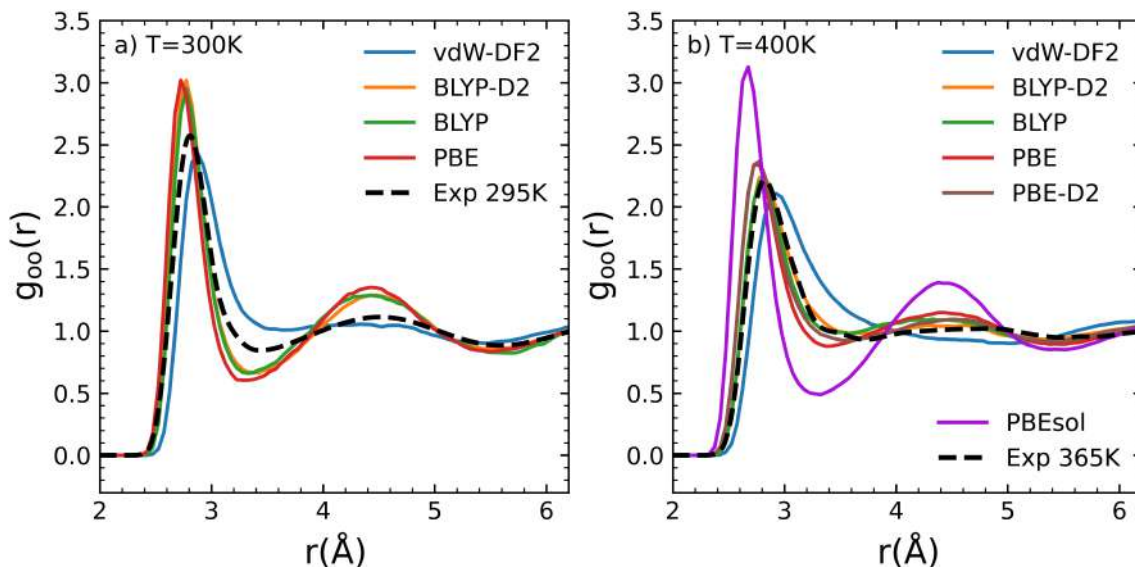


Figure III.2: Fonction de distribution radiale oxygène-oxygène obtenues pour différentes fonctionnelles à a) 300K et b) 400K. Les courbes en pointillées noires sont les fonctions de distribution radiale obtenues expérimentalement par [103]

augmenté lentement la température.

Les simulations à 300K avec les fonctionnelles PBE et BLYP font $30ps$. Celle avec la fonctionnelle BLYP à 400K fait $100ps$ et nous avons conservé les 70 dernières picosecondes pour analyse. Pour les autres simulations, le temps de simulation est de $120ps$ et les 100 dernières picosecondes ont été utilisées pour les analyses. Les configurations ont été sauvegardées tous les 10 pas.

III.3 Résultats

On présente dans un premier temps les résultats obtenus pour les dynamiques moléculaires à 300K et 400K.

On souhaite déterminer les paramètres de simulation permettant de décrire correctement la sphère de coordination du calcium. Mais, comme nous l'avons vu partie III.1, les études expérimentales donnent des valeurs de la coordinence très variables rendant impossible la comparaison simple des résultats des simulations avec les valeurs expérimentales. On commence alors par étudier les propriétés de l'eau, considérant qu'il est nécessaire de bien les décrire pour avoir une description correcte de l'environnement proche du calcium.

III.3.1 Propriétés de l'eau

Les fonctions de distributions radiale oxygène-oxygène, $g_{OO}(r)$, sont présentées figure III.2 et comparées aux courbes expérimentales obtenues par diffraction au rayon X par [103] à 295K et 365K, qui est la température maximale pour laquelle nous avons trouvé des données expérimentales.

Pour toutes les fonctionnelles, un premier maximum est observé à environ 2.8Å et un second, sauf pour vdW-DF2 à 400K, à environ 4.5Å correspondant à la première et deuxième sphère de coordinence de l'oxygène. Entre ces deux sphères, on trouve un minimum à environ 3.5Å (voir Tableau III.3). Pour la fonctionnelle vdW-DF2 à 400K, la deuxième sphère de coordinence disparaît.

La température influe très légèrement sur la position du premier maximum avec une augmentation de 0.05Å entre 300K et 400K pour toutes les fonctionnelles. Les données expérimentales de [103], entre 295K et 365K, montrent que le premier maximum se situe entre 2.80Å et 2.82Å. Pour cette donnée, c'est la fonctionnelle BLYP, avec ou sans termes de dispersion de Grimme, qui donne le meilleur accord avec l'expérience. Les fonctionnelles PBEsol et PBE (avec ou sans terme de correction Grimme D2) donnent une position inférieure alors que la fonctionnelle vdW-DF2 donne une valeur supérieure.

Concernant le premier minimum, les données expérimentales montrent que sa position augmente avec la température allant de 3.4Å, à ~300K, à 3.7 à ~360K. Cette augmentation est aussi observée dans nos simulations. L'ajout de correction de Grimme aux fonctionnelles PBE et BLYP augmente aussi cette valeur de ~0.10Å pour BLYP et ~0.20Å pour PBE. A 300K, les fonctionnelles PBE et BLYP donnent une valeur de 3.33Å, plus faible que celle obtenue expérimentalement. Pour BLYP-D2, on obtient une valeur de 3.43Å, proche de la valeur expérimentale alors que la fonctionnelle vdW-DF2 surestime la valeur avec un minimum à 3.63Å.

Enfin, pour la position de la deuxième sphère de coordinence, bien que les données expérimentales semblent montrer une augmentation avec la température, cette tendance n'est pas observée dans nos simulations. Comme précédemment, l'ajout de terme de dispersion de Grimme D2 augmente sa valeur de ~0.10Å pour les fonctionnelles PBE et BLYP. Toutes les fonctionnelles, sauf vdW-DF2, donnent une position légèrement inférieure à la valeur expérimentale.

La différence entre les fonctions de distribution radiale obtenues pour les différentes fonctionnelles est plus flagrante lorsque l'on regarde l'amplitude des extrema. A 300K, la fonctionnelle vdW-DF2 sous-estime l'amplitude des ces extrema comparé à l'expérience, ce qui est la signature d'une eau sous-structurée. Au contraire, pour les fonctionnelles PBE, BLYP et BLYP-D2, l'amplitude des extrema est surestimée donnant une eau sur-structurée.

A 400K, les valeurs montrent que pour la fonctionnelle PBEsol on observe une eau sur-structurée alors que pour la fonctionnelle vdW-DF2 elle est sous-structurée. Les fonctionnelles PBE, PBE-D2, BLYP et BLYP-D2 donnent des amplitudes en bien meilleur accord avec les valeurs expérimentales. La fonctionnelle BLYP donne tout de même des amplitudes plus faibles que la fonctionnelle PBE, et reproduit mieux la $g_{OO}(r)$.

La température semble jouer un rôle important avec un écart aux valeurs expérimentales très important à 300K. En particulier, pour les fonctionnelles PBE et BLYP, avec ou sans correction de Grimme, la sur-structuration de l'eau observée à 300K disparaît à 400K. Pour regarder plus en détail ce phénomène, on trace figure III.3 les fonctions de déplacement quadratique moyen de l'oxygène, MSD_O .

Dynamiques moléculaires							
Fonctionnelle	T(K)	$g_{OO}^{\max,1}$	$r_{OO}^{\max,1}$ (Å)	$g_{OO}^{\min,1}$	$r_{OO}^{\min,1}$ (Å)	$g_{OO}^{\max,2}$	$r_{OO}^{\max,2}$ (Å)
PBEsol	400	3.13	2.68	0.49	3.33	1.39	4.38
PBE	300	3.02	2.73	0.60	3.33	1.35	4.43
	400	2.34	2.78	0.88	3.38	1.15	4.43
BLYP	300	2.95	2.78	0.66	3.33	1.29	4.43
	400	2.21	2.83	0.98	3.58	1.10	4.43
PBE-D2	400	2.37	2.78	0.92	3.58	1.09	4.53
BLYP-D2	300	3.02	2.78	0.66	3.43	1.29	4.48
	350	2.55	2.78	0.87	3.53	1.13	4.53
	400	2.24	2.83	0.98	3.68	1.05	4.46
	450	2.09	2.83	0.99	3.98	-	-
	500	1.98	2.83	-	-	-	-
	550	1.93	2.88	-	-	-	-
	600	1.90	2.88	-	-	-	-
vdW-DF2	300	2.43	2.88	1.00	3.63	1.06	4.33
	400	2.10	2.93	0.90	5.03	-	-
Données expérimentales							
Ref	T(K)	$g_{OO}^{\max,1}$	$r_{OO}^{\max,1}$ (Å)	$g_{OO}^{\min,1}$	$r_{OO}^{\min,1}$ (Å)	$g_{OO}^{\max,2}$	$r_{OO}^{\max,2}$ (Å)
[103]	295	2.57	2.81	0.84	3.44	1.11	4.50
[105]	298	2.75	2.74	0.78	3.37	1.16	4.51
[103]	307	2.50	2.80	0.87	3.38	1.09	4.53
	312	2.50	2.81	0.88	3.47	1.08	4.51
	323	2.43	2.81	0.90	3.51	1.07	4.56
	334	2.37	2.82	0.92	3.58	1.05	4.60
	342	2.34	2.83	0.93	3.62	1.04	4.59
	354	2.22	2.82	0.93	3.70	1.03	4.63
	365	2.21	2.82	0.93	3.73	1.02	4.85

Table III.3: Positions et amplitudes du premier et deuxième maxima de la fonction de distribution radiale de l'oxygène, ($r_{OO}^{\max,1}$; $g_{OO}^{\max,1}$) et ($r_{OO}^{\max,2}$; $g_{OO}^{\max,2}$), ainsi que du premier minimum, ($r_{OO}^{\min,1}$; $g_{OO}^{\min,1}$) obtenues pour les différentes dynamiques moléculaire ainsi que les valeurs des données expérimentales de [103, 105].

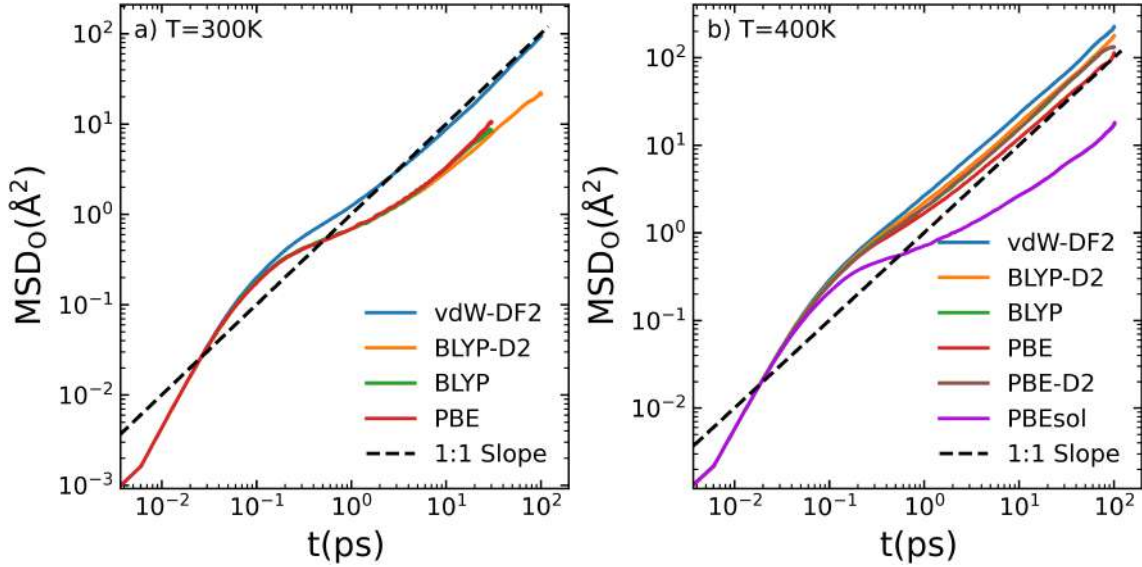


Figure III.3: Fonction de déplacement quadratique moyen de l'oxygène, en échelle logarithmique, obtenues pour différentes fonctionnelles à a) 300K et b) 400K. Les courbes en pointillées noires sont des droites de pente égale à 1 caractéristiques du régime diffusif.

Pour toutes les simulations, on observe un régime balistique puis un régime transitoire et enfin un régime diffusif. Néanmoins, le régime diffusif semble être atteint après des temps de simulation différents en fonction de la fonctionnelle utilisée. On calcule le coefficient de dispersion de l'oxygène D_O à l'aide de la formule d'Einstein :

$$MSD_O(t) = 6D_O t \iff D_O = \frac{MSD_O(t)}{6t} \quad (\text{III.1})$$

Cette relation est vraie en régime diffusif. Si celui-ci est atteint, on devrait observer la convergence du coefficient de diffusion. Les courbes se trouvent figure III.4. La convergence du coefficient de diffusion est observée pour toutes les fonctionnelles.

A partir de ces courbes, on calcule les coefficients de dispersion en prenant la moyenne des valeurs entre 20 et 80 picosecondes. Pour les trajectoires à 300K avec les fonctionnelles BLYP et PBE, qui font au total 30ps, la valeur du coefficient de diffusion est déterminée en prenant la moyenne sur les 10 dernières picosecondes de la trajectoire. Pour la trajectoire à 400K avec la fonctionnelle BLYP, on prend la moyenne entre 20 et 50 picosecondes. Les résultats sont comparés à la littérature tableau III.4.

Les valeurs ainsi obtenues se situent entre 0.04 et $0.15 \text{Å}^2 \cdot \text{ps}^{-1}$ à 300K et 0.03 et $0.36 \text{Å}^2 \cdot \text{ps}^{-1}$ à 400K et sont du même ordre de grandeur que celles trouvées pour d'autres simulations de l'eau dans la littérature. Le tableau III.4 montre qu'il est difficile de conclure sur l'effet des différents paramètres de simulation sur le coefficient de diffusion. Il est aussi important de noter que, les MDS_O étant obtenues par moyenne des MSD_O calculées sur des intervalles de temps de 0ps à T , tels que $T \leq 100\text{ps}$, voir annexe A, les valeurs du coefficient de diffusion en fin de trajectoire sont moins bien estimées.

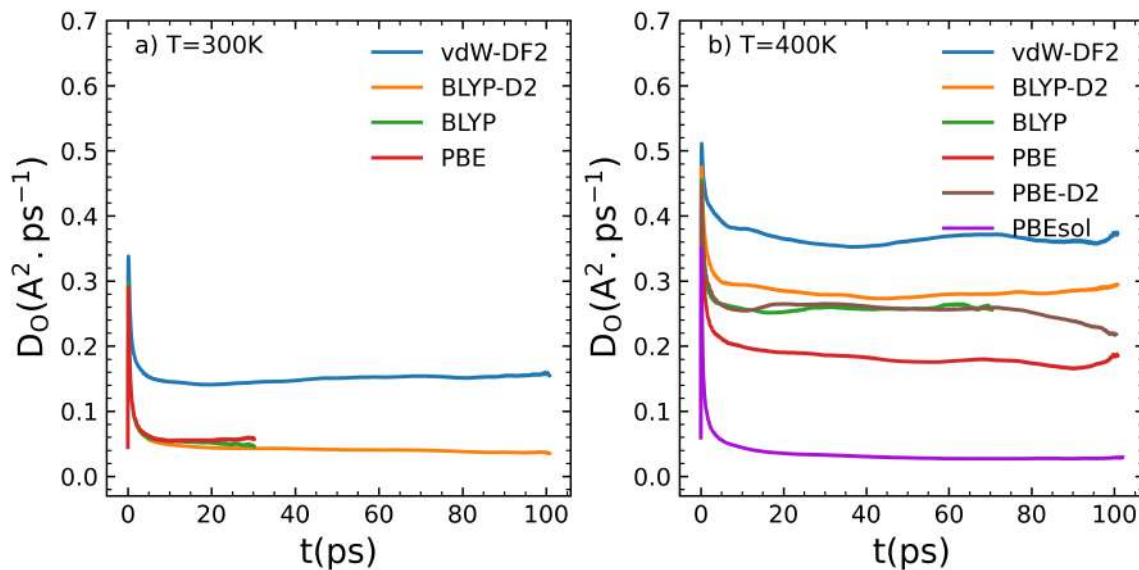


Figure III.4: Coefficients de diffusion de l'oxygène, en échelle logarithmique, obtenus pour différentes fonctionnelles à a) 300K et b) 400K.

Finale­ment, bien qu'avec la fonctionnelle vdW-DF2, on obtient une eau sous-structurée, les autres fonctionnelles donnent quant à elle une eau sur-structurée. Cette sur-structuration est plus importante pour les fonctionnelles PBEsol et PBE comparée à la fonctionnelle BLYP confirmant les conclusions de la revue [47]. L'ajout de termes de dispersion de Grimme D2 donne des positions des extrema plus grandes améliorant les $g_{OO}(r)$. On observe finalement que plus l'eau est structurée, plus le coefficient de diffusion est faible, avec des valeurs très inférieures à la valeur expérimentale à 300K [108] qui est de $0.219\text{\AA}^2\cdot\text{ps}^{-1}$. Il est aussi important de mentionner que la présence de $\text{Ca}_{(\text{aq})}^{2+}$ change la densité du liquide et peut donc influencer la diffusion de l'oxygène. Enfin, la méthode, et notamment le choix du thermostat, peut aussi avoir une influence sur les valeurs du coefficient de diffusion.

On poursuit notre étude en regardant maintenant l'effet du choix de la fonctionnelle à 300K et 400K sur l'environnement du calcium.

III.3.2 Environnement du calcium

Les fonctions de distribution radiale calcium-oxygène $g_{\text{CaO}}(r)$, ainsi que leurs intégrales permettant d'obtenir la coordinence, pour les différentes fonctionnelles testées à 300K et 400K se trouvent figure III.5.

Aux deux températures, deux sphères de coordinence peuvent être observées à des distances Ca-O d'environ 2.4\AA et 4.5\AA (voir Tableau III.5). A 300K, la position de la première sphère de coordinence est la même pour toutes les fonctionnelles et se trouve à 2.38\AA sauf pour la fonctionnelle vdW-DF2 où elle se trouve à 2.43\AA . A 400K, elle se situe à 2.33\AA pour la fonctionnelle PBEsol et ne change pas pour les

Functionelle	Ref	sim-alg	$N_{\text{H}_2\text{O}}$	T(K)	$t_{\text{simu}}(\text{ps})$	$D(\text{\AA}^2.\text{ps}^{-1})$
PBEsol		NVT/CP	62	400	100	0.029
PBE	[2]	NVE/CP	54	291	20	0.03
	[50]	NVE/CP	54	295	20	0.01
	[34]	NVT/CP	64	300	20	0.020
		NVT/CP	62	300	30	0.049
	[97]	NVT/BO	64	306	20	0.01
	[74]	NVE/CP	64	314	50-117	0.03
	[102]	NVT/CP	32	325	20	0.012
	[120]	NVE/CP,BO	32	337	20	0.06
	[113]	NVT/CP	32	350	10	0.047
		NVT/CP	62	400	100	0.18
BLYP	[50]	NVE/CP	32	293	20	0.01
	[38]	NVE/BO	64	300	11-32	0.02
		NVT/CP	62	300	30	0.049
	[70]	NVE/CP	32	309	60	0.06
	[66]	NVT/BO	128	317	57	0.020
	[74]	NVE/CP	64	319	50-117	0.10
	[6]	NVE/NVT/CP/BO/MC	64	323	20	0.01
	[120]	NVE/CP,BO	32	324	20	0.05
	[113]	NVT/CP	32	350	10	0.048
	[8]	NVE/BO	64	353	40	0.114
		NVT/CP	62	400	70	0.26
PBE-Gd2	[74]	NVE/CP	64	324	50-117	0.06
		NVT/CP	62	400	100	0.26
BLYP-Gd2		NVT/CP	62	300	100	0.041
	[74]	NVE/CP	64	321	50-117	0.16
	[8]	NVE/BO	64	328	40	0.071
		NVT/CP	62	350	100	0.14
		NVT/CP	62	400	100	0.28
		NVT/CP	62	450	100	0.43
		NVT/CP	62	500	100	0.57
		NVT/CP	62	550	100	0.70
		NVT/CP	62	600	100	0.74
vdW-DF2		NVT/CP	62	300	100	0.15
		NVT/CP	62	400	100	0.36
Experimental	[108]			298.2		0.219

Table III.4: Coefficients de diffusion de l'oxygène obtenus à partir de nos simulations (en **gras**) et obtenues dans la littérature pour des simulation de dynamiques moléculaires ab-initio Bohr-Oppenheimer (BO) ou Car-Parrinello (CP) et monte-carlo (MC) de l'eau, pour les fonctionnelles utilisées dans notre étude. $N_{\text{H}_2\text{O}}$ est le nombre de molécules d'eau du système étudié, T la température, t_{simu} le temps de simulation et sim-alg la méthode utilisée. NVE et NVT correspondent aux ensembles microcanonique et canonique.

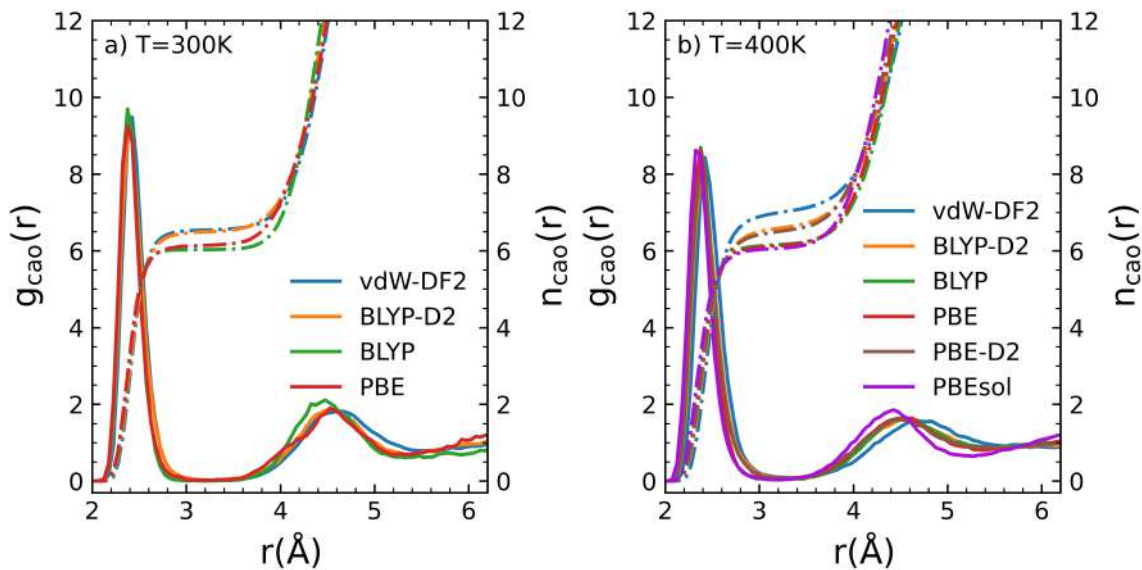


Figure III.5: Fonctions de distribution radiale calcium-oxygène (trait plein) et leurs intégrales (en pointillé) obtenues pour différentes fonctionnelles à a) 300K et b) 400K.

autres.

La coordinnence moyenne peut être estimée à partir de la valeur de l'intégrale de la RDF Ca-O prise à la distance correspondant au minimum des $g_{CaO}(r)$ entre la première et la deuxième sphère de coordinnence. Nous appellerons cette valeur $CN^{(1)}$ par la suite.

La coordinnence $CN^{(1)}$ varie en fonction des différentes fonctionnelles (voir Tableau III.6). A 300K, $CN^{(1)}$ est égale à 6.0, 6.14, 6.50 et 6.52 pour PBE, BLYP, BLYP-D2 et vdW-DF2, respectivement. A 400K, la variation des valeurs est encore plus grande allant de 6.05 à 7.02. Pour les deux températures, un effet clair est observé par l'ajout de termes de dispersion augmentant la valeur de $CN^{(1)}$.

Entre 300K et 400K, la température ne semble pas avoir un effet très important sur la valeur de $CN^{(1)}$ sauf pour la fonctionnelle vdW-DF2. Cet effet peut être attribué à la différence trouvée sur la position de la première sphère de coordinnence et à la distance moyenne Ca-O (proches voisins). Plus cette distance est grande, plus il y a de molécules d'eau dans la sphère de coordinnence.

Au vu des résultats, nous nous sommes demandés si la température avait une influence sur la valeur de la coordinnence et si les différences observées pouvaient aussi être le résultat d'un problème d'échantillonnage de nos simulations aux basses températures. C'est pourquoi nous avons décidé de faire une étude plus approfondie avec la fonctionnelle qui nous a donné de meilleurs résultats pour la description de l'eau, c'est à dire BLYP-D2.

Dynamiques moléculaires							
Fonctionnelle	T(K)	$g_{\text{CaO}}^{\text{max},1}$	$r_{\text{CaO}}^{\text{max},1}$ (Å)	$g_{\text{CaO}}^{\text{min},1}$	$r_{\text{CaO}}^{\text{min},1}$ (Å)	$g_{\text{CaO}}^{\text{max},2}$	$r_{\text{CaO}}^{\text{max},2}$ (Å)
PBEsol	400	8.62	2.33	0.03	3.18	1.86	4.43
PBE	300	9.25	2.38	0.02	3.03	1.90	4.53
	400	8.63	2.38	0.03	3.18	1.64	4.58
BLYP	300	9.69	2.38	0.01	3.03	2.11	4.48
	400	8.70	2.38	0.06	2.98	1.60	4.53
PBE-D2	400	8.14	2.38	0.08	3.33	1.63	4.48
BLYP-D2	300	8.99	2.38	0.02	3.28	1.87	4.58
	350	8.71	2.38	0.05	3.23	1.71	4.48
	400	7.94	2.28	0.08	3.28	1.61	4.53
	450	7.63	2.38	0.10	3.33	1.50	4.63
	500	7.05	2.38	0.13	3.28	1.50	4.63
	550	6.93	2.38	0.18	3.43	1.38	4.48
vdW-DF2	600	6.60	2.38	0.19	3.33	1.37	4.73
	300	9.48	2.43	0.02	3.23	1.82	4.63
	400	8.43	2.43	0.08	3.33	1.58	4.63

Table III.5: Positions et amplitudes du premier et deuxième maximum de la fonction de distribution radiale de calcium-oxygène, $(r_{\text{CaO}}^{\text{max},1}; g_{\text{CaO}}^{\text{max},1})$ et $(r_{\text{CaO}}^{\text{max},2}; g_{\text{CaO}}^{\text{max},2})$, ainsi que du premier minimum, $(r_{\text{CaO}}^{\text{min},1}; g_{\text{CaO}}^{\text{min},1})$ obtenues pour les différentes dynamiques moléculaires.

III.3.3 Etude en température

Avant de s'intéresser plus en détail à l'étude de l'échantillonnage de nos simulations, on commence par présenter, comme pour les deux parties précédentes, les résultats obtenus sur la structure de l'eau et l'environnement du calcium avec la fonctionnelle BLYP-D2 à des températures allant de 300K à 600K.

III.3.3.1 Structure de l'eau

Les fonctions de distribution radiale oxygène-oxygène $g_{\text{OO}}(r)$ pour la fonctionnelle BLYP-D2 à 300K, 350K, 400K, 450K, 500K, 550K et 600K sont présentées figure III.6 et les valeurs des positions et amplitudes des extrema se trouve Tableau III.3. Comme vu dans la partie III.3.1, on observe que les positions des extrema augmentent avec la température et que leurs amplitudes diminuent.

Les coefficients de diffusion calculés à partir des MSDs, figures III.7, se trouvent figure III.8. Les moyennes des coefficients de diffusion prisent entre 20ps et 80ps se trouvent tableau III.4. On observe que le régime diffusif est atteint après ~ 20 ps. Les valeurs des coefficient de diffusion augmentent significativement avec la température passant de $0.041 \text{Å}^2 \cdot \text{ps}^{-1}$ à $0.74 \text{Å}^2 \cdot \text{ps}^{-1}$. Les différentes valeurs du coefficient de diffusion aux différentes températures nous donnent la possibilité d'étudier plus précisément son évolution avec la température. On trace, figure III.9, le coefficient de diffusion, en échelle logarithmique, en fonction de l'inverse de la température. On observe deux régimes, un premier pour des températures variant de 300K à 400K et un second pour des températures variant de 450K à 600K. Deux ajustements

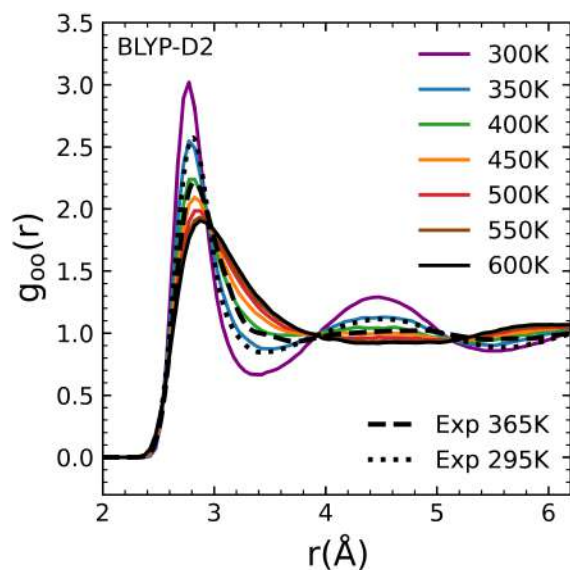


Figure III.6: Fonctions de distribution radiale oxygène-oxygène obtenues avec la fonctionnelle BLYP-D2 à différentes températures. Les courbes en pointillées noires sont les fonctions de distribution radiale obtenues expérimentalement par [103].

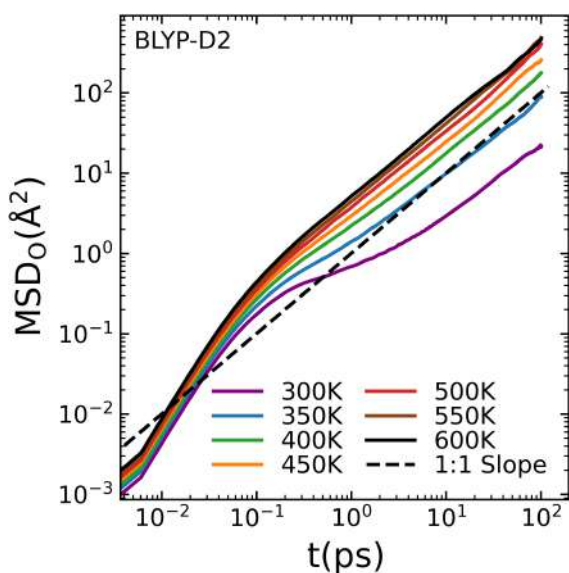


Figure III.7: Fonctions de déplacement quadratique moyen de l'oxygène obtenues avec la fonctionnelle BLYP-D2 à différentes températures.

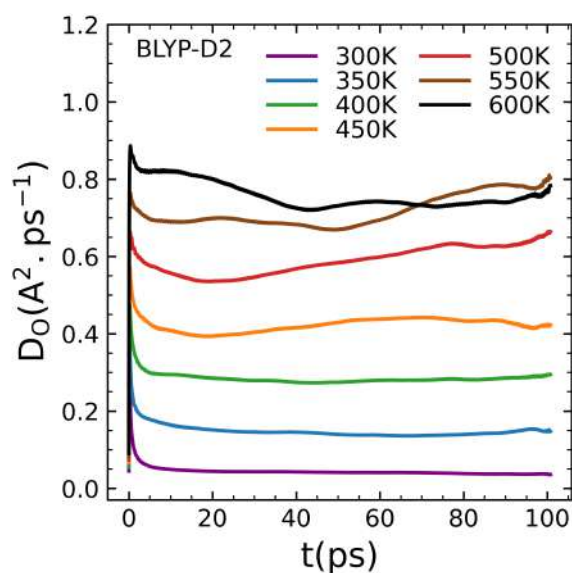


Figure III.8: Coefficient de diffusion de l'oxygène en fonction du temps obtenues avec la fonctionnelles BLYP-D2 à différentes températures.

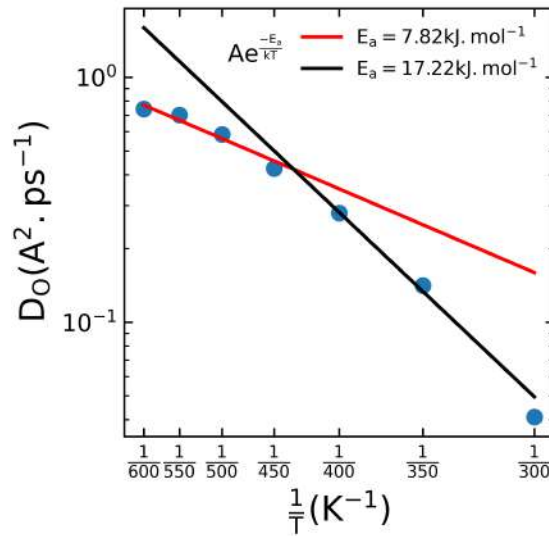


Figure III.9: Coefficients de diffusion obtenus avec la fonctionnelle BLYP-D2 en fonction de l'inverse de la température. Les courbes en rouge et noire correspondent à un ajustement par une fonction exponentielle suivant la loi d'Arrhenius.

exponentiels des données montrent que les coefficients de diffusion suivent une loi d'Arrhenius sur ces deux plages de températures :

$$D = A e^{-\frac{E_a}{kT}} \quad (\text{III.2})$$

avec A un coefficient, E_a l'énergie d'activation, k la constante de Boltzmann et T la température.

Les valeurs des énergies d'activation sont de $E_a = 17.22 \text{ kJ} \cdot \text{mol}^{-1}$ pour l'ajustement entre 300K et 400K et de $E_a = 7.82 \text{ kJ} \cdot \text{mol}^{-1}$ entre 450K et 600K. Les valeurs expérimentales de l'énergie d'activation sont de $17.8 \text{ kJ} \cdot \text{mol}^{-1}$ [108] et $18.41 \text{ kJ} \cdot \text{mol}^{-1}$ [123], calculées pour des variations de température de 278.2K à 308.2K et 278.2K à 318.2K, respectivement. Ces données sont en accord avec la valeur obtenue à partir de nos simulations entre 300K et 400K. L'observation de deux régimes peut s'expliquer par le fait qu'il est possible que la loi d'Arrhenius ne soit valable que sur des plages réduites de température. Cela peut aussi venir d'un problème d'échantillonnage aux basses températures et du choix du thermostat. Il faut aussi noter que nos simulations sont effectuées à volume constant et atteignent des températures très élevées, rarement prises en compte dans les expériences. Les conclusions pourraient aussi être différentes à pression constante.

III.3.3.2 Environnement du calcium

Les courbes des différentes $g_{CaO}(r)$, ainsi que leurs intégrales, aux différentes températures se trouvent figure III.10. Les positions et les amplitudes des extrema des $g_{CaO}(r)$ se trouvent tableau III.5. On observe, comme dans la partie précédente, que la position de la première sphère de coordinence ne dépend pas de la température et on trouve une valeur de 2.38 \AA . En revanche, les positions de la deuxième sphère de coordinence, ainsi que celles des minimum se situant entre les deux sphères augmentent

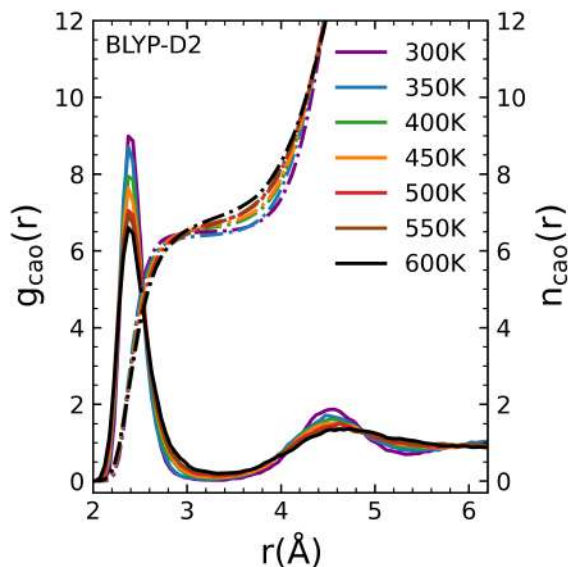


Figure III.10: Fonctions de distribution radiale, et leurs intégrales calcium-oxygène obtenue à différentes températures avec la fonctionnelle BLYP-D2.

quant à eux avec la température. Finalement, comme observé précédemment à 300K et 400K, les amplitudes des extrema diminuent avec la température et les largeurs des sphères de coordination augmentent. Les valeurs de la coordination déterminées à partir des intégrales des $g_{CaO}(r)$, $CN^{(1)}$ (voir tableau III.6), augmentent avec la température entre 350K et 600K passant de 6.40 à 6.83. Une valeur intermédiaire de 6.50 est trouvée pour la trajectoire à 300K.

Etant donné que le calcium possède plusieurs complexes aqua stables, même à température ambiante, on pourrait s'attendre à ce que les simulations à hautes températures permettent d'explorer plus efficacement et rapidement les configurations de coordinences différentes mais que la valeur de la coordination moyenne reste inchangée. Néanmoins, les résultats, entre 350K et 600K, semblent montrer que la coordination augmente avec la température, alors que la valeur trouvée à 300K dévie de cette tendance. Cela montre, en plus des conclusions de la partie III.3.2, qu'il est nécessaire d'étudier plus en détail la convergence de la valeur de la coordination. Ceci est en accord avec les conclusions des auteurs de [73] qui ont déclaré que de longues trajectoires pourraient être nécessaires pour une bonne convergence de la coordination.

III.3.4 Evolution temporelle de la coordination

Pour répondre au problème de convergence soulevé, il est nécessaire de mesurer l'évolution de la coordination durant nos simulations. Cela peut être fait en comptant le nombre de molécules d'eau, à chaque pas de temps, se trouvant dans la première sphère de coordination, c'est à dire le nombre d'oxygènes se trouvant à une distance inférieure $r_{CaO}^{\min,1}$, correspondant au minimum des $g_{CaO}(r)$ entre les deux sphères

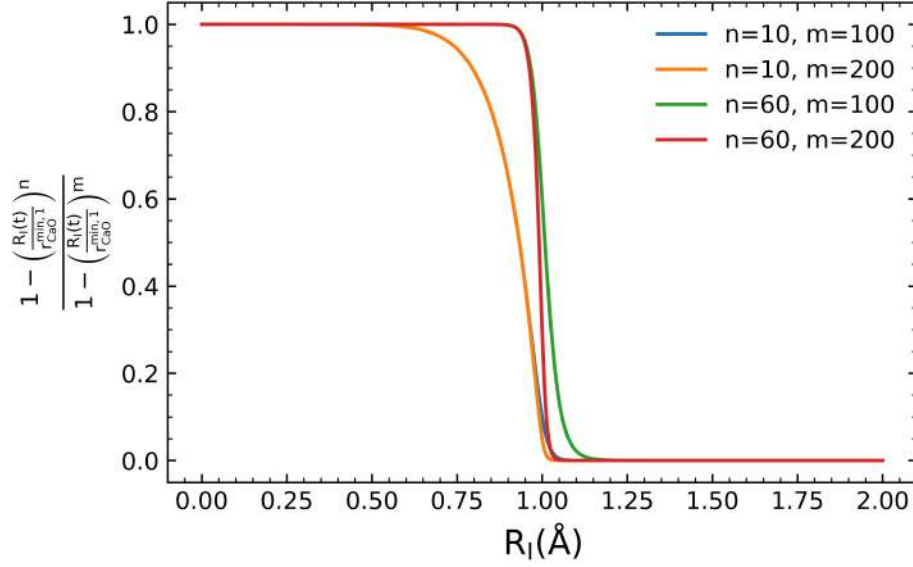


Figure III.11: Coordinence, en fonction de la distance R_I d'un atome, définie par l'équation III.3 pour différentes valeurs des paramètres n et m . La valeur de $r_{\text{CaO}}^{\text{min},1}$ est prise égale à 1.

de coordinence. Néanmoins, pour éviter de se retrouver dans le cas où un atome d'oxygène se trouve très proche de la valeur de $r_{\text{CaO}}^{\text{min},1}$, et entrant et sortant alors de nombreuses fois de la sphère de coordinence sur de courts intervalles de temps, entraînant ainsi de nombreux changements de coordinence, nous avons choisi de travailler avec une expression analytique de la coordinence :

$$CN(t) = \sum_{I=1}^{N_O} \frac{1 - \left(\frac{R_I(t)}{r_{\text{CaO}}^{\text{min},1}}\right)^n}{1 - \left(\frac{R_I(t)}{r_{\text{CaO}}^{\text{min},1}}\right)^m} \quad (\text{III.3})$$

où N_0 est le nombre d'atomes d'oxygènes, les R_I leurs distances à l'atome de calcium et n et m des paramètres.

Pour comprendre l'intérêt de définir une telle fonction, il est possible de la tracer, pour un atome I, en fonction de la distance R_I . La figure III.11 montre son évolution pour différents choix de paramètres n et m . Lorsqu'un atome se trouve à une distance bien inférieure à $r_{\text{CaO}}^{\text{min},1}$, la valeur de la fonction est de 1 alors que, à l'inverse, lorsqu'un atome se situe à une distance bien supérieure, elle est de 0. Entre ces deux valeurs, la fonction est continue diminuant l'effet d'aller-retour d'un atome d'oxygène qui se situerait à une distance $\sim r_{\text{CaO}}^{\text{min},1}$. Cela permet en plus une influence moins importante de la valeur de $r_{\text{CaO}}^{\text{min},1}$ qui a une incertitude non négligeable dû aux valeurs quasi-constantes des $g_{\text{CaO}}(r)$ entre les deux sphères de coordinence (voir figures III.5 et III.10). Plusieurs tests ont montré que le choix des paramètres n et m avait peu d'influence sur nos résultats finaux des coordinences moyennes et nous avons choisi de prendre $n = 60$ et $m = 200$. On présente, figure III.12 l'évolution de la coordinence en fonction du temps pour la trajectoire avec la fonctionnelle BLYP à 300K et pour la fonctionnelle BLYP-D2 à 600K. Ainsi que la coordinence moyenne

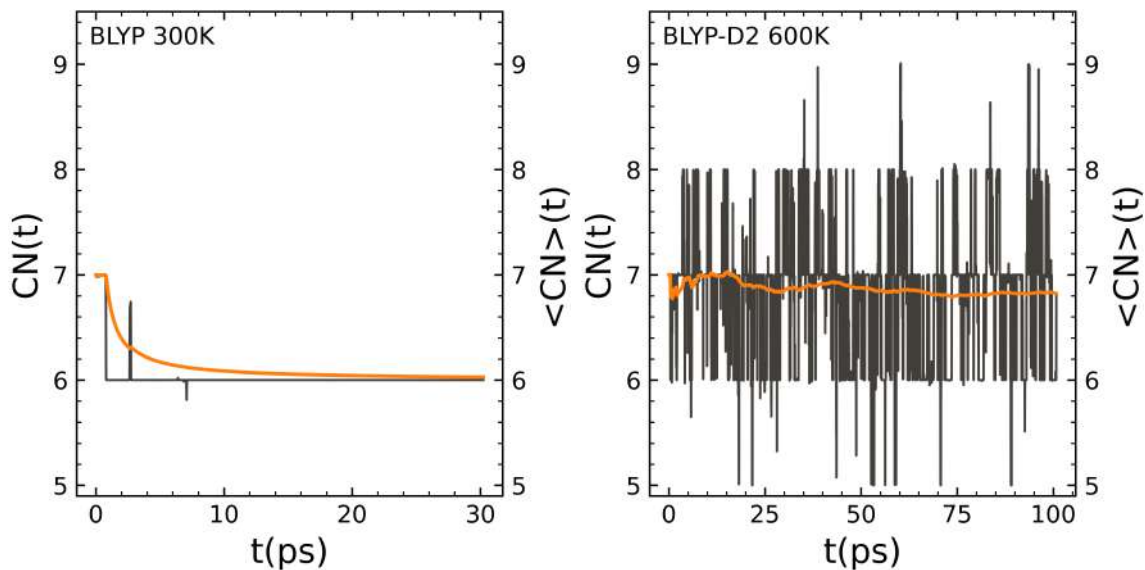


Figure III.12: Evolution de la coordinence en fonction du temps (en noir) ainsi que de sa moyenne (en orange) pour la trajectoire avec la fonctionnelle BLYP à 300K (à gauche) et avec la fonctionnelle BLYP-D2 à 600K (à droite).

en fonction du temps :

$$\langle CN \rangle (t) = \frac{1}{N_{steps}} \sum_{s=0}^{N_{steps}} CN(s\Delta t) \quad (\text{III.4})$$

où N_{steps} correspond au nombre de pas de la trajectoire au temps t et Δt le pas de temps. Pour les autres trajectoires, les figures se trouvent en annexe B.1. Les deux courbes montrent deux comportements très différents. A 300K, pour la fonctionnelle BLYP, après quelques picoseconde, la coordinence n'évolue plus et reste égale à 6 sur le reste de la trajectoire. Cela est en accord avec la faible diffusion de l'atome d'oxygène observée précédemment et signifie que le système n'explore pas suffisamment de configurations et reste dans son état initial (ou proche de son état de plus basse énergie) durant la simulation. A l'inverse, pour la fonctionnelle BLYP-D2 à 600K, la coordinence varie rapidement entre des valeurs de 6 et 7 et explore des configurations de coordinence 5, 8 et 9 sur de courts laps de temps. On peut, en l'état, supposer que le système est bien équilibré et qu'il a perdu la mémoire de son état initial.

Les courbes des évolutions de la moyenne de la coordinence, $\langle CN \rangle (t)$, semblent en revanche converger dans les deux cas. On appelle la dernière valeur $\langle CN \rangle (t)$, $CN^{(2)}$.

L'étude de ces deux exemples montrent que les valeurs de $CN^{(2)}$, se trouvant tableau III.6, ne peuvent pas être considérées comme justes dans tous les cas, même si la courbe $\langle CN \rangle (t)$ converge. Il est nécessaire d'observer un grand nombre d'"échanges", c'est à dire de molécules d'eau passant de la première à la

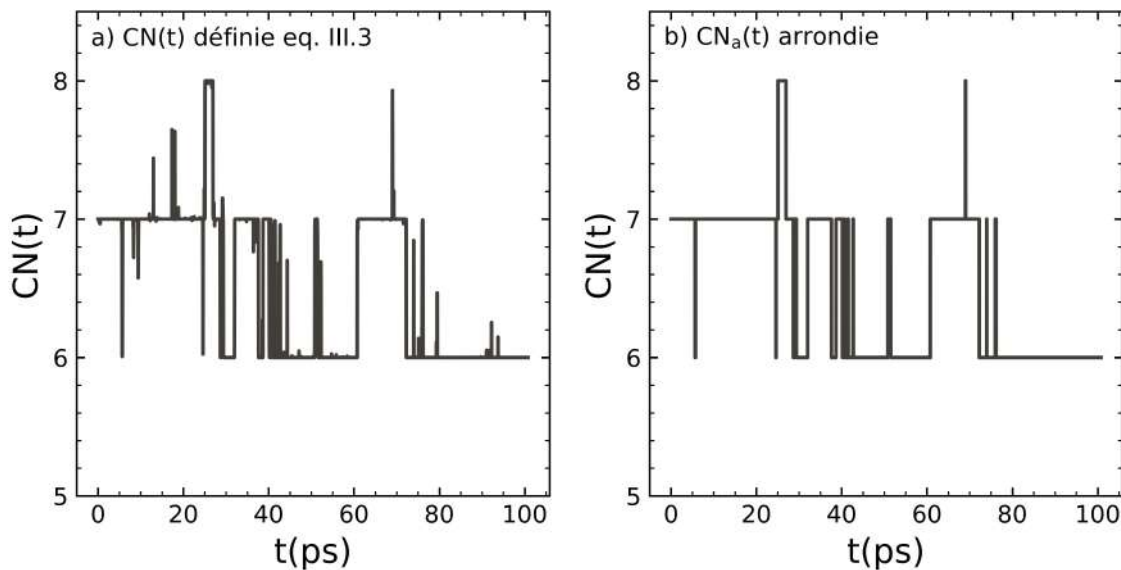


Figure III.13: Evolution de la coordinence en fonction du temps $CN(t)$ définie par l'équation III.3 à gauche et son évolution après "arrondissement" à droite pour la fonctionnelle BLYP-D2 à 300K.

deuxième sphère de coordinence et inversement, pour considérer que la trajectoire a échantillonné suffisamment de configurations. Encore une fois, pour les trajectoires à basses températures montrant des problèmes de diffusion, cela peut nécessiter de très longues trajectoires. Pour regarder plus en détail les différences observées, nous allons mesurer le nombre d'échanges ainsi que le temps de résidence moyen d'une molécule d'eau dans la première sphère de coordinence.

III.3.5 Echanges et temps de résidence

L'évolution temporelle de la coordinence en fonction du temps définie équation III.3 donne des valeurs continues. Pour compter le nombre d'échanges le long d'une trajectoire, c'est à dire le passage d'une coordinence à l'autre, il est nécessaire de définir un critère. On considère alors qu'il y a changement de coordinence lorsque la coordinence passe d'une valeur entière X à une valeur $X \pm 0.8$. Par exemple, si le système se trouve dans une configuration où la coordinence est égale à 6, on compte un échange si la coordinence à un temps supérieure a pour valeur 6.8 ou 5.2. Dans ce cas, la nouvelle coordinence est prise égale à 7 ou 5 respectivement. Cela revient à "arrondir" $CN(t)$, avec un critère de 0.8, et on note les nouvelles valeurs obtenues $CN_a(t)$. Pour la configuration au temps $t = 0ps$, la coordinence est prise égale à la valeur entière la plus proche. Un exemple de la modification obtenue est représenté figure III.13 pour la fonctionnelle BLYP-D2 à 300K. A partir des courbes de $CN_a(t)$ obtenues, on compte le nombre d'échanges le long d'une trajectoire N_{ex} en comptant le nombre de fois où la coordinence change. Dans ce cas, on fait aussi l'hypothèse qu'il n'y a pas de cas où une molécule d'eau entre dans la sphère de coordinence et une autre en sort, en même temps. Cette hypothèse a été testée sur plusieurs trajectoires.

Les valeurs des N_{ex} sont reportées tableau III.6. Comme on pouvait s'y attendre,

Fonctionnelle	T(K)	t_{simu} (ps)	CN ⁽¹⁾	CN ⁽²⁾	N_{ex}	MRT(ps)
PBEsol	400	100	6.05	6.05	51	9.84
PBE	300	30	6.14	6.11	17	10.84
	400	100	6.12	6.11	54	11.33
BLYP	300	30	6.03	6.03	1	180.89
	400	70	6.16	6.15	33	18.67
PBE-D2	400	100	6.49	6.46	136	4.77
BLYP-D2	300	100	6.50	6.51	29	22.41
	350	100	6.40	6.39	65	9.84
	400	100	6.60	6.60	126	5.23
	450	100	6.66	6.65	171	3.89
	500	100	6.69	6.68	214	3.15
	550	100	6.68	6.75	308	2.19
	600	100	6.83	6.82	388	1.77
vdW-DF2	300	100	6.54	6.54	33	19.82
	400	100	7.02	6.97	146	4.81

Table III.6: Coordinences CN⁽¹⁾ et CN⁽²⁾, nombres d'échanges N_{ex} et temps de résidence moyen MRT des trajectoires effectuées à différentes températures, T, pour les différentes fonctionnelles. Le temps de simulations est t_{simu} .

N_{ex} augmente avec la température et est plus élevé pour les trajectoires où l'eau diffuse le plus. Les écarts de valeurs sont importants avec seulement un échange compté pour la trajectoire BLYP à 300K et 388 pour celle avec BLYP-D2 à 600K. A partir de ces échanges, il est possible de déterminer le temps de résidence moyen, MRT , en utilisant la méthode développée par [59] :

$$MRT = \frac{CN^{(1)}t_{simu}}{N_{ex}} \quad (\text{III.5})$$

où t_{simu} est le temps de simulation.

On note que dans la littérature, il est d'usage de considérer qu'un échange est "réel" si la molécule d'eau reste dans son nouveau état, c'est à dire dans la sphère de coordinence ou en dehors de celle-ci, plus longtemps qu'un temps caractéristique t^* . Empêchant ainsi de compter des molécules qui ne sont pas entièrement entrées ou sorties de la sphère de coordinence. [64] propose de prendre ce temps égal à 2ps, correspondant au temps de réorientation de la molécule d'eau. Néanmoins, comme ce temps dépend de la température, nous avons choisi de le prendre égal à 0 par soucis de comparaison.

Les temps de résidence trouvés varient de ~ 5 ps à ~ 20 ps pour les trajectoires entre 300K et 400K avec une valeur de 180ps pour la trajectoire avec la fonctionnelle BLYP à 300K. Si on admet qu'il est nécessaire d'observer un nombre important

d'échanges, de l'ordre de la centaine, pour obtenir une valeur de la coordinence correcte, cela demanderait, dans la majeure partie des cas, des trajectoires de plusieurs centaines de picosecondes. A plus haute température, les MRTs sont de l'ordre de quelques picosecondes et on peut alors espérer que dans ces cas, la convergence soit atteinte.

Pour tenter une analyse plus quantitative et espérer trouver un critère de convergence, nous avons développé une nouvelle approche du calcul de l'incertitude de la coordinence.

III.3.6 Estimation de l'incertitude/critère de convergence

Le but de cette partie est double. On cherche tout d'abord un critère de convergence de la coordinence. On s'attend effectivement qu'à partir d'un certain temps de simulation, ou qu'au bout d'un certain nombre d'échanges, la valeur de la coordinence moyenne n'évolue plus et que son écart-type converge. Nous effectuons notre analyse sur les trajectoires effectuées avec la fonctionnelle BLYP-D2, à différentes températures, pour lesquelles nous avons observé plusieurs dizaines d'échanges et donc, pour lesquelles on peut supposer que l'espace des configurations a été correctement échantillonné (nous reviendrons sur cette affirmation par la suite). Cela nous permettra donc, en plus, d'étudier plus en profondeur l'influence de la température sur la valeur finale de la coordinence.

Pour cela, considérant qu'une dynamique moléculaire est simplement une méthode d'échantillonnage de l'espace des phases, on fait l'hypothèse suivante : l'évolution de la coordinence en fonction du temps $CN_a(t)$ donne un ensemble de valeurs

$$CN_a(0), CN_a(\Delta t), \dots, CN_a(s\Delta t), \dots, CN_a((N_{steps} - 1)\Delta t) \quad \text{avec } s = 0, \dots, N_{steps} - 1 \quad (\text{III.6})$$

correspondant à une chaîne de Markov.

C'est à dire, qu'à chaque pas de temps s , la valeur de la coordinence au pas $s + 1$ ne dépend que de sa valeur au pas s donné par une probabilité de transition entre une coordinence i et une coordinence j , P_{ij} . Cette probabilité de transition, pour $i \neq j$, correspond à la probabilité de franchir la barrière de potentiel entre deux valeurs de la coordinence.

A partir de cette hypothèse, et en considérant de plus que la chaîne de Markov est stationnaire, réversible et irréductible, il est possible de calculer un estimateur de l'écart-type à partir de la méthode dite "initial positive sequence" [44]. On présente ici les différentes étapes nécessaires à la construction de cet estimateur. Pour plus de détails, le lecteur est renvoyé à [44, 43].

L'écart type de la coordinence au pas s , σ_s , peut être déterminé à partir de fonctions d'autocorrelation de $CN_a(t)$:

$$\gamma_{s,k} = \gamma_{s,-k} = \frac{1}{s} \sum_{i=0}^{s-1-k} (CN_a(i\Delta t) - \langle CN_a(s\Delta t) \rangle) (CN_a((i+k)\Delta t) - \langle CN_a(s\Delta t) \rangle) \quad (\text{III.7})$$

En définissant :

$$\Gamma_{s,k} = \gamma_{s,2k} + \gamma_{s,2k+1} \quad (\text{III.8})$$

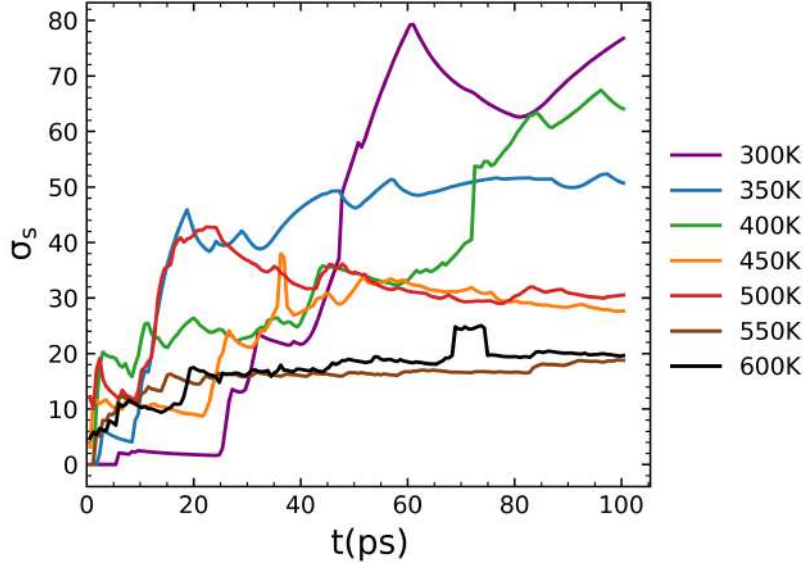


Figure III.14: Evolution de l'écart-type σ pour les trajectoires BLYP-D2 à différentes température.

on obtient σ_s en sommant Γ :

$$\sigma_s = -\gamma_{s,0} + 2 \sum_{k=1}^m \Gamma_{s,k} \quad (\text{III.9})$$

où m est choisi tel que :

$$\Gamma_{s,k} > 0 \quad k = 0, \dots, m \quad (\text{III.10})$$

La figure III.14 montre l'évolution de l'écart-type pour les trajectoires aux différentes températures. On trace aussi figure III.15 l'évolution de l'écart-type en fonction du nombre d'échanges. Lorsque l'on regarde l'évolution de l'écart-type au cours de la simulation, on s'attend à partir d'une valeur nulle (à l'instant initial il n'y a qu'une seule configuration), puis à ce que l'écart-type augmente lorsque les premiers échanges sont observés pour ensuite converger. Pour les trajectoire de 450K à 600K les écarts-types sont convergés. A 550K et 600K, la convergence est rapide, après environ 150 échanges. Pour les trajectoires à 450K et 500K, il semble qu'il soit nécessaire d'observer au moins 200 échanges.

Concernant les trajectoires de 300K à 400K, les nombres d'échanges totaux au cours des simulations étant inférieurs au nombre d'échanges nécessaires à la convergence des écarts-types des trajectoires aux plus hautes températures, les coordinences ne sont pas considérées comme convergées. Une explication possible est que les distributions des valeurs de la convergence obtenues ne sont pas stationnaires et/ou réversibles. En effet, on peut le voir, figure III.16, pour la trajectoire à 400K en regardant l'évolution de la moyenne de la coordinence au cours du temps dans les deux sens temporels de la simulation $\langle CN_a(t) \rangle$ et $\langle CN_a(t_{simu} - t) \rangle$. En effet, on devrait observer que les deux courbes, au bout d'un certain temps $t < t_{simu}$, convergent vers la même valeur. Ce qui n'est pas le cas pour la trajectoire à 400K contrairement à celle à 500K.

On peut noter qu'on observe aussi des "bump" sur les courbes des évolutions des

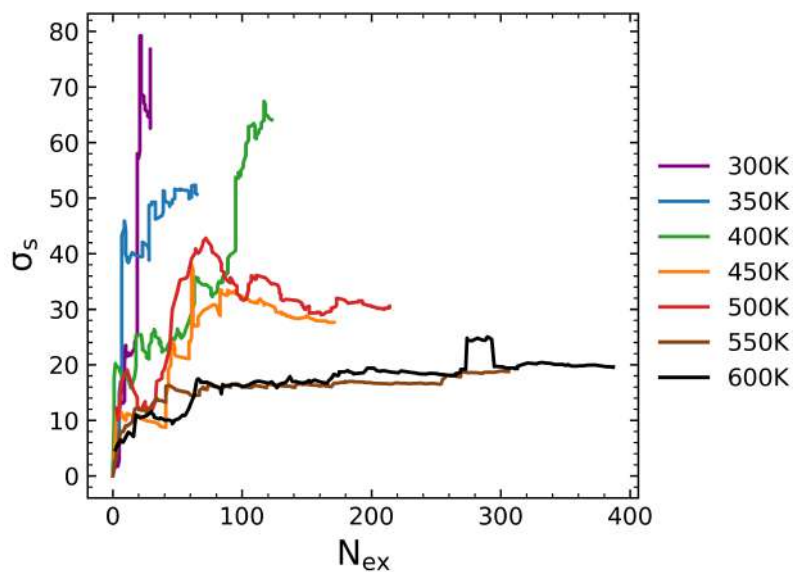


Figure III.15: Evolution de l'écart-type σ en fonction du nombre d'échanges pour les trajectoires BLYP-D2 à différentes températures.

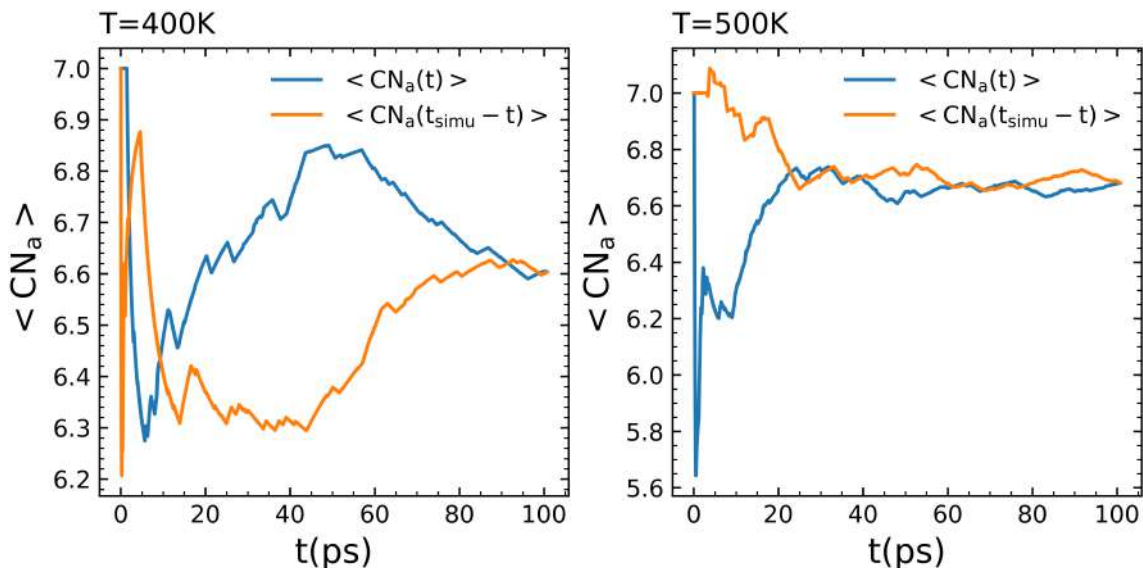


Figure III.16: Evolution de la moyenne de la coordinnce $\langle CN_a(t) \rangle$ et $\langle CN_a(t_{simu} - t) \rangle$ pour la trajectoire BLYP-D2 à 400K (gauche) et 500K (droite).

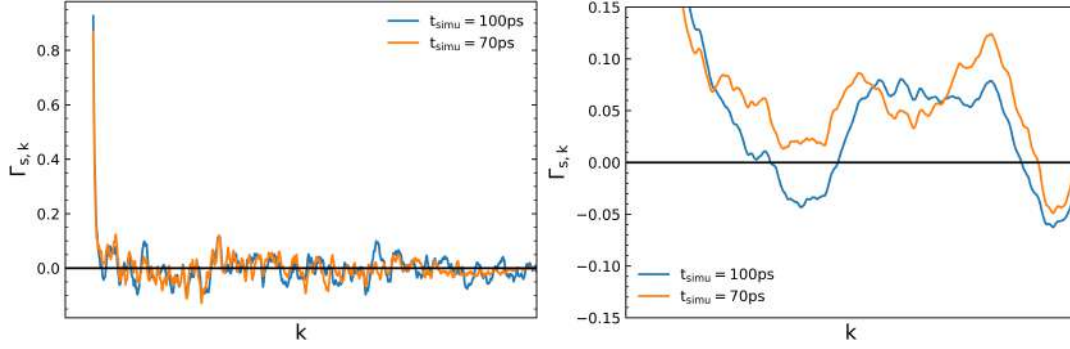


Figure III.17: $\Gamma_{s,k}$ pour calculer pour toute la trajectoire à 600K ($t_{simu} = 100ps$), et après 70ps ($t_{simu} = 70ps$).

T(K)	120ps			dernières 100ps			dernières 60ps			dernières 30ps		
	CN ⁽²⁾	σ	u(CN ⁽²⁾)	CN ⁽²⁾	σ	u(CN ⁽²⁾)	CN ⁽²⁾	σ	u(CN ⁽²⁾)	CN ⁽²⁾	σ	u(CN ⁽²⁾)
300	6.51	69.8	0.31	6.51	76.77	0.38	6.21	46.2	0.29	-	-	-
350	6.44	51.9	0.23	6.39	50.68	0.25	6.39	50.5	0.32	6.39	45.7	0.41
400	6.63	58.4	0.26	6.60	64.06	0.31	6.53	56.5	0.36	6.36	24.8	0.22
450	6.71	29.7	0.13	6.65	27.63	0.14	6.66	25.9	0.16	6.70	20.7	0.19
500	6.71	29.4	0.13	6.68	30.51	0.15	6.73	30.6	0.19	6.77	32.0	0.29
550	6.72	20.5	0.09	6.75	18.76	0.09	6.66	21.4	0.14	6.65	24.7	0.22
600	6.86	21.7	0.10	6.82	19.61	0.10	6.81	22.2	0.14	6.90	23.6	0.21

Table III.7: Coordinences moyennes (CN) et leurs écart-types σ obtenues avec la fonctionnelle BLYP-D2 pour différentes températures T. Les valeurs sont calculées en prenant en compte toutes la trajectoire, en enlevant les 20 premières picosecondes, sur la dernière moitié de la trajectoire et sur le dernier quart de la trajectoire. A 300K, les valeurs pour le dernier quart de la trajectoire ne sont pas calculées puisque seulement 3 échanges sont observés sur cette intervalle. Les incertitudes $u(CN^{(2)})$, ont été calculées en utilisant la relation $2\sigma/\sqrt{N_{steps}}$, où N_{steps} .

écarts-types, par exemple à environ 70ps pour la trajectoire à 600K ou à environ 35ps sur la trajectoire à 450K. Cela vient du fait que les $\Gamma_{s,k}$ décroissent jusqu'à devenir bruitées. Comme la somme equation III.9 se fait sur les valeurs positives de $\Gamma_{s,k}$, il se peut que dans certain cas, la courbe passe par zero après le début de la zone bruitée entraînant un "saut" sur les courbes de $\sigma_s(t)$ (voir figure III.17).

Pour terminer l'étude de la convergence de la coordinnce, on calcule les valeurs de la coordinnce moyenne et de son écart-type σ (la dernière valeur de σ_s) sur différents intervalles de temps de la trajectoire (120ps, 100ps, 60ps et 30ps). Les valeurs sont reportées tableau III.7. Si l'échantillonnage des configurations de différentes coordinnces n'est pas efficace, la valeur de la coordinnce moyenne devrait changer significativement pour les différents intervalles de temps, notamment entre le début et la fin de la trajectoire où la trajectoire devrait être plus équilibrée. A 400K, la différence entre la coordinnce moyenne prise sur l'ensemble de la trajectoire, qui est de 6.6, et sur les dernières 30 picosecondes, qui est de 6.4, semble indiquer un échantillonnage insuffisant. La variation de la valeur de l'écart-type est encore plus drastique. Le même phénomène est observé à 300K. A l'inverse, les trajectoires à

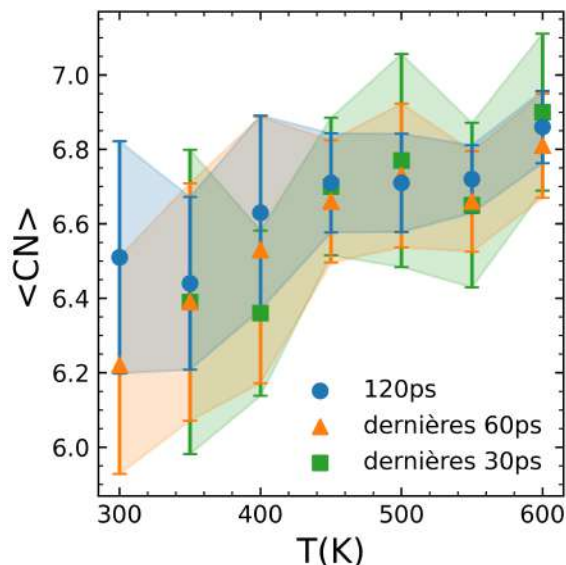


Figure III.18: Evolution de la coordinnence moyenne en fonction de la température pour la fonctionnelle BLYP-D2 calculé sur toute la trajectoire (120ps), sur la dernière moitié de la trajectoire et le dernier quart de la trajectoire. Les zones de couleurs correspondent à l'incertitude calculée comme étant égale à $2\sigma/\sqrt{N_{steps}}$, où N_{steps} est le nombre de pas de la trajectoire sur l'intervalle de temps considéré.

hautes températures montrent des variations de la coordinnence moyenne et de l'écart-type plus faible, suggérant que les valeurs obtenues sont convergées. A 350K, les valeurs varient peu mais, compte tenu du doute sur la convergence de la coordinnence pour la trajectoire à 400K, il est difficile de conclure que la trajectoire à 350K est équilibrée.

La figure III.18 montre la variation de la coordinnence moyenne $CN^{(2)}$ en fonction de la température, calculée pour différents intervalles de temps, avec son incertitude calculée comme étant égale à $2\sigma/\sqrt{N_{steps}}$, où N_{steps} est le nombre de pas de la trajectoire sur l'intervalle de temps considéré. Bien que la coordinnence semble augmenter avec la température, compte tenu des remarques sur l'échantillonnage insuffisant des trajectoires aux températures inférieures à 450K, il est difficile de conclure sur une dépendance de la coordinnence avec la température.

Compte tenu de ces résultats, on peut en déduire que la coordinnence moyenne est comprise entre 6 et 7, avec un intervalle de confiance de ± 0.2 pour des températures supérieures à 450K et, ± 0.3 pour des températures inférieures.

III.4 Conclusion et perspectives

L'étude des fonctions de distribution radiale entre le calcium et les oxygènes seule ne permet pas de discriminer la meilleure fonctionnelle pour décrire la sphère de coordination du calcium. Cependant, pour décrire correctement l'environnement du calcium aqueux, il faut d'abord reproduire les propriétés de l'eau. Les fonctions de distribution radiale oxygène-oxygène à 300K montrent que, sauf pour la fonctionnelle PBEVdW, l'eau est sur-structurée. Les fonctions de distribution quadratique

moyenne indiquent que cette sur-structuration est liée à un problème de diffusion. Nous avons tenté de rendre compte de l'effet de la température sur les coefficients de diffusion par une loi d'Arrhenius, mais cela montre essentiellement que ces coefficients de diffusion ne suivent pas cette loi entre 300 et 600K.

A 400K, les fonctions de distribution radiale oxygène-oxygène obtenues avec les fonctionnelles PBE et BLYP sont en meilleur accord avec les données expérimentales notamment lorsque des termes de dispersion, Grimme D2, sont utilisées. La trajectoire avec la fonctionnelle BLYP-D2 donne des résultats légèrement meilleurs. Ceci nous donne un argument pour dire que, pour l'eau pure, la fonctionnelle GGA qui décrit le mieux le système est BLYP-D2. De petites différences par rapport aux valeurs expérimentales peuvent être affectées par le traitement classique des noyaux dans nos simulations. Cela pourrait aussi affecter la valeur de la coordinence mais demande, par exemple, d'effectuer des dynamiques moléculaires d'intégrales de chemin qui sont très coûteuses.

Le choix de la fonctionnelle utilisée est important pour la description de la sphère de coordinence du calcium puisque nous avons observé que les fonctionnelles donnant une eau sur-structurée donnaient aussi des distances Ca-O (proches voisins) plus faibles et donc diminuaient la valeur de la coordinence.

Une fois la fonctionnelle BLYP-D2 choisie comme étant celle qui donnait de meilleurs résultats, nous avons étudié l'environnement du calcium à différentes températures. Nos résultats montrent qu'obtenir une valeur convergée de la coordinence à basse température est difficile. L'utilisation d'outils d'analyse basés sur les chaînes de Markov nous a montré qu'à hautes températures, il était nécessaire d'observer au moins 200 échanges pour obtenir une valeur fiable. Si on peut étendre cette observation aux autres fonctionnelles, cela signifierait donc, compte tenu des valeurs de MRTs trouvées, qu'il serait nécessaire d'avoir des temps de simulations de l'ordre de quelques nanosecondes pour des températures $\lesssim 400K$. On note que les temps de simulations du calcium en solution aqueuse trouvés dans la littérature ne s'approchent jamais d'une telle valeur, émettant un doute quant à fiabilité des résultats. Le problème de convergence soulevé peut aussi être vu à partir des résultats de [124]. Pour la même fonctionnelle, PBE-D3, des trajectoires à 300K de 50ps, 200ps, 500ps donnent une coordinence de 6.0, alors qu'il est admis que le calcium a plusieurs complexes aqueux stables. On peut donc se demander si le nombre d'échanges de ces simulations était suffisamment grand pour obtenir un bon échantillonnage des configurations.

L'utilisation de potentiels de type "machine learning" pourrait nous permettre d'effectuer de longues simulations. Une autre méthode serait aussi d'effectuer des simulations avec une dynamique contrainte comme cela a été fait par [5]. Ces simulations permettent aussi d'avoir accès à l'énergie libre en fonction de la coordinence montrant plusieurs minima locaux pour des valeurs de coordinence entre 5 et 8. Ces résultats peuvent être comparés à ceux trouvés par [63] pour l'ion Mg^{2+} . En effet, ils trouvent qu'une seule coordinence est accessible à température ambiante, contrairement au calcium, et les problèmes soulevés dans cette partie seraient donc non pertinents pour cet ion.

Pour les autres ions formant plusieurs complexes stables, comme Na^+ ou K^+ , les MRTs trouvées dans la littérature [124] sont bien plus faibles que pour le calcium

permettant une convergence de la coordinence pour des temps de simulations bien plus faibles. A titre de comparaison, [124] trouvent des MRTs allant de 1.0ps à 1.3ps pour Na^+ et 0.4ps à 0.5ps pour K^+ dépendant de la fonctionnelle utilisée, alors que qu'ils trouvent des MRTs allant de 12.7ps à 71.7ps pour les différentes fonctionnelles considérées pour leur étude de Ca^{2+} .

Au vu des résultats de notre étude, nous considérons que les valeurs de la coordinence obtenues pour la fonctionnelles BLYP-D2 sont convergées pour des températures supérieures à 450K. On estime ces valeurs à 6.82, 6.75, 6.68 et 6.65, à 600K, 550K, 500K et 450K respectivement. Aux températures plus basses, il n'est pas possible à partir de nos simulations d'obtenir une valeur fiable. Néanmoins, basé sur nos valeurs non-convergées à 300K, 350k et 400K, la température semble avoir un effet sur la coordinence. La valeur de CN à 300K pourrait alors être légèrement inférieure à celle trouvée.

Génération d'un potentiel machine learning pour le calcium en solution aqueuse

Ce chapitre est consacré à la génération d'un potentiel machine learning (ML) pour le calcium en solution aqueuse. Ce potentiel ML sera utilisé pour les dynamiques moléculaires d'intégrales de chemin à partir desquelles nous calculerons le fractionnement isotopique. On présente, dans une première partie, IV.1, l'utilisation du code MLACS permettant de générer un potentiel ML à moindre coup à l'aide d'un échantillonnage efficace de l'espace des phases de l'ensemble canonique. Bien que cette méthode nous ait donné de bons premiers résultats, nous verrons qu'elle a encore certaines limites, notamment sur le choix des descripteurs et pour les systèmes composés de plusieurs éléments chimiques. C'est pourquoi, on développe, dans une deuxième partie, IV.2, la génération d'un potentiel ML à partir du code N2P2 basé sur les réseaux de neurones. Cette méthode nous a donné des résultats proches de ceux de l'ab-initio. Une dernière partie, IV.3, est alors consacrée à l'utilisation de ce potentiel ML pour des dynamiques moléculaires (MLMD) et à la comparaison des résultats obtenus avec ceux de l'AIMD. On verra que la possibilité d'effectuer de longues trajectoires MLMD permet de compléter l'étude faite au chapitre III.

Dans la suite, tous les calculs DFT sont faits avec la fonctionnelle BLYP-D2 avec les mêmes paramètres qu'au chapitre III. De même, lorsque l'on parlera des résultats AIMD, on fera référence aux différentes trajectoires obtenues avec la fonctionnelle BLYP-D2.

Table des matières

IV.1 Echantillonnage canonique assisté par Machine Learning (MLACS)	92
IV.1.1 Protocole	92
IV.1.2 Résultats	94
IV.2 Potentiel Machine learning et réseau de neurones	103
IV.2.1 Choix des paramètres	103
IV.2.2 Entraînements et premiers résultats	115
IV.3 Coordinence du calcium	118
IV.3.1 Protocole	118
IV.3.2 Résultats	118

IV.1 Echantillonnage canonique assisté par Machine Learning (MLACS)

La construction d'un potentiel ML dépend grandement du choix des configurations de la base de données. Dans cette partie, on présente la génération d'un potentiel ML à l'aide du code MLACS [21], qui est une méthode d'apprentissage actif, servant à obtenir rapidement une base de données représentative de la distribution canonique. Cet échantillonnage se fait à l'aide d'une procédure auto-cohérente. Il est ensuite possible de récupérer le potentiel ML de la dernière boucle auto-cohérente (voir partie I.1.3.4). Les problèmes d'échantillonnage vus au chapitre III font de MLACS une méthode de choix dans le but d'obtenir un potentiel ML pour notre système.

IV.1.1 Protocole

Le code MLACS a déjà montré son efficacité pour des solides tel que le silicium et le zirconium et pour l'uranium liquide [21]. Néanmoins, le code n'étant pas adapté aux cas où le système est composé de molécules, nous avons dû, dans un premier temps, implémenter cette possibilité.

Dans notre cas, la base de données de départ est composée d'une seule configuration d'un calcium Ca^{2+} et 62 molécules d'eau dans une boîte cubique de côté égale à 12.41Å, qui est le même système qu'au chapitre III.

Lors de la première boucle auto-cohérente, un premier potentiel ML est généré à partir de cette seule configuration. Ce potentiel donne, au vu de la taille de la base de données, de mauvais résultats, et les MLMD effectuées avec celui-ci ne sont pas stables.

Le code MLACS offre deux possibilités pour résoudre ce problème. Tout d'abord, il est possible de définir un potentiel de référence. L'ajustement du potentiel ML se fait alors sur la différence entre les valeurs des énergies et des forces obtenues avec ce potentiel de référence et celles des calculs DFT. Comme potentiel de référence, nous avons choisi le modèle SPC/fw pour l'eau et un potentiel de type Lennard-Jones pour l'interaction Ca-O (voir tableau I.1). Les MLMDs se font alors avec un potentiel qui est la somme du potentiel de référence et du potentiel ML.

Le code MLACS possède ensuite deux modes: équilibration et production. On commence d'abord la simulation en mode équilibration et après un certain nombre de boucles auto-cohérentes, N_{confeq} , on rentre en mode production pour $N_{confprod}$ boucles donnant un total de $N_{boucle} = N_{confeq} + N_{confprod}$ boucles auto-cohérentes. La différence entre ces deux modes est le nombre de pas de la MLMD servant à l'obtention de nouvelles configurations. On récupère en effet la dernière configuration de la MLMD qui est ensuite ajoutée à la base de donnée à chaque fin de boucle. Comme au début de la simulation, les potentiels sont générés avec une base de données réduites, on choisit un nombre de pas faible lorsque l'on est en phase

n°équilibration	N_{confeq}	$N_{stepseq}$	Δt (fs)
1	50	50	0.01
2	50	50	0.05
3	50	100	0.05
4	50	100	0.1
5	50	250	0.1
6	50	500	0.1

Table IV.1: Ensemble des valeurs utilisées pour les équilibrations multiples et successives.

d'équilibration. Une fois la base de données suffisamment grande pour obtenir un potentiel ML "convenable", on choisit un nombre de pas grand pour les MLMDs effectuées en phase de production afin d'obtenir le meilleur échantillonnage possible.

Après plusieurs tests, même en procédant ainsi, des problèmes de stabilité des MLMDs ont toujours été observés. Nous avons alors choisi d'effectuer des équilibrations multiples. Cette méthode nous permet alors aussi de jouer sur le pas de temps des MLMDs. La logique est la suivante. Supposons que l'on veuille que, lors de la phase de production, le pas de temps des MLMD soit de $\Delta t = 0.1 fs$ pour un nombre de pas total, $N_{stepsprod}$, d'au moins 1000. On commence une première équilibration pour obtenir un nombre, N_{confeq} , de configurations obtenues à partir de MLMD, ayant comme paramètres un petit Δt et un faible nombre de pas $N_{stepseq}$. On fait alors de nouvelles équilibrations en augmentant progressivement Δt et $N_{stepseq}$ jusqu'à des valeurs proches de celles de la production. On présente tableau IV.1 les valeurs utilisées pour les paramètres des équilibrations.

La version du code MLACS utilisée propose deux types de descripteurs, SO(4) et SO(3) et l'ajustement des potentiels ML se fait par regression linéaire. Nous avons testé ces deux descripteurs avec une valeur typique de $j = 8$ pour les descripteurs SO(4). Plusieurs valeurs de n_{max} , l et α ont été testées pour les descripteurs SO(3) (voir partie I.1.3.2). Dans les deux cas, on utilise un R_c de 6\AA .

Pour juger de la qualité du potentiel ML obtenu, on peut suivre l'évolution des RMSEs de l'énergie par atome et des forces au cours de la procédure auto-cohérente. On souhaite que les RMSEs convergent vers des valeurs les plus faibles possibles. On s'intéressera en particulier à la RMSE(F) des forces car ces dernières sont utilisées pour le calcul du fractionnement isotopique. A titre de comparaison, les forces s'exerçant sur les atomes de notre système sont typiquement de l'ordre de $0.5\text{-}2.0 eV/\text{\AA}$.

Notre grandeur d'intérêt étant la coordinence du calcium, on peut aussi juger du bon échantillonnage fait par MLACS en regardant les fonctions de distribution radiale des configurations de la base de données. Effectivement, si la base de données ne donne pas une coordinence moyenne proche de celle obtenue lors de dynamiques moléculaires AIMD, il est peu probable que le potentiel puisse reproduire correcte-

ment l'environnement du calcium et il en va de même pour l'eau. Enfin, il est possible d'effectuer des MLMDs avec le potentiel ML obtenu lors de la dernière boucle auto-cohérente.

Typiquement, les auteurs de [21] montrent que, grâce à leur méthode, un nombre réduit de boucles auto-cohérentes est nécessaire pour obtenir un bon échantillonnage (environ 100). Dans notre cas, nous sommes allés jusqu'à 800 configurations (300 d'équilibration et 500 de production).

Dans la suite, toutes les MLMDs ont été effectuées avec le code LAMMPS dans l'ensemble NVT à une température de 400K avec des thermostats de Nose-Hoover de longueur de chaînes égales à 5. Le choix de la température a été fait dans le but de pouvoir comparer les résultats avec ceux de l'AIMD pour lesquels on estime que l'échantillonnage a été suffisant.

IV.1.2 Résultats

On présente dans un premier temps les résultats obtenus avec les descripteurs SO(4) puis ceux obtenus avec le descripteurs SO(3).

IV.1.2.1 Descripteurs SO(4)

Nous présentons deux cas pour lesquels des descripteurs de type SO(4) ont été utilisés. Dans les deux cas, on réalise un même ensemble d'équilibrations multiples. Après avoir obtenu les 300 N_{confeq} de l'équilibration, on lance 500 boucles auto-cohérentes en mode production. Dans un cas, le nombre de pas des MLMDs de la production est égale à $N_{stepsprod} = 1000$ et pour le second cas on choisit un nombre de pas de $N_{stepsprod} = 10000$. Pour les deux productions, le pas de temps des MLMDs est de $\Delta t = 0.1 fs$.

Les figures IV.1 montrent les RMSEs des énergies par atome, RMSE(E), et des forces, RMSE(F), obtenues pour les deux simulations. On observe d'abord une augmentation des RMSEs, dû au fait que le nombre de configurations dans la base de données augmente et que les premiers potentiels ML sont mauvais, puis les RMSE convergent. On obtient une valeur finale pour la RMSE(E) de 0.0022eV/at pour les deux simulations et une valeur finale de la RMSE(F) de 0.2090eV/Å pour la simulation avec $N_{stepsprod} = 1000$, plus faible que pour la simulation avec $N_{stepsprod} = 10000$ qui est de 0.2148eV/Å. Cela correspond au bon ordre de grandeur des valeurs des RMSEs obtenues lors de la construction d'un potentiel ML.

Il est aussi possible de comparer les deux simulations en regardant les distributions de fonction radiale, $g(r)$, des configurations obtenues lors de la production afin de vérifier le bon échantillonnage. On les compare de plus aux $g_{CaO}(r)$ et $g_{OO}(r)$ obtenues par AIMD du chapitre III et à celles obtenues à partir d'une dynamique moléculaire utilisant les potentiels de référence, utilisant les mêmes paramètres que les MLMDs et d'une durée de 100ps. Les deux simulations donnent des $g(r)$ très similaires. Concernant les $g_{CaO}(r)$, elles sont bien meilleures que celles obtenues avec le potentiel de référence. Elles donnent néanmoins une coordinence un peu plus faible que celle obtenue par AIMD et la première sphère de coordination est plus structurée. En revanche, bien que la position du premier maximum de la $g_{OO}(r)$

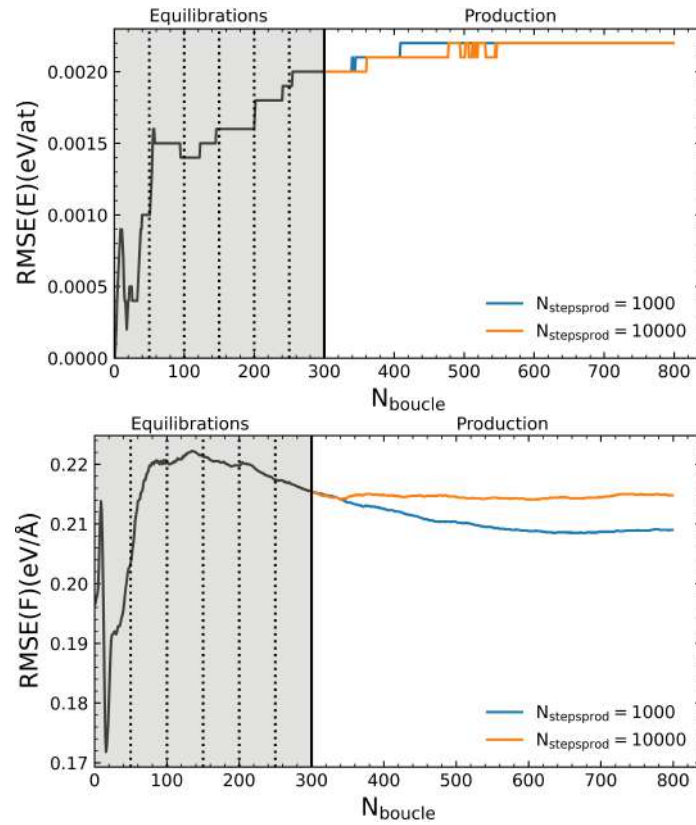


Figure IV.1: RMSE de l'énergie et des forces obtenues pour une même équilibration multiple et deux productions avec un nombre de pas de MLMD différent.

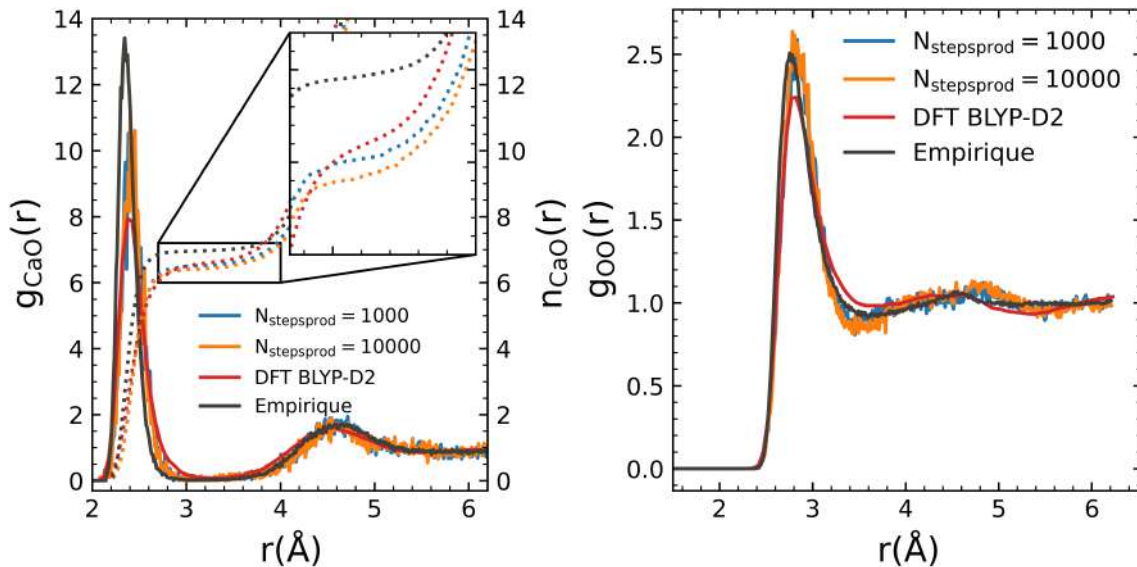


Figure IV.2: Fonction de distribution radiale Ca-O et O-O des configurations obtenues lors des productions avec $N_{stepsprod} = 1000$ et $N_{stepsprod} = 10000$. Elles sont comparés aux résultats de la dynamique moléculaire ab-initio et de la dynamique moléculaire le potentiel de référence. La température est fixée à 400K.

soit proche de celle obtenue par AIMD, on obtient une eau plus structurée avec des amplitudes des extrema plus grandes qu'en AIMD.

Pour juger plus précisément de la qualité du potentiel, on a effectué une MLMD de 500ps avec le dernier potentiel de la simulation avec $N_{stepsprod} = 1000$ qui donnait une meilleure RMSE(F). L'ensemble des $g(r)$ obtenues se trouvent figure IV.3. Bien que la $g_{CaO}(r)$ montre que l'on obtient une valeur de la coordinence de 6.57 très proche de la valeur AIMD qui est de 6.60, on observe, là encore, que la première sphère de coordinence est moins large et a une plus grande amplitude. Le potentiel ML n'est alors pas en mesure de reproduire l'environnement du calcium. Cela est encore plus visible lorsque l'on regarde la $g_{CaH}(r)$. Concernant l'eau, le potentiel ML donne des positions des extrema des $g_{OO}(r)$ et $g_{OH}(r)$ proches de ceux obtenus en AIMD sauf pour la $g_{HH}(r)$ où la position du second maximum est la même que celle donnée par le potentiel de référence.

Les amplitudes des extrema de la $g_{OO}(r)$ montrent que la MLMD donne une eau plus structurée qu'en AIMD.

Au vu des ces résultats, il est raisonnable de penser qu'il est possible d'obtenir un potentiel ML capable de donner des résultats proches de ce que l'on obtient avec la fonctionnelle BLYP-D2. Néanmoins, la sur-structuration de l'eau, qui comme nous l'avons vu au chapitre III est liée à des problèmes de diffusion et d'échantillonnage, ainsi que le fait que la première sphère de coordinence du calcium est moins étendue qu'en ab-initio suscite des doutes quant à la possibilité de reproduire correctement les propriétés du fractionnement isotopique avec ce potentiel. D'autres tests, avec différents paramètres, ne nous ont pas permis d'améliorer ces résultats et nous avons donc cherché à obtenir un meilleur potentiel ML en testant un autre type de descripteurs : les descripteurs SO(3).

IV.1.2.2 Descripteurs SO(3)

L'utilisation de descripteurs de type SO(3) nécessite la donnée de trois paramètres: n_{max} , l_{max} et α (voir partir I.1.3.2). Le code MLACS utilise comme paramètres par défaut: $n_{max} = 4$, $l_{max} = 3$ et $\alpha = 1.0$. Afin de choisir le jeu de paramètres le plus approprié, nous avons étudié l'influence de la variation d'un paramètre en laissant les autres à leurs valeurs par défaut.

Variation de n_{max}

Les RMSEs des énergies par atome et des forces obtenues pour des simulations avec comme valeur $n_{max} = 3, 4, 5, 6$ se trouvent figure IV.4. Elles montrent que plus la valeur de n_{max} est élevée, plus les RMSEs convergent vers de faibles valeurs.

Les fonctions de distribution radiale des configurations obtenues lors de la production se trouvent figure IV.5 et sont comparées à celles obtenues en AIMD et avec le potentiel de référence. Le nombre limité de configurations rend les $g(r)$ bruitées et la comparaison difficile. Néanmoins, peu de différences sont observées pour les différentes $g_{CaO}(r)$. La simulation avec $n_{max} = 6$ donne toute de même des amplitudes des extrema des $g_{OO}(r)$ en meilleur accord avec l'AIMD.

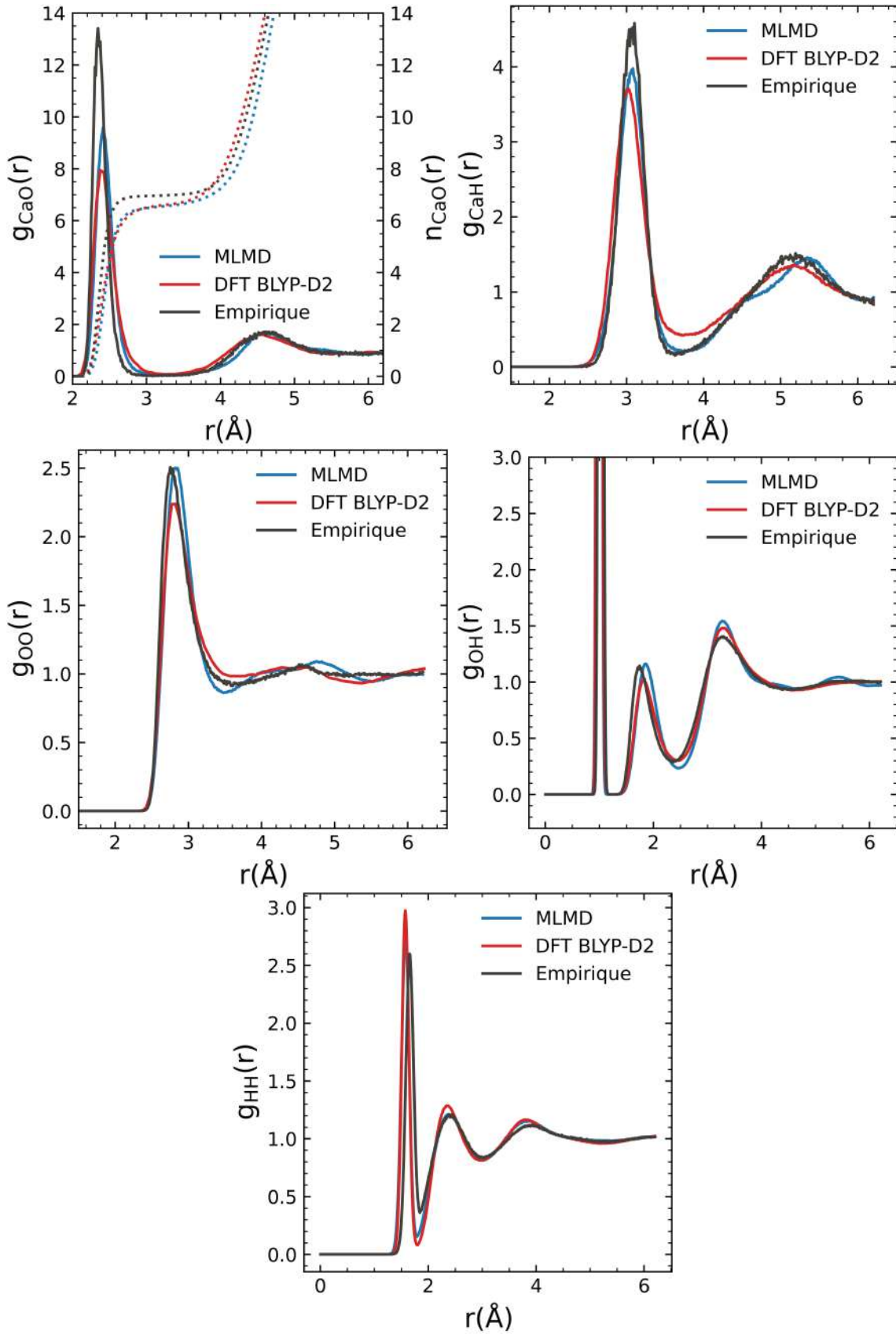


Figure IV.3: Fonction de distribution radiale obtenues à partir de la MLMD à 400K avec le potentiel final de la simulation avec $N_{\text{stepsprod}} = 1000$. Elles sont comparées aux résultats de la dynamique moléculaire ab-initio et de la dynamique moléculaire le potentiel de référence.

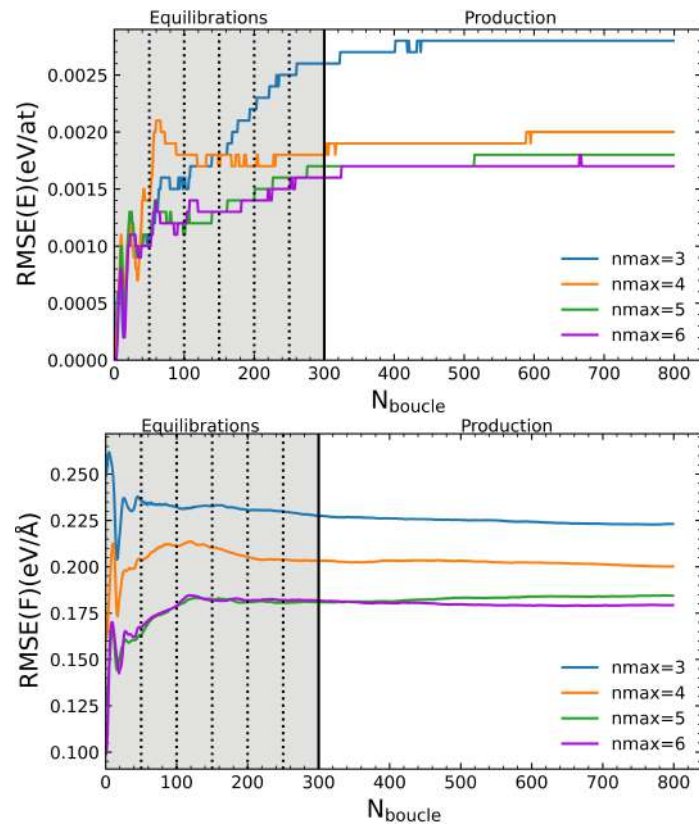


Figure IV.4: RMSE de l'énergie par atome et des forces obtenues pour $l_{max} = 3$, $\alpha = 1.0$ et différents n_{max} . Les MLMDs sont effectuées à 400K.

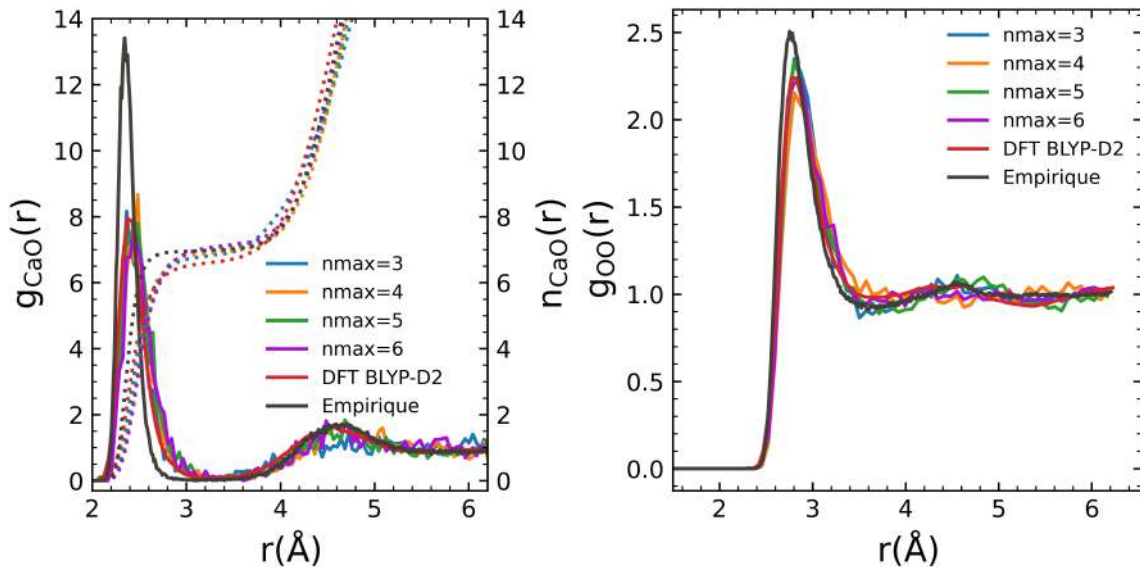


Figure IV.5: Fonctions de distribution radiale Ca-O et O-O des configurations obtenues lors des productions pour $l_{max} = 3$, $\alpha = 1.0$ et différentes valeurs de n_{max} et comparées aux résultats des dynamiques moléculaires ab-initio et avec le potentiel de référence. Les MLMDMS sont effectuées à 400K.

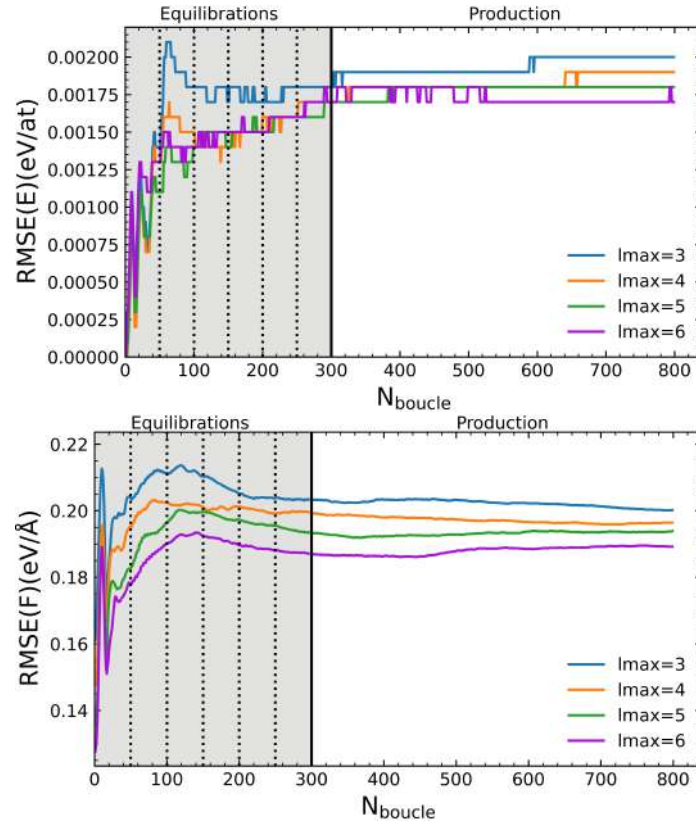


Figure IV.6: RMSE de l'énergie par atome et des forces obtenues pour $n_{max} = 4$, $\alpha = 1.0$ et différents l_{max} . Les MLMDs sont effectuées à 400K.

Variation de l

La même étude est faite pour des valeurs de $l_{max} = 3, 4, 5$ et 6 . Les RMSEs obtenues se trouvent figure IV.7 et les valeurs convergées sont plus faibles dans le cas où $l_{max} = 6$ est choisi.

Concernant les $g(r)$ des configurations échantillonnées pendant la phase de production, figure IV.7, elles montrent que plus la valeur de l_{max} est élevée plus l'eau est structurée. Pour la $g_{CaO}(r)$, il est difficile de tirer une conclusion même si la simulation à $l_{max} = 3$ semble donner une sphère de coordinence plus élargie.

Variation d' α

On teste des valeurs de $\alpha = 1.0, 1.5, 1.8, 2.0$ et 2.5 . Les RMSEs, figure IV.8, donnent des valeurs convergées comparables sauf pour $\alpha = 1$ qui donne des valeurs plus élevées.

De la même façon que pour le paramètre l_{max} , la $g_{OO}(r)$ est plus structurée plus la valeur de α est élevée. La même conclusion peut être observée sur les $g_{CaO}(r)$ même si le résultat à $\alpha = 2.0$ ne suit pas cette tendance. La valeur de $\alpha = 1.8$ donnant un meilleur accord avec l'AIMD.

Finalement, cette analyse nous a permis de déterminer que, parmi les différents tests effectués, une valeur de $n_{max} = 6$ était celle qui donnait de meilleurs résultats. Concernant les valeurs de l_{max} et α , si de faibles valeurs sont choisies pour ces paramètres

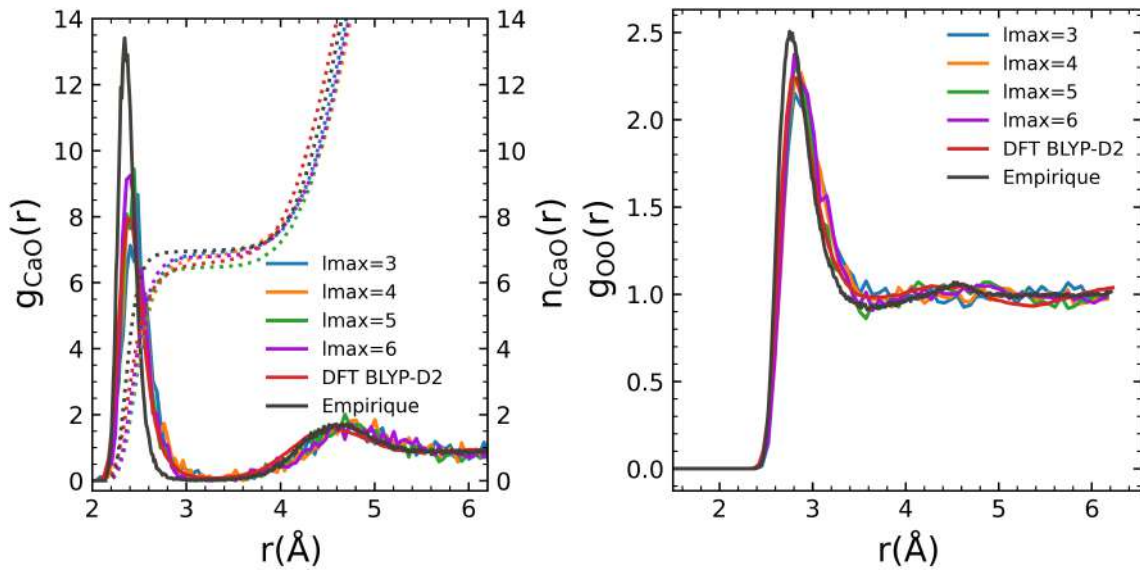


Figure IV.7: Fonctions de distribution radiale Ca-O et O-O des configurations obtenues lors des productions pour $n_{max} = 4$, $\alpha = 1.0$ et différentes valeurs de l_{max} et comparées aux résultats des dynamiques moléculaires ab-initio et avec le potentiel de référence. Les MLMDs sont effectuées à 400K.

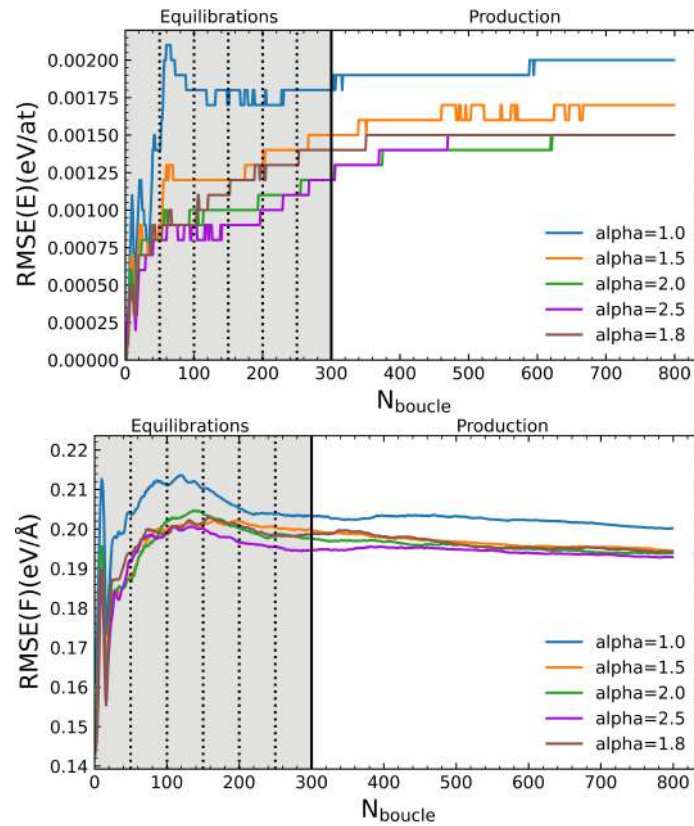


Figure IV.8: RMSE de l'énergie par atome et des forces obtenues pour $n_{max} = 4$, $l_{max} = 3$ et différents α . Les MLMDs sont effectuées à 400K.

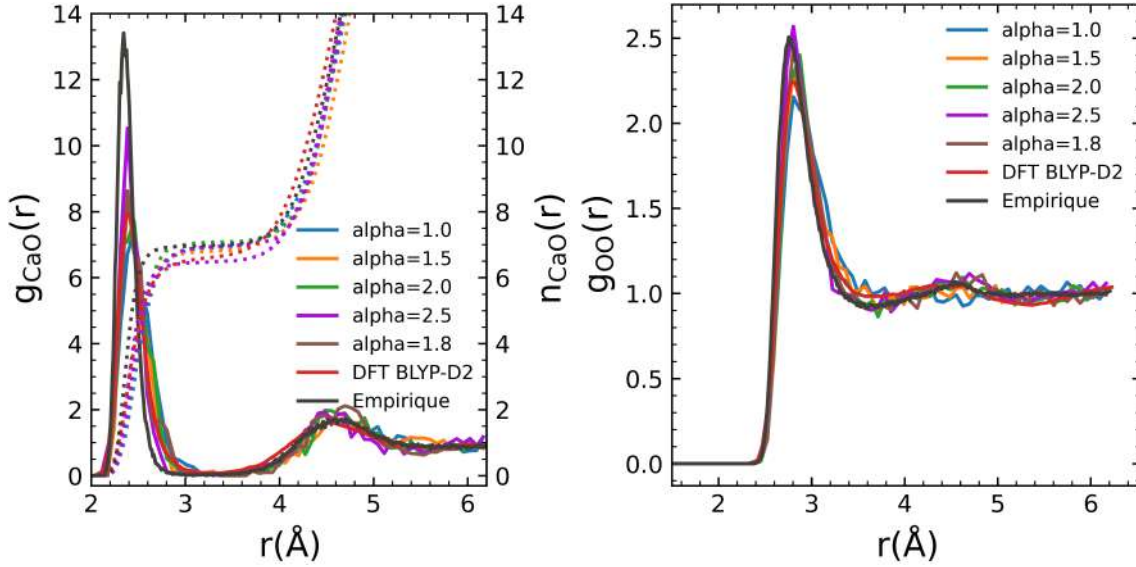


Figure IV.9: Fonctions de distribution radiale Ca-O et O-O des configurations obtenues lors des productions pour $n_{max} = 4$, $l_{max} = 3$ et différentes valeurs de α et comparées aux résultats des dynamiques moléculaires ab-initio et avec le potentiel de référence. Les MLMDs sont effectuées à 400K.

($l_{max} = 3$, $\alpha = 1.0$), l'eau est moins structurée. C'est une donnée importante suite aux conclusions tirées du chapitre précédent. Néanmoins, il semblerait qu'une valeur plus élevée, notamment pour le paramètre α , soit nécessaire pour mieux reproduire la première sphère de coordination du calcium obtenue en AIMD.

Nous avons donc choisi d'étudier un autre jeu de paramètre: $n_{max} = 6$, $l_{max} = 6$ et $\alpha = 1.8$.

Résultats avec $n_{max}, l_{max}, \alpha = (6,6,1.8)$ et $(6,3,1.0)$

Les valeurs de $n_{max} = 6$ et $l_{max} = 6$ ont été choisies car ce sont les valeurs donnant les meilleurs RMSEs. Quant à la valeur de $\alpha = 1.8$, c'est celle qui semble le mieux reproduire la première sphère de coordination du calcium. Dans la suite, nous comparons les résultats obtenus avec ceux de la simulation avec: $n_{max} = 6$, $l_{max} = 3$ et $\alpha = 1.0$, qui selon nos conclusions, est la simulation donnant l'eau la moins structurée.

Toujours selon la même procédure, les RMSEs pour ces deux simulations sont présentées figure IV.10. Les RMSEs convergent vers une valeur plus faible pour la simulation ($n_{max} = 6$, $l_{max} = 6$, $\alpha = 1.8$). Néanmoins, lorsqu'on analyse l'ensemble des $g(r)$, figure IV.11, bien que le choix de ces paramètres semble permettre de reproduire l'environnement du calcium avec des $g_{CaO}(r)$ et $g_{CaH}(r)$ proches de celles obtenues en AIMD, et une valeur de la coordination semblable, la $g_{OO}(r)$ donne une eau plus structurée. Au contraire, avec le choix de paramètres ($n_{max} = 6$, $l_{max} = 3$, $\alpha = 1.0$), on obtient une bonne $g_{OO}(r)$. Concernant, la $g_{OH}(r)$, dans les deux cas, les amplitudes des maxima sont comparables à l'AIMD, la simulation ($n_{max} = 6$, $l_{max} = 6$, $\alpha = 1.8$) a un second minimum à une distance plus grande que l'AIMD avec une amplitude plus faible. A l'inverse, bien qu'avec, là encore, un second minimum avec une amplitude faible, c'est cette simulation qui reproduit le mieux la $g_{HH}(r)$ puisque la simulation ($n_{max} = 6$, $l_{max} = 3$, $\alpha = 1.0$) donne une seconde sphère de coordination

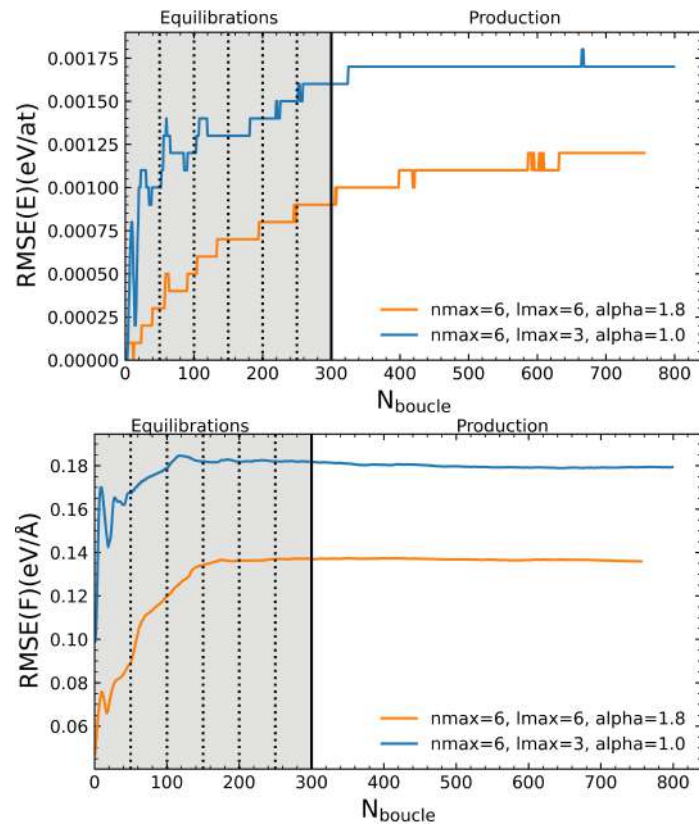


Figure IV.10: RMSE de l'énergie et des forces obtenues pour différents $n_{max} = 6$, $l_{max} = 6$, $\alpha = 1.8$ et $n_{max} = 6$, $l_{max} = 3$, $\alpha = 1.0$.

H-H quasi-identique au potentiel de référence.

A ce stade, il nous semble inutile de poursuivre d'autres tests dans le but d'améliorer les résultats. En effet, malgré les différents descripteurs, ainsi que les différents paramètres testés, nous n'obtenons pas de simulations permettant de reproduire à la fois l'eau liquide et l'environnement du calcium. Nous pensons que cela vient du fait que le code MLACS, en l'état actuel, ne permet pas d'utiliser des descripteurs différents en fonction de l'espèce chimique. Les environnements atomiques du calcium, des oxygènes et des hydrogènes étant différents, cela expliquerait les résultats.

Toutefois, le développeur de MLACS, ainsi que ses contributeurs, devraient mettre en place d'autres descripteurs et la possibilité d'utiliser des réseaux de neurones. Au vu des résultats, nous pensons aussi que le code MLACS est en mesure de fournir un potentiel de l'eau liquide bien meilleur que les potentiels empiriques, et ce, avec un coût très faible en temps de calcul.

Nous cherchons à construire un potentiel permettant de décrire correctement le fractionnement isotopique qui dépend grandement de la structure du système. C'est pourquoi nous souhaitons générer un potentiel donnant de meilleurs résultats. Néanmoins, les potentiels obtenus avec MLACS peuvent servir à des applications ne demandant pas une telle précision.

Nous avons alors choisi de générer un potentiel ML, pour notre système, à l'aide d'un réseau de neurones en utilisant le code N2P2.

IV.2 Potentiel Machine learning et réseau de neurones

L'intérêt de l'utilisation de MLACS est son aptitude à échantillonner efficacement l'espace des phases. La construction d'une base de données peut être complexe et demander de longs temps de calculs pour obtenir des configurations échantillonnant l'ensemble de la PES que l'on souhaite reproduire. Néanmoins, suite à l'étude du chapitre III, nous avons à notre disposition de longues trajectoires AIMD obtenues avec la fonctionnelle BLYP-D2 et ce pour sept températures. Nous avons donc choisi de récupérer des configurations de ces AIMD afin de constituer notre base de données et générer un potentiel machine learning basé sur les réseaux de neurones à l'aide du code N2P2 [101].

Comme dans la partie précédente, les paramètres (fonction d'activation, descripteurs, nombre de couches, ...) doivent être testés afin de déterminer le meilleur choix à effectuer.

IV.2.1 Choix des paramètres

Nous séparons les paramètres en deux catégories. Ceux correspondant au réseau de neurones, par exemple le choix du nombre de couches ou des descripteurs, et ceux correspondant aux paramètres de l'algorithme d'optimisation des poids.

Dans un premier temps, nous décrivons les paramètres testés puis nous présenterons

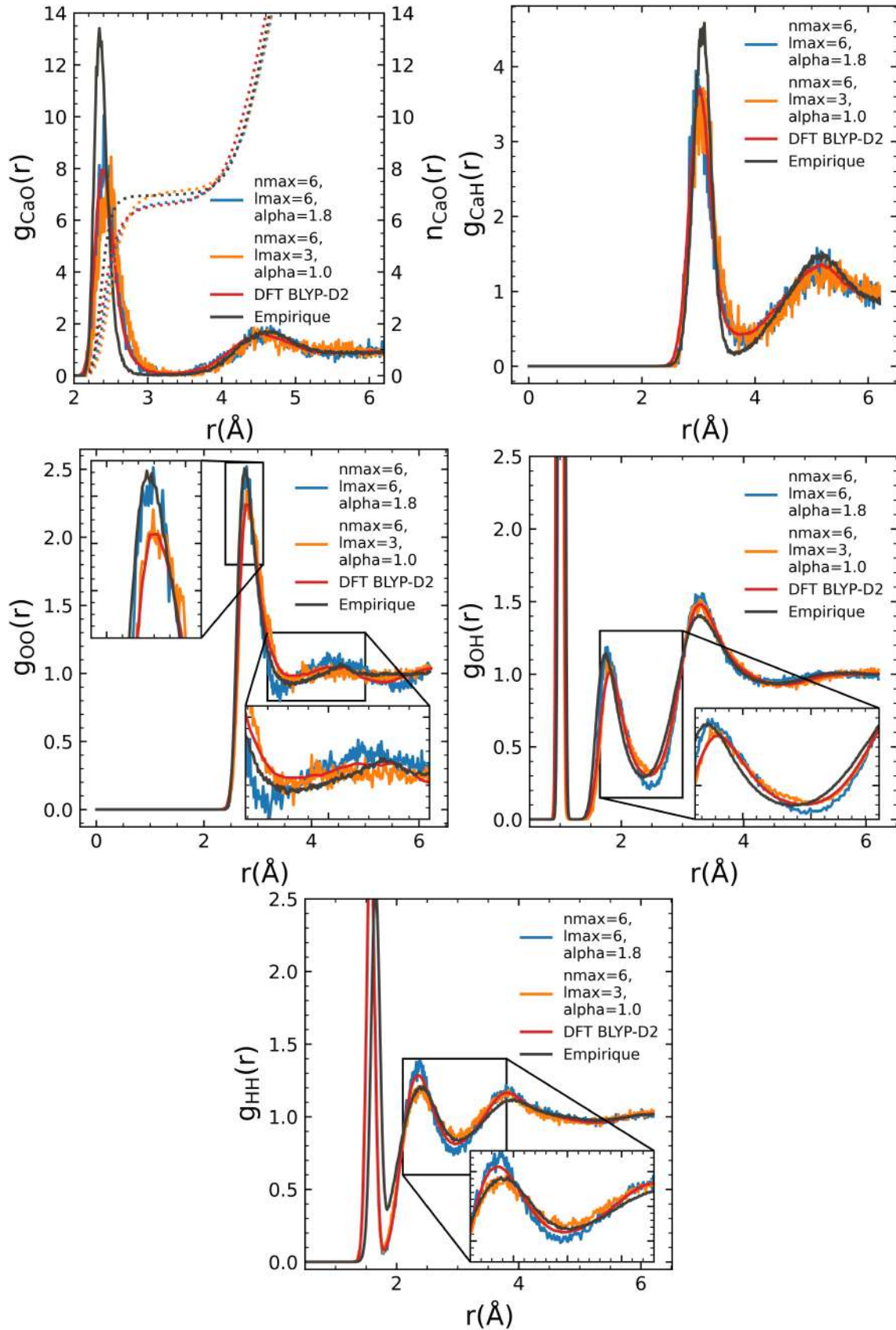


Figure IV.11: Fonctions de distribution radiale des configurations obtenues lors des productions pour différentes $n_{\text{max}} = 6$, $l_{\text{max}} = 6$, $\alpha = 1.8$ et $n_{\text{max}} = 6$, $l_{\text{max}} = 3$, $\alpha = 1.0$ et comparées aux résultats des dynamiques moléculaires ab-initio et avec le potentiel de référence.

les résultats de l'étude. La base de données utilisée pour les tests, sauf pour celui étudiant l'influence de la base de données, est composée de 600 configurations, provenant de nos AIMD, soit 100 configurations régulièrement espacées à chaque température (300K, 350K, 400K, 450K, 500K, 550K et 600K).

IV.2.1.1 Description des paramètres

Paramètres du filtre de Kalman

Pour l'optimisation des poids et des biais, \vec{w} , du réseau de neurones, nous avons utilisé un algorithme basé sur les filtres de Kalman implémenté dans N2P2 et développé par [100]. Cette méthode permet une optimisation plus efficace que celles basées sur l'algorithme du gradient. Pour une description complète de l'algorithme, le lecteur est renvoyé à [100]. On présente ici le principe général permettant de comprendre les tests effectués.

Notre base de données est composée de 600 configurations, c'est à dire 600 énergies et $600 \times 3 \times N_{\text{at}}$ forces, où N_{at} est le nombre d'atomes du système. A chaque époque e , on fournit à l'algorithme toutes les énergies et une fraction des forces choisies aléatoirement. Par exemple, dans la suite, lorsque l'on parlera d'un ratio 1:10, cela signifiera que l'algorithme d'optimisation utilise 600 énergies et 6000 forces par époque. L'algorithme traite séparément les énergies et les forces. Par époque, il va effectuer autant d'itérations qu'il y a d'énergies et de forces. Une itération correspond à :

- Calcul de l'erreur $\xi(e)$ tel que :

$$\xi(e) = E^{\text{data}} - E(\vec{w}(e)) \quad \text{pour une énergie} \quad (\text{IV.1})$$

$$\xi(e) = \beta (F^{\text{data}} - F(\vec{w}(e))) \quad \text{pour une force} \quad (\text{IV.2})$$

où β est un paramètre permettant de contrôler l'importance relative des forces par rapport aux énergies.

- Mise à jour les poids :

$$\vec{w}(e+1) = \vec{w}(e) + \vec{K}(e)\xi(e) \quad (\text{IV.3})$$

où $\vec{K}(e)$ est alors un vecteur dépendant de trois paramètres, $\eta(e)$, $q(e)$ et ϵ .

Le paramètre ϵ permet d'initialiser le vecteur \vec{K} et est pris égale à 10^{-2} . Les paramètres $\eta(e)$ et $q(e)$ varient en fonction de l'époque tels que :

$$q(e) = \max(q_0 e^{e/u\tau_q}, q_{\min}) \quad (\text{IV.4})$$

$$\eta(e) = \min(\eta_0 e^{e/u\tau_\eta}, \eta_{\max}) \quad (\text{IV.5})$$

avec u le nombre d'itérations par époque et η_0 , η_{\max} , τ_η , q_0 , q_{\min} et τ_q des paramètres. Le rôle de $q(e)$ est d'introduire un bruit, empêchant de rester bloqué dans un minimum local.

Finalement, au lieu d'effectuer chaque itération de l'algorithme à la suite, il est possible de les diviser en s taches. A chaque tache est donné un sous ensemble d'énergies et de forces et les calculs se font en parallèle.

Paramètres du réseau de neurones

Tout d'abord, il est possible de normaliser la base de données avant l'entraînement. Cette normalisation se fait telle que les énergies de la base de données normalisées, E^{data*} , ont une moyenne nulle: $\langle E^{data*} \rangle = 0$. Ensuite, un mode "ref" assure que l'écart-type des énergies et forces normalisées, F^{data*} , soient égale à 1 ($\sigma_{E^{data*}} = \sigma_{F^{data*}} = 1$). Un deuxième mode "force" assure quant à lui que les écart-types des forces normalisées ainsi que des forces prédites, F^{NN*} , lors d'une étape d'initialisation aléatoire du réseau de neurones soient égales à 1 ($\sigma_{F^{NN}} = \sigma_{F^{data*}} = 1$). Les descripteurs G peuvent être mise à l'échelle. En gardant les notations du code N2P2, il est possible d'effectuer les opérations suivantes :

- Aucune mise à l'échelle : $G_{scaled} = G$
 - ST_SCALE : $G_{scaled} = \frac{G - G_{min}}{G_{max} - G_{min}}$
 - ST_CENTER : $G_{scaled} = G - \langle G \rangle$
 - ST_SCALECENTER : $G_{scaled} = \frac{G - \langle G \rangle}{G_{max} - G_{min}}$
 - ST_SCALESIGMA : $G_{scaled} = \frac{G - \langle G \rangle}{\sigma_G}$
- où σ_G est l'écart type des valeurs du descripteur.

$\langle G \rangle$ est la moyenne des valeurs du descripteur et G_{max} et G_{min} ses valeurs maximale et minimale. Les opérations de mise à l'échelle permettent d'éviter des valeurs de la fonction d'activation proches de 1 ou -1 et d'exploiter au mieux sa non linéarité.

Quatre fonctions d'activation ont été testées :

- la fonction tangente hyperbolique, $f_a(x) = \tanh(x)$;
- la fonction softplus, $f_a(x) = \ln(1 + e^x)$;
- la fonction logistique, $f_a(x) = 1/(1 + e^{-x})$;
- et la fonction ReLU, $f_a(x) = \max(0, x)$.

Concernant les fonctions de cutoff, en gardant encore une fois les notations de N2P2, il est possible de choisir entre :

- CT_HARD : $f_c(R) = 1$;
- CT_COS : $f_c(R) = 1/2(\cos(\pi R) + 1)$;
- CT_TANHU : $f_c(R) = \tanh^3(1 - R/R_c)$;
- CT_TANH : $f_c(R) = (e^1 + 1/e^1)/(e^1 - 1/e^1) \tanh^3(1 - R/R_c)$, $f_c(0) = 1$;
- CT_EXP : $f_c(R) = \exp(1 - 1/(1 - R^2))$;
- CT_POLY1 : $f_c(R) = (2R - 3)R^2 + 1$;
- CT_POLY2 : $f_c(R) = ((12 - 6R)R - 10)R^3 + 1$;
- CT_POLY3 : $f_c(R) = (R(R(20R - 70) + 84) - 3)R^4 + 1$;
- CT_POLY4 : $f_c(R) = (R(R((35 - 70R)R - 540) + 420) - 126)R^5 + 1$.

Il est aussi possible d'effectuer une opération de normalisation à une valeur d'entrée du neurone i de la couche j , y_i^j . Cette opération consiste à diviser la combinaison linéaire des poids et valeurs des neurones de la couche $j - 1$ par le nombre de neurones de cette couche. Ainsi l'équation I.81 devient:

$$y_i^j = f_i^j \left(\frac{b_i^j + \sum_{k=1}^{N_{j-1}} w_{ki}^{j-1,j} y_k^{j-1}}{N_{j-1}} \right) \quad (\text{IV.6})$$

Finalement, après avoir testé différents descripteurs, nous regarderons l'influence du nombre de couches et du nombre de neurones par couche ainsi que l'influence de

la taille de la base de données.

L'ensemble des commandes de N2P2 à définir dans le fichier d'entrée, associées aux différents paramètres présentés, se trouvent en annexe tableau 5.

IV.2.1.2 Résultats

Pour l'ensemble des tests, on choisit le paramètre optimal, à partir des valeurs des RMSE, et on utilise ce paramètre pour les tests suivants, en faisant l'hypothèse que la détermination du choix d'un paramètre n'a pas d'influence sur les autres.

Pour commencer les tests, on choisit, arbitrairement, un réseau composé de deux couches de dix neurones (n10n10) avec une fonction d'activation, $f_a(x)$, de type tangente hyperbolique, et la fonction de cutoff, $f_c(R)$, CT_TANH, avec $R_C = 6\text{Å}$.

Des descripteurs de type atome-centré (G) (voir partie I.1.3.2) sont utilisés. Pour l'oxygène et l'hydrogène, on les choisit identiques à ceux de l'étude [100], montrant de bons résultats. En ce qui concerne le calcium, on utilise une méthode développée par [41] et implémentée dans un code se trouvant à l'adresse https://github.com/flobuch/n2p2/tree/symfunc_paramgen/tools/python/symfunc_paramgen/examples. Cette méthode donne, pour les descripteurs à deux corps calcium-oxygène et calcium-hydrogène, un ensemble de neuf gaussiennes dont les centres sont uniformément répartis entre 0Å et 6Å . Pour les descripteurs à trois corps, on en obtient 12 pour chaque triplet d'atomes. L'ensemble des paramètres des descripteurs se trouvent en annexe C.2 tableaux 6, 7 et 8.

Les descripteurs, sont mis à l'échelle en utilisant l'opération ST_SCALECENTER.

Filtre de Kalman

Nous avons décidé d'effectuer les mêmes tests que les auteurs de [100]. Ils utilisent une valeur "typique" de $\tau_\eta = \tau_q = \ln(10)^{-1}$ et montrent qu'il y a peu d'influence sur le choix des paramètres q_0 et q_{min} que l'on prendra alors égale à 10^{-2} et 10^{-6} respectivement.

Les tests qui suivent ont été effectués pour chaque ensemble de paramètres (s , ratio énergie:forces, β , η_{max}), avec $\eta_0 = 10^{-2}$, tel que :

$$\begin{aligned} s &\in \{16, 32, 64\} \\ \text{ratio énergie:force} &\in \{1 : 10, 1 : 20, 1 : 50\} \\ \beta &\in \{2, 5, 10, 20\} \\ \eta_{max} &\in \{0.01, 0.03, 0.1, 0.3, 1.0\} \end{aligned}$$

Pour chaque ensemble de paramètres, on réalise trois optimisations de 30 époques avec des poids initiaux différents. On récupère ensuite parmi ces 3x30 époques les valeurs de RMSEs en énergie par atome et forces les plus basses.

Les résultats des RMSEs des énergies par atome sont présentés figure IV.12. Les RMSE en énergie par atome obtenues varient de 0.77 meV/at à 2.02meV/at. Un nombre de tâches s de 16 ou 32 donne des résultats similaires et meilleurs qu'avec 64 tâches. Il est préférable d'utiliser une faible valeur β ainsi que de grandes valeurs de η_{max} . Pour ces valeurs ($\beta = 2$, $\eta_{max} = 0.3 - 1.0$), la valeur du ratio énergie:forces ne semble pas avoir une influence significative.

Pour les forces, les RMSEs, figure IV.13, varient entre de 55.26 meV/Å et 167.41

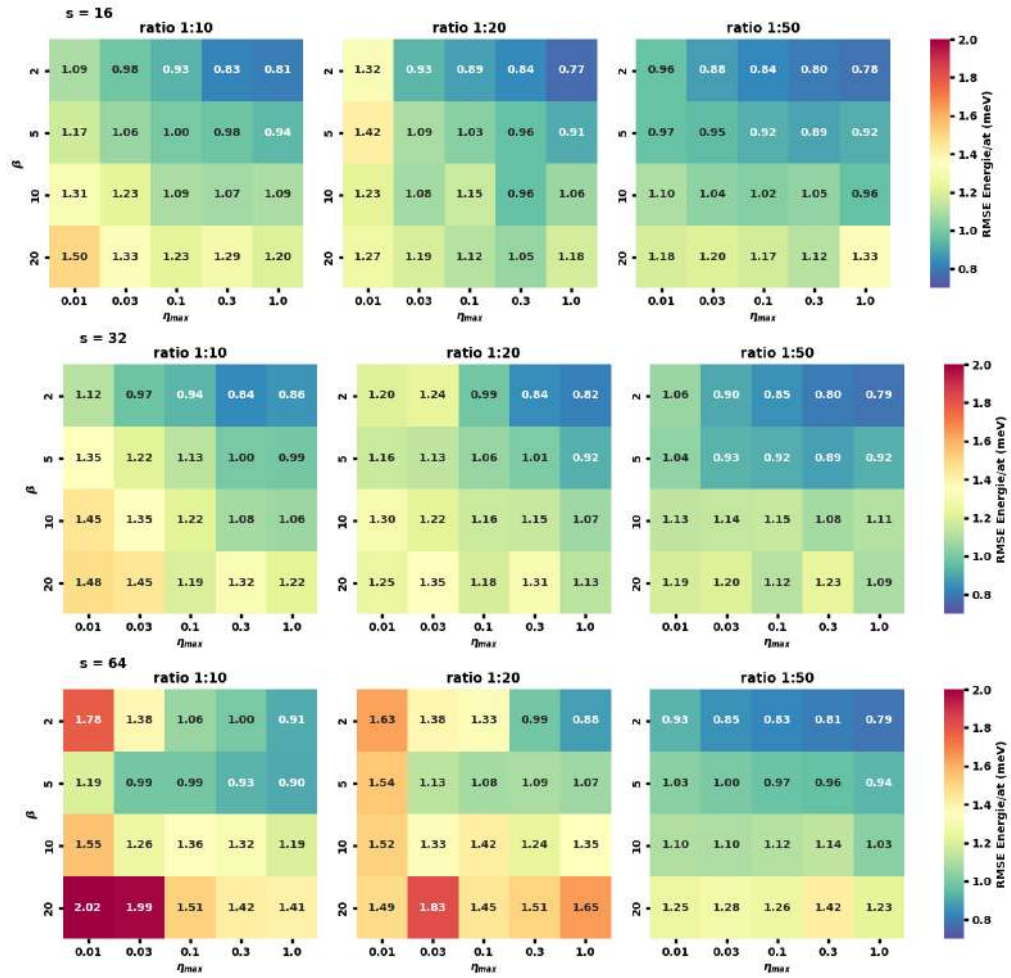


Figure IV.12: RMSEs des énergie par atomes les plus faibles obtenues pour différents ensemble de paramètres (s , ratio énergie:forces, β , η_{max})

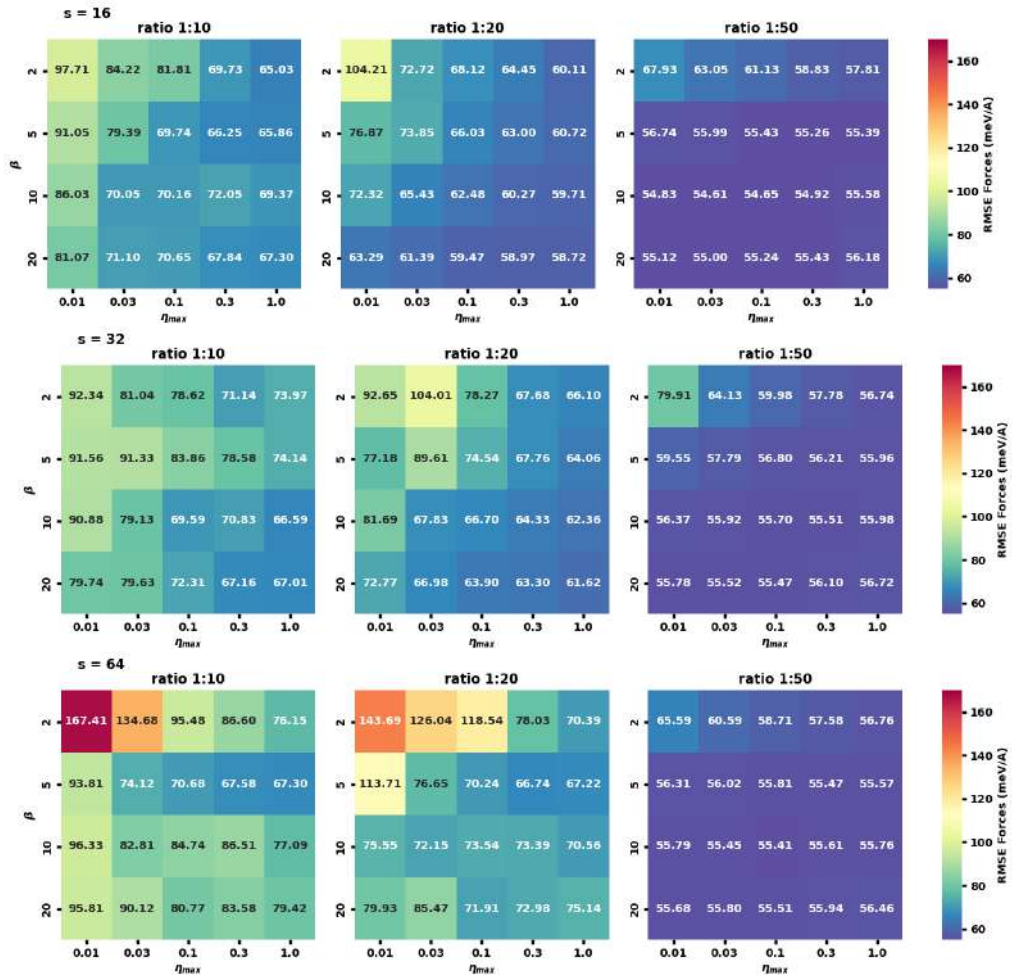


Figure IV.13: RMSEs des forces les plus faibles obtenues pour différents ensemble de paramètres (s , ratio énergie:forces, β , η_{max})

meV/Å et la conclusion est la même concernant le nombre de tâches. De même, des valeurs de η_{max} de 0.3 ou 1.0 donnent de meilleurs résultats. Contrairement à l'énergie, la valeur de ratio énergie:forces est importante avec de meilleurs résultats pour un ratio 1:50. Pour ce ratio, avec $s = 16 - 32$ et $\eta_{max} = 0.3 - 1.0$ la valeur de β ne modifie pas significativement les valeurs avec tout de même des résultats plus faibles pour $\beta = 5$.

Dans la suite, on choisira alors $s = 16$, $\beta = 5$, $\eta_{max} = 0.3$ et un ratio de 1:50.

Etant donné la grande amélioration des RMSEs des forces lorsque qu'un ratio énergie:forces faible est utilisé, on verra aussi dans la partie suivante que nous avons testé le cas d'un ratio 1:5049

Paramètres du réseau de neurones

Nous effectuons ici différents tests concernant l'architecture du réseau de neurones. Pour ces tests, on lance, sauf précisé, cinq entraînements, de 10 époques par paramètre, avec une initialisation des poids du réseau de neurones différente. On trace ensuite la moyenne des RMSE de l'énergie par atome et des forces de ces cinq entraînements

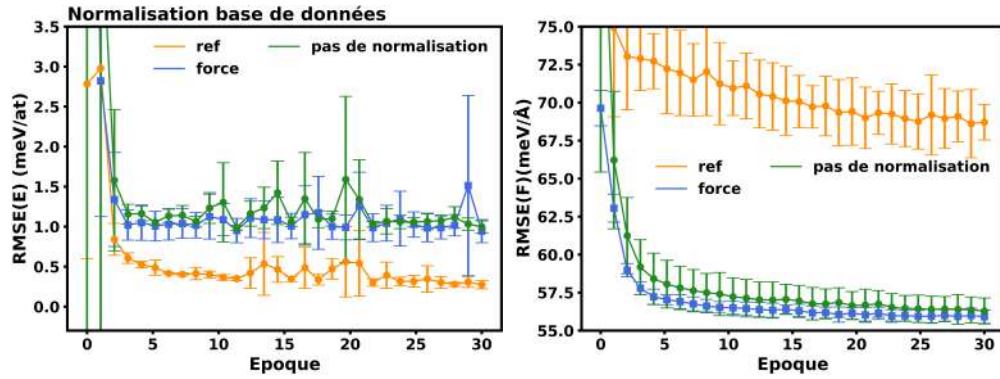


Figure IV.14: RMSE(E) et RMSE(F) pour le test des différents mode de normalisation de la base de donnée.

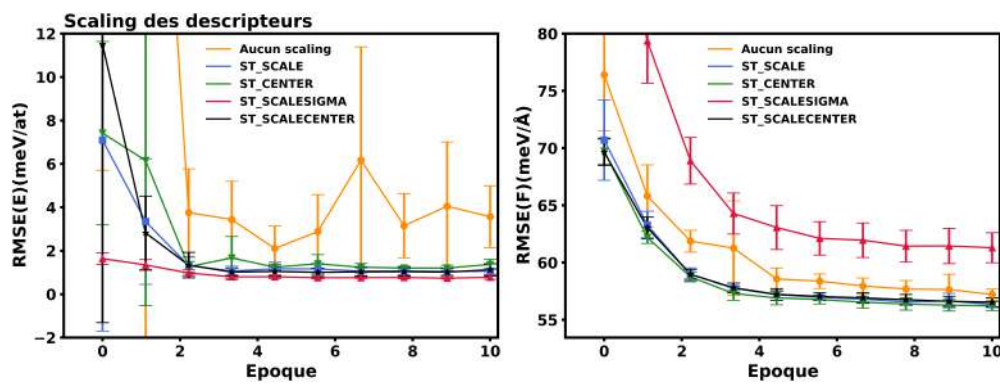


Figure IV.15: RMSE(E) et RMSE(F) des tests du scaling des descripteurs.

avec les barres d'erreur prisent égales à l'écart-type. On choisit ensuite le paramètre donnant des RMSEs les plus faibles.

Concernant la normalisation de la base de données, ou pour chaque entraînement 30 époques ont été réalisés, le mode "ref" donne de meilleurs RMSE(E) mais de mauvaises RMSE(F) (voir figue IV.14). Concernant le mode "force", on obtient des RMSE convergeant vers des valeurs semblables au cas où l'on effectue aucune normalisation mais tout de même une moyenne des RMSE(F) des différents entraînements plus faible. On choisira alors ce mode par la suite.

On regarde ensuite l'utilisation d'une opération de mise à l'échelle des descripteurs. Les résultats montrent que l'utilisation de ces opérations réduit considérablement les RMSEs (voir figure IV.15). L'utilisation de ST_SCALESIGMA donne de meilleurs valeurs de RMSE(E) mais de mauvaises valeurs de RMSE(F). Pour les trois autres opérations, la différence est très faible. On choisit par la suite de garder la mise à l'échelle ST_SCALECENTER.

Les fonctions d'activation testées, tangente hyperbolique, softplus, logistique et la fonction ReLU donnent des RMSE(E) (voir figure IV.16) qui convergent vers des valeurs similaires. En revanche, pour les forces la fonction ReLU donne des RMSE(F) presque deux fois supérieures aux autres fonctions, qui quant à elles donnent des valeurs très proches. Nous avons donc décidé de garder la fonction d'activation tangente hyperbolique.

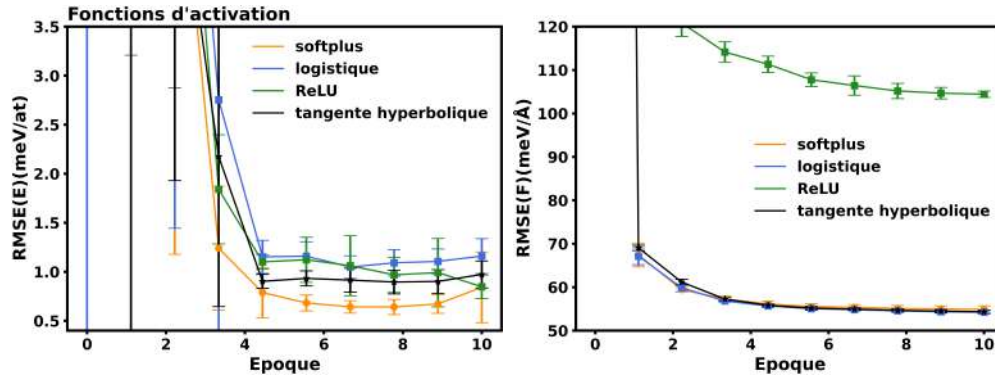


Figure IV.16: RMSE(E) et RMSE(F) des tests du choix de la fonction d'activation.

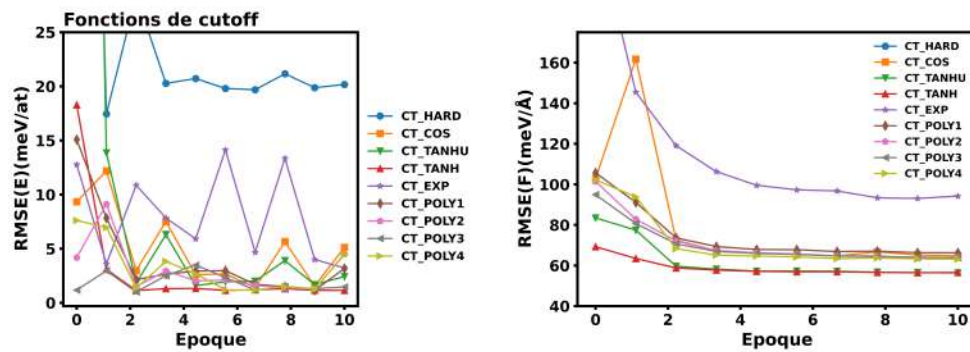


Figure IV.17: RMSE(E) et RMSE(F) des tests du choix de la fonction de cutoff.

Pour le choix de la fonction de cutoff, cette fois-ci, un seul entraînement pour les neuf fonctions de cutoff testées a été réalisé. Mise à part pour les fonctions de cutoff CT_HARD et CT_EXP, le choix de la fonction de cutoff ne semble pas influencer grandement les résultats (voir figure IV.17). Les fonctions de cutoff CT_TANHU et CT_TANH donnent des RMSE(F) convergeant vers des valeurs plus faibles et on gardera pour la suite la fonction CT_TANH.

La normalisation des valeurs d'entrées des neurones est généralement utilisée pour améliorer et accélérer l'entraînement, néanmoins, dans notre cas, on obtient de moins bonnes RMSE(E) et la RMSE(F) converge plus lentement (voir figure IV.18). Nous avons donc décidé de ne pas utiliser cette normalisation.

Jusqu'à maintenant, les tests ont été effectués pour un réseau de neurones de 2 couches composées de 10 neurones chacun (n10n10). On teste alors l'influence de l'augmentation du nombre de neurones par couche (n20n20 et n40n40) et de l'augmentation du nombre de couches (n10n10n10 et n20n20n20). Les résultats de ces tests se trouvent figure IV.19. Les résultats pour n10n10 et n10n10n10 donnent des résultats quasi-identiques. La même conclusion est observée pour n20n20 et n20n20n20. On en conclut qu'ajouter une couche à notre réseau de neurones n'améliore pas les résultats. En revanche, l'ajout de neurones dans une couche donne, notamment pour la RMSE(F), des valeurs convergées plus faibles. Entre n20n20 et n40n40, l'écart est tout de même petit, avec une amélioration de $1\text{meV}/\text{Å}$ au bout

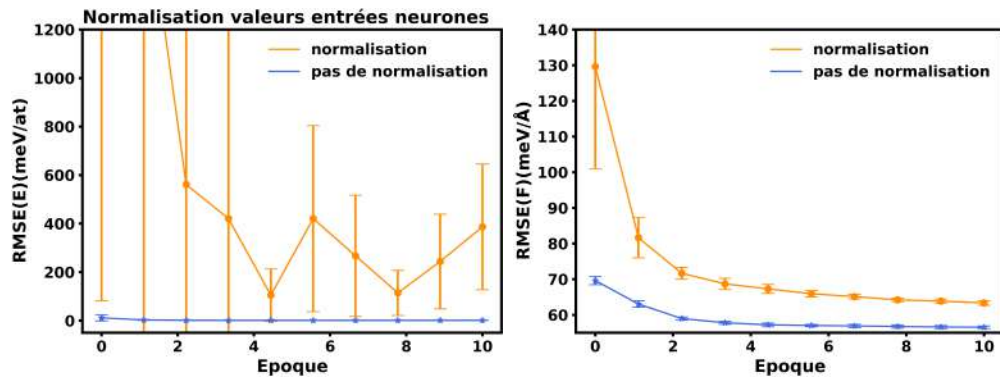


Figure IV.18: RMSE(E) et RMSE(F) de la normalisation des valeurs d'entrées des neurones.

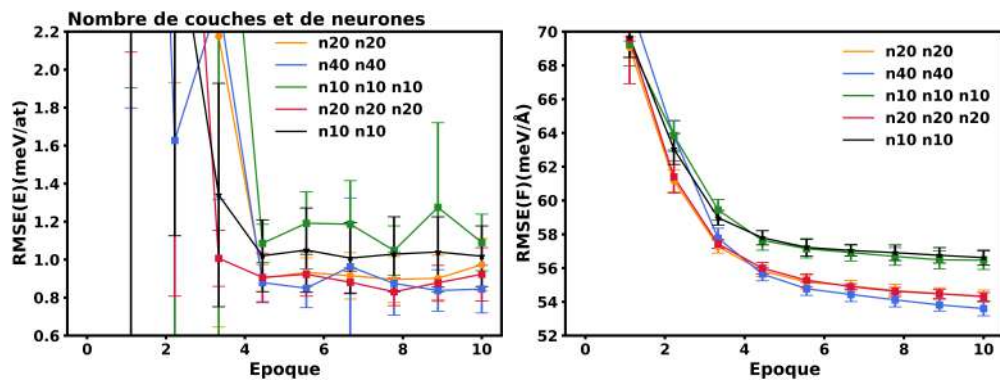


Figure IV.19: RMSE(E) et RMSE(F) des tests obtenues pour différents nombres de couches du réseau de neurones et pour différents nombres de neurones par couche.

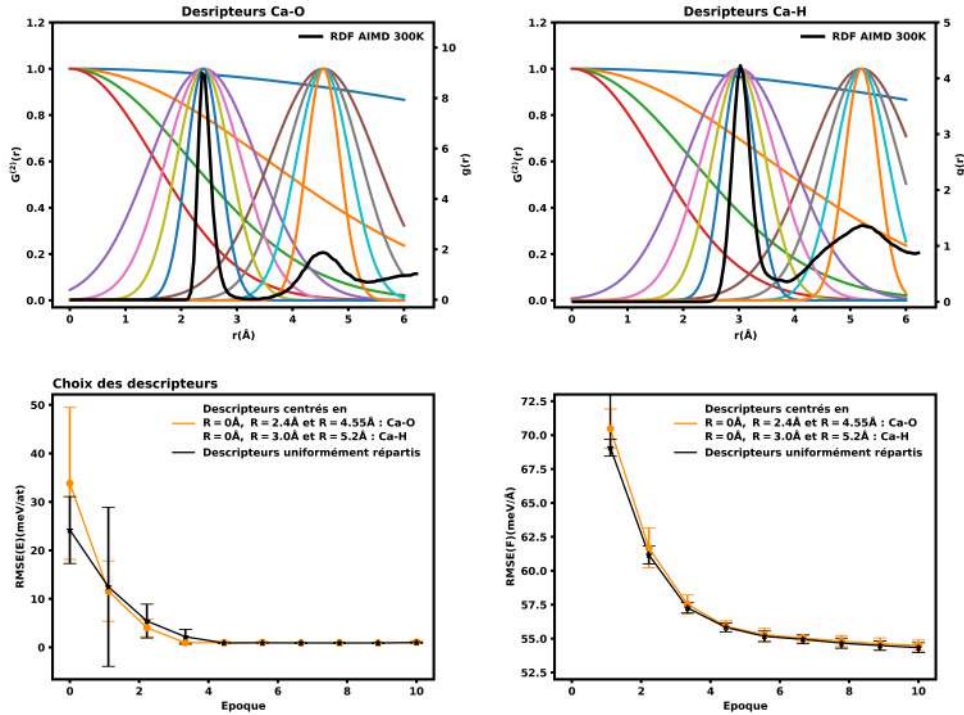


Figure IV.20: En haut, les descripteurs Ca-O et Ca-H centrés en $R = 0\text{\AA}$ et sur les positions des sphères de coordinence. En bas, les RMSE(E) et RMSE(F) obtenues et comparé au cas où les descripteurs sont uniformément répartis entre $R = 0\text{\AA}$ et $R = 6\text{\AA}$.

de 10 époques. Comme l'entraînement, et l'utilisation, du réseau de neurones n40n40 demandent un temps de calcul beaucoup plus long comparé au réseau de neurones n20n20, nous avons décidé de garder ce dernier pour la suite.

Il nous reste à tester si la modification des descripteurs permet l'obtention de meilleurs résultats. Concernant les descripteurs à deux corps Ca-O et Ca-H, au lieu de prendre des gaussiennes de même largeur uniformément réparties entre $R = 0\text{\AA}$ et $R = 6\text{\AA}$, nous avons testé l'utilisation de gaussiennes de différentes largeurs centrées en $R = 0\text{\AA}$ et sur les positions de la première et deuxième sphère de coordinence de Ca-O et Ca-H (voir figure IV.20). Les RMSEs obtenues se trouvent figure IV.20 et les résultats ne sont pas améliorés. Nous avons donc choisi de continuer avec les descripteurs à deux corps précédemment choisis. Concernant les descripteurs à trois corps, nous avons testé d'augmenter leur nombre en ajoutant, pour chaque valeur de η et λ des descripteurs avec $\xi = 12, 24$ et 48 . Là encore, les RMSEs (figure IV.21) ne sont pas améliorées. Ainsi, dans la suite, les descripteurs utilisés sont les mêmes que ceux définis au début de cette partie.

Finalement, avant de lancer des entraînements avec les paramètres choisis, nous effectuons une étude de la convergence des RMSEs en fonction de la taille de la base de données. On teste des bases de données composées de 600 (6x100), 1500 (6x250), 2500 (6x500), 4800 (6x800) et 6000 (6x1000) configurations. Dans chaque cas, les bases de données sont construites en récupérant, à intervalle régulier, un nombre

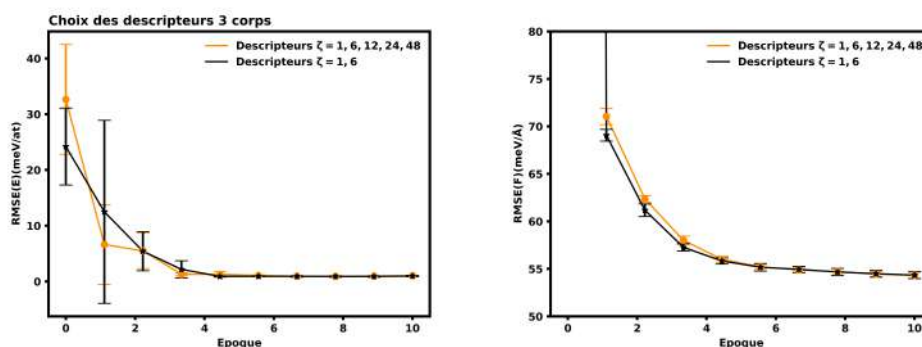


Figure IV.21: RMSE(E) et RMSE(F) obtenues avec différents descripteurs à 3 corps.

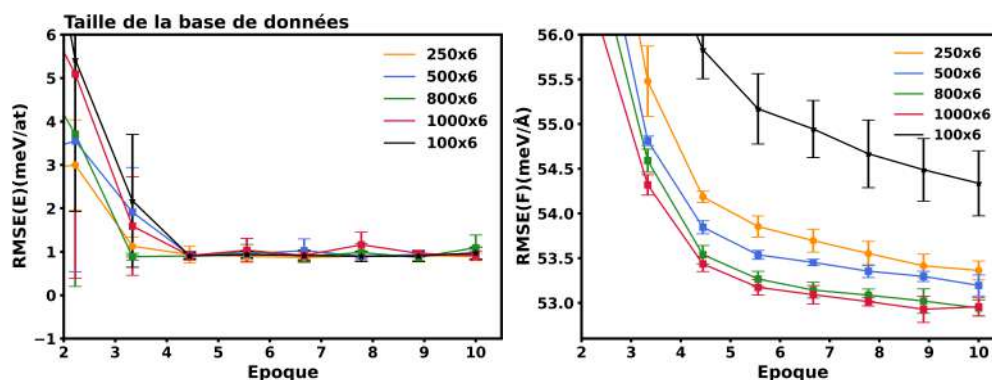


Figure IV.22: RMSE(E) et RMSE(F) obtenues pour différentes tailles de la base de données.

égal de configurations, provenant des AIMDs aux températures 300K, 350K, 400K, 450K, 500K, 550K et 600K.

Concernant la RMSE(E), figure IV.22, il est difficile de conclure quant à l'effet de la taille de la base de données. Pour les RMSE(F), toujours figure IV.22, plus la base de données est grande, plus les valeurs obtenues sont basses. Si on compare les résultats pour une base de données composée de 4800 configurations et 6000 configurations, les RMSE(F) convergent vers des valeurs similaires. On considère alors qu'il n'est pas nécessaire d'utiliser une base de données plus grande que 6000 configurations et c'est cette base de données qui sera utilisée par la suite.

Ces premiers tests nous ont permis de déterminer les paramètres permettant d'obtenir un apprentissage d'un potentiel machine learning rapide et efficace. Nous lançons alors dans la suite, plusieurs entraînements, à partir desquels nous effectuerons des premières dynamiques moléculaires que nous comparerons aux AIMD afin de déterminer le potentiel machine learning le plus approprié pour la suite de nos études.

Pour tous les autres paramètres, nous avons utilisé les valeurs de [100], dont les inputs se trouvent dans le dossier exemple du code N2P2.

	RMSE(E)(meV/at)		RMSE(F)(meV/Å)	
	train	test	train	test
entraînement1_1	1.55	1.67	52.26	52.55
entraînement1_2	0.90	0.93	52.24	52.74
entraînement1_3	0.98	1.00	52.12	52.98
entraînement1_4	0.83	0.80	52.25	52.72
entraînement1_5	0.90	0.86	52.34	52.88
entraînement2	1.39	1.37	55.40	55.29

Table IV.2: RMSEs des différentes entraînements pour l'époque où la $RMSE(F)_{train}$ est la plus faible.

IV.2.2 Entraînements et premiers résultats

Nous lançons cinq entraînements de potentiel machine learning, avec les paramètres déterminés dans la partie précédente, avec cinq poids et biais initiaux différents. On note ces entraînements "entraînement1_X" avec $X=1, \dots, 5$. Compte tenu de l'importance du choix du ratio énergie:forces que nous avons observé dans la partie précédente, nous lançons aussi un entraînement avec un ratio 1:549 que nous notons "entraînement2". A noter que pour cet entraînement, le scaling des descripteurs utilisé est ST_SCALESIGMA et que nous n'avons pas normalisé la base de données. Le nombre d'époques d'un entraînement est limité par le "wall time" du cluster de calcul utilisé. En effet, même si le code N2P2 permet de relancer un apprentissage à partir des poids et biais déterminés lors d'un précédent entraînement, il ne garde pas en mémoire les valeurs de $\vec{K}(e)$ utilisé équation IV.3. Pour les entraînements1_X, nous effectuons 100 époques. Pour l'entraînement2, seulement 9 époques, le nombre de forces utilisées pour l'optimisation des poids d'une époque étant très grand, les calculs sont très longs.

Pour chaque entraînement, 10% de la base de données, soit 600 configurations, est utilisé comme test.

A partir des RMSEs obtenues au cours de l'entraînement, on cherche l'époque pour laquelle on obtient la RMSE(F) la plus faible. En effet, on privilégie les forces devant l'énergie car elles sont utilisées pour le calcul du fractionnement isotopique. Les RMSE(E) et RMSE(F) de cette époque se trouvent tableau IV.2 pour chaque entraînement, notées par l'indice "train", avec les RMSEs obtenues sur les configurations réservées au test, notées avec l'indice "test".

Concernant les entraînements1_X, les $RMSE(F)_{train}$ et $RMSE(F)_{test}$ obtenues sont de l'ordre de 52-53meV/Å. Les RMSE(E) sont quant à elles de l'ordre de 0.80-1.00 meV/at, avec tout de même des valeurs de ~ 160 meV/at pour l'entraînement1_1. Pour l'entraînement2, les RMSEs sont légèrement plus élevées, mais le nombre d'époque plus faible, avec des valeurs de ~ 55 meV/Å pour les forces et ~ 1.4 meV/at pour les énergies par atome.

Il est difficile à ce stade de déterminer quel potentiel donnera des meilleurs résultats pour la suite. On effectue alors, pour chaque potentiel une dynamique

Dynamique moléculaire	$g_{OO}^{\max,1}$	$r_{OO}^{\max,1}(\text{Å})$	$g_{OO}^{\min,1}$	$r_{OO}^{\min,1}(\text{Å})$	$g_{OO}^{\max,2}$	$r_{OO}^{\max,2}(\text{Å})$
entraînement1_1	2.32	2.80	0.96	3.63	1.07	4.52
entraînement1_2	2.41	2.80	0.91	3.64	1.11	4.50
entraînement1_3	2.37	2.78	0.95	3.55	1.08	4.37
entraînement1_4	2.43	2.79	0.90	3.53	1.11	4.42
entraînement1_5	2.46	2.79	0.89	3.51	1.13	4.46
entraînement2	2.26	2.81	0.98	3.63	1.05	4.40
AIMD	2.24	2.82	0.98	3.68	1.05	4.46

Table IV.3: Positions et amplitudes du premier et deuxième maximum de la fonction de distribution radiale de l'oxygène, $(r_{OO}^{\max,1}; g_{OO}^{\max,1})$ et $(r_{OO}^{\max,2}; g_{OO}^{\max,2})$, ainsi que du premier minimum, $(r_{OO}^{\min,1}; g_{OO}^{\min,1})$ obtenues pour les différents MLMD à 400K avec les potentiels ML provenant des différents entraînements ainsi que les valeurs des données ab-initio.

moléculaire de 500ps à 400K avec les mêmes paramètres que dans la partie IV.1.2.1. Comme on souhaite être en mesure de reproduire correctement l'environnement du calcium, on trace figure IV.23 les $g(r)$ Ca-O et Ca-H et les compare à celles obtenues en AIMD, ainsi que les $g(r)$ de l'eau. L'observation des $g_{CaO}(r)$ et $g_{CaH}(r)$ nous permet tout de suite de voir que les potentiels entraînement1_3 et entraînement1_4 sont incapables de reproduire l'environnement du calcium. Pour ces $g(r)$, les autres potentiels donnent des résultats quasi-identiques et proches de l'AIMD avec tout de même une amplitude du premier maxima légèrement supérieure donnant une coordinence un peu plus faible que celle obtenue en AIMD.

Concernant les $g(r)$ de l'eau, les potentiels des entraînement1_X, donnent des amplitudes des extrema supérieures à l'AIMD, donnant une eau plus structurée. En revanche, le potentiel de l'entraînement2 reproduit extrêmement bien les valeurs de l'AIMD. En effet, si on regarde en particulier la $g_{OO}(r)$, dont les valeurs des positions des extrema ainsi que leurs amplitudes se trouvent tableau IV.3, on remarque que, bien que les positions des extrema soient bien reproduites pour tous les entraînements, l'entraînement2 donne des amplitudes identiques à l'AIMD pour le premier minima et le deuxième maxima et très proche pour le premier maxima avec une valeur de 2.26 comparée à 2.24 pour l'AIMD. Pour les entraînement1_X, toutes les amplitudes sont supérieures à celles de l'entraînement2.

Nous avons montré que les potentiels ainsi générés pouvaient être en mesure de reproduire les données de l'AIMD. Il se trouve que c'est le potentiel provenant de l'entraînement2 qui donne de meilleurs résultats, malgré le fait que l'on obtienne des RMSEs supérieures aux autres entraînements. Dans la suite, c'est ce potentiel que nous choisirons pour nos études et que l'on appellera "le" potentiel machine learning.

A l'aide de ce potentiel ML, nous sommes maintenant en mesure d'effectuer de longues trajectoires MLMD.

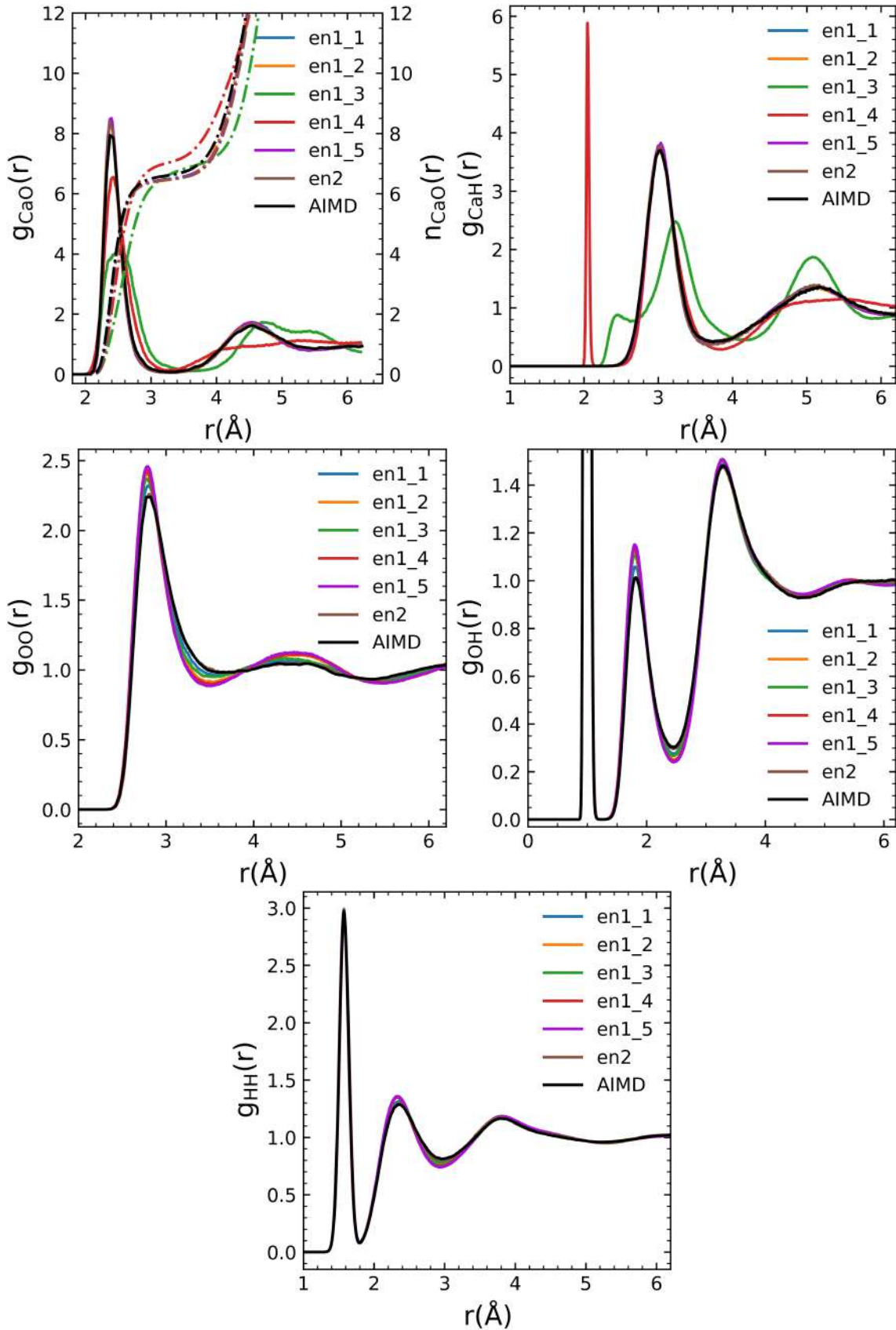


Figure IV.23: Fonctions de distribution radiale des configurations obtenues à partir des différentes MLMD à T=400K avec les potentiels des entraînement1_X (en1_X), X=1,...,5, et entraînement2 (en2) comparées à celles obtenues en AIMD.

IV.3 Coordinnence du calcium

L'objectif de cette partie est double. Nous souhaitons tout d'abord continuer à tester le potentiel ML, notamment à différentes températures. Cela nous permettra en plus de poursuivre l'étude de la coordinnence du calcium du chapitre III.

IV.3.1 Protocole

Pour tester le potentiel ML à différentes températures et étudier la convergence de la coordinnence du calcium, nous avons effectué des MLMDs en effectuant la procédure suivante:

1. A partir d'une même configuration initiale, on lance trois MLMDs de 500ps à 600K avec des vitesses initiales différentes.
2. On récupère les dernières configurations et on lance trois MLMDs de 2ns à 600K.
3. On récupère les dernières configurations et on lance trois MLMDs de 500ps à une température de départ de 600K et une température de fin de 550K.
4. On récupère les dernières configurations et on lance trois MLMDs de 2ns à 550K.

On répète ensuite les étapes 3 et 4 jusqu'à obtenir trois MLMDs de 2ns pour des températures de 600K, 550K, 500K, 450K, 400K, 350K et 300K. Toutes les MLMDs sont effectuées dans l'ensemble NVT avec des thermostats de type Nose-Hoover de longueur 5. Le pas de temps de simulation pour les étapes 1 et 3 est de $\Delta t = 0.1\text{fs}$ et pour les étapes 2 et 4 de $\Delta t = 0.06047\text{fs}$, le même pas de temps que pour les AIMD.

IV.3.2 Résultats

Dans cette partie nous gardons les mêmes notations que dans la partie III.

Pour plus de lisibilité, l'ensemble des $g(r)$, pour les trois MLMDs à chaque température, se trouve en fin de chapitre figures IV.32, IV.33, IV.34, IV.35, IV.36, IV.37 et IV.38. On remarque que le potentiel ML reproduit extrêmement bien la structure de l'eau obtenue en AIMD. Concernant la diffusion de l'eau, les coefficients de diffusion de l'oxygène en fonction du temps pour une MLMD aux différentes températures se trouvent figure IV.24 et les MSDs à partir desquelles ces coefficients ont été calculés sont en annexe C.3 figure 19.

Les coefficients de diffusion convergent rapidement, avec tout de même des variations observées, de l'ordre de $\sim 0.5\%$ de la valeur moyenne, en fin de trajectoire à 550K et 600K. A partir de ces courbes, on détermine le coefficient de diffusion en moyennant les valeurs prises entre 500ps et 1500ps. Sur cette interval de temps, les valeurs sont convergées et on exclut les 500 dernières picosecondes où on peut observer des variations de la valeur du coefficient de diffusion possiblement dû à la méthode utilisée pour calculer les MSDs. Les résultats se trouvent tableau IV.4 avec ceux de l'AIMD. Les coefficients de diffusion obtenus en MLMD sont systématiquement supérieurs à ceux de l'AIMD. A 300K et 350K, la différence est faible, de 0.013 et 0.071 respectivement. Aux autres températures, on obtient des coefficients de

T(K)	Dynamique moléculaire	D($\text{\AA}^2 \cdot \text{ps}^{-1}$)
300	MLMD n°0	0.057
	AIMD	0.041
350	MLMD n°0	0.211
	AIMD	0.140
400	MLMD n°0	0.543
	AIMD	0.280
450	MLMD n°0	0.825
	AIMD	0.430
500	MLMD n°0	1.140
	AIMD	0.570
550	MLMD n°0	1.393
	AIMD	0.700
600	MLMD n°0	1.738
	AIMD	0.740

Table IV.4: Coefficient de diffusion des oxygènes obtenues à partir de nos simulations MLMD et AIMD.

diffusion d'une valeur deux fois supérieure. La différence avec les valeurs de l'AIMD n'a pas été investiguée en détail. Elle pourrait venir des paramètres de simulation et du thermostat. Pour les MLMDs, la longueur des chaînes de Nose-Hoover est de 5, alors qu'en AIMD elle est de 4.

A partir de ces valeurs, on trace l'évolution du coefficient de diffusion en fonction de l'inverse de la température figure IV.25. Comme pour l'AIMD, on observe que les valeurs entre 450K et 600K et entre 300K et 400K suivent une loi d'Arrhenius. Entre 450K et 600K, on obtient une valeur d'énergie d'activation de $E_a = 10.88 kJ.mol^{-1}$ et entre 300K et 400K, une énergie d'activation de $E_a = 22.05 kJ.mol^{-1}$. Les valeurs des énergies d'activations sont du même ordre de grandeur que celles obtenues en AIMD mais supérieures de $\sim 3 kJ.mol^{-1}$ et $\sim 5 kJ.mol^{-1}$ respectivement.

La différence entre les $g(r)$ des MLMDs et de l'AIMD sont plus visibles sur les g_{CaO} et g_{CaH} . En effet, même si les deux sphères de coordinence sont bien reproduites, sauf à 300K où elles sont légèrement plus structurées, on peut remarquer des valeurs de l'intégrale de $g_{CaO}(r)$ inférieure à celles de l'AIMD entre les deux sphères de coordinence. Les valeurs de la coordinence, $CN^{(1)}$, extraites à partir de ces intégrales se trouvent tableau IV.6 avec celles de l'AIMD. Les différences observées sont de l'ordre de ~ 0.4 , ~ 0.1 , ~ 0.1 , ~ 0.1 , ~ 0.05 , ~ 0.05 , ~ 0.1 à 300K, 350K, 400K, 450K, 500K, 550K et 600K respectivement. Il est possible que le potentiel ML donne des valeurs de la coordinence plus faible qu'avec la fonctionnelle BLYP-D2. Néanmoins, les différences plus importantes observées aux plus faibles températures peuvent aussi s'expliquer par les problèmes d'échantillonnage des AIMD discutés aux chapitre III.

Pour étudier la convergence de la coordinence et l'efficacité de l'échantillonnage, nous avons calculé, pour toutes les MLMDs, les évolutions de la coordinence en fonction du temps $CN(t)$ et $CN_a(t)$. Les figures se trouvent en annexe C.3, figures 20, 21, 22,

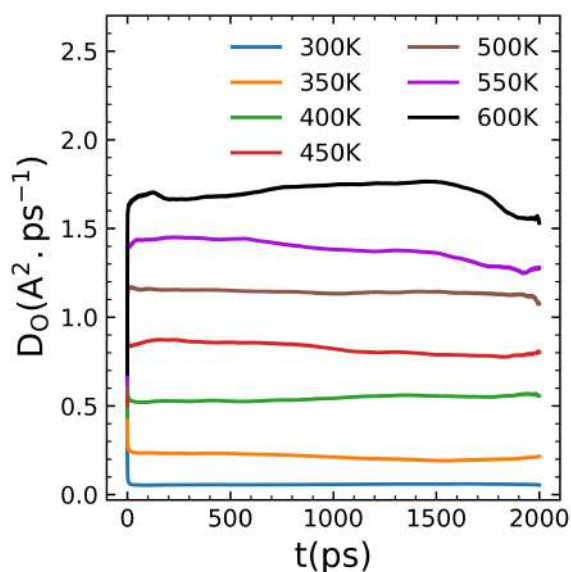


Figure IV.24: Coefficient de diffusion des atomes d'oxygène en fonction du temps des MLMDs à différentes températures.

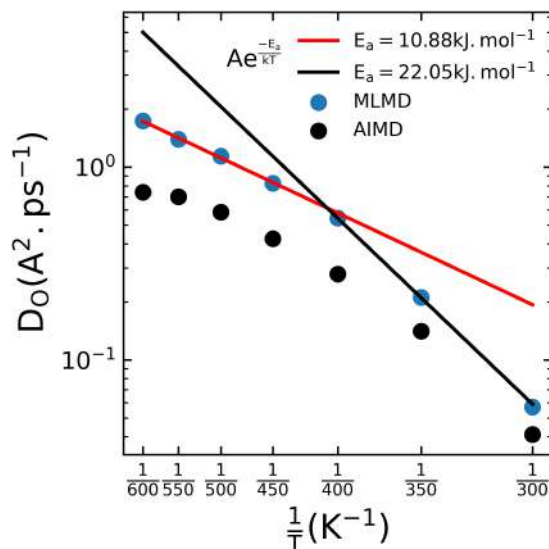


Figure IV.25: Coefficients de diffusion des atomes d'oxygène des MLMDs en fonction de l'inverse de la température. Les courbes en rouge et noir correspondent à un ajustement par une fonction exponentielle.

23, 24,25 et 26. A partir de ces courbes, nous avons déterminé le nombre d'échanges N_{ex} et le nombre d'échanges moyen sur un intervalle de 100ps N_{ex}^{100ps} (voir tableau IV.6). Les N_{ex}^{100ps} obtenues à partir des MLMDs sont plus faibles qu'en AIMD. En revanche, même à 300K, on observe au minimum 250 échanges. Nous avons montré au chapitre III qu'il était nécessaire d'observer au moins 200 échanges pour avoir une valeur de la coordinence convergée, indiquant ainsi, a priori, que nos simulations MLMDs ont un temps de simulations assez long pour avoir un bon échantillonnage. On peut l'observer en calculant, encore une fois en considérant que les $CN_a(t)$ forment une chaîne de Markov, les écart-types, σ , des valeurs de la coordinence à partir de la relation III.9. Les figures IV.26 et IV.27 montrent les σ en fonction du temps et en fonction du nombre d'échanges. Toutes les courbes montrent un comportement asymptotique et, sauf à 300K, les écart-types convergent. A 300K, les écart-types continuent à diminuer. On peut alors supposer qu'à cette température on obtient une valeur de σ qui surestime la vraie valeur. Dans tous les cas, les courbes commencent leur convergence/décroissance au bout de quelques centaines d'échanges.

De plus, si les MLMDs ont permis un bon échantillonnage des configurations, alors les distributions des valeurs de la coordinence obtenues devraient nous donner, pour une température donnée, la même valeur de la coordinence. On trace alors, pour les trois MLMD de même température, l'évolution de la coordinence moyenne en fonction du temps $\langle CN \rangle (t)$ avec la valeur des incertitudes, calculées à partir de σ ($2\sigma/\sqrt{N_{steps}}$), figures IV.28 et IV.29. A 400K, 500K, 550K, 600K, toutes les MLMDs donnent des valeurs de la coordinence en accord avec leurs incertitudes. Pour les MLMDs à 300K, 350K, 450K, on observe qu'une MLMD a une valeur de

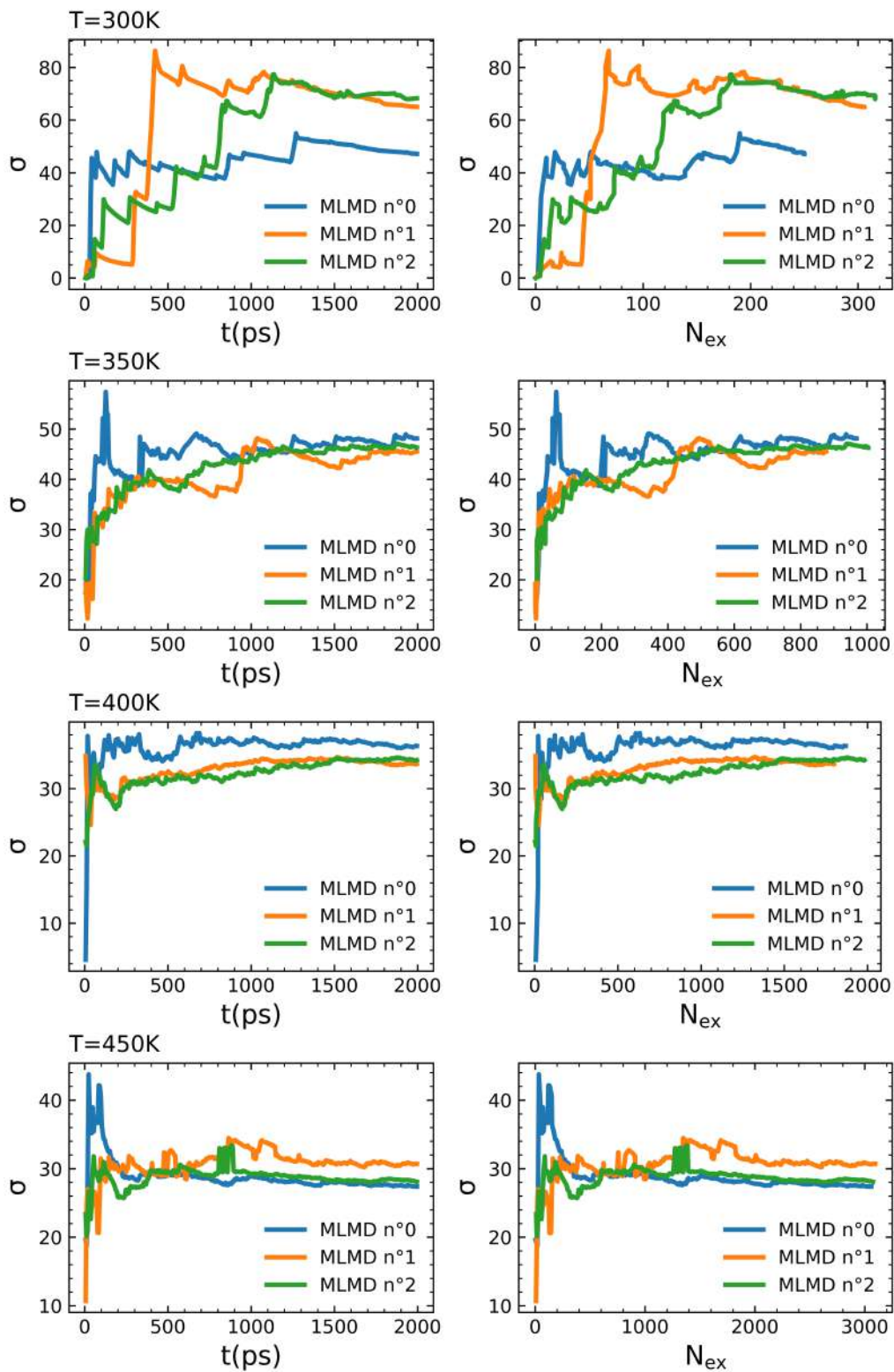


Figure IV.26: Evolution de l'écart-type en fonction du temps et du nombre d'échanges N_{ex} pour les MLMDs à 300K, 350K, 400K et 450K.

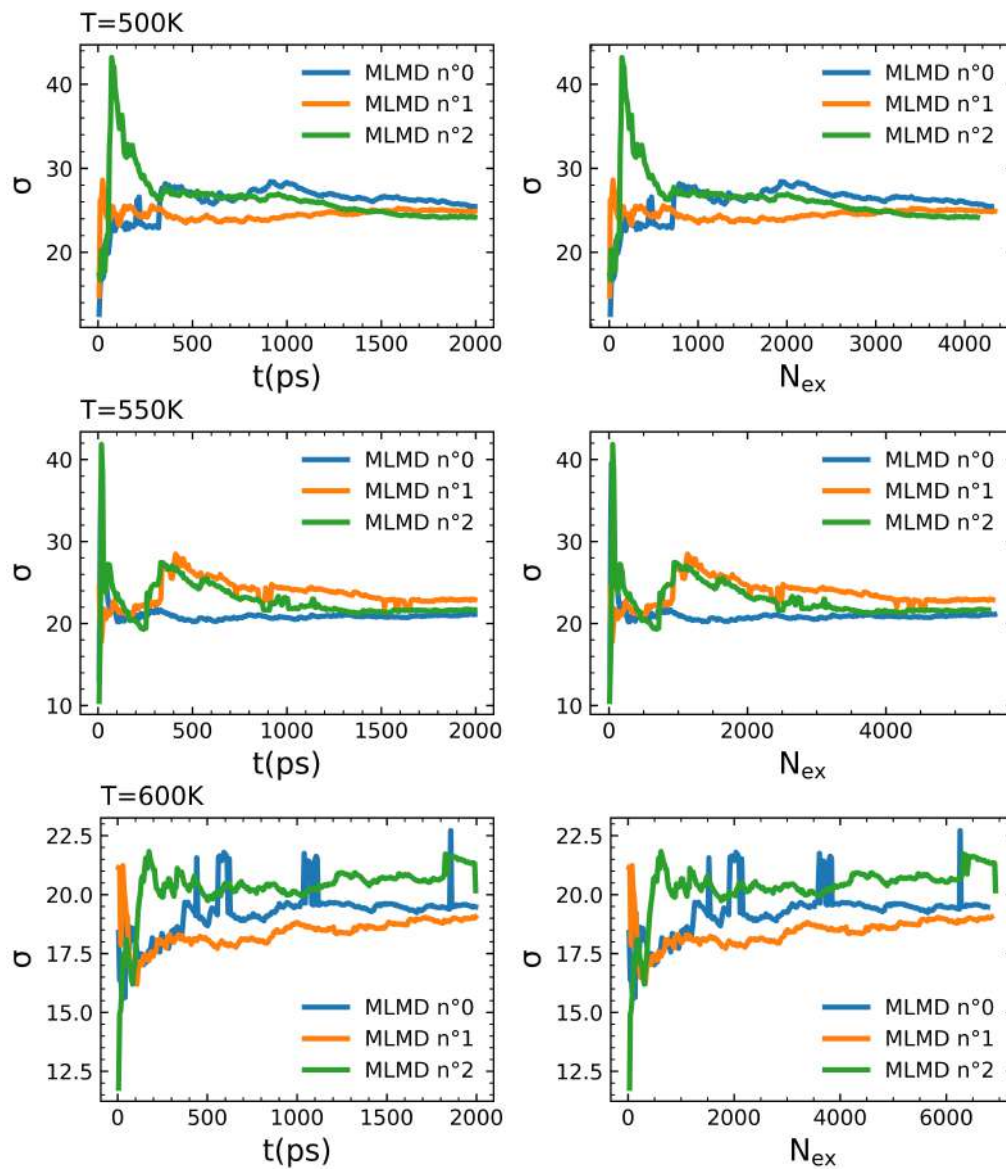


Figure IV.27: Evolution de l'écart-type en fonction du temps et du nombre d'échanges N_{ex} pour les MLMDs à 500K, 550K et 600K.

la coordinence très légèrement en désaccord avec les autres (< 0.01). Les faibles écarts des valeurs obtenues nous permettent néanmoins d'être confiant quant aux coordinences finales obtenues.

D'après notre approche basée sur les chaînes de Markov, nous avons, à chaque température, trois chaînes. Il existe un paramètre, \hat{R} , qui permet de diagnostiquer la convergence d'une grandeur à partir de chaînes de Markov indépendantes par l'intermédiaire d'écart-types entre les différentes chaînes [121]:

$$\hat{R} = \sqrt{\frac{A}{W}} \quad (\text{IV.7})$$

avec :

$$A = \frac{N_{steps} - 1}{N_{steps}} W + \frac{1}{N_{steps}} B \quad (\text{IV.8})$$

$$B = \frac{N_{steps}}{M - 1} \sum_{m=1}^M \left(\langle CN_a^m(t) \rangle - \langle CN_a^{all}(t) \rangle \right)^2 \quad (\text{IV.9})$$

$$\text{avec: } \langle CN_a^m(t) \rangle = \frac{1}{N_{steps}} \sum_{i=0}^{N_{steps}-1} CN_a^m(i\Delta t), \quad \langle CN_a^{all}(t) \rangle = \frac{1}{M} \sum_{i=1}^M \langle CN_a^m(t) \rangle \quad (\text{IV.10})$$

$$W = \frac{1}{M} \sum_{m=1}^M s_m^2, \quad \text{avec: } s_m^2 = \frac{1}{N_{steps} - 1} \sum_{i=0}^{N_{steps}-1} (CN_a^m(i\Delta t) - \langle CN_a^m(t) \rangle)^2 \quad (\text{IV.11})$$

$$(\text{IV.12})$$

où M est le nombre de chaînes de Markov. Pour un processus ergodique, \hat{R} tend vers 1 lorsque N_{steps} tend vers l'infini. Les auteurs de [42] proposent de considérer que la convergence est atteinte lorsque $\hat{R} < 1.1$, alors que les auteurs de [121] proposent de prendre comme limite $\hat{R} < 1.01$. Il est aussi possible de couper les chaînes de Markov afin de diagnostiquer des problèmes de non-stationnarité. On applique la relation aux trois chaînes de Markov et on obtient les valeurs tableau IV.5. Pour les MLMDs à 300K et 350K, les valeurs sont inférieures à 1.1. Aux autres températures, elles sont inférieures à 1.01. Ces valeurs sont un autre indicateur que pour les MLMDs à des température supérieures à 400K, la convergence est atteinte. A 300K et 350K, selon le critère de [121], il sera nécessaire de poursuivre les trajectoires mais, la valeur des écart-types étant surestimée, nous avons décidé de garder ces valeurs.

Finalement, un point important du chapitre III était la question de la dépendance de la coordinence avec la température. On trace alors, figure IV.30, les coordinences finales, $CN^{(2)}$, des MLMDs en fonction de la température. Dans le cas des MLMDs, on observe plus clairement, surtout aux basses températures, que la coordinence croît avec la température. Les valeurs des coordinences des MLMDs sont en accord avec celles de l'AIMD sauf à 600K où elles sont légèrement inférieures. C'est aussi à cette température que les configurations des MLMDs sont les plus susceptibles d'avoir des environnements atomiques éloignés de ceux de la base de données qui

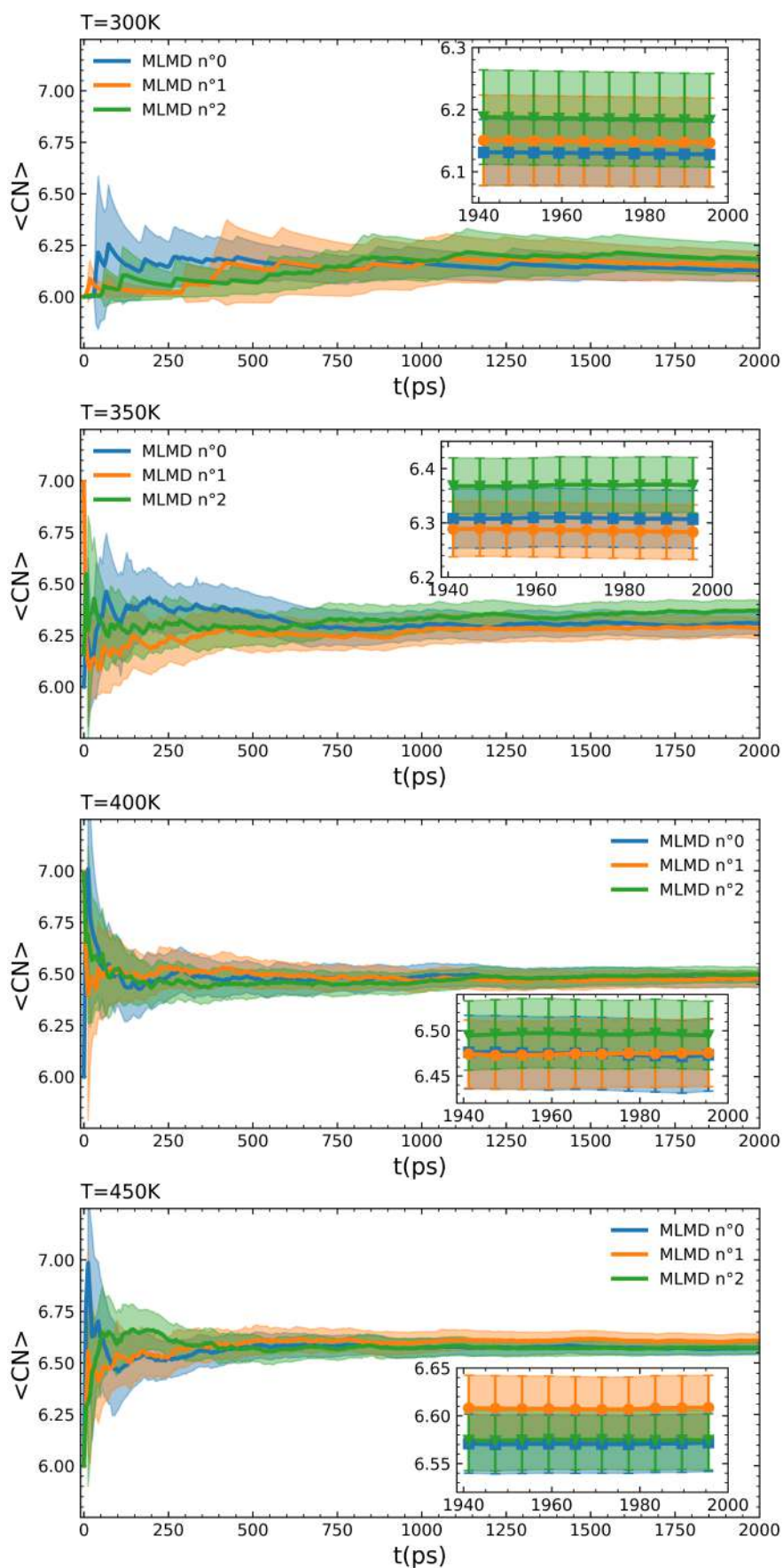


Figure IV.28: Evolution de la coordinence moyenne en fonction du temps des MLMD à 300K, 350K, 400K et 450K. La zone de couleur correspondent aux incertitudes calculées comme étant égales à $2\sigma/\sqrt{N_{steps}}$.

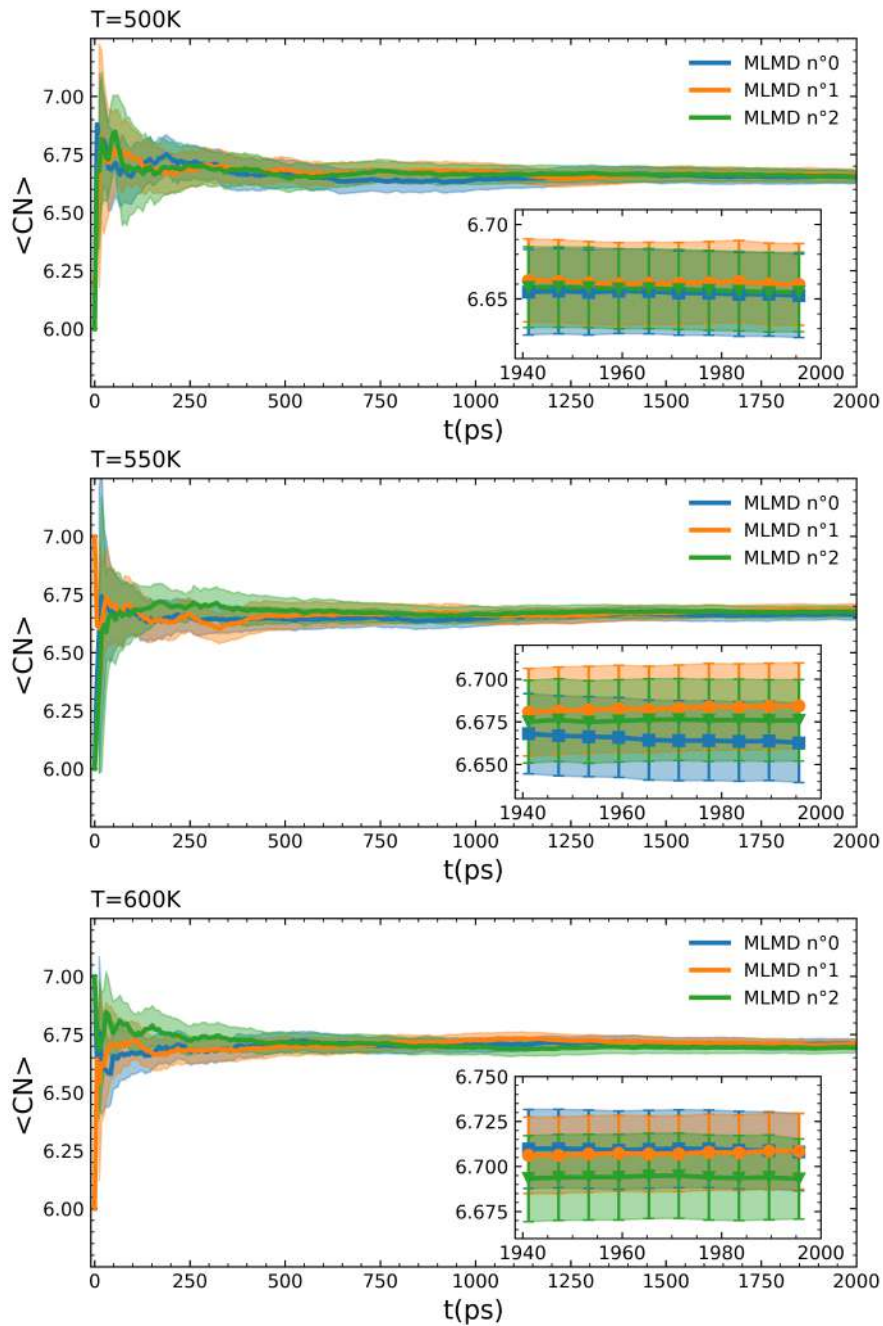
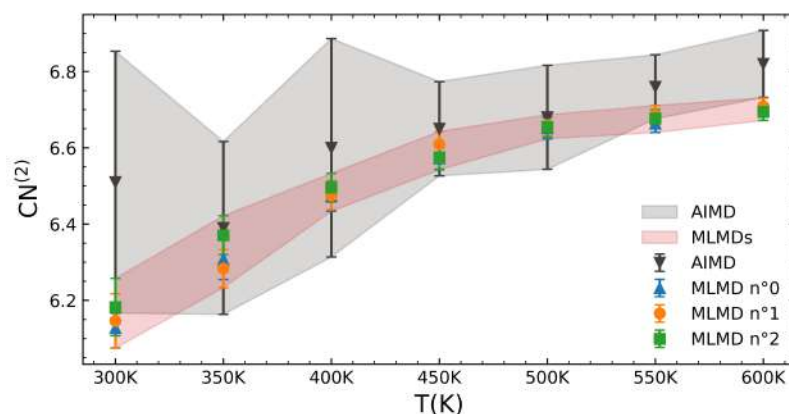


Figure IV.29: Evolution de la coordinence moyenne en fonction du temps des MLMD à 500K, 550K et 600K. La zone de couleur correspondent aux incertitudes calculées comme étant égales à $2\sigma/\sqrt{N_{steps}}$.

T(K)	\hat{R}
300	1.0505
350	1.0442
400	1.0026
450	1.0046
500	1.0002
550	1.0010
600	1.0006

Table IV.5: Valeur de \hat{R} en pour différentes températures.Figure IV.30: Variation de la coordinence des MLMD et de l'AIMD en fonction de la température. Les zones de couleur correspondent à la valeur maximale et minimale en considérant l'erreur sur la valeur de la coordinence égale à $2\sigma/\sqrt{N_{steps}}$.

a servi à l'entraînement du potentiel ML. Il se peut alors que les interpolations du potentiel ML sur ces configurations soient plus mauvaises qu'aux autres températures. Néanmoins, au vu des résultats aux autres températures, on peut confirmer que la coordinence dépend effectivement de la température et ainsi conforter les observations et conclusions du chapitre III.

L'étape suivante est alors d'utiliser le potentiel ML pour l'étude du fractionnement isotopique. Comme nous l'avons vu au chapitre II, le fractionnement isotopique dépend des modes de vibration du système. On peut alors, calculer les fonctions d'auto-corrélation des vitesses, VAF, et surtout la densité d'état des phonons, VDOS, obtenue par transformée de Fourier des VAF. Les résultats obtenus pour les MLMD à 300K se trouvent figure IV.31, avec ceux de l'AIMD. Les VDOS de l'hydrogène montrent quatre bandes. Une à $\sim 100cm^{-1}$ correspondant aux translations, une à $\sim 600cm^{-1}$ due aux rotations, une à $\sim 1600cm^{-1}$ provenant des déformations et une dernière à $\sim 3300cm^{-1}$ venant des modes de vibrations des élongations. Les deux premières bandes obtenues à partir des trajectoires AIMD et MLMD ont des positions et largeurs proches. Les trajectoires MLMD donnent une bande de déformation moins large qu'en AIMD alors que pour la bande d'élongation, sa position est inférieure

T(K)	Simulation	$t_{\text{simu}}(\text{ps})$	$CN^{(1)}$	$CN^{(2)}$	$u(CN^{(2)})$	N_{ex}	$N_{\text{ex}}^{100\text{ps}}$
300	AIMD	100	6.50	6.51	0.38	29	29
	MLMD n°0	2000	6.12	6.12	0.05	250	12.5
	MLMD n°1	2000	6.14	6.15	0.07	306	15.3
	MLMD n°2	2000	6.13	6.15	0.08	318	15.9
350	AIMD	100	6.40	6.39	0.25	65	65
	MLMD n°0	2000	6.31	6.30	0.05	969	48.4
	MLMD n°1	2000	6.29	6.28	0.05	877	43.8
	MLMD n°2	2000	6.38	6.37	0.05	1004	50.2
400	AIMD	100	6.60	6.60	0.31	126	126
	MLMD n°0	2000	6.46	6.47	0.04	1874	93.7
	MLMD n°1	2000	6.47	6.48	0.04	1808	90.4
	MLMD n°2	2000	6.49	6.50	0.04	1985	99.2
450	AIMD	100	6.66	6.65	0.14	171	171
	MLMD n°0	2000	6.56	6.57	0.03	3066	153.3
	MLMD n°1	2000	6.59	6.61	0.03	3102	155.1
	MLMD n°2	2000	6.58	6.57	0.08	3076	153.8
500	AIMD	100	6.69	6.68	0.15	214	214
	MLMD n°0	2000	6.64	6.65	0.03	4318	215.9
	MLMD n°1	2000	6.62	6.66	0.03	4344	217.2
	MLMD n°2	2000	6.63	6.66	0.03	4153	207.6
550	AIMD	100	6.68	6.5	0.09	308	308
	MLMD n°0	2000	6.63	6.66	0.02	5592	279.6
	MLMD n°1	2000	6.62	6.68	0.03	5558	277.9
	MLMD n°2	2000	6.63	6.68	0.02	5500	275.0
600	AIMD	100	6.83	6.82	0.10	388	388
	MLMD n°0	2000	6.67	6.70	0.02	6802	340.1
	MLMD n°1	2000	6.70	6.71	0.02	6870	343.5
	MLMD n°2	2000	6.65	6.69	0.02	6944	347.2

Table IV.6: Coordinences $CN^{(1)}$, obtenues à partir de l'intégrale de la $g_{CaO}(r)$, et $CN^{(2)}$, obtenues en prenant la moyenne des valeurs de $CN(t)$, ainsi que le nombre d'échanges N_{ex} , le nombre d'échanges durant 100ps $N_{\text{ex}}^{100\text{ps}}$ et l'erreur sur la valeur de la coordinnence, $\text{mathrmu}(CN^{(2)})$, égale à $2\sigma/\sqrt{N_{\text{steps}}}$ pour des MLMDs et de l'AIMD à différentes températures.

($\sim 100\text{cm}^{-1}$) à celle de l'AIMD. Concernant les VDOS de l'oxygène, nous obtenons un bon accord entre l'AIMD et les MLMD. Finalement, les modes de vibrations du calcium en AIMD correspondent à des nombres d'onde compris entre 0 et $\sim 500\text{cm}^{-1}$ avec un maximum à $\sim 100\text{cm}^{-1}$ bien reproduit par le potentiel ML.

Nous avons, dans cette partie, utilisé deux méthodes permettant de construire un potentiel ML. La première, utilisant le code MLACS, nous a permis de montrer les limitations de cette méthode dans le cas d'un système composé de plusieurs éléments chimiques. Néanmoins, nous pensons que cela est dû à l'impossibilité, pour le moment, de définir différents types de descripteurs pour chaque élément. De plus, les résultats ont montré qu'il était possible d'obtenir un potentiel ML donnant une structure de l'eau et de l'environnement du calcium meilleures que celles obtenues avec les potentiels empiriques et avec un temps de calcul faible. La précédente étude, du chapitre III, nous a permis de construire une base de données, avec des configurations provenant d'AIMD à différentes températures, sans calculs supplémentaires. Cette base de données a été utilisée pour générer un potentiel ML se basant sur les réseaux de neurones à partir du code N2P2. Les résultats obtenus montrent un bon accord avec l'AIMD, notamment en ce qui concerne la structure de l'eau et la coordinence du calcium. Ce potentiel nous a aussi permis de poursuivre l'étude du chapitre III et d'étudier l'influence de la température sur la coordinence du calcium. Les valeurs finales de la coordinence montrent une dépendance croissante avec la température. Ces résultats se basent sur des trajectoires de 2ns, plus longues que les trajectoires AIMD de notre étude et de la littérature, permettant d'obtenir un meilleur échantillonnage des configurations.

De plus, le potentiel ML reproduit les modes de vibrations du calcium, permettant de l'utiliser pour l'étude du fractionnement isotopique.

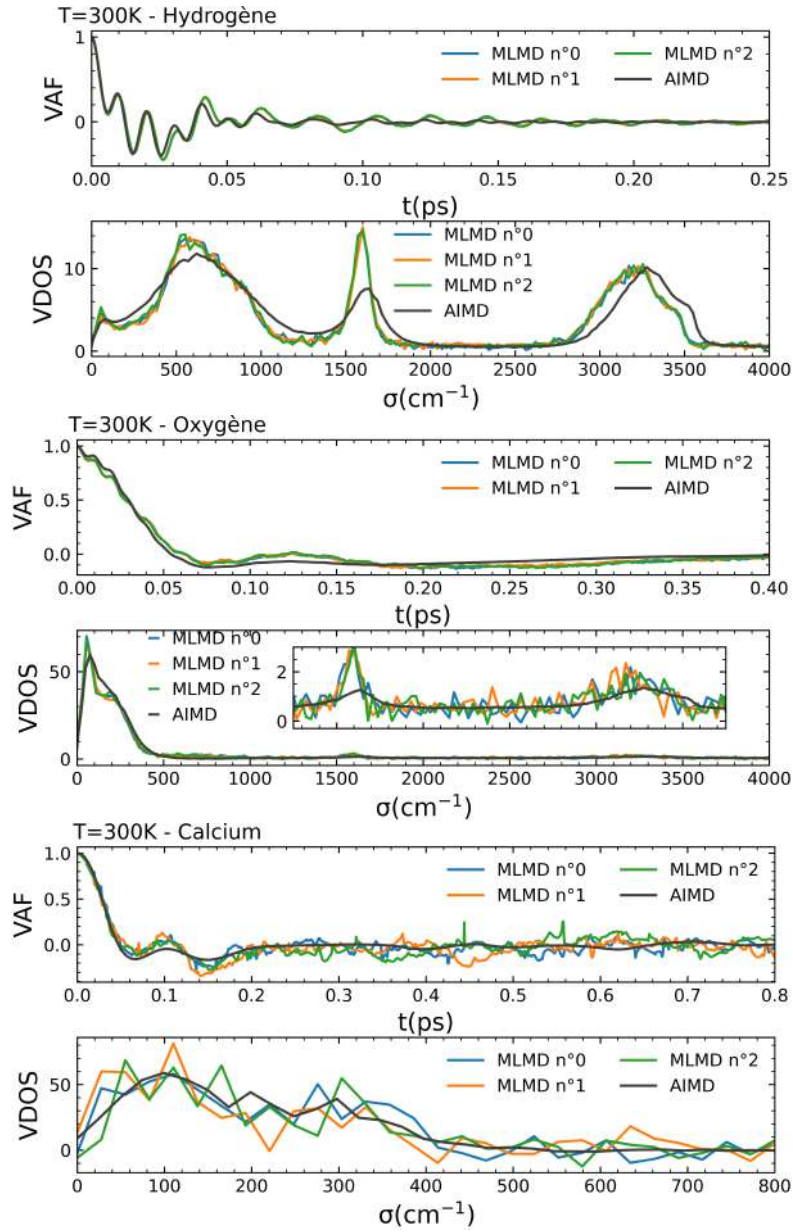


Figure IV.31: Fonction d'auto-corrélation des vitesses et les densités d'états vibrationnelles de l'hydrogène, de l'oxygène et du calcium des MLMD et l'AIMD à 300K.

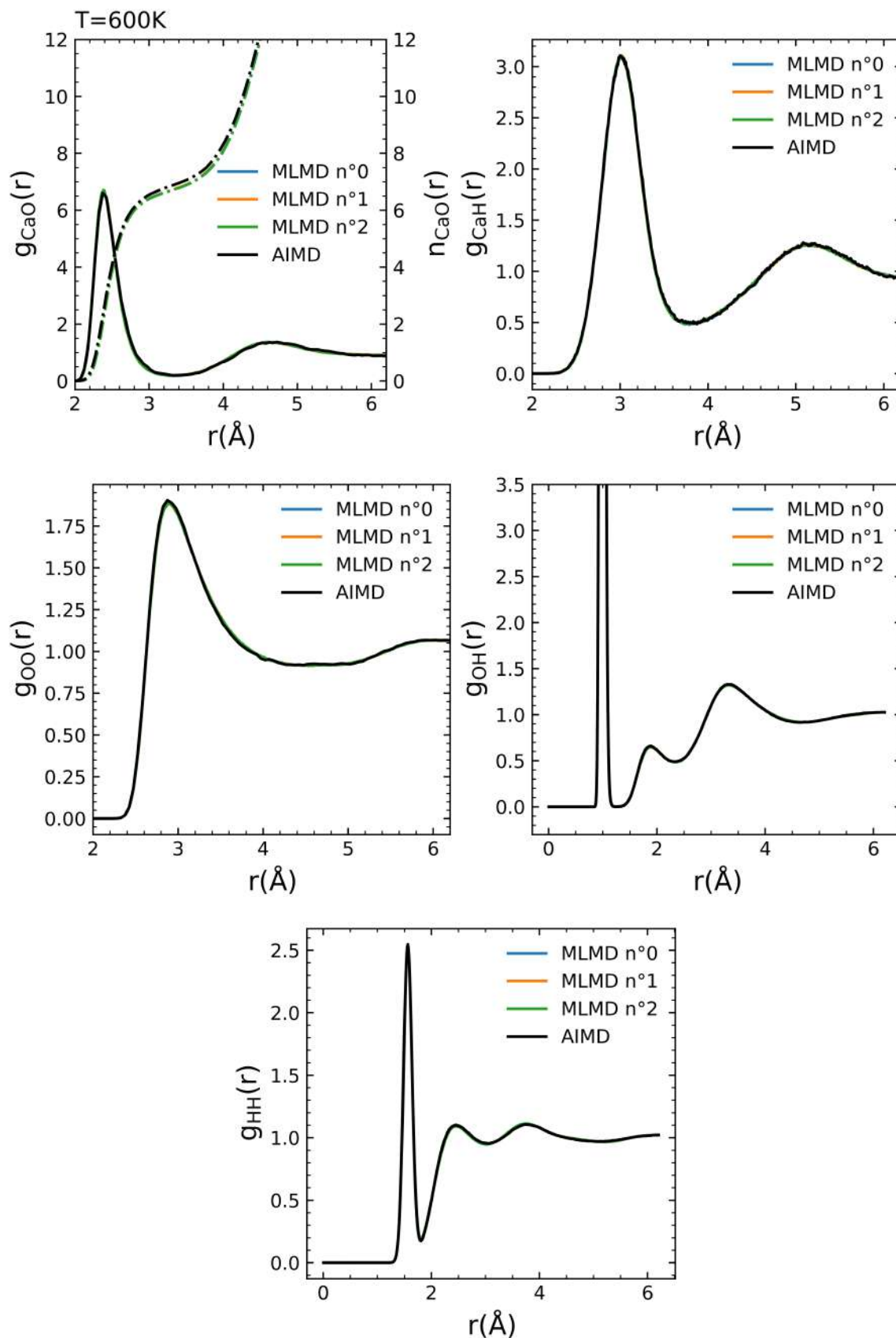


Figure IV.32: Fonctions de distribution radiale des configurations obtenues à partir des différentes MLMD à $T=600\text{K}$ et comparées à celles obtenues en AIMD.

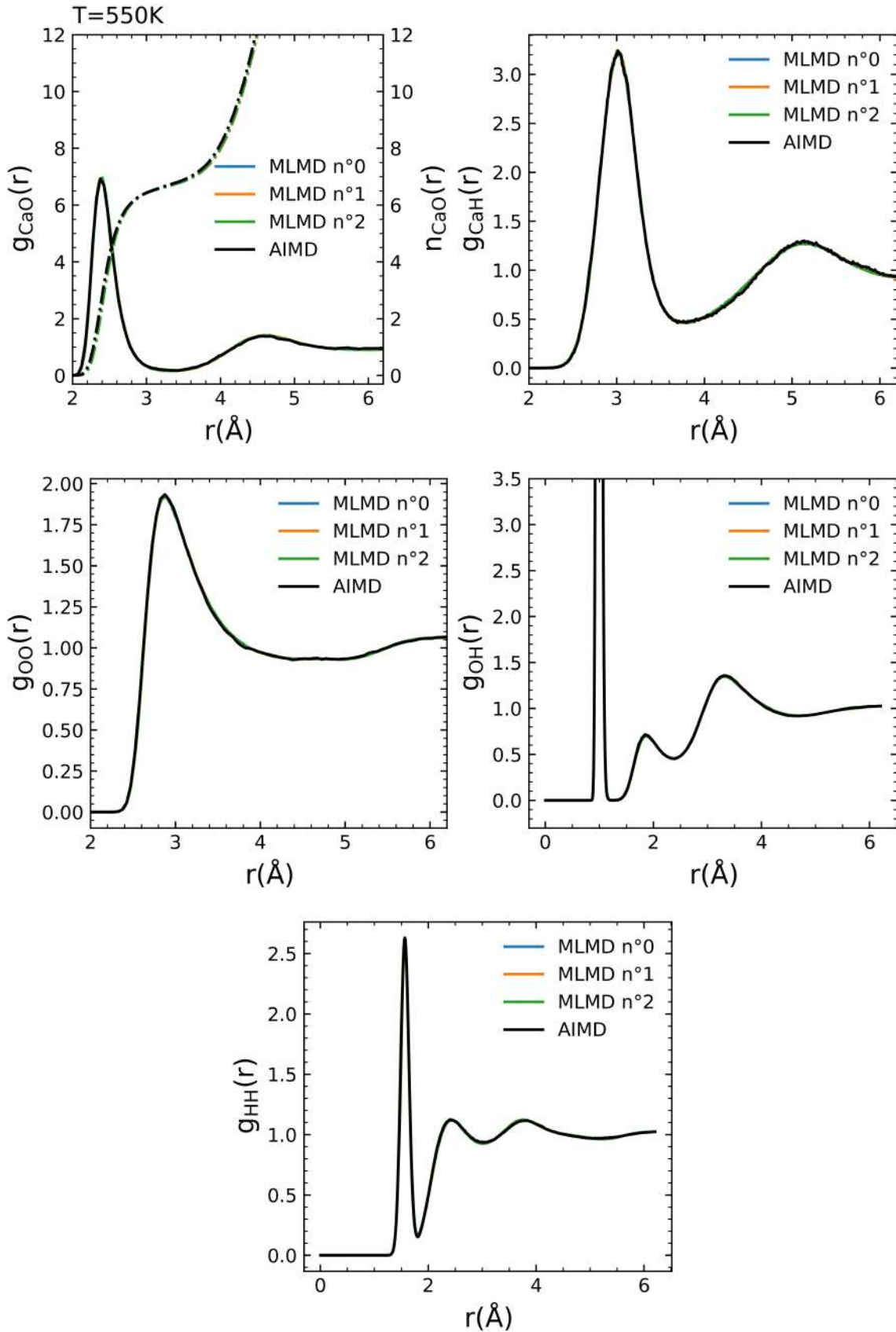


Figure IV.33: Fonctions de distribution radiale des configurations obtenues à partir des différentes MLMD à T=550K et comparées à celles obtenues en AIMD.

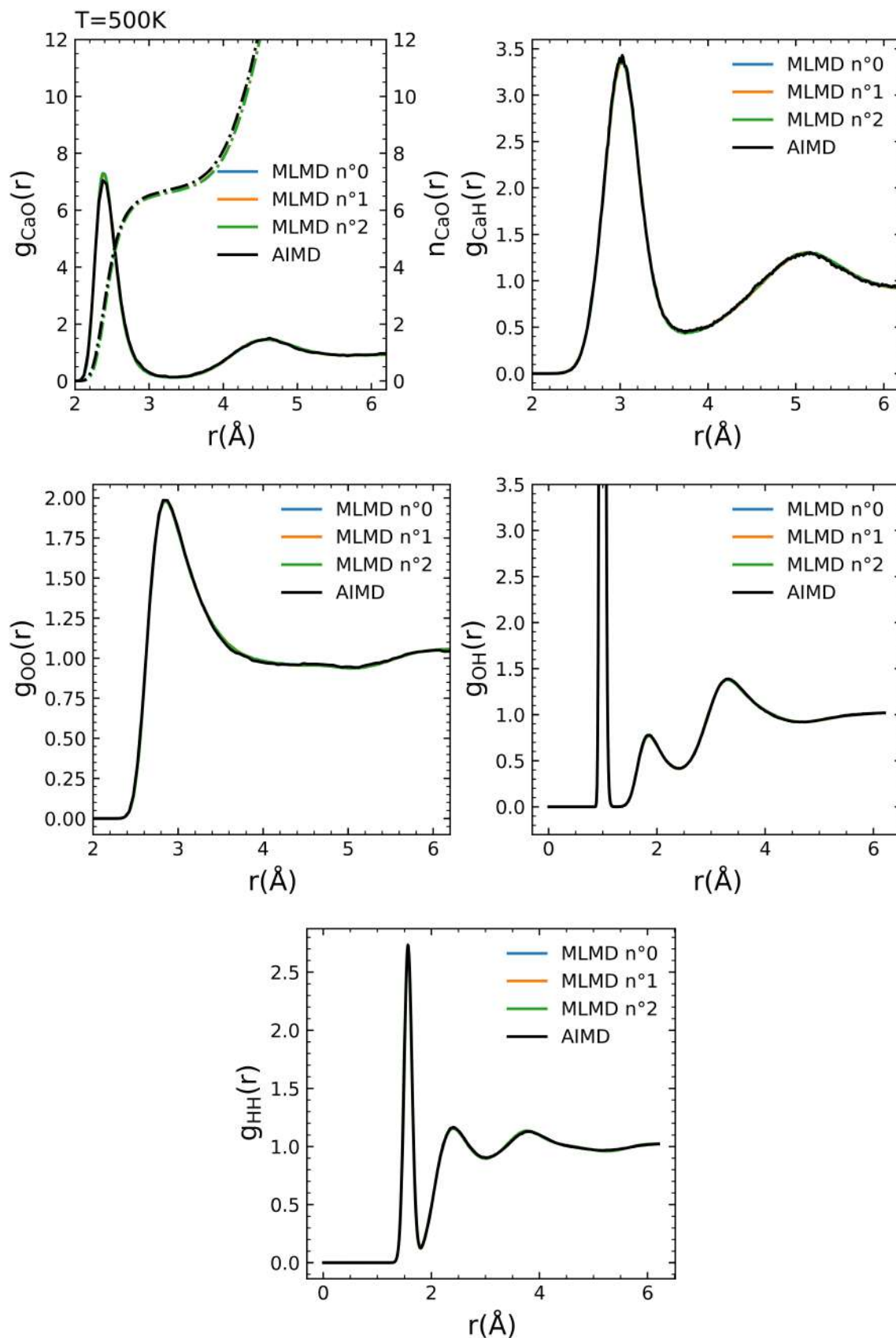


Figure IV.34: Fonctions de distribution radiale des configurations obtenues à partir des différentes MLMD à T=500K et comparées à celles obtenues en AIMD.

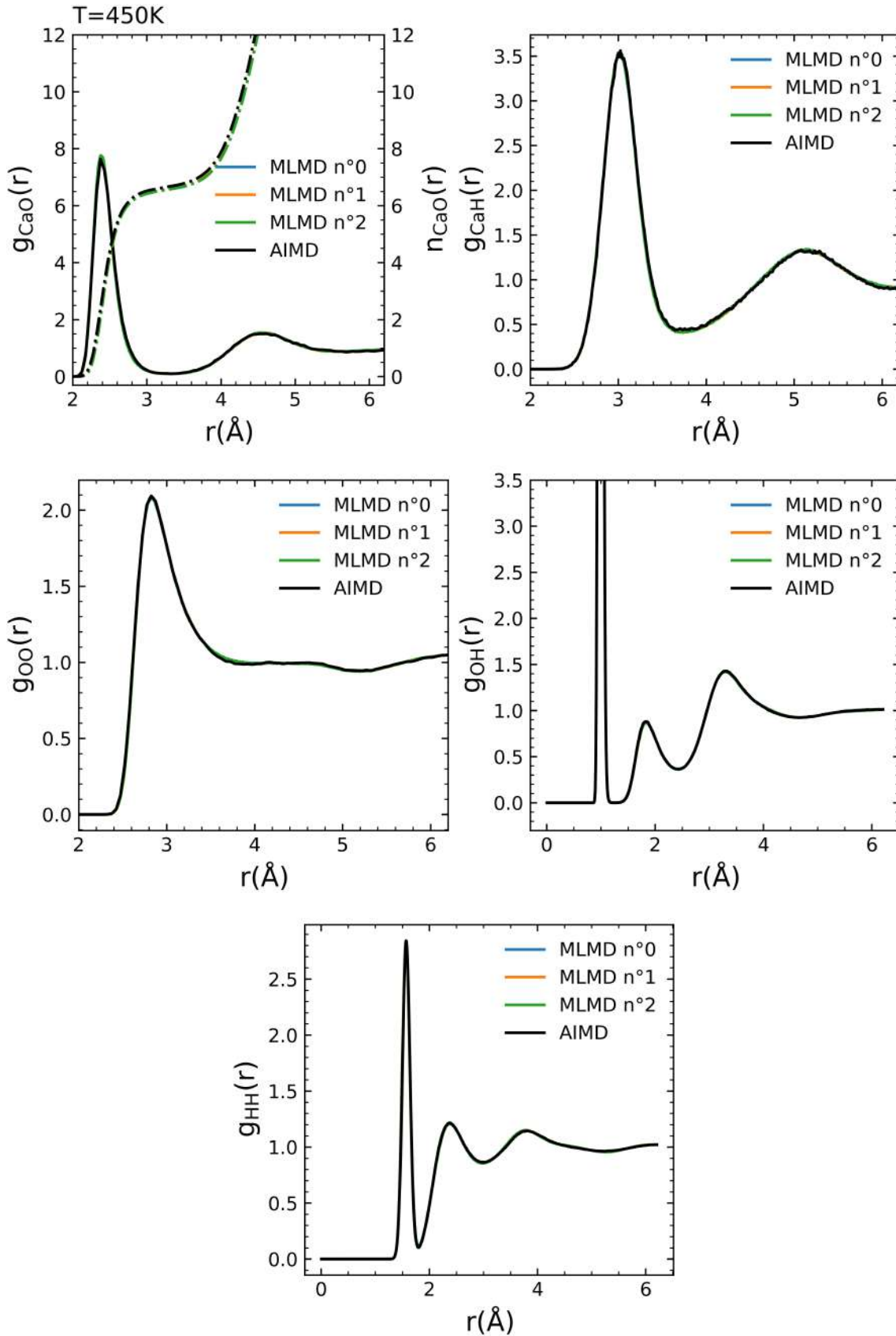


Figure IV.35: Fonctions de distribution radiale des configurations obtenues à partir des différentes MLMD à $T=450\text{K}$ et comparées à celles obtenues en AIMD.

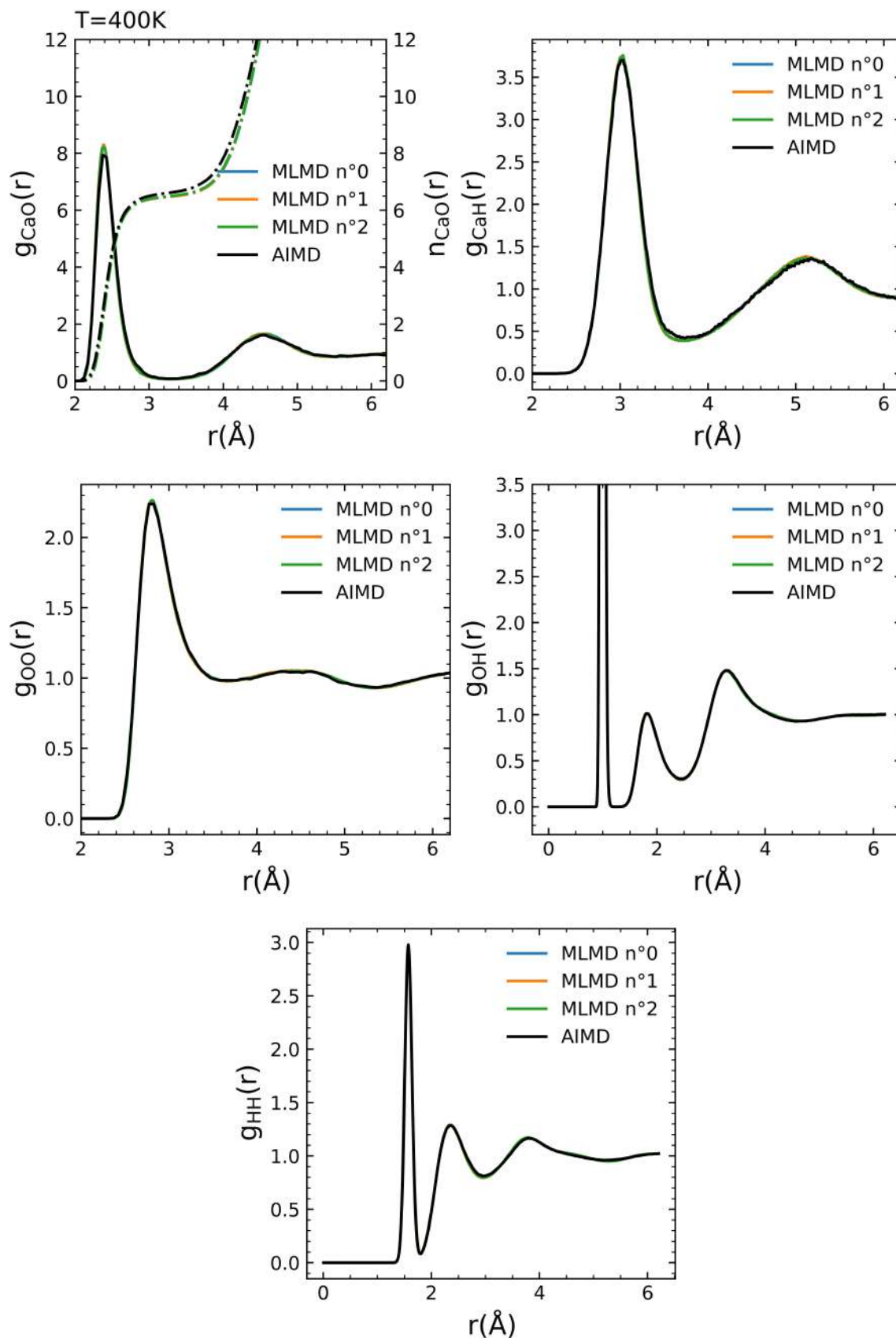


Figure IV.36: Fonctions de distribution radiale des configurations obtenues à partir des différentes MLMD à $T=400\text{K}$ et comparées à celles obtenues en AIMD.

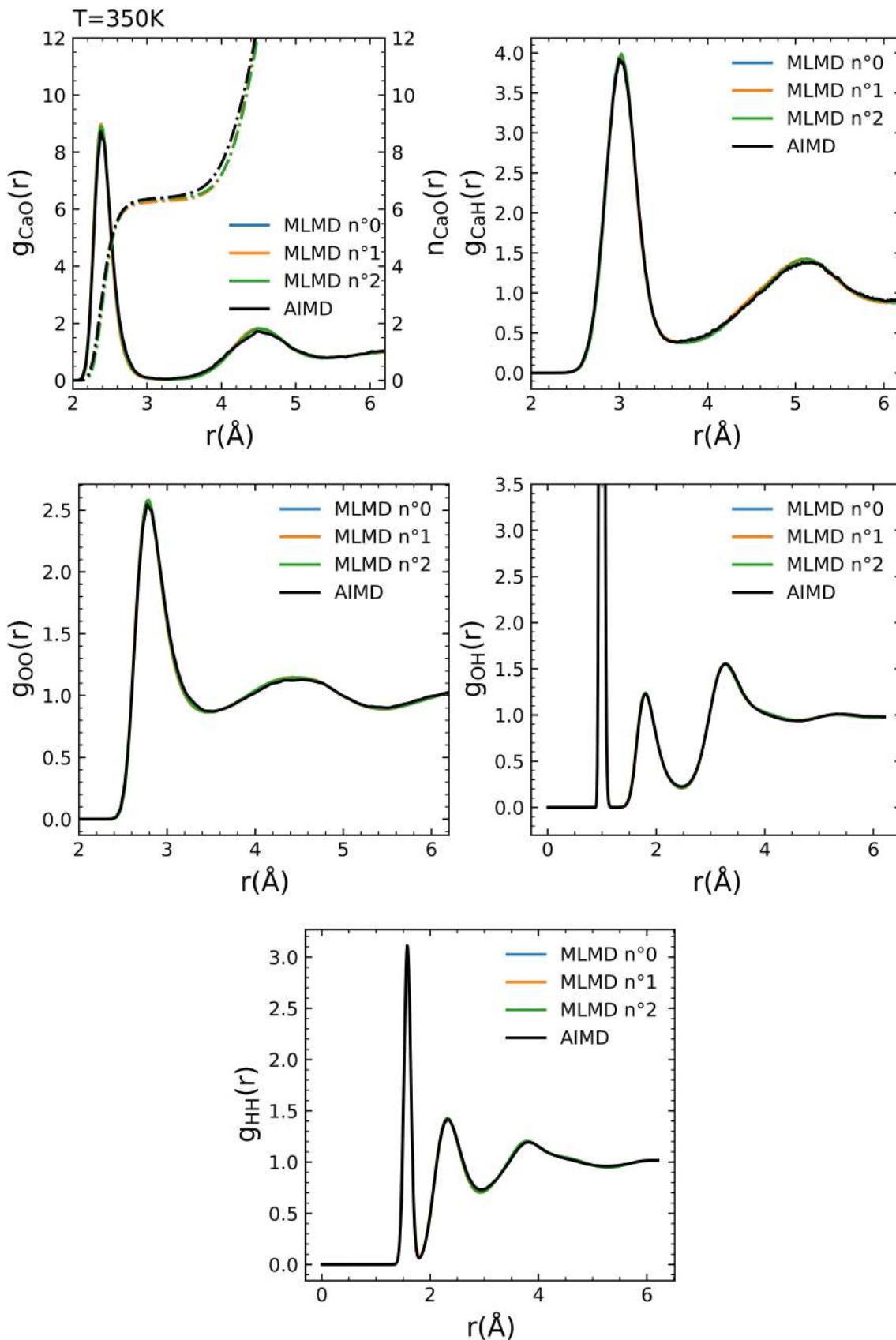


Figure IV.37: Fonctions de distribution radiale des configurations obtenues à partir des différentes MLMD à T=350K et comparées à celles obtenues en AIMD.

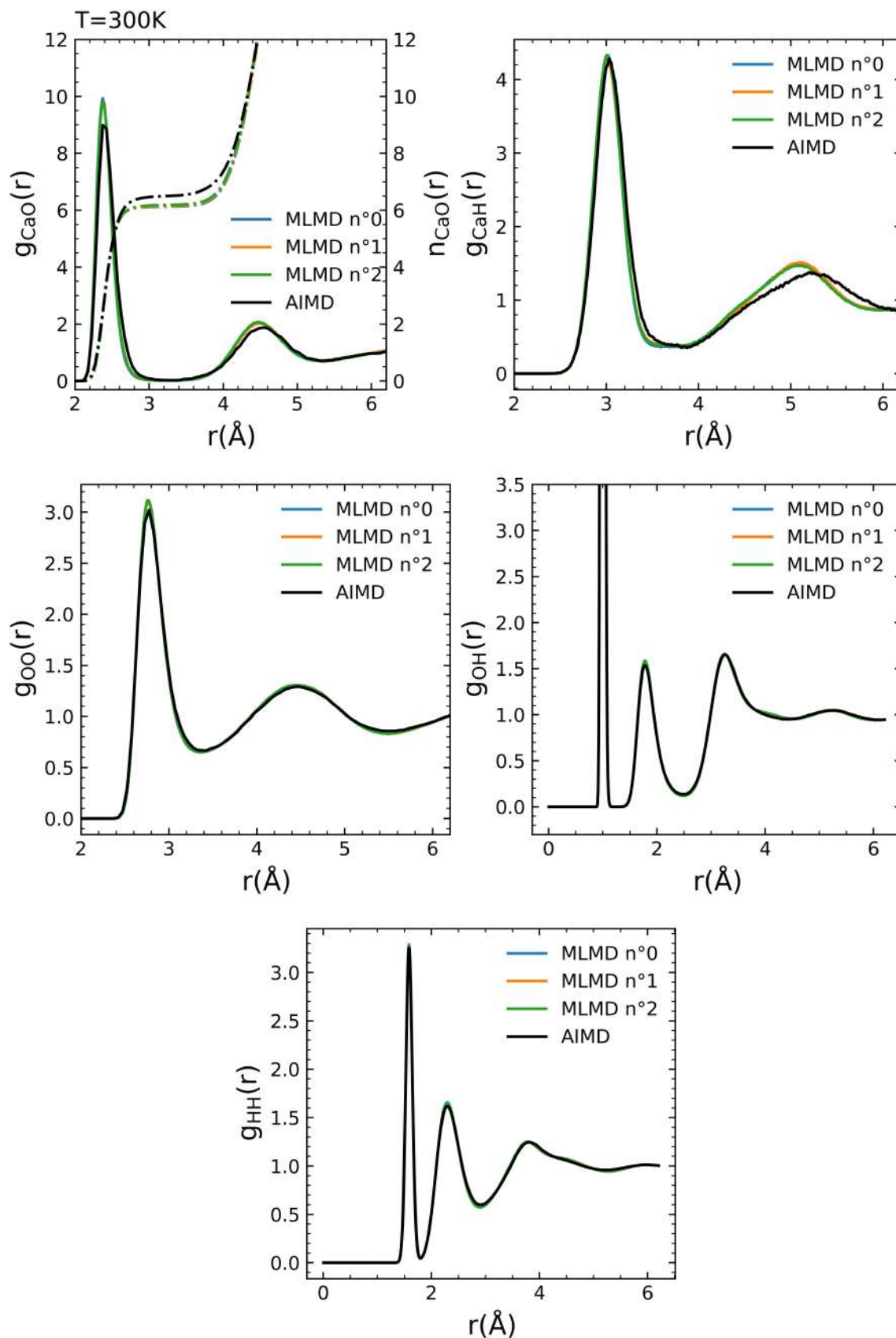


Figure IV.38: Fonctions de distribution radiale des configurations obtenues à partir des différentes MLMD à T=300K et comparées à celles obtenues en AIMD.

Calcul du fractionnement isotopique du calcium en solution aqueuse

Ce dernier chapitre est consacré au calcul du fractionnement isotopique entre le calcium ^{40}Ca et ^{44}Ca en solution aqueuse. On présente, dans une première partie, V.1, le calcul du facteur beta \mathcal{B} à partir d'une méthode employée pour le calcul du fractionnement isotopique des solides reposant sur l'approximation harmonique. Ces calculs sont effectués avec le potentiel ML de la partie précédente et en ab-initio avec la fonctionnelle BLYP-D2. Une deuxième partie, V.2, présente les résultats obtenus à partir des dynamiques moléculaires d'intégrales de chemin (PIMD) effectuées avec le potentiel ML. On y étudie aussi l'influence de la prise en compte des effets nucléaires quantiques sur la coordinence du calcium.

Table des matières

V.1 Approximation harmonique	137
V.1.1 Protocole	138
V.1.2 Résultats	139
V.2 Dynamique moléculaire d'intégrales de chemin	144
V.2.1 Protocole	144
V.2.2 Résultats	145

V.1 Approximation harmonique

De précédents calculs du fractionnement isotopique du calcium en solution aqueuse ont déjà été obtenus par les auteurs de [29] et [80] en considérant que les complexes aqueux peuvent être modélisés comme des groupes moléculaires isolés. L'objectif ici est de calculer le fractionnement isotopique du calcium en solution aqueuse, en prenant en compte l'ensemble du système, en se basant sur des méthodes utilisées pour le calcul de fractionnement isotopique des solides [35] employant l'approximation harmonique. Cela nous permettra aussi de comparer les valeurs obtenues avec celles résultants des dynamiques moléculaires d'intégrales de chemin et ainsi d'évaluer l'erreur commise lorsque l'approximation harmonique est employée.

V.1.1 Protocole

Le calcul du fractionnement isotopique a été effectué à la fois dans le cadre de la DFT, avec la fonctionnelle BLYP-D2, et avec le potentiel ML construit au chapitre IV. Dans les deux cas, nous avons à notre disposition des configurations atomiques issues des dynamiques moléculaires du chapitre III et du chapitre IV. Pour un sous ensemble de ces configurations, on détermine leur matrice dynamique que l'on diagonalise pour obtenir les modes de vibrations propres du système. Enfin, le facteur beta \mathcal{B} est obtenu à partir de l'équation II.21.

Calculs ab-initio

Contrairement au cas d'un solide, la dynamique moléculaire d'un liquide nécessite d'avoir un système assez grand pour éviter les effets de taille. Ainsi, l'obtention de la matrice dynamique par calcul ab-initio a un coût numérique important et son calcul, pour chaque configuration de la trajectoire, ou pour un grand nombre de configurations, est impossible.

Nous avons donc sélectionné un ensemble réduit de configurations régulièrement espacées le long de la trajectoire à 400K. Les configurations sont extraites de la trajectoire à 400K car nous avons montré qu'aux températures plus basses, l'échantillonnage des configurations était insuffisant. Cela permet ainsi d'obtenir plusieurs configurations de coordinences différentes.

Pour l'ensemble des 12 configurations retenues, l'obtention de la matrice dynamique se fait en trois étapes [35] :

- On effectue un calcul d'optimisation structurelle à 0K à l'aide du code PWscf de Quantum Espresso [46] afin d'obtenir la configuration correspondant au minimum de l'énergie potentielle. Si lors de l'optimisation, la coordinence change, la configuration est écartée. Cela a été observé pour une configuration.
- On récupère la configuration de l'étape précédente et on effectue un calcul auto-cohérent (voir partie I.1.1.6) afin d'obtenir l'état fondamental électronique.
- A partir de l'état fondamental atomique et électronique de la configuration, la matrice dynamique est déterminée au point Γ à l'aide du code PHonon de Quantum Espresso. Le calcul de la matrice dynamique uniquement au point Γ est justifié par la grande taille du système.

Pour les étapes 1 et 2, une énergie de cutoff pour les fonctions d'onde de 80Ry a été appliquée et une énergie de cutoff de 320Ry pour la densité électronique. L'optimisation s'arrête lorsque l'énergie totale est inférieure à $5.10^{-5}\text{Ry}(= 6,8.10^{-4}\text{eV})$ et que les forces résiduelles sont inférieures à $5.10^{-4}\text{Ry.Bohr}^{-1}(=1,3.10^{-2}\text{eV.}\text{\AA}^{-1})$. Une fois la matrice dynamique obtenue, on la diagonalise à l'aide du code dynmat de Quantum Espresso pour une masse du calcium de 39.9625u et 43.95548u, correspondant aux masses de l'isotope 40 et 44 du calcium. On obtient ainsi les fréquences propres du système pour les deux isotopes.

Bien que la taille du système permette d'effectuer le calcul uniquement au point Γ , le temps de simulation reste long. Nous avons alors aussi employé une méthode développée par [68] et adaptée par [35] qui consiste à considérer que seuls les modes de vibrations des atomes proches de l'isotope du calcium contribuent au fractionnement isotopique. Cela correspond aux molécules d'eau de la première sphère de

coordination (définie par la distance $r_{CaO}^{min,1}$, voir chapitre III). Dans ce cas, on effectue une optimisation structurale partielle en gardant fixe l'ensemble des molécules d'eau ne se trouvant pas dans la première sphère de coordination du calcium en utilisant le paramètre "*if_pos*" de code PWscf. Il est ensuite possible de spécifier au code PHonon un sous-ensemble d'atomes pour lesquels le calcul est fait.

Une fois la matrice dynamique ainsi obtenue, il est alors nécessaire, avant de la diagonaliser, de la modifier afin de spécifier les atomes qui sont fixés. Pour cela, on remplace leur masse par une masse très grande.

Dans la suite, afin de différencier les résultats obtenus avec optimisation et calcul de phonon partiel, on utilisera l'indice "*part*".

Potentiel ML

Les calculs du facteur beta avec le potentiel ML ont été réalisés avec le code LAMMPS [109]. Les trajectoires des dynamiques moléculaires de la partie IV.3 étant plus longues, permettant ainsi un meilleur échantillonnage des configurations, 100 configurations régulièrement espacées ont été récupérées de la trajectoire MLMD n°0 à 300K. Pour chaque configuration, on lance deux calculs de matrice dynamique. Un pour une masse du calcium de 39.9625u et un pour une masse de 43.95548u. Dans les deux cas:

1. on effectue une optimisation structurale à l'aide de la commande "*minimize*". L'optimisation s'arrête lorsque l'énergie du système est inférieure à 10^{-12} eV ou que les forces résiduelles sont inférieures à 10^{-12} eV.Å⁻¹.
2. On détermine les matrices dynamiques à partir de la commande "*dynamical_matrix*".

Les matrices dynamiques sont ensuite diagonalisées et les facteurs beta calculés à l'aide des modes de vibrations propres.

Afin de comparer avec les calculs ab-initio, cette opération a aussi été réalisée pour 100 configurations régulièrement espacées de la trajectoire MLMD n°0 à 400K.

V.1.2 Résultats

Afin de comparer les facteurs beta \mathcal{B} obtenus en ab-initio après optimisation et calcul de la matrice dynamique en prenant en compte tous les atomes, ou seulement ceux de la première sphère de coordination \mathcal{B}_{part} , on trace figure V.1 $1000\ln\mathcal{B}$ en fonction de $1000\ln\mathcal{B}_{part}$. Si les deux méthodes produisent les mêmes résultats, les points du graphe devraient s'aligner le long de la droite $1000\ln\mathcal{B} = 1000\ln\mathcal{B}_{part}$. Pour les configurations de coordination égale à 6, les deux méthodes donnent des résultats proches, avec un plus grand désaccord trouvé de 4%. En revanche, pour les configurations de coordination 7, la détermination du facteur beta à partir d'une optimisation et calcul de phonon partiel ne reproduit pas ceux obtenus en considérant l'ensemble du système, avec des désaccords allant jusqu'à 13%. Ces résultats indiquent une influence non négligeable de la seconde sphère de coordination.

Les différences de résultats peuvent provenir du calcul partiel des phonons, ou bien de l'optimisation partielle. Cette dernière est supposée préserver au mieux la structure liquide. De plus amples investigations seraient nécessaires.

Par la suite, on comparera les résultats obtenus avec le potentiel ML avec ceux de

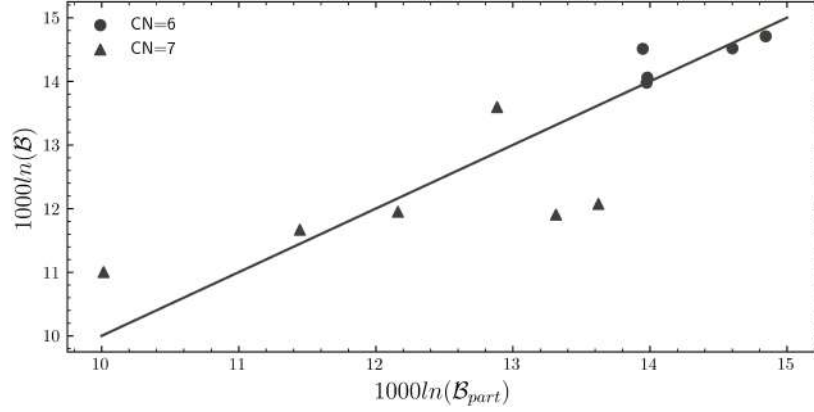


Figure V.1: $1000\ln\mathcal{B}_{part}$ obtenue après optimisation structurale et calcul des modes de vibrations pour les atomes de la première sphère de coordination en fonction de $1000\ln\mathcal{B}$ obtenue après optimisation structurale et calcul des modes de vibrations de l'ensemble de atomes. La droite $1000\ln\mathcal{B} = 1000\ln\mathcal{B}_{part}$ correspond au cas où les deux méthodes donnent le même résultats. La coordinnence (CN) de la configuration est déterminée par le nombre d'atomes d'oxygène à une distance inférieure à $r_{CaO}^{min,1}$.

l'ab-initio où les modes de vibration ont été calculés en prenant en compte l'ensemble du système.

Les figures V.2 et V.3 montrent les résultats des facteurs beta obtenus avec le potentiel ML, pour les configurations des trajectoires à 300K et 400K, en fonction de la coordinnence et de la distance moyenne calcium-oxygène, r_{CaO}^{mean} , des oxygènes de la première sphère de coordination. On y trouve aussi les résultats des calculs ab-initio. Que ce soit avec le potentiel ML, ou par calcul ab-initio, on observe une diminution du facteur beta avec la coordinnence et la distance r_{CaO}^{mean} . Ce résultat est cohérent avec une favorisation des isotopes lourds pour de plus faibles coordinnences.

Pour le calcul du facteur beta des configurations de la trajectoire à 300K, pour une même valeur de la coordinnence (déterminée par le nombre d'atome d'oxygène à une distance inférieure à $r_{CaO}^{min,1}$), les calculs ab-initio sont systématiquement supérieurs à ceux déterminés avec le potentiel ML. Pour les configurations de la trajectoire à 400K, on observe le même phénomène pour les configurations de coordinnence égale à 6. Lorsque l'on regarde la variation de $1000\ln\mathcal{B}$ en fonction de la distance r_{CaO}^{mean} , on observe que, pour une même valeur de la coordinnence, la variation des valeurs de r_{CaO}^{mean} est plus grande pour les calculs utilisant le potentiel ML. Pour des valeurs de r_{CaO}^{mean} proches, l'écart des valeurs de $1000\ln\mathcal{B}$, entre les calculs ab-initio et avec le potentiel ML, sont de l'ordre de ~ 1 pour les configurations de coordinnence 6. Pour les configurations de coordinnence 7, il est plus difficile de conclure, avec des écarts de l'ordre de $\sim 0.5-1$ pour les configurations extraites de la trajectoire à 300K et des résultats proches pour les configurations extraites de la trajectoire à 400K. Pour chaque valeur de coordinnence, on reporte table V.1, la valeur moyenne des $1000\ln(\mathcal{B})$. L'incertitude est prise comme étant égale à l'écart-type divisé par la racine carrée du nombre de valeurs. Pour le potentiel ML, les valeurs de $1000\ln(\mathcal{B})$ pour une coordinnence de 6 et un coordinnence de 7 des configurations extraites des trajectoires

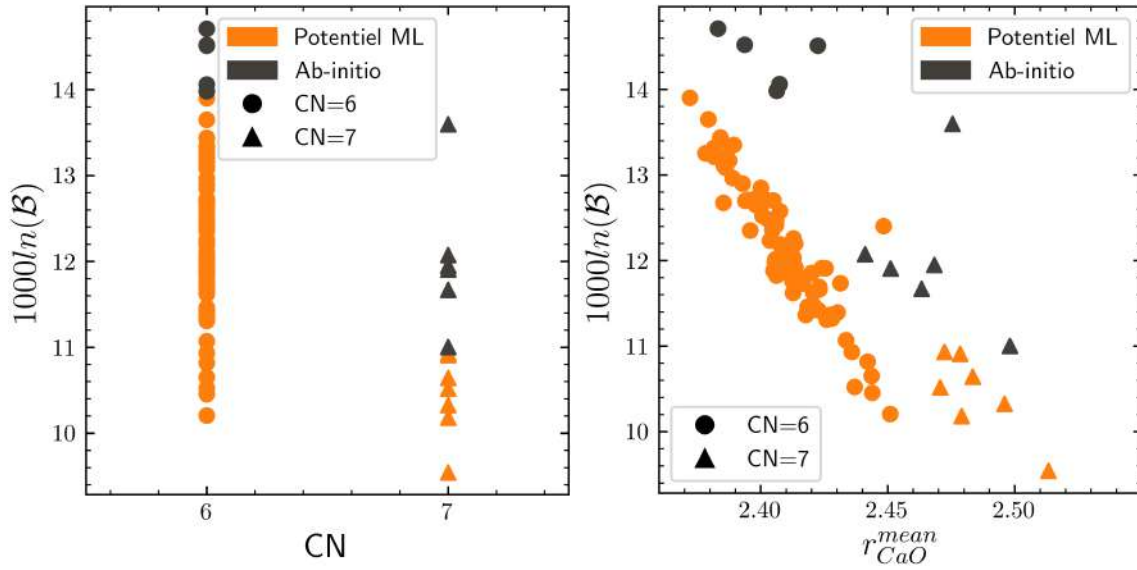


Figure V.2: $1000\ln(\mathcal{B})$ en fonction de la coordinnce (gauche) et de la distance moyenne Ca-O (droite) pour les oxygènes de la première sphère de coordinnce obtenus par calcul ab-initio et avec le potentiel ML. Pour le calcul du facteur \mathcal{B} avec le potentiel ML, les configurations proviennent de la trajectoire MLMD à 300K.

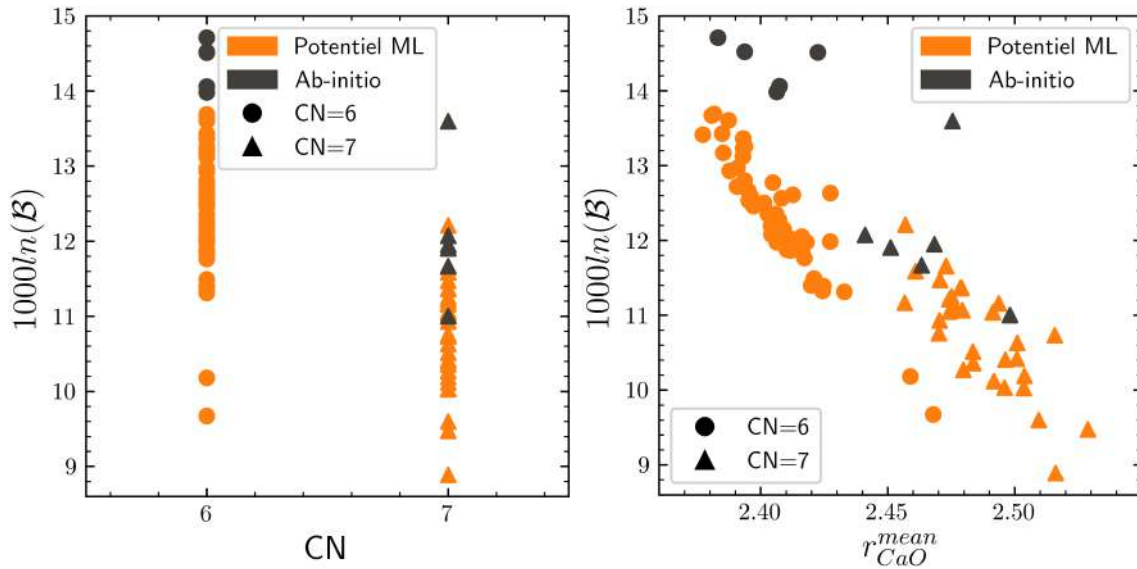


Figure V.3: $1000\ln(\mathcal{B})$ en fonction de la coordinnce (gauche) et de la distance moyenne Ca-O (droite) pour les oxygènes de la première sphère de coordinnce obtenus par calcul ab-initio et avec le potentiel ML. Pour le calcul du facteur \mathcal{B} avec le potentiel ML, les configurations proviennent de la trajectoire MLMD à 400K.

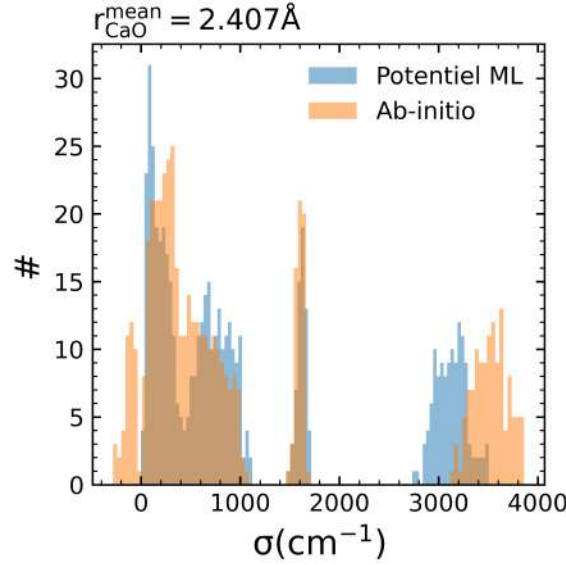


Figure V.4: Histogramme des nombres d'ondes des modes propres obtenus par calcul ab-initio et avec le potentiel ML pour deux configurations ayant une distance $r_{CaO}^{mean}=2.407\text{\AA}$ pour le calcium ^{40}Ca .

à 300K et 400K sont en accord. Les résultats des calculs ab-initio sont supérieurs à ceux résultant du potentiel ML avec une valeur de 14.4 ± 0.2 contre 12.2 ± 0.1 et 12.2 ± 0.2 pour une coordinence de 6 et de 12.0 ± 0.3 contre 10.4 ± 0.1 et 10.7 ± 0.2 pour une coordinence de 7. Ceci peut être dû à trois choses. La première est que le nombre de configurations dont les facteurs beta calculés en ab-initio sont extraits, est faible. Davantage de calculs pourrait donner une variation des valeurs plus grande. La deuxième raison peut être le choix des valeurs de l'énergie et des forces limites pour l'arrêt de l'optimisation structurale. Pour les calculs avec le potentiel ML, ces valeurs ont été testées afin d'obtenir convergence des résultats et un nombre de fréquences négatives faible. En revanche, pour les calculs ab-initio, les valeurs utilisées sont des valeurs usuelles et leurs modifications n'ont pas été investiguées. On peut voir, si l'on trace l'histogramme des nombres d'ondes σ obtenus avec le potentiel ML et par calcul ab-initio (voir figure V.4), pour deux configurations ayant des distances r_{CaO}^{mean} identiques, que pour le calcul ab-initio, un nombre de fréquences négatives non négligeable est obtenu indiquant que l'optimisation structurale n'a pas été suffisamment poussée.

Concernant les différences pour des valeurs du nombre d'ondes de l'ordre de $\sim 3000\text{cm}^{-1}$, elles correspondent à celles déjà observées, pour les modes de déformation, sur les VDOS de l'hydrogène au chapitre précédent figure IV.31 et ne devraient pas influencer le fractionnement isotopique. Finalement, il est possible que le potentiel ML ne puisse pas reproduire correctement les modes de vibrations propres du système.

Finalement, la valeur moyenne de l'ensemble des $1000\ln\mathcal{B}$ des calculs ab-initio, de 13.1, est plus élevée que celles obtenues avec le potentiel ML qui sont de 12.0 et 11.7, pour les configurations extraites des trajectoires à 300K et 400K respectivement (voir tableau V.1). Il est important de noter que les coordinences

Méthode	Moyenne $1000 \ln \mathcal{B}$			
	CN=6	CN=7	Totale	CN _{configurations}
Ab-initio	14.4±0.2	12.0±0.3	13.1±0.4	6.65
Potentiel ML configuration trajectoire à 300K	12.2±0.1	10.4±0.1	12.0±0.1	6.08
Potentiel ML configuration trajectoire à 400K	12.3±0.2	10.7±0.2	11.7±0.2	6.38

Table V.1: Valeurs moyennes de $1000 \ln \mathcal{B}$ pour les configurations de coordinence 6, 7, 8 ainsi que la valeur moyenne totale obtenues par calcul ab-initio et avec le potentiel ML pour des configurations récupérées des trajectoires MLMD à 300K et 400K. CN_{configurations} correspond à la moyenne des coordinences des configurations utilisées pour le calcul du facteur beta.

moyennes des configurations utilisées, reportées tableau V.1, pour le calcul du facteur beta peuvent varier de celles des trajectoires complètes. Pour les calculs ab-initio, elle est de 6.65, légèrement plus élevée que la valeur de 6.51 de l'AIMD. Cela suggère que la valeur du facteur beta serait, en ab-initio, supérieure à 13.09. Sans oublier le fait que, comme discuté au chapitre III, la valeur de la coordinence de l'AIMD à 300K comporte une grande incertitude. Pour une étude plus approfondie, il serait nécessaire d'effectuer un plus grand nombre de calcul ab-initio pour des configurations extraites d'une longue trajectoire AIMD, où la valeur de la coordinence est correctement convergée.

Concernant les calculs avec le potentiel ML, pour les configurations de la trajectoire à 300K, la valeur moyenne de la coordinence des configurations est de 6.08, proche de celle de la MLMD qui est de 6.12. On peut alors présumer que la valeur obtenue est fiable. Pour les calculs avec les configurations de la trajectoire à 400K, la valeur moyenne de la coordinence est de 6.38, plus élevée que la coordinence de la MLMD à 300K, ce qui peut expliquer la valeur plus faible de la moyenne de $1000 \ln \mathcal{B}$ trouvée comparé à celle des configurations de la trajectoire à 300K.

Le fractionnement isotopique du calcium à partir de calculs ab-initio, pour différentes fonctionnelles, a été étudié plus amplement par J.Aufort, pour des fractionnement entre solide et liquide (travaux en attente de publication). Dans cette étude, pour la fonctionnelle BLYP-D2, les valeurs moyennes de $1000 \ln \mathcal{B}$ trouvées, pour des configurations de coordinence 6 et de coordinence 7, sont de 13.2 et 11.7, respectivement. La valeur moyenne de la coordinence des configurations étudiées est de 6.6 avec une valeur moyenne totale de $1000 \ln \mathcal{B}$ de 12.6. Ces résultats montrent une diminution des valeurs du facteur beta de 1.2 pour les configurations de coordinence 6 et de 0.3 pour les configurations de coordinence 7 comparée aux calculs ab-initio présentés dans cette partie. Ils sont en meilleur accord, mais toujours plus élevés, avec les résultats obtenus avec le potentiel ML.

V.2 Dynamique moléculaire d'intégrales de chemin

Le calcul du fractionnement isotopique à partir de dynamiques moléculaires d'intégrales de chemin (PIMD) permet de s'affranchir de l'approximation harmonique. On présente les résultats obtenus, en utilisant le potentiel ML et les compare à ceux de la partie précédente afin d'évaluer les effets anharmoniques éventuels.

V.2.1 Protocole

Les simulations de PIMD ont été effectuées à l'aide du code I-PI [22] couplé avec le code Lammmps [109] pour le calcul des forces. Le système étudié est composé d'un ion calcium Ca^{2+} et de 62 molécules d'eau dans une boîte cubique de 12.41Å, identique aux systèmes des études des chapitres III et IV. On travaille dans l'ensemble canonique avec un thermostat de Langevin "PILE_G" (voir partie I.3.4), contrôlé par le paramètre τ_0 . Le pas de temps Δt a été fixé égal à 0.25fs. Les forces ont été calculées à partir du potentiel interatomique ML généré au chapitre IV.

Contrairement aux études précédentes où des thermostats de type Nose-Hoover ont été utilisés, le code I-PI utilise des thermostats de type Langevin. C'est pourquoi, afin de tester l'efficacité de l'échantillonnage des configurations et la convergence de la coordinence avec ces thermostats, nous avons, dans un premier temps, effectué des simulations de 2ns pour un nombre de beads égal à 1 pour trois différentes valeurs de τ_0 (10, 50 et 100). Les résultats ont été comparés à ceux de la partie IV.3.

Nous avons ensuite effectué des simulations à 400K, où la convergence de l'estimateur du fractionnement isotopique, ainsi que de la coordinence, est plus rapide, afin de tester les différents paramètres des simulations. Nous avons étudié l'influence du nombre de beads, de la valeur du paramètres τ_0 et du choix de la masse du calcium. Pour l'étude de l'influence du nombre de beads, une valeur de $\tau_0=10$ a été choisie ainsi qu'une masse du calcium de 41.88765u, correspondant à une valeur du paramètre de masse $\lambda=0.5$ en utilisation la paramétrisation de l'équation II.44. Les simulations ont été effectuées pour des valeurs de nombre beads égales à 4, 8, 16, 32 et 64 pour un temps total de simulation de 500ps.

Une même masse a été choisie pour l'étude de l'influence du paramètre τ_0 avec un nombre de beads égal à 8 pour des temps de simulations de 500ps.

Finalement, l'influence de la masse du calcium a été étudié pour $\tau_0=10$, pour des simulations comprenant 8 beads et des masses de 39.96259u, 40.90819u, 41.88785u, 42.9020u et 43.95548u, correspondant à des valeurs de λ de 0, 0.25, 0.5, 0.75 et 1, toujours avec la paramétrisation de l'équation II.44.

Enfin, une simulation à 300K avec 8 beads a été effectuée, pour une masse du calcium de 41.88785u et une valeur de $\tau_0=10$. Les résultats sont comparés à ceux obtenus en ab-initio.

Pour l'ensemble des simulations, on récupère les positions et les forces tous les 100pas. Une simulation, utilisant les potentiels empiriques Ca-O et q-SPC/fw présenté tableau I.1, à 300K, a été effectuée dans le but de comparer les résultats avec ceux du potentiel ML et de juger de sa nécessité.

Pour l'ensemble des simulations, la masse des atomes d'hydrogène a été fixée à

MLMD	CN ⁽²⁾	u(CN ²)
NH n°0	6.12	0.05
NH n°1	6.15	0.07
NH n°2	6.15	0.08
Langevin $\tau_0 = 10$	6.12	0.05
Langevin $\tau_0 = 50$	6.25	0.10
Langevin $\tau_0 = 100$	6.15	0.05

Table V.2: Coordinences moyennes finales CN⁽²⁾ des MLMDs à 300K utilisant un thermostat de Nose-Hoover (NH) et des MLMDs à 300K utilisant un thermostat de type Langevin pour différentes valeurs de τ_0 . Les incertitudes u(CN²) sont calculées à partir de la relation $2\sigma/\sqrt{N_{steps}}$ où N_{steps} est le nombre de pas de la dynamique moléculaire.

1.007940u et celle des atomes d'oxygène à 15.999400u.

V.2.2 Résultats

Comparaison des thermostats

On présente les résultats obtenus avec le thermostat de type Langevin, pour des valeurs de τ_0 égales à 10, 50 et 100, à 300K, et les compare avec ceux obtenus pour les trois MLMDs à 300K effectuées avec les thermostats de Nose-Hoover de la partie IV.3.2. On souhaite, pour les deux thermostats, obtenir une structure de l'eau identique et surtout, une coordinence moyenne du calcium identique.

Les fonctions de distribution radiale obtenus sont présentées figure V.13. Pour toutes les simulations, l'ensemble des $g(r)$ semblent identiques. Comme pour les MLMD du chapitre précédent, on calcule l'évolution de la coordinence en fonction du temps, $CN(t)$, à partir de la formule III.3. A partir de ces données, on utilise la même méthode que dans la partie III, en considérant que l'ensemble des valeurs $CN(t)$ forme une chaîne de Markov, pour déterminer l'écart-type, σ , de la coordinence moyenne. On trouve en annexe , figure 27, les évolutions des $CN(t)$ obtenues. La figure V.5 présente l'évolution de la valeur moyenne, en fonction du temps, pour les trois simulations effectuées pour $\tau_0=10,50$ et 100, ainsi que celles des trois MLMD où un thermostat de Nose-Hoover a été utilisé. Les valeurs des coordinences finales CN⁽²⁾, ainsi que leurs incertitudes, se trouvent tableau V.2. On observe que pour une valeur de $\tau_0=10$ et $\tau_0=100$, les valeurs finales de la coordinence moyenne sont de 6.12 et 6.15, respectivement, similaires à celles des trois MLMDs utilisant ses thermostats de Nose-Hoover (6.12, 6.15 et 6.15). Pour la simulation avec $\tau_0=50$, on obtient une valeur finale de la coordinence moyenne légèrement supérieure, de 6.25, et qui converge moins rapidement. Néanmoins, en prenant en compte les incertitudes, la valeur obtenue est en accord avec les autres simulations.

L'utilisation de thermostat de type Langevin semble permettre d'obtenir les mêmes résultats qu'avec des thermostats de Nose-Hoover. En revanche, il est difficile de conclure quant à l'influence du paramètre τ_0 sur nos simulations.

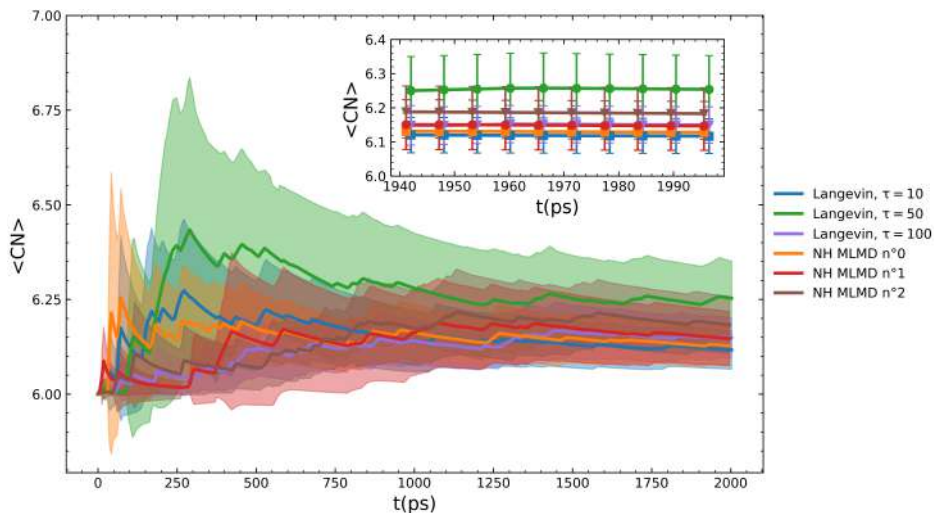


Figure V.5: Evolution de la coordinence moyenne pour les MLMD utilisant un thermostat de Nose-Hoover (NH) et celles utilisant un thermostat de type Langevin pour trois valeurs de τ_0 . Les incertitudes sont calculées à partir de l'écart-type et de la formule $2\sigma/\sqrt{N_{steps}}$, où N_{steps} est le nombre de pas de la simulation. L'encart en haut à droite correspond à un zoom sur les dernières 60ps des simulations.

Choix du nombre de beads P

On teste tout d'abord le nombre de beads P qu'il est nécessaire d'utiliser dans nos simulations de PIMD afin de caractériser les effets quantiques. Pour cela, on effectue différentes simulations à 400K, avec $\tau_0=10$, pour 5 valeurs différentes de P : 4, 8, 16, 32 et 64. Pour chacune de ces simulations de 500ps, la valeur de $1000\ln(\mathcal{B})$ en fonction du temps est calculée à partir de la méthode présentée partie II.3.2 et les résultats se trouvent figure V.6 Pour des valeurs de P de 8, 16, 32 et 64, les $1000\ln(\mathcal{B})$ convergent vers des valeurs similaires, autour de ~ 6.4 , après environ 250ps. Concernant la simulation à P=4, la valeur de $1000\ln(\mathcal{B})$ converge aussi, après environ 250ps, mais vers une valeur plus faible d'environ 6.15. On en conclut que, pour prendre en compte les effets quantiques nucléaires, il est nécessaire d'avoir au moins 8 beads et c'est cette valeur qui sera utilisée par la suite.

Choix du paramètre τ_0

Afin d'étudier plus en détail l'influence du paramètre τ_0 sur les simulations de PIMD, et sur la valeur de $1000\ln(\mathcal{B})$, on effectue des PIMD, à 400K, pour des valeurs de τ_0 de 10, 50, 100, 200, 1000 et 2000 avec une masse du calcium de 41.88785u.

On regarde tout d'abord la structure de l'eau obtenue ainsi que celle de la première sphère de coordinence du calcium et les compare à celles obtenues avec la MLMD n°0 à 400K de la partie IV.3. Les fonctions de distribution radiale se trouvent figure V.14. Pour l'eau, les PIMD donnent une même structure peu importe la valeur de τ_0 choisie. Les $g_{oo}(r)$ sont proches de celles obtenues en MLMD. En revanche, on observe une diminution de la valeur des amplitudes des extrema, ainsi qu'un élargissement des pics des $g_{hh}(r)$ et pour le premier maxima des $g_{oh}(r)$. Le même phénomène est observé pour le premier maxima des $g_{cah}(r)$. Ceci est cohérent avec le fait que les effets quantiques nucléaires sont plus importants pour l'hydrogène de par sa faible

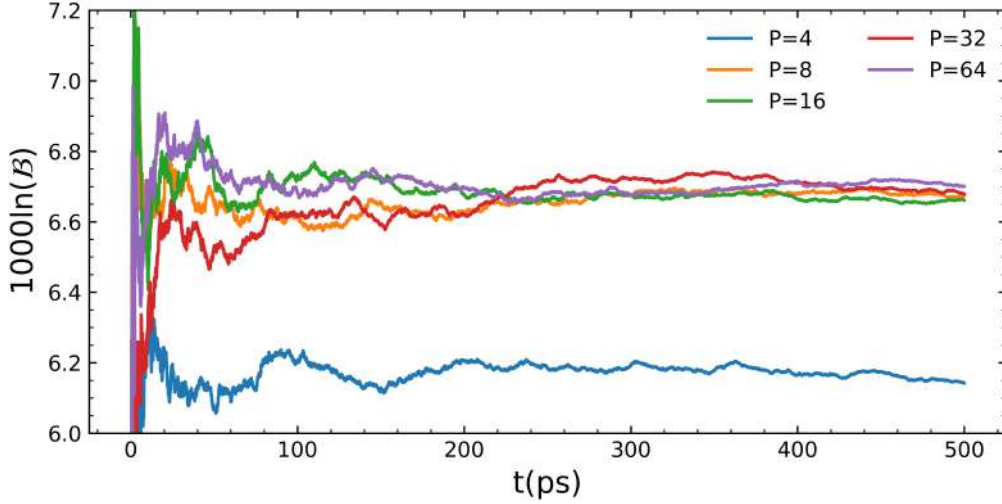


Figure V.6: Evolution de $1000\ln(\mathcal{B})$ en fonction du temps pour des simulations avec un nombre de beads P différent. Les simulations ont été effectuées à 400K avec le potentiel ML.

masse. L'élargissement des pics des $g(r)$ traduit le fait que la délocalisation de l'atome d'hydrogène est plus grande que pour l'oxygène et le calcium et correspond dans nos simulations à des chemins ayant une extension spatiale plus grande. Concernant les $g_{cao}(r)$, bien qu'il y ait peu de différence entre l'allure des courbes PIMD et MLMD, les intégrales PIMD donnent des valeurs de la coordinence plus faibles qu'en MLMD. Pour chaque PIMD, on calcule l'évolution de la coordinence en fonction du temps $CN(t)$, pour chaque pas de temps, en prenant la valeur moyenne des coordinences de chaque bead et en utilisant la formule III.3. On en déduit les évolutions des coordinences moyennes en fonction du temps $\langle CN \rangle (t)$ que l'on trace figure V.7. La convergence de la coordinence moyenne est observée à partir de 250ps, sauf pour la PIMD avec $\tau_0=100$ où la coordinence converge après environ 375ps. Pour l'ensemble des PIMD, une coordinence moyenne finale, $CN^{(2)}$, de ~ 6.3 est obtenue, avec une valeur minimale de 6.28 pour la PIMD avec $\tau_0=100$ et une valeur maximale de 6.34 pour $\tau_0=1000$. L'ensemble des valeurs se trouve table V.4. L'écart des valeurs ne semble pas dépendre du paramètre τ_0 et pourrait être dû à un temps de simulation trop court pour un échantillonnage correct.

Pour les trajectoires MLMDs à 400K, des coordinences moyennes de ~ 6.48 ont été obtenues, correspondant à des valeurs supérieures de 0.1 à 0.2. Les effets quantiques nucléaires pourraient alors avoir un effet sur la coordinence du calcium en la diminuant.

Concernant les valeurs de $1000\ln(\mathcal{B})$, on trace leurs évolutions temporelles figure V.8. La convergence des valeurs de $1000\ln(\mathcal{B})$ est observée après environ 300ps. Afin d'étudier la variation des valeurs en fonction du paramètre τ_0 , on trace aussi figure V.8 la valeur de $1000\ln(\mathcal{B})$ obtenue en fin de simulation. Les PIMDs, avec comme valeurs de τ_0 100, 200 et 2000, donnent des $1000\ln(\mathcal{B})$ proches, de 6.63, 6.64 et 6.60, respectivement (voir tableau V.3). Pour la PIMD avec un valeur de τ_0 égale à 10, un obtient un résultat supérieur, de 6.67. Concernant les PIMD effectuées pour des valeurs de τ_0 de 50 et 1000, on obtient des valeurs plus faibles de 6.54. Encore une

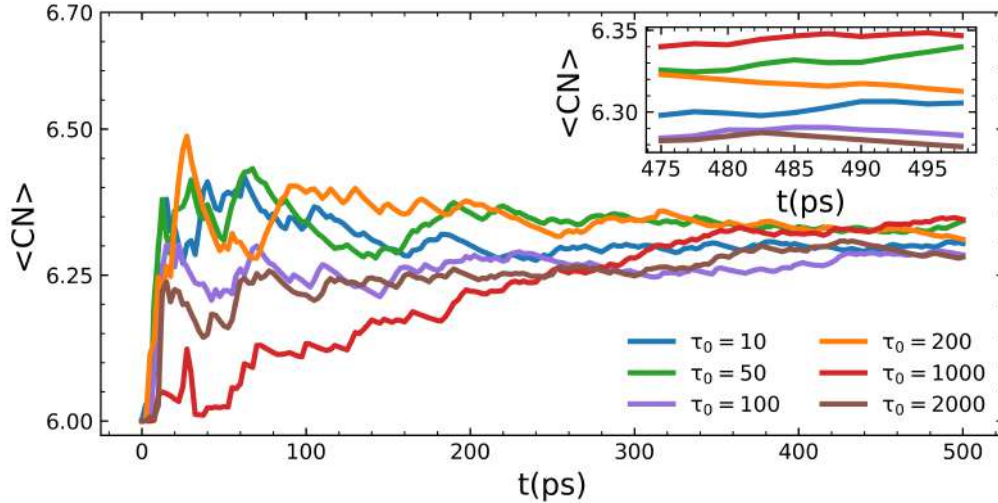


Figure V.7: Evolution de la coordinence moyenne en fonction du temps les PIMD à 400K pour différentes valeurs de τ_0 .

τ_0	$1000 \ln \mathcal{B}$
10	6.67
50	6.54
100	6.63
200	6.64
1000	6.54
2000	6.60

Table V.3: Valeur des $1000 \ln \mathcal{B}$ pour les PIMD à 400K pour différentes valeurs de τ_0 .

fois, il est difficile de conclure quant à l'influence du paramètre τ_0 sur nos résultats. Ces différences de résultats indiquent, compte tenu du temps de simulation, que l'erreur faite sur la valeur de $1000 \ln(\mathcal{B})$ est de l'ordre de 0.1. On remarque de plus que les plus faibles valeurs de $1000 \ln(\mathcal{B})$ correspondent aussi aux deux simulations où la valeur de la coordinence finale est la plus élevée. Ceci est en accord avec le fait qu'une coordinence élevée favorise les isotopes lourds. Des simulations plus longues pourraient améliorer la convergence des valeurs de la coordinence et donc celles des facteurs beta.

Dans la suite, on utilisera une valeur de $\tau_0=10$.

Influence de la masse du calcium

Comme expliqué dans la partie II.3.2, la détermination de la variation de l'énergie libre en fonction de la masse se fait par intégration thermodynamique. Cela nécessite d'effectuer plusieurs PIMDs pour des valeurs de la masse du calcium différentes. Néanmoins, pour un système proche d'un système harmonique, la paramétrisation du paramètre de masse λ suivante :

$$\frac{1}{\sqrt{M(\lambda)}} = \frac{1-\lambda}{\sqrt{M^{40Ca}}} + \frac{\lambda}{\sqrt{M^{44Ca}}} \quad (\text{V.1})$$

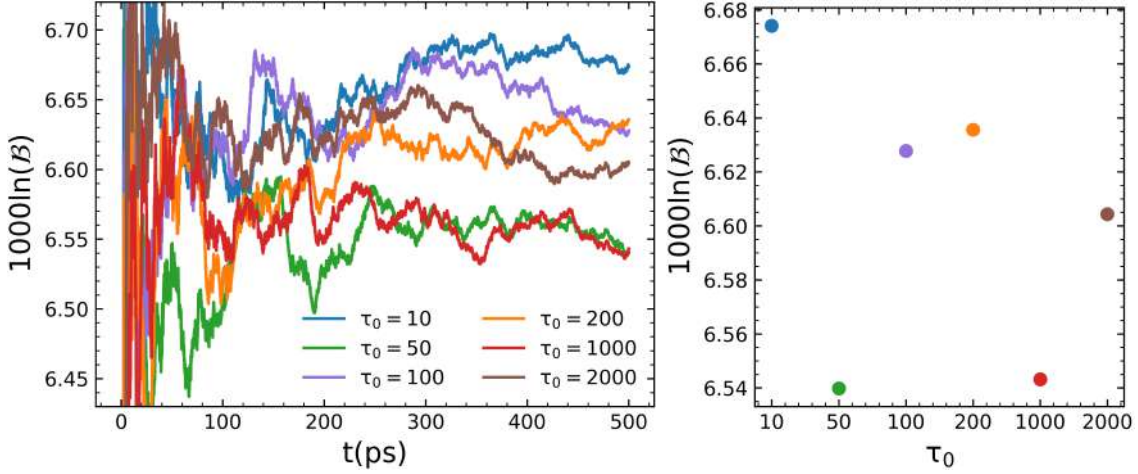


Figure V.8: Evolution temporelle de $1000 \ln(\mathcal{B})$ des PIMD pour différentes valeurs de τ_0 (à gauche). La figure de droite présente la valeur de $1000 \ln(\mathcal{B})$ obtenue en fin de simulation en fonction de τ_0 .

devrait donner une variation quasi-linéaire du facteur beta en fonction de λ . Si cela est le cas, il n'est plus nécessaire de faire l'intégration pour un ensemble de valeur de λ et on peut déterminer le facteur beta à partir d'une seule simulation avec $\lambda=1/2$. On souhaite vérifier si cela est le cas pour notre système et on effectue des PIMD de 500ps, à 400K, avec 8 beads et une valeur de τ_0 égale à 10 pour des masses du calcium de 39.96259u, 40.90819u, 41.88785u, 42.9020u et 43.95548u, correspondant à des valeurs de λ de 0, 0.25, 0.5, 0.75 et 1. Comme pour l'étude de l'influence du paramètre τ_0 , on calcule la coordinence en fonction du temps $CN(t)$ et on trace figure V.9 l'évolution temporelle de la coordinence moyenne en fonction du temps. On obtient des valeurs finales de la coordinence $CN^{(2)}$ inférieures à celles des MLMD. Ces valeurs sont reportées tableau V.4 et varient de 6.30 pour $\lambda=0.5$ à 6.38 pour $\lambda=0.0$ et $\lambda=0.75$.

Les évolutions des $1000 \ln(\mathcal{B})$ en fonction du temps, pour les différents PIMD, se trouvent figure V.10. Pour les PIMDs ayant des valeurs de λ de 0.00, 0.25 et 0.5, les valeurs de $1000 \ln(\mathcal{B})$ convergent après environ 300ps. Pour celles ayant pour valeur de λ 0.75 et 1.00, les courbes décroissent entre ~ 200 ps et ~ 350 -400ps puis convergent, indiquant que de plus longues simulations seraient nécessaires pour améliorer les résultats finaux. On trace figure V.10 la variation des $1000 \ln(\mathcal{B})$ obtenues en fin de simulation en fonction de λ . Les barres d'erreur correspondent à une valeur de 0.1, qui est l'ordre de grandeur de l'erreur sur $1000 \ln(\mathcal{B})$ trouvé lors de l'étude de l'influence du paramètre τ_0 . On effectue un ajustement par une droite affine des valeurs. La valeur, correspondant à $\lambda=0.5$, est celle qui dévie le plus de cet ajustement et peut être expliquée, encore une fois, par le fait que c'est pour cette PIMD que la valeur de la coordinence moyenne la plus faible a été trouvée. En effet, un temps de simulation plus long aurait permis un meilleur échantillonnage des configurations de coordinence différentes. Il aurait alors été possible que la valeur de la coordinence augmente et se rapproche de celle trouvée pour les autres valeurs de λ . Une coordinence plus grande donnerait une valeur plus faible de $1000 \ln(\mathcal{B})$ et un meilleur accord avec l'ajustement serait alors trouvé. Finalement, ces études montrent que, comme pour

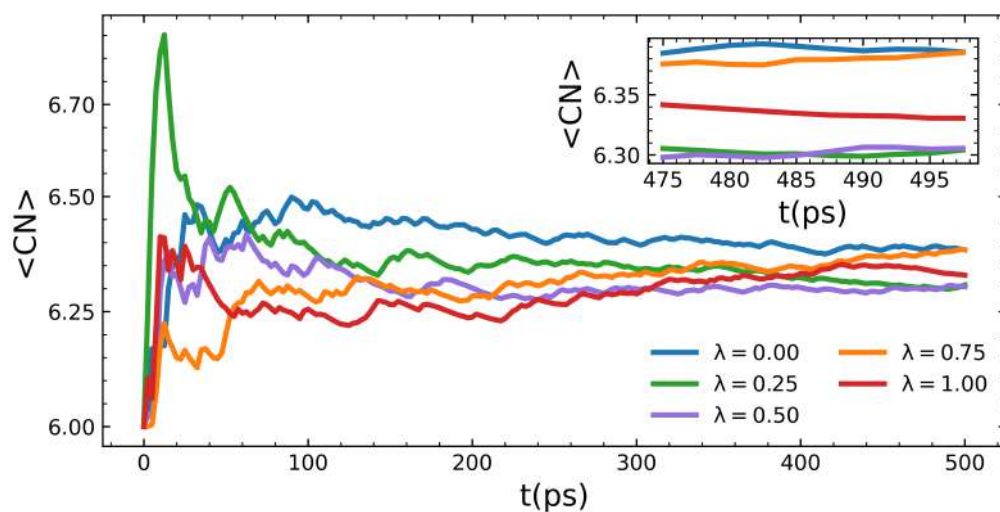


Figure V.9: Evolution de la coordinence moyenne en fonction du temps les PIMD à 400K pour différentes valeurs de τ_0 .

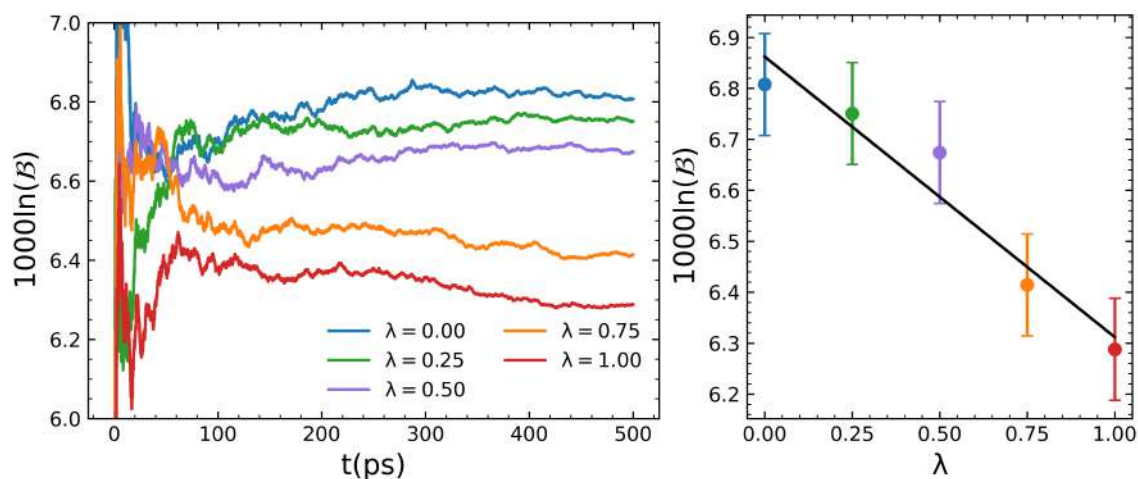


Figure V.10: Evolution temporelle de $1000 \ln(\mathcal{B})$ des PIMD pour différentes valeurs de λ (à gauche). La figure de droite présente la de $1000 \ln(\mathcal{B})$ obtenue en fin de simulation en fonction de λ . Les barres d'erreur sont prises égale à 0.1.

MLMD	CN ⁽²⁾	PIMD	CN ⁽²⁾	PIMD	CN ⁽²⁾
NH n°0	6.47	P = 8, $\lambda=0.5$		P=8, $\tau_0=10$	
NH n°1	6.48	$\tau=10$	6.30	$\lambda=0.0$	6.38
NH n°2	6.49	$\tau=50$	6.34	$\lambda=0.25$	6.31
		$\tau=100$	6.28	$\lambda=0.50$	6.30
		$\tau=200$	6.31	$\lambda=0.75$	6.38
		$\tau=1000$	6.34	$\lambda=1.00$	6.33
		$\tau=2000$	6.28		

Table V.4: Valeurs des coordinences moyennes $CN^{(2)}$ extraites des MLMD utilisant un thermostat de Nose-Hoover (NH) et des PIMD pour différentes valeurs de τ_0 et du paramètre de masse λ déterminé à partir de l'équation II.44. Pour toutes les simulations, la température a été fixée à 400K.

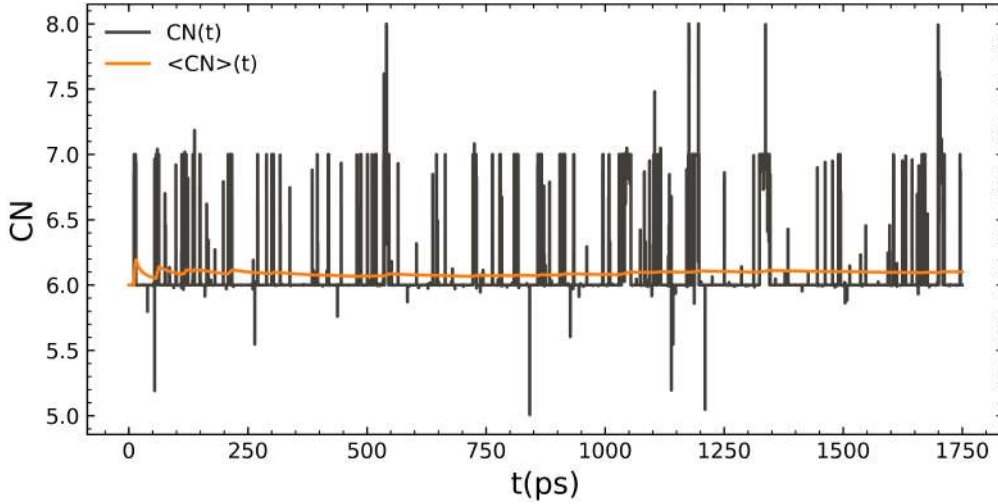


Figure V.11: Evolution temporelle de la coordinnence en fonction du temps $CN(t)$ et de la coordinnence moyenne $\langle CN \rangle(t)$ pour la PIMD à 300K.

la coordinnence, il est important que les simulations échantillonnent correctement l'espace des configurations afin d'obtenir une valeur du facteur beta correcte.

Compte tenu de ces résultats, et de ce constat, on considère qu'il est possible de déterminer le facteur beta à partir d'une seule simulation de PIMD avec $\lambda=1/2$.

Calcul du facteur beta à 300K

Les paramètres de simulation ayant été testés, on présente les résultats obtenus pour une simulation PIMD à 300K, de 1.750ns, avec une valeur de τ_0 égale à 10 et une valeur de λ de 0.5.

Les fonctions de distribution radiale se trouvent figure V.15. Les mêmes différences que précédemment sont observées pour les $g_{cah}(r)$, $g_{oo}(r)$, $g_{oh}(r)$ et $g_{hh}(r)$. La $g_{cao}(r)$, semble indiquer une coordinnence moyenne légèrement plus faible qu'en MLMD. Pour l'observer plus en détail, on calcule l'évolution de la coordinnence en fonction du temps $CN(t)$ ainsi que l'évolution de la coordinnence moyenne $\langle CN \rangle(t)$ que l'on trace figure V.11. On observe qu'au cours de la trajectoire, des configurations de

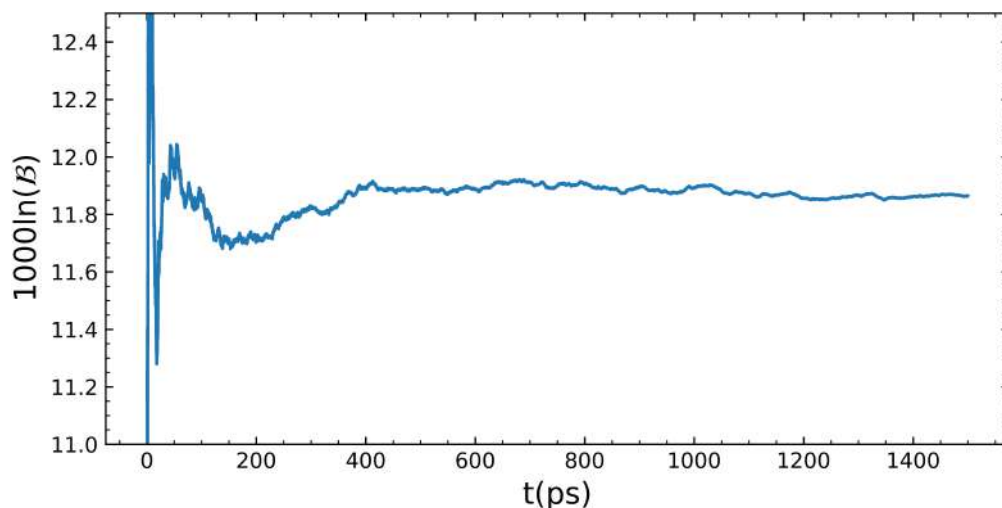


Figure V.12: Evolution de $1000\ln(\mathcal{B})$ de la PIMD à 300K

coordinnence 7 et 8 sont explorées mais sur des intervalles de temps très courts. La valeur finale de la coordinnence moyenne est de 6.10, plus faible de 0.2-0.4 que les coordinnences des MLMD. L'écart entre les coordinnences obtenues en PIMD et MLMD est moins important que celui observé à 400K.

La figure V.12 montre l'évolution de $1000\ln(\mathcal{B})$. La convergence de $1000\ln(\mathcal{B})$ est obtenue après environ 500ps. La valeur de $1000\ln(\mathcal{B})$ en fin de simulation est de 11.86. La valeur trouvée est très légèrement inférieure à celle obtenue par les calculs utilisant l'approximation harmonique (qui est de 12.0 pour les configurations de la trajectoire à 300K).

Les PIMD effectuées avec le potentiel ML ont permis de mettre en évidence l'influence de la prise en compte des effets quantiques nucléaires sur la coordinnence du calcium. Des coordinnences plus faibles sont trouvées à partir de PIMD par rapport aux MLMD, avec un écart plus important à 400K comparé à 300K. Les valeurs du facteur beta trouvées sont proches de celles obtenues en utilisant l'approximation harmonique indiquant que l'anharmonicité du système est très faible. L'utilisation du potentiel ML a permis d'effectuer des PIMD, trop coûteuses en ab-initio, et d'obtenir des résultats du facteur beta meilleurs qu'avec l'utilisation de potentiels empiriques (voir figure 28 en annexe D).

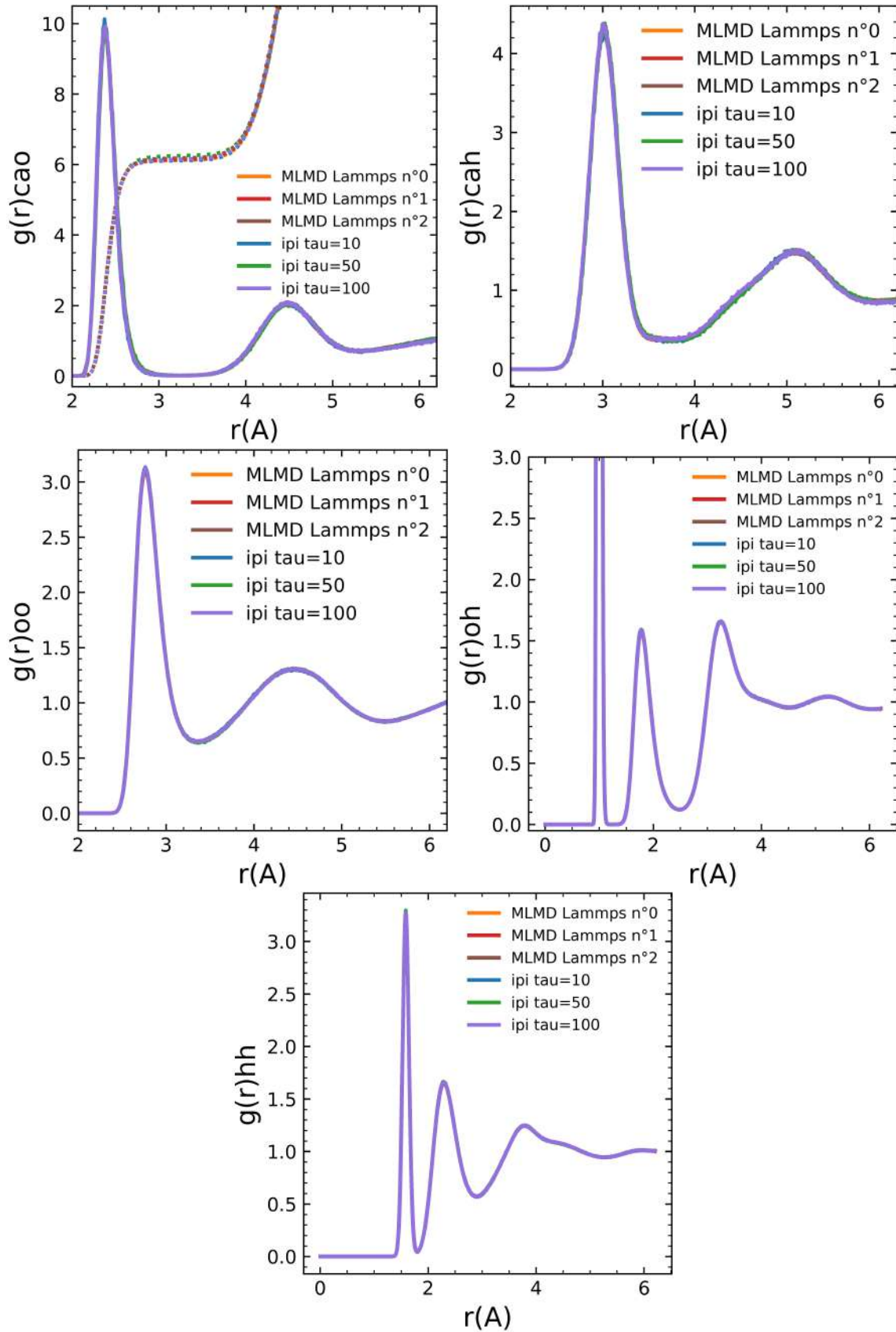


Figure V.13: Fonctions de distribution radiale des MLMD à 300K obtenues avec un thermostat de Nose-Hoover (MLMD Lammmps) et avec un thermostat de Langevin (ipi) pour différentes valeurs de τ_0 .

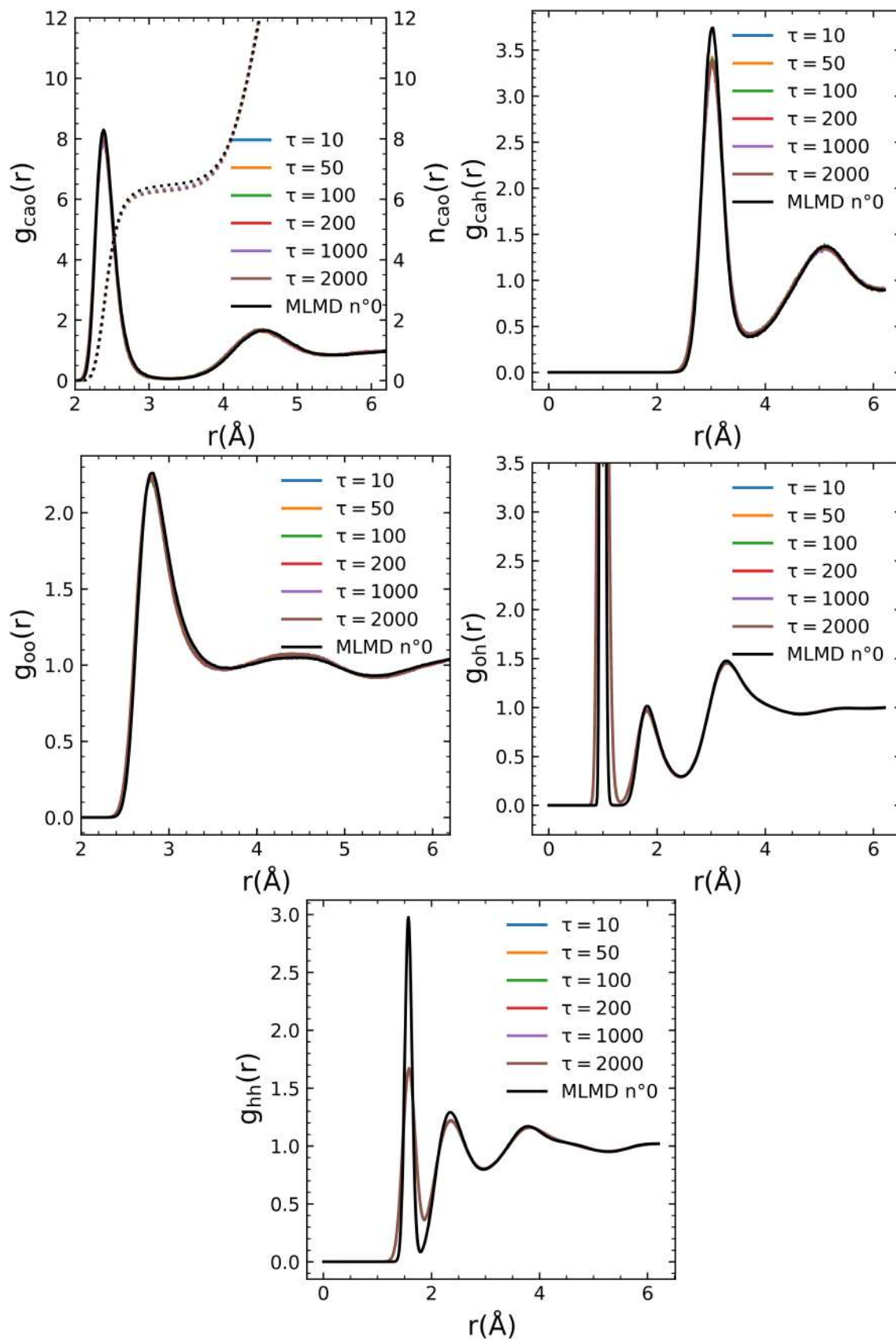


Figure V.14: Fonctions de distribution radiale des PIMD à 400K, avec 8 beads, pour différentes valeur de τ_0 et comparées à celles obtenue en MLMD.

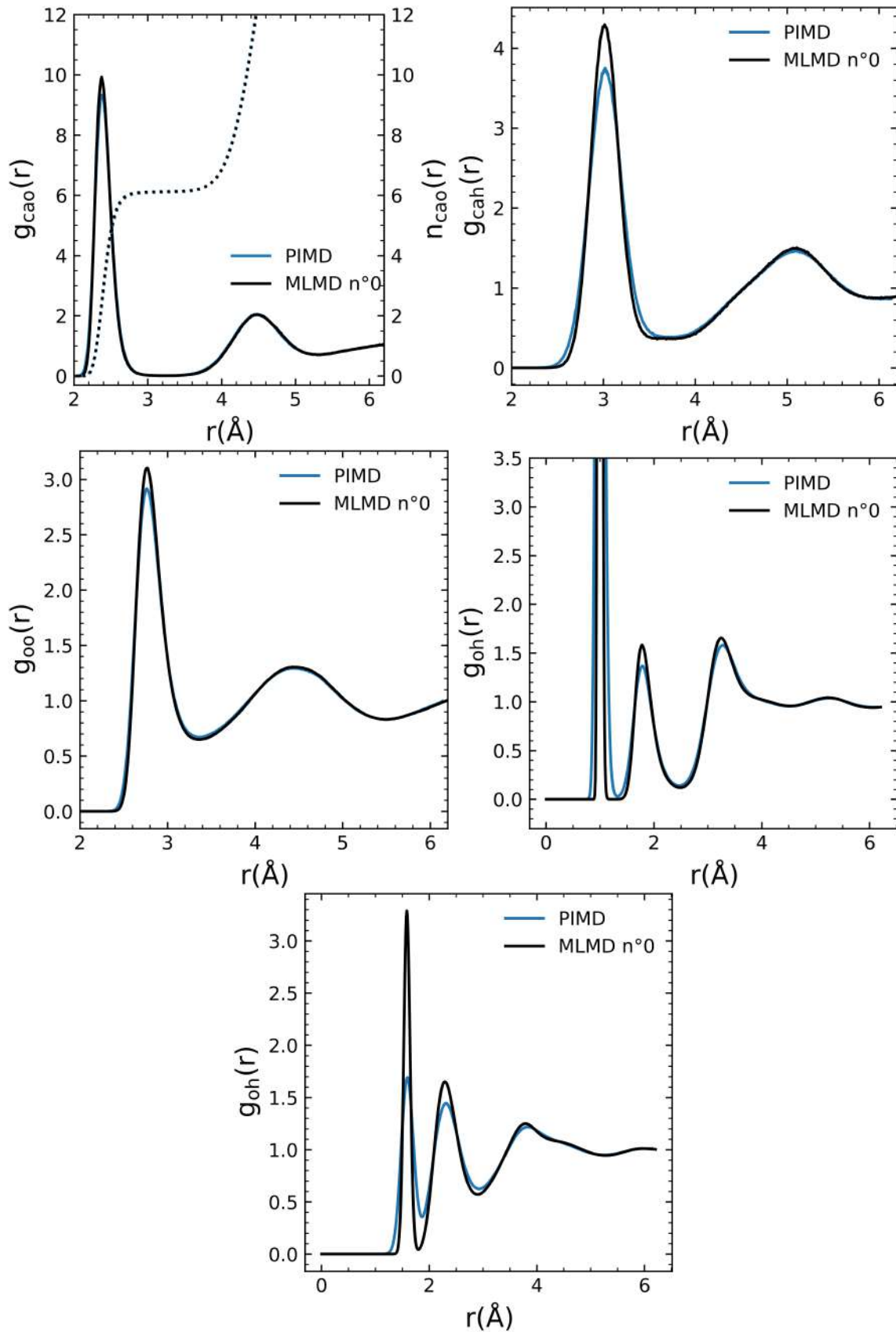


Figure V.15: Fonctions de distribution radiale de la PIMD à 300K, avec 8 beads, comparées à celles obtenue en MLMD.

Conclusion générale

L'étude du fractionnement isotopique du calcium à l'équilibre permet de mieux comprendre son cycle bio-géologique dans la zone critique ainsi que les processus s'y produisant. Lors de ce cycle, le calcium est présent sous forme aqueuse. La détermination de son fractionnement comporte alors deux défis. Le premier est que le fractionnement isotopique dépend de la structure autour du calcium, et donc de sa coordinence. Les précédentes études expérimentales et numériques n'ont toujours pas permis de trouver un consensus sur sa valeur. Le second est que la détermination du fractionnement isotopique se fait usuellement, pour les solides, en employant l'approximation harmonique. Cette approximation n'est, a priori, pas justifiée pour les liquides.

Dans ce travail de thèse, nous avons, dans un premier temps, étudié la coordinence du calcium en solution aqueuse à partir de simulations de dynamiques moléculaire ab-initio. Dans cette étude, différentes fonctionnelles GGA ont été étudiées. En se basant sur l'analyse de la structure de l'eau nous avons conclu que la fonctionnelle BLYP, avec l'ajout de termes de dispersion de Grimme, était, parmi les fonctionnelles étudiées, celle qui était la plus susceptible de reproduire au mieux l'environnement du calcium. Cette étude nous a aussi permis de montrer que les temps de simulation des dynamiques moléculaires usuellement utilisés ne permettaient pas d'effectuer un échantillonnage correct des configurations, en particulier à température ambiante. Nous avons mis en place une procédure, basée sur les chaînes de Markov, dans le but de déterminer un critère de convergence. Nous espérons que cela permettra de comprendre les différentes valeurs de la coordinence de la littérature. Enfin, lors de cette étude, des trajectoires effectuées à différentes températures ont montré une valeur de la coordinence croissante avec la température.

Des méthodes alternatives, comme les dynamiques contraintes ("Umbrella sampling") ou des simulations "parallel tempering" pourraient permettre de résoudre les problèmes d'échantillonnage des dynamiques moléculaires ab-initio. Une autre méthode est d'effectuer de longues simulations, avec un potentiel de type machine learning.

Dans le but d'effectuer des simulations de dynamique moléculaire d'intégrales de chemin, et de compléter l'étude précédente, une deuxième partie de ce travail de thèse a été consacrée à la génération d'un potentiel de type machine learning pour le calcium en solution aqueuse. Une première méthode, basée sur l'apprentissage actif et un échantillonnage efficace de l'espace des phases de l'ensemble canonique, nous a permis d'obtenir un premier potentiel pour un temps de calcul réduit. Les résultats obtenus avec ce potentiel ne nous ont pas permis de reproduire fidèlement la structure de l'eau ainsi que l'environnement du calcium des simulations ab-initio dont dépend le fractionnement isotopique. Grâce à la grande base de données que nous avons à notre disposition, suite à l'étude de la coordinence du calcium, nous

avons généré un potentiel de type machine learning à réseau de neurones donnant des résultats structurels très proches des simulations ab-initio avec la fonctionnelle BLYP-D2. De longues simulations, à différentes températures, ont permis d'obtenir des valeurs plus précises de la coordinence du calcium montrant, comme en ab-initio, une dépendance croissante avec la température.

Finalement, une dernière partie de ce travail a été consacrée à l'étude du fractionnement du calcium en solution aqueuse. Tout d'abord, à partir d'une méthode utilisée pour les solides utilisant l'approximation harmonique. La détermination du facteur \mathcal{B} a été faite à la fois à partir de calculs ab-initio et à partir du potentiel machine learning. Les résultats montrent qu'avec le potentiel machine learning, les valeurs obtenues ne reproduisent pas exactement celles des calculs ab-initio. Cela pourrait venir de la méthode de calcul, et notamment de l'optimisation structurelle, ou bien du faible nombre de configurations utilisées pour les calculs ab-initio. Une autre explication serait que le potentiel machine learning ne serait pas en mesure de décrire correctement les états vibrationnels du système et qu'il serait nécessaire de l'améliorer. Soit, par exemple, en poursuivant l'apprentissage, ou en utilisant des méthodes d'apprentissage actif pour compléter notre base de données.

Des dynamiques moléculaires d'intégrales de chemin ont été effectuées à l'aide de ce potentiel machine learning. Dans un premier temps, des simulations à 400K, nous ont permis de choisir les paramètres de simulation. Ces simulations ont aussi permis de montrer que la convergence de la valeur du facteur beta était liée à la convergence de la coordinence et que des simulations suffisamment longues étaient nécessaires pour un échantillonnage correct. Enfin, une simulation à 300K nous a permis de déterminer une valeur du facteur \mathcal{B} légèrement plus faible que celle obtenue à partir de l'approximation harmonique. Cela suggère que pour notre système les effets anharmoniques sont faibles.

Les valeurs de la coordinence extraites de ces simulations sont plus faibles que celles des dynamiques moléculaires ab-initio, indiquant que les effets quantiques nucléaires pourraient avoir un effet sur cette propriété.

Finalement, nous avons montré que l'utilisation d'un potentiel machine learning pour la détermination du fractionnement isotopique d'une espèce en solution à partir de simulations de dynamique moléculaire d'intégrales de chemin permettait de réduire considérablement les temps de calculs. Cette procédure peut être appliquée à différents systèmes avec des conditions thermodynamiques différentes afin d'étudier d'autres phénomènes géologiques.

Annexes

A Outils d'analyse des trajectoires de dynamique moléculaire

On présente dans cette partie les outils utilisés pour analyser les trajectoires de dynamique moléculaire et de dynamique moléculaire d'intégrales de chemin.

Fonction de distribution radiale

La fonction de distribution radiale, $g_{XY}(r)$, entre deux atomes X et Y , est une fonction de corrélation permettant de décrire la distribution des atomes Y par rapport aux atomes X . Elle donne la probabilité de trouver un atome Y à une distance r de l'atome X et est donnée par :

$$g_{XY}(r) = \frac{dn_{XY}}{4\pi r^2 dr \rho_X} \quad (2)$$

où dn_{XY} est le nombre d'atomes Y à une distance comprise entre r et $r + dr$ et ρ_X est la densité d'atome X .

En pratique, on calcule les distances entre les atomes X et Y de toutes les configurations extraites de la trajectoire de la dynamique moléculaire et on construit un histogramme à partir de la formule :

$$g_{XY}(r) = \frac{1}{(N(N-A))N_{steps}} \sum_{s=0}^{N_{steps}-1} \sum_{i=1}^{N_X} \sum_{j=1}^{N_Y} \delta(r - |\vec{r}_i^X - \vec{r}_j^Y|) \quad (3)$$

$$\text{avec : } A = 0 \text{ si } i = j \text{ et } A = 1 \text{ sinon} \quad (4)$$

où N_{steps} est le nombre de pas de la trajectoire, \vec{r}_i^X et \vec{r}_j^Y les positions des atomes X et Y et N_X et N_Y leurs nombres. Pour les dynamiques moléculaires d'intégrales de chemin, le calcul se fait en considérant toutes les configurations de chaque bead.

Dans cette thèse, les fonctions de distribution radiale ont été calculées avec le code VMD [62] pour un $dr = 0.01\text{\AA}$. La détermination des positions des extrema a été faite à partir de la fonction python `scipy.signal.argrelmax` et `scipy.signal.argrelmin` de la librairie `scipy`.

Le nombre moyen d'atomes, N_a , situés à une distance comprise entre $r = 0$ et $r = r_a$ peut être déterminé à partir de l'intégrale de la fonction de distribution radiale :

$$N_a = \int_0^{r_a} g(r) dr \quad (5)$$

C'est de cette façon qu'ont été déterminées les coordonnées que l'on a notée CN⁽¹⁾.

Fonction de déplacement quadratique moyen

La fonction de déplacement quadratique moyen, MSD_X , d'un atome X , permet de quantifier l'espace exploré par les atomes. Son expression est donnée par :

$$MSD_X(t; \tau) = \frac{1}{N_X} \sum_{i=1}^{N_X} \left(\vec{r}_i^X(t + \tau) - \vec{r}_i^X(\tau) \right)^2 \quad (6)$$

où $\vec{r}_i^X(t + \tau)$ est la position de l'atome X au temps $t + \tau$ et $\vec{r}_i^X(\tau)$ la position de référence.

La fonction de déplacement quadratique moyen peut être calculée pour différentes valeurs de τ telles que $0 \leq \tau \leq t_{simu}$, où t_{simu} est le temps de simulation. Il est ensuite possible de les moyenner, tel que :

$$MSD_X(t) = \frac{1}{A(t)} \sum_{\tau=0}^{t_{simu}} MSD_X(t; \tau) \quad (7)$$

où $A(t)$ est un coefficient de normalisation dépendant du temps.

Fonction d'autocorrélation des vitesses

La fonction d'autocorrélation des vitesses, VAF_X , d'un atome X , donne des informations sur les propriétés dynamiques du système. Elle a pour expression :

$$VAF_X(t; \tau) = \frac{1}{N_X} \sum_{i=1}^{N_X} \vec{v}_i^X(\tau) \cdot \vec{v}_i^X(t + \tau) \quad (8)$$

où τ est l'origine du temps choisi et \vec{v}_i^X les vitesses des atomes X . Dans cette thèse, les VAF_X ont été calculées sur plusieurs intervalles de temps de 1ps et ont ensuite été moyennées.

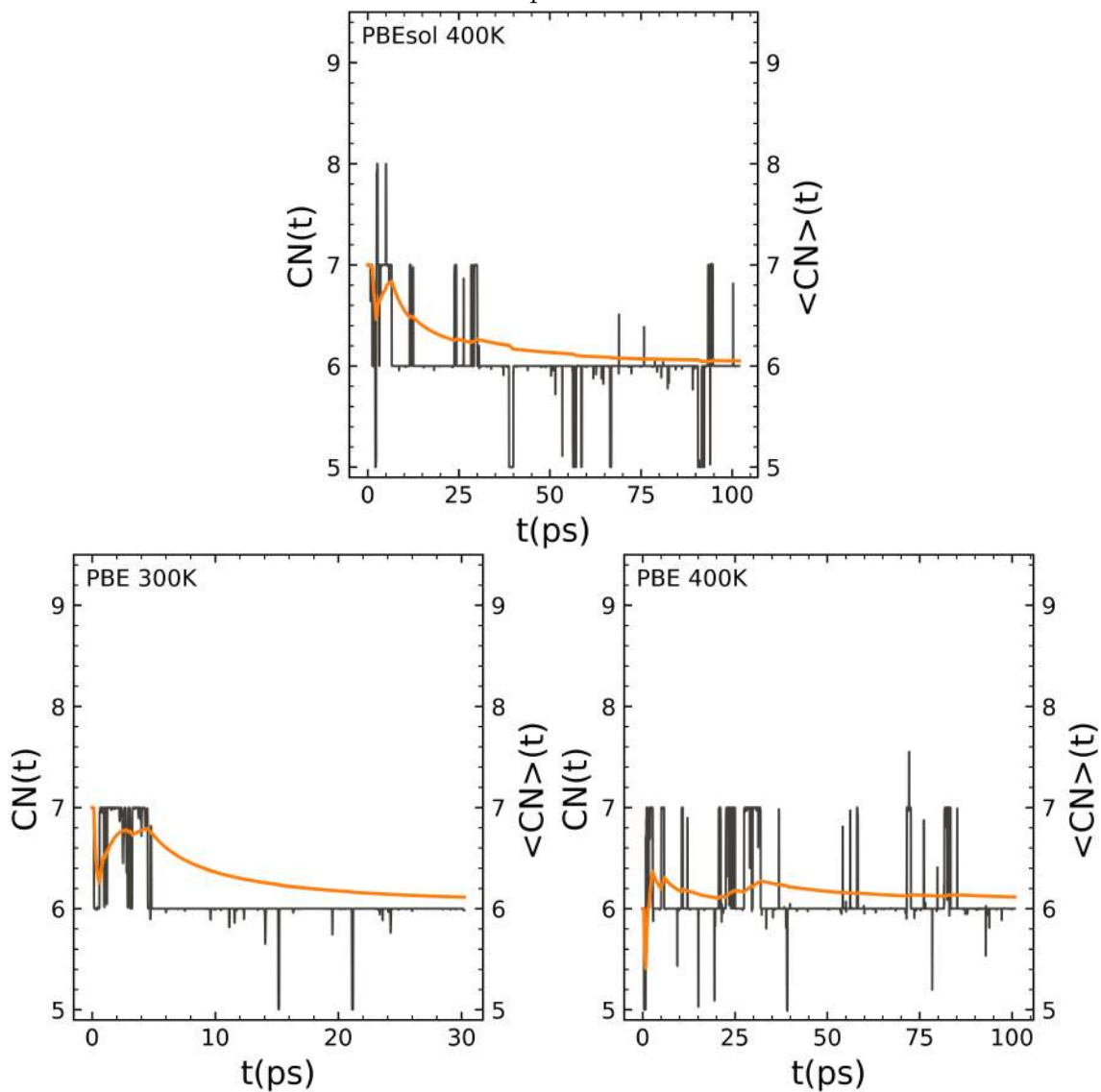
Le mouvement des atomes étant une combinaison des modes normaux, il est possible d'accéder aux fréquences de ces modes par transformée de Fourier.

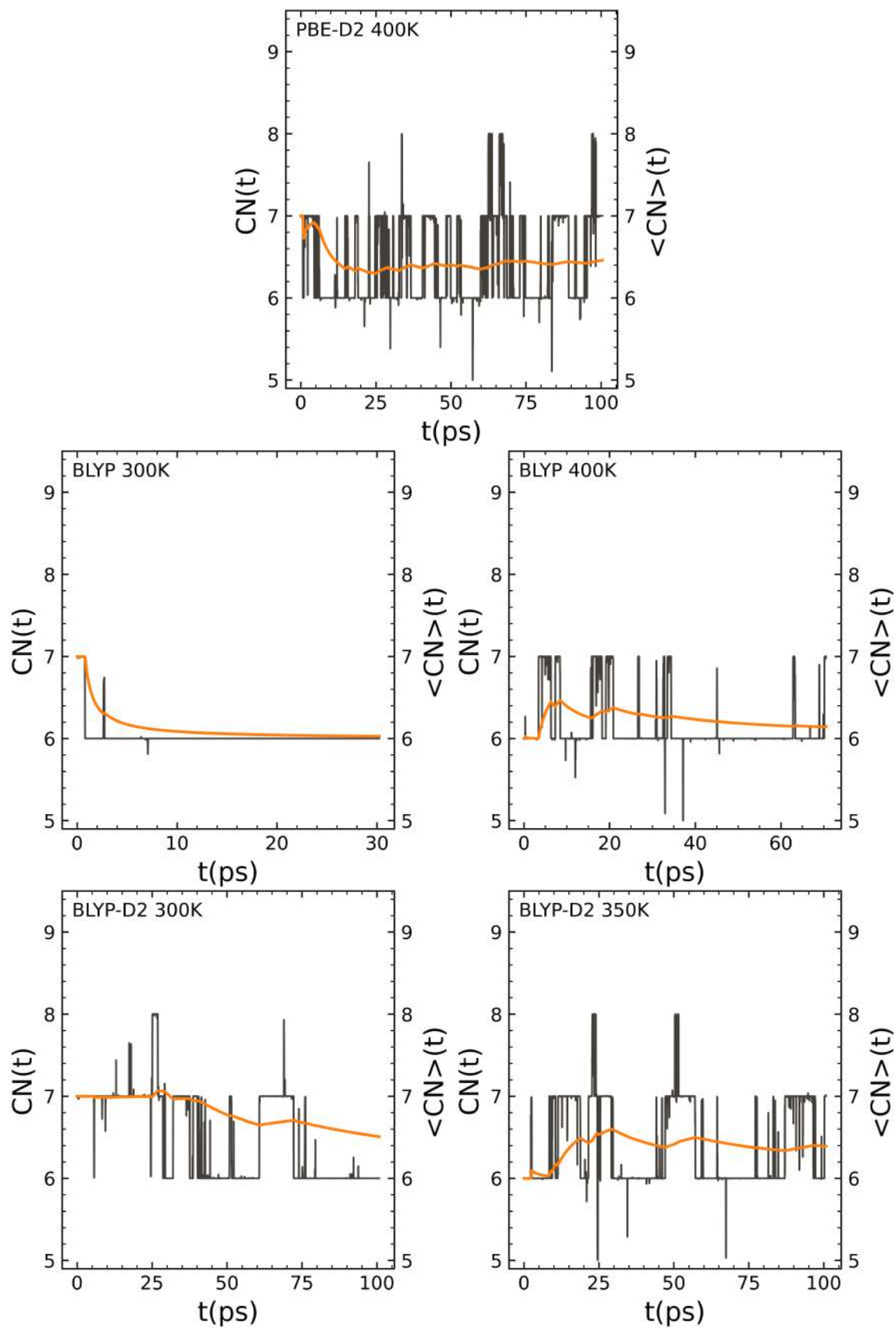
B Compléments du chapitre 3

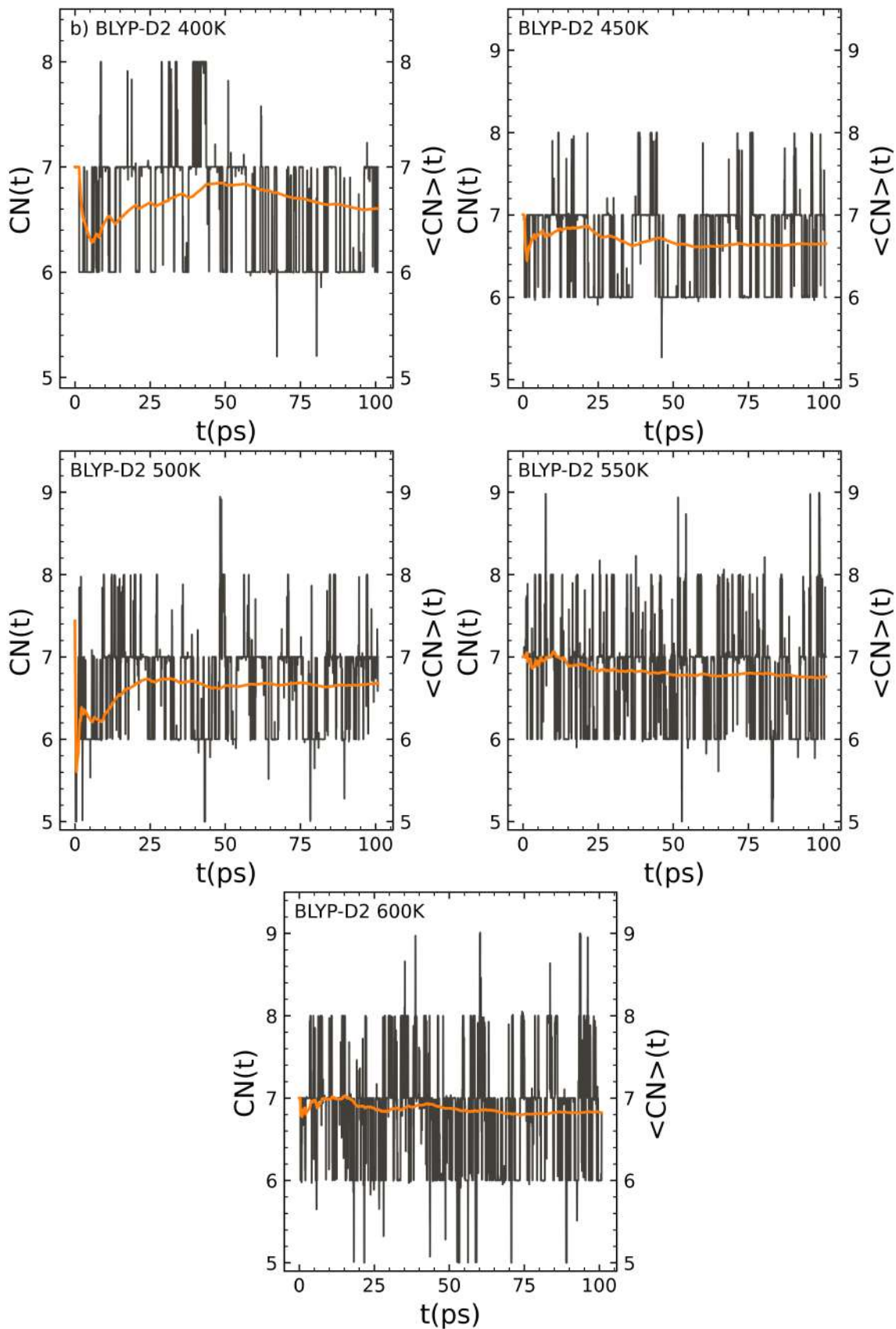
B.1 Coordinnence en fonction du temps

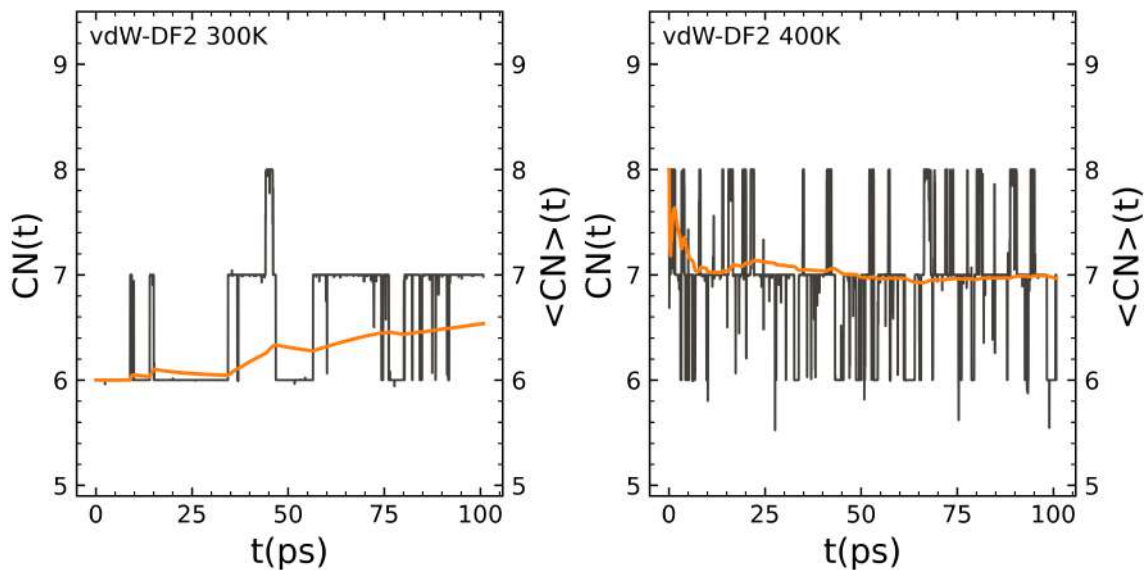
On présente figure 16 l'ensemble des coordinnences en fonction du temps calculées pour les différentes fonctionnelles à différentes températures.

Figure 16: Evolution de la coordinnence en fonction du temps pour les différentes fonctionnelles étudiées à différentes températures.









B.2 Fonctions de distribution radiale des simulations avec un contre-ion et pour des concentrations différentes

On présente dans cette partie les fonctions de distribution radiale Ca-O et O-O obtenues avec la présence d'un contre-ion Cl^- pour des dynamiques moléculaires à 400K avec la fonctionnelle BLYP-D2. Un premier système est composé d'un ion Ca^{2+} d'un ion Cl^- et de 62 molécules d'eau dans une boîte cubique de côté 12.41Å. Un second système est composé d'un ion Ca^{2+} d'un ion Cl^- et de 124 molécules d'eau, permettant d'étudier l'influence de la concentration, dans une boîte cubique de côté 15.49Å donnant une même densité de l'eau que pour le premier système. Tous les autres paramètres sont les mêmes que ceux détaillés partie III.2 avec un temps de simulation de 120ps pour le premier système et de 30ps pour le second. La figure 17 montre les fonctions de distribution radiale pour le premier système et comparé au cas où aucun contre-ion n'est présent. Quant à la figure 18, elle montre la différence entre les fonctions de distribution radiale entre le premier et le second système avec la présence d'un contre-ion. Pour les deux systèmes, la structure de l'eau est similaire. En revanche, des différences sont observées lorsque l'on regarde la g_{cao} . En effet, on trouve une coordinnence moyenne, déterminée à partir de l'intégrale de la g_{cao} , plus élevée que pour le système sans présence d'un contre-ion. Pour le premier système, composé d'un ion Ca^{2+} d'un ion Cl^- et de 62 molécules d'eau, la coordinnence $\text{CN}^{(1)}$ obtenue est de 6.70, alors qu'elle est de 6.61 sans ion Cl^- . La concentration semble aussi avoir une influence, avec une coordinnence moyenne $\text{CN}^{(1)}$ de 6.97 obtenue pour le système avec un ion Ca^{2+} d'un ion Cl^- et 124 molécules d'eau.

Néanmoins, il est difficile de conclure sur l'influence réelle de la présence d'un ion Cl^- ou de la concentration au vu des problèmes d'échantillonnage des configurations de coordinnence différentes discuté au chapitre III et, pour ces systèmes, l'étude n'a pas été poursuivie.

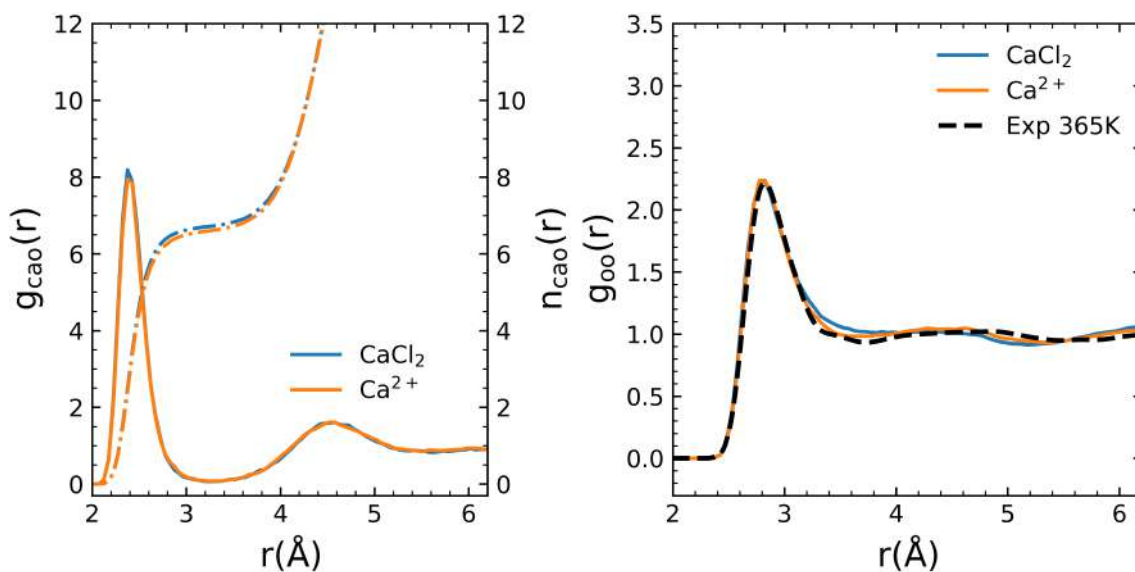


Figure 17: Fonction de distribution radiale Ca-O (à gauche) et O-O (à droite) pour la fonctionnelle BLYP-D2 à 400K avec et sans contre-ion Cl^- .

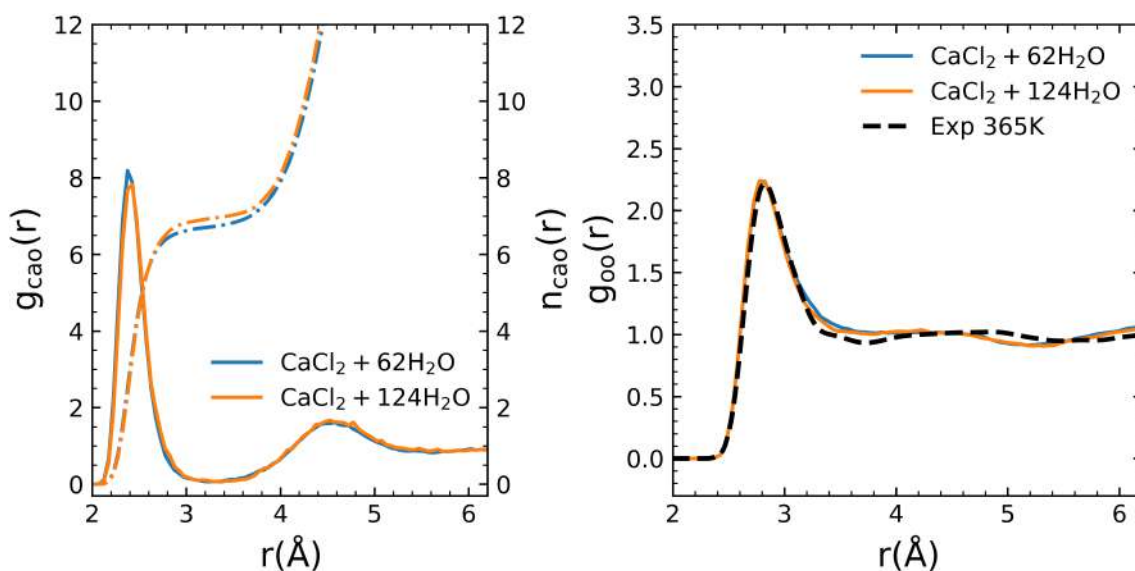


Figure 18: Fonction de distribution radiale Ca-O (à gauche) et O-O (à droite) pour la fonctionnelle BLYP-D2 à 400K avec un contre-ion Cl^- à deux différentes concentration.

C Compléments du chapitre 4

C.1 Equivalence entre les paramètres du réseau de neurones testés et les commandes N2P2

Paramètres	Commandes N2P2
Filtre de kalman	
ratio energies:forces	<i>short_energy_fraction</i> <i>short_force_fraction</i>
β	<i>force_weight</i>
ϵ	<i>kalman_epsilon</i>
q_0	<i>kalman_q0</i>
τ_q	<i>kalman_qtau</i>
q_{min}	<i>kalman_qmin</i>
η_0	<i>kalman_eta</i>
τ_η	<i>kalman_etatau</i>
η_{max}	<i>kalman_etamax</i>
ϵ	<i>kalman_epsilon</i>
Réseau de neurones	
normalisation de la base de données	<i>normalize_data_set</i>
mise à l'échelle des descripteurs	<i>scale_symmetry_functions</i> <i>scale_symmetry_functions_sigma</i> <i>center_symmetry_functions</i> <i>scale_min_short</i> <i>scale_max_short</i>
fonction d'activation	<i>global_activation_short</i>
fonction de cutoff	<i>cutoff_type</i>
normalisation valeurs d'entrées des neurones	<i>normalize_nodes</i>
nombre de couches et de neurones	<i>global_nodes_short</i>

Table 5: Commandes N2P2 associées aux paramètres du filtre de Kalman et du réseau de neurones testés.

C.2 Descripteurs utilisés pour le réseau de neurones

Les descripteurs utilisés sont de type atome-centré. Pour les descripteurs à deux corps, ils sont définis en utilisant la commande :

$$symfunction_short\ I\ 2\ J\ \eta\ R_s\ R_c$$

où I est l'élément central, J l'élément voisin et η , R_s et R_c les grandeurs définies équations I.65. Pour les descripteurs à trois corps, ils sont définis en utilisant la commande :

$$symfunction_short\ I\ 3\ J\ K\ \eta\ \lambda\ \zeta\ R_c\ R_s$$

où I est l'élément central, J et K les éléments voisins et η , λ , ζ , R_c et R_s les grandeurs définies équations ??.

Element I	Element J	Element K	$\eta(\text{\AA}^{-1})$	$R_s(\text{\AA})$	$R_c(\text{\AA})$	λ	ζ
Deux corps							
H	H	-	0.0040	0.00	6.0	-	-
H	H	-	0.0400	0.00	6.0	-	-
H	H	-	0.1070	0.00	6.0	-	-
H	H	-	0.2140	0.00	6.0	-	-
H	H	-	0.5370	1.00	6.0	-	-
H	H	-	1.0700	1.00	6.0	-	-
H	H	-	2.1400	1.00	6.0	-	-
H	H	-	5.3700	1.00	6.0	-	-
H	O	-	0.0040	0.00	6.0	-	-
H	O	-	0.0400	0.00	6.0	-	-
H	O	-	0.1070	0.00	6.0	-	-
H	O	-	0.2140	0.00	6.0	-	-
H	O	-	0.5370	0.48	6.0	-	-
H	O	-	1.0700	0.48	6.0	-	-
H	O	-	2.1400	0.48	6.0	-	-
H	O	-	5.3700	0.48	6.0	-	-
H	Ca	-	0.8889	0.00	6.0	-	-
H	Ca	-	0.8889	0.75	6.0	-	-
H	Ca	-	0.8889	1.50	6.0	-	-
H	Ca	-	0.8889	2.25	6.0	-	-
H	Ca	-	0.8889	3.00	6.0	-	-
H	Ca	-	0.8889	3.75	6.0	-	-
H	Ca	-	0.8889	4.50	6.0	-	-
H	Ca	-	0.8889	5.25	6.0	-	-
H	Ca	-	0.8889	6.00	6.0	-	-
Trois corps							
H	O	H	0.7100	0.0	6.0	1.0	1.0
H	O	H	0.2500	0.0	6.0	1.0	1.0
H	O	H	0.2500	0.0	6.0	-1.0	1.0
H	O	H	0.1070	0.0	6.0	1.0	1.0
H	O	H	0.1070	0.0	6.0	-1.0	1.0
H	O	H	0.0400	0.0	6.0	1.0	4.0
H	O	H	0.0400	0.0	6.0	-1.0	4.0
H	O	O	0.1070	0.0	6.0	1.0	1.0
H	O	O	0.1070	0.0	6.0	-1.0	1.0
H	O	O	0.0400	0.0	6.0	1.0	4.0
H	O	O	0.0400	0.0	6.0	-1.0	4.0
H	Ca	H	0.05556	0.0	6.0	1.0	1.0
H	Ca	H	0.05556	0.0	6.0	-1.0	1.0
H	Ca	H	0.05556	0.0	6.0	1.0	6.0
H	Ca	H	0.05556	0.0	6.0	-1.0	6.0
H	Ca	H	0.05556	3.0	6.0	1.0	1.0
H	Ca	H	0.05556	3.0	6.0	-1.0	1.0
H	Ca	H	0.05556	3.0	6.0	1.0	6.0
H	Ca	H	0.05556	3.0	6.0	-1.0	6.0
H	Ca	H	0.05556	6.0	6.0	1.0	1.0
H	Ca	H	0.05556	6.0	6.0	-1.0	1.0
H	Ca	H	0.05556	6.0	6.0	1.0	6.0
H	Ca	H	0.05556	6.0	6.0	-1.0	6.0
H	Ca	O	0.05556	0.0	6.0	1.0	1.0
H	Ca	O	0.05556	0.0	6.0	-1.0	1.0
H	Ca	O	0.05556	0.0	6.0	1.0	6.0
H	Ca	O	0.05556	0.0	6.0	-1.0	6.0
H	Ca	O	0.05556	3.0	6.0	1.0	1.0
H	Ca	O	0.05556	3.0	6.0	-1.0	1.0
H	Ca	O	0.05556	3.0	6.0	1.0	6.0
H	Ca	O	0.05556	3.0	6.0	-1.0	6.0
H	Ca	O	0.05556	6.0	6.0	1.0	1.0
H	Ca	O	0.05556	6.0	6.0	-1.0	1.0
H	Ca	O	0.05556	6.0	6.0	1.0	6.0
H	Ca	O	0.05556	6.0	6.0	-1.0	6.0

Table 6: Descripteurs atome-centrés de l'hydrogène.

Element I	Element J	Element K	$\eta(\text{\AA}^{-1})$	$R_s(\text{\AA})$	$R_c(\text{\AA})$	λ	ζ
Deux corps							
O	O	-	0.0040	0.00	6.0	-	-
O	O	-	0.0400	0.00	6.0	-	-
O	O	-	0.1070	0.00	6.0	-	-
O	O	-	0.2140	0.00	6.0	-	-
O	O	-	0.5370	1.00	6.0	-	-
O	O	-	1.0700	1.00	6.0	-	-
O	O	-	2.1400	1.00	6.0	-	-
O	O	-	5.3700	1.00	6.0	-	-
O	H	-	0.0040	0.00	6.0	-	-
O	H	-	0.0400	0.00	6.0	-	-
O	H	-	0.1070	0.00	6.0	-	-
O	H	-	0.2140	0.00	6.0	-	-
O	H	-	0.5370	0.48	6.0	-	-
O	H	-	1.0700	0.48	6.0	-	-
O	H	-	2.1400	0.48	6.0	-	-
O	H	-	5.3700	0.48	6.0	-	-
O	Ca	-	0.8889	0.00	6.0	-	-
O	Ca	-	0.8889	0.75	6.0	-	-
O	Ca	-	0.8889	1.50	6.0	-	-
O	Ca	-	0.8889	2.25	6.0	-	-
O	Ca	-	0.8889	3.00	6.0	-	-
O	Ca	-	0.8889	3.75	6.0	-	-
O	Ca	-	0.8889	4.50	6.0	-	-
O	Ca	-	0.8889	5.25	6.0	-	-
O	Ca	-	0.8889	6.00	6.0	-	-
Trois corps							
O	H	H	0.2500	0.0	6.0	1.0	1.0
O	H	H	0.2500	0.0	6.0	-1.0	1.0
O	H	H	0.1070	0.0	6.0	1.0	1.0
O	H	H	0.1070	0.0	6.0	-1.0	1.0
O	H	H	0.0400	0.0	6.0	1.0	4.0
O	H	H	0.0400	0.0	6.0	-1.0	4.0
O	O	H	0.1070	0.0	6.0	1.0	1.0
O	O	H	0.1070	0.0	6.0	-1.0	1.0
O	O	H	0.0400	0.0	6.0	1.0	4.0
O	O	H	0.0400	0.0	6.0	-1.0	4.0
O	Ca	H	0.05556	0.0	6.0	1.0	1.0
O	Ca	H	0.05556	0.0	6.0	-1.0	1.0
O	Ca	H	0.05556	0.0	6.0	1.0	6.0
O	Ca	H	0.05556	0.0	6.0	-1.0	6.0
O	Ca	H	0.05556	3.0	6.0	1.0	1.0
O	Ca	H	0.05556	3.0	6.0	-1.0	1.0
O	Ca	H	0.05556	3.0	6.0	1.0	6.0
O	Ca	H	0.05556	3.0	6.0	-1.0	6.0
O	Ca	H	0.05556	6.0	6.0	1.0	1.0
O	Ca	H	0.05556	6.0	6.0	-1.0	1.0
O	Ca	H	0.05556	6.0	6.0	1.0	6.0
O	Ca	H	0.05556	6.0	6.0	-1.0	6.0
O	Ca	O	0.05556	0.0	6.0	1.0	1.0
O	Ca	O	0.05556	0.0	6.0	-1.0	1.0
O	Ca	O	0.05556	0.0	6.0	1.0	6.0
O	Ca	O	0.05556	0.0	6.0	-1.0	6.0
O	Ca	O	0.05556	3.0	6.0	1.0	1.0
O	Ca	O	0.05556	3.0	6.0	-1.0	1.0
O	Ca	O	0.05556	3.0	6.0	1.0	6.0
O	Ca	O	0.05556	3.0	6.0	-1.0	6.0
O	Ca	O	0.05556	6.0	6.0	1.0	1.0
O	Ca	O	0.05556	6.0	6.0	-1.0	1.0
O	Ca	O	0.05556	6.0	6.0	1.0	6.0
O	Ca	O	0.05556	6.0	6.0	-1.0	6.0

Table 7: Descripteurs atome-centrés de l'oxygène.

Element I	Element J	Element K	$\eta(\text{\AA}^{-1})$	$R_s(\text{\AA})$	$R_c(\text{\AA})$	λ	ζ
Deux corps							
Ca	H	-	0.8889	0.00	6.0	-	-
Ca	H	-	0.8889	0.75	6.0	-	-
Ca	H	-	0.8889	1.50	6.0	-	-
Ca	H	-	0.8889	2.25	6.0	-	-
Ca	H	-	0.8889	3.00	6.0	-	-
Ca	H	-	0.8889	3.75	6.0	-	-
Ca	H	-	0.8889	4.50	6.0	-	-
Ca	H	-	0.8889	5.25	6.0	-	-
Ca	H	-	0.8889	6.00	6.0	-	-
Ca	O	-	0.8889	0.00	6.0	-	-
Ca	O	-	0.8889	0.75	6.0	-	-
Ca	O	-	0.8889	1.50	6.0	-	-
Ca	O	-	0.8889	2.25	6.0	-	-
Ca	O	-	0.8889	3.00	6.0	-	-
Ca	O	-	0.8889	3.75	6.0	-	-
Ca	O	-	0.8889	4.50	6.0	-	-
Ca	O	-	0.8889	5.25	6.0	-	-
Ca	O	-	0.8889	6.00	6.0	-	-
Trois corps							
Ca	H	H	0.05556	0.0	6.0	1.0	1.0
Ca	H	H	0.05556	0.0	6.0	-1.0	1.0
Ca	H	H	0.05556	0.0	6.0	1.0	6.0
Ca	H	H	0.05556	0.0	6.0	-1.0	6.0
Ca	H	H	0.05556	3.0	6.0	1.0	1.0
Ca	H	H	0.05556	3.0	6.0	-1.0	1.0
Ca	H	H	0.05556	3.0	6.0	1.0	6.0
Ca	H	H	0.05556	3.0	6.0	-1.0	6.0
Ca	H	H	0.05556	6.0	6.0	1.0	1.0
Ca	H	H	0.05556	6.0	6.0	-1.0	1.0
Ca	H	H	0.05556	6.0	6.0	1.0	6.0
Ca	H	H	0.05556	6.0	6.0	-1.0	6.0
Ca	O	O	0.05556	0.0	6.0	1.0	1.0
Ca	O	O	0.05556	0.0	6.0	-1.0	1.0
Ca	O	O	0.05556	0.0	6.0	1.0	6.0
Ca	O	O	0.05556	0.0	6.0	-1.0	6.0
Ca	O	O	0.05556	3.0	6.0	1.0	1.0
Ca	O	O	0.05556	3.0	6.0	-1.0	1.0
Ca	O	O	0.05556	3.0	6.0	1.0	6.0
Ca	O	O	0.05556	3.0	6.0	-1.0	6.0
Ca	O	O	0.05556	6.0	6.0	1.0	1.0
Ca	O	O	0.05556	6.0	6.0	-1.0	1.0
Ca	O	O	0.05556	6.0	6.0	1.0	6.0
Ca	O	O	0.05556	6.0	6.0	-1.0	6.0
Ca	O	H	0.05556	0.0	6.0	1.0	1.0
Ca	O	H	0.05556	0.0	6.0	-1.0	1.0
Ca	O	H	0.05556	0.0	6.0	1.0	6.0
Ca	O	H	0.05556	0.0	6.0	-1.0	6.0
Ca	O	H	0.05556	3.0	6.0	1.0	1.0
Ca	O	H	0.05556	3.0	6.0	-1.0	1.0
Ca	O	H	0.05556	3.0	6.0	1.0	6.0
Ca	O	H	0.05556	3.0	6.0	-1.0	6.0
Ca	O	H	0.05556	6.0	6.0	1.0	1.0
Ca	O	H	0.05556	6.0	6.0	-1.0	1.0
Ca	O	H	0.05556	6.0	6.0	1.0	6.0
Ca	O	H	0.05556	6.0	6.0	-1.0	6.0

Table 8: Descripteurs atome-centrés du calcium.

C.3 Résultats complémentaires des dynamiques moléculaires avec le potentiel machine learning

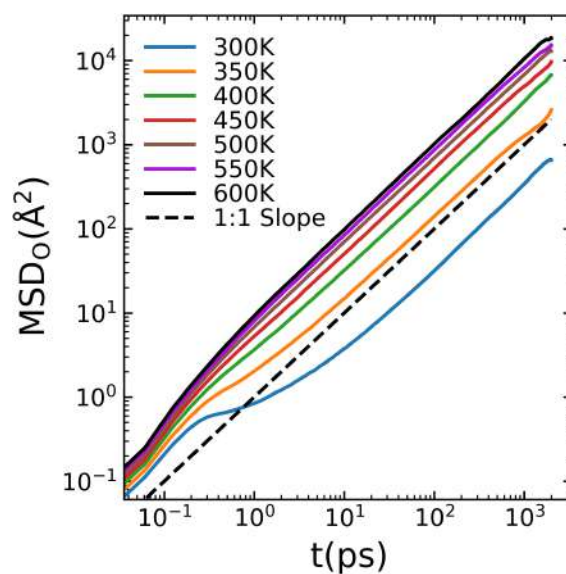


Figure 19: Déplacement quadratique moyen des MLMDs n°0 à différentes températures.

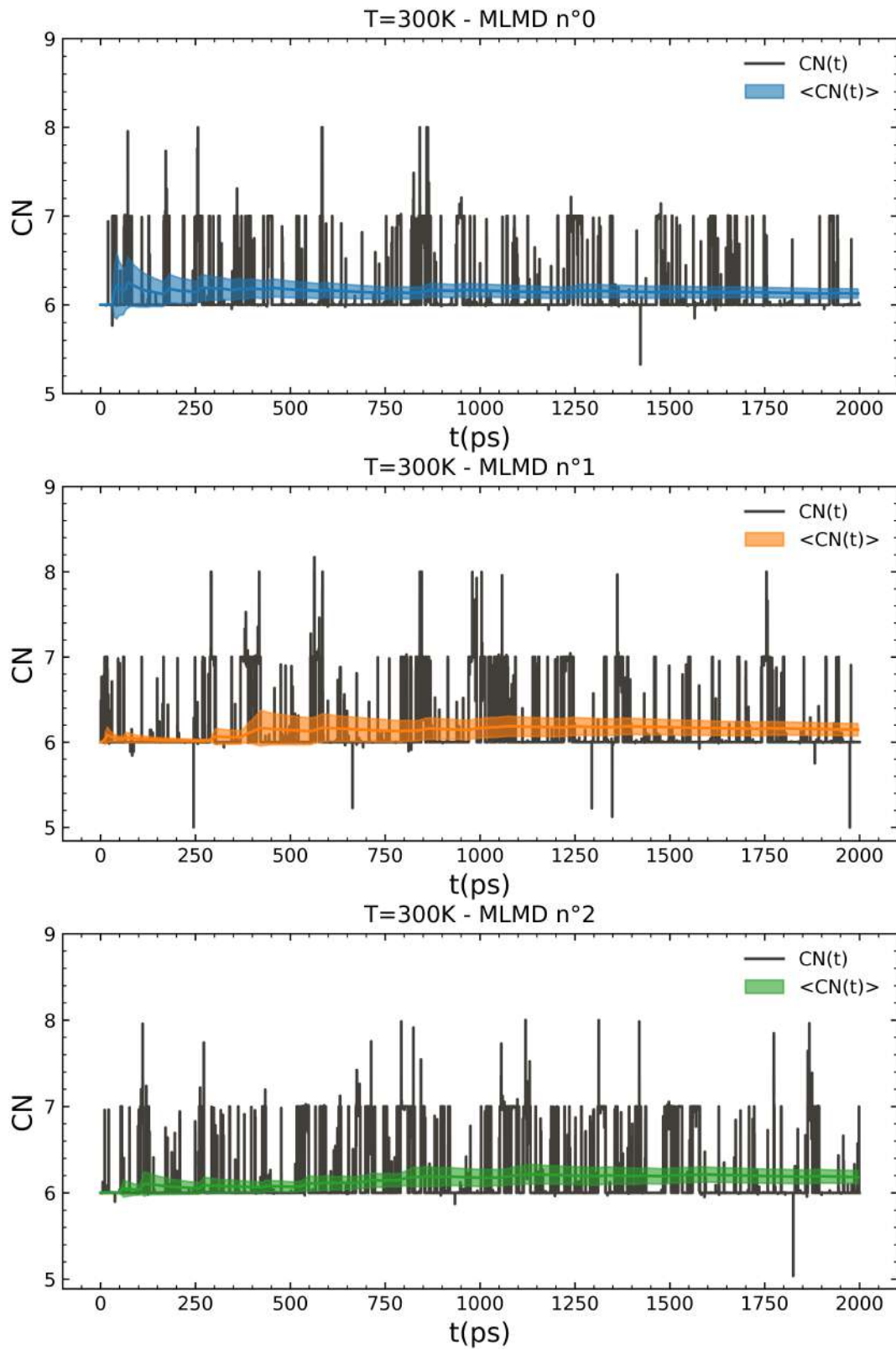


Figure 20: Evolution temporelle de la coordinnce des MLMDs à 300K.

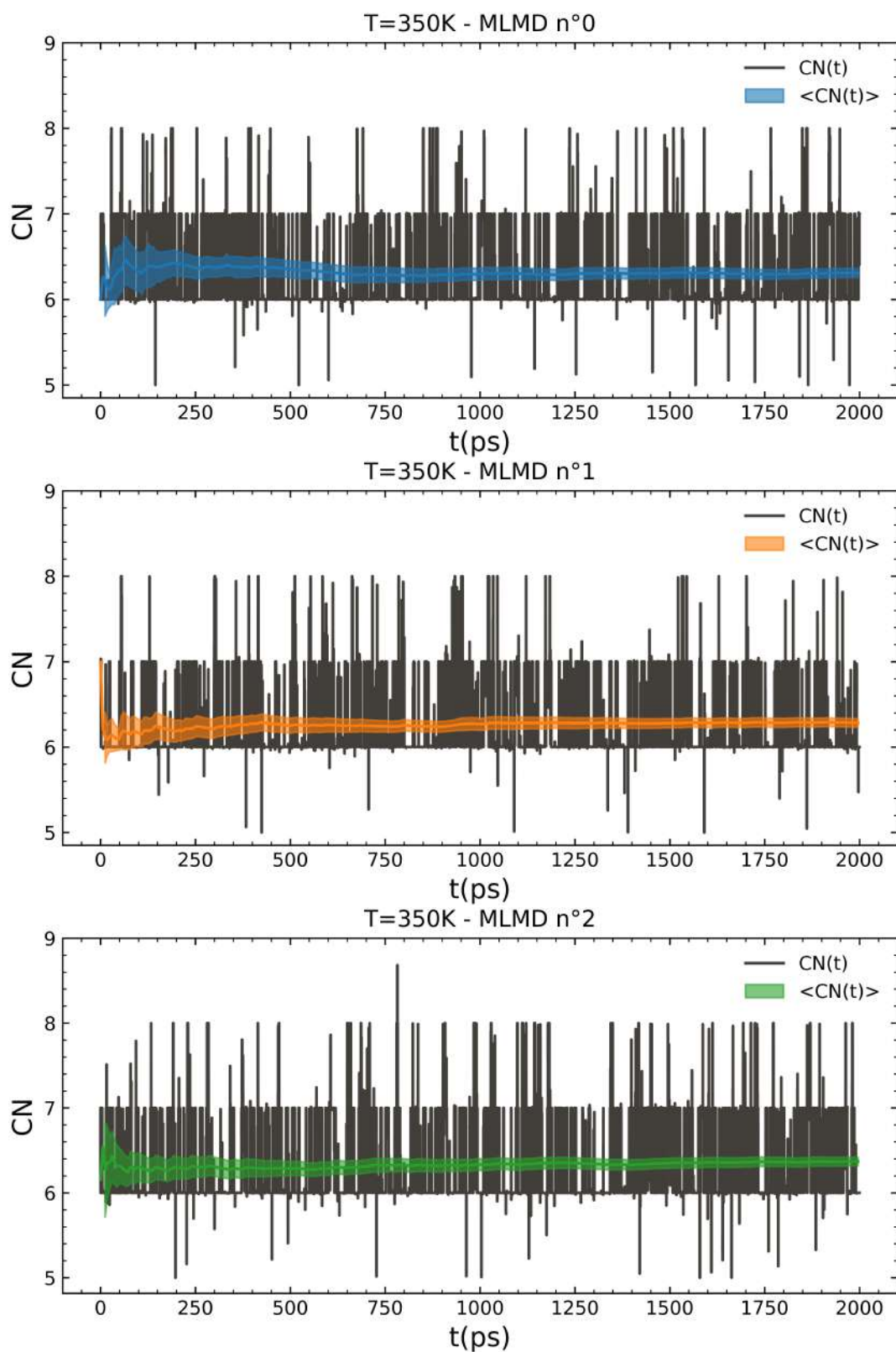


Figure 21: Evolution temporelle de la coordinnce des MLMDs à 350K.

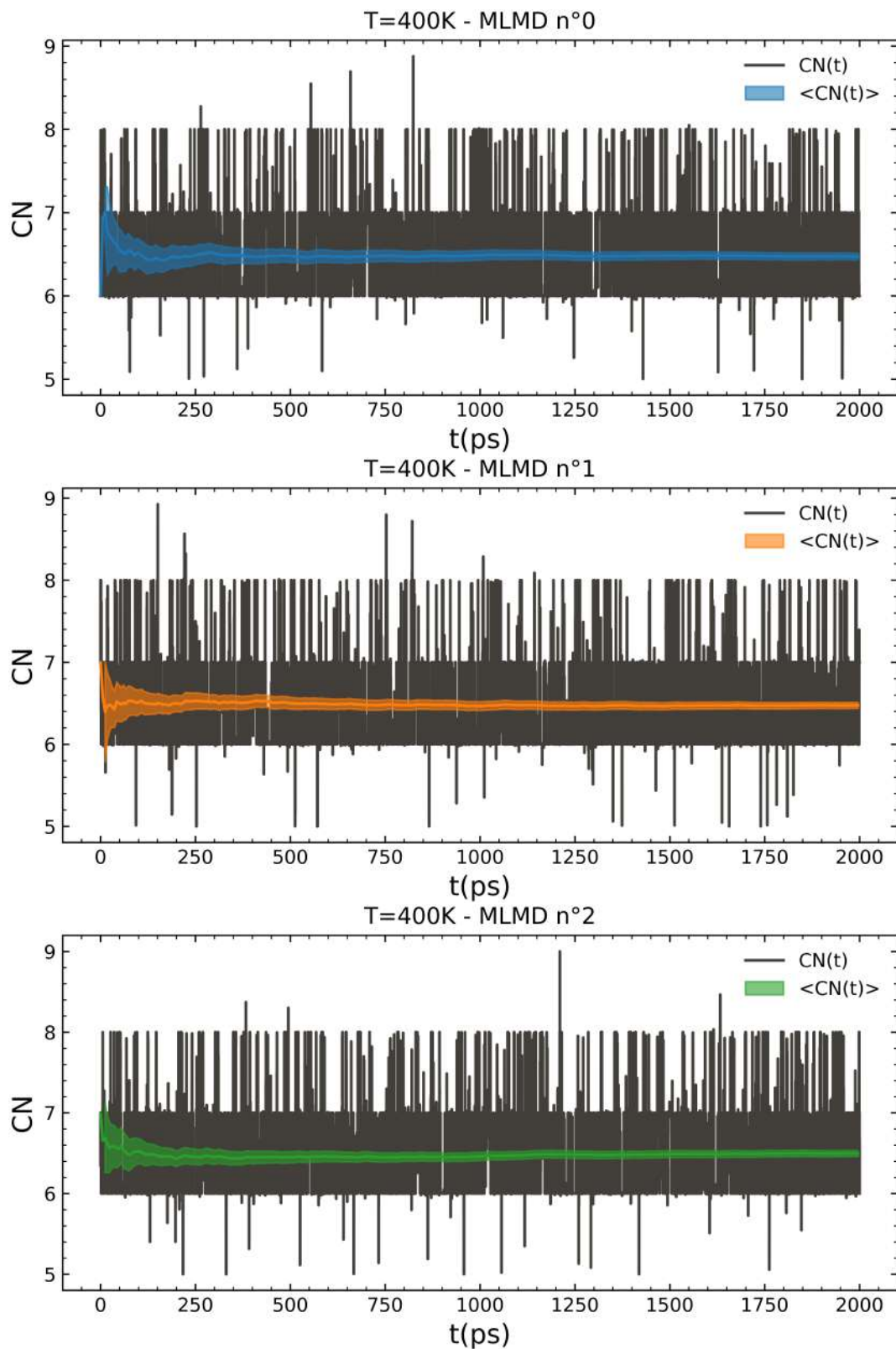


Figure 22: Evolution temporelle de la coordinnce des MLMDs à 400K.

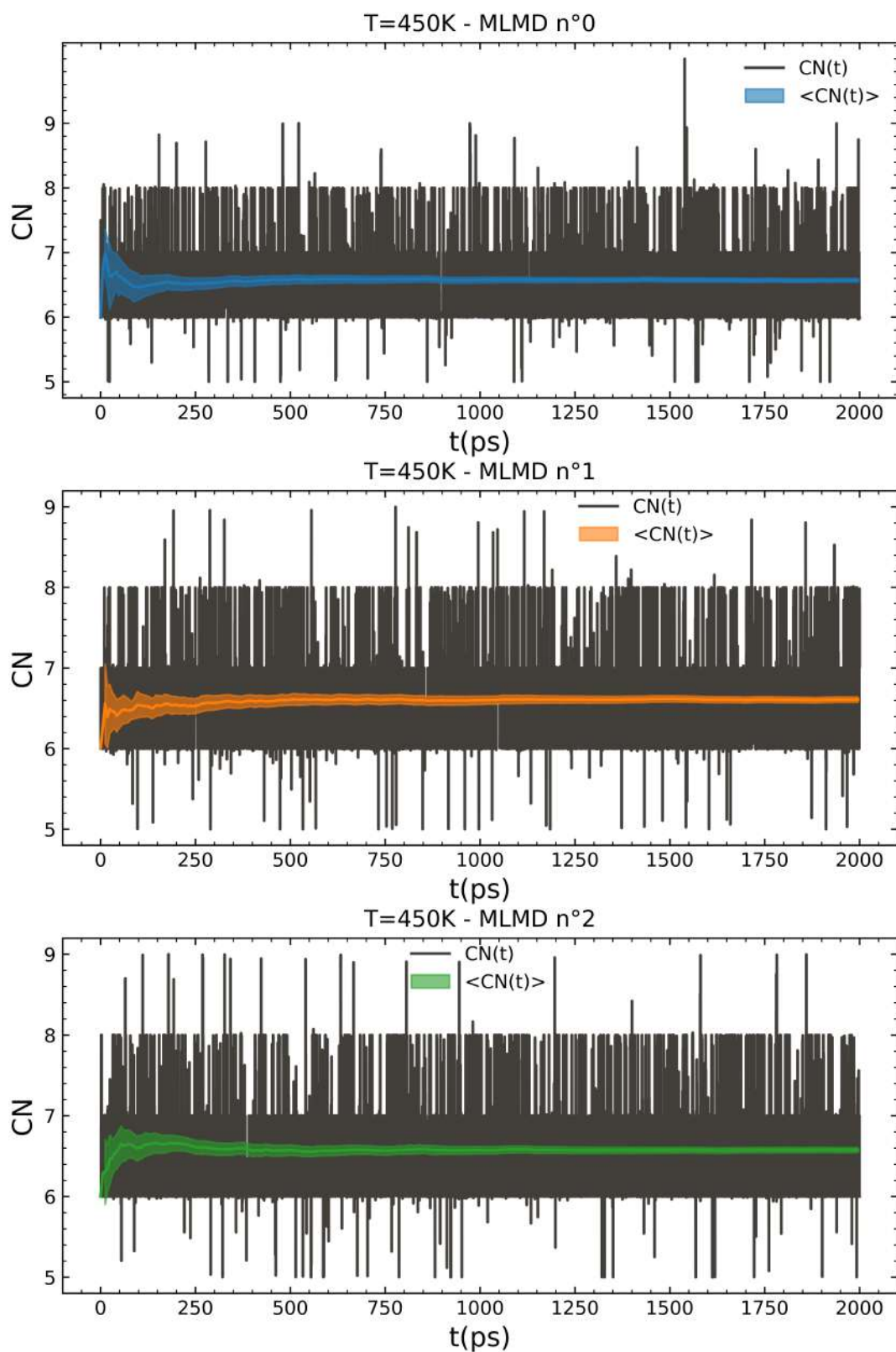


Figure 23: Evolution temporelle de la coordinnce des MLMDs à 450K.

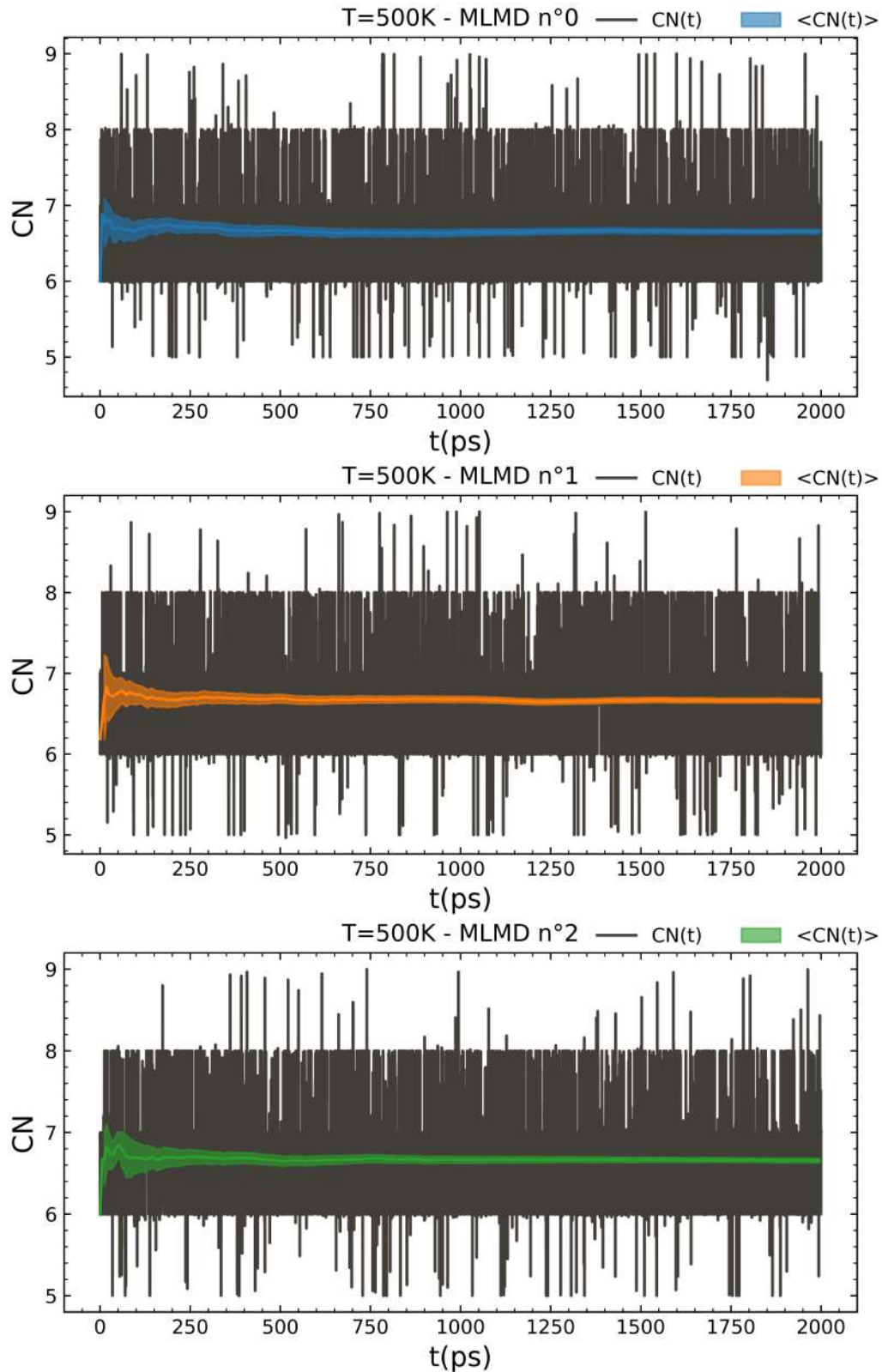


Figure 24: Evolution temporelle de la coordinnce des MLMDs à 500K.

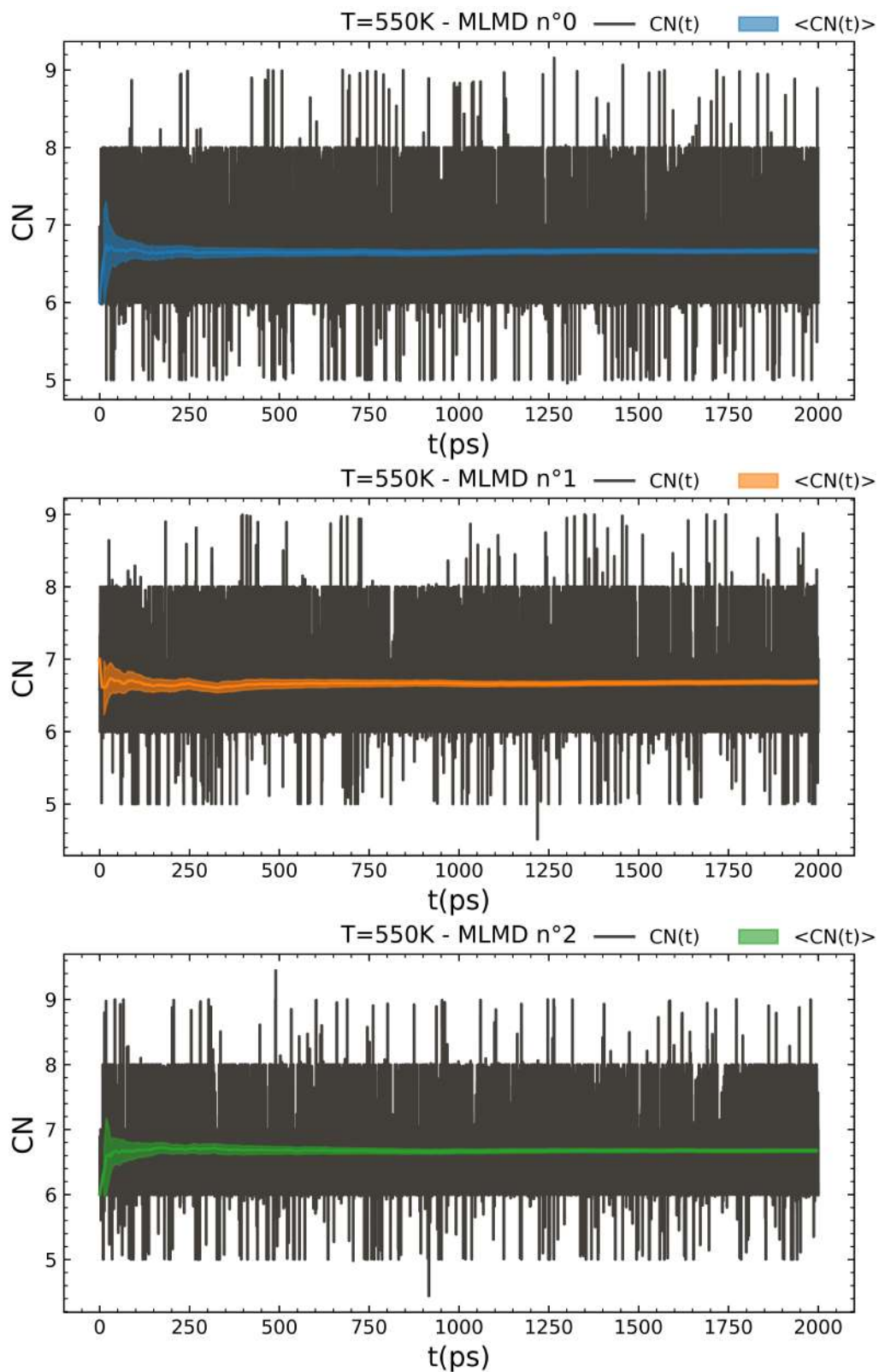


Figure 25: Evolution temporelle de la coordinnce des MLMDs à 550K.

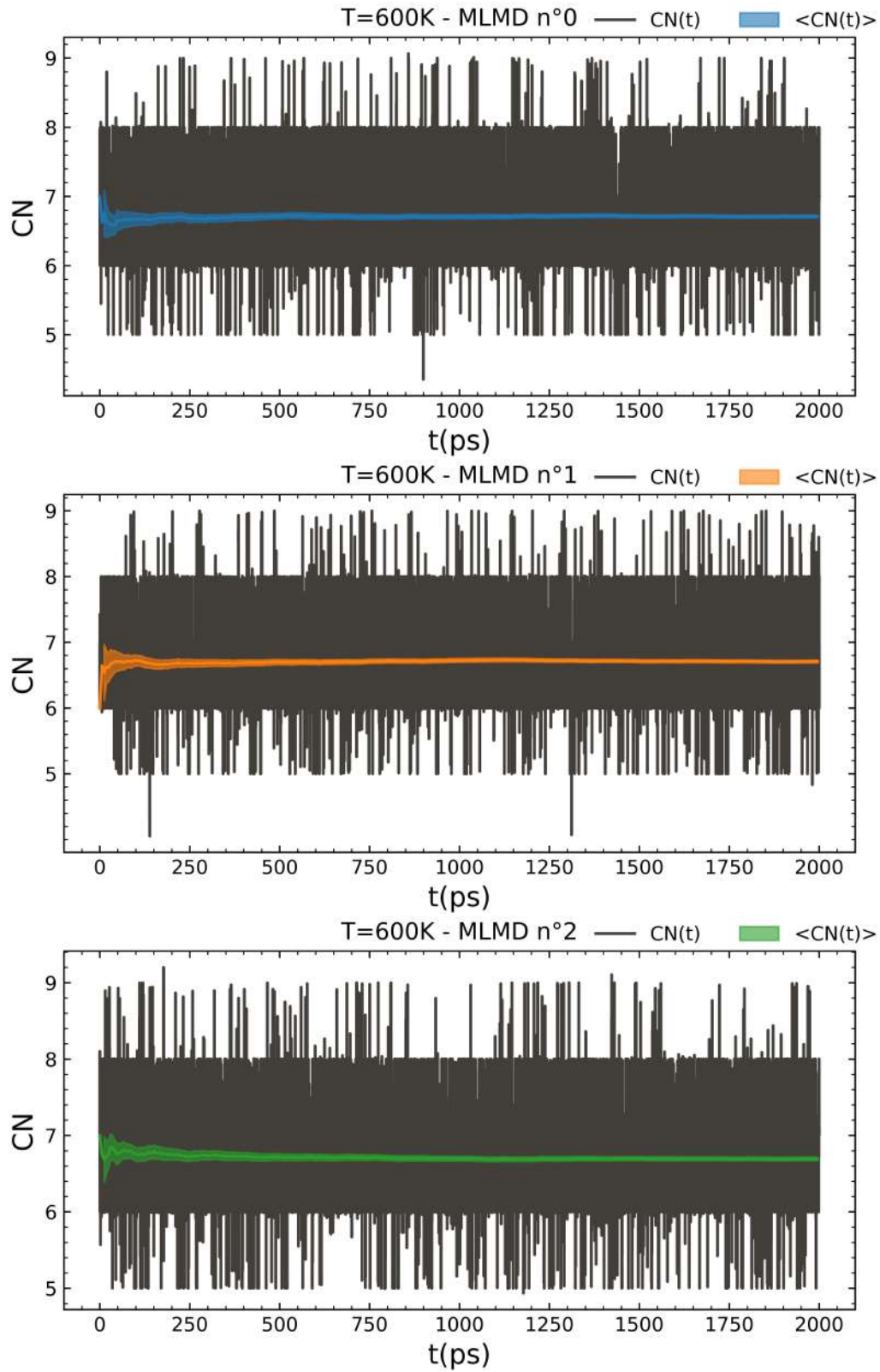


Figure 26: Evolution temporelle de la coordinnce des MLMDs à 600K.

D Compléments du chapitre 5

Evolution de la coordinence en fonction du temps des MLMDs avec le code I-PI et un thermostat de Langevin

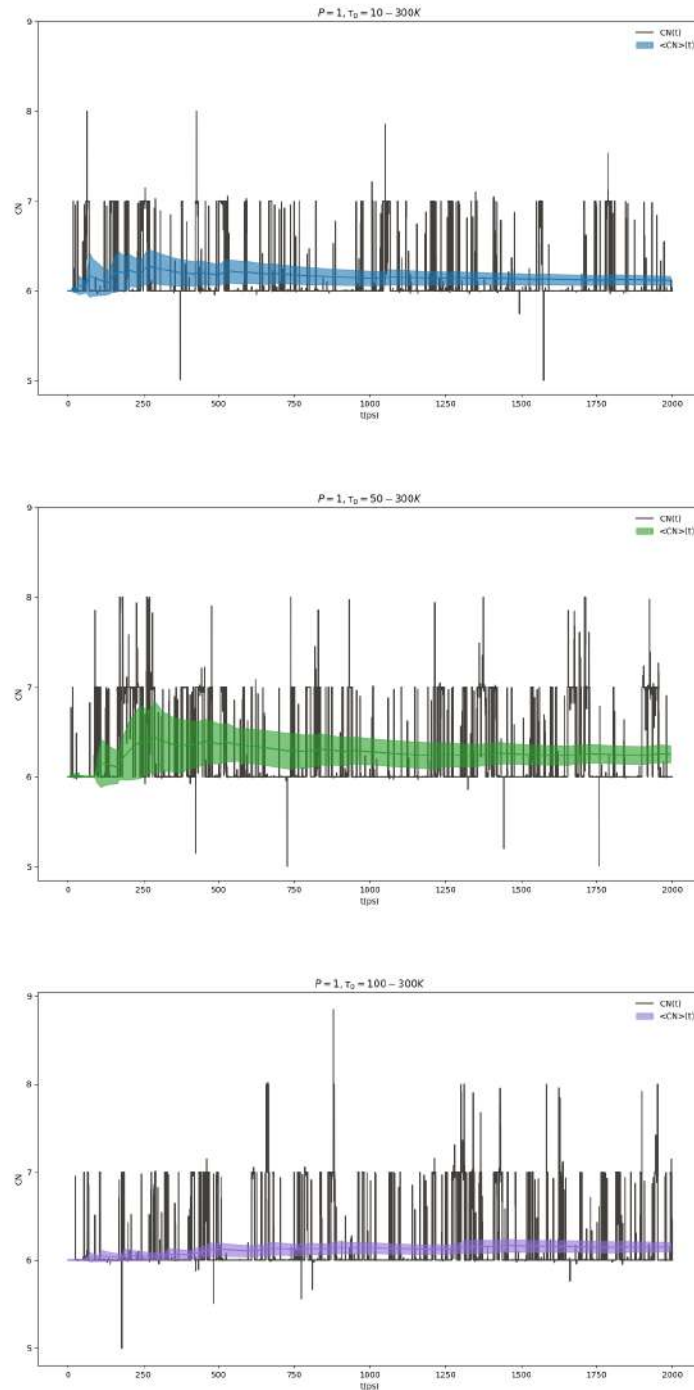


Figure 27: Evolution de la coordinence en fonction du temps $CN(t)$, ainsi que de sa moyenne $\langle CN \rangle(t)$, pour les MLMDs effectuées avec le code I-PI et des thermostats de Langevin pour différentes valeurs de τ_0 à 300K.

Facteur beta déterminé en utilisant des potentiels empiriques

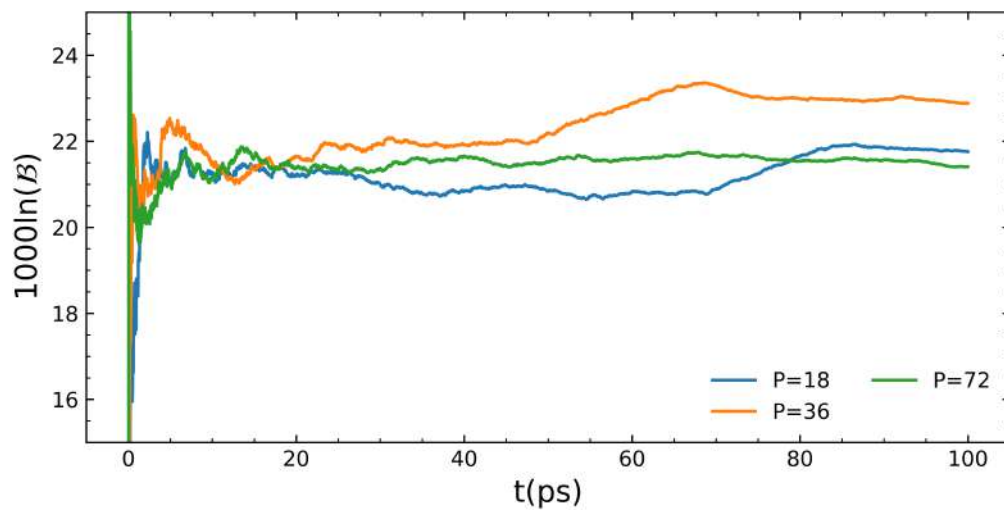


Figure 28: $1000 \ln(\mathcal{B})$ des PIMDs à 300K, avec $\tau_0=100$ et $\lambda=0.5$, utilisant le modèle q-SPC/fw de l'eau et un potentiel de type Lennard-Jones pour l'interaction Ca-O (voir tableau I.1.)

Bibliographie

- [1] James N. Albright. X-Ray Diffraction Studies of Aqueous Alkaline-Earth Chloride Solutions. *The Journal of Chemical Physics*, 56(8):3783–3786, April 1972.
- [2] Markus Allesch, Eric Schwegler, François Gygi, and Giulia Galli. A first principles simulation of rigid water. *The Journal of Chemical Physics*, 120(11):5192–5198, March 2004.
- [3] Michael A. Antonelli, Donald J. DePaolo, John N. Christensen, Jörn-Frederik Wotzlaw, Nicholas J. Pester, and Olivier Bachmann. Radiogenic ^{40}Ca in Seawater: Implications for Modern and Ancient Ca Cycles. *ACS Earth Space Chem.*, 5(9):2481–2492, September 2021.
- [4] Yaspal S. Badyal, Adrian C. Barnes, Gabriel J. Cuello, and John M. Simonson. Understanding the Effects of Concentration on the Solvation Structure of Ca^{2+} in Aqueous Solution. II: Insights into Longer Range Order from Neutron Diffraction Isotope Substitution. *J. Phys. Chem. A*, 108(52):11819–11827, December 2004.
- [5] Marcel D. Baer and Christopher J. Mundy. Local Aqueous Solvation Structure Around Ca^{2+} During $\text{Ca}^{2+} \cdots \text{Cl}^-$ Pair Formation. *J. Phys. Chem. B*, 120(8):1885–1893, March 2016. Publisher: American Chemical Society.
- [6] Marcel D. Baer, Christopher J. Mundy, Matthew J. McGrath, I.-F. Will Kuo, J. Ilja Siepmann, and Douglas J. Tobias. Re-examining the properties of the aqueous vapor–liquid interface using dispersion corrected density functional theory. *The Journal of Chemical Physics*, 135(12):124712, September 2011.
- [7] I. Bakó, J. Hutter, and G. Pálinkás. Car–Parrinello molecular dynamics simulation of the hydrated calcium ion. *The Journal of Chemical Physics*, 117(21):9838–9843, December 2002.
- [8] Arindam Bankura, Anwesa Karmakar, Vincenzo Carnevale, Amalendu Chandra, and Michael L. Klein. Structure, Dynamics, and Spectral Diffusion of Water from First-Principles Molecular Dynamics. *J. Phys. Chem. C*, 118(50):29401–29411, December 2014.
- [9] Stefano Baroni, Stefano De Gironcoli, Andrea Dal Corso, and Paolo Giannozzi. Phonons and related crystal properties from density-functional perturbation theory. *Rev. Mod. Phys.*, 73(2):515–562, July 2001.
- [10] Albert P. Bartók, Risi Kondor, and Gábor Csányi. On representing chemical environments. *Phys. Rev. B*, 87(18):184115, May 2013. arXiv:1209.3140 [cond-mat, physics:physics].

- [11] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38(6):3098–3100, September 1988.
- [12] Jörg Behler. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *The Journal of Chemical Physics*, 134(7):074106, February 2011.
- [13] Jörg Behler. Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations. *Phys. Chem. Chem. Phys.*, 13(40):17930, 2011.
- [14] Jörg Behler. First Principles Neural Network Potentials for Reactive Simulations of Large Molecular and Condensed Systems. *Angew. Chem. Int. Ed.*, 56(42):12828–12840, October 2017.
- [15] M. I. Bernal-Uruchurtu and I. Ortega-Blake. A refined Monte Carlo study of Mg²⁺ and Ca²⁺ hydration. *The Journal of Chemical Physics*, 103(4):1588–1598, July 1995.
- [16] Stuart Bogatko, Emilie Cauët, Eric Bylaska, Gregory Schenter, John Fulton, and John Weare. The Aqueous Ca²⁺ System, in Comparison with Zn²⁺, Fe³⁺, and Al³⁺: An Ab Initio Molecular Dynamics Study. *Chem. Eur. J.*, 19(9):3047–3060, February 2013.
- [17] M. Born and R. Oppenheimer. Zur Quantentheorie der Molekeln. *Ann. Phys.*, 389(20):457–484, 1927.
- [18] Yan Bottinga. Calculation of fractionation factors for carbon and oxygen isotopic exchange in the system calcite-carbon dioxide-water. *J. Phys. Chem.*, 72(3):800–808, March 1968.
- [19] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1):014101, January 2007.
- [20] R. Car and M. Parrinello. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.*, 55(22):2471–2474, November 1985.
- [21] Aloïs Castellano, François Bottin, Johann Bouchet, Antoine Levitt, and Gabriel Stoltz. A *ab initio* canonical sampling based on variational inference. *Phys. Rev. B*, 106(16):L161110, October 2022.
- [22] Michele Ceriotti, Joshua More, and David E. Manolopoulos. i-PI: A Python interface for *ab initio* path integral molecular dynamics simulations. 2014. Publisher: arXiv Version Number: 1.
- [23] Michele Ceriotti, Michele Parrinello, Thomas E. Markland, and David E. Manolopoulos. Efficient stochastic thermostating of path integral molecular dynamics. *The Journal of Chemical Physics*, 133(12):124104, September 2010.

- [24] David Chandler and Peter G. Wolynes. Exploiting the isomorphism between quantum theory and classical statistical mechanics of polyatomic fluids. *The Journal of Chemical Physics*, 74(7):4078–4095, April 1981.
- [25] Melanie B. Channon, Gwyneth W. Gordon, Jennifer L.L. Morgan, Joseph L. Skulan, Scott M. Smith, and Ariel D. Anbar. Using natural, stable calcium isotopes of human blood to detect and monitor changes in bone mineral balance. *Bone*, 77:69–74, August 2015.
- [26] Vladimir I. Chizhik, Andrei V. Egorov, Maria S. Pavlova, Maria I. Egorova, and Aleksey V. Donets. Structure of hydration shell of calcium cation by NMR relaxation, Car-Parrinello molecular dynamics and quantum-chemical calculations. *Journal of Molecular Liquids*, 224:730–736, December 2016.
- [27] Nan-Chin Chu, Gideon M. Henderson, Nick S. Belshaw, and Robert E.M. Hedges. Establishing the potential of Ca isotopes as proxy for consumption of dairy products. *Applied Geochemistry*, 21(10):1656–1667, October 2006.
- [28] Christopher A. Colla, Joshua Wimpenny, Qing-Zhu Yin, James R. Rustad, and William H. Casey. Calcium-isotope fractionation between solution and solids with six, seven or eight oxygens bound to Ca(II). *Geochimica et Cosmochimica Acta*, 121:363–373, November 2013.
- [29] Christopher A. Colla, Joshua Wimpenny, Qing-Zhu Yin, James R. Rustad, and William H. Casey. Calcium-isotope fractionation between solution and solids with six, seven or eight oxygens bound to Ca(II). *Geochimica et Cosmochimica Acta*, 121:363–373, November 2013.
- [30] S. Cummings, J. E. Enderby, and R. A. Howe. Ion hydration in aqueous CaCl₂ solutions. *J. Phys. C: Solid State Phys.*, 13(1):1–8, January 1980. Publisher: IOP Publishing.
- [31] Volker L. Deringer, Miguel A. Caro, and Gábor Csányi. Machine Learning Interatomic Potentials as Emerging Tools for Materials Science. *Adv. Mater.*, 31(46):1902765, November 2019.
- [32] Saaketh Desai, Samuel Temple Reeve, and James F. Belak. Implementing a neural network interatomic model with performance portability for emerging exascale architectures, February 2020. arXiv:2002.00054 [cond-mat, physics:physics].
- [33] Devis Di Tommaso and Nora H. de Leeuw. First Principles Simulations of the Structural and Dynamical Properties of Hydrated Metal Ions Me²⁺ and Solvated Metal Carbonates (Me = Ca, Mg, and Sr). *Crystal Growth & Design*, 10(10):4292–4302, October 2010.
- [34] Robert A. DiStasio, Biswajit Santra, Zhaofeng Li, Xifan Wu, and Roberto Car. The individual and collective effects of exact exchange and dispersion interactions on the *ab initio* structure of liquid water. *The Journal of Chemical Physics*, 141(8):084502, August 2014.

- [35] Romain Dupuis. *Calculs réalistes du fractionnement isotopique du silicium et du lithium pour des équilibres impliquant des phases liquides*. PhD thesis, Université Paul Sabatier, 2014.
- [36] P. P. Ewald. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.*, 369(3):253–287, 1921.
- [37] Matthew S. Fantle and Edward T. Tipper. Calcium isotopes in the global biogeochemical Ca cycle: Implications for development of a Ca isotope proxy. *Earth-Science Reviews*, 129:148–177, February 2014.
- [38] M. V. Fernández-Serra and Emilio Artacho. Network equilibration and first-principles liquid water. *The Journal of Chemical Physics*, 121(22):11136–11144, December 2004.
- [39] R. P. Feynman. Forces in Molecules. *Phys. Rev.*, 56(4):340–343, August 1939.
- [40] John L. Fulton, Steve M. Heald, Yaspal S. Badyal, and J. M. Simonson. Understanding the Effects of Concentration on the Solvation Structure of Ca²⁺ in Aqueous Solution. I: The Perspective on Local Structure from EXAFS and XANES. *J. Phys. Chem. A*, 107(23):4688–4696, June 2003.
- [41] M. Gastegger, L. Schwiedrzik, M. Bittermann, F. Berzsényi, and P. Marquetand. wACSF—Weighted atom-centered symmetry functions as descriptors in machine learning potentials. *The Journal of Chemical Physics*, 148(24):241709, June 2018.
- [42] Andrew Gelman and Donald B. Rubin. Inference from Iterative Simulation Using Multiple Sequences. *Statist. Sci.*, 7(4), November 1992.
- [43] Charles Geyer. Introduction to Markov Chain Monte Carlo. In *Handbook of Markov Chain Monte Carlo*.
- [44] Charles J. Geyer. Practical Markov Chain Monte Carlo. *Statist. Sci.*, 7(4), November 1992.
- [45] Paolo Giannozzi. Notes on pseudopotential generation.
- [46] Paolo Giannozzi, Stefano Baroni, Nicola Bonini, Matteo Calandra, Roberto Car, Carlo Cavazzoni, Davide Ceresoli, Guido L Chiarotti, Matteo Cococcioni, Ismaila Dabo, Andrea Dal Corso, Stefano De Gironcoli, Stefano Fabris, Guido Fratesi, Ralph Gebauer, Uwe Gerstmann, Christos Gougoussis, Anton Kokalj, Michele Lazzeri, Layla Martin-Samos, Nicola Marzari, Francesco Mauri, Riccardo Mazzarello, Stefano Paolini, Alfredo Pasquarello, Lorenzo Paulatto, Carlo Sbraccia, Sandro Scandolo, Gabriele Sclauzero, Ari P Seitsonen, Alexander Smogunov, Paolo Umari, and Renata M Wentzcovitch. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys.: Condens. Matter*, 21(39):395502, September 2009.

- [47] Michael J. Gillan, Dario Alfè, and Angelos Michaelides. Perspective: How good is DFT for water? *The Journal of Chemical Physics*, 144(13):130901, April 2016.
- [48] Feliciano Giustino. *Materials modelling using density functional theory: properties and predictions*. Oxford university press edition, 2014.
- [49] Stefan Grimme. Density functional theory with London dispersion corrections. *WIREs Comput Mol Sci*, 1(2):211–228, March 2011.
- [50] Jeffrey C. Grossman, Eric Schwegler, Erik W. Draeger, François Gygi, and Giulia Galli. Towards an assessment of the accuracy of density functional theory for first principles simulations of water. *The Journal of Chemical Physics*, 120(1):300–311, January 2004.
- [51] Nikolaus Gussone, Anne-Desiree Schmitt, Alexander Heuser, Frank Wombacher, Martin Dietzel, Edward Tipper, and Martin Schiller. *Calcium Stable Isotope Geochemistry*. Advances in Isotope Geochemistry. Springer Berlin Heidelberg, Berlin, Heidelberg, 2016.
- [52] Scott Habershon, David E. Manolopoulos, Thomas E. Markland, and Thomas F. Miller. Ring-Polymer Molecular Dynamics: Quantum Effects in Chemical Dynamics from Classical Trajectories in an Extended Phase Space. *Annu. Rev. Phys. Chem.*, 64(1):387–413, April 2013.
- [53] Randall W. Hall and B. J. Berne. Nonergodicity in path integral molecular dynamics. *The Journal of Chemical Physics*, 81(8):3641–3643, October 1984.
- [54] Hellmann Hans. Einführung in die Quantenchemie. page 185, 1937.
- [55] Anna L. Harrison, Alexander Heuser, Volker Liebetrau, Anton Eisenhauer, Jacques Schott, and Vasileios Mavromatis. Equilibrium Ca isotope fractionation and the rates of isotope exchange in the calcite-fluid and aragonite-fluid systems at 25 °C. *Earth and Planetary Science Letters*, 603:117985, February 2023.
- [56] M. F. Herman, E. J. Bruskin, and B. J. Berne. On path integral Monte Carlo simulations. *The Journal of Chemical Physics*, 76(10):5150–5155, May 1982.
- [57] N. A. Hewish, G. W. Neilson, and J. E. Enderby. Environment of Ca²⁺ ions in aqueous solvent. *Nature*, 297(5862):138–139, May 1982.
- [58] R. W. Hockney and J. W. Eastwood. *Computer Simulation Using Particles*. January 1981. Publication Title: Computer Simulation Using Particles ADS Bibcode: 1981csup.book.....H.
- [59] Thomas S. Hofer, Hung T. Tran, Christian F. Schwenk, and Bernd M. Rode. Characterization of dynamics and reactivities of solvated ions by ab initio simulations. *J. Comput. Chem.*, 25(2):211–217, January 2004.
- [60] P. Hohenberg and W. Kohn. Inhomogeneous Electron Gas. *Phys. Rev.*, 136(3B):B864–B871, November 1964.

- [61] William G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31(3):1695–1697, March 1985.
- [62] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38, February 1996.
- [63] Takashi Ikeda, Mauro Boero, and Kiyoyuki Terakura. Hydration properties of magnesium and calcium ions from constrained first principles molecular dynamics. *The Journal of Chemical Physics*, 127(7):074503, August 2007.
- [64] R. W. Impey, P. A. Madden, and I. R. McDonald. Hydration and mobility of ions in solution. *J. Phys. Chem.*, 87(25):5071–5083, December 1983.
- [65] Farideh Jalilehvand, Daniel Spångberg, Patric Lindqvist-Reis, and Kersti Hermansson. Hydration of the Calcium Ion. An EXAFS, Large-Angle X-ray Scattering, and Molecular Dynamics Simulation Study. *J. Am. Chem. Soc.*, 123(3):431–441, January 2001. Publisher: American Chemical Society.
- [66] Romain Jonchiere, Ari P. Seitsonen, Guillaume Ferlat, A. Marco Saitta, and Rodolphe Vuilleumier. Van der Waals effects in *ab initio* water at ambient and supercritical conditions. *The Journal of Chemical Physics*, 135(15):154503, October 2011.
- [67] W. Kohn and L. J. Sham. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.*, 140(4A):A1133–A1138, November 1965.
- [68] Piotr M. Kowalski and Sandro Jahn. Prediction of equilibrium Li isotope fractionation between minerals and aqueous solutions at high P and T: an efficient *ab initio* approach. 2011. Publisher: arXiv Version Number: 1.
- [69] Chengteh Lee, Weitao Yang, and Robert G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37(2):785–789, January 1988.
- [70] Hee-Seung Lee and Mark E. Tuckerman. Dynamical properties of liquid water from *ab initio* molecular dynamics performed in the complete basis set limit. *The Journal of Chemical Physics*, 126(16):164501, April 2007.
- [71] Kyuho Lee, Éamonn D. Murray, Lingzhu Kong, Bengt I. Lundqvist, and David C. Langreth. Higher-accuracy van der Waals density functional. *Phys. Rev. B*, 82(8):081101, August 2010.
- [72] G. Licheri. X-ray diffraction study of the average solute species in CaCl₂ aqueous solutions. *J. Chem. Phys.*, 64(6):2437, 1976.
- [73] Felice C. Lightstone, Eric Schwegler, Markus Allesch, François Gygi, and Giulia Galli. A First-Principles Molecular Dynamics Study of Calcium in Water. *ChemPhysChem*, 6(9):1745–1749, September 2005.

- [74] I-Chun Lin, Ari P. Seitsonen, Ivano Tavernelli, and Ursula Rothlisberger. Structure and Dynamics of Liquid Water from *ab Initio* Molecular Dynamics—Comparison of BLYP, PBE, and revPBE Density Functionals with and without van der Waals Corrections. *J. Chem. Theory Comput.*, 8(10):3902–3910, October 2012.
- [75] Ondrej Marsalek, Pei-Yang Chen, Romain Dupuis, Magali Benoit, Merlin Méheut, Zlatko Bačić, and Mark E. Tuckerman. Efficient Calculation of Free Energy Differences Associated with Isotopic Substitution Using Path-Integral Molecular Dynamics. *J. Chem. Theory Comput.*, 10(4):1440–1453, April 2014.
- [76] Glenn J. Martyna, Michael L. Klein, and Mark Tuckerman. Nosé–Hoover chains: The canonical ensemble via continuous dynamics. *The Journal of Chemical Physics*, 97(4):2635–2643, August 1992.
- [77] L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.*, 30(13):2157–2164, October 2009.
- [78] Tünde Megyes, Tamás Grósz, Tamás Radnai, Imre Bakó, and Gábor Pálinkás. Solvation of Calcium Ion in Polar Solvents: An X-ray Diffraction and *ab Initio* Study. *J. Phys. Chem. A*, 108(35):7261–7271, September 2004.
- [79] Frédéric Moynier and Toshiyuki Fujii. Calcium isotope fractionation between aqueous compounds relevant to low-temperature geochemistry, biology and medicine. *Sci Rep*, 7(1):44255, March 2017.
- [80] Frédéric Moynier and Toshiyuki Fujii. Calcium isotope fractionation between aqueous compounds relevant to low-temperature geochemistry, biology and medicine. *Sci Rep*, 7(1):44255, March 2017.
- [81] Merlin Méheut. *Calcul prédictif du facteur de partage isotopique entre minéraux dans le cadre de la théorie de la fonctionnelle de la densité*. PhD thesis, Université Pierre et Marie Curie, 2008.
- [82] Mor M. Naor, Keith Van Nostrand, and Christoph Dellago. Car–Parrinello molecular dynamics simulation of the calcium ion in liquid water. *Chemical Physics Letters*, 369(1-2):159–164, February 2003.
- [83] Shuichi Nosé. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics*, 81(1):511–519, July 1984.
- [84] Francesco Paesani, Wei Zhang, David A. Case, Thomas E. Cheatham, and Gregory A. Voth. An accurate and simple quantum model for liquid water. *The Journal of Chemical Physics*, 125(18):184507, November 2006.
- [85] G. Pastore, E. Smargiassi, and F. Buda. Theory of *ab initio* molecular-dynamics calculations. *Phys. Rev. A*, 44(10):6334–6347, November 1991.

- [86] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.*, 77(18):3865–3868, October 1996.
- [87] John P. Perdew, Adrienn Ruzsinszky, Gábor I. Csonka, Oleg A. Vydrov, Gustavo E. Scuseria, Lucian A. Constantin, Xiaolan Zhou, and Kieron Burke. Restoring the Density-Gradient Expansion for Exchange in Solids and Surfaces. *Phys. Rev. Lett.*, 100(13):136406, April 2008.
- [88] M. M. Probst, T. Radnai, K. Heinzinger, P. Bopp, and B. M. Rode. Molecular dynamics and x-ray investigation of an aqueous calcium chloride solution. *J. Phys. Chem.*, 89(5):753–759, February 1985.
- [89] Paolo Raiteri, Raffaella Demichelis, and Julian D. Gale. Thermodynamically Consistent Force Field for Molecular Dynamics Simulations of Alkaline-Earth Carbonates and Their Aqueous Speciation. *J. Phys. Chem. C*, 119(43):24447–24458, October 2015.
- [90] Bruno Reynard and Vincent Balter. Trace elements and their isotopes in bones and teeth: Diet, environments, diagenesis, and dating of archeological and paleontological samples. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 416:4–16, December 2014.
- [91] L.M. Reynard, G.M. Henderson, and R.E.M. Hedges. Calcium isotopes in archaeological bones and their relationship to dairy consumption. *Journal of Archaeological Science*, 38(3):657–664, March 2011.
- [92] P. Richet, Y. Bottinga, and M. Janoy. A review of hydrogen, carbon, nitrogen, oxygen, sulphur, and chlorine stable isotope enrichment among gaseous molecules. *Annual Review of Earth and Planetary Sciences*, 5:65–110, January 1977. ADS Bibcode: 1977AREPS...5...65R.
- [93] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, October 1986.
- [94] E. A. Schauble. Applying Stable Isotope Fractionation Theory to New Systems. *Reviews in Mineralogy and Geochemistry*, 55(1):65–111, January 2004.
- [95] Anne-Désirée Schmitt, Natalia Borrelli, Damien Ertlen, Sophie Gangloff, François Chabaux, and Margarita Osterrieth. Stable calcium isotope speciation and calcium oxalate production within beech tree (*Fagus sylvatica* L.) organs. *Biogeochemistry*, 137(1-2):197–217, January 2018.
- [96] Anne-Désirée Schmitt, Nathalie Vigier, Damien Lemarchand, Romain Millot, Peter Stille, and François Chabaux. Processes controlling the stable isotope compositions of Li, B, Mg and Ca in plants, soils and waters: A review. *Comptes Rendus Geoscience*, 344(11-12):704–722, November 2012.

- [97] Eric Schwegler, Jeffrey C. Grossman, François Gygi, and Giulia Galli. Towards an assessment of the accuracy of density functional theory for first principles simulations of water. II. *The Journal of Chemical Physics*, 121(11):5400–5409, September 2004.
- [98] C. F. Schwenk and B. M. Rode. Ab initio QM/MM MD simulations of the hydrated Ca²⁺ ion. *Pure and Applied Chemistry*, 76(1):37–47, January 2004.
- [99] Christian F. Schwenk, Hannes H. Loeffler, and Bernd M. Rode. Molecular dynamics simulations of Ca²⁺ in water: Comparison of a classical simulation including three-body corrections and Born–Oppenheimer *ab initio* and density functional theory quantum mechanical/molecular mechanics simulations. *The Journal of Chemical Physics*, 115(23):10808–10813, December 2001.
- [100] Andreas Singraber, Tobias Morawietz, Jörg Behler, and Christoph Dellago. Parallel Multistream Training of High-Dimensional Neural Network Potentials. *J. Chem. Theory Comput.*, 15(5):3075–3092, May 2019.
- [101] Andreas Singraber, Mpbircher, Sam Reeve, David W.H. Swenson, Jérémy Lauret, and Philippedavid. CompPhysVienna/n2p2: Version 2.1.4, May 2021.
- [102] P. H.-L. Sit and Nicola Marzari. Static and dynamical properties of heavy water at ambient conditions from first-principles molecular dynamics. *The Journal of Chemical Physics*, 122(20):204510, May 2005.
- [103] L. B. Skinner, C. J. Benmore, J. C. Neufeind, and J. B. Parise. The structure of water around the compressibility minimum. *The Journal of Chemical Physics*, 141(21):214507, December 2014.
- [104] Pavel Smirnov, Motoyuki Yamagami, Hisanobu Wakita, and Toshio Yamaguchi. An X-ray diffraction study on concentrated aqueous calcium nitrate solutions at subzero temperatures. *Journal of Molecular Liquids*, 73-74:305–316, November 1997.
- [105] A.K. Soper. The radial distribution functions of water and ice from 220 to 673 K and at pressures up to 400 MPa. *Chemical Physics*, 258(2-3):121–137, August 2000.
- [106] Daniel Spångberg, Kersti Hermansson, Patric Lindqvist-Reis, Farideh Jalilehvand, Magnus Sandström, and Ingmar Persson. Model Extended X-ray Absorption Fine Structure (EXAFS) Spectra from Molecular Dynamics Data for Ca²⁺ and Al³⁺ Aqueous Solutions. *J. Phys. Chem. B*, 104(45):10467–10472, November 2000.
- [107] William C. Swope, Hans C. Andersen, Peter H. Berens, and Kent R. Wilson. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *The Journal of Chemical Physics*, 76(1):637–649, January 1982.

- [108] Kazuko Tanaka. Measurements of self-diffusion coefficients of water in pure water and in aqueous electrolyte solutions. *J. Chem. Soc., Faraday Trans. 1*, 71(0):1127, 1975.
- [109] Aidan P. Thompson, H. Metin Aktulga, Richard Berger, Dan S. Bolintineanu, W. Michael Brown, Paul S. Crozier, Pieter J. In 'T Veld, Axel Kohlmeyer, Stan G. Moore, Trung Dac Nguyen, Ray Shan, Mark J. Stevens, Julien Tranchida, Christian Trott, and Steven J. Plimpton. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Computer Physics Communications*, 271:108171, February 2022.
- [110] Aidan P. Thompson, Laura P. Swiler, Christian R. Trott, Stephen M. Foiles, and Garritt J. Tucker. A Spectral Analysis Method for Automated Generation of Quantum-Accurate Interatomic Potentials. *Journal of Computational Physics*, 285:316–330, March 2015. arXiv:1409.3880 [cond-mat].
- [111] A.P. Thompson, L.P. Swiler, C.R. Trott, S.M. Foiles, and G.J. Tucker. Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials. *Journal of Computational Physics*, 285:316–330, March 2015.
- [112] Teodora Todorova, Philippe H. Hünenberger, and Jürg Hutter. Car–Parrinello Molecular Dynamics Simulations of CaCl_2 Aqueous Solutions. *J. Chem. Theory Comput.*, 4(5):779–789, May 2008.
- [113] Teodora Todorova, Ari P. Seitsonen, Jürg Hutter, I-Feng W. Kuo, and Christopher J. Mundy. Molecular Dynamics Simulation of Liquid Water: Hybrid Density Functionals. *J. Phys. Chem. B*, 110(8):3685–3691, March 2006.
- [114] H. F. Trotter. On the product of semi-groups of operators. *Proc. Amer. Math. Soc.*, 10(4):545–551, 1959.
- [115] N. Troullier and José Luriaas Martins. Efficient pseudopotentials for plane-wave calculations. *Phys. Rev. B*, 43(3):1993–2006, January 1991.
- [116] Mark E Tuckerman. *Statistical Mechanics Theory and Molecular Simulation*.
- [117] Mark E. Tuckerman, Bruce J. Berne, Glenn J. Martyna, and Michael L. Klein. Efficient molecular dynamics and hybrid Monte Carlo algorithms for path integrals. *The Journal of Chemical Physics*, 99(4):2796–2808, August 1993.
- [118] Harold C. Urey. The thermodynamic properties of isotopic substances. *J. Chem. Soc.*, page 562, 1947.
- [119] David Vanderbilt. Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. *Phys. Rev. B*, 41(11):7892–7895, April 1990.
- [120] Joost VandeVondele, Fawzi Mohamed, Matthias Krack, Jürg Hutter, Michiel Sprik, and Michele Parrinello. The influence of temperature and density

- functional models in *ab initio* molecular dynamics simulation of liquid water. *The Journal of Chemical Physics*, 122(1):014515, January 2005.
- [121] Aki Vehtari, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. Rank-normalization, folding, and localization: An improved \widehat{R} for assessing convergence of MCMC. *Bayesian Anal.*, 16(2), June 2021. arXiv:1903.08008 [stat].
- [122] Loup Verlet. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.*, 159(1):98–103, July 1967.
- [123] Jui Hsin Wang, Charles V. Robinson, and I. S. Edelman. Self-diffusion and Structure of Liquid Water. III. Measurement of the Self-diffusion of Liquid Water with H^2 , H^3 and O^{18} as Tracers¹. *J. Am. Chem. Soc.*, 75(2):466–470, January 1953.
- [124] Xiangwen Wang, Dimitrios Toroz, Seonmyeong Kim, Simon L. Clegg, Gun-Sik Park, and Devis Di Tommaso. Density functional theory based molecular dynamics study of solution composition effects on the solvation shell of metal ions. *Phys. Chem. Chem. Phys.*, 22(28):16301–16313, 2020.
- [125] Yujie Wu, Harald L. Tepper, and Gregory A. Voth. Flexible simple point-charge water model with improved liquid-state properties. *The Journal of Chemical Physics*, 124(2):024503, January 2006.
- [126] Howard Yanxon, David Zagaceta, Binh Tang, David S Matteson, and Qiang Zhu. PyXtal_ff: a python library for automated force field generation. *Mach. Learn.: Sci. Technol.*, 2(2):027001, June 2021.
- [127] David Zagaceta, Howard Yanxon, and Qiang Zhu. Spectral neural network potentials for binary alloys. *Journal of Applied Physics*, 128(4):045113, July 2020.

Titre : Modélisation atomistique du calcium en solution appliquée à l'étude du fractionnement isotopique

Mots clés : dynamique moléculaire, Effets quantiques nucléaires, fractionnement isotopique

Résumé : La composition en isotopes stables du calcium de systèmes géochimiques permet d'obtenir des informations concernant les processus survenant dans le cycle biochimique du calcium de la zone critique. Ces processus sont encore mal connus et les simulations atomistiques permettent d'étudier le fractionnement isotopique entre deux systèmes, notamment entre le calcium en solution aqueuse, Ca^{2+} , et des minéraux. Ce fractionnement dépend de la structure autour de Ca^{2+} et en particulier de sa première sphère de coordination. Cependant la valeur de la coordinence est toujours sujette à débat expérimentalement et théoriquement. Dans cette thèse, une première partie a donc été consacrée à l'étude de la coordinence de Ca^{2+} à l'aide de dynamiques moléculaires ab-initio (AIMD) et de la théorie de la fonctionnelle de la densité (DFT), le but étant de déterminer les paramètres numériques d'influence. Plusieurs fonctionnelles ont été testées confirmant la nécessité d'ajouter des termes de dispersion pour reproduire la structure de l'eau. La diffusion, ainsi que la variation du nombre de molécules d'eau dans la sphère de coordination, ont été étudiées pour des simulations effectuées à plusieurs températures ce qui a permis de mettre en évidence des problèmes d'équilibration à température ambiante. Cela nous a poussé à mettre en place des protocoles de validation de l'équilibration en utilisant des chaînes de Markov. Les résultats montrent aussi une légère dépendance en température de la coordinence de Ca^{2+} . Les propriétés du fractionnement isotopique ont ensuite été calculées dans le cadre de l'approximation harmonique. Pour les liquides, l'approximation harmonique peut engendrer des erreurs. Pour s'en affranchir, une méthode consiste à réaliser des dynamiques moléculaires d'intégrales de chemin (PIMD). Les simulations PIMD ab initio sont coûteuses en temps de calcul et l'utilisation de potentiels empiriques ne permet pas de reproduire correctement les propriétés du fractionnement isotopique de Ca^{2+} . Une seconde partie s'est donc concentrée sur l'utilisation de méthodes "machine learning" pour construire des potentiels interatomiques ayant une précision bien supérieure aux potentiels empiriques avec un temps de calcul réduit par rapport à la DFT. Les méthodes utilisant des réseaux de neurones ont donné de meilleurs résultats. Finalement, des PIMD ont été réalisées avec ces potentiels. Le fractionnement isotopique a été calculé par intégration thermodynamique et comparé aux résultats obtenus en utilisant l'approximation harmonique.

Title: Atomistic modelling of calcium in solution applied to the study of isotopic fractionation

Key words: molecular dynamics, quantum nuclear effects, isotopic fractionation

Abstract: The stable calcium isotope composition of geochemical systems provides information about processes occurring in the biochemical calcium cycle of the critical zone. These processes are still poorly understood, and atomistic simulations make it possible to study the isotopic fractionation between two systems, in particular between calcium in aqueous solution, Ca^{2+} , and minerals. This fractionation depends on the structure around Ca^{2+} and in particular on its first coordination sphere. However, the value of the coordination sphere is still subject to debate, both experimentally and theoretically. The first part of this thesis was therefore devoted to the study of Ca^{2+} coordination using ab-initio molecular dynamics (AIMD) and density functional theory (DFT), with the aim of determining the influencing numerical parameters. Several functionals were tested, confirming the need to add dispersion terms to reproduce the structure of water. Diffusion, as well as the variation in the number of water molecules in the coordination sphere, were studied for simulations carried out at several temperatures, revealing equilibration problems at room temperature. This led us to set up equilibration validation protocols using Markov chains. The results also show a slight temperature dependence of Ca^{2+} coordination. Isotopic fractionation properties were then calculated using the harmonic approximation. For liquids, harmonic approximation can lead to errors. To overcome this, one method is to perform path integral molecular dynamics (PIMD). Ab initio PIMD simulations are costly in terms of computation time, and the use of empirical potentials does not correctly reproduce the properties of Ca^{2+} isotope fractionation. A second part therefore focused on the use of "machine learning" methods to construct interatomic potentials with a much higher accuracy than empirical potentials, with a reduced computation time compared to DFT. Methods using neural networks gave better results. Finally, PIMD was performed with these potentials. Isotopic fractionation was calculated by thermodynamic integration and compared with those obtained using the harmonic approximation.