Ab Initio Simulations of Activated Processes: Application to CO_2 Reactions and a Photovoltaic Material

THÈSE N^{O} 5918 (2013)

PRÉSENTÉE LE 25 OCTOBRE 2013
À LA FACULTÉ DES SCIENCES DE BASE
CENTRE EUROPÉEN DE CALCUL ATOMIQUE ET MOLÉCULAIRE
PROGRAMME DOCTORAL EN PHYSIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Grégoire GALLET

acceptée sur proposition du jury:

Prof. H. M. Rønnow, président du jury
Prof. W. Andreoni, Dr F. Pietrucci, directeurs de thèse
Prof. R. Car, rapporteur
Prof. C. Dellago, rapporteur
Dr C. Pignedoli, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2013

Life is tough
and then you graduate.
— Tajel, Piled Higher and Deeper

Pour Mamie…

Acknowledgements

I would like to spend a few words thanking those who helped to complete my thesis and supported me during this long and passionate journey: first and foremost, my thesis advisors: Prof. Wanda Andreoni and Dr Fabio Pietrucci who took the time to read, improve, correct, synthetize and read time and time again the present work. During these few years, they helped me unweil the wonders of the world of research and in their company I learned that the path taken to answer a question is at least as important as the actual answer.

I wish also to thank my family, of course: my grandmother, Mamie, who left us during the completion of my Ph. D., who had always been there for me and whom I miss dearly today. My mum who always supported me even though she is closer to the humanities than hard sciences. I am not forgetting the rest of the family: my dad, Maxime, Juliette, Boudou, Claire-Lise and Thierry, Claude, Nadine and the girls, and Isabelle and Philippe.

My best friends (women first and in alphabetical order so as not to make anyone jealous): Nathalie and Turan, Alexis, Julien, Steve, Valentin and Vincent among which we can find, to this day, one and a half Ph. D. at least twice many engineers and a few exceptionnaly talented entrepreneurs; those guys know when to work hard and when to party and I could not have dreamt better friends to spend the more or less ten years that separate us from when we were young and we thought we knew everyhing.

Lausanne, October 25th 2013 Grégoire A. Gallet

Abstract

This thesis is devoted to the study of activated processes at the nanoscale by means of state-of-the-art computer simulations. On the one hand, we are interested in addressing specific aspects of chemico-physical processes of great practical importance, namely the early steps of wet carbon dioxide capture, and the diffusion of an impurity in cadmium telluride. On the other hand, we assess the limitations of current methodologies and propose new tools and protocols to extend their range of applicability.

The simulation of the dynamics of activated processes at the atomistic level is a delicate issue facing two major difficulties: (i) an accurate description of the changes happening in the electronic structure has to be obtained ; and (ii) a statistically meaningful portion of the configuration space has to be explored, including scarcely populated regions like transition states. By employing finite temperature *ab initio* molecular dynamics empowered by metadynamics, it is possible to successfully overcome these challenges. We apply these methods and develop general robust protocols for these simulations and for the analysis of the results.

Applications refer to CO_2 chemical reactions, which are relevant to the science of carbon capture, and sulfur diffusion in CdTe, a photovoltaic material of current great interest. An unprecedented characterization of these processes is obtained, leading to new understanding of the experimentally observed behavior, well beyond the reach of state-of-the-art calculations. For example, our simulations disclose the active role played by water in the capture of carbon dioxide, and reveal the easiness by which sulfur is trapped in a CdTe lattice.

Keywords : activated process, chemical reaction, *ab initio* molecular dynamics, enhanced sampling, metadynamics, carbon dioxide, monoethanolamine, water, cadmium telluride, atom diffusion in solids, impurities in semiconductors, structural clustering, collective variables, permutation symmetry

Résumé

Le travail de recherche présenté dans cette thèse de doctorat porte sur l'étude des processus nanoscopiques activés au moyen de simulations numériques avancées. Nous nous intéresserons pour partie à des processus physico-chimiques particulièrement importants tels que les étapes initiales de la capture du dioxyde de carbone (CO_2) en solution ainsi qu'à la diffusion d'une impureté dans le tellurure de cadmium (CdTe), un matériau employé dans certaines cellules photovoltaïques. Nous explorerons également les limites de méthodes actuelles et proposerons de nouveaux outils et protocoles pour étendre leur champ d'application.

La simulation au niveau atomique de la dynamique des processus activés est une tâche délicate confrontée à deux obstacles majeurs : (i) les changements survenus dans la structure électronique doivent pouvoir être décrits avec précision ; et (ii) l'espace des configurations, incluant également des régions peu peuplées tels que les états de transitions, doit pouvoir être exploré de façon statistiquement satisfaisante.

L'utilisation de la méthode, dite *ab initio*, de dynamique moléculaire à température finie et augmentée de la métadynamique permet d'attaquer ces deux difficultés de façon efficace et de les surmonter. Nous appliquerons ces techniques et développerons des procédures générales et robustes, à la fois pour les simulations en tant que telles, ainsi que pour l'analyse des résultats.

Les applications portent sur les réactions chimiques du CO_2 ainsi que sur la diffusion du souffre dans CdTe. En caractérisant ces processus au-delà des limites atteintes par les simulations actuelles, nous pouvons améliorer la compréhension des observations expérimentales. A titre d'exemple, nos calculs ont révélé que le solvant joue un rôle prépondérant lors de la capture du CO_2 et que le souffre peut très facilement se faire piéger dans le réseau de tellurure de cadmium.

Mots clés : processus activés, réactions chimiques, dynamique moléculaire *ab initio*, exploration améliorée, métadynamique, dioxyde de carbone, monoethanolamie, eau, tellurure de cadmium, diffusion d'atomes dans les solides, impuretés dans les semi-conducteurs, agrégations de structures, variables collectives, symétrie de permutation

Table of Contents

Acknowledgements **v**

Abstract (English/Français) **vii**

Table of Contents **xii**

List of Figures **xiv**

List of Tables **xv**

1 Introduction **1**

2 Fundamentals of Theory and Simulation Methods **3**
- 2.1 Molecular Hamiltonian and the Stationary Schrödinger Equation 3
- 2.2 Density Functional Theory 6
  - 2.2.1 Practical Implementations of DFT 7
    - Various Approximations to the Exchange and Correlation Functional . . 7
    - The Pseudopotentials and Plane-Wave Scheme 12
- 2.3 Molecular Dynamics 13
  - 2.3.1 Classical Molecular Dynamics 14
  - 2.3.2 *Ab Initio* Molecular Dynamics 15
    - Car-Parrinello Molecular Dynamics 15
    - Born-Oppenheimer Molecular Dynamics 16
  - 2.3.3 Molecular Dynamics in the Canonical Ensemble 16
- 2.4 Activated Processes 18
  - 2.4.1 Reaction Coordinates, Order Parameters and Collective Variables 19
  - 2.4.2 Techniques for the Study of Activated Processes 23
    - Nudged Elastic Band 23
    - Umbrella Sampling and the Weighted Histogram Analysis Method 24
    - Metadynamics 25

3 Advances in Method Development **29**
- 3.1 Protocol for Studying Finite Temperature Chemical Reactions in Gas Phase . . 29
- 3.2 Structural Clustering of Chemical Reactions in Solutions 30
  - 3.2.1 The Permutation Invariant Vector 31

3.2.2 Test Case: Na^+-Cl^- Ion Pair Dissociation in Solution 33
3.2.3 Conclusion 37

4 Applications 39
4.1 Gas Phase Reaction of Carbon Dioxide and Water: the Formation of Carbonic Acid 41
4.1.1 Results from the Potential Energy Surface 41
4.1.2 Reaction Dynamics and Free Energy Surfaces 45
The Crucial Choice of Collective Variables and its Validation 45
Convergence of the Free Energy Surfaces 46
Free Energy Landscapes and Comparison with the Potential Energy Surface 47
4.1.3 Conclusions 53
4.2 Carbon Dioxide - Monoethanolamine Reactions in Liquid Water 54
4.2.1 Conclusion 58
4.3 Impurity Diffusion in Cadmium Telluride 58
4.3.1 Relevant Impurity Configurations 59
4.3.2 Sulfur and Tellurium Migration from Metadynamics 62
4.3.3 *Post*-Metadynamics NEB calculations 63
4.3.4 Conclusions 64

A Computational Details 67
A.1 Sodium Chloride in Solution 67
A.2 Carbonic Acid Formation 67
A.3 Monoethanolamine and Carbon Dioxide 68
A.4 Sulfur Diffusion in Cadmium Telluride 69

B Additional Information: Sulfur diffusion in Cadmium Telluride 71

Bibliography 73

Curriculum Vitae 87

List of Figures

2.1 Perdew's ladder of xc functionals . . . 8
2.2 Energy profile of a reaction . . . 18
2.3 Principle of the nudge elastic band method . . . 24
2.4 Principle of metadynamics . . . 26

3.1 Definition of the permutation invariant vector (PIV) . . . 32
3.2 PIV - Clustering Algorithms . . . 32
3.3 PIV - two algorithms for the centers . . . 33
3.4 NaCl - free energy surface . . . 34
3.5 Different coordination functions . . . 34
3.6 NaCl - distribution of $\mathcal{D}_{AB}$ for various PIV . . . 35
3.7 NaCl - distribution of the clusters on the FES . . . 36
3.8 NaCl - committor probability of the clusters . . . 37

4.1 H_2CO_3 - structures of R_n, TS_n and P_n . . . 42
4.2 H_2CO_3 - committor analysis . . . 46
4.3 H_2CO_3 - parallel growth of free energy profiles . . . 47
4.4 H_2CO_3 - distribution of distances (R_3 and R'_3) . . . 48
4.5 H_2CO_3 - cluster centers $n = 3$ . . . 48
4.6 H_2CO_3 - free energy surfaces and dynamics . . . 50
4.7 H_2CO_3 - traditional reaction coordinates . . . 51
4.8 H_2CO_3 - temperature dependance of the entropic contributions . . . 53
4.9 MEA - zwitterion in solution . . . 55
4.10 MEA - free energy profiles . . . 55
4.11 MEA - water cage . . . 56
4.12 MEA - orientation of H_2O . . . 56
4.13 MEA - cluster analysis . . . 57
4.14 CdTe - supercell . . . 59
4.15 High symmetry sites in zinceblende . . . 60
4.16 CdTe:S - S-Te dumbbell (A) . . . 60
4.17 CdTe:S (A) - Kohn-Sham state . . . 61
4.18 CdTe:2S - two different structures . . . 61
4.19 CdTe:S - S trapping (B) . . . 62

4.20 CdTe:S- (B) - Kohn-Sham state 62
4.21 CdTe:S - FES ($A \rightarrow B$) 63
4.22 CdTe:S - dynamical ensemble (A) 63
4.23 CdTe:S - FES (S diffusion) 64
4.24 CdTe:S - mechanism (S diffusion) 64
4.25 CdTe:S - transition state (S diffusion) 65

B.1 CdTe - lattice constant and bulk modulus 72
B.2 CdTe - vibrational density of states 72

List of Tables

4.1 H_2CO_3 - structure deviation from MP2 43
4.2 H_2CO_3 - binding energies 43
4.3 H_2CO_3 - energy barrier heights 44
4.4 H_2CO_3 - free energy barrier heights 49
4.5 H_2CO_3 - enthalpy barrier heights 52
4.6 H_2CO_3 - entropic constribution to the barriers 52

1 Introduction

In this thesis I study activated physico-chemical processes using state-of-the-art atomistic computer simulations with a twofold aim: obtaining new insights into particularly challenging issues, and investigating the power and limitations of the available methodologies.

A specific introduction to the applications is postponed to Chapter 4. The topics we tackle are especially important in the current search for new energy sources: chemical reactions of carbon dioxide in gas phases and aqueous solutions, and impurity diffusion in CdTe, a material of wide technological use, especially as photovoltaic material in solar cells. The challenging problems these systems present are indeed of general importance. For example, they concern the understanding of the role of a solvent in a chemical reaction from the microscopic viewpoint or the physical mechanisms that govern the diffusion of impurities and dopants in a real material.

A computational approach of the above processes at the atomistic level requires methods that ensure both an accurate description of interatomic forces and techniques allowing for a statistically sufficient sampling of the relevant regions of the configuration space. Concerning the former, bond breaking and formation in chemical reactions imply a quantum mechanical method. To date, density functional theory (DFT) [1, 2] is still the method of choice, when one deals with large-scale systems. Validation and verification of the specific functionals was carried out whenever possible. Simulations of the time evolution of atomistic models rely on molecular dynamics (MD) techniques [3, 4]. In 1985, the seminal paper by Car and Parrinello (CP) [5], opened the field of *ab initio* MD by evaluating under a unified scheme the motion of the nuclei and the electronic structure at the DFT level. Further implementations of the concepts and techniques introduced [6] have allowed to extend the reach of CP molecular dynamics to increasingly complex systems, from small atom clusters to the nanoscale, and to access the physics of complicated phenomena, with applications relevant to diverse disciplines. Nowadays this is certainly the method of choice for the study of the physics of real materials and of chemical reactions in condensed phases.

Regarding the sampling problem, very long MD simulations could, in principle, access all the relevant regions of the configuration space. This thorough exploration, however, is hampered by the uneven nature of the underlying free energy surface (FES). Indeed, configurations whose free energy differs from the local minimum by $\gg k_B T$ are practically unaccessible within the timescale of *ab initio* MD. Consequently, activated processes occurring in Nature on long timescales (even larger than 1 s), for instance chemical reactions, require the use of enhanced sampling techniques. Different schemes have been developed to address this issue. We will mainly employ a flexible and successful technique, metadynamics [7], that relies on a history-dependent bias added to the FES allowing to accelerate the crossing of free energy barriers as well as reconstructing the free energy landscape at the end of the calculation. Metadynamics requires the choice of a few reaction coordinates able to capture the important changes undergone by the system as it progresses along the reaction pathway. Determining good reaction coordinates is in itself a non-trivial task and the use of inappropriate ones can dramatically affect the results of the simulations. Great care has been taken in all the simulations presented here, as it will become clear from the presentation.

Through the research carried out in this thesis, new protocols and analysis tools have been developed for atomistic simulations in condensed matter phases and also new unprecedented understanding have been obtained into the physics and chemistry of the systems we have studied.

My work has benefited from the CPMD [8] code, which is especially flexible and efficient, as well as of the use of the IBM BlueGene systems, namely today's most powerful supercomputers.

Outline

In Chapter 2, we present a short reminder of the fundamentals of the theory on which computations are based (density-functional-theory based molecular dynamics (the Car-Parrinello method)) and of the current methods used to simulate activated processes.

In Chapter 3 we describe a general protocol to approach the study of chemical reactions in gas phase by mean of *ab initio* MD and introduce a new approach to extract representative structures from MD trajectories taking into account the permutation symmetry of identical atoms.

In Chapter 4 we discuss the applications mentioned above: the formation and dissociation of carbonic acid (H_2CO_3) in the gas phase, the early steps of the formation of a stable carbamate from CO_2 dissolved in amine solutions, and sulfur diffusion in cadmium telluride. Progress and insights from new methodological advancements will be there included.

Conclusions will be given separately for each research.

2 Fundamentals of Theory and Simulation Methods

This Chapter presents the theoretical methods employed in this thesis for the simulation of activated processes at the atomistic level. After briefly summarizing the foundations of density functional theory (DFT) and DFT-based (*ab initio*) molecular dynamics (MD), I will describe some of the algorithms currently used to enhance configuration sampling and thus extend the reach of *ab initio* MD by accelerating the occurrence of rare events.

2.1 Molecular Hamiltonian and the Stationary Schrödinger Equation

Let us consider a system of N atoms, namely N nuclei of coordinates $\{\boldsymbol{R}_I\}$ and n electrons of coordinates $\{\boldsymbol{r}_i\}$. The corresponding Hamiltonian can be written as:

$$\hat{\mathcal{H}} = \underbrace{-\sum_{i=1}^{n}\frac{\hbar^2}{2m_e}\nabla_i^2}_{\hat{\mathcal{T}}^{\mathrm{el}}} \underbrace{-\sum_{I=1}^{N}\frac{\hbar^2}{2M_I}\nabla_I^2}_{\hat{\mathcal{T}}^{\mathrm{nucl}}} \underbrace{+\frac{1}{2}\sum_{i,j}^{n}\frac{e^2}{|\boldsymbol{r}_i-\boldsymbol{r}_j|}}_{\hat{\mathcal{V}}^{\mathrm{int}}} \underbrace{-\sum_{i,I}^{n,N}\frac{Z_I e^2}{|\boldsymbol{r}_i-\boldsymbol{R}_I|}}_{\hat{\mathcal{V}}^{\mathrm{ext}}} \underbrace{+\frac{1}{2}\sum_{I,J}^{N}\frac{Z_I Z_J e^2}{|\boldsymbol{R}_I-\boldsymbol{R}_J|}}_{\hat{\mathcal{U}}}. \tag{2.1}$$

In this thesis we will be mainly concerned with solving the time-independent Schrödinger equation:

$$\hat{\mathcal{H}}\Phi = \mathcal{E}\Phi, \tag{2.2}$$

where $\Phi = \Phi(\{\boldsymbol{r}\}, \{\boldsymbol{R}\})$ is the wavefunction describing the electron-nuclei system and $\mathcal{E}$, its the energy. This problem does not have an analytical solution for anything but trivial systems.

The Born-Oppenheimer (BO) approximation decreases the complexity of Eq. 2.2 by decoupling nuclear and electronic degrees of freedom. This is achieved by introducing the following ansatz for the wavefunction:

$$\Phi\left(\{\boldsymbol{r}\},\{\boldsymbol{R}\}\right)=\sum_{\mu=0}^{\infty}\varphi_{\mu}\left(\{\boldsymbol{R}\}\right)\cdot\Psi_{\mu}\left(\{\boldsymbol{r}\};\{\boldsymbol{R}\}\right), \tag{2.3}$$

where the φ_μ and Ψ_μ are the nuclear and electron wavefunctions respectively. The latter satisfy orthonormality conditions

$$\int \mathrm{d}^n\boldsymbol{r}\,\Psi_\nu^*\Psi_\mu=\delta_{\mu\nu} \tag{2.4}$$

and depend parametrically on $\{\boldsymbol{R}\}$.

One can demonstrate [9] that the ansatz (Eq. 2.3) reduces Eq. 2.2 with the Hamiltonian (Eq. 2.1) to the following equations:

$$\mathcal{E}\varphi_\nu=\left(\hat{\mathcal{T}}^{\mathrm{nucl}}+E_\nu+\hat{\mathcal{U}}\right)\varphi_\nu-\sum_\mu A_{\nu\mu}\varphi_\mu, \tag{2.5}$$

where $E_\nu(\boldsymbol{R})$ is the eigenvalue of the Schrödinger equation for the electronic system:

$$\underbrace{\left(\hat{\mathcal{T}}^{\mathrm{el}}+\hat{\mathcal{V}}^{\mathrm{int}}+\hat{\mathcal{V}}^{\mathrm{ext}}\right)}_{\hat{\mathcal{H}}^{\mathrm{el}}(\{\boldsymbol{r}\};\{\boldsymbol{R}\})}\Psi_\nu=E_\nu(\{\boldsymbol{R}\})\Psi_\nu. \tag{2.6}$$

In Eq. 2.5, $A_{\nu\mu}$ couples the electronic and nuclear variables

$$A_{\nu\mu}\varphi_\mu=\sum_I\frac{\hbar^2}{2M_I}\int \mathrm{d}^n\boldsymbol{r}\,\Psi_\nu^*\left(\varphi_\mu\nabla_I^2\Psi_\mu+2\nabla_I\varphi_\mu\cdot\nabla_I\Psi_\mu\right). \tag{2.7}$$

On the basis of the different scale of the nuclear and electronic masses

$$\frac{m_e}{M_I}\ll 1, \tag{2.8}$$

one can demonstrate that $A_{\nu\mu}$ can be neglected for most purposes. This is the BO approximation.

Therefore Eq. 2.5 reduces to an equation of the nuclear wavefunctions:

$$\mathcal{E}\varphi_\nu(\{\boldsymbol{R}\})=\left(\hat{\mathcal{T}}^{\mathrm{nucl}}+\hat{\mathcal{U}}+E_\nu(\{\boldsymbol{R}\})\right)\varphi_\nu(\{\boldsymbol{R}\}). \tag{2.9}$$

The nuclei obey a Schrödinger equation with the potential term containing $E_\nu(\{\boldsymbol{R}\})$, *i.e.*, the electrons provide the potential energy surface onto which the nuclei evolve. We will consider simulations in which the electrons will be in their ground state ($\Psi_0(\{\boldsymbol{r}\},\{\boldsymbol{R}\})$).

In Hartree-Fock theory (HF), the electronic Schrödinger equation (Eq. 2.6) is solved by approximating the many electron wavefunction Ψ by a Slater determinant of n single electron orbitals $\psi_i(\boldsymbol{r}_j,\sigma_j)$. For a singlet state (total spin = 0) where opposite spins are paired, we can thus write:

$$\Psi(\{\boldsymbol{r}\}) = \frac{1}{\sqrt{n!}}\begin{vmatrix} \psi_1(\boldsymbol{r}_1) & \cdots & \psi_n(\boldsymbol{r}_1) \\ \vdots & \ddots & \vdots \\ \psi_1(\boldsymbol{r}_n) & \cdots & \psi_n(\boldsymbol{r}_n) \end{vmatrix}, \tag{2.10}$$

where we have omitted the spin variables.
Ψ is antisymmetric under permutation of electrons by construction in accordance with Pauli's exclusion principle. No other correlations are included and the electrons are independent. By plugging the ansatz (Eq. 2.10) into the electronic Schrödinger equation (Eq. 2.6) and minimizing the expectation value of the electronic Hamiltonian with respect to the variation of the single-particle wavefunctions,

$$0 = \frac{\delta\left(\langle\Psi|\hat{\mathcal{H}}^{\mathrm{el}}|\Psi\rangle - \sum_j \Lambda_{jk}(\langle\psi_j|\,\psi_k\rangle - \delta_{ij})\right)}{\delta\psi_i} \tag{2.11}$$

under orthonormalization constraints, we can show that the ψ_i obey

$$\hat{\mathcal{F}}_i\psi_i = \epsilon_i\psi_i, \tag{2.12}$$

where the ψ_i were rotated so that $\Lambda_{ij} = \epsilon_i\delta_{ij}$, and

$$\hat{\mathcal{F}}_i = \hat{h}_i + \sum_j(\hat{J}_j - \hat{K}_j) \tag{2.13}$$

is the Fock operator with

$$\hat{h}_i = -\frac{\hbar^2}{2m_e}\nabla_i + \sum_I \frac{Z_I e^2}{|\boldsymbol{r}_i - \boldsymbol{R}_I|}; \tag{2.14}$$

$$\hat{J}_j\psi_i(\boldsymbol{r}) = e^2\int \mathrm{d}\boldsymbol{r}'\frac{|\psi_j(\boldsymbol{r}')|^2}{|\boldsymbol{r}-\boldsymbol{r}'|}\psi_i(\boldsymbol{r}); \tag{2.15}$$

$$\hat{K}_j\psi_i(\boldsymbol{r}) = e^2\int \mathrm{d}\boldsymbol{r}'\frac{\psi_j^*(\boldsymbol{r}')\psi_i(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}\psi_j(\boldsymbol{r}). \tag{2.16}$$

The $\hat{J}_j\psi_i$ term (Eq. 2.15) is the Hartree potential which corresponds to the classical Coulomb interaction of a charge distribution $|\psi_j|^2$, and $\hat{K}_j$ is the exchange operator. This has no classical counterpart. It arises only from the antisymmetric character of the electronic wavefunction. In particular the self interaction in Eq. 2.15 is exactly compensated in Eq. 2.16. Hartree-Fock is the starting point of more advanced approximation to the solution of the multi-electrons Schrödinger equation, which include electron-electron correlation. For example, the Møller-Plesset theory which is a second order perturbation theory or the coupled cluster methods that use a combination of Slater determinants as basis set.

2.2 Density Functional Theory

Density functional theory (DFT) is a radically different approach to post Hartree-Fock methods because it focuses on the electronic density

$$\rho(\boldsymbol{r}) = n\int \mathrm{d}\boldsymbol{r}_2\ldots\mathrm{d}\boldsymbol{r}_n|\Psi_0(\boldsymbol{r},\boldsymbol{r}_2,\ldots,\boldsymbol{r}_n)|^2, \quad n = \int \mathrm{d}\boldsymbol{r}\rho(\boldsymbol{r}) \tag{2.17}$$

rather than on the electronic wavefunction. A major advantage is that the number of degrees of freedom is reduced from $3n$ to 3.
Two fundamentals theorems demonstrated by Hohenberg and Kohn (HK) [1] lay the foundations of density functional theory[1]:

HK1 For any system of interacting electrons, the external potential $v^{\text{ext}}(\boldsymbol{r})$ is determined uniquely, except for a constant, by the ground state electronic density $\rho(\boldsymbol{r})$.
As a corollary, the Hamiltonian is fully determined, except for a constant shift of the energy. It follows that the many-electron wavefunction for all states (ground and excited) are determined. Therefore, all properties of the system are completely determined given only the ground state density $\rho(\boldsymbol{r})$

HK2 A universal functional

$$\tilde{E}[\rho] = T[\rho] + E^{\mathrm{H}}[\rho] + E^{\mathrm{XC}}[\rho] \tag{2.18}$$

of the density $\rho(\boldsymbol{r})$ can be defined. $T[\rho]$ is the kinetic energy functional and $E^{\mathrm{H}}[\rho]$ is the Hartree functional. The exchange and correlation (xc) functional, $E^{\mathrm{XC}}[\rho]$ includes the exchange contribution and correlation among the electrons. For any v^{ext}, the energy functional can be written as:

$$E[\rho] = \tilde{E}[\rho] + V^{\text{ext}}[\rho], \text{ where } V^{\text{ext}}[\rho] = \int \mathrm{d}\boldsymbol{r}\, v^{\text{ext}}(\boldsymbol{r})\rho(\boldsymbol{r}). \tag{2.19}$$

The density minimizing the functional $E[\rho]$ is density of the ground state. The corresponding value of $E[\rho]$ is the ground state energy.

The many-electron problem is generally solved by employing an ansatz originally formulated by Kohn and Sham (KS) [2]. They assumed the existence of an auxiliary fictitious system, in which the electrons are non interacting by definition, and which has the same energy density as the system of interacting electrons.
The many-electron wavefunction can be written as a Slater determinant (Eq. 2.10) of the one particle Kohn-Sham wavefunctions ψ_i, and $E^H[\rho]$ and $V^{\text{ext}}[\rho]$ are easily expressed in terms of the electronic density

$$\rho(\boldsymbol{r}) = \sum_{i=1}^{n} |\psi_i(\boldsymbol{r})|^2. \tag{2.20}$$

---

[1] The formulation hereafter is taken almost verbatim from Ref.[9]

The KS kinetic energy functional is

$$T_{\mathrm{KS}}[\rho] = -\sum_{i=1}^{n} \langle\psi_i| \frac{\hbar^2\nabla_i}{2m_e} |\psi_i\rangle . \tag{2.21}$$

Following, HK2, we can carry out the following minimization procedure

$$\frac{\delta\left(E_{\mathrm{KS}}[\rho] - \sum_{j,k}\Lambda_{jk}\left(\langle\psi_j|\,\psi_k\rangle - \delta_{jk}\right)\right)}{\delta\psi_i} = 0 \tag{2.22}$$

and obtain the Kohn Sham equations

$$\left(-\frac{\hbar^2}{2m_e}\nabla^2 + v_{\mathrm{KS}}^{\mathrm{eff}}(\boldsymbol{r})\right)\psi_i(\boldsymbol{r}) = \epsilon_i\psi_i(\boldsymbol{r}), \tag{2.23}$$

where we defined

$$v_{\mathrm{KS}}^{\mathrm{eff}}(\boldsymbol{r}) = e^2\int \mathrm{d}\boldsymbol{r}'\,\frac{\rho(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}\underbrace{-e^2\sum_I\frac{Z_I}{|\boldsymbol{r}-\boldsymbol{R}_I|}}_{v^{\mathrm{ext}}(\boldsymbol{r})}+\underbrace{\frac{\delta E_{\mathrm{KS}}^{\mathrm{XC}}[\rho]}{\delta\rho(\boldsymbol{r})}}_{v_{\mathrm{KS}}^{\mathrm{XC}}(\boldsymbol{r})} \tag{2.24}$$

We carried out a rotation of the orbitals so that $\Lambda_{ij} = \epsilon_i\delta_{ij}$. Solving Eq. 2.23 self-consistently will yield the ground state density and energy which are the same than the ones of the physical system.

The Kohn-Sham exchange and correlation functional $E_{\mathrm{KS}}^{\mathrm{XC}}[\rho]$ and its potential $v_{\mathrm{KS}}^{\mathrm{XC}}(\boldsymbol{r})$ are unknown. Therefore one has to rely on an approximation. This is the problem one encounters in applying DFT.

2.2.1 Practical Implementations of DFT

Various Approximations to the Exchange and Correlation Functional

Perdew has proposed to classify the approximations to the exchange-correlation functional of DFT on a ladder (Perdew's ladder; Figure 2.1) ranging from local-spin density to "exact exchange and partial exact correlation" [10].

We shall recall some of these approximations that will be used in this thesis. For the sake of simplicity, we will neglect the dependence on spin variables.

- **Local density approximation (LDA)**
  The first rung is the local density approximation, in which the xc functional is a local functional of the electron density $\rho(\boldsymbol{r})$:

  $$E_{\mathrm{LDA}}^{\mathrm{XC}}[\rho] = \int \mathrm{d}\boldsymbol{r}\,\rho(\boldsymbol{r})\epsilon_{\mathrm{HEG}}^{\mathrm{XC}}(\rho(\boldsymbol{r})), \tag{2.25}$$

  where the xc energy density $\epsilon_{\mathrm{HEG}}^{\mathrm{XC}}$ is a local function of $\rho(\boldsymbol{r})$. At each $\rho(\boldsymbol{r})$, this function is

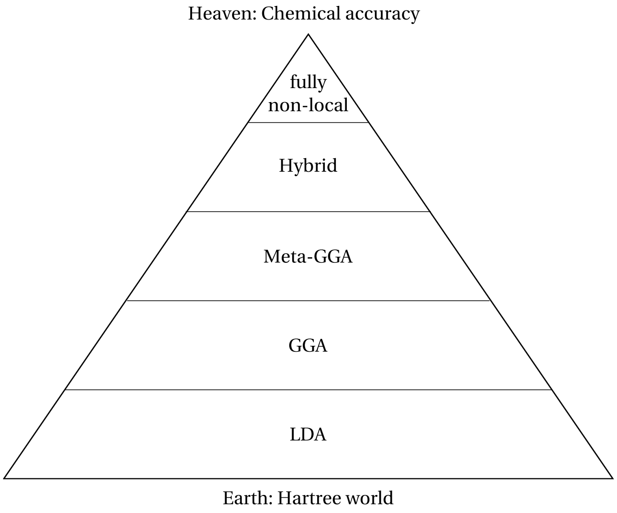


Figure 2.1 – Perdew's ladder of xc functionals; adapted from [10].

that of the homogeneous electron gas (HEG). Therefore, for the exchange term [11] :

$$\epsilon^{\mathrm{X}}_{\mathrm{HEG}}(\rho) = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3}\rho^{1/3} \tag{2.26}$$

and for the correlation term:

- in the low density limit ($r_s \to \infty$) [12]

$$\epsilon^{\mathrm{C}}_{\mathrm{HEG}}(\rho) \overset{r_s \to \infty}{\to} \frac{a_1}{r_s} + \frac{a_2}{r_s^{3/2}} + \frac{a_3}{r_s^2} + \dots \tag{2.27}$$

- in the high density limit ($r_s \to 0$) [13]

$$\epsilon^{\mathrm{C}}_{\mathrm{HEG}}(\rho) \overset{r_s \to 0}{\to} b_1 \log r_s + b_2 + r_s\left(b_3 \log r_s + b_4 r_s\right), \tag{2.28}$$

where r_s is the Wigner-Seitz radius

$$\frac{4}{3}\pi r_s^3 = \rho^{-1}. \tag{2.29}$$

Typical densities of interest for DFT calculations are in the range $0.8 < r_s < 10$. LDA calculations uses interpolation formulas of Monte Carlo computations of correlation energies of the HEG over a wide range of densities [14, 15, 16, 17].

By construction, LDA works well in systems with a slowly varying density but indeed its range of validity goes well beyond it. This was explained on the basis of fundamental conditions that the exchange and correlation functionals must satisfy and that LDA does satisfy, being grounded on the HEG (*e.g.*, Ref. [18]).

- **Semi-local density functionals**

  The most natural extension of LDA consists in introducing partial nonlocality via a (truncated) Taylor expansion of the energy density (GEA, Gradient Expansion Approximation). The lowest order in this expansion (second order in the gradient), called GE2, was applied and shown to provide an overcorrection to LDA for realistic densities and thus worsen the results. (see, *e.g.*, Ref. [19, 20]).

$$E^{\mathrm{XC}}_{\mathrm{GGA}}[\rho] = \int \mathrm{d}\boldsymbol{r}\,\rho(\boldsymbol{r})\epsilon^{\mathrm{XC}}(\rho,\nabla\rho) \tag{2.30}$$

Starting from GEA, a new approach was introduced by Perdew and collaborators, in a series of papers [21, 22, 23, 24, 25] which they named Generalized Gradient Approximation (GGA). In the GGA, the energy density depends only on ρ and $\nabla\rho$ (see Eq. 2.30) ; it introduces cutoffs to correct off the long-range behavior of GEA and is constructed in such a way that several exact conditions on the exchange and correlation contributions be separately satisfied. Remarkably, this construction is from first-principles, like LDA, namely it does not include any empirical parameter. Moreover, GGA reduces to LDA for slowly varying density $\rho(\boldsymbol{r})$.

GGA significantly improves on LDA especially in the prediction of atomization energies, the cohesive energies of solids and in the description of inhomogeneous systems, like surfaces or adsorbates. In this thesis, we will apply one of the latest and most widely used formulations of GGA functionals by Perdew and coworkers (Perdew-Burke-Ernzerhof (PBE) [26]).

Other functionals have expressions consistent with GGA but are semi-empirical. We refer here to Becke's approximation for the exchange functional [27]. This is mainly used in conjunction with the Lee-Yang-Parr correlation [28] functional that has a different origin. This combination goes under the acronym BLYP. We will also present calculations using this approximation.

PBE

The original GGA functional was derived in numerical form and reduced to an analytical expression by Perdew and Wang [24]. A further simplification lead to the so called PBE functional [26].

For the correlation functional:

$$E^{\mathrm{C}}_{\mathrm{PBE}}[\rho] = \int \mathrm{d}\boldsymbol{r}\,\rho\left(\epsilon^{\mathrm{C}}_{\mathrm{HEG}}(\rho) + H^{\mathrm{C}}(\rho,t)\right), \quad t = \frac{\nabla\rho}{2k_s\rho}, \tag{2.31}$$

t is a dimensionless density gradient, $1/k_s$ is the Thomas-Fermi screening wavelength, and

$$H^{\mathrm{C}}(\rho,t) = \frac{a_2\,m_e}{\hbar^2}\log\left(1 + \frac{a_1}{a_2}t^2\left(\frac{1+At^2}{1+At^2+A^2t^4}\right)\right), \tag{2.32}$$

where

$$A = \frac{a_1}{a_2}\frac{1}{\exp(-\epsilon^{\mathrm{C}}_{\mathrm{HEG}}/(a_2\,m_e/\hbar^2)) - 1} \tag{2.33}$$

with $a_1 = 0.066725$ and $a_2 = (1-\log 2)/\pi^2$.

Similarly for the exchange we have:

$$E^{\mathrm{X}}_{\mathrm{PBE}}[\rho] = \int \mathrm{d}\boldsymbol{r}\,\rho(\boldsymbol{r})\epsilon^{\mathrm{X}}_{\mathrm{HEG}}(\rho)F^{\mathrm{X}}(s), \quad s = 2(3\pi^2/2)^{1/3}\sqrt{r_s}\,t, \tag{2.34}$$

s is a dimensionless gradient, and

$$F^{\mathrm{X}}(s) = 1 + b_1 - \frac{b_1}{1+\frac{b_2 s^2}{b_1}} \tag{2.35}$$

with $b_1 = 0.804$ and $b_2 = a_2\pi^2/3$.

BLYP

Becke proposed to correct the exchange density from LDA (Eq. 2.26)

$$E^{\mathrm{X}}_{\mathrm{B}}[\rho] = \int \mathrm{d}\boldsymbol{r}\,\rho\epsilon^{\mathrm{X}}_{\mathrm{HEG}}(\rho)\left(1+\frac{4}{3}\left(\frac{\pi}{3}\right)^{1/3} a_1\frac{x^2}{1+6a_1 x\sinh^{-1}x}\right), \quad x = \frac{|\nabla\rho|}{\rho^{4/3}}. \tag{2.36}$$

This expression has the advantage of being simple while respecting the known limit for the exchange energy density. a_1 is obtained via a least square fit on the exact Hartee-Fock exchange of noble gases ($a_1 = 0.0042$ a.u.)

Starting from the Hartree-Fock formulation and introducing Jastrow factors to describe correlation effects, Colle and Salvetti derived a simplified expression for the energy due to dynamical correlation, depending on one parameter only which was fitted to helium [29]. The well-known LYP correlation functional $E^{\mathrm{C}}_{\mathrm{LYP}}$ was obtained from a simplification and further reformulation of the Colle-Salvetti formula in terms of ρ and $\nabla\rho$ only.

$$E^{\mathrm{C}}_{\mathrm{LYP}}[\rho] = -\int \mathrm{d}\boldsymbol{r}\frac{b_1}{1+b_4\rho^{-1/3}}\left(\rho + b_2\rho^{-2/3}\left(C_F\rho^{5/3} - 2t_W + \frac{t_W}{9} + \frac{\nabla^2\rho}{18}\right)\mathrm{e}^{-b_3\rho^{-1/3}}\right), \tag{2.37}$$

where

$$C_F = \frac{3}{10}(3\pi)^{2/3} \quad \text{and} \quad t_W = \frac{1}{8}\left(\frac{|\nabla\rho|^2}{\rho} - \nabla^2\rho\right). \tag{2.38}$$

t_W is the kinetic energy density, and $b_1 = 0.04918$, $b_2 = 0.132$, $b_3 = 0.2533$ and $b_4 = 0.349$.

- **Beyond GGA**
  On the third rung of Perdew's ladder we find meta-GGA functionals (*e.g.*, Ref. [30, 31, 32, 33]) that are explicitly dependent on the Kohn-Sham kinetic energy (Eq. 2.21).

The more advanced rung includes the so-called hybrid functionals, that combine GGA expressions with the exact Hartree-Fock exchange (Eq. 2.16) with weight generally empirically adjusted to, *e.g.*, molecular properties.
In this thesis we will present calculations using both a hybrid functional born from PBE (PBE0 [34]) (Eq. 2.39) and one from BLYP (B3LYP [35]) (Eq. 2.40).

$$E^{\mathrm{XC}}_{\mathrm{PBE0}} = E^{\mathrm{XC}}_{\mathrm{PBE}} + a_0 \left(E^{\mathrm{X}}_{\mathrm{HF}} - E^{\mathrm{X}}_{\mathrm{PBE}}\right) \tag{2.39}$$

In PBE0, the only adjustable parameter $a_0 = 0.25$ is not empirical but derived from Perdew's theoretical arguments in Ref. [36].

The B3LYP functional is written as:

$$E^{\mathrm{XC}}_{\mathrm{B3LYP}} = (1-a_0)E^{\mathrm{X}}_{\mathrm{LDA}} + a_0 E^{\mathrm{X}}_{\mathrm{HF}} + a_x \Delta E^{\mathrm{X}}_{\mathrm{B}} + a_c E^{\mathrm{C}}_{\mathrm{LYP}} + (1-a_c)E^{\mathrm{C}}_{\mathrm{VMN}}, \tag{2.40}$$

where $E^{\mathrm{C}}_{\mathrm{VMN}}$ is the Vosko-Wilk-Nusair expression for the correlation energy of the HEG [15] and $a_0 = 0.20$, $a_x = 0.72$, $a_c = 0.81$ [37].

The accuracy and shortcomings of all xc functionals presented here are well documented (see [38, 9]) For example, neither LDA nor GGA cancel the electron self-interaction coming from the Hartree term in Eq. 2.15 and obviously hybrid functional do only partially. Semi-empirical corrections are often employed, *e.g.*, Ref. [16, 39, 40, 41].
Extensions of GGA functionals and also new strategies for the construction of the DFT functionals continue to be developed. Here we mention only one: the series of functionals due to Truhlar and his collaborators [42] born from a combination of the meta-GGA and Hartree-Fock exchange. Different parametrizations characterize the different classes of functionals, which are each optimized to represent certain classes of systems or physical quantities. We will use one of them (Ref. [43]) as reference in one of the applications presented in this thesis.
Another problem concerns dispersion interactions which are inherently non-local and thus not accounted for in GGA.

Grimme's corrections to GGA functionals
Simple semi-empirical corrections to GGA were introduced by Grimme in a series of papers [44, 45, 46] via the addition of a long-range van-der-Waals-like attractive potential E_{disp} to the Kohn-Sham total energy :

$$E_{\mathrm{disp}} = -\frac{s_6}{2}\sum_{I,J}^{N} \frac{C_6^{IJ}}{d_{IJ}^6} f_{\mathrm{damp}}(d_{IJ}), \quad d_{IJ} = |\boldsymbol{R}_I - \boldsymbol{R}_J|. \tag{2.41}$$

In this thesis, we will employ the so called D2 version [45], in which the dispersion coefficients were calculated as

$$C_6^{IJ} = \sqrt{C_6^I C_6^J}, \tag{2.42}$$

where the single atom coefficients C_6^I are computed from the ionization potential and polarizability.

The s_6 parameter in Eq. 2.41 is a scaling constant that is functional-dependent. It is obtained via fitting to a set of binding energies and intramolecular distances of non-covalently bound complexes obtained with high-level quantum chemistry methods. The damping function f_{damp} is required to make vanish E_{disp} at short interatomic distances. More recently Grimme *et al.* [46] proposed a new approximation (D3) in which the effect of the local environement is accounted for. Moreover, the level of empiricism in the values of the dispersion coefficients is reduced by deriving them for time-dependent density functional theory calculations. In this way, the accuracy was shown to improve on a few selected systems, *e.g.*, S22 set [47].

The Pseudopotentials and Plane-Wave Scheme

The calculations presented in this thesis use the pseudopotential (PP) method, which reduces the electronic problem to the valence subspace, and plane-waves as basis set for the expansion of the valence wavefunctions.

The basic assumption from which the pseudopotential method develops is the frozen-core approximation, in which one neglects the redistribution of the core electron density. The change of core wavefunctions caused by changes in the chemical environment, *e.g.*, processes like bond breaking or formation, can be safely neglected in most cases. The pseudopotential method, however, goes one step further, and provides a technique to solve exclusively for the valence wavefunctions. This is done by accounting for the contribution of the core electrons via an effective potential describing their interaction with the valence electrons. In this way, for each l subshell, the valence states are described by pseudo-wavefunctions, and the one with lowest energy is nodeless. The validity of a pseudopotential scheme is measured in terms of what is called "transferability", namely it should be able to represent the valence properties of a given atom in different chemical environments. The seminal paper by Hamann *et al.* [48] established the criteria on which one should construct such an accurate atomic pseudopotential from an all-electron (AE) calculation.

For each angular momentum l, not only

$$\epsilon_l^{\text{PP}} = \epsilon_l^{\text{AE}} \tag{2.43}$$

but also the radial pseudo-wavefunction should satisfy the following conditions, as close as possible

$$\psi_l^{\text{PP}}(r) = \psi_l^{\text{AE}}(r), \quad \text{for } r > r_c, \tag{2.44}$$

where ψ_l^{AE} is the result of an all-electron calculation and

$$\int_0^{r_c} \mathrm{d}r\, r^2 |\psi_l^{\text{AE}}|^2 = \int_0^{r_c} \mathrm{d}r\, r^2 |\psi_l^{\text{PP}}|^2, \tag{2.45}$$

where r_c is a selected effective "core" radius, typically between the outmost node and the outmost maximum of $\psi_l^{\mathrm{AE}}(r)$ or $r\psi_l^{\mathrm{AE}}(r)$. (Eq. 2.45) is the norm conservation condition, which guarantees that the integrated probability density within the "core" region be preserved. Clearly the choice of r_c governs the degree of accuracy of the pseudopotential and at the same time its hardness.

Different techniques have been proposed for the derivation of pseupdopotentials; either in numerical or analytical form. We will make use of two most popular schemes: one introduced by Troullier and Martins (TM)[49], the other by Goedecker, Tetter and Hutter (GTH) [50, 51]. TM is numerical; its construction enforces also continuity of the pseudopotential up to the fourth derivative at r_c and flatness about $r = 0$. GTH is analytical, with an expression separable in real and reciprocal space.

Although not restricted to it, the introduction of a pseudopotential becomes necessary when plane waves are used as basis set for the expansion of the one-electron wavefunctions:

$$\psi_i(\boldsymbol{r}) = \sum_{\boldsymbol{g}} c_i(\boldsymbol{g}) \mathrm{e}^{i\boldsymbol{g}\cdot\boldsymbol{r}} \tag{2.46}$$

In practice a cutoff must be established, which depends on the hardness of the pseudopotential.
Plane-waves have several advantages, and in particular: (i) they are a complete set, which allows one to verify the convergence of the calculations in terms of the cutoff; (ii) the Hamiltonian matrix elements have often an analytical expression; (iii) the forces acting on the ions can also be expressed easily. Moreover (iv) they are extremely convenient in *ab initio* (Car-Parrinello) molecular dynamics, because the basis functions do not change when the atoms move.

2.3 Molecular Dynamics

The aim of the molecular dynamics (MD) technique is the simulation of the time evolution of an atomistic system obtained by numerically integrating the equation of motion (see, *e.g.*, [4] and reference therein). We will assume that the nuclei are classical objects subject only to the laws of classical mechanics. The difference between classical and *ab initio* MD is in the treatment of the electrons, which, in the latter case, are explicitly included in the calculation and participate in the shaping of the potential energy surface for the nuclei.
The basic idea behind MD is to approach an atomic system very much like it is done in a real experiment where the system is prepared in some state (certain temperature, pressure, etc.) and then properties are measured. In MD, a set of N atoms is selected, and their evolution is simulated, and with the help of statistical mechanics, the trajectory of these atoms (*i.e.*, the set of position and velocities) can be related to observables, *e.g.*, temperature, diffusion

coefficients, or free energy differences.

2.3.1 Classical Molecular Dynamics

Classical MD denotes a dynamics driven by classical forces on the atoms. The electrons are not explicitly considered and the atoms evolve on a potential energy surface given by an empirical force field. Classical MD is employed in a broad range of fields including simulations of materials, nanostructures, solutions and complex biomolecules.
The Newtonian equations of motion can be derived from the Lagrangian:

$$\mathcal{L}(\{\boldsymbol{R}\},\{\dot{\boldsymbol{R}}\}) = \sum_I \frac{M_I}{2}\dot{\boldsymbol{R}}_I^{\,2} - U(\{\boldsymbol{R}\}) \tag{2.47}$$

by means of the Euler-Lagrange equations:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathcal{L}}{\partial \dot{\boldsymbol{R}}_I} = \frac{\partial \mathcal{L}}{\partial \boldsymbol{R}_I} \tag{2.48}$$

which yield

$$M_I \ddot{\boldsymbol{R}}_I = -\nabla_I U. \tag{2.49}$$

The integration of the equations of motion starts from initial conditions $\{\boldsymbol{R}(0), \dot{\boldsymbol{R}}(0)\}$ with the velocity drawn, *e.g.*, from a Maxwellian distribution at a given temperature.
In practice, the trajectories $\{\boldsymbol{R}(t)\}$ and $\{\dot{\boldsymbol{R}}(t)\}$ are obtained by numerical integration, and the system evolution is computed at discrete times $\{\boldsymbol{R}(n\Delta t), \dot{\boldsymbol{R}}(n\Delta t)\}_{n\in\mathbb{N}}$.
The velocity-Verlet algorithm [52] is often employed to carry out this integration. Let us consider one particle of mass M in one dimension x for convenience. The position at the next step $x(t+\Delta t)$ is obtained from the current position $x(t)$ and from the position at the previous step $x(t-\Delta t)$:

$$x(t+\Delta t) = 2x(t) - x(t-\Delta t) + \frac{F(t)}{M}\Delta t^2 + \mathcal{O}(\Delta t^4), \quad F(t) = -\left.\frac{\mathrm{d}U(x)}{\mathrm{d}x}\right|_{x(t)}. \tag{2.50}$$

The velocities are not used to iterate Eq. 2.50; instead, they are computed separately:

$$v(t+\Delta t) = v(t) + \frac{F(t+\Delta t)+F(t)}{2M}\Delta t + \mathcal{O}(\Delta t^3). \tag{2.51}$$

We note that in order to compute the velocities at time t the forces at time $t+\Delta t$ are needed, therefore in practice the velocities at step t are computed only once the new positions have been computed from Eq. 2.50.
The velocity-Verlet algorithm is widely employed because it is rather simple and consumes only a limited amount of resources. Moreover, it can be derived from Liouville's description of classical mechanics [53]. Thus, it is area preserving in phase-space and, provided that the timestep Δt is small enough, it will only exhibit a moderate long-term energy drift. This is

a desirable feature because the forces in Eq. 2.49 derive from a potential and therefore the total mechanical energy must be conserved. This is important if one is interested in time dependent properties and not only in configurational averages.
The timestep Δt for integrating the equations of motion, has to be smaller than the period of the natural thermal motion of the particles; typically $\Delta t \sim 0.1 - 1$ fs.

2.3.2 *Ab Initio* Molecular Dynamics

Ab initio MD takes the electrons into account explicitly. Hence, this approach is suitable to study processes where electrons are expected to play a significant role, *e.g.*, during chemical reactions where covalent bonds are broken and formed.

Car-Parrinello Molecular Dynamics

The Car-Parrinello (CP) method [5] was born as the unification of DFT and molecular dynamics. The CP Lagrangian (Eq. 2.52) incorporates both electronic and ionic degrees of freedom, adopts the KS total energy as potential energy and simplifies the solution of the electronic problem via a fictitious kinetic energy of the electronic wavefunctions ψ_i.

$$\mathcal{L}(\{\boldsymbol{R}\},\{\dot{\boldsymbol{R}}\},\{\psi\},\{\dot{\psi}\}) = \sum_I \frac{1}{2} M_I \dot{\boldsymbol{R}}_I^2 + \sum_i \mu \langle\dot{\psi}_i|\dot{\psi}_i\rangle - E_{\mathrm{KS}}(\{\psi\};\{\boldsymbol{R}\}) + \sum_{i,j} \Lambda_{ij} \left(\langle\psi_i|\psi_j\rangle - \delta_{ij}\right) - U(\{\boldsymbol{R}\}). \tag{2.52}$$

The introduction of the Lagrange multiplier guarantees the orthonormalization of the wavefunctions.
The corresponding Euler-Lagrange equations read:

$$\begin{cases} M_I \ddot{R}_I & = -\nabla_I (E_{\mathrm{KS}} + U) \\ \mu \ddot{\psi}_i & = -\frac{\delta E_{\mathrm{KS}}}{\delta \psi_I^*} + \sum_j \Lambda_{ij} \psi_i \end{cases} \tag{2.53}$$

If the electrons are in the ground state, then the second equation amounts to the Kohn-Sham equation (Eq. 2.23).
μ is a key parameter, acting as a fictitious mass associated to the electronic variables and is meant to be adjusted to guarantee a BO dynamics, or in other words to prevent energy from flowing from the ionic to the electronic variables. If their respective vibrational density of states do not overlap the transfer of energy will be sufficiently slow and adiabatic separation will be achieved. It can be shown [54] that the lowest frequency of the vibrational spectrum of the ψ can be written as:

$$\omega_{\mathrm{min}}^{\mathrm{el}} \propto \sqrt{\frac{2E_{\mathrm{gap}}}{\mu}}, \tag{2.54}$$

where E_{gap} is the energy gap between the highest occupied electronic KS state and the lowest unoccupied KS state.
The bigger the difference between $\omega^{\text{el}}_{\text{min}}$ and the largest phonon frequency of the ions ($\omega^{\text{ion}}_{\text{max}}$) the better the adiabatic decoupling between electronic and ionic degrees of freedom. Hence the crucial role of the value one selects for μ. Indeed the whole electronic spectrum depends on $\mu^{-1/2}$. One more choice is however as important. As we have seen in Section 2.3.1, the timestep Δt for the integration of the equations of motion should always be smaller than the period of oscillation of the fastest degree of freedom. Therefore, in solving the CP equations of motion :

$$\Delta t \propto \sqrt{\mu}. \tag{2.55}$$

In its original, and most widely used, version, the CP Lagrangian was restricted to systems with a "sufficiently" large energy gap. However, a number of extensions have been made to allow application also to metallic systems [55, 56]. The simplest one makes use of a thermostat to maintain the kinetic energy at fixed values, namely to maintain the electronic variables "cold" [55].

Born-Oppenheimer Molecular Dynamics

The CP-MD method solves for the electronic and ionic variables on the same footing. One can, however, adopt another procedure, called BO-MD, which is conceptually simpler but (generally) less efficient in practice.
In the BO-MD one solves for the electronic and ionic variables in sequence, and the electrons are kept on the BO surface. The Newton equations for the ions can be written as

$$M_I \ddot{\boldsymbol{R}}_I = -\nabla_I (E_{\text{KS},0} + U) \tag{2.56}$$

where $E_{\text{KS},0}$ is the KS energy in the ground state.
At each step in the evolution of the ions, the electronic problem is solved (Eq. 2.6), thus providing $E_{\text{KS},0}$. The positions of the ions are then updated. In turn, this generates a new external potential for the electrons. The procedure is iterated.
The time step for the solution of the Newton equation can be taken larger than for the CP-MD, but this is not always an advantage, given that one must also solve the KS equation (Eq 2.23) at each step.

2.3.3 Molecular Dynamics in the Canonical Ensemble

Experiments are often carried out at constant temperature and fixed volume, *i.e.*, they sample the canonical ensemble. In both classical and *ab initio* molecular dynamics the total energy is conserved and therefore both schemes naturally sample the microcanonical ensemble, *i.e.*, the probability $P(\{\boldsymbol{R}\})$ of exploring a configuration $\{\boldsymbol{R}\}$ is proportional to

$$P(\{\boldsymbol{R}\}) \propto \delta\left(E - U(\{\boldsymbol{R}\})\right). \tag{2.57}$$

Where E is the (constant) energy of the simulation and $U(\{\boldsymbol{R}\})$ the energy of the configuration. Instead, in the canonical ensemble, the probability of a configuration is given by its Boltzmann weight:

$$P(\{\boldsymbol{R}\}) \propto \mathrm{e}^{-\beta U(\{\boldsymbol{R}\})}, \quad \beta = \frac{1}{k_\mathrm{B} T}. \tag{2.58}$$

Many thermostats exist that allow to extend MD to sample the configuration space at an average constant temperature. The most naive one is the velocity rescaling algorithm that rescales all the velocities in order to have

$$< E_\mathrm{kin} >= \frac{3N}{2} k_\mathrm{B} T \tag{2.59}$$

as soon as the average kinetic energy goes out of a specified region.
This simple approach however does not reproduce the fluctuations in temperature

$$\sigma_T^2 = \frac{2T}{3N} \tag{2.60}$$

which exist in the canonical ensemble since, in between rescaling steps, the sampled ensemble is microcanonical.
A less abrupt scheme was introduced by Berendsen *et al.* [57]: rather than having a sudden rescaling, they proposed to have the velocities scaled so that the average kinetic energy would exponentially decay or grow towards the desired temperature. This scheme although less drastic than velocity rescaling, is still unable to reproduce the fluctuations of the canonical ensemble.
Andersen [58] proposed a thermostat scheme that couples the system to a heat bath. A particle randomly colliding with it has its velocity reset to one drawn from a Maxwell-Boltzmann distribution at the desired temperature. Employing this thermostat will induce faster decorrelations.
Nosé-Hoover chain thermostat [59, 60, 61], contrary to Andersen's, is deterministic. It was shown to sample Boltzmann's distribution and to reproduce the fluctuations associated with the canonical ensemble correctly. It introduces additional degrees of freedom ζ in the equations of motion to ensure a sampling of the canonical distribution. The "mass" of these ζ is chosen so as to overlap the power spectra of the system with those of the ζ in order to maximize the exchange of energy between them [62].
In this thesis, we will employ this thermostat almost exclusively.

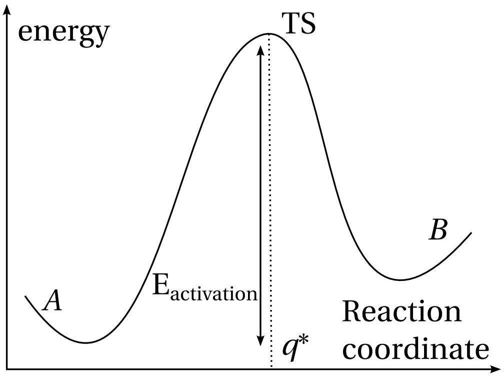


Figure 2.2 – A sketch of the energy profile of the process in Eq. 2.61. $E_{\text{activation}}$ is the activation energy and q^* indicates the coordinate of the transition state.

2.4 Activated Processes

Physico-chemical transition processes can be schematized as

$$A \leftrightarrows B. \tag{2.61}$$

Where A and B are distinct metastable states, *i.e.*, the system is almost always found in either A or B. The rate of transition from A to B is a particularly important characteristic of the reaction. It is related to the frequency at which the systems visits a high energy, scarcely populated state, located between A and B, the transition state (TS).

To determine the transition rate $k_{A\to B}$ we first consider the evolution of the population in the two states:

$$\dot{P}_A = -k_{A\to B}P_A + k_{B\to A}P_B \tag{2.62}$$
$$\dot{P}_B = +k_{A\to B}P_A - k_{B\to A}P_B. \tag{2.63}$$

The next step requires the introduction of the concept of reaction coordinates. They are coordinates that track the progress of the reaction as the system passes from A to B. In particular they allow to draw energy profiles like in Figure 2.2. On this sketch, TS has coordinate q^*.

A crucial assumption of transition state theory is that if the system crosses q^* to go from A to B, it will not recross it (from B to A) before it has reached equilibrium in B. Moreover, a transition from A to B happens when the system reaches q^* with $\dot{q} > 0$. Hence, the rate $k_{A\to B}$ is given by the ratio (similarly for $k_{B\to A}$)

$$k_{A\to B} = \frac{\langle \dot{q}\theta(\dot{q})\delta(q-q^*)\rangle}{\langle \theta(q-q^*)\rangle}. \tag{2.64}$$

The expectation value $<\theta(q-q^*)>$ counts the number of states in A. Eq. 2.64 can be derived from Equations 2.62 and 2.63 in the framework of linear response (see, *e.g.*, Chapter 16 of [4]). Explicitly calculating the averages in Eq. 2.64 in the canonical ensemble yields

$$k_{A\to B} \propto \mathrm{e}^{-\beta(E_{\mathrm{TS}}-E_A))} \tag{2.65}$$

and we define the activation energy $E_{\text{activation}} = E_{\text{TS}} - E_A$. Based on experimental data, van 't Hoff and Arrhenius arrived at the same expression already in the late 19th century. The proportionality prefactor in Eq. 2.65 is independent of the temperature.

Physico-chemical processes in Nature may be very slow because they require overcoming barriers with $E_A \gg k_{\text{B}}T$, *i.e.*, much larger than the typical size of the thermal energy fluctuations. Consequently, the timescale associated with these reactions can be several orders of magnitude higher than the timescale associated with the fastest motion of the atoms. For instance, O – H bond stretching in a water molecule has a period of about 10 fs while a chemical reaction such as the formation of carbonic acid requires about twenty minutes at 300 K [63].

Activated processes are a serious pitfall for molecular dynamics because, as we know, the timestep for integrating the equations of motion in MD has to be smaller than the timescale related to the fastest degrees of freedom. To date, the total simulation time available to classical MD ranges from nanoseconds to milliseconds while it is of the order of of 100 ps for *ab initio* MD. A whole class of processes is therefore completely unaccessible to standard molecular dynamics. Several techniques exist and that allow to overcome this intrinsic limitation and we detail a few of them in the upcoming paragraphs.

2.4.1 Reaction Coordinates, Order Parameters and Collective Variables

Molecular dynamics simulations of chemical reactions may involve an important number N of particles, especially if the reaction takes place in a solvent. The volume of the configuration space grows exponentially with the number of particles; it is hopeless to explore its entirety in a reasonable time, even more so if barriers are present.

As previously mentioned, a reaction coordinate q smoothly tracks the evolution of the reaction when passing from A to B. As we discussed, this concept is especially useful to represent the energy change along the reaction path and can be further used to identify the transition state (q^* in Figure 2.2).

In the canonical ensemble, the probability of observing the system in a configuration with a given value for q follows directly from the Boltzmann weight:

$$P(s) = \frac{1}{\mathcal{Q}} \int \mathrm{d}^N \boldsymbol{R} \mathrm{e}^{-\beta U(\boldsymbol{R})} \delta(q(\boldsymbol{R}) - s), \quad \mathcal{Q} = \int \mathrm{d}^N \boldsymbol{R} \mathrm{e}^{-\beta U(\boldsymbol{R})} a. \tag{2.66}$$

The free energy is a fundamental quantity and it assumes the minimum value at equilibrium. It can be defined as:

$$F(s) = -k_{\text{B}} T \log P(s) \Rightarrow P(s) \propto \mathrm{e}^{-\beta F(s)} \tag{2.67}$$

The free energy (F_Ω), enthalpy (H_Ω) and entropy (S_Ω) of a region Ω of the configuration phase space can be estimated by:

$$F_\Omega = -k_\mathrm{B} T \log \frac{\int_\Omega \mathrm{e}^{-\beta F} \mathrm{d}s}{\mathcal{Z}}, \ \mathcal{Z} = \int \mathrm{d}s \mathrm{e}^{-\beta F} \tag{2.68}$$

$$H_\Omega = \frac{\int_\Omega \mathrm{d}^N \boldsymbol{R} U(\{\boldsymbol{R}\}) \mathrm{e}^{-\beta U(\{\boldsymbol{R}\})}}{\int_\Omega \mathrm{d}^N \boldsymbol{R} \mathrm{e}^{-\beta U(\{\boldsymbol{R}\})}} \tag{2.69}$$

$$T S_\Omega = H_\Omega - F_\Omega \tag{2.70}$$

Actually, the thermodynamic definition of the enthalpy is

$$H = E + pV \tag{2.71}$$

with V the volume and p the pressure. In our simulations, *e.g.*, chemical reactions in vacuum or in dilute solution, the changes in pV term can typically be ignored and enthalpy and energy changes coincide. Often the partition function $\mathcal{Z}$ cannot be directly obtained. Nonetheless, it is usually not relevant as it amounts to fixing an absolute reference for the free energy, whereas the important quantities are the free energy difference between two regions Ω and Σ; in that case

$$\Delta F = F_\Omega - F_\Sigma = -k_\mathrm{B} T \log \frac{\int_\Omega \mathrm{e}^{-\beta F} \mathrm{d}s}{\int_\Sigma \mathrm{e}^{-\beta F} \mathrm{d}s} \tag{2.72}$$

and the computation of $\mathcal{Z}$ is not required.

The use of reaction coordinates appears very powerful to simplify the study of an activated process and extract properties from a molecular dynamics simulation. Until now however we did not specify how reaction coordinates can be found or constructed.
By employing functions of the atomic coordinates (collective variables (CVs)), it is possible to approximate a reaction coordinate. CVs can have a very simple functional form, *e.g.*, the interatomic distance but can also be much more complicated, *e.g.*, the path CVs [64] which are introduced in the following.

The choice of CV might be very delicate. To illustrate this issue we take the example of sodium chloride (NaCl) dissociation in liquid water [65]. The system has three local minima the contact ion pair (A), the solvent separated ion pair (B) and the dissociated ion pair (C). In a seminal work by Geissler *et al.* [65] it was shown that employing solely the distance between Na^+ and Cl^- was not sufficient to describe the nature of the transition of the transitions states between A and B. And that in fact, solvent related degrees of freedom had to be taken into account explicitly.
They also discuss a stringent test of the CV, namely committor analysis. It consists in starting many trajectories differing by the initial random velocities from all the configurations at q^*; if

q^* indeed is a transition state then, for each configuration, half of the trajectories must end up in A and the other half in B. If this is not satisfied, *i.e.*, if the majority of the configurations at q^* are solely committed to either A or B, then the CV employed to compute q^* are proven unable to correctly account for all relevant degrees of freedom participating in the reaction. The $\mathrm{Na}^+ - \mathrm{Cl}^-$ distance is sufficient to distinguish configurations in A from configuration in B, *i.e.*, it is a good *order parameter*. It is an inadequate reaction coordinate because it fails to identify the precise nature of the transition state and to successfully pass the committor test.

In the following paragraphs we present two kinds of CVs that we used when studying activated process.

Coordination Functions

This type of CV counts the number of contacts between two sets of atoms (Ω, Σ) based on the Cartesian distances ($d_{IJ} = |R_I - R_J|$) between their elements

$$\mathcal{CN}(\{\boldsymbol{R}\}) = \sum_{I\in\Omega, J\in\Sigma} cn\left(d_{IJ}\right). \tag{2.73}$$

Two definitions for the pairwise coordination function cn are often employed.

- The symmetric coordination (Fermi-Dirac like)

$$cn(x) = \frac{1}{1+\mathrm{e}^{\lambda(x-d_0)}}. \tag{2.74}$$

  d_0 is the critical distance before we can consider the atoms to be coordinated and after which they are not anymore, λ controls the steepness of the function.

- The asymmetric coordination

$$cn(x) = \frac{1-\left(\frac{x}{d_0}\right)^p}{1-\left(\frac{x}{d_0}\right)^{p+q}}, \quad p, q > 0. \tag{2.75}$$

  Similarly to the symmetric coordination, d_0 specifies the critical region. Two additional parameter (p and q) control the steepness of the head and the tail of the coordination function respectively (see Figure 3.5 for specific examples).

Path collective variables

The path CVs were introduced by Branduardi *et al.* [64] and are constructed from a predetermined path $\{\boldsymbol{R}_0\}$ from A to B

- the first function, s, measures the progress along the path:

$$s(\{\boldsymbol{R}\}) = \lim_{\lambda\to\infty} \frac{\int_0^1 \mathrm{d}t\, t \cdot \mathrm{e}^{-\lambda\mathcal{D}(\{\boldsymbol{R}_0(t)\},\{\boldsymbol{R}\})}}{\int_0^1 \mathrm{d}t\, \mathrm{e}^{-\lambda\mathcal{D}(\{\boldsymbol{R}_0(t)\},\{\boldsymbol{R}\})}} \tag{2.76}$$

and $\mathcal{D}$ is a measure of how much the configuration $\{\boldsymbol{R}\}$ differs from the reference structure $\{\boldsymbol{R}_0(t)\}$. The limit, $\lambda \to \infty$, is taken to ensure that the largest contribution to the integral comes from the structure in the reference path for which $\mathcal{D}$ is at a minimum, *i.e.*, the t' for which $\{\boldsymbol{R}_0(t')\}$ is most similar to $\{\boldsymbol{R}\}$.

- the second function, z, measures the distance away from the reference path:

$$z(\{\boldsymbol{R}\}) = -\lim_{\lambda\to\infty} \lambda^{-1} \log \int_0^1 \mathrm{d}t \mathrm{e}^{-\lambda\mathcal{D}(\{\boldsymbol{R}_0(t)\},\{\boldsymbol{R}\})}. \tag{2.77}$$

The measure $\mathcal{D}$ can be, for instance, the mean square difference between all pairs of interatomic distances computed in $\{\boldsymbol{R}^A\}$ and in $\{\boldsymbol{R}^B\}$

$$\mathcal{D}\left(\{\boldsymbol{R}^A\},\{\boldsymbol{R}^B\}\right) = \frac{2}{N(N-1)} \sum_{J>I}^{N} \left(|\boldsymbol{R}_I^A - \boldsymbol{R}_J^A| - |\boldsymbol{R}_I^B - \boldsymbol{R}_J^B|\right)^2. \tag{2.78}$$

It is important to note that $\mathcal{D}$ should contain information on all the relevant atoms, and therefore can be used to describe complicated cooperative processes. The z-coordinate is a measure of the distance in a direction orthogonal to the path. Therefore, if enough atoms are included in the path, the risk of not being able to capture an essential aspect of the reaction is reduced .

From the computational point of view, the continuous path $\{\boldsymbol{R}_0(t)\}$ has to be replaced by a finite set of $\mathcal{N}_0$ structures $\left[\{\boldsymbol{R}_0^{(k)}\}\right]_{k=1}^{\mathcal{N}_0}$ and the path CVs become

$$s(\{\boldsymbol{R}\}) = \frac{1}{\mathcal{Z}} \sum_{k=1}^{\mathcal{N}_0} k \cdot \mathrm{e}^{-\lambda\mathcal{D}\left(\{\boldsymbol{R}_0^{(k)}\},\{\boldsymbol{R}\}\right)} \tag{2.79}$$

with

$$\mathcal{Z} \equiv \mathcal{Z}(\{\boldsymbol{R}\}) = \sum_{k=1}^{\mathcal{N}_0} \mathrm{e}^{-\lambda\mathcal{D}\left(\{\boldsymbol{R}_0^{(k)}\},\{\boldsymbol{R}\}\right)} \tag{2.80}$$

and

$$z(\{\boldsymbol{R}\}) = -\lambda^{-1} \log \mathcal{Z}(\{\boldsymbol{R}\}). \tag{2.81}$$

The structures in the reference path should be chosen $\mathcal{D}$-equidistant and λ should be such that $\lambda\mathcal{D}(\{\boldsymbol{R}_0^{(k)}\},\{\boldsymbol{R}_0^{(k+1)}\}) \simeq 1$ in order to avoid a discontinuous evolution of the CVs.

In Section 3.2 we will extend this idea to construct a metric which remain invariant under permutation of identical atoms.

2.4.2 Techniques for the Study of Activated Processes

Nudged Elastic Band

If one is only interested in a description of activated processes based on the potential energy surface (PES) ($T = 0$), finding the transition state requires a good algorithm able to locate saddle points between reactant and product basins. The nudged elastic band (NEB) methods [66] aims at finding the minimum energy path (MEP) and transition state between known stationary points of the PES (reactants and products). NEB is particularly easy to employ because it does not rely on reaction coordinates, or rather the reaction coordinate is computed within the NEB framework from a set of $\mathcal{N}$ configurations ($\{\boldsymbol{R}^{(k)}\}_{k=1}^{\mathcal{N}}$):

$$q^{(k)} = q^{(k-1)} + \mathcal{D}\left(\boldsymbol{R}^{(k)}, \boldsymbol{R}^{(k-1)}\right), \quad q^{(1)} \overset{!}{=} 0 \tag{2.82}$$

and one does not have to guess it *a priori.* Traditionally the Euclidean distance is employed as metric

$$\mathcal{D}\left(\boldsymbol{R}^{(k)}, \boldsymbol{R}^{(k-1)}\right) = \sqrt{\frac{1}{3N}\sum_{J=1}^{N}(\boldsymbol{R}_J^{(k)} - \boldsymbol{R}_J^{(k-1)})^2}. \tag{2.83}$$

Furthermore, the location of the saddle point is found by computing only energy gradients and not the Hessian matrix as it is instead the case for other TS location algorithms (*e.g.*, the partitioned rational function optimizer (P-RFO) in [67]) reducing therefore the costs of computation if many atoms are included.

NEB works by optimizing a path (composed of $\mathcal{N}$ configurations with NEB-coordinates $q^{(1)}, \ldots, q^{(\mathcal{N})}$ along the NEB path (Eq. 2.82)) connecting A and B until all atoms have zero forces in the direction orthogonal to the path. This is achieved by computing effective forces on the atoms: the parts orthogonal to the path derive from the PES while instead, the parallel components are replaced by forces coming from springs linking the configurations together, and such that, at equilibrium, all configurations are equidistant along the path:

$$\boldsymbol{F}_J^{(k),\mathrm{NEB}} = \boldsymbol{F}_J^{(k),\perp} + \boldsymbol{F}_J^{(k),\mathrm{spring},\parallel}; \tag{2.84}$$

only the parallel components of the spring forces are considered to alleviate the problem of corner cutting.

Based on the NEB forces a steepest descent algorithm can be employed to find the MEP.

Generally the highest energy configuration, *i.e.*, the one closest to the saddle point, is set to not feel the force from the spring and the parallel component has simply an inverted sign to the one coming from the PES [68]

$$\boldsymbol{F}_J^{(\mathrm{CI})} = \boldsymbol{F}_J^{(\mathrm{CI}),\perp} - \boldsymbol{F}_J^{(\mathrm{CI}),\parallel}. \tag{2.85}$$

Therefore this configuration is allowed to climb up the energy gradient and is referred to as a *climbing image* (CI).

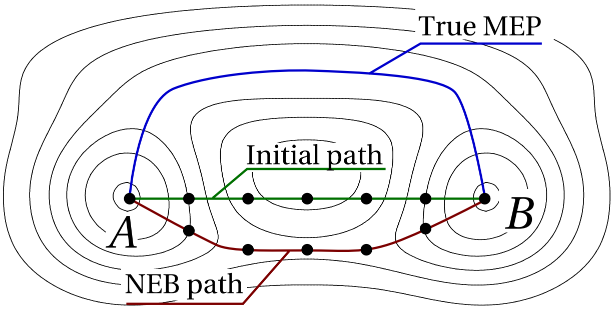


Figure 2.3 – NEB of the MEP converges to a path with higher barrier than the true MEP.

A limitation of the NEB method is that the relaxed configuration of both reactants and products have to be known beforehand; fortunately this is not so uncommon when studying physico-chemical processes with static methods.
Figure 2.3 illustrates a more serious shortcoming: the initial path (the green lines) between A and B consists in a few configurations taken from an interpolation between reactants and products. Carrying out a minimization procedure based on the forces in Equations 2.84 and 2.85 will converge to the red path which indeed goes through a saddle point; as such this might also be relevant although it is not the true MEP (the blue line).
An alternative formulation based on the reparametrization of the path and the redistribution of the configurations has also been proposed [69]. However, we will not employ it in this thesis.

Umbrella Sampling and the Weighted Histogram Analysis Method

Although NEB is a very powerful technique when employed with care, it cannot include finite temperature effects beyond the harmonic approximation and in particular it does not provide free energy barriers.
Umbrella sampling (US) [70] uses finite temperature MD to incorporate those effects and a biasing potential to overcome the sampling problem associated with the poor exploration of high free energy regions. Like many other enhanced sampling techniques it requires to define a reaction coordinate which is used to apply an *ad hoc* biasing potential $W(q)$. Clearly employing $W(q) = -F(q)$ would guarantee the most effective sampling of the reaction coordinate. However, $F(q)$ is not known beforehand and a practical strategy is to perform multiple simulations, in each one applying a bias potential tailored to focus the simulation in a given window of q. Typical form is

$$W_i(q) = \frac{1}{2}k(q-q_i)^2 \tag{2.86}$$

If windows overlap, the weighted histogram analysis method (WHAM) can be employed.
The probability in the i^{th} window is

$$P_i(s) = \frac{1}{\mathcal{Z}_i}\int \mathrm{d}^N\boldsymbol{R}\mathrm{e}^{-\beta(U(\boldsymbol{R})+W_i(q(\boldsymbol{R})))}\delta(q(\boldsymbol{R})-s) \tag{2.87}$$

The unbiased probability can be estimated as a linear combination of [71, 72]

$$P^{\mathrm{e}} = \sum_{i=1}^{n} \pi_i \mathrm{e}^{\beta W_i} \frac{\mathcal{Z}_i}{\mathcal{Z}} P_i, \tag{2.88}$$

where the weights π_i can be determined by minimizing the statistical uncertainty of P^{e} computed from a histogram distribution in each window assuming they follow a Poisson distribution. The derivation holds if the sampling of the distribution is performed at sufficiently large time intervals. The estimated probability is then

$$P^{\mathrm{e}} = \sum_i \alpha n_i P_i = \frac{\sum_i n_i P_i}{\sum_i \mathrm{e}^{-\beta W_i} n_i \mathcal{Z}/\mathcal{Z}_i}. \tag{2.89}$$

The ratio $\mathcal{Z}/\mathcal{Z}_i$ has to be computed self consistently.

Whenever the reaction coordinate is known to a good approximation, umbrella sampling is an extremely powerful technique whose results can in principle be systematically improved by collecting more statistics in the windows and make the estimate in Eq. 2.88 more accurate. Although, it is possible to include more than one reaction coordinate in theory, it becomes rapidly too expensive in practice because the number of bins grows exponentially with the number of reaction coordinates.

Metadynamics

Metadynamics (MTD) [7] relies on the same basic idea as umbrella sampling: it biases the potential along a CV to enhance the sampling of high free energy regions. However it differs significantly from US because first MTD does not employ multiple simulations in different windows and second the bias is constructed in a different way. In metadynamics, the bias is a history dependent quantity: as the system evolves, it deposits small hills of potential energy (traditionally Gaussian) in the CV-space discouraging the system to revisit the same region. At time t, the bias at q is

$$W(q;t) = \sum_{i=1}^{N_{\mathrm{hills}}(t)} h \mathrm{e}^{-\left(\frac{q-q(t_i)}{\sigma_i}\right)^2}, \quad t_i = i \cdot \tau, \tag{2.90}$$

where τ is the time between hill depositions.
Under the assumption that the CV is well suited to describe the reaction, the added bias will gradually compensate for the underlying well of the free energy landscape and drive the system to undergo transitions. Once the barrier has been crossed, the system relaxes to the next unvisited minimum and starts filling it as well. After some time t_{filling}, all the wells have been filled and the system should diffuse on a more or less flat landscape (Figure 2.4). In fact

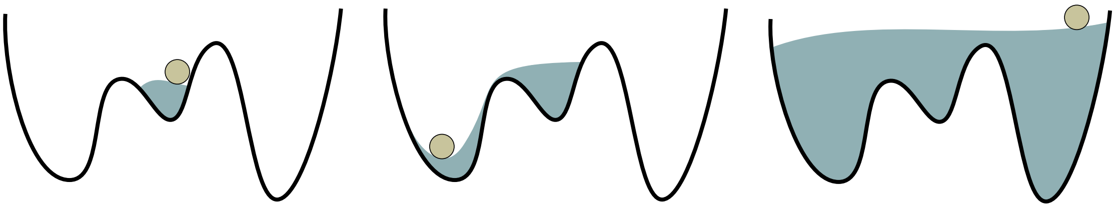


Figure 2.4 – Sketch of the principle behind metadynamics. The system deposits positive hills of potential energy as it moves in the free energy landscape. The sum of added hills is the blue filling. At the end the system only feels a flat potential and diffuses freely.

the central assumption of metadynamics is that

$$-\lim_{t\to\infty} W(q;t) \simeq F(q) + \text{const.} \tag{2.91}$$

It was first shown empirically correct by extensive testing on simple one dimensional systems with Langevin dynamics and the exact reaction coordinate [73]. Furthermore, the authors derived an expression for the instantaneous error ($\epsilon(t)$) on the reconstructed free energy surface (FES):

$$\epsilon(t) = \sqrt{\int \mathrm{d}q \left(W(q;t) - F(q) + C\right)^2} \propto \sqrt{\frac{h\cdot\tau}{D\beta}}, \tag{2.92}$$

where D is the diffusion coefficient associated with the system and C is an unphysical shift constant chosen to optimally align the profiles. The expression in Eq. 2.91 was formally proven true in Ref. [74] under the same assumptions, and the same error expression as in Eq. 2.92 was derived. In more complex systems, comparisons between reconstructed free energy surface from MTD against those obtained with other free energy calculation techniques, *e.g.*, umbrella sampling typically show good agreement within a few kcal/mol (see, *e.g.*, [75, 76]) These fluctuations in the reconstructed FES can be averaged out by computing the following [77]

$$F(q) \simeq -W(q;t_{\text{filling}}) - \frac{1}{t_{\text{tot}} - t_{\text{filling}}} \int_{t_{\text{filling}}}^{t_{\text{tot}}} \mathrm{d}t W(q;t), \tag{2.93}$$

where t_{tot} is the total time of the simulation. Crespo *et al.* [78] empirically showed that, even for a system lacking adiabatic separation between slow and fast degrees of freedom, performing such an average will reduce the error

$$\epsilon(t) \propto t_{\text{tot}}^{-1/2}, \tag{2.94}$$

which is the same law governing the convergence of reconstructed FES from umbrella sampling if the compensating bias is $W(q) = -F(q)$ (ideal case).

A practical way to estimate the statistical error in a finite simulation is by comparing profiles obtained with Eq. 2.93 in different time windows: a parallel growth of $W(q)$ is an indication that the systems reached a stationary state in which $W(q)$ compensates the underlying FES. Well tempered metadynamics [79] is a different approach aiming at reducing the fluctuations by progressively reducing the size of the hills. This technique, however, will not be employed in this thesis.

Often in a simulation, it is possible to realize if the CVs are not capturing one or more crucial slow degrees of freedom: the reconstructed free energy surface will in this case exhibit strong hysteresis oscillating between different profiles and the system will never reach a diffusive behavior.

Similarly to umbrella sampling, the volume of the space of the CV also grows exponentially with the number of CVs making the use of more than ~ 2 CV in the same simulation expensive computationally. Nonetheless, bias exchange metadynamics is an elegant way to circumvent this issue [80]. It works by simulating several replicas of the system each with different CV and allowing them to periodically exchange the bias direction. The FES is then reconstructed through a weighted histogram analysis [81].

3 Advances in Method Development

3.1 Protocol for Studying Finite Temperature Chemical Reactions in Gas Phase

While it is customary to obtain robust and reproducible results on the potential energy surface with traditional quantum chemistry methods, it is more difficult in the case of the free energy surface from MD. There are multiple reasons: (i) temperature introduces stochastic fluctuations, (ii) many techniques exist for enhanced sampling (see Section 2.4.2) with no consensus on a "optimal" one and often they require to choose collective variables that approximate the true reaction coordinate which already is a non-trivial task (see Section 2.4.1). And, finally, (iii) the statistical uncertainty on the results is seldom estimated.

Here we address these issues and, based on the study of gas-phase CO_2-water reactions in Section 4.1, we come to propose a general protocol that is able to provide a reliable and reproducible comparison of potential energy surfaces (PES) and free energy surfaces (FES).

1. The suitability of DFT for the characterization of the reaction has to be determined by comparing the results with those of higher level quantum chemistry methods.
2. *Ab initio* molecular dynamics empowered by metadynamics and the path collective variables can then be used to accelerate the reaction and reconstruct the FES. The use of the path CVs is strongly encouraged especially if the process studied is highly cooperative.

   A physically sound putative path for the path CVs can in principle be generated by starting short molecular dynamics simulations close to the transition state of the PES and extracting configurations located at equal distance from one another between reactants and products. If not, one can employ an interpolation between the stable states.

   The MD simulations have to be long enough to allow multiple recrossings to take place.
3. The convergence of the reconstructed FES is then measured by comparing the profiles in different time windows and checking for parallel growth.

4. A committor analysis (see Section 2.4.1) on the structures located at the saddle point of the FES is then performed to validate the CVs *a posteriori*.

New mechanisms, present only at finite temperature, can be uncovered by a thorough comparison of the structures in the reactant and product basins as well as in the transition state region against those computed from the PES. This analysis can be simplified by the use of structural clustering algorithms (see next section). Similarly, the calculations of ΔF, ΔH and $T\Delta S$ (Eqs. 2.68-2.70) provide an account of the energetics at finite temperature with an uncertainty estimated on the basis of a statistical argument.
Extension of this procedure to, *e.g.*, chemical reactions in solution, is problematic unless a suitable metric is employed. This issue is discussed in the next section.

3.2 Structural Clustering of Chemical Reactions in Solutions

The importance of having good reaction coordinates to study activated processes and retrieve, *e.g.*, free energy barriers has clearly emerged. A crucial property of a good reaction coordinate is the ability to resolve important chemical variations of the system (*e.g.*, bond breaking and formation) irrespective of trivial geometric differences like, for instance, the reorganisation of identical solvent molecules. To this end, we propose a novel metric [82] to compare two different atomic configurations of the same system including solvent molecules diffusing in a disordered way and possibly participating to reactions. The metric can be employed to cluster together similar structures and therefore simplify the study of long MD trajectories. The peculiarity of this metric is that it relies on a space of coordinates that is invariant under the permutation of identical objects. This might seem trivial since it is well known that Nature possesses such a symmetry; however, the customarily used Cartesian atomic coordinates in an MD simulation do not because atoms, in this case, are labelled. Hence, configurations extracted from a molecular dynamics simulation of a solute in equilibrium with the solvent will typically display large distances in Cartesian space despite belonging to the same physical metastable state of the system.
Because of this, automatizing the identification of structures that truly differ from one another from those that only differ by unimportant or unphysical details like the labelling of atoms is difficult and therefore hampers the analysis of the species found along a MD trajectory. Powerful techniques developed for simulations of complex biomolecules (*e.g.*, proteins) are able to distinguish among different conformations of the macromolecule [83, 84, 85, 86, 87]. In that case, however, first the solvent is ignored and, second, the atoms are ordered and are therefore distinguishable which avoids having issues related to the permutation symmetry.
In disordered systems or when covalent bonds are formed or broken (during a chemical reaction), this ordering no longer exists and therefore it becomes difficult to distinguish between structures that are truly different and those whose topology (physically equivalent) only differs by the unphysical labelling of the atoms. The need for a permutation invariant measure is manifest. We note that in principle it is possible to lift this kind of degeneracy by systematically permuting the atoms and checking the similarity (*e.g.*, with root mean square displacement)

[88, 89]. However in practice the number of permutations grows with the factorial of the number of particles and therefore this approach is unrealistic to compute for anything but the simplest systems.

The metric presented in the next sections is general enough that it could, in principle, be applied to a wide range of system including chemical reactions in disordered systems. Ultimately it could also serve as a distance for the path collective variables ($\mathcal{D}$ in Eq. 2.78).

3.2.1 The Permutation Invariant Vector

Clustering molecular dynamics trajectories requires the ability to compute a distance $\mathcal{D}_{AB}$ between two different configurations ($\{\boldsymbol{R}^A\}$ and $\{\boldsymbol{R}^B\}$) of the same system. This kind of distance is also employed in the path collective variables (Section 2.4.1) where $\mathcal{D}$ was chosen as the mean square difference between all pairs of interatomic distances.

We introduce a metric that is focused on a given range of distances (encompassing, for instance, bond breaking and formation) and is quite insensitive to structural differences at longer range (due to, *e.g.*, disorder in the system). We start by constructing a vector $\boldsymbol{v}$ of $N_{\text{pairs}} = N(N-1)/2$ components

$$\boldsymbol{v} = \text{sort}\left(\mathcal{F}_{IJ}\right) = \text{sort}\left(\mathcal{F}(|\boldsymbol{R}_I - \boldsymbol{R}_J|)\right), \tag{3.1}$$

where $\mathcal{F}_{IJ}$ is a monotonically decreasing function of the interatomic distance $d_{IJ} = |\boldsymbol{R}_I - \boldsymbol{R}_J|$ (*e.g.*, coordination functions cn in Equation 2.73). Its purpose is to focus the metric on the range of interest. $\boldsymbol{v}$ includes all the elements from the upper triangular part of the matrix $\mathcal{F}$. These elements are sorted within each block characterized by the pairing of the same elements. In Figure 3.1, these blocks are: the O-O (in red), the O-H in yellow, and the H-H in gray. This sorting operation ensures the invariance of $\boldsymbol{v}$ under the permutation of identical atoms and $\boldsymbol{v}$ can be dubbed a permutation invariant vector (PIV).

We note that the unmodified all-to-all interatomic distance matrix provides a redundant description of the geometry the system (aside from, *e.g.*, chirality). The focusing and sorting operations (Eq. 3.1) amount to a reduction of the information contained in the PIV. The number of components in the PIV is always constant; however based on the function $\mathcal{F}$, the number of elements with a negligible value ($\simeq 0$) will vary. The dimensionality of the PIV scales as N^2 and therefore employing it may become computationally expensive for very large N. Nonetheless, the intrinsic limited range of $\mathcal{F}$ makes possible the use of, *e.g.*, neighbor lists which should reduce the computational cost and effectively make it $\sim \mathcal{O}(N)$.

Finally, the metric $\mathcal{D}_{AB}$ is constructed from the Euclidean distance between the PIVs ($\boldsymbol{v}^A$ and $\boldsymbol{v}^B$)

$$\mathcal{D}_{AB} = |\boldsymbol{v}^A - \boldsymbol{v}^B| = \sqrt{\sum_{k=1}^{N_{\text{pairs}}} \left(v_k^A - v_k^B\right)^2}. \tag{3.2}$$

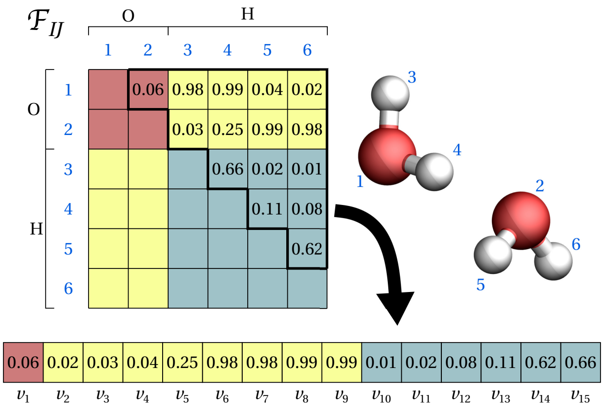

$\mathcal{F}_{IJ}$
O
H
1 2 3 4 5 6
0.06 0.98 0.99 0.04 0.02
0.03 0.25 0.99 0.98
0.66 0.02 0.01
0.11 0.08
0.62
0.06 0.02 0.03 0.04 0.25 0.98 0.98 0.99 0.99 0.01 0.02 0.08 0.11 0.62 0.66
v_1 v_2 v_3 v_4 v_5 v_6 v_7 v_8 v_9 v_{10} v_{11} v_{12} v_{13} v_{14} v_{15}

Figure 3.1 – A simple system of two water molecules illustrating the definition of the PIV: the upper diagonal part of the 6 × 6 symmetric coordination matrix $\mathcal{F}$ is rearranged into a vector of 15 components v_k. The entries from each matrix block corresponding to O−O (in red), O−H (in yellow), and H−H in (gray) are sorted in ascending order. $\mathcal{F}$ from Eq. 2.75 with $x_0 = 1.7$ Å, $p = q = 6$.

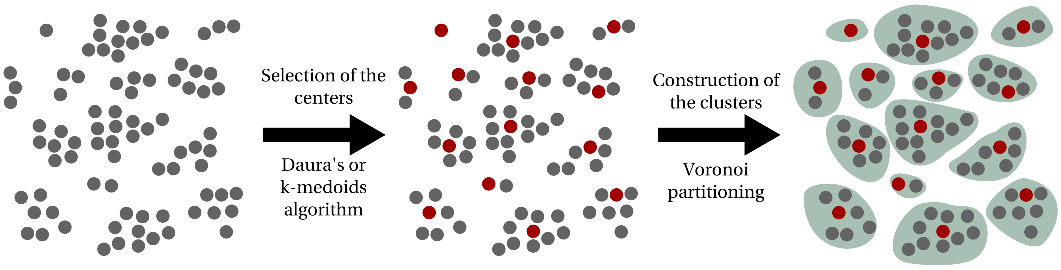
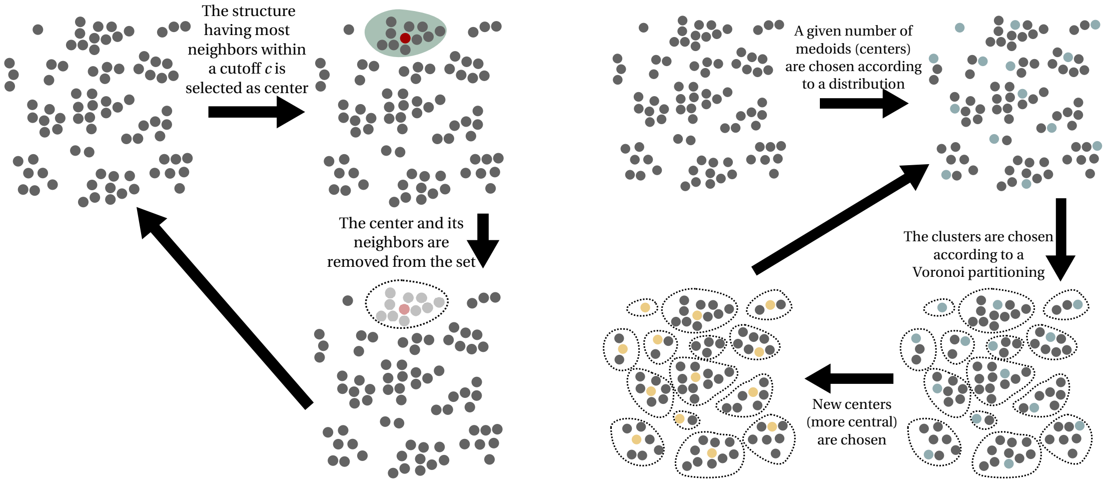

Selection of the centers
Daura's or k-medoids algorithm
Construction of the clusters
Voronoi partitioning

Figure 3.2 – Schematic view of a structural clustering analysis.

$\mathcal{D}$ could also be referred to as the frame-to-frame distance matrix. It can be to cluster together structures having a small distance (in this N_{pairs} dimensional space). Two clustering algorithms were considered to select the cluster centers:

- The simple method from Daura *et al.* [90] which requires choosing a cutoff c for $\mathcal{D}$. For each frame A, the number of neighbors, *i.e.*, the number of structures B such that $\mathcal{D}_{AB} < c$ is determined. The frame A with the most neighbors is chosen as the cluster center C_1 and it is removed along with all its neighbors from the set of structures. The procedure is iterated until no frame is left in the set (see Figure 3.3 (a)).
  The cutoff c implicitly controls the number of clusters: the larger the cutoff, the less clusters.

- The k-medoids algorithm [91] (with k-means++ initialization [92]) takes the desired number (k) of clusters rather than a cutoff as input.
  The cluster centers are chosen as follow: the first center C_1 is chosen randomly among all the structures. The second one (C_2) is chosen based on a probability distribution proportional to $\mathcal{D}^2_{C_1C_2}$. The remaining centers C_j are chosen among the remaining structures with probability $\propto \mathcal{D}^2_{C_iC_j}$ where C_i is the center (already chosen) closest to C_j ($\mathcal{D}_{C_iC_j}$ is at a minimum).
  Given a set of centers, the members are determined by Voronoi partitioning (see Figure 3.3 (b)). Then, within each cluster, a new center is determined as the one with the smallest average distance with all the other members. The procedure is iterated over


The structure having most neighbors within a cutoff *c* is selected as center
The center and its neighbors are removed from the set
A given number of medoids (centers) are chosen according to a distribution
The clusters are chosen according to a Voronoi partitioning
New centers (more central) are chosen

(a) Daura's algorithm

(b) k-medoids algorithm

Figure 3.3 – Schematic representation of two algorithms for choosing the cluster centers. See text for more details.

until the centers do not change anymore.

Once the centers have been determined the clusters are constructed based on a Voroni partitioning, *i.e.*, the configuration A is attributed to the center C for which $\mathcal{D}_{AC}$ is at a minimum. This method has been applied to chemical reactions in gas phase (Section 4.1.2) as well as in solution (Section 4.2). We further benchmarked it on the well documented cases of Na^+-Cl^- ion pair dissociation in solution (see Section 3.2.2).
A code performing this analysis, `piv_clustering.x`, can be downloaded freely from the sourceforge website [93].

3.2.2 Test Case: Na^+-Cl^- Ion Pair Dissociation in Solution

NaCl in solution is a well studied system (*e.g.*, [65, 94]) and offers a sound test case for assessing the robustness of the new structural clustering approach proposed in Section 3.2.
In this system, the solvent does not behave as a spectator in the dissociation/association reactions and beside the $Na^+ - Cl^-$ distance, one needs a second collective variable to account for the reorganization of the solvation shells (see Ref. [65] and Section 2.4.1). Following Ref. [7], we select the potential energy calculated between Na^+ and the surrounding water molecules (Eq. A.1). For additional computational details, we refer to Appendix A.1.

The systems has three metastable states: the contact ion pair, the solvent separated ion pair and the fully dissociated ion pair (respectively A, B and C in Figure 3.4) with relative equilibrium populations of 1%, 8% and 67%. Figure 3.4, shows the free energy surface obtained directly from the equilibrium MD simulations.

Figure 3.4 – Free energy landscape as a function of the Na$^+$-Cl$^-$ distance and of the electrostatic energy between Na$^+$ and water. Contour lines are plotted every $k_B T$. Metastable basins A, B and C correspond to the contact ion pair, the solvent separated ion pair and the fully solvated ions respectively. The dots are the centers of the 40 clusters obtained with the k-medoids algorithm. The inset magnifies the transition state region between A and B.

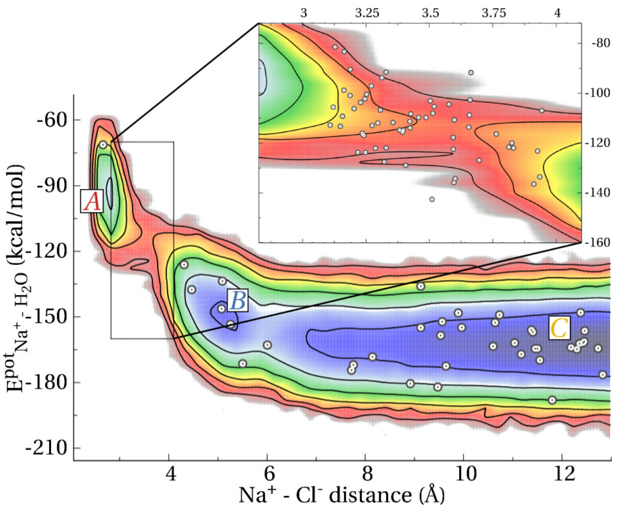
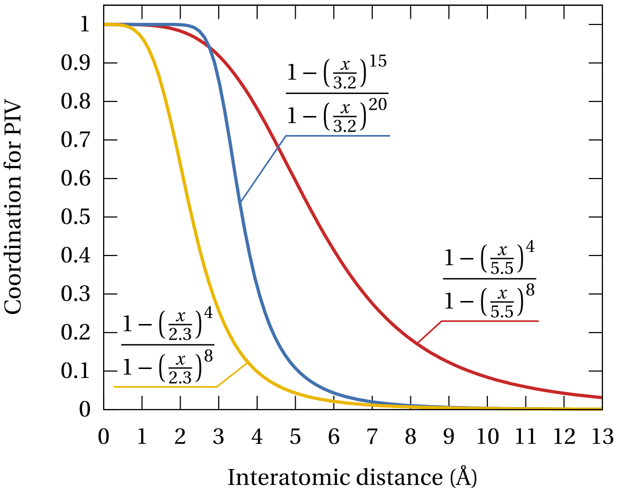


Figure 3.5 – Coordination functions computed from the interatomic distance x and employed to focus the PIV on the appropriate length scale in the different system presented.



The all-to-all coordination matrix $\mathcal{F}$ focused the search for clusters in appropriate range of interatomic distances, typically below ~ 9 Å, and removed the contribution from far away atoms that, in solution, do not add relevant chemical information but mainly noise. Moreover, since the hydrogen atoms do not participate in the breaking or formation of covalent bonds, we assume that they can be discarded when constructing the PIV. From a computational point of view, the reduction in the dimensionality of the PIV amounts to an important gain in computational speed. In the absence of hydrogen, the shortest distance between a pair of atoms in this system is ~ 2.5 Å. Therefore we construct $\mathcal{F}$ to be focused in the $[2.5:9.0]$ Å range by employing the asymmetric coordination function (Equation 2.75) and the parameters $d_0 = 5.5$ Å, $p = 4$ and $q = 4$ (see red curve in Figure 3.5).

We note that if we were to employ the interatomic distances (d_{IJ}) instead (*e.g.*, Ref. [95]) of the $\mathcal{F}_{IJ}$ in Equation 3.1 we would precisely bring noise and spoil the results of the clustering. This was verified by extracting structures from the basins A, B, and C and computing the

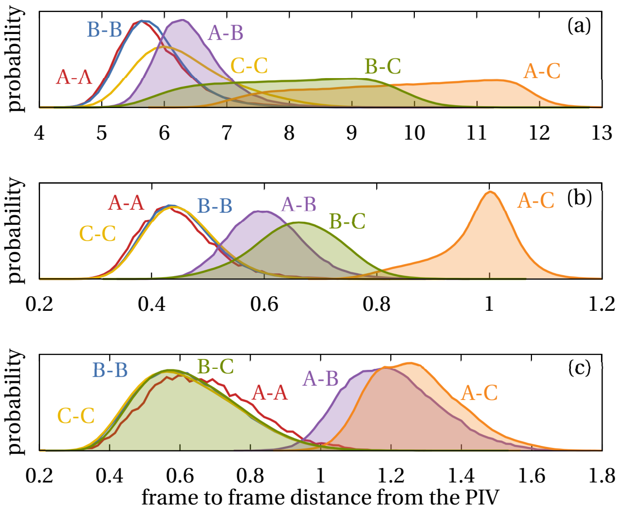


Figure 3.6 – Distribution of intra- and inter basin distances computed from three different definitions of $\mathcal{F}$. (a) plain interatomic distances ($\mathcal{F}_{IJ} = d_{IJ}$). (b) coordination focused on a 2.5-9.0 Å range (red curve in Figure 3.5) and (c) coordination focused on a 2.5-5.0 Å range (blue curve in Figure 3.5).

distance $\mathcal{D}$ between all possible structure once with a PIV constructed from d_{IJ} and once from $\mathcal{F}_{IJ}$. Figure 3.6 (a) and (b) shows the respective distributions of frame-to-frame distances $\mathcal{D}$. The overlap between the $A-A$ distribution and the $A-B$ distribution in Figure 3.6 (a) is problematic because it means that a significant portion of structures belonging to basin B have a similar distance from a structure in A than from another structure in B. This overlap is substantially reduced if $\mathcal{F}$ is used to compute the PIV instead (Figure 3.6 (b)).

The resulting forty centers from the clustering of 1 μs long trajectory (more than 10000 frames) with the k-medoids algorithm are reported as dots in Figure 3.4: they are well distributed over the FES and we see from Figure 3.7 that all the members assigned to a specific center can be determined to belong to either A, B or C. This shows that tens of thousands of structures can be reduced to only a few tens which significantly simplifies the analysis and visualization of the ensemble of structures.

Since the k-medoids algorithm is randomly initialized, it will sometimes fail to locate at least one center in A because this state represents only 1% of the total population and therefore we advise care in employing this algorithm in presence of scarcely populated metastable states. From this point of view, Daura *et al.* 's algorithm [90] is more reliable. Finally, we note that the different numbers of clusters in basins A, B and C can be related to the difference in their relative population.

For the $A \leftrightarrow B$ transition state region, a finer analysis is instructive. We extracted all the structures (424) having a Na$^+$-Cl$^-$ distance between 3.10-3.95 Å as it is the range where the transition state region is located. We already stated that only a few of these structures will behave as true transition state because of the non-trivial solvent-related degree of freedom and we shall see if our clustering approach is able to capture this behavior. This time the PIV was constructed from a slightly more tuned coordination function focusing the range in 2.5-5 Å (blue curve in Figure 3.5); as shown in Figure 3.6 (c), this function fails to distinguish basin B from basin C but there is almost no overlap between A and B which is very important since the

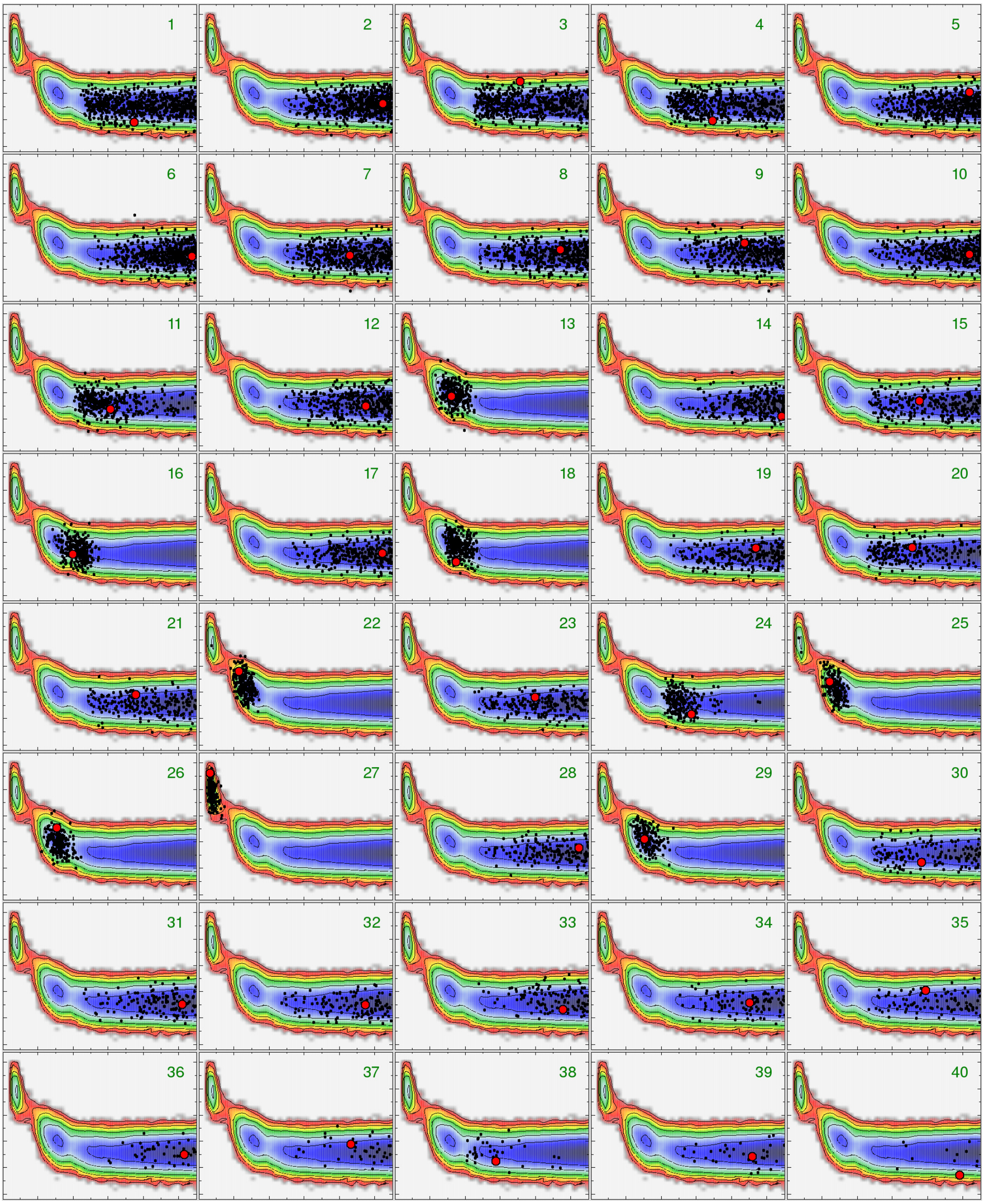


Figure 3.7 – Cluster members (black points) and centers (red points) for all 40 clusters of the NaCl solution trajectory obtained using a PIV focused in the 2.5-9 Å range (red curve in Figure 3.5) and the k-medoids algorithm. The FES is the same as in Figure 3.4.

TS is located between these two basins. Indeed, in the absence of a clear separation between A and B, there is no hope of correctly identifying the transition state because, by definition, structures from A and B will be mixed together. In particular, we see from Figure 3.6(a), that employing the interatomic distance rather than $\mathcal{F}$ provides a poorer separation of the A and B basins.

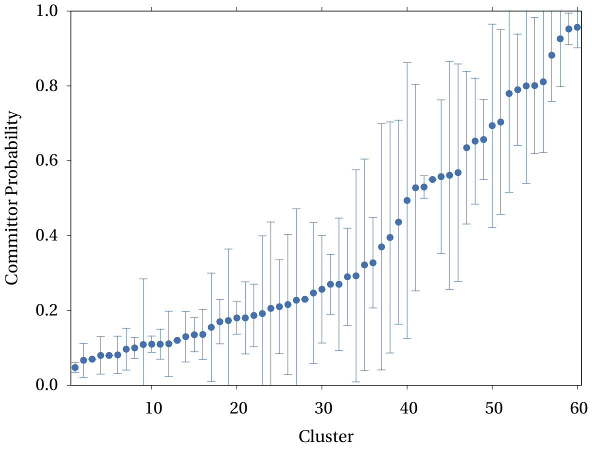


Figure 3.8 – Distribution of probabilities to fall in basin B (the solvent-separated Na^+-Cl^- pair) in the set of members of each cluster from the region in the inset of Figure 3.4. The circles denote the average committor probability with RMSD as error bars.

The sixty centers obtained with the k-medoids algorithm are depicted as dot in the inset of Figure 3.4. For each individual cluster, we performed a committor analysis (see Section 2.4.1) on its members by starting for each one 100 MD runs of 4 ps with Boltzmann distributed ($T = 300$ K) initial velocities. The probability to fall into basin B for each cluster averaged over the cluster members is reported in Figure 3.8. A majority of the clusters is committed either to A or B and only a few lie between 0.4 and 0.6 which are the ones exhibiting a transition state-like behavior.

The use of the PIV based metric allowed us to recover known features of the NaCl dissociation reactions in solution. A more thorough analysis of the different clusters reported in Figure 3.7 and at the transition state could reveal unforeseen characteristics of the process. This will be the subject of a later study.

3.2.3 Conclusion

The simple approach we introduced allowed us to cluster MD trajectories without requiring a deep knowledge of the processes beforehand. It can be used to simplify the analysis, interpretation, and visualization of the results. It was shown to be particularly effective for chemical reactions in solution where the clusters exhibited consistent structural and kinetic properties. Two other applications of the methods are present in this thesis; one is a reaction in the gas phase (Section 4.1.2) and the other is in solution (Section 4.2).

4 Applications

In this Chapter we apply the simulation methods described in Chapters 2 and 3 to the study of reactions that are known to require a non-negligible barrier in order to take place. We choose to investigate specific aspects of problems of great current interest, namely the early steps in the capture of carbon dioxide in solution and the diffusion of an impurity (S) into a photovoltaic material (CdTe).

Reducing the concentration of atmospheric carbon dioxide is one of the main challenges of this century [96, 97]. Carbon capture and sequestration encompasses techniques able to either remove it directly from the atmosphere or to prevent its release from point sources such as fossil-fuels (coal and oil) based power plants or industrial facilities. The most mature technology for CO_2 post-combustion capture goes through a cyclic process and employs amine solutions [98]. First, the gas passes through the aqueous solution in which CO_2 binds to the organic molecules in different forms through different chemical reactions depending on the nature of the amine. Then higher temperature scrubbing allows CO_2 to be released in a controlled environment for final storage, and the amine solutions are regenerated by heating so as to restart the capture step. The commonly used apparatus uses monoethanolamine (MEA) at fixed concentration (30 wt.%), which is highly performant for the capture. However, release and regeneration imply a high energy penalty [99, 100]. Many efforts are being made worldwide to try and devise new solvents or processes to improve the overall efficiency of this operation [101, 102, 103].

There is a huge literature on the performance of different amines, under diverse conditions, as well as on the description of the kinetics of the multi-step reactions involved. On the other hand, there are only few simulations aimed at characterizing their pathways and energetics. There is ample room for extensive and more accurate computational studies.

We have approached this problematics from scratch. First, we have cared about the suitability of our computational scheme to represent the CO_2-water interaction [104]. We focused on characterizing the reactions occurring in the gas phase, *i.e.*, the formation and dissociation of a carbonic acid molecule (H_2CO_3) with increasing number of water molecules. This system offered us the possibility to study in detail these reactions with calculations of the potential energy surface (PES) as well as via metadynamics (MTD) and the reconstructed free-energy surface (FES). Moreover the PES characteristics obtained with a variety of exchange-correlation functionals could be directly compared with high level quantum-chemical calculations [105, 106].

Then we studied one of the early steps of the CO_2 capture in a (dilute) MEA solution [76][1], namely the formation of the carbamate via a zwitterion. The latter was never observed experimentally but has long been postulated to act as the intermediate of the reaction [107]. Modeling and simulations of this reaction are only a few and of limited sophistication. In particular, the solute-solvent interactions are often described with continuum models [108, 109, 110, 111], whose validity fails when the solvent participates actively to the reaction.

Reducing the amount of CO_2 in the atmosphere also implies finding clean ways to produce energy. Solar cells are among the most promising solutions. Silicon provides the current solution to the demands of the market, both in the crystalline and amorphous form.

Cadmium telluride is a semiconductor with a direct band gap E_{gap} of 1.50 eV at 300 K [112], which is optimal for sunlight harvesting , and high absorption coefficient ([113]). The optimal junction is with n-type cadmium sulfide (CdS), which acts as the window material having a gap of 2.4 eV at 300 K [114]. The material characteristics that are believed to mostly influence the performance are the morphology of the CdTe/CdS interface and trapping of impurities and dopants at the grain boundaries [115, 116, 117].
Understanding from theory relies on studies of the electronic structure of bulk CdTe with native defects and typical dopants [118, 119, 120] and on very few calculations of simplified models of grain boundaries [121]. Recently, diffusion of Cu and Ag impurities in the bulk phase has been investigated with nudged elastic band (NEB) calculations [122].

Our simulations refer to sulfur, which is an unavoidable impurity penetrating from the interface and mainly diffusing along the grain boundaries. We studied the dynamics of diffusion of sulfur in bulk CdTe at very low concentration (one over 100 anions), in the absence and in the presence of native defects, with the aim of finding the most probable pathways, and of understanding the underlying mechanisms and the induced changes in electronic structure. Again we attempted a comparison of the outcome of MTD with that of NEB calculations. [123]

---

[1] The simulations presented in this section were performed by Ciro A. Guido. My contribution was mainly to the analysis of the results.

4.1 Gas Phase Reaction of Carbon Dioxide and Water: the Formation of Carbonic Acid

4.1.1 Results from the Potential Energy Surface

The PES of the systems composed of a carbon dioxide molecule and up to three water molecules was first investigated in order to assess the suitability of DFT employing different xc functionals (namely PBE, BLYP and B3LYP) to describe the interactions between CO_2 and H_2O. This included the comparison with higher level quantum chemistry methods [105, 106]. The limited size of the system (at most 12 atoms) makes all electrons calculations of PES possible and we carried out B3LYP and meta-hybrid M06 [43] calculations in this approach. For more computationally demanding simulations such as *ab initio* MD we will employ the pseudopotential/plane wave approach exclusively.
The transition states of the PES were determined by using the partitioned rational-function-optimizer (P-RFO) [67]; additional computational details can be found in Appendix A.2.

The formation and dissociation of carbonic acid follows

$$CO_2 + nH_2O \leftrightharpoons H_2CO_3 + (n-1)H_2O, \quad n = 1,2,3 \tag{4.1}$$

The PES of these reactions have been studied extensively at different levels of theory. The most accurate ones were obtained by Nguyen and collaborators (CCSD(T) energy on MP2 geometry) [106]. In agreement with previous studies [124, 125, 105], they found a substantial barrier reduction for H_2CO_3 formation and dissociation upon the addition of water molecules. The direct participation of $n = 3$ H_2O molecules was stressed as being particularly important. DFT studies were also carried out with various approximations for the xc functionals [126, 127, 128]. We will employ these results to perform a critical analysis of the PES obtained in our calculations and validate our DFT scheme.

In Figure 4.1, we report a sketch of the optimized structures of the reactants (R_n), products (P_n) and transition states (TS_n) for the reactions in Eq. 4.1. All methods agree on the description of the structures, as indicated by the small root mean square deviations from MP2 geometries (Table 4.1).
In the reactants, the H_2O molecules do not tend to form hydrogen bonds with CO_2. The formation of a water cluster is favored when more than one water molecules are present (dimer for $n = 2$ and closed/open trimer for R_3 and R'_3 respectively). For $n = 1$, the transition state breaks apart the water molecule while carbon dioxide bends to eventually form carbonic acid. For $n > 1$, the mechanisms are qualitatively different: the TS consists in a ring of water molecules in which a proton can circulate more easily. For $n = 3$, two mechanisms were considered; TS_3 has all water molecules actively participating and exchanging protons along the ring. In TS'_3, one behaves as a spectator and neither accepts nor donates a proton. The latter mechanism is

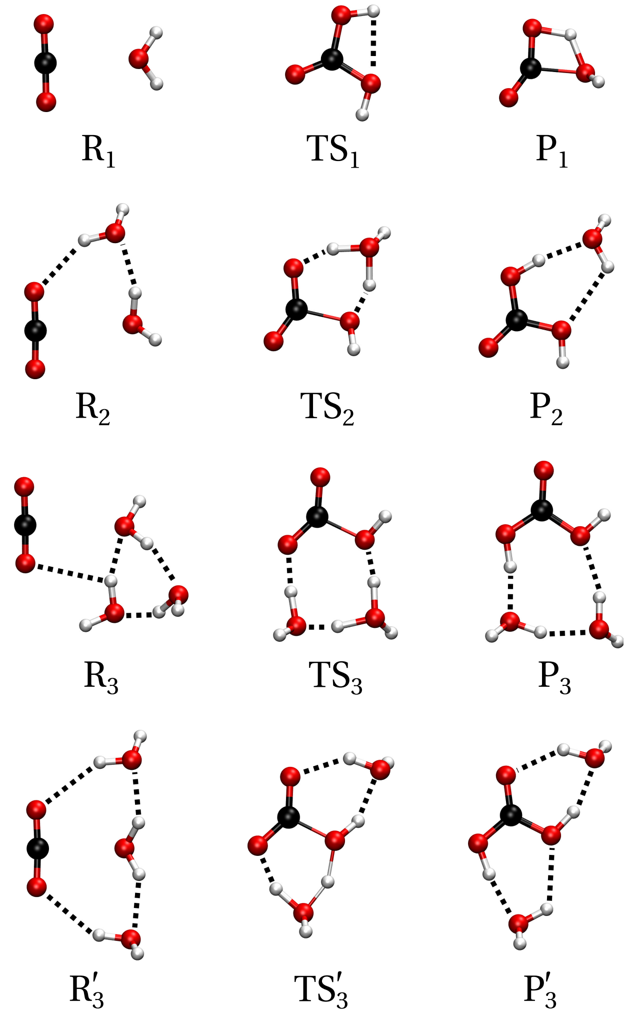


Figure 4.1 – Reactants (R_n), transition states (TS_n) and products (P_n) for the reaction in Eq. 4.1. $n = 1,2,3$. The black ball is carbon, the red are oxygen and the withe hydrogen.

similar to the one observed for $n = 2$.

The structures being very close in different xc functional schemes, a comparison of the binding energies (BE)

$$\mathrm{BE}(X_n) = E(X_n) - ((E(\mathrm{CO_2}) + nE(\mathrm{H_2O}))\,, \quad X = \{\mathrm{R}, \mathrm{TS}, \mathrm{P}\} \tag{4.2}$$

should provide precious insights on the actual capabilities of the different functionals to account for the interactions of the different complexes correctly. We adopt the convention of having a negative binding energy when the system is bound. The term *binding energy* for a transition state might be slightly misleading since it is usually not bound but we formally refer to the definition in Eq. 4.2. Further down, we shall also mention *barrier heights* (BH) for the forward and backward reactions (formation and dissociation of H_2CO_3), we will actually be

Table 4.1 – Root mean square deviation of DFT geometries with respect to MP2 geometries in Ref. [106] (Å). TS_3 geometry not reported in Ref. [106].

System	PBE^a	BLYP^a	B3LYP^a	B3LYP$^{a'}$	M06$^{a'}$
R_1	0.007	0.010	0.008	0.010	0.003
TS_1	0.001	0.004	0.002	0.001	0.002
P_1	0.001	0.002	0.001	0.001	0.001
R_2	0.007	0.008	0.005	0.003	0.007
TS_2	0.003	0.005	0.003	0.001	0.002
P_2	0.004	0.005	0.004	0.004	0.004
R_3	0.012	0.015	0.011	0.011	0.025
P_3	0.032	0.032	0.032	0.033	0.032
R'_3	0.011	0.012	0.008	0.006	0.008
TS'_3	0.003	0.005	0.004	0.006	0.004
P'_3	0.005	0.005	0.005	0.005	0.003

^{a} Pseudopotential/plane waves. Cutoff 100 Ry $^{a'}$ All electrons. Aug-cc-pVTZ basis set

Table 4.2 – Binding energies (kcal/mol) referred to the isolated CO_2 and H_2O molecules, including ZPE.

System	PBE^a	BLYP^a	B3LYP^a	B3LYP$^{a'}$	M06$^{a'}$	MP2^b	CCSD(T)^b
R_1	−1	abs < 1	abs < 1	−1	−2	−2.0	−2.1
TS_1	40	46	49	48	49	48.3	48.5
P_1	8	14	11	10	11	11.4	8.6
R_2	−6	−3	−4	−4	−6	−7.0	−7.2
TS_2	16	25	26	26	26	25.1	25.0
P_2	2	9	6	5	4	4.2	1.6
R_3	−12	−8	−9	−10	−12	−14.2	−14.4
TS_3	5	16	15	15	15	14.0	13.9
P_3	−6	3	−1	−1	−2	−3.0	−5.7
R'_3	−9	−5	−7	−7	−9	−10.9	−11.3
TS'_3	6	18	17	17	16	14.3	13.8
P'_3	−9	2	−3	−4	−6	−5.6	−8.5

$^{a,a'}$ Same as Table 4.1 ^{b} Ref. [106]: single-point calculations on MP2 geometries. aug-cc-pVTZ basis set

Table 4.3 – Barrier heights (kcal/mol) for the formation and dissociation of carbonic acid and their decrease (in italics) due to the addition of one water molecule (from n to $n+1$).

(a) formation

n	PBE^a	BLYP^a	B3LYP^a	B3LYP$^{a'}$	M06$^{a'}$	MP2^b	CCSD(T)^b
1	41	46	49	49	51	50	51
2	22	29	30	30	32	32	32
1−2	*19*	*18*	*19*	*19*	*19*	*18*	*19*
3	17	24	25	25	28	28	28
2−3	*5*	*5*	*5*	*5*	*4*	*4*	*4*
3′	15	23	23	24	25	25	25
2−3′	*7*	*5*	*6*	*6*	*7*	*7*	*7*

(b) dissociation

n	PBE^a	BLYP^a	B3LYP^a	B3LYP$^{a'}$	M06$^{a'}$	MP2^b	CCSD(T)^b
1	32	32	37	38	39	37	40
2	15	16	20	21	22	21	23
1−2	*17*	*16*	*17*	*17*	*17*	*16*	*17*
3	11	12	17	16	18	17	20
2−3	*4*	*4*	*4*	*4*	*4*	*4*	*4*
3′	15	17	20	21	21	20	22
2−3′	abs < *1*	abs < *1*	*1*	abs < *1*	*1*	abs < *1*	abs < *1*

$^{a,a'}$ Same as Table 4.1 ^{b} same as Table 4.2

referring to $E(\mathrm{TS}_n) - E(\mathrm{R}_n)$ and $E(\mathrm{TS}_n) - E(\mathrm{P}_n)$ respectively. These BH will always be positive. The zero-point energy (ZPE) is included in the BE and BH from the potential energy surfaces. All methods predict the same ZPE contribution for the reactants, products and transition states within 1 kcal/mol.

By inspection of the BEs in Table 4.2, we remark that the reactant R_1 is a shallow minimum for all methods, suggesting a limited interaction between CO_2 and H_2O. This is further confirmed by the softness of the modes associated with the $\mathrm{C}_{\mathrm{CO}_2} - \mathrm{O}_{\mathrm{H}_2\mathrm{O}}$ distance (~ 150 cm^{-1}). The same remark can be made for R_2 and R_3 where the interactions (*e.g.*, hydrogen bonds) within the water moiety (dimer or trimer) account for the major part of the binding energy.

From Tables 4.3 (a) and (b) we note that all methods predicts the same catalytic effect caused by water addition (within 1 kcal/mol). This remarkable quantitative agreement suggests that the main contribution to the reduction of either formation or dissociation barriers is mainly due to electrostatic effects.

Besides these similarities, we note that PBE strongly underestimates (~ 10 kcal/mol) the BE of the transition states and the formation or dissociation BH. PBE being non-empirical leads us to conclude that the relative good agreement in the reactants and products is fortuitous and can be attributed to the known tendency of PBE to strengthen intermolecular bonds in water [129].

The closeness of the BE obtained for B3LYP in both the pseudopotential and all-electrons schemes allows us to validate the former even though, the pseudopotential employed was originally derived for the BLYP xc functional. We also note the good agreement of the BE computed with the hybrid B3LYP and meta-hybrid M06 functionals with the MP2 values.

4.1.2 Reaction Dynamics and Free Energy Surfaces

With a solid characterization of the potential energy surfaces for the reactions in Eq. 4.1 we carry on with a study of the formation and dissociation of carbonic acid at finite temperature by means of *ab initio* MD. We employ metadynamics to accelerate the occurrence of the transitions and allow the system to overcome the high free energy barriers suggested by the PES. The limited size of the systems makes the calculations of long MD trajectories possible (> 100 ps) and much can be learned on how finite temperature effects can modify the reactions mechanisms and energetics.

The study of these reactions at finite temperature was the basis to construct the protocol detailed in Section 3.1.

The Crucial Choice of Collective Variables and its Validation

The choice of appropriate collective variables is central if one wants to track the reaction progress and reconstruct free energy surfaces (see Section 2.4.1).

In fact, prior to devising the scheme in Section 3.1, we had made several unsuccessful attempts employing simple collective variables such as the $C - O_{H_2O}$ coordination or the number of covalent bridges between C-O-H. These CVs were successful in distinguishing reactants and products (*i.e.*, good order parameters) and even to simulate formation and dissociation reactions. However, the structures at the saddle point of the FES failed the committor test, *i.e.*, the structures at the saddle point were mostly committed to the reactants or to the products (Section 2.4.1). Moreover, the free energy surfaces reconstructed from the metadynamics resulted in barriers higher than the ones from the potential energy surfaces by ~ 20 kcal/mol. This finding could not be reasonably explained. Despite fine tuning of the parameters entering in the CV definition, these issues were never satisfactorily solved.

To capture the very cooperative nature of the mechanisms involved, we decided to employ the path CVs instead (see Section 2.4.1). We could have chosen a linear interpolation between reactants and products for the reference paths, However, obtaining more physical paths was relatively easy and computationally inexpensive: they were generated from short unbiased MD runs started at the PES transition states and forming continuous trajectories from reactant to product. A few equidistant (~ 10 – 20) frames were selected from the trajectories. The paths obtained with the PBE xc functional were also employed in the simulations with the BLYP and B3LYP xc functionals.

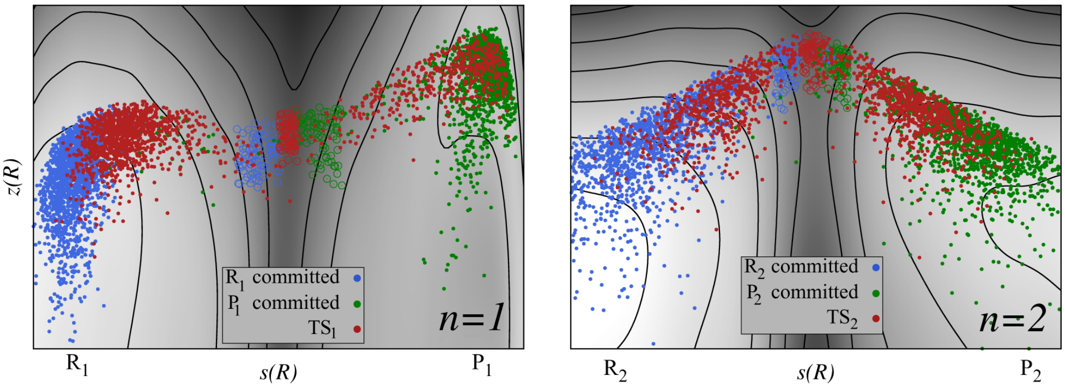


Figure 4.2 – Committor analysis for the $n = 1$ and $n = 2$ systems (BLYP). Start (empty circles) and end (filled) points of the unbiased trajectories.

These CVs performed extremely well with respect to our previous choice: they allowed us to simulate easily multiple re-crossings from reactants and back. In addition, they were *a posteriori* validated by passing the stringent test of committor analysis.

Figure 4.2 shows the results of the committor analysis for the BLYP $n = 1$ and $n = 2$ cases. We selected a region around the saddle point of the free energy surfaces (Ω^{TS}) and chose two different regions, one more towards the reactants (Ω^{R}) and the other more towards the products (Ω^{P}), which should be committed either to the reactants or the products respectively. In Figure 4.2, structures extracted from the metadynamics trajectories and falling into the regions Ω^{TS}, Ω^{R} and Ω^{P} are represented as empty red, blue and green circles respectively. From each of these configurations, we started twenty short MD runs with initial velocities drawn from a Maxwell-Boltzmann distribution at 300 K. The end points of these simulation are represented as filled circles in Figure 4.2.

For $n = 1$, the configurations in the region Ω^{R} were committed to the products (p_R) with $p_R = 100\%$ while those in the region Ω^{P} were committed to the products (p_P) with $p_P = 98\%$. In Ω_{TS}, the probabilities were $p_R = 70\%$ and $p_P = 30\%$. For $n = 2$ we had $p_R = 99\%$, $p_P = 98\%$ and $p_R = 39\%$ and $p_P = 53\%$ respectively. Qualitatively similar results were found for the systems with $n = 3$.

Convergence of the Free Energy Surfaces

In metadynamics simulations of chemical reactions with *ab initio* molecular dynamics it is customary to estimate free energy barriers heights after the first crossing (see, *e.g.*, Ref. [130]) rather than after having observed multiple forward and backward transitions. With this approach, parallel growth of the surfaces cannot be observed, *i.e.*, the reconstructed FES do not grow parallel to itself (Section 2.4.2). Therefore, the results may suffer from uncertainties which can not be estimated on the basis of a statistical argument. In principle however, given an appropriate choice of collective variables, parallel growth can be achieved if the system is

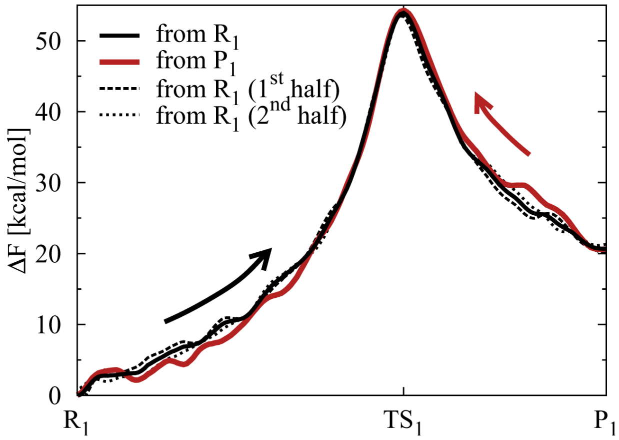


Figure 4.3 – Free energy profile along the reaction path on the reconstructed FES for the reaction in Eq. 4.1 with $n = 1$ (BLYP): from reactants (solid black) and products (solid red) and from the two halves of the simulations started from the reactants (dashed and dotted black lines).

small enough that long MD trajectories can be generated. Having both several re-crossings and parallel growth of the reconstructed FES is highly desirable if one aims at the same level of robustness and reproducibility as in PES calculations.

All the free energy surfaces discussed in the following were reconstructed at the end of a MTD simulation (~ 200 ps) after about ten forward/backward reactions had been observed. Parallel growth was realised in all cases and for all the xc functionals.

As an example, we refer to the similarity of the two free energy profiles reconstructed at different times (dashed and dotted lines in Figure 4.3). In general, the root mean square deviation between two FES reconstructed at two different filling time (Eq. 2.93) provides an estimate of the statistical error of ~ 2 kcal/mol.

Having already fulfilled both the committor test and the parallel growth requirement, the path CVs scheme was even further validated by checking the reproducibility of the results. For $n = 1$, a second simulation was initialized in the products P_1 instead of the reactants. Excellent agreement was found between the reconstructed FES (see the alignment of the red and black curves of the profile along the reaction pathway in Figure 4.3).

Free Energy Landscapes and Comparison with the Potential Energy Surface

The previous section established that the chosen path CV produced reliable results. In particular, we showed that reconstructed free energy surfaces were converged. We are, therefore, confident that the calculations of the free energy barriers, the characterization of the reactants and products basins as well as of the transition state regions, and the description of the dynamics of the reaction are meaningful and that comparing them with the results from the potential energy surfaces is possible.

The structures in the product basins and the transition state regions were in close agreement with their equilibrium counterparts and the structural parameters oscillated around the equi-

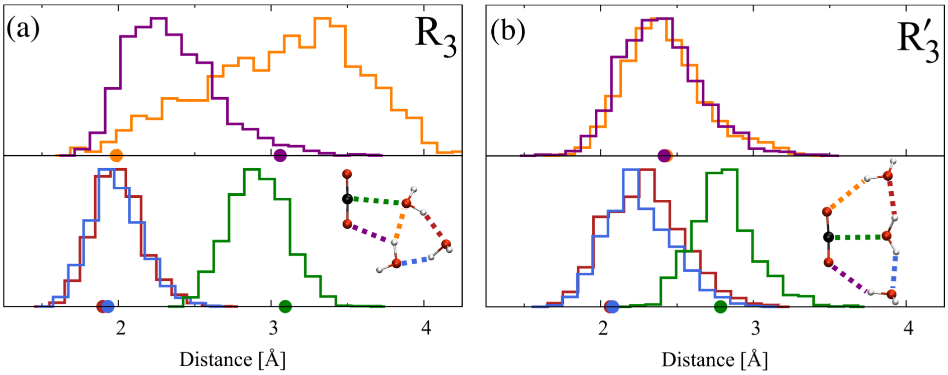


Figure 4.4 – $n = 3, 3'$: distribution of relevant distances in the reactants and products basins. The optimized geometry from Figure 4.1 is reported in inset. The dot represent the value of the distance of the optimized geometry.

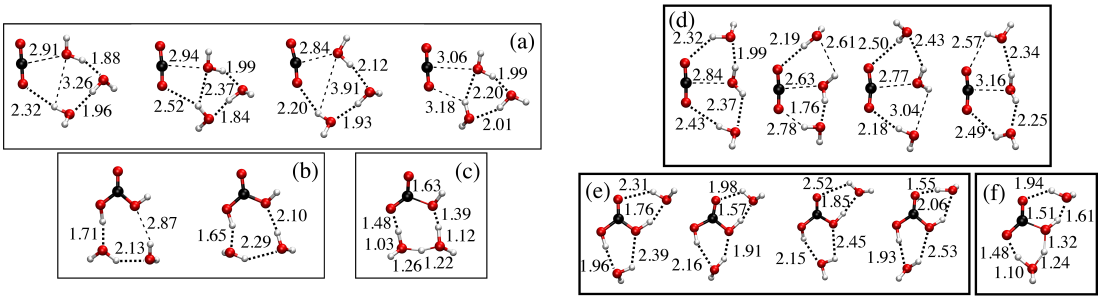


Figure 4.5 – $n = 3$ cluster centers for the (a)/(d) reactant basin R_3/R'_3, (b)/(e) product basin P_3/P'_3 and (c)/(f) transition state region TS_3/ TS'_3. Distances in Å.

librium value. The same was true for the reactants with $n = 1, 2$. For $n = 3$, however, there were a few differences for the reactants. The optimized R_3 from the PES was made of a closed water trimer having no hydrogen bond with CO_2 (Figure 4.1), while the structures in the reactant basin of the FES mostly were made of an open water chain forming one hydrogen bond with carbon dioxide (see Figure 4.4 (a) for the distribution of distances and Figure 4.5 (a) for an example of the configuration). Also shown in Figure 4.4 (b) is the comparison of distances for R'_3 showing the agreement between statics values (filled circles) and their dynamical distribution.

These configurations were further investigated employing structural clustering based on the permutation invariant vector and Daura's algorithm (see Section 3.2). Being in the gas phase, no noise can come from, *e.g.*, far away solvent molecules and therefore it is not necessary to restrict the PIV to a specific range. We used simply the plain interatomic distances to construct it. The cutoff for the clustering algorithm was chosen such that the first four clusters of the reactant basin encompassed more than 50 % of all the structures in the basin.
The cluster centers from the reactant and product basins for $n = 1, 2$ differed mainly in the orientation of the water molecules; both TS_1 and TS_2 had only one cluster and the structure of their respective center were similar to the PES transition state.

Table 4.4 – Free energy barrier heights (kcal/mol) for the formation and dissociation of carbonic acid at room temperature and decrease (in italics) due to the addition of one water molecule (from n to $n+1$).

	Formation			Dissociation		
n	PBE	BLYP	B3LYP	PBE	BLYP	B3LYP
1	46	51	54	31	33	38
2	29	32	31	18	20	19
1 − 2	*17*	*20*	*23*	*14*	*13*	*19*
3	26	31		16	19	
2 − 3	*3*	*1*		*1*	*1*	
$3'$	19	26		15	16	
2 − 3′	*10*	*6*		*3*	*4*	

In the case with three water molecules, the richness of conformations in R_3 was confirmed and the most populated was indeed the "open water chain"; the other centers all showed a different degree of openness in the chain and the geometry of the static reactant was also present (Figure 4.5 (a)). The product P_3 exhibited a narrower clustering distribution with the first two clusters encompassing already 80 % of the 1600 configurations (Figure 4.5 (b)) and TS_3 was reduced to one cluster (Figure 4.5 (c)) as in the case of $n = 1$ and $n = 2$. R'_3 exhibited less deviation from the geometry at 0 K, the cluster analysis revealed that the four most populated cluster are more compact than for R_3: two water molecules form a hydrogen bond with CO_2 and the third one serves as a bridge between them (Figure 4.5 (d)). The P'_3 clusters are closely related to each other with the carbonic acid separating the two remaining water molecules (Figure 4.5 (e)).

In Figure 4.6, we report the free energy surfaces and the evolution of a few structural parameters along the reaction pathway (BLYP xc functional; similar FES were obtained for PBE with $n = 1,2,3$ and B3LYP with $n = 1,2$). For $n = 3$, we performed two distinct simulations whose paths were generated by a trajectory shot either from TS_3 or from TS'_3. In principle, both mechanisms could have been found in a single simulation by virtue of the second collective variable of the path CV (z; Eq. 2.81) which allows the system to explore different TS from the one included in the path. However, a reasonably precise free energy surface for different mechanisms can only be obtained if the two different paths are employed. We report the values for the free energy barriers for formation and dissociation in Table 4.4. We recall that the root mean square alignment of FES reconstructed at two different filling times was ~ 2 kcal/mol and within this error, the catalytic effect of adding water molecules is quite independent of the xc functionals.

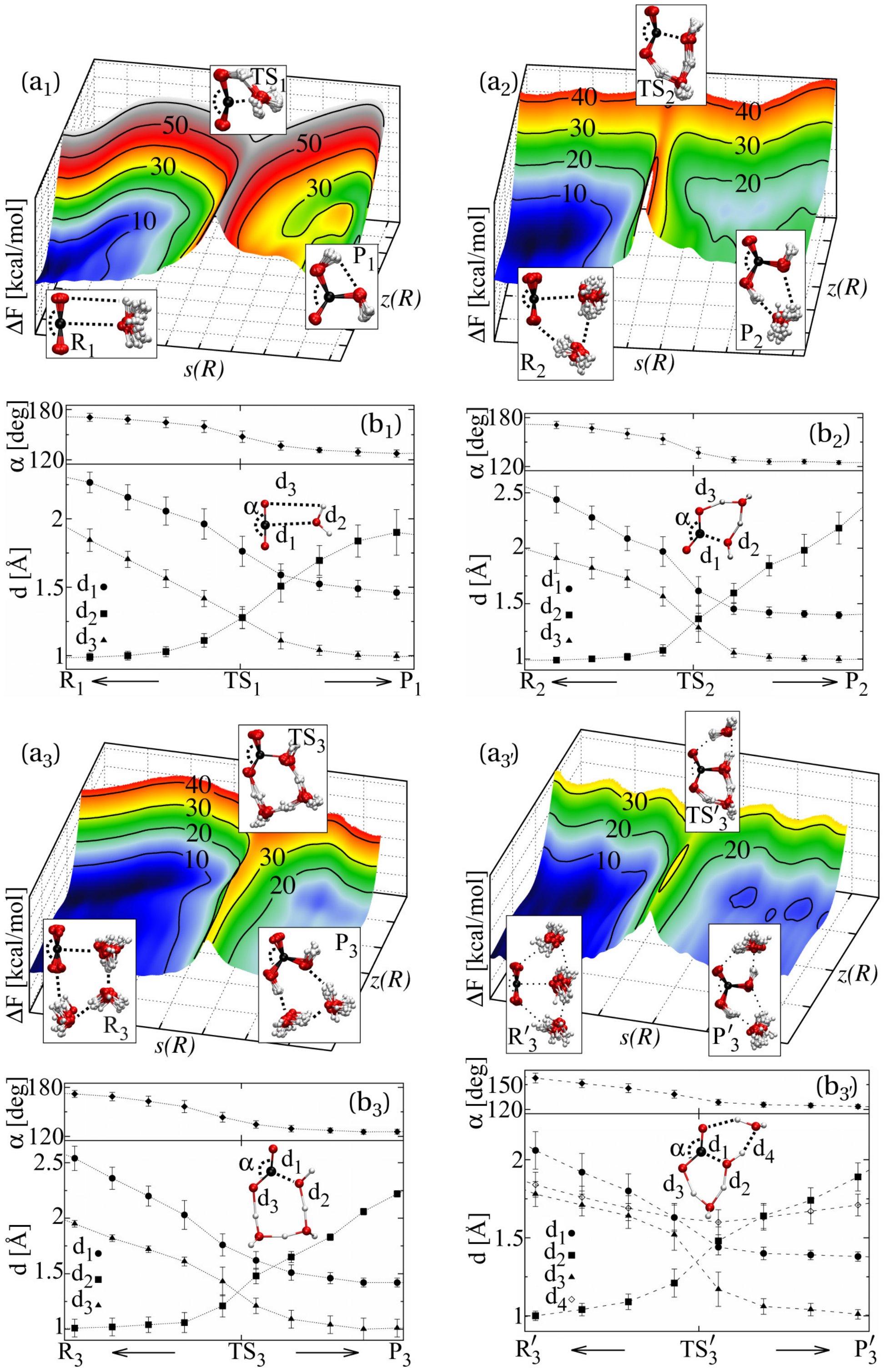


Figure 4.6 – (a_n) Free energy surfaces for the reactions in Eq. 4.1 and (b_n) evolutions of relevant structural parameters along the reaction pathway (BLYP). $n = 1, 2, 3, 3'$. Distance in Å and angles in degrees. Error bars denote the RMSD.

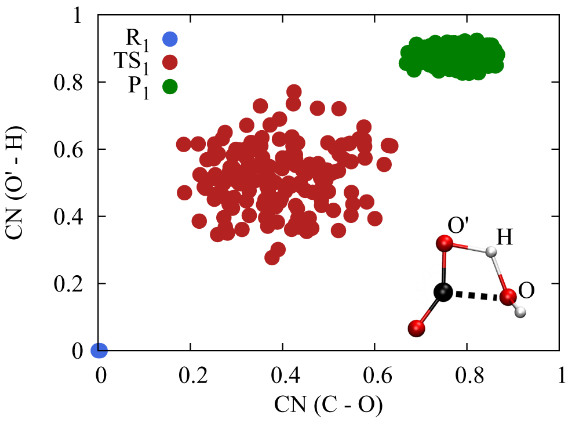


Figure 4.7 – BLYP reactants, products and transition identified during our simulation and plotted in the CV space used in Ref. [130].

Previous metadynamics simulations of the dissociation of carbonic acid were published in Ref. [130]. The authors reported a dissociation free energy barrier of 37 kcal/mol in the BLYP xc functional and the difference of 4 kcal/mol with respect to our result can easily be traced back to the different pseudopotential scheme or box size. In addition the number they report is computed as a first passage barrier passage and the simulation does not include a formation reaction.

In fact, their choice of collective variables (coordination of $O_{CO_2} - H_{H_2O}$ and of $O_{H_2O} - H_{H_2O}$; Eq 2.74), although intuitively sound, is flawed for the simulations of H_2CO_3 formation. To illustrate this statement we refer to Figure 4.7 in which our reactant R_1, transition state TS_1 and product P_1 configurations are reported in their CV space: the TS occupies almost a quarter of the total CV space while the product basins shrinks and the reactants collapse to a single point whence $\partial_q W = 0$ (Eq. 2.90) in this region. Carrying out a metadynamics simulation there is ineffective because the forces from the bias are null and therefore the system is not driven towards the transition state.

Car-Parrinello KS energies of the structures in the basins (as reported along the MTD run) were used to estimate the enthalpy (Eq. 2.69). From calculations on 13000 configurations extracted from the trajectories at room temperature, a difference of ~ 2 kcal/mol on average was noted between the CP-KS energies and the "true" KS energies obtained after quenching the electronic degrees freedom to the BO surface

The remarkable agreement between the enthalpy barriers (in Table 4.5) and the corresponding static value (in square brackets) gives us confidence in ascribing the difference between free energy and enthalpy barriers to entropic effects. We note that the static values are reported without ZPE contributions because ions are treated classically within our MD scheme.

Entropy contributions to the free energy barriers could be directly calculated by $T\Delta S = \Delta H - \Delta F$ (Eq. 2.70) without having to separate its contribution into vibrational, rotational and translational entropy as it is customarily done to estimate entropy form the PES (see, *e.g.*,

Table 4.5 – Enthalpy barrier height (kcal/mol) for the formation and dissociation of carbonic acid estimated from room temperature DFT-MD and decrease (in italics) due to the addition of one water molecule[a] (from n to $n+1$).

	Formation			Dissociation		
n	PBE	BLYP	B3LYP	PBE	BLYP	B3LYP
1	42 [41]	48 [47]	50 [50]	33 [34]	34 [35]	38 [41]
2	24 [23]	31 [30]	31 [31]	17 [17]	19 [16]	22 [17]
1 − 2	*18* [*18*]	*17* [*18*]	*20* [*19*]	*16* [*17*]	*15* [*16*]	*16* [*17*]
3	20 [18]	28 [26]		13 [14]	17 [16]	
2 − 3	*4* [*4*]	*3* [*4*]		*4* [*4*]	*2* [*4*]	
$3'$	16 [15]	24 [23]		19 [17]	18 [20]	
2 − 3′	*8* [*8*]	*8* [*6*]		*−2* [abs < *1*]	*1* [abs < *1*]	

[a] In square brackets: results from the PES. Cutoff 70 Ry; no ZPE

Table 4.6 – Entropy contribution ($T\Delta S$) (kcal/mol) to the free-energy barriers at room-temperature. MD: estimate from MTD and Eq. 2.70. H: from PES, vibrational component computed in the harmonic approximation (see Ref. [131]).

	Formation						Dissociation					
	PBE		BLYP		B3LYP		PBE		BLYP		B3LYP	
n	MD	H	MD	H	MD	H	MD	H	MD	H	MD	H
1	−4	−4	−3	−3	−4	−3	2	1	1	1	1	1
2	−5	−6	−1	−5	abs < 1	−4	abs < 1	−1	−1	−1	3	−1
3	−6	−7	−3	−5			−3	−1	−2	−1		
$3'$	−2	−8	−2	−7			4	−1	2	−1		

Ref [131]). The comparison of results from the two methods is possible as was the comparison of finite temperature enthalpy with the Kohn-Sham energy from the PES. Results are reported in Table 4.6; the estimated error on the MD values is 3 kcal/mol. The vibrational entropy is computed within the harmonic approximation, *i.e.*, the modes are populated according to a Boltzmann distribution since the ions are treated classically.

The dynamical reactants R$'_3$ (and to a lesser extent the products P$'_3$) exhibit sizable deviations from the harmonic entropy (~ 6 kcal/mol; beyond our error bars of 3 kcal/mol). This fact, in addition to the discrepancy noted in the structural description of the reactants R$_3$, suggest that the sole study of the potential energy surface is insufficient to faithfully describe the reaction mechanisms for $n \geq 3$.

Temperature dependence of the entropic term was also investigated by carrying out an additional simulation at $T = 500$ K for the $n = 1$ case. Within our accuracy of 3 kcal/mol, the entropic contribution to the free energy dissociation barrier remained the same while it in-

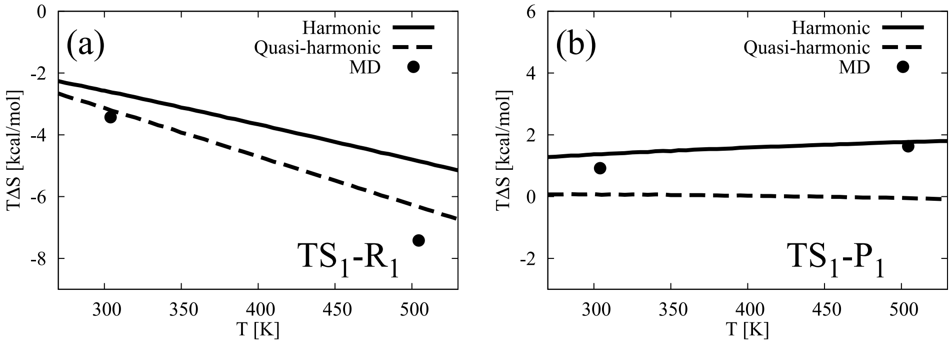


Figure 4.8 – Entropic contribution (in kcal/mol) to free-energy barriers (temperature in K on the x-axis): from *ab initio* MD (circles) and from standard calculation using harmonic (solid) and quasi-harmonic (dashed) approximations for the vibrational entropy.

creased it significantly (6 kcal/mol) for the formation. Entropy tends to stabilize the reactants which is consistent with the very shallow nature of this minimum.

4.1.3 Conclusions

In this section we have uncovered interesting aspects of the formation and dissociation reactions of carbonic acid in the gas phase. In particular, we have shown that the catalytic effect of water addition is quantitatively independent on the xc functional and is in agreement with the results from previous quantum chemistry calculations [106, 128]. This remained true, to a large extent, also at room temperature.

Moreover, comparison of the potential and free energy surfaces of the processes revealed that finite temperature effects were limited on the reactions with one or two water molecules. In the case of CO_2 interacting with three water molecules, substantial changes were identified: notably the reactant basin cannot be described as a single configuration but as a statistical ensemble of quasi-degenerate configurations.

The rapid decrease of the catalytic effect and of the free energy barrier height towards the experimental value in solution [63] is however not sufficient to conclude that the chemistry of solvated CO_2 is accurately described by the inclusion of only a few water molecules. Indeed, experiments [63] and standard molecular dynamics simulations with metadynamics have shown that the formation of carbonic acid may proceed via more complex pathways in liquid water. [132, 133, 134].

Finally, we also verified that the protocol introduced in Section 3.1 produces robust and reproducible results.

4.2 Carbon Dioxide - Monoethanolamine Reactions in Liquid Water

Computational details can be found in Appendix A.3.

The formation of a zwitterion (a neutral molecule with both a negative and positive charge) composed of a CO_2 molecule (carrier of the negative charge) bound to a monoethanolamine (MEA) molecule (carrier of the positive charge) is a long postulated intermediate in the reaction from solvated CO_2-MEA to stable carbamate in solution [107]:

$$\underbrace{HOCH_2CH_2NH_2}_{\text{MEA}} + CO_{2\,(aq.)} \overset{CO_2\text{ release}}{\leftrightarrow} \underbrace{HOCH_2CH_2NH_2CO_2}_{\text{zwitterion}} \overset{\text{deprotonation}}{\leftrightarrow} \underbrace{HOCH_2CH_2NHCO_2^-}_{\text{carbamate}} + H_3O^+. \tag{4.3}$$

The question of the zwitterion stability emerges because its very existence eludes experimental confirmation. Two possible routes can be envisaged for its decomposition: it can either release CO_2 and go back to the MEA (left-hand side of Eq. 4.3) or lose one of the protons attached to the nitrogen and form a carbamate (right-hand side of Eq. 4.3).

We investigated the dependence of these decomposition mechanisms in liquid water by means of *ab initio* MD including explicitly 122 water molecules.

The description of liquid water by commonly used GGA functionals such as PBE and BLYP was thoroughly investigated by many authors and it is an established fact that both fail at reproducing correct physical properties (see, *e.g.*, Ref. [135, 136, 137, 138] and references therein). For instance they tend to underestimate the equilibrium density by 10-20% and overstructure the liquid (radial distribution functions at 300 K have too high peaks and too shallow wells). The inclusion of van der Waals effects to the functionals was shown to mitigate these issues [139, 140, 141]. In particular, the empirical scheme presented in Section 2.2.1 softens the structure and improves the density to almost match the experimental datum.

In our Car-Parrinello simulations, the nuclei are treated as classical objects, and therefore we neglect the effects of their quantum nature. Indeed, the inclusion of nuclear quantum effects can be done using the path-integral method [142]. In this case, the radial distribution functions of liquid water become less structured and the momentum distribution of the proton becomes broader [143] in closer agreement with experimental data. In the case of ice, a more recent study [144] has identified the zero-point motion as the major quantum contribution to the dynamics of the proton.

In order to address the issue of the zwitterion existence in solution (Figure 4.9) we simulated its evolution at finite temperature and on our timescale (~ 20 ps) it remained whole.

The relative stability of the zwitterion within the *ab initio* MD framework implies the use of enhanced sampling schemes (Section 2.4) to facilitate the zwitterion decomposition and we employed both metadynamics and umbrella sampling .

Free energy profiles for the CO_2 release and deprotonation are reported in Figure 4.10 (a) and (b) respectively; the inset shows a snapshot of the transition state region with the distance used as collective variable indicated in yellow. The free energy barriers corresponding to the first

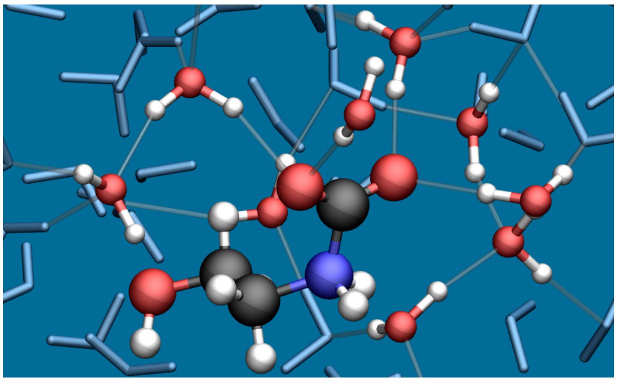
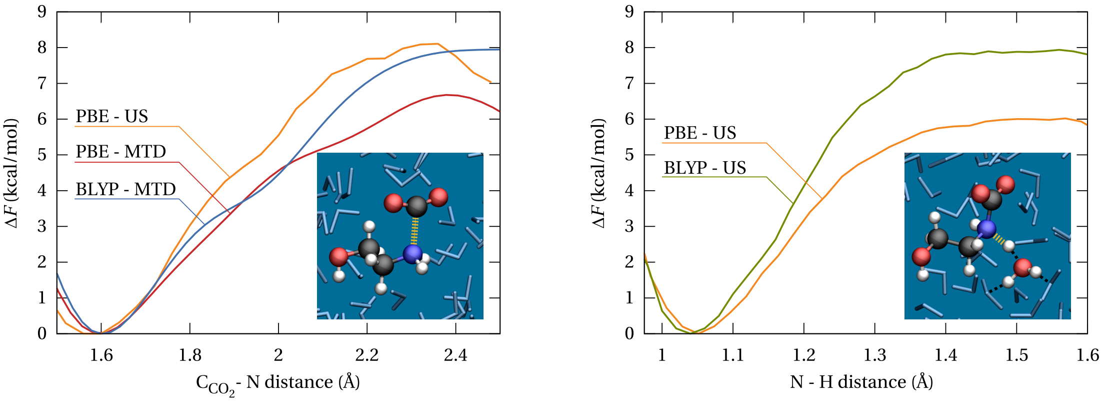


Figure 4.9 – The zwitterion formed from a monoethanolamine molecule and CO_2 molecule solvated by water. Black balls are carbon, red balls: oxygen, blue ball: nitrogen, white balls: hydrogen and blue sticks are further water molecules.



(a) CO_2 release

(b) Formation of the carbamate via N deprotonation

Figure 4.10 – Free energy profiles for the two decomposition routes of the MEA-CO_2 zwitterion. Only the first passage barrier is reported. Inset: snapshot at the transition state region; the dashed yellow line indicates the distance used as collective variable in the enhanced sampling scheme.

passage for both reactions were found to be in the range of 6-8 kcal/mol. Therefore they are competitive. The agreement between the results obtained with the two different functionals and also between metadynamics and umbrella sampling was remarkable: the free energy barriers were the same within 2 kcal/mol, namely, within the accuracy of the calculations. The enthalpic barriers, calculated from Eq. 2.69, turned out to be 24 ± 3 kcal/mol and 13 ± 3 kcal/mol for CO_2 release and deprotonation respectively. These values indicate substantial entropic contributions of ~ 17 kcal/mol and ~ 6 kcal/mol respectively that reduce the free energy barriers at room temperature.

These large entropic effects come from the reorganization of the solvent and the use of explicit water molecules allows us to reveal its importance in the reactions. We can, for instance, provide a detailed analysis of the modifications undergone by the hydrogen bond network when carbon dioxide is released or when the carbamate is formed. We adopt the definition of Luzar *et al.* for the hydrogen bond [145].

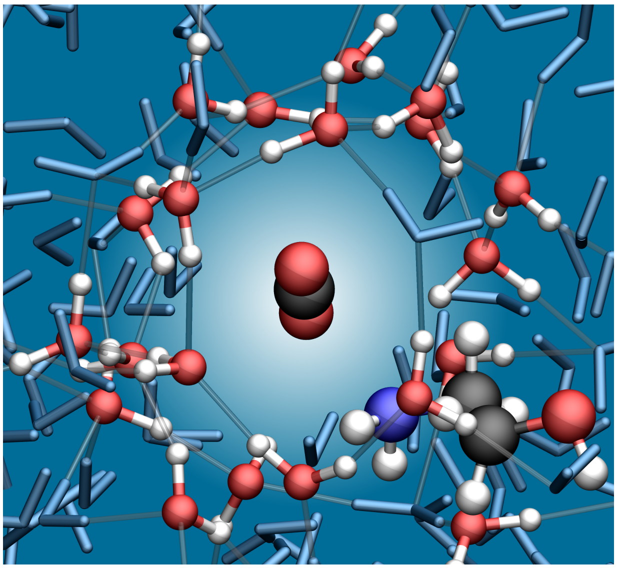
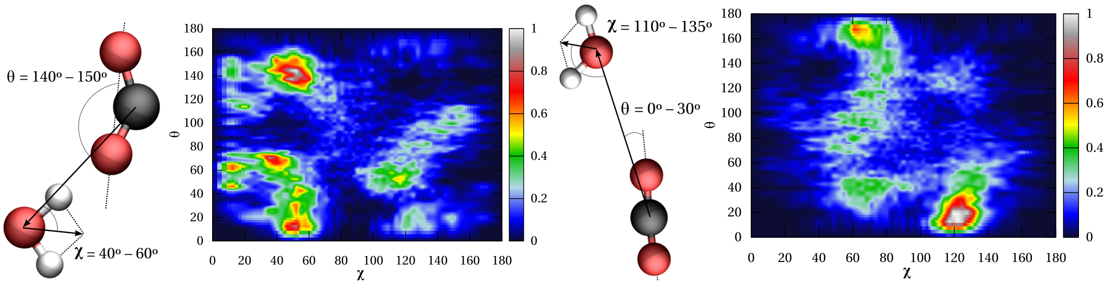


Figure 4.11 – The water cage forming around CO_2. The hydrogen bond network was significantly modified with respect to the reactants in Figure 4.9.



(a) $\chi-\theta$ distribution in the reactants

(b) $\chi-\theta$ distribution at the transition state

Figure 4.12 – Definition of the θ and χ (following [147]) angles and their distribution in the reactants (a) and the transition state (b). All water in the first shell are included.

When captured by the amine, the CO_2 molecule is bent (140°) and forms 3 ± 0.7 hydrogen bonds on average with the surrounding water molecules (Figure 4.9). As it is released it progressively loses them: 1 ± 0.7 remain at the transition state and they vanish in the product. Free CO_2 is linear and repels water (see Figure 4.11). The caging effect is further confirmed by the orientation of the water molecules surrounding it (Figure 4.12). This particular distribution of $\theta-\chi$ angles has previously been related to hydrophobicity [146, 147].

The changes in the solvent are not confined to the sole interaction with CO_2. We noted that, on passing from the reactant to the transition state, the number of hydrogen bonds within the first solvation shell increased from 0.4 ± 0.2 to 1 ± 0.1 on average per water molecule while it decreased from 2.3 ± 0.3 to 1.5 ± 0.2 for hydrogen bonds between first and second shells.

The solvent reorganisation is an interesting mechanism to investigate with the permutation

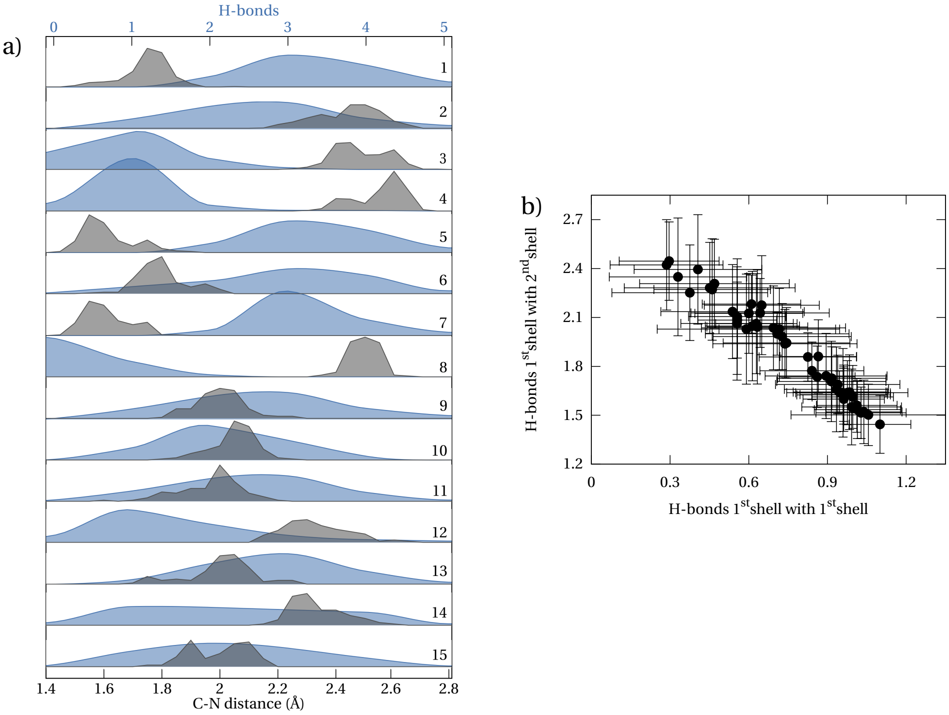


Figure 4.13 – (a) distribution of carbon-nitrogen distance (black lines) for the 15 largest clusters, compared with a smoothed out distribution of the discrete number of hydrogen bonds (defined according to Ref. [145]) between CO_2 and the surrounding water molecules (blue lines). (b) Number (average and RMSD) of water-water hydrogen bonds per molecule within the first solvation shell and between the first and second shell for each of the 50 clusters.

invariant vector and the cluster analysis introduced in Section 3.2. From the umbrella sampling simulations of the CO_2 release we extracted thousands of configurations and applied the k-medoids algorithm to obtain 50 clusters from a PIV focused on a range of 1-5 Å (yellow curve in Figure 3.5). Figure 4.13 (a) shows a comparison of the distribution of the $C_{CO_2} - N_{MEA}$ distance (black line) with the smoothed out distribution of the number of hydrogen bonds between CO_2 and water molecules (blue line) for the fifteen largest clusters. There is an anticorrelation between increasing C-N distance and the number of hydrogen bonds in the clusters and the limited width of the distributions indicate that our metric is quite successful in extracting clusters whose member configurations are structurally consistent. The PIV also captures the changes beyond the first solvation shell: the anticorrelation between the number of hydrogen bonds within the first shell and the number of hydrogen bonds between the first and second shell is well reproduced (Figure 4.13 (b)).

During the deprotonation reaction, one water molecule directly participates in the breaking of covalent bonds because the proton leaving the NH_2 site is transferred to it to form a hydronium cation (H_3O^+). Substantial changes also take place in the hydrogen bond network surrounding the proton-accepting water molecule. For instance the number of solvent molecules forming a hydrogen bond with it decreases from 2.9 ± 0.7 to 2.2 ± 0.5 on average while the bonds themselves are strengthened (H-bond distance goes from 1.82 Å to 1.62 Å). Diffusion of the proton in water was not investigated further as it is *per se* a delicate issue [148, 149].

4.2.1 Conclusion

Our calculations indicate that the zwitterion is metastable in solution and could act as an intermediate in the formation reaction of a carbamate from monoethanolamine and carbon dioxide.

Both mechanisms inducing the dissociation of the zwitterion, namely the CO_2 release and nitrogen deprotonation, involve a non-trivial contribution from the solvent molecules, also beyond the first solvation shell. This effect largely explains the sizable entropic contribution to the free energy barrier favoring the decomposition of the zwitterion.

Our results also show that one cannot justify the use of either implicit solvent models [110] reducing the effect of the solvent to a change in the electrostatic potential or of QM/MM schemes including only a very small number of water molecules (hybrid explicit-implicit models [111]).

Therefore, although computationally much more expensive, the explicit treatment of the solvent appears to be crucial for an understanding of the mechanisms responsible for the reactions studied in this work and more generally for the simulation of CO_2 capture in any aqueous solutions.

4.3 Impurity Diffusion in Cadmium Telluride

In this section we report on an investigation of the diffusion of sulfur in bulk CdTe, at low concentration. First, we describe relevant impurity configurations and characterize the signature of sulfur on the Kohn-Sham spectrum. The former were obtained either with geometry optimization or from the trajectories of *ab initio* molecular dynamics. Then, we will discuss the results of metadynamics as well as NEB simulations.

Under standard conditions cadmium telluride adopts a zincblende structure and is a semiconductor with a 1.50 eV direct band gap at room temperature [112].

Our model is a $3 \times 3 \times 3$ cubic supercell (Figure 4.14) (216 atoms). Simulations were carried out within the PBE xc functional scheme; in order to estimate the dependence of the calculated quantities on the xc functional, we also performed some calculations using the PBE0 functional. As usual for covalent systems, we find that PBE predicts a lower density (by 7%) and softer phonons than the real system. In this case, the experimental spectrum [150] taken from inelastic neutron diffraction at room temperature, ranges from 3.7 (30) to 19.2 (155) meV (cm^{-1}). The vibrational density of states (VDOS) calculated at $T = 0$ (harmonic) appears

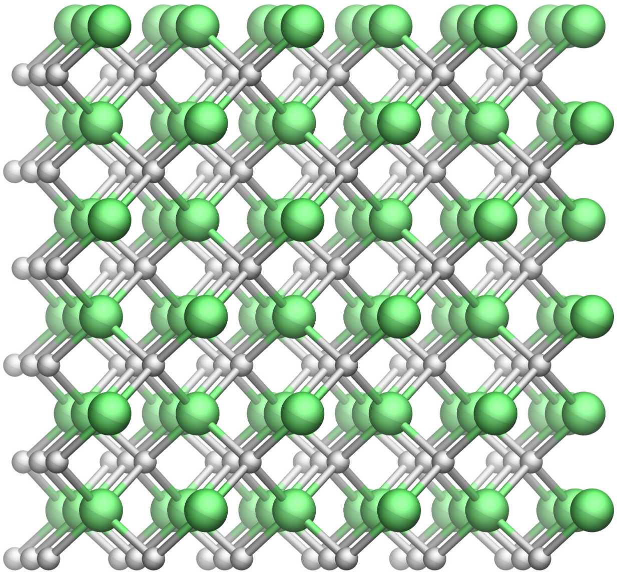


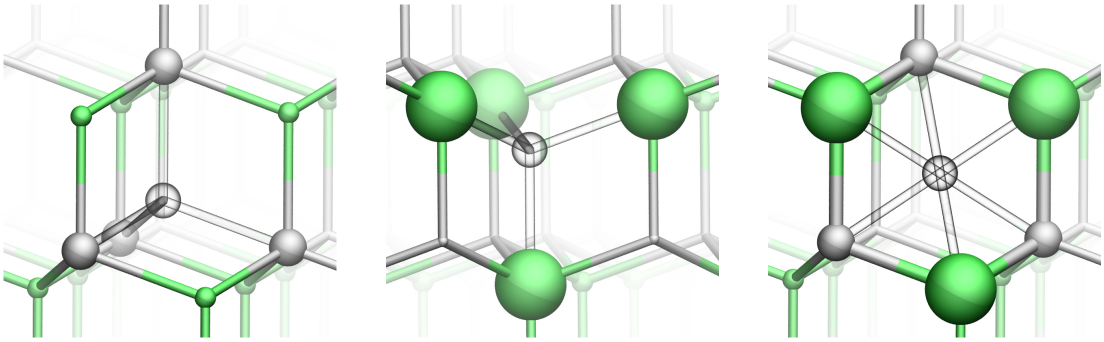
Figure 4.14 – CdTe. The supercell model. Silver and green spheres represent the cadmium and tellurium atoms respectively.

to match perfectly with experiment but this is only an accident due to the softening from PBE. The room temperature VDOS (accounting also for anharmonicity) is shifted down by $\sim 30\ \mathrm{cm}^{-1}$ (~ 4 meV). This is the same shift between experiment and the room temperature VDOS. Therefore the apparent red-shift of the theory is a consequence of the PBE underestimate, and the temperature effect is relevant. This may imply that sulfur diffusion is facilitated in the PBE model. Comparison of PBE results for structural and vibrational properties with experiment is reported in the Appendix B. It is also well known that PBE strongly underestimates the energy gap, namely from ~ 1.5 to 0.5 eV. We analyze the dependence of our results on the xc functional, through additional calculations made within the hybrid PBE0 functional.

4.3.1 Relevant Impurity Configurations

The zincblende structure has high symmetry sites available to an impurity atom: the centers of the two tetrahedrons formed by four Cd (T(Cd)) or four Te (T(Te)) (Figure 4.15 (a) and (b) respectively) and the center (M) of the hexagon formed by three Cd and three Te atoms (Figure 4.15 (c)). These geometries are generally considered relevant for the diffusion in zincblende crystals, and in particular also in recent calculations of noble metal impurities in CdTe [122].

By allowing relaxation of all Cd and Te coordinates but constraining the sulfur atom at the symmetrical positions, we obtain a pseudo-energy-minimum only for the T(Cd) and T(Te) geometries, but 1.3 eV and 1.6 eV higher than the ground state respectively. Once also the S coordinates are allowed to relax, in any of these cases, the system ends in the same structure, with a S-Te$^{(1)}$ dumbbell (inset of Figure 4.16. This will be referred to as the *A* state. There, the S-Te$^{(1)}$ bond-length is 2.45 Å and Te$^{(1)}$ is displaced from its lattice site by 1.17 Å. The S-Te$^{(1)}$ bond is at an angle α of 38° with respect to the direction of the Cd-Te bond in the lattice. As can be seen from the displacement plot in Figure 4.16 the presence of the S impurity causes



(a) Site at the center of a Cd tetrahedron T(Cd)

(b) Site at the center of a Te tetrahedron T(Te)

(c) Hexagonal (M) site

Figure 4.15 – Three high-symmetry empty sites in a zincblende crystal.

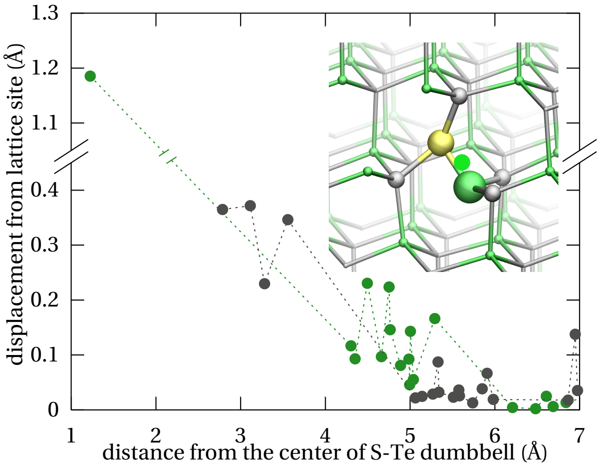


Figure 4.16 – CdTe:S - *A* state. Structure and displacement from the lattice sites of Cd and Te atoms as a function of their distance from the S-Te center. Colors are the same as in Figure 4.14; sulfur is represented by a yellow sphere. The ideal Te lattice site is indicated by a green circle.

sizable local changes, *e.g.*, the Cd nearest neighbors are displaced by ~ 0.4 Å.
The signature of the S-Te$^{(1)}$ pair can be seen in the Kohn-Sham spectrum by the emergence of two levels near the top of the valence band, split by 0.2eV. These correspond to the combination of p-states on S and Te, localized at the S-Te pair.

When we introduced an additional sulfur atom, we observed that it formed a weakly bonded S dimer (Figure 4.18 (a)) (2.07 Å bond-length) at close distance from Te (2.47 Å), respectively. We call it "S dimer". This configuration is favored with respect to having two separate S-Te dumbbells (Figure 4.18 (b)) by only 0.1 eV. The signature of two separate S-Te dumbbells on the KS spectrum does not differ from that of the single pair. The "S dimer", instead, gives rise to a localized state in the gap split by 0.2 eV from the top of the Te valence band.

Our MD simulations (see below) revealed that another state (*B*) is particularly important. There, S is a substitutional impurity at 0.18 Å from the Te lattice site and a Te-Te dumbbell

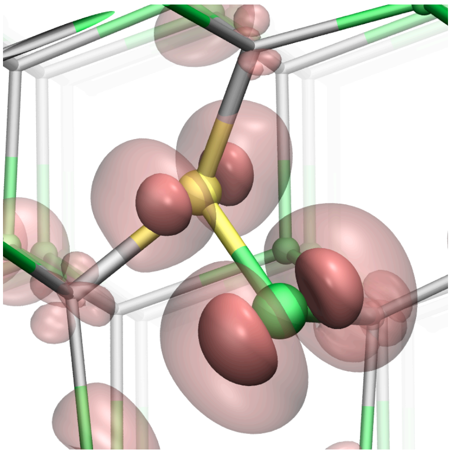
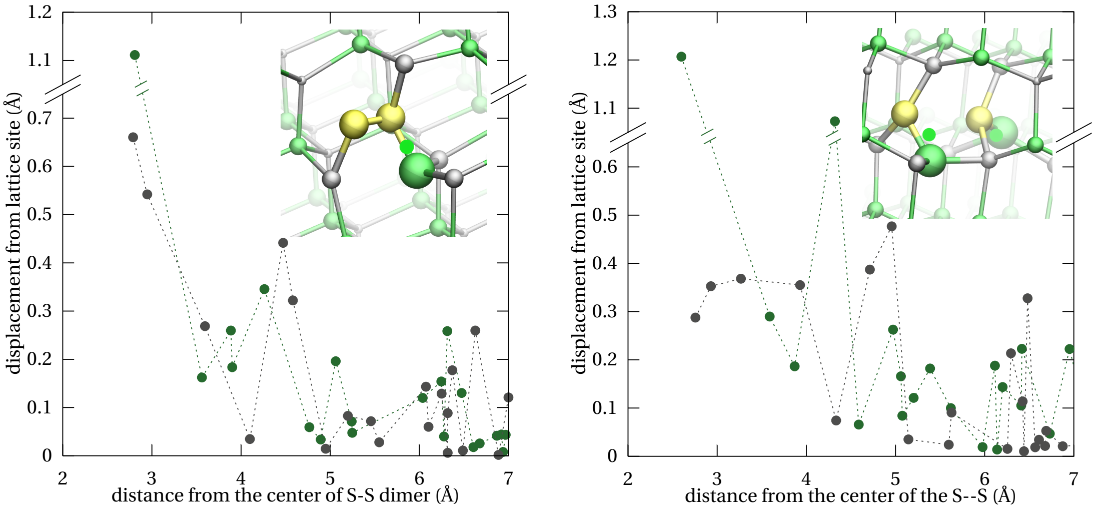


Figure 4.17 – CdTe:S *A* state. Probability density corresponding to an impurity state (see text). Isosurface values of 0.001 (light pink) and 0.01 (solid pink).



(a) A S-S dimer in CdTe

(b) Two S-Te dumbbells loctated on different Te sites

Figure 4.18 – The case of two sulfur atoms. Structure and displacement from the lattice sites of Cd and Te atoms as a function of their distance the center of the S-S distance vector. Colors as in Figure 4.16.

forms around another Te lattice site (inset of Figure 4.19), separated by 2.78 Å. B is higher than A by 0.4 eV. The displacement plot in Figure 4.19 shows again substantial changes in the local environment around the center of the Te-Te pair, especially the Cd nearest neighbors (up to 0.5 Å).

The analysis of the KS states shows a picture much like *A*: two localized states on the Te-Te pair in the valence band, split by 0.2 eV, of which one 0.2 eV above the valence band (Figure 4.20). No other significant changes are shown otherwise in the valence band.

In summary, we remark that, apart from the formation of new localized impurity states, the variations induced in the valence band by the presence of the sulfur atom are minor. PBE0

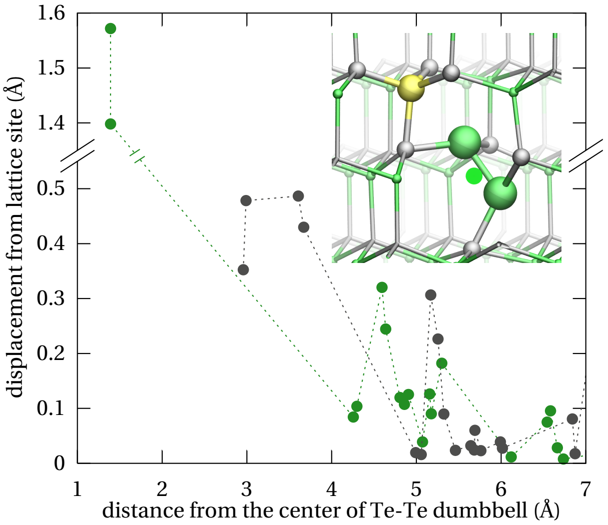


Figure 4.19 – CdTe:S *B* state - S trapping. Structure and displacement from the lattice sites of Cd and Te atoms as a function of their distance from the center of the Te-Te dumbbell. Colors as in Figure 4.16.

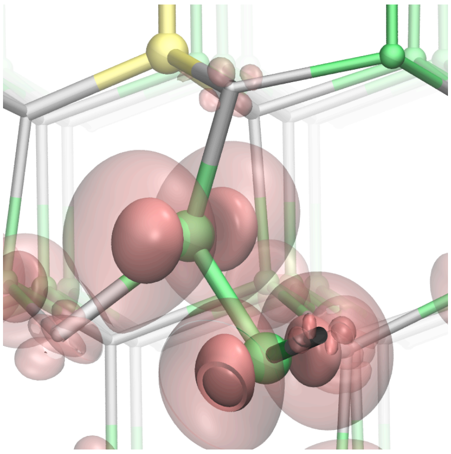


Figure 4.20 – CdTe:S *B* state. Probability density corresponding to an impurity state (see text). Isosurface values of 0.001 (light pink) and 0.01 (solid pink).

applied to the PBE geometries did not alter the KS electronic structure and also predicted the same changes of the KS levels from one state to the other, within the accuracy of our calculations. Also the decrease of the energy gap does not change in the two xc schemes : it is negligible in the three cases considered above.
Major effects can be seen, instead, on the position of the $3s$ state of sulfur, which is split from the bottom of the $5s$ Te band by ~ 2 eV when hosted as interstitial impurity (A) and ~ 1 eV when substitutional (B).

4.3.2 Sulfur and Tellurium Migration from Metadynamics

Path-metadynamics assumed as reference the trajectory obtained from simple NEB simulations either between A (Figure 4.16) and an equivalent state (S-Te dumbbell) about a different Te site or between states A and B (Figure 4.19). However, it went away from it and always found new pathways. Starting from the A configuration and heating up the system to room temperature for 1 ps, the changes were limited; in particular α is $38 \pm 5°$. Successive MTD

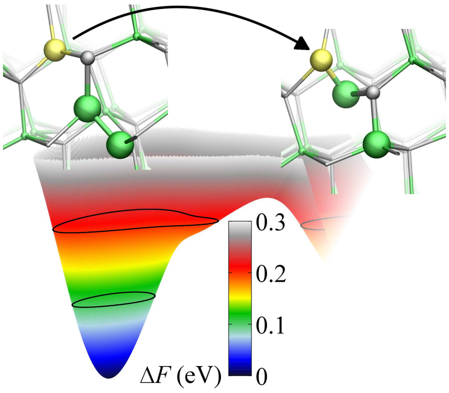


Figure 4.21 – CdTe:S. Reconstructed free energy surface for the process $A \rightarrow B$. Estimate of the barrier fro the first passage.

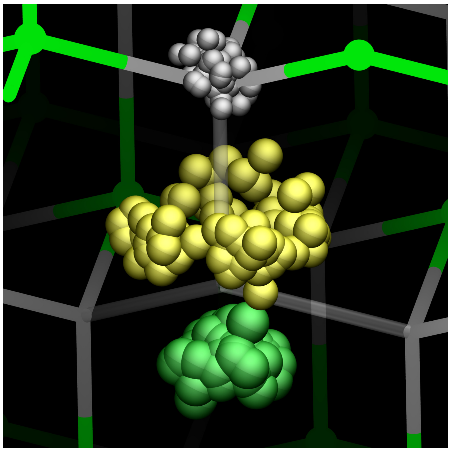


Figure 4.22 – CdTe:S. The configurations explored by an equilibrium MD, started at state A.

revealed an unforeseen scenario, in which S replaces Te in the lattice and Te starts to diffuse via a kick-out mechanism. The first step of this process (A to B) has a ΔF of 0.3 ± 0.2 eV. The second step (Te kick-out by Te) was barrierless.

In order to study the diffusion of S, the A to B path had to be prevented by constraining the Te atoms, closest to the Te forming the S-Te dumbbell, to the center of their respective Cd cage. In this case room-temperature unbiased CP-MD (20 ps) explored a larger basin for the "reactant" which can be described as an ensemble of A states with the S-Te$^{(1)}$ pair jumping from one geometry to another equivalent with α ranging from 25 to 50.
Then, the reconstruction of the FES resulted in a free energy barrier of $\Delta F = 0.6$ eV $+/-0.2$ (Figure 4.23) which we can consider higher than the one relative to the $A \rightarrow B$ mechanism. The FES in Figures 4.21 and 4.23 are reconstructed after the first transition, *i.e.*, the free energy of the end state was not estimated.

4.3.3 *Post*-Metadynamics NEB calculations

For the sake of comparison, we have performed NEB search starting from the trajectories resulting from the above simulations. The final NEB pathway remained close to the initial

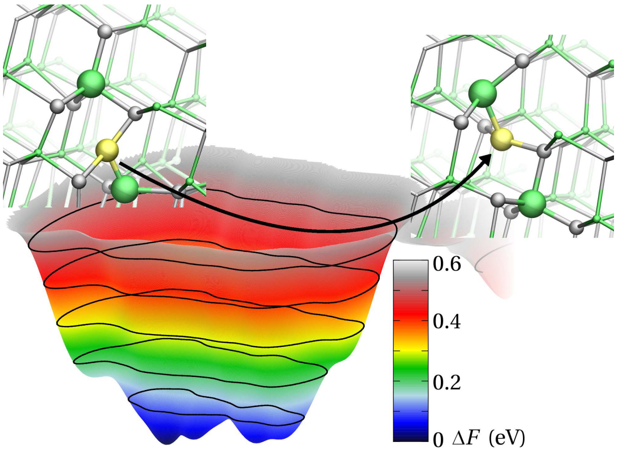


Figure 4.23 – CdTe:S. Reconstructed free energy surface for the diffusion of sulfur. Estimate of the barrier fro the first passage.

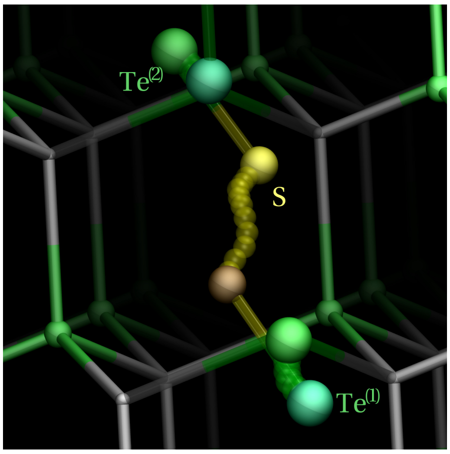


Figure 4.24 – CdTe:S. The path followed by sulfur to go from A to an equivalent state around a different Te lattice site ($S - Te^{(1)} \to S - Te^{(2)}$). Color as in Figure 4.16.

guess in both cases studied.

For the $A \to B$, the energy barrier turns out to be 0.4 eV, namely close to our estimate for the FEB. Within the uncertainty of our computations, we cannot assess the effect of temperature in this case.

For the case of diffusion from one cell to the next, instead, the energy barrier (following the path in Figure 4.24) was 0.2 eV, which can be considered somewhat lower than the FEB (~ 0.6 eV +/- 0.2 eV). One possible explanation is that in this case the ensemble of A states describing the "reactant basin" had a larger spread in configuration space (Figure 4.22) (for example α changed from 25 to 50 degrees) so as to stabilize it through entropic effects.

As expected, at the "transition state" (Figure 4.25), S is equidistant (2.70 Å) from the two Te, which are displaced by 0.45 Å from their respective lattice sites. The nearest Cd is at a distance of 2.42 Å from S and it relaxes away from it by 0.45 Å with respect to its ideal site.

4.3.4 Conclusions

The study presented here has revealed unforeseen characteristics of the CdTe system and its interaction with sulfur. In particular, it was clearly shown that S has a tendency to be trapped in

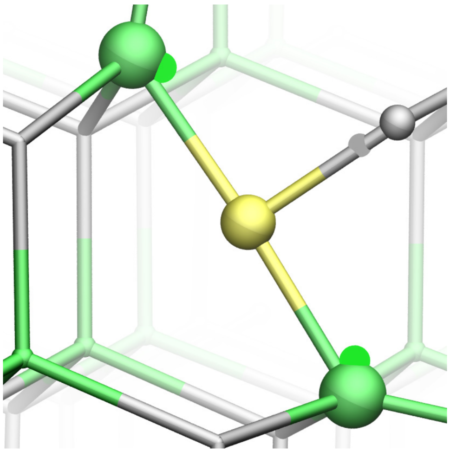


Figure 4.25 – CdTe:S. Snapshot showing a configuration on the PES near the "transition state". Colors as in Figure 4.16; a gray dot indicates the Cd lattice site.

lattice sites, replacing Te. Moreover we could compare the output of two widely used methods like NEB and MTD. NEB failed to be predictive. However, once used as a post-MTD tool, it gave similar quantitative results as MTD and helped understand the effects of temperature.

It is not possible to reconcile these results with experiment yet. Scattered information exists on thin films, in which CdTe is polycrystalline and grain-boundaries make the S concentration profile very inhomogeneous [151, 152]. In bulk CdTe, instead, SIMS measurements of the concentration profile are available at high temperature ($372° - 675°$ C) [153, 154]. There, the (traditional) analysis resulted in two slopes for the Arrhenius plots of the diffusion constant as a function of $1/T$, indicating that two different mechanisms were dominating, depending on temperature, with activation barriers of 1.0 ± 0.04 and 1.7 ± 0.2 eV.
We remark that the model we adopted for CdTe is "ideal" namely no defects were introduced and temperature is low. This makes comparison with experiment not possible yet. Indeed defects are pervasive in CdTe and not easily controllable. It is known that Cd vacancies and Te at the anti-site are the most common native defects. Preliminary results obtained with MTD on CdTe samples including Cd vacancies clearly indicated that they help S trapping. Indeed the ground state corresponds to a substitutional S and a Te in the antisite, which requires overcoming a free energy barrier of the order of 1 eV to escape from it. These calculations were performed in the GGA-PBE approximation. In this case, however, we expect that the use of hybrid functionals is important for a quantitative assessment of diffusion processes, because exact exchange tends to modify the localization of the electronic states in defective structures. This is the on-going step of this research.

A clear message from our simulations so far is that S diffusion is not determined by one single mechanism and the influence of the defects in the crystal is very important and not trivial. The tendency to be trapped is certainly strong and will tend to slow down the diffusion drastically.

A Computational Details

A.1 Sodium Chloride in Solution

We employed classical molecular dynamics and the Gromacs 4.5.5 code [155] to simulate NaCl in solution. We took the force field parameters from Joung and Cheatham to describe the sodium and chlorine ions [156] and those of the TIP3P model for the water molecules [157]; these parameters had previously been validated by comparison against both experiment [158, 159] and *ab initio* MD [94]. We used the particle mesh Ewald method [160, 161] for computing Coulomb interactions and the cutoff for both short and long range Lennard-Jones interactions was 8 Å. The timestep for the integration of the equations of motion was 0.5 fs.

The system consisted in one Na-Cl pair and 574 water molecules (0.1 M concentration) in a cubic box of 26 Å edge periodically repeated. We equilibrated it at $T = 300$ K and $P = 1$ atm for 1 ns by using the stochastic velocity rescaling thermostat from Bussi *et al.* [162] which efficiently samples the canonical ensemble and Berendsen's barostat [57] both with $\tau = 5$ ps. The production runs consisted in 10 independent unbiased runs of 100 ns each in the NVT ensemble.

As CV, we considered both the Na^+-Cl^- distance and a quantity associated with the solvent around the Na^+ ion: its electrostatic potential energy with the surrounding water molecules (Ω_{H_2O} within a 6.5 Å radius) [7]

$$E^{\mathrm{pot}} = \frac{Q_{\mathrm{Na}^+}}{4\pi\epsilon_0} \sum_{i\in\Omega_{\mathrm{H_2O}}} \left(\frac{Q_{\mathrm{O}}}{||\boldsymbol{R}_{\mathrm{O}_i} - \boldsymbol{R}_{\mathrm{Na}^+}||} + \frac{Q_{\mathrm{H}}}{||\boldsymbol{R}_{\mathrm{H}_i^1} - \boldsymbol{R}_{\mathrm{Na}^+}||} + \frac{Q_{\mathrm{H}}}{||\boldsymbol{R}_{\mathrm{H}_i^1} - \boldsymbol{R}_{\mathrm{Na}^+}||} \right) \tag{A.1}$$

where $Q_{\mathrm{Na}^+} = 1$ e, $Q_{\mathrm{O}} = -0.834$ e, $Q_{\mathrm{H}} = -0.417$ e and ϵ_0 is the vacuum permittivity.

A.2 Carbonic Acid Formation

We employed the GGAs PBE [26] and BLYP [27, 28] and hybrid B3LYP [15, 37, 35] to approximate the c functional (see Section 2.2.1). All the calculations at that level made use of the Troullier-Martins norm conserving pseudopotentials [49] and a plane waves set to expand the

Kohn-Sham orbitals up to a 100 Ry kinetic energy cutoff (see Section 2.2.1). A periodic box of 16 Å was employed and an electrostatic decoupling scheme [163] was used to prevent periodic image interactions. The code was CPMD [8].

Geometries were relaxed employing the L-BFGS algorithm [164, 165, 166, 167] while the transition states were found with the partitioned rational-function-optimizer (P-RFO) [67]; no particular symmetries of either stable or saddle points were enforced. The structures were converged so that the maximum component of the force was below 10^{-4} Ha/bohr Vibrational frequencies were computed by linear response and, in all cases, there were no imaginary frequencies except for the transition state which only had one.

For the sake of comparison we also carried out all electrons calculations with the B3LYP and the meta-hybrid M06 [43] functionals with an aug-cc-PVTZ basis set. The code used in this case was the NWChem software [168].

Metadynamics [7] simulations were carried out with CPMD patched with PLUMED [169] the plug-in for free energy calculations compatible with a number of *ab initio* and classical MD codes. We employed the path CV [64].

For computational efficiency reasons, we reduced the kinetic energy cutoff of the plane waves from 100 Ry for the calculations on the PES to 70 Ry. We checked that this would not change the relative energies by more than 2 kcal/mol (*i.e.*, within our accuracy).

The electrons were given a mass of 500 au, and the equations of motion were integrated with the velocity-Verlet algorithm on a 0.1 fs timestep. The temperature was controlled by the Nosé-Hoover chain thermostat [59, 60, 61]. We quenched the electronic degrees of freedom to the Born-Oppenheimer surface every 1 ps (10’000 MD steps) to ensure that they would not warm up due to the bias from metadynamics.

A.3 Monoethanolamine and Carbon Dioxide

The zwitterion was placed in a box of 15.6 Å edge and surrounded by 122 water molecules (0.46 M solution). DFT was employed with both PBE and BLYP xc functionals empirically corrected for dispersion effects following Grimme’s scheme (D2) [45]. Norm-conserving Troullier-Martins pseudopotentials were used to simulate core-valence interactions. The Kohn-Sham orbitals were represented by a plane wave basis set up to a cutoff of 80 Ry. The code employed was CPMD.

The fictitious electronic mass for CP-MD was 500 a.u., the equation of motions were integrated with a 0.1 fs timestep by the velocity-Verlet algorithm. All hydrogen atoms were replaced by deuterium. The temperature was set to 300 K and controlled by a Nosé-Hoover chain thermostat.

We employed both metadynamics and umbrella sampling [70] to retrieve the free energy profile with PLUMED. The reaction coordinates were either the $C_{CO_2} - N$ distance or the $N - H$ distance for CO_2 release or deprotonation respectively. A supplementary MTD simulation of the deprotonation employing the N – H coordination instead of the distance as CV was also carried out which yielded essentially the same results. During the metadynamics runs, quenching to the Born-Oppenheimer surface was done every picosecond. During the umbrella sampling calculations, statistics in each window was collected over a time frame extending up to 10 ps after equilibration and the free energy profile was retrieved employing the weighted histogram analysis method described in Section 2.4.2 as implemented in the code of Grossfield [170].

A.4 Sulfur Diffusion in Cadmium Telluride

Calculations were carried out at DFT level with CPMD in a cell of 19.92 Å, with Goedecker-Tetter-Hutter pseudopotentials [50] and a kinetic energy cutoff of 80 Ry for the plane wave basis set. The xc functional was PBE. All reported PBE equilibrium geometries were optimized so that the maximum force component on the ions were below 10^{-4} Ha/bohr. The NEB calculations were carried out employing the CPMD-based script developed by Sbraccia [171]. The electronic degrees of freedom were given a fictitious mass of 500 a.u. and the equation of motion were integrated using the velocity-Verlet algorithm on a 0.1 fs timestep. In order to accelerate the dynamics, isotopes in the correct ratio were employed: the mass of cadmium was set to 10.000 a.m.u. while tellurium was 11.352 a.m.u. and sulfur 2.853 a.m.u. The temperature was controlled by a Nosé-Hoover chain thermostat. The electronic degrees of freedom were quenched to the Born-Oppenheimer surface periodically (~ 1 ps).

The various activated events discussed were computed with metadynamics (PLUMED) and the path CV. with a reference path generated either with NEB or from linear interpolation between the equilibrium structures.

B Additional Information: Sulfur diffusion in Cadmium Telluride

The lattice constants and bulk moduli for PBE and PBE0 are determined by interpolation of the calculated values for $E(V)$ (see Figure B.1) using the Murnaghan equation of state

$$E(V) = E_0 + \frac{B_0 \cdot V}{B_0'} \left(\frac{\left(\frac{V_0}{V}\right)^{B_0'}}{B_0' - 1} + 1 \right) - \frac{B_0 V_0}{B_0' - 1} \tag{B.1}$$

where E_0, V_0, B_0 and B_0' respectively are the energy, volume, bulk modulus and its V-derivative at the equilibrium unit cell. The results are $a_0 = 6.64$ Å (6.51 Å) and $B_0 = 35$ GPa (43 GPa) for PBE (PBE0). Experimental values are 6.48 Å and 42 GPa [172].

From unbiased molecular dynamics of bulk CdTe, we can retrieve the vibrational density of states as the Fourier transform of the velocity auto-correlation function (VACF) (Figure B.2 (b)). In practice, the VACF assumes discrete values:

$$c_j = \langle \frac{1}{N} \sum_{I=1}^{N} \boldsymbol{v}_I(0) \cdot \boldsymbol{v}_I(j \cdot \Delta t) \rangle , \quad j = 0, \ldots, m-1 \tag{B.2}$$

since vibrational frequencies are real, the Fourier transform can be computed as a discrete cosine transform

$$\tilde{c}_k = \sum_{j=0}^{m-1} c_j \cos\left(\frac{\pi}{N} \left(j + 1/2 \right) k \right), \quad k = 0, \ldots, m-1 \tag{B.3}$$

It is interesting to compare this finite temperature density of states with the density of phonon states computed within the harmonic approximation at $T = 0$ K (Figure B.2 (a)) employing density functional perturbation theory [173]. Finite temperature effects turn out to be important already at $T = 300$ K resulting in an overall redshift of about 25 cm^{-1}.
We also report the results from neutron diffraction experiments at room temperature [150] (Figure B.2 (c)). Both the overall shape and the peak positions in the low frequency part are

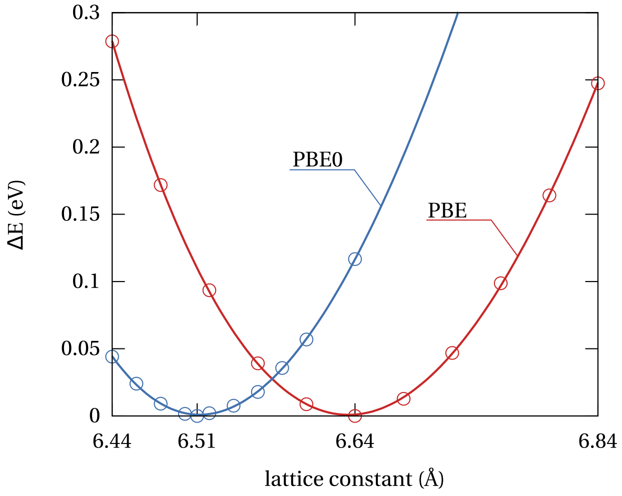


Figure B.1 – PBE (red) and PBE0 (blue) energy per unit cell (minimum set to 0 for each of the xc functional separately) as function of the lattice parameter. Continuous line is a fit to Murnaghan equation (Eq. B.1)

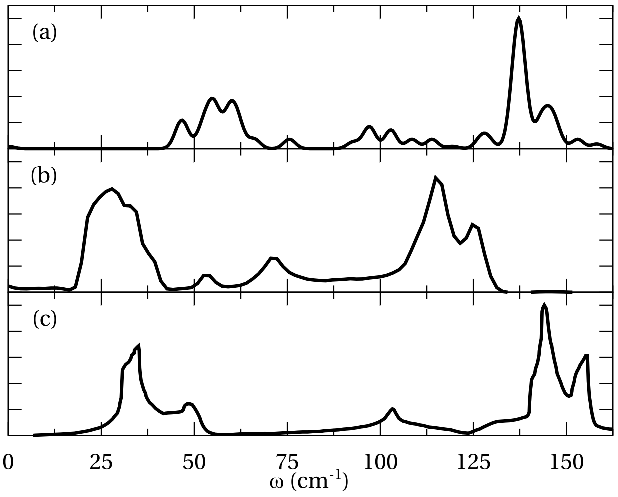


Figure B.2 – Vibrational density of states of bulk CdTe. (a) Harmonic approximation at $T = 0$ K. (b) Fourier transform of the velocity-velocity autocorrelation function at $T = 300$ K. (c) Neutron diffraction experiment at $T = 300$ K, adapted from Ref. [150]

well reproduced by the finite temperature calculation. The latter displays, however, a sizable redshift (about 30 cm^{-1}) in the high frequency part that can be traced back to the softening of bonds by GGA-PBE.

Bibliography

[1] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864, 1964.

[2] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133, 1965.

[3] A. Rahman. Correlations in the motion of atoms in liquid argon. *Phys. Rev.*, 136:A405, 1964.

[4] D. Frenkel and B. Smit. *Understanding Molecular Simulation.* Academic Press, second edition edition, 2002.

[5] R. Car and M. Parrinello. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.*, 55:2471, 1985.

[6] D. Marx and J. Hutter. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods.* Cambridge University Press, 2009.

[7] A. Laio and M. Parrinello. Escaping free-energy minima. *Proc. Natl Acad. Sci USA*, 99:12562, 2002.

[8] CPMD. Copyright IBM Corp 1990–2013 and Copyright MPI für Festköperforschung Stuttgart 1997–2001.

[9] R. M. Martin. *Electronic Structure.* Cambridge University Press, 2004.

[10] J. P. Perdew and K. Schmidt. Jacob's ladder of density functional approximations for the exchange-correlation energy. *AIP Conference Proceedings*, 577:1, 2001.

[11] P. A. M. Dirac. Note on exchange phenomena in the Thomas atom. *Math. Proc. Cambridge Philo. Soc.*, 26:376, 1930.

[12] W. J. Carr. Energy, specific heat, and magnetic properties of the low-density electron gas. *Phys. Rev.*, 122:1437, 1961.

[13] M. Gell-Mann and K. A. Brueckner. Correlation energy of an electron gas at high density. *Phys. Rev.*, 106:364, 1957.

[14] D. M. Ceperley and B. J. Alder. Ground state of the electron gas by a stochastic method. *Phys. Rev. Lett.*, 45:566, 1980.

[15] S. H. Vosko, L. Wilk, and M. Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.*, 58:1200, 1980.

[16] J. P. Perdew and A. Zunger. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B*, 23:5048, 1981.

[17] G. Ortiz and P. Ballone. Correlation energy, structure factor, radial distribution function, and momentum distribution of the spin-polarized uniform electron gas. *Phys. Rev. B*, 50:1391, 1994.

[18] R. O. Jones and O. Gunnarsson. The density functional formalism, its applications and prospects. *Rev. Mod. Phys.*, 61:689, 1989.

[19] S.-K. Ma and K. A. Brueckner. Correlation energy of an electron gas with a slowly varying high density. *Phys. Rev.*, 165:18, 1968.

[20] F. Herman, J. P. Van Dyke, and I. B. Ortenburger. Improved statistical exchange approximation for inhomogeneous many-electron systems. *Phys. Rev. Lett.*, 22:807, 1969.

[21] D. C. Langreth and J. P. Perdew. Theory of nonuniform electronic systems. i. analysis of the gradient approximation and a generalization that works. *Phys. Rev. B*, 21:5469, 1980.

[22] J. P. Perdew. Accurate density functional for the energy: Real-space cutoff of the gradient expansion for the exchange hole. *Phys. Rev. Lett.*, 55:1665, 1985.

[23] J. P. Perdew and Y. Wang. Accurate and simple density functional for the electronic exchange energy: Generalized gradient approximation. *Phys. Rev. B*, 33:8800, 1986.

[24] J. P. Perdew. *Electronic Structure of Solids '91*. Akademie Verlag, Berlin, 1991.

[25] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais. Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Phys. Rev. B*, 46:6671, 1992.

[26] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865, 1996.

[27] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38:3098, 1988.

[28] C. Lee, W. Yang, and R. G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37:785, 1988.

[29] R. Colle and O. Salvetti. Approximate calculation of the correlation energy for the closed shells. *Theor. chim. acta*, 37:329, 1975.

[30] J. P. Perdew, S. Kurth, A. Zupan, and P. Blaha. Accurate density functional with correct formal properties: A step beyond the generalized gradient approximation. *Phys. Rev. Lett.*, 82:2544, 1999.

[31] J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria. Climbing the density functional ladder: Nonempirical meta–generalized gradient approximation designed for molecules and solids. *Phys. Rev. Lett.*, 91:146401, 2003.

[32] Y. Zhao and D. G. Truhlar. A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *J. Chem. Phys.*, 125:194101, 2006.

[33] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, L. A. Constantin, and J. Sun. Workhorse semilocal density functional for condensed matter physics and quantum chemistry. *Phys. Rev. Lett.*, 103:026403, 2009.

[34] C. Adamo and V. Barone. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.*, 110:6158, 1999.

[35] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.*, 98:11623, 1994.

[36] J. P. Perdew, M. Ernzerhof, and K. Burke. Rationale for mixing exact exchange with density functional approximations. *J. Chem. Phys.*, 105:9982, 1996.

[37] A. D. Becke. Density-functional thermochemistry. iii. the role of exact exchange. *J. Chem. Phys.*, 98:5648, 1993.

[38] R. G. Parr and W. Yang. *Density-Functional Theory of Atoms and Molecules.* New York: Oxford University Press, 1989.

[39] U. Lundin and O. Eriksson. Novel method of self-interaction corrections in density functional calculations. *Intl J. Quantum Chem*, 81:247, 2001.

[40] W. M. Temmerman, A. Svane, L. Petit, M. Lüders, P. Strange, and Z. Szotek. Pressure induced valence transitions in f -electron systems. *Phase Transitions*, 80:415, 2007.

[41] W.M. Temmerman, L. Petit, A. Svane, Z. Szotek, M. Lüders, P. Strange, J.B. Staunton, I.D. Hughes, and B.L. Gyorffy. Chapter 241 the dual, localized or band-like, character of the 4f-states. volume 39 of *Handbook on the Physics and Chemistry of Rare Earths*, pages 1 – 112. Elsevier, 2009.

[42] Y. Zhao and D. G. Truhlar. Density functionals with broad applicability in chemistry. *Acc. Chem. Res.*, 41:157, 2008.

[43] Y. Zhao and D. G. Truhlar. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four m06-class functionals and 12 other functionals. *Theor. Chem. Acc.*, 120:215, 2008.

[44] S. Grimme. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J. Comput. Chem.*, 25:1463, 2004.

[45] S. Grimme. Semiempirical gga-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.*, 27:1787, 2006.

[46] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.*, 132:154104, 2010.

[47] P. Jurecka, J. Sponer, J. Cerny, and P. Hobza. Benchmark database of accurate (mp2 and ccsd(t) complete basis set limit) interaction energies of small model complexes, dna base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.*, 8:1985, 2006.

[48] D. R. Hamann, M. Schlüter, and C. Chiang. Norm-conserving pseudopotentials. *Phys. Rev. Lett.*, 43:1494, 1979.

[49] N. Troullier and J. L. Martins. Efficient pseudopotentials for plane-wave calculations. *Phys. Rev. B*, 43:1993, 1991.

[50] S. Goedecker, M. Teter, and J. Hutter. Separable dual-space gaussian pseudopotentials. *Phys. Rev. B*, 54:1703, 1996.

[51] M. Krack. Pseudopotentials for H to Kr optimized for gradient-corrected exchange-correlation functionals. *Theor. Chem. Acc.*, 114:145, 2005.

[52] L. Verlet. Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Phys. Rev.*, 159:98, 1967.

[53] M. E. Tuckerman, B. J. Berne, and G. J. Martyna. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.*, 97:1990, 1992.

[54] G. Pastore, E. Smargiassi, and F. Buda. Theory of *ab initio* molecular-dynamics calculations. *Phys. Rev. A*, 44:6334, 1991.

[55] P. E. Blöchl and M. Parrinello. Adiabaticity in first-principles molecular dynamics. *Phys. Rev. B*, 45:9413, 1992.

[56] A. Alavi, J. Kohanoff, M. Parrinello, and D. Frenkel. *Ab Initio* molecular dynamics with excited electrons. *Phys. Rev. Lett.*, 73:2599, 1994.

[57] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684, 1984.

[58] H. C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.*, 72:2384, 1980.

[59] S. Nosé. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.*, 81:511, 1984.

[60] W. G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31:1695, 1985.

[61] G. J. Martyna, M. L. Klein, and M. E. Tuckerman. Nosé–hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.*, 97:2635, 1992.

[62] M. E. Tuckerman and M. Parrinello. Integrating the Car-Parrinello equations. i. basic integration techniques. *J. Chem. Phys.*, 101:1302, 1994.

[63] X. Wang, W. Conway, R. Burns, N. McCann, and M. Maeder. Comprehensive study of the hydration and dehydration reactions of carbon dioxide in aqueous solution. *J. Phys. Chem. A*, 114:1734, 2010.

[64] D. Branduardi, F. L. Gervasio, and M. Parrinello. From A to B in free energy space. *J. Chem. Phys.*, 126:054103, 2007.

[65] P. L. Geissler, C. Dellago, and D. Chandler. Kinetic pathways of ion pair dissociation in water. *J. Phys. Chem. B*, 103:3706, 1999.

[66] G. Mills, H. Jónsson, and G. K. Schenter. Reversible work transition state theory: application to dissociative adsorption of hydrogen. *Surface Sci.*, 324:305, 1995.

[67] S. R. Billeter, A. Curioni, and W. Andreoni. Efficient linear scaling geometry optimization and transition-state search for direct wavefunction optimization schemes in density functional theory using a plane-wave basis. *Comput. Mat. Sci*, 27:437, 2003.

[68] G. Henkelman, B. P. Uberuaga, and H. Jonsson. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.*, 113:9901, 2000.

[69] W. E, W. Ren, and E. Vanden-Eijnden. String method for the study of rare events. *Phys. Rev. B*, 66:052301, 2002.

[70] G.M. Torrie and J.P. Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.*, 23:187, 1977.

[71] A. M. Ferrenberg and R. H. Swendsen. Optimized Monte Carlo data analysis. *Phys. Rev. Lett.*, 63:1195, 1989.

[72] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *J. Comput. Chem.*, 13:1011, 1992.

[73] A. Laio, A. Rodriguez-Fortea, F. L. Gervasio, M. Ceccarelli, and M. Parrinello. Assessing the accuracy of metadynamics. *J. Phys. Chem. B*, 109:6714, 2005.

[74] G. Bussi, A. Laio, and M. Parrinello. Equilibrium free energies from nonequilibrium metadynamics. *Phys. Rev. Lett.*, 96:090601, 2006.

[75] F. L. Gervasio, A. Laio, and M. Parrinello. Flexible docking in solution using metadynamics. *J. Am. Chem. Soc.*, 127:2600, 2005.

[76] C. A. Guido, F. Pietrucci, G. A. Gallet, and W. Andreoni. The fate of a zwitterion in water from ab initio molecular dynamics: Monoethanolamine (MEA)-CO_2. *J. Chem. Theory Comput.*, 9:28, 2013.

[77] A. Laio and F. L Gervasio. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep. Prog. Phys.*, 71:126601, 2008.

[78] Y. Crespo, F. Marinelli, F. Pietrucci, and A. Laio. Metadynamics convergence law in a multidimensional system. *Phys. Rev. E*, 81:055701, 2010.

[79] A. Barducci, G. Bussi, and M. Parrinello. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.*, 100:020603, 2008.

[80] S. Piana and A. Laio. A bias-exchange approach to protein folding. *J. Phys. Chem. B*, 111:4553, 2007.

[81] F. Marinelli, F. Pietrucci, A. Laio, and S. Piana. A kinetic model of trp-cage folding from multiple biased molecular dynamics simulations. *PLoS Computational Biology*, 5, 2009.

[82] G. A. Gallet and F. Pietrucci. Structural cluster analysis of chemical reactions in solution. *J. Chem. Phys.*, 139:074101, 2013.

[83] M. E. Karpen, D. J. Tobias, and C. L. Brooks. Statistical clustering techniques for the analysis of long molecular dynamics trajectories: analysis of 2.2-ns trajectories of YPGDV. *Biochemistry*, 32:412, 1993.

[84] J. Shao, S. W. Tanner, N. Thompson, and T. E. Cheatham. Clustering molecular dynamics trajectories: 1. characterizing the performance of different clustering algorithms. *J. Chem. Theory Comput.*, 3:2312, 2007.

[85] A. Rajan, P. L. Freddolino, and K. Schulten. Going beyond clustering in MD trajectory analysis: An application to villin headpiece folding. *PLoS ONE*, 5:e9890, 2010.

[86] P. Cossio, A. Laio, and F. Pietrucci. Which similarity measure is better for analyzing protein structures in a molecular dynamics trajectory? *Phys. Chem. Chem. Phys.*, 13:10421, 2011.

[87] J. P. G. L. M. Rodrigues, M. Trellet, C. Schmitz, P. Kastritis, E. Karaca, A. S. J. Melquiond, and A. M. J. J. Bonvin. Clustering biomolecular complexes by residue contacts similarity. *Proteins: Struct., Funct., Bioinf.*, 80:1810, 2012.

[88] S. Goedecker, W. Hellmann, and T. Lenosky. Global minimum determination of the Born-Oppenheimer surface within density functional theory. *Phys. Rev. Lett.*, 95:055501, 2005.

[89] A. Sadeghi, S. A. Ghasemi, M. A. Lill, and S. Goedecker. Metrics for measuring distances in configuration spaces. *preprint arXiv:1302.2322*, 2013.

[90] X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren, and A. E. Mark. Peptide folding: When simulation meets experiment. *Angew. Chem. Int. Ed.*, 38:236, 1999.

[91] L Kauffman and P. J. Rousseeuw. *Finding groups in data.* John Willey & Sons, New York, 1990.

[92] D. Arthur and S. Vassilvitskii. *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, chapter k-means++: The Advantages of Careful Seeding, page 1027. Society for Industrial and Applied Mathematics Philadelphia, PA, USA, 2007.

[93] G. A. Gallet and F. Pietrucci. PIV clustering. http://sourceforge.net/projects/pivclustering.

[94] J. Timko, D. Bucher, and S. Kuyucak. Dissociation of NaCl in water from ab initio molecular dynamics simulations. *J. Chem. Phys.*, 132:114510, 2010.

[95] H. ur Rehman, M. Springborg, and Y. Dong. Structural and electronic properties of Sin, Gen, and SinGen clusters. *J. Phys. Chem. A*, 115:2005, 2011.

[96] B. Metz, O. Davidson, H. C. de Coninck, M. Loos, and L. A. Meyer, editors. *IPCC Special Report on Carbon Dioxide Capture and Storage.* Cambridge University Press, 2005.

[97] R. A. Kerr. Global warming is changing the world. *Science*, 316:188, 2007.

[98] G. T. Rochelle. Amine scrubbing for CO_2 capture. *Science*, 325:1652, 2009.

[99] L. M. Romeo, I. Bolea, and J. M. Escosa. Integration of power plant and amine scrubbing to reduce CO_2 capture costs. *Appl. Therm. Eng.*, 28:1039, 2008.

[100] M. R. M. Abu-Zahra, L. H. J. Schneiders, J. P. M. Niederer, P. H. M. Feron, and G. F. Versteeg. CO_2 capture from power plants: Part i. a parametric study of the technical performance based on monoethanolamine. *Int. J. Greenhouse Gas Control*, 1:37, 2007.

[101] G. Puxty, R. Rowland, A. Allport, Q. Yang, M. Bown, R. Burns, M. Maeder, and M. Attalla. Carbon dioxide postcombustion capture: A novel screening study of the carbon dioxide absorption performance of 76 amines. *Environ. Sci. Technol.*, 43:6427, 2009.

[102] L. Raynal, P.-A. Bouillon, A. Gomez, and P. Broutin. From MEA to demixing solvents and future steps, a roadmap for lowering the cost of post-combustion carbon capture. *Chem. Eng. J.*, 171:742, 2011.

[103] F. Porcheron, A. Gibert, P. Mougin, and A. Wender. High throughput screening of CO_2 solubility in aqueous monoamine solutions. *Environ. Sci. Technol.*, 45:2486, 2011.

[104] G. A. Gallet, F. Pietrucci, and W. Andreoni. Bridging static and dynamical descriptions of chemical reactions: An ab initio study of CO_2 interacting with water molecules. *J. Chem. Theory Comput.*, 8:4029, 2012.

[105] C. S. Tautermann, A. F. Voegele, T. Loerting, I. Kohl, A. Hallbrucker, E. Mayer, and K. R. Liedl. Towards the experimental decomposition rate of carbonic acid (H_2CO_3) in aqueous solution. *Chem. Eur. J.*, 8:66, 2002.

[106] M. T. Nguyen, M. H. Matus, V. E. Jackson, V. T. Ngan, J. R. Rustad, and D. A. Dixon. Mechanism of the hydration of carbon dioxide: Direct participation of H_2O versus microsolvation. *J. Phys. Chem. A*, 112:10386, 2008.

[107] M. Caplow. Kinetics of carbamate formation and breakdown. *J. Am. Chem. Soc.*, 90:6795, 1968.

[108] B. Arstad, R. Blom, and O. Swang. CO_2 absorption in aqueous solutions of alkanolamines: Mechanistic insight from quantum chemical calculations. *J. Phys. Chem. A*, 111:1222, 2007.

[109] E. F. da Silva and H. F. Svendsen. Computational chemistry study of reactions, equilibrium and kinetics of chemical CO_2 absorption. *Int. J. Greenhouse Gas Control*, 1:151, 2007.

[110] H.-B. Xie, Y. Zhou, Y. Zhang, and J. K. Johnson. Reaction mechanism of monoethanolamine with CO_2 in aqueous solution from molecular modeling. *J. Phys. Chem. A*, 114:11844, 2010.

[111] B. Han, C. Zhou, J. Wu, D. J. Tempel, and H. Cheng. Understanding CO_2 capture mechanisms in aqueous monoethanolamine via first principles simulations. *J. Phys. Chem. Lett.*, 2:522, 2011.

[112] T. S. Jeong and P. Y. Yu. Temperature dependence of the photocurrent in p-type CdTe. *J. Kor. Phys. Soc.*, 43:1101, 2003.

[113] O. Vigil-Galán, E. Marín, J.-S. Hernández, E. Saucedo, C. M. Ruiz, G. Contreras-Puente, and A. Calderón. A study of the optical absorption in CdTe by photoacoustic spectroscopy. *J. Mat. Sci.*, 42:7176, 2007.

[114] K. Ravichandran and P. Philominathan. Comparative study on structural and optical properties of CdS films fabricated by three different low-cost techniques. *Appl. Surf. Sci.*, 255:5736, 2009.

[115] Y. Yan, D. Albin, and M. M. Al-Jassim. Do grain boundaries assist S diffusion in polycrystalline CdS/CdTe heterojunctions? *Appl. Phys. Lett.*, 78:171, 2001.

[116] D. Wang, Z. Hou, and Z. Bai. Study of interdiffusion reaction at the CdS/CdTe interface. *J. Mater. Res.*, 26:697, 2011.

[117] J. Wang, G. Tang, X. S. Wu, and M. Gu. Stable structure and effects of sulfur in CdTe/CdS heterojunctions. *Surf. Interface Anal.*, 44:434, 2012.

[118] S.-H. Wei and A. Zunger. Calculated natural band offsets of all II-VI and III-V semiconductors: Chemical trends and the role of cation d orbitals. *Appl. Phys. Lett.*, 72:2011, 1998.

[119] S.-H. Wei, S. B. Zhang, and A. Zunger. First-principles calculation of band offsets, optical bowings, and defects in CdS, CdSe, CdTe, and their alloys. *J. Appl. Phys.*, 87:1304, 2000.

[120] S.-H. Wei and S. B. Zhang. Chemical trends of defect formation and doping limit in II-VI semiconductors: the case of CdTe. *Phys. Rev. B*, 66:155211, 2002.

[121] L. Zhang, J. L. F. Da Silva, J. Li, Y. Yan, T. A. Gessert, and S.-H. Wei. Effect of copassivation of Cl and Cu on CdTe grain boundaries. *Phys. Rev. Lett.*, 101:155501, 2008.

[122] J. Ma and S.-H. Wei. Origin of novel diffusions of Cu and Ag in semiconductors: The case of CdTe. *Phys. Rev. Lett.*, 110:235901, 2013.

[123] G. A. Gallet, F. Pietrucci, and W. Andreoni. 2013. (manuscript in preparation).

[124] M. T. Nguyen, G. Raspoet, L. G. Vanquickenborne, and P. T. Van Duijnen. How many water molecules are actively involved in the neutral hydration of carbon dioxide? *J. Phys. Chem. A*, 101:7379, 1997.

[125] T. Loerting, C. Tautermann, R. T. Kroemer, I. Kohl, A. Hallbrucker, E. Mayer, and K. R. Liedl. On the surprising kinetic stability of carbonic acid (H_2CO_3). *Angew. Chem. Int. Ed.*, 39:891, 2000.

[126] C.-L. Lin and S.-Y. Chu. Comparative study between carbonic and sulfurous acids for dissociation reaction. *J. Chin. Chem. Soc.*, 49:777, 2002.

[127] N.R. Jena and P.C. Mishra. An ab initio and density functional study of microsolvation of carbon dioxide in water clusters and formation of carbonic acid. *Theor. Chem. Acc.*, 114:189, 2005.

[128] J. Baltrusaitis and V. H. Grassian. Carbonic acid formation from reaction of carbon dioxide and water coordinated to $Al(OH)_3$: A quantum chemical study. *J. Phys. Chem. A*, 114:2350, 2010.

[129] J. Ireta, J. Neugebauer, and M. Scheffler. On the accuracy of DFT for describing hydrogen bonds: Dependence on the bond directionality. *J. Phys. Chem. A*, 108:5692, 2004.

[130] P. P. Kumar, A. G. Kalinichev, and R. J. Kirkpatrick. Dissociation of carbonic acid: Gas phase energetics and mechanism from ab initio metadynamics simulations. *J. Chem. Phys.*, 126:204315, 2007.

[131] D. A. McQuarrie and J. D. Simon. *Molecular Thermodynamics.* University Science Books: Sausalito, CA, 1999.

[132] A. Stirling and I. Pápai. H_2CO_3 forms via HCO_3^- in water. *J. Phys. Chem. B*, 114:16854, 2010.

[133] A. Stirling. HCO_3 – formation from CO_2 at high ph: Ab initio molecular dynamics study. *J. Phys. Chem. B*, 115:14683, 2011.

[134] M. Galib and G. Hanna. Mechanistic insights into the dissociation and decomposition of carbonic acid in water via the hydroxide route: An ab initio metadynamics study. *J. Phys. Chem. B*, 115:15024, 2011.

[135] J. C. Grossman, E. Schwegler, E. W. Draeger, F. Gygi, and G. Galli. Towards an assessment of the accuracy of density functional theory for first principles simulations of water. *J. Chem. Phys.*, 120:300, 2004.

[136] E. Schwegler, J. C. Grossman, F. Gygi, and G. Galli. Towards an assessment of the accuracy of density functional theory for first principles simulations of water. ii. *J. Chem. Phys.*, 121:5400, 2004.

[137] I.-F. W. Kuo, C. J. Mundy, M. J. McGrath, J. I. Siepmann, J. VandeVondele, M. Sprik, J. Hutter, B. Chen, M. L. Klein, F. Mohamed, M. Krack, and M. Parrinello. Liquid water from first principles: Investigation of different sampling approaches. *J. Phys. Chem. B*, 108:12990, 2004.

[138] H.-S. Lee and M. E. Tuckerman. Structure of liquid water at ambient temperature from ab initio molecular dynamics performed in the complete basis set limit. *J. Chem. Phys.*, 125:154507, 2006.

[139] J. Schmidt, J. VandeVondele, I.-F. W. Kuo, D. Sebastiani, J. I. Siepmann, J. Hutter, and C. J. Mundy. Isobaric-isothermal molecular dynamics simulations utilizing density functional theory: An assessment of the structure and density of water at near-ambient conditions. *J. Phys. Chem. B*, 113:11959, 2009.

[140] I.-C. Lin, A. P. Seitsonen, M. D. Coutinho-Neto, I. Tavernelli, and U. Rothlisberger. Importance of van der Waals interactions in liquid water. *J. Phys. Chem. B*, 113:1127, 2009.

[141] J. Wang, G. Roman-Perez, J. M. Soler, E. Artacho, and M.-V. Fernandez-Serra. Density, structure, and dynamics of water: The effect of van der Waals interactions. *J. Chem. Phys.*, 134:024516, 2011.

[142] D. Marx and M. Parrinello. Ab initio path-integral molecular dynamics. *Zeitschrift für Physik B Condensed Matter*, 95:143, 1994.

[143] J. A. Morrone and R. Car. Nuclear quantum effects in water. *Phys. Rev. Lett.*, 101:017801, 2008.

[144] L. Lin, J. A. Morrone, R. Car, and M. Parrinello. Momentum distribution, vibrational dynamics, and the potential of mean force in ice. *Phys. Rev. B*, 83:220302, 2011.

[145] A. Luzar and D. Chandler. Effect of environment on hydrogen bond dynamics in liquid water. *Phys. Rev. Lett.*, 76:928, 1996.

[146] J. C. Grossman, E. Schwegler, and G. Galli. Quantum and classical molecular dynamics simulations of hydrophobic hydration structure around small solutes. *J. Phys. Chem. B*, 108:15865, 2004.

[147] M. Allesch, E. Schwegler, and G. Galli. Structure of hydrophobic hydration of benzene and hexafluorobenzene from first principles. *J. Phys. Chem. B*, 111:1081, 2007.

[148] P. L. Geissler, C. Dellago, D. Chandler, J. Hutter, and M. Parrinello. Autoionization in liquid water. *Science*, 291:2121, 2001.

[149] A. Hassanali, F. Giberti, J. Cuny, T. D. Kühne, and M. Parrinello. Proton transfer through the water gossamer. *Proc. Natl Acad. Sci USA*, 110:13723, 2013.

[150] J. M. Rowe, R. M. Nicklow, D. L. Price, and K. Zanio. Lattice dynamics of cadmium telluride. *Phys. Rev. B*, 10:671, 1974.

[151] M. K. Herndon, A. Gupta, V. Kaydanov, and R. T. Collins. Evidence for grain-boundary-assisted diffusion of sulfur in polycrystalline CdS/CdTe heterojunctions. *Appl. Phys. Lett.*, 75:3503, 1999.

[152] I. Riech, J. L. Peña, O. Ares, A. Rios-Flores, V. Rejón-Moo, P. Rodríguez-Fragoso, and J. G. Mendoza-Alvarez. Effect of annealing time of $CdCl_2$ vapor treatment on CdTe/CdS interface properties. *Semicond. Sci. Technol.*, 27:045015, 2012.

[153] D. W. Lane, G. J. Conibeer, D. A. Wood, K. D. Rogers, P. Capper, S. Romani, and S. Hearne. Sulfur diffusion in CdTe and the phase diagram of the CdS-CdTe pseudo-binary alloy. *J. Cryst. Growth*, 197:743, 1999.

[154] D. W. Lane, K. D. Rogers, J. D. Painter, D. A. Wood, and M. E. Ozsan. Structural dynamics in CdS-CdTe thin films. *Thin Solid Films*, 361–362:1, 2000.

[155] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, 4:435, 2008.

[156] I. S. Joung and T. E. Cheatham. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B*, 112:9020, 2008.

[157] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926, 1983.

[158] I. S. Joung and T. E. Cheatham. Molecular dynamics simulations of the dynamic and energetic properties of alkali and halide ions using water-model-specific ion parameters. *J. Phys. Chem. B*, 113:13279, 2009.

[159] C. Zhang, S. Raugei, B. Eisenberg, and P. Carloni. Molecular dynamics in physiological solutions: Force fields, alkali metal ions, and ionic strength. *J. Chem. Theory Comput.*, 6:2167, 2010.

[160] T. Darden, D. York, and L. Pedersen. Particle mesh Ewald: An $n \cdot \log(n)$ method for Ewald sums in large systems. *J. Chem. Phys.*, 98:10089, 1993.

[161] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. A smooth particle mesh Ewald method. *J. Chem. Phys.*, 103:8577, 1995.

[162] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *J. Chem. Phys.*, 126:014101, 2007.

[163] G. J. Martyna and M. E. Tuckerman. A reciprocal space based method for treating long range interactions in ab initio and force-field-based calculations in clusters. *J. Chem. Phys.*, 110:2810, 1999.

[164] C. G. Broyden. The convergence of a class of double-rank minimization algorithms 1. general considerations. *IMA J. of Appl. Math.*, 6:76, 1970.

[165] R. Fletcher. A new approach to variable metric algorithms. *Comput. J.*, 13:317, 1970.

[166] D. Goldfarb. A family of variable-metric methods derived by variational means. *Math. Comput.*, 24:23, 1970.

[167] D. F. Shanno. Conditioning of quasi-newton methods for function minimization. *Math. Comput.*, 24:647, 1970.

[168] M. Valiev, E.J. Bylaska, N. Govind, K. Kowalski, T.P. Straatsma, H.J.J. Van Dam, D. Wang, J. Nieplocha, E. Apra, T.L. Windus, and W.A. de Jong. NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Comput. Phys. Comm.*, 181:1477, 2010.

[169] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia, and M. Parrinello. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Comm.*, 180:1961, 2009.

[170] A. Grossfield. WHAM: the weighted histogram analysis method, v. 2.0.6. http://membrane.urmc.rochester.edu/content/wham.

[171] C. Sbraccia. *Computer Simulation of Thermally Activated Processes.* PhD thesis, Scuola International Superiore di Studi Avanzati, 2005.

[172] *Landolt-Börnstein: Numerical Data and Functional Relationships in Science and Technology - Crystal and Solid State Physics*, volume 18. Springer, Berlin, 1984.

[173] S. Baroni, S. de Gironcoli, A. Dal Corso, and P. Giannozzi. Phonons and related crystal properties from density-functional perturbation theory. *Rev. Mod. Phys.*, 73:515, 2001.

GRÉGOIRE A. GALLET

Rue du Maupas, 11
1004 Lausanne
Switzerland

C3PN – Chair of Computational Chemical Physics at the Nanoscale
École **P**olytechnique **F**édérale de **L**ausanne
PH H2 454
1015 Lausanne
Switzerland

Tel. : +41-(21) 693 19 04
eMail : gregoire.gallet@epfl.ch

EDUCATION

2009–2013	**Ph.D.,** École Polytechnique Fédérale de Lausanne Thesis: *"Ab Initio* Simulations of Activated Processes: Application to CO_2 Reactions and a Photovoltaic Material Thesis director and co-director: Prof. **Wanda Andreoni** and Dr **Fabio Pietrucci** *Quantum chemistry methods, molecular dynamics and enhanced sampling schemes.*
2007–2009	**Master of Science MSc in Physics**, École Polytechnique Fédérale de Lausanne
2004–2007	**Bachelor of Science BSc in Physics**, École Polytechnique Fédérale de Lausanne Exchange student, Kunglia Tekniska Högskolan – Stockholm (2006-2007)

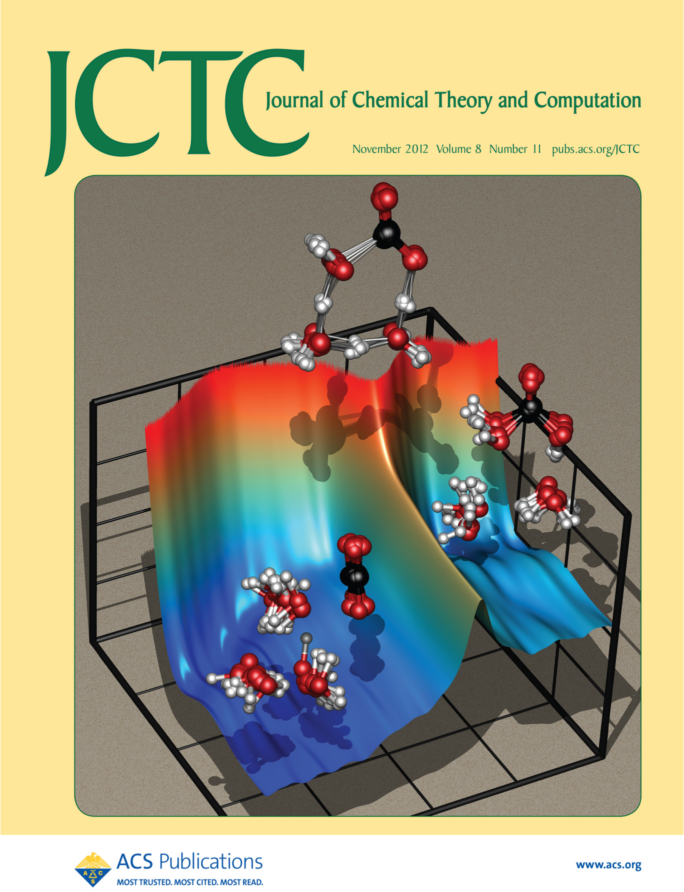
PUBLICATIONS AND CONFERENCES

- **Structural Cluster Analysis of Chemical Reactions in Solution**
  Grégoire A. Gallet and Fabio Pietrucci
  J. Chem. Phys. **139** 074101 (2013) – cover article
- **The Fate of a Zwitterion in Water from *ab Initio* Molecular Dynamics: Monoethanolamine (MEA)-CO_2**
  Ciro A. Guido, Fabio Pietrucci, Grégoire A. Gallet and Wanda Andreoni
  J. Chem. Theory Comput. **9** 28 (2013)
- **Bridging Static and Dynamical Description of Chemical Reactions: An *ab Initio* Study of CO_2 Interacting with Water Molecules**
  Grégoire A. Gallet, Fabio Pietrucci and Wanda Andreoni
  J. Chem. Theory Comput **8** 4029 (2012) – cover article

- **Dynamical Process in CdTe and CdTe:S : New Inisghts from *ab Initio* Molecular Dynamics Simulations**
  Article in preparation
  Poster at Material Research Society Fall Meeting 2012, *Boston*
  Poster at Energy from the Sun: Computational Chemists and Physicists Take up the Challenge 2012, *Chia Laguna*
- **The Fate of a Zwitterion in Water from *ab Initio* Molecular Dynamics: Monoethanolamine (MEA)-CO_2**
  Poster at Simulating Activated Processes in Physics and Chemisty: Theoretical Foundations 2013,*Villars-sur-Ollon*
- **The Reaction of Carbon Dioxide with Water Clusters: an Ab Initio Metadynamics Study**
  Poster at Deutsche Physicalische Gesellschaft Spring Meeting 2012, *Berlin.*
  Poster at DFT and Beyond – Hands on Tutorial Workshop 2011, *Berlin.*
  Poster at Solving the Schrödinger Equation – From Electronic Structure to Quantum Dynamics 2011,*Villars-sur-Ollon.*
  Contributed talk at American Physical Society March Meeting 2011, *Dallas.*
  Poster at Towards Reality in Nanoscale Materials 2010, *Levi.*

TECHNICAL SKILLS

- Familiar with the standard programming language: Fortran, C, C++, bash scripting, basics of Python and MPI
- Experience on large computing facilities: BlueGene, Cray, Linux clusters
- Specific codes: CPMD, Plumed, NWChem, Quantum Espresso

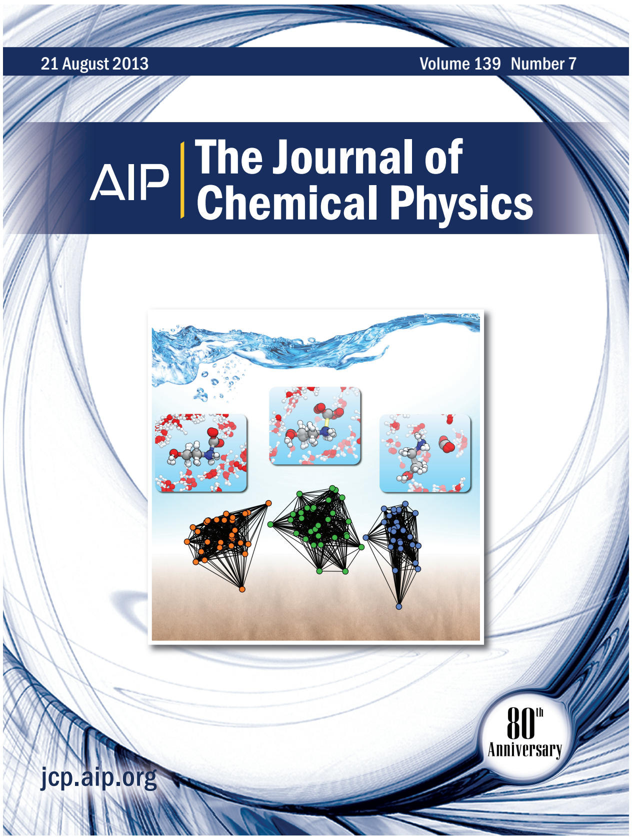


JCTC Journal of Chemical Theory and Computation

November 2012 Volume 8 Number 11 pubs.acs.org/JCTC

ACS Publications
MOST TRUSTED. MOST CITED. MOST READ.

www.acs.org



21 August 2013 Volume 139 Number 7

AIP The Journal of Chemical Physics

jcp.aip.org

80th Anniversary