

Solvent Effects on Chemistry with Alcohols An Ab Initio Study

Academisch Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam,
op gezag van de Rector Magnificus
prof. mr. P. F. van der Heijden ten overstaan van een
door het college voor promoties ingestelde commissie,
in het openbaar te verdedigen in de Aula der Universiteit
op vrijdag 25 april 2003 te 10.00 uur

door

Titus Sebastiaan van Erp

geboren te Tilburg

promotiecommissie

promotor: prof. dr. B. Smit

co-promotor: dr. E. J. Meijer

overige leden:

prof. dr. E.-J. Baerends

prof. dr. D. Frenkel

prof. dr. A. Fasolino

dr. M. Sprik

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

The research reported in this thesis was carried out at the Department of Chemical Engineering, Faculty of Science, University of Amsterdam (Nieuwe Achtergracht 166, 1018 WV, Amsterdam, The Netherlands) with financial support from the council for chemical sciences of the Netherlands Organization for Scientific Research (NWO-CW).

cover thesis:

"Les routes des molécules dansant"

"The pathways of the dancing molecules"

"Het pad der dansende moleculen"

oil on canvas

T. S. van Erp, 2003

Preface

In principle, the most fundamental laws in chemistry are known since 1926, when Schrödinger postulated his wave equation. This basic law of quantum mechanics is capable of describing all inter- and intramolecular interactions with great accuracy. Except for metals, relativistic corrections are mostly negligible and therefore the relativistic Dirac equation (1928) is of less importance. One could therefore argue that chemistry is not a science in itself. It just follows the laws of quantum mechanics and is therefore nothing but applied physics. However, as a cook does not stop when he has gathered all his ingredients, science does not stop when its basic laws are known. The 'cooking part' can still be a hell of a job. Up to now, theoretical chemists and physicists are by no means capable to replace the job of the experimentalists. If industries try to invent a new catalyst for a certain chemical process, they usually do not rely on theoretical calculations. Much more efficient are the methods of an experimentalist: a bit of intuition and a lot of trial and error. The reason for this is that theoretical calculations take enormous amounts of time when systems are complex. Mathematical analytical solutions of the Schrödinger equation only exist for the most simple systems, like the hydrogen atom. If the total number of particles is more than two, one has to rely on approximate numerical calculations. Of course, computers are an enormously powerful tool, but the number of atoms and electrons in most chemical systems is so large that even calculations on powerful supercomputers can take years.

Still, if a theoretical calculation can be performed, it can give much more insight on how the chemical process evolves and why a certain type of molecule is a good catalyst and another is not. This makes it worth studying these systems on a theoretical basis. However, the computational costs strongly limit the system size and simulation time. Gas-phase reactions containing only a small number of atoms are therefore usually the subject of these studies. Most chemical reactions in nature, in industry, or in laboratory experiments do not occur in the gas-phase, but occur in the presence

of a solvent like water. An accurate description of liquid water requires at least about 20 water molecules. In the last years an enormous progress has been made in the possibilities for theoretical calculations. This is partly due to the increase in computer power, but more important were the creative inventions of smart algorithms. Density Functional Theory and the Car-Parrinello method have, by use of minor approximations, decreased the computational costs by orders of magnitude. In combination with statistical methods to simulate rare events, the understanding of molecular behavior of more and more complex chemical reactions is in our reach.

I started my PhD four years ago and the final result is this doctoral thesis. The main question was to investigate the solvent effect on chemical reactions, in particular chemical reactions with alcohols in water. One of the conclusions of this research is that solvent dynamics is indeed very important. Molecular fluctuations in liquid water, yielding special structures between water molecules, can have an enhanced polarizing effect on solute molecules. This effect can facilitate or even initiate a chemical event. This thesis also contains a related but different subject. During the research, while studying methods for rare events, I got a new idea to improve upon the Transition Path Sampling method. In cooperation with Daniele Moroni and Peter Bolhuis we turned this idea into a completely new method for the calculation of rate constants. This new method, that we named Transition Interface Sampling, has become an important additional subject of this thesis.

Of course, this work could never have come to a successful end without the help and support of many people during the last four years, for which I am very grateful. Therefore, I would like to thank a number of people: First of all, my supervisor Evert Jan Meijer for sharing his detailed knowledge and broad experience on the art of Ab Initio molecular dynamics. Secondly, my promotor Berend Smit for having given me the opportunity for this PhD and for his stimulating support. Jan Willem Handgraaf, my most direct colleague in this field, for our fruitful collaboration during the last four years. Peter Bolhuis and Daniele Moroni for developing Transition Interface Sampling to a promising method. All the members of the manuscript commission for carefully and critically reading this manuscript. Daniele Moroni and Sofia Calero for being my paranims. For correcting parts of this thesis I would like to thank David Dubbeldam, my father Herman van Erp, and my mother Elize Schade. For solving many of my computer problems I owe a lot to Thijs Vlugt, Gooitzen Zwanenburg and Jochem Wichers Hoeth. I thank Arent Pelster for his help on the digital part of the cover. Moreover, for so-

cial support I am grateful to a countless number of people among which are: all my colleagues and friends of the ITS department, all my friends of Amsterdam and Nijmegen and of all other places in the world, the Weesperflat people, all the people from the Crea oil painting classes, especially Caroline of course, family (in particular my aunt Marianne and uncle Jacques for their hospitality in Bloemendaal the first half year of my PhD), my brother Pepijn and his wife Kitty, coming over for my defense from Tanzania, and finally my parents; thanks for everything.

Contents

1	Introduction	1
1.1	Structure of the Thesis	1
1.2	The Solvent	2
1.3	Alcohols and Aqueous Alcohol Solutions	2
1.4	Chemistry with Alcohols	4
1.5	Ab Initio Molecular Dynamics	5
1.5.1	DFT	6
1.5.2	Car-Parrinello	10
1.5.3	Pseudopotentials	12
1.6	Simulating Rare Events	13
1.6.1	Constrained Dynamics	13
1.6.2	Path Sampling	15
2	Hydration of Methanol in Water	19
2.1	Introduction	19
2.2	Methods and Validation	20
2.3	Solvation	23
2.3.1	Structure	24
2.3.2	Dynamics	26
2.3.3	Discussion	28
3	Aqueous Solvation of Ethanol and Ethylene	29
3.1	Introduction	30
3.2	Computational Methods	32
3.3	Gas-Phase Complexes	33
3.4	Solvation Structure	40
3.4.1	Ethanol Solvation	40
3.4.2	Ethylene Solvation	42
3.4.3	Water Structure	42

3.4.4	Hydrogen Bonds	44
3.5	Solute Dynamics	45
3.6	Polarization	46
3.7	Conclusions	48
4	Hydration/Dehydration Reaction between Ethanol and Ethene	51
4.1	Introduction	51
4.2	Computational Methods	55
4.3	Mechanism	57
4.4	Energetics	59
4.5	Structural Properties	62
4.6	Electronic Structure	63
4.7	Hydrogen Bonds	67
4.8	Conclusions	69
5	Transition Interface Sampling	71
5.1	Introduction	71
5.2	Theory	73
5.2.1	Transition State Theory and the Calculation of Rate Constants	73
5.2.2	Transition Path Sampling	76
5.2.3	Transition Interface Sampling	77
5.3	The Transition Interface Sampling Algorithm	83
5.4	Numerical Results	85
5.4.1	The Model	85
5.4.2	Methodology	86
5.4.3	System with High Energy Barrier	87
5.4.4	System with Low Energy Barrier	92
5.5	Conclusion	101
5.5.1	Appendix A: Proof of Flux Equivalence	102
5.5.2	Appendix B: Flux Relation	102
5.5.3	Appendix C: Alternative Transmission Coefficient	103
6	Non-Catalyzed Hydration of Ethylene	107
6.1	Introduction	107
6.2	Theory and Methods	108
6.3	Results and Conclusion	109
	Bibliography	111

CONTENTS	vii
-----------------	------------

Samenvatting	121
Publications	127
Curriculum Vitae	129

Chapter 1

Introduction

1.1 Structure of the Thesis

The starting point of my work as a PhD student was to investigate the influence of the solvent on chemical reactions with alcohols. Since the last decade this type of research has gained an increasing popularity. The reason for this is that the increase of computer power as well as the invention of smart algorithms have made it possible to perform full electronic calculations of much larger systems than before. It has been conjectured that the influence of the solvent plays an important role in chemical processes, especially for water, which is known to be one of the most complex liquids [1]. To investigate this, knowledge about the solvation properties of the reactant and product molecules is essential. Therefore, in chapter 2 we studied the solvation of methanol, the smallest alcohol. Chapter 3 deals with the solvation of ethylene and ethanol in water. The latter can be considered as an important pre-study for the hydration/dehydration reaction between ethylene and ethanol in acid aqueous solution, which is investigated in chapter 4. In chapter 5 we introduce the new method of Transition Interface Sampling (TIS) for the calculation of rate constants in rare event simulations. We derive the central expression of the TIS theory and apply the method on a simple molecular dynamics (MD) simulation. In chapter 6 we end with a simulation that combines the TIS method with *ab initio* methods to study the direct gas-phase hydration of ethylene. In this introduction I will continue by giving a short overview of the subject and a short introduction to the used methods.

1.2 The Solvent

Due to the still increasing computer power and the invention of efficient algorithms, the Car-Parrinello [2] method being the prime example, it is nowadays possible to study the molecular evolution of chemical reactions in solvents on an *ab initio* level. Before, only chemical reactions in the gas-phase were suitable for simulations, as the number of particles required for a correct description of a solvent was beyond the computational capabilities. However, in solution many chemical processes are fundamentally different from the gas-phase. The solvent is able to influence the chemical process in many ways. One of the known effects is that the solvent structure can stabilize or destabilize transition states, and can in this way lower or enhance reaction barriers.

A way to mimic these effects is to replace the solvent by a continuous field. However, this approximation is by no means capable of describing all the complexities of a solvent. Especially for the most common solvent, liquid water, the continuous field description would be a too strong simplification. Water has a unique structure due to its ability to form hydrogen bonds. In liquid water each molecule is on average hydrogen bonded to four neighboring water molecules, yielding a kind of tetrahedral structure (see Fig. 1.1). Although this structure is very dynamical with a constant process of breaking and making of hydrogen bonds, the local structure in one fixed snapshot of liquid water looks very similar to a crystal. This typical hydrogen bonded structure is responsible for the fast proton transfer in liquid water, the strong polarizing effect on solute molecules, and many other characteristics [1].

Nowadays *ab initio* methods, like the Car-Parrinello method, are able to simulate systems with 30 to 60 water molecules. To give an indication for the computational costs of these kind of systems: a 10 ps (10^{-12} s) simulation with 30 water molecules takes approximately a two weeks non-stop calculation on a cluster of 20 computers running in parallel. This is sufficient to describe the structural and dynamical properties of liquid water and aqueous solutions.

1.3 Alcohols and Aqueous Alcohol Solutions

Primary alcohols (methanol, ethanol, propanol,...) consist of a hydroxyl (OH) and an alkyl (C_nH_{2n+1}) group. The hydroxyl group has the ability to form hydrogen bonds. In that way, alcohols are similar to water. However,

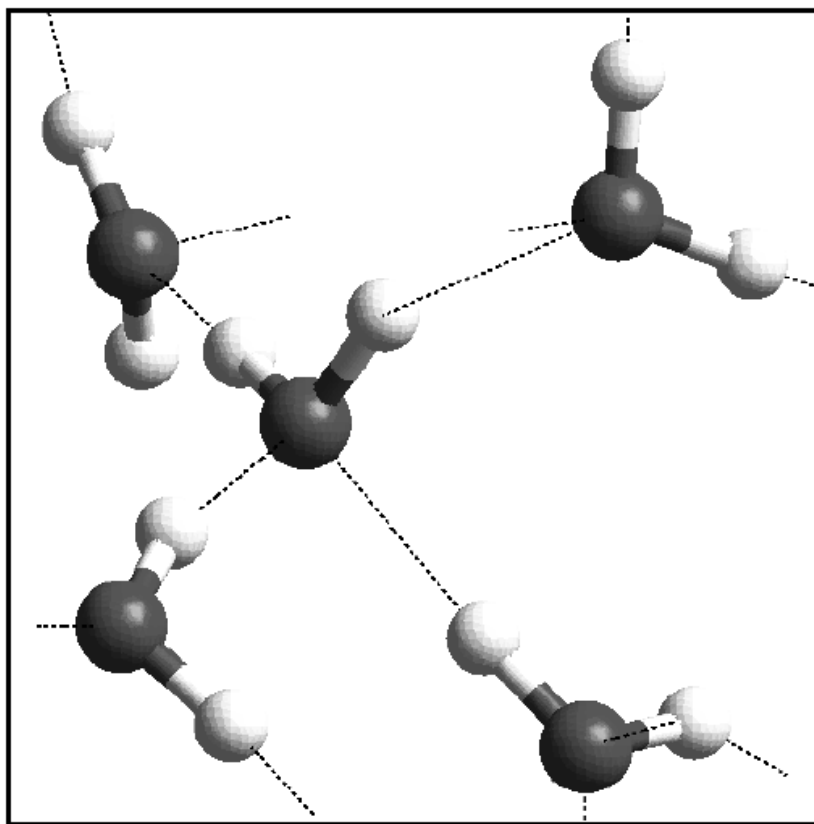


Figure 1.1: Typical snapshot of liquid water. For clarity, only five water molecules are shown. Oxygens are dark grey. Hydrogens are light grey. The dashed lines indicate the hydrogen bonds. The middle water molecule is connected via four hydrogen bonds to four neighboring water molecules. In this hydrogen bonded network, the middle water molecule acts two times as proton donor (to the upper waters) and two times as donor acceptor (from the lower waters). These four hydrogen bonds yield locally the typical tetrahedral structure of liquid water.

the alkyl group is apolar, what makes the molecular structure of liquid alcohol considerably different from that of liquid water. Whereas water forms a tetrahedral network, liquid alcohol has the tendency to form chains, rings or small clusters.

Since the hydroxyl group can participate in the hydrogen bonded water network, small alcohols are well soluble in water. The molecular structures of alcohol-water mixtures are subject of many scientific debates. Some

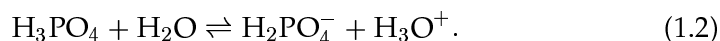
scientist believe that, in dilute alcohol solutions, water molecules orient themselves around the hydrophobic alkyl group forming a kind of hydrogen bonded cage. This effect, often referred as hydrophobic hydration [3, 4], increases the local water structure and decreases the mobility of water molecules in the vicinity of the alkyl group. However, the conclusions of different experimental studies do not all agree on this point. Most experimental measurements, such as X-ray or neutron diffraction experiments, only give indirect information on molecular structures. Computer simulations are therefore essential for a better understanding of these structures and for a good interpretation of the measured results. Still, the picture of alcohol/water mixtures is far from complete.

1.4 Chemistry with Alcohols

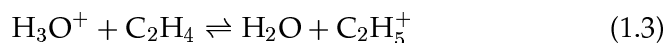
Chemical reactions with alcohols in an aqueous environment are important for many biological and industrial processes. Examples of such reactions are the conversion of ethanol into acetaldehyde in biological systems and the hydration of ethylene for industrial ethanol production. The latter reaction is examined in chapter 4 and 6. This process implies the following addition reaction :



Surprisingly, the reverse reaction is also applied by industry, but on a much smaller scale. For western countries, ethylene is relatively cheap as it is a product from the cracking process of crude oil. Many developing countries do not have the large supply of fossile fuels, but do have large amounts of ethanol from fermentation of molasses. For them the reverse reaction of (1.1) is the most economical way to produce ethylene. The reaction barrier of reaction (1.1) is high, making the industrial ethanol synthesis energetically expensive. However, the addition of an acid catalyst can lower the energetic cost significantly. Phosphoric acid has proven to give the best catalytic properties. This due to its effectiveness and its subduction of possible side products, but basically, the mechanism of the catalytic process is the same for all acids. The important part is the positively charged proton that will split off when the acid is solvated in water:



The H_3O^+ hydronium plays the catalytic role in the reaction process and makes an alternative pathway possible:





If the reaction steps (1.3) and (1.4) follow subsequently, then one speaks of the Ad_E2 mechanism. It is known that for small alcohols, like ethanol, the Ad_E3 mechanism is more favorable for the hydration reaction [5,6]. In that mechanism, step (1.3) and (1.4) happen simultaneously and protonated ethylene C_2H_5^+ is not present as intermediate reaction product. Step (1.3) and (1.4) are the rate determining steps as (1.5) has a much lower reaction barrier. The acid catalyzed mechanism for the hydration of ethene (1.3-1.5) lowers the activation barrier by approximately 30 kcal/mol compared to reaction (1.1). This implies an enhancement of the reaction rate by many orders of magnitude. Although, qualitatively the mechanism shown here is known for a long time, little is known about how this process exactly evolves in the environment of an aqueous solution. *Ab initio* molecular dynamics simulations can provide valuable insight towards a better understanding of this process.

1.5 Ab Initio Molecular Dynamics

Ab initio is Latin for 'from the beginning'. In the same context, *Ab initio* MD (AIMD) is often called 'first principles' MD. It means that this method is derived from the fundamental laws of physics. In that sense, this method is different than force field MD. In an AIMD simulation, forces on the ions are obtained by a full electronic structure calculation based on the Schrödinger equation of quantum mechanics. Force-field MD tries to mimic these forces without performing these expensive electronic structure calculations. It uses empirical potentials describing the forces on pairs of atoms or molecules as function of their relative positions. This potential is obtained by postulating an ansatz with a few free parameters. These can be fitted to, for example, available experimental data for equilibrium structures. Obviously, for those cases force field MD is not truly independent to experiment. Still, it can yield valuable insights that one can not obtain from experimental measurements. Besides, force field MD is many times faster than AIMD.

In turn, AIMD has some crucial advantages. In liquids, the total force on a molecule is not a simple summation in terms of pair interactions between this molecule and its neighbors. The force between two molecules can significantly be influenced by the presence of a third or a fourth nearby situated molecule as this can change the complete electronic state. In po-

lar liquids, it is known that dipole moments of individual molecules can strongly fluctuate and depend sensitively on the polarization of the surrounding molecules. In principle, these effects can be incorporated in standard MD by use of more advanced force fields, such as many body potentials and polarizable force fields (for water, see e.g. Ref. [7]). However, the design of such extended force fields is not an easy job. In AIMD these effects come naturally as the problem is treated starting from its quantum mechanical origin. More importantly, AIMD gives the possibility to study chemical reactions. The making and breaking of chemical bonds are accompanied with huge changes in the electronic density. The forces in intermediate stages, during the chemical event, cannot be known a-priori, which makes the design of accurate force fields for most situations an impossible task.

Straightforward AIMD is not feasible for large systems because of its enormous computational effort. Fortunately, innovative algorithms have drastically decreased the computational expense using only a few approximations. In this respect, Density Functional Theory (DFT) [8, 9] and the Car-Parrinello algorithm [2] were major breakthroughs. Of course, the increased computer power was also of importance. These developments have opened new possibilities for research, that were out of reach before.

1.5.1 DFT

The foundation of Density Functional Theory is the Hohenberg-Kohn (HK) theorem [8], which states that, for the electronic ground state, there is a one-to-one relation between the complex multi-dimension wavefunction and the electronic density:

$$\Psi_0(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \dots, \mathbf{r}_N) \leftrightarrow \rho_0(\mathbf{r}). \quad (1.6)$$

Here, Ψ_0 and ρ_0 are respectively the wavefunction and the electron density of a N -electron system in its ground state, \mathbf{r} denotes the general space coordinates (x, y, z) and \mathbf{r}_i denotes the space coordinate of electron i . Relation (1.6) implies that, if the system is in its electronic groundstate, knowledge of the electron density is in principle sufficient to obtain the N -electron wavefunction and thus all the information on the system. Consequently, the total electronic energy E_e can be expressed as a functional¹ of the density ρ_0 only:

$$E_e[\rho_0] \equiv \langle \Psi_0 | H_e | \Psi_0 \rangle = T[\rho_0] + U_{ee}[\rho_0] + U_{ext}[\rho_0]. \quad (1.7)$$

¹A functional $A[f(x)]$ is an extended type of function, that does not depend on a single input parameter or on a set of input parameters, but has a whole function $f(x)$ as argument.

Here, H_e is the Hamiltonian of the electron system, $T[\rho_0]$ and $U_{ee}[\rho_0]$ denote the electron kinetic energy and the electron-electron interaction energy and $U_{ext}[\rho_0]$ is the electron energy due to an external potential, usually the Coulomb interaction between the atomic nuclei and electrons. In principle, the electronic ground state density $\rho_0(\mathbf{r})$ can now be found by finding the minimum of the functional $E_e[\rho]$ as $E_e[\rho] \geq E_e[\rho_0]$ for any $\rho(\mathbf{r}) \neq \rho_0(\mathbf{r})$. To find the density that minimizes $E_e[\rho]$, we can use a variational approach. This requires taking the functional derivative² of $E_e[\rho]$ to solve $\frac{\delta E_e[\rho]}{\delta \rho} = 0$. The problem with Eq. (1.7) is that the HK theorem only proves that that unique functional $E_e[\rho]$ exists, but the functional itself remains unknown. There is no straightforward recipe to obtain the shape of this functional, but based on various principles one can design approximate energy functionals. Almost all existing functionals are designed within the framework of the Kohn-Sham (KS) formalism [9], where the interacting many-electron system is mapped onto a system of non-interacting electrons, yielding the exact same electron density. This formalism relies on the assumption, that the density ρ_0 can be expressed in terms of N orthonormal Kohn-Sham orbitals ψ_i

$$\rho(\mathbf{r}) = \sum_{i=1}^N \psi_i^*(\mathbf{r})\psi_i(\mathbf{r}), \quad (1.8)$$

that satisfy the Kohn-Sham equations (atomic units)

$$\left[-\frac{1}{2}\nabla^2 + v_s(\mathbf{r}) \right] \psi_i(\mathbf{r}) = \epsilon_i \psi_i(\mathbf{r}), \quad (1.9)$$

with $v_s(\mathbf{r})$ an effective potential, referred to as the Kohn-Sham potential, and ϵ_i the energy eigenvalues of the KS orbitals. If the assumption is valid, then the decomposition of Eq. (1.8) is unique yielding another one-to-one relation

$$\rho_0(\mathbf{r}) \leftrightarrow v_s(\mathbf{r}). \quad (1.10)$$

Relation (1.10) implies that solving Eqs. (1.9) is equivalent to finding the minimum of Eq. (1.7) by solving $\frac{\delta E_e[\rho]}{\delta \rho} = 0$. Again, the exact Kohn-Sham potential $v_s(\mathbf{r})$ is unknown. It is however much easier to find approximate expressions for $v_s(\mathbf{r})$ than for $E_e[\rho]$ directly. To achieve this, it is convenient

²The function derivative $\frac{\delta}{\delta f}$ of a functional $A[f(x)]$ gives back a function $g(x)$ as follows: $g(x) \equiv \frac{\delta A[f(x)]}{\delta f} \Rightarrow g(x') = \lim_{\delta f \rightarrow 0} \frac{A[f(x) + \delta(x-x')\delta f] - A[f(x)]}{\delta f}$, where $\delta(x-x')$ is the Dirac-delta function.

to separate $v_s(\mathbf{r})$ into following contributions:

$$v_s(\mathbf{r}) = v_{ext}(\mathbf{r}) + v_H(\mathbf{r}) + v_{xc}(\mathbf{r}), \quad (1.11)$$

with $v_{ext}(\mathbf{r})$, $v_H(\mathbf{r})$, and $v_{xc}(\mathbf{r})$ the external potential, the Hartree potential and the exchange-correlation potential, respectively. The potential $v_{ext}(\mathbf{r})$ corresponds to $U_{ext}[\rho_0]$ in Eq. (1.7). For an isolated system with m ions, $v_{ext}(\mathbf{r})$ is equal to:

$$v_{ext}(\mathbf{r}) = - \sum_{I=1}^m \frac{Z_I}{|\mathbf{r} - \mathbf{R}_I|}, \quad (1.12)$$

with \mathbf{R}_I the position and Z_I the charge of ion I .

The Hartree potential $v_H(\mathbf{r})$ approximates the Coulomb repulsion between the electrons in the following way:

$$v_H(\mathbf{r}) \equiv \int d\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}. \quad (1.13)$$

Note that Eq. (1.13) cannot be exactly equal to the Coulomb electron repulsion, as in the true Coulomb electron interaction, an electron cannot repel itself.

The exchange-correlation potential $v_{xc}(\mathbf{r})$ can be expressed as the functional derivative of the the exchange-correlation energy:

$$v_{xc}(\mathbf{r}) \equiv \frac{\delta E_{xc}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})}. \quad (1.14)$$

This exchange-correlation energy $E_{xc}[\rho]$ basically corrects all the other terms and is the only remaining unknown part of the problem. As $v_s(\mathbf{r})$ depends explicitly on the electron density $\rho(\mathbf{r})$ itself via relation (1.13) and (1.14), the KS Eqs. (1.9) must be solved iteratively in a self-consistent field approach.

The physical meaning of the non-interacting electron system, satisfying relation (1.9) and with total wavefunction $\Psi_s = \det|\psi_1, \psi_2, \dots, \psi_N|$, is that it has the same electron density as the real system. Other quantities do not have to be the same. For instance, the energy is not a simple sum of the KS electron energies ϵ_i , but must be obtained by putting the density of Eq. (1.8) back into Eq. (1.7), which has now an easier form:

$$E[\rho_0] = T_s[\rho_0] + U_{ext}[\rho_0] + U_H[\rho_0] + E_{xc}[\rho_0], \quad (1.15)$$

with $T_s[\rho]$, $U_{ext}[\rho]$ and $U_H[\rho]$ defined as:

$$T_s[\rho] \equiv -\frac{1}{2} \sum_{i=1}^N \int d\mathbf{r} \psi_i^*(r) \nabla^2 \psi_i(r), \quad (1.16)$$

$$U_H[\rho] \equiv \frac{1}{2} \int d\mathbf{r} \int d\mathbf{r}' \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}, \quad (1.17)$$

$$U_{ext}[\rho] \equiv \int d\mathbf{r} \rho(\mathbf{r}) v_{ext}(\mathbf{r}), \quad (1.18)$$

where the $*$ in the Eq. (1.16) denotes the complex conjugate.

The main achievement of the KS formalism is that $T[\rho_0]$ in Eq. (1.7) is now replaced by $T_s[\rho_0]$ in Eq. (1.15), which can be found by solving the one-electron equations (1.9), and putting the obtained KS orbitals $\{\psi\}$ into Eq. (1.16). This is formally correct as the KS assumption implies that $\psi_i = \psi_i[\rho]$. The above procedure looks rather cumbersome, but practice has shown that attempts to approximate $T[\rho_0]$ directly, the so-called orbital-free methods, are (still) rather inaccurate.

The remaining exchange-correlation energy can now be expressed as:

$$E_{xc}[\rho_0] \equiv T[\rho_0] - T_s[\rho_0] + U_{ee}[\rho_0] - U_H[\rho_0]. \quad (1.19)$$

This energy corrects for the fact that an electron cannot repel itself and for the fact that the electron motions are not truly independent. The exchange-correlation energy can be approximated by the local density approximation (LDA) based on the results of a uniform electron gas:

$$E_{xc}[\rho] = \int d\mathbf{r} \rho(\mathbf{r}) e_{xc}^{\text{uniform}}(\rho(\mathbf{r})), \quad (1.20)$$

with $e_{xc}^{\text{uniform}}(\rho)$ the exchange-correlation energy density function of the uniform electron gas that is known to a high accuracy from Monte Carlo calculations [10]. In spite of its approximate nature, LDA works surprisingly well in many applications, in particular in solid state physics. However, for a quantitative study of chemical systems, that often require an accuracy of a few kcal/mol, LDA generally fails. One can improve LDA by adding terms which depend on the spatial derivatives of the electronic density, commonly referred to as generalized gradient approximations (GGA). The main drawback of DFT is that there is no systematic way to improve the energy functional. Inventing new functionals is a science in itself, but the true functional is not known. In practice, you have to choose, among the existing functionals, a functional that gives the best possible description for your system. For our research, we used the Becke-Lee-Yang-Parr (BLYP) functional [11, 12] that has shown to give a good description of liquid water [13, 14].

DFT has proven to be more accurate than, for instance, the Hartree-Fock method and to be much faster than other methods of comparable accuracy. This has made DFT one of the most important methods in quantum chemistry.

1.5.2 Car-Parrinello

To obtain a manageable algorithm, AIMD methods use the following two approximations. First they consider the nuclei (ions) as classical particles and only the electrons are treated quantum mechanically. Secondly, the Born-Oppenheimer approximation is applied, where one assumes that the electronic state will remain constantly in the ground state. This implies that the electrons will adapt instantaneously to the motion of the ions. For most systems these two approximations are very accurate.

We can now write down the equations of motion for the ions:

$$m_I \ddot{\mathbf{R}}_I = - \frac{d}{d\mathbf{R}_I} \left[V_{\text{ion-ion}}(\{\mathbf{R}\}) + E_e[\rho_0] \right], \quad (1.21)$$

with m_I the mass of ion I and $V_{\text{ion-ion}}(\{\mathbf{R}\})$ the Coulomb interaction between the ions. Eq. (1.21) can be simplified by use of the following relation:³

$$\begin{aligned} \frac{d}{d\mathbf{R}_I} E_e[\rho_0] &= \frac{\partial}{\partial \mathbf{R}_I} E_e[\rho_0] + \int d\mathbf{r} \frac{\delta E_e[\rho_0]}{\delta \rho_0} \frac{d\rho_0(\mathbf{r})}{d\mathbf{R}_I} \\ &= \frac{\partial}{\partial \mathbf{R}_j} E_e[\rho_0] = \frac{\partial}{\partial \mathbf{R}_j} U_{\text{ext}}[\rho_0], \end{aligned} \quad (1.22)$$

where we have used $\frac{\delta E[\rho]}{\delta \rho} = 0$ if $\rho = \rho_0$. This is the DFT-variant of what is called the Hellmann-Feynman theorem which is valid for a system in its electronic ground state. This theorem states that the force, as well as the energy, can be calculated for a given atomic configuration without recalculating the electronic states, or finding their derivatives. This makes force calculations simpler and allows to rewrite Eq. (1.21) as follows:

$$m \ddot{\mathbf{R}}_I = - \sum_{J < I}^M \frac{(\mathbf{R}_J - \mathbf{R}_I) Z_I Z_J}{|\mathbf{R}_J - \mathbf{R}_I|^3} + \sum_I^M \int d\mathbf{r} \frac{\rho_0(\mathbf{r})(\mathbf{r} - \mathbf{R}_I) Z_I}{|\mathbf{r} - \mathbf{R}_I|^3}. \quad (1.23)$$

In principle, this equation together with Eqs. (1.8) and (1.9) allows to perform an AIMD simulation. It requires the numerical integration of Eq.

³The difference between the total derivative $\frac{d}{dx}$ and the partial derivative $\frac{\partial}{\partial x}$ on a function $a(x, f(x))$ is the following: $\frac{d}{dx} a(x, f(x)) = \lim_{dx \rightarrow 0} \frac{a(x+dx, f(x+dx)) - a(x, f(x))}{dx}$ whereas $\frac{\partial}{\partial x} a(x, f(x)) = \lim_{dx \rightarrow 0} \frac{a(x+dx, f(x)) - a(x, f(x))}{dx}$. The total derivative and the partial derivative can be related via the so called chain rule: $\frac{d}{dx} a(x, f(x)) = \frac{\partial a}{\partial x} + \frac{\partial a}{\partial f} \frac{df}{dx}$. Similarly, for a functional $A[f]$, that has an explicit dependence on a variable λ and in addition an implicit dependence on λ via the function $f(x)$, the chain rule is as follows: $\frac{d}{d\lambda} A[f] = \frac{\partial A[f]}{\partial \lambda} + \int dx \frac{\delta A[f]}{\delta f} \frac{df(x)}{d\lambda}$. This is used in step 2 of Eq. 1.22.

(1.23) by making small time steps like in standard MD. However, after each time step the time-consuming iterative calculation of Eq. (1.9) has to be performed to update the electronic density. Moreover, numerical calculations of the electronic structure will never be completely converged. Consequently, the forces determined by Eq. (1.23) will be slightly in error. Experience has shown that this error accumulates, yielding a non-conservation of the energy accompanied by a deviation of the calculated trajectory from the true one. In order to keep this deviation within acceptable bounds the electronic structure calculation requires a very high degree of convergence [15]. This gives rise to a strong computational effort.

In 1985 Car and Parrinello adopted an alternative approach [2], that avoids the expensive iterative calculation of Eq. (1.9) at every time step. First, one expands the KS orbitals ψ_i into a proper basis set $\{\phi_\alpha\}$, usually plane waves

$$\psi_i = \sum_{\alpha} c_{i\alpha} \phi_{\alpha} \quad (1.24)$$

Now, instead of finding the basis set coefficients $c_{i\alpha}$ by iteratively solving Eq. (1.9), Car-Parrinello MD (CPMD) uses a dynamical optimization. Basically, this implies that the basis-set coefficients are considered as one-dimensional particles with an associated mass μ yielding a similar equation for $c_{i\alpha}$ as Eq. (1.23) for the ions:

$$\mu \ddot{c}_{i\alpha} = -\frac{\partial E_e}{\partial c_{i\alpha}^*} + \sum_j \Lambda_{ij} c_{j\alpha}. \quad (1.25)$$

Here Λ_{ij} are the so-called Lagrange multipliers which constrain the KS orbitals to remain orthonormal. In addition to Eqs. (1.23) and (1.25) there is also an equation for the time evolution for these Lagrange multipliers, that we will not give explicitly. The electronic mass μ has no physical meaning, but it is required in order to define the dynamics for the electronic coefficients $c_{i\alpha}$. By choosing the mass μ small enough, the electronic coefficients will experience sufficiently large acceleration ensuring that the electrons can quickly adapt to the motions of the nuclei and that the deviations of the electronic coefficients from the true ground state remain sufficiently small. On the other hand, a too small value of μ requires a very small time step. So in practice one has to find for each system a kind of optimum.

At each MD time step, the ion motions and the basis set coefficients will change according to Eq. (1.23) and (1.25). During the simulation, the electron density will remain close, though not exactly equal, to the true density

ρ_0 , yielding forces on the ions that are again not exactly correct. However, the average force in CPMD has proven to be very accurate. This is because the deviations are, in contrary to straightforward AIMD, not systematic. Consequently, the ionic motion calculated from a CPMD simulation does match the true motion to a high accuracy, also over longer time [16]. This makes CPMD a very efficient way to perform AIMD simulations.

1.5.3 Pseudopotentials

Due to the very steep potential in the region near the nucleus, and due to the orthogonality condition between different states, the KS orbitals exhibit rapid oscillations close to the nuclei. A good description of these wavefunctions would require, in a plane-wave expansion, a very large basis set. So, if we would include all electrons in a system explicitly using the full Coulomb potential of the nuclei, the computational costs would be prohibitive. Fortunately, it is known that the core electrons states are almost independent of the environment surrounding the atom and that only the valence electrons participate strongly in the interactions between atoms. This allows to make a simplification in the description of the electronic state. Using the fact that the core electron states are almost fixed, we can define, for each atomic species, a pseudopotential, that incorporates the effects of the nucleus and core electrons [17–19]. These pseudopotentials must be constructed such, that they correctly describe the interaction of the nucleus and the core electrons with the valence electrons outside the core region. The valence pseudowavefunctions corresponding to this modified potential do not have the rapid oscillations of the true wavefunctions in the core region. This drastically reduces the number of basis set coefficients required for their representation. Then, the calculation only explicitly considers the valence electrons in terms of these pseudowavefunction solutions that approach the true wavefunction outside a core radius r_c . The enhancement in efficiency due to the pseudopotentials is considerable, as it ‘decreases’ the number of electrons in the system and secondly allows for a much smaller basis set for the remaining valence electrons.

Different schemes exist for the construction of pseudopotentials. The type of pseudopotential used for the research in this thesis are the so-called semi-local norm-conserving Martins-Troullier pseudopotentials [20]. For a more detailed overview of pseudopotentials, I refer to Ref. [21,22].

1.6 Simulating Rare Events

The above section dealt with the problem of how to make an AIMD calculation computationally feasible. The methods discussed in this section are also important for standard MD simulations. Given the system size and simulation time accessible by MD simulations, most chemical reactions should be considered as rare events. It is not possible to simulate these reactions in a direct way by imitating the same conditions under which one measures rates experimentally. In experiments, rate constants are obtained by measuring the macroscopic densities of reactant and product states over a long period. These macroscopic densities, corresponding to billions of molecules, will show an exponential relaxation with a decay according to the rate of reaction. MD simulations can only describe a few hundreds of molecules during simulation times of nanoseconds (the situation for AIMD is even much worse). The probability to detect one single event in these simulations is completely negligible. Therefore, additional techniques are required to determine reaction rates by molecular simulation. Many methods have been devised to this purpose, yielding crucial information on reaction barriers, rate constants and mechanisms of reaction. In the next section we will give a very short introduction to the constrained dynamics method, also called thermodynamic integration, that is used in chapter 4. In section 1.6.2 we will introduce the Path Sampling methods. In chapter 5 we will improve the original Transition Path Sampling method by a new method, that we call Transition Interface Sampling. This new method is used in chapter 6 in combination with AIMD, for the calculation of the non-catalyzed gas-phase hydration of ethylene.

1.6.1 Constrained Dynamics

Transition State Theory (TST) estimates the reaction rate as well as the activation energy of a chemical reaction from the free energy profile along a well chosen reaction coordinate. Mostly, this reaction coordinate will be something like the length of a bond that has to be broken or to be formed, but in general it can be any complex function of the coordinates of all particles. Given a reaction coordinate $\lambda(\{\mathbf{R}\})$, one can obtain the free energy change $\Delta F(\lambda)$ along the reaction path using the technique of thermodynamic integration (see, e.g., Ref. [23]). Along this path from reactant ($\lambda(\{\mathbf{R}\}) = \lambda_A$) to product state ($\lambda(\{\mathbf{R}\}) = \lambda_B$), one can then identify the transition state dividing surface ($\lambda(\{\mathbf{R}\}) = \lambda^*$ with $\lambda_A < \lambda^* < \lambda_B$), that is a local maximum of the free energy function, called the free energy barrier. If the reaction coor-

dinate is well chosen, this barrier corresponds to the true reaction barrier or activation energy of the reaction. This free energy barrier can be expressed as follows:

$$\Delta F(\lambda^*) = F(\lambda^*) - F(\lambda_A) = \int_{\lambda_A}^{\lambda^*} d\lambda' \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_{\lambda'}. \quad (1.26)$$

Here H is the Hamiltonian of the complete system. The brackets denote an ensemble average, and the subscript indicates that the conditional average is evaluated at a value $\lambda(\{\mathbf{R}\}) = \lambda'$ of the reaction coordinate. The term $-\left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_{\lambda'}$ is usually called the mean force. To evaluate this conditional ensemble average in a MD simulation, one can use the method of constraint where the dynamics of the system is performed with the reaction coordinate fixed at the specified value λ' . The mean force is in that case approximately equal to the constraint force (Lagrange multiplier) needed to keep the constraint at its fixed value. It is not exactly equal as the constraint introduces a bias in the ensemble. However, it is possible to correct for this bias and to obtain the exact mean force and free energy barrier [24,25].

Although, the free energy barrier is then exact, it still depends on the choice of the reaction coordinate. Only for an ideal reaction coordinate, the free energy barrier corresponds to the true reaction barrier. For an ideal reaction coordinate, any trajectory crossing the transition state dividing surface will remain at that side for a long time. This is exactly the assumption made in TST. In practice, one can usually not find such a coordinate, and fast recrossings are inevitable. In those situations the free energy barrier will serve as an underestimate of the true activation energy, whereas the TST rate expression will serve as an overestimate of the true reaction rate. In many cases, one can correct for the incorrect reaction coordinate by calculating a transmission coefficient that requires an additional simulation. Still, the choice of the reaction coordinate must be approximately correct. Otherwise, the corresponding transmission coefficient can be extremely low, yielding an inaccurate calculation.

In high dimensional complex systems, the choice of reaction coordinate can be extremely difficult and usually requires detailed *a priori* knowledge of the transition mechanism. For those cases, one can use Transition Path Sampling that was specifically designed to overcome these limitations, though with a price of a high computational cost. However, the new path sampling method introduced in this thesis, gives a strong improvement of the efficiency compared to Transition Path Sampling. This will make path sampling a good alternative to the thermodynamic integration method for many systems.

1.6.2 Path Sampling

The original path sampling method, called Transition Path Sampling, is based on the calculation of a time correlation function $k_{AB}^{TPS}(t)$, that, for a true exponential decay, will exhibit a plateau yielding the true reaction rate of the system:

$$k_{AB}^{TPS}(t) \equiv \frac{\langle h_A(0) \dot{h}_B(t) \rangle}{\langle h_A \rangle}, \quad k_{AB} = k_{AB}^{TPS}(T). \quad (1.27)$$

Here h_A and h_B are the characteristic functions of the two stable states, the product and the reactant state respectively, and T is a value in the plateau time region. The characteristic function of each state is either one, if the system is inside, or zero, if the system is outside the stable state region. Of course, the shape of the function $k_{AB}^{TPS}(t)$ depends on how one defines the stable states. However, as long as the choice of the state definitions is reasonable, the plateau value will not sensitively depend on this.

To calculate Eq. (1.27) in TPS, one samples a large number of paths, starting in A and ending in B , by use of a Monte Carlo algorithm that employs the so called shooting move (illustrated in Fig. 1.2). One of the advantages of this method compared to the constrained dynamics method is that the reaction evolves in a more natural way yielding a better understanding of the actual mechanism of the molecular rearrangements during a reactive event. Besides, if TST does not work due to a very low transmission coefficient, path sampling can still be applied. However, for most systems it is computationally more demanding than thermodynamic integration.

In chapter 5 we introduce a new algorithm called Transition Interface Sampling (TIS) that improves the efficiency of Transition Path Sampling. The idea is to introduce new characteristic functions based on a different kind of states, the *overall states*. These *overall states* exist next to the two stable states. Now, different from stable state definitions, a system belongs to *overall state* \mathcal{A} or \mathcal{B} not only depending on its instantaneous configuration, but also on its past behavior. Overall state \mathcal{A} covers all phase space points lying inside stable region A , but also all phase space points that visit A , before reaching B when the equations of motion are integrated backward in time. Similarly, state \mathcal{B} comprises stable state B and all phase points, coming directly from this state in the past (not via A). These two states span the complete phase space and they do not sensitively depend on the definitions of the stable states. This allows to write a rate expression similar

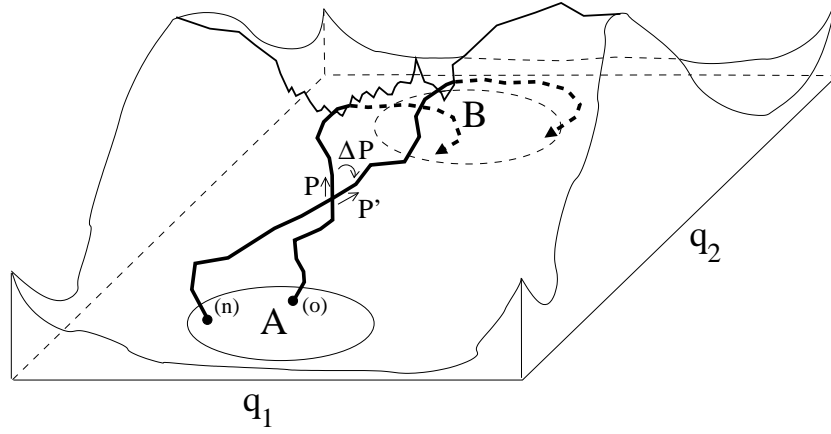


Figure 1.2: Illustration of the shooting move. The picture shows the free energy landscape as function of q_1 and q_2 , which represent two projections of the degrees of freedom in the system. The two circles, located in the valleys of this mountain landscape represents the two stable states A and B , that are separated by a free energy barrier. Two paths are shown that start in A and end in B . One path (n) is obtained from the other (o) via a MC shooting move. This works as follows. Along the old trajectory, one takes a random time slice at which one adds random changes ΔP to the momenta of all particles. This configuration with the new momenta $P' = P + \Delta P$ is then integrated forward and backward in time yielding the new trajectory (n). As long as the momenta changes are small, the chance that the new path will reach both stable states will be considerable. Then, by replacing the old path by the new one and starting the same procedure again, one can collect a whole set of different paths, all going from A to B .

to Eq. (1.27):

$$k_{AB} = \frac{\langle h_A(0) \dot{h}_B(0) \rangle}{\langle h_A \rangle}, \quad (1.28)$$

This expression has significant advantages over Eq. (1.27). Again, the evaluation of Eq. (1.28) is performed by sampling a number of paths via a similar kind of shooting moves. However, unlike in TPS where each path must have a fixed length T , here the path lengths are variable and each path can be limited to the strict minimum, necessarily for the determination of the h_A and h_B characteristic functions. Secondly, the ensemble average in Eq. (1.27) consists of both positive as negative terms, whereas expression (1.28) consists of positive terms only, resulting in a faster convergence of the MC algorithm. The third point requires a bit more knowledge about

the two path sampling methods, for which I refer to chapter 5. To increase the statistical accuracy, TPS combines the MC algorithm with a kind of free energy calculation by use of umbrella sampling techniques. In TIS this is replaced by a series of so-called interface ensemble averages, yielding another improvement to the total efficiency compared to TPS.

Chapter 2

Hydration of Methanol in Water

A DFT-based Molecular Dynamics Study¹

We studied the hydration of a single methanol molecule in aqueous solution by first-principle DFT-based molecular dynamics simulation. The calculations show that the local structural and short-time dynamical properties of the water molecules remain almost unchanged by the presence of the methanol, confirming the observation from recent experimental structural data for dilute solutions. We also see, in accordance with this experimental work, a distinct shell of water molecules that consists of about 15 molecules. We found no evidence for a strong tangential ordering of the water molecules in the first hydration shell.

2.1 Introduction

The solvation of alcohols in water has been studied extensively [26]. It is of fundamental interest in physics, chemistry and biology, but also of importance in technical applications. The characteristic hydroxyl group allows alcohols to form hydrogen bonds and is responsible for the good solubility of the smaller alcohols. In contrast, the alkyl group is hydrophobic and does not participate in the hydrogen bonding network of water. The presence of both hydrophobic and hydrophilic groups make the microscopic picture of solvation of alcohol in water a non-trivial and therefore interesting matter.

Understanding the solvation of methanol in water is a prerequisite for

¹This chapter is based on: Titus S. van Erp, Evert Jan Meijer, "Hydration of methanol in water. A DFT-based molecular dynamics study", *Chemical Physics Letters* 333 , 290-296, (2001).

the study of chemistry of alcohols in aqueous solution. Important examples of such reactions are the conversion of ethanol into acetaldehyde in biological systems or the industrial ethanol production by acid-catalyzed hydration of ethylene. An accurate microscopic understanding of the mechanism and kinetics of such reactions is of fundamental interest. However, presently, this picture is still far from complete. Density Functional Theory (DFT) based Molecular Dynamics simulation has proved to be a promising tool provide such an insight. An accurate calculation of the chemical bonding is incorporated via a DFT-based electronic structure calculations. The effect of temperature and solvent on the reactive events is implicitly accounted for via the Molecular Dynamics technique. The implementation of DFT-based MD as proposed by Car and Parrinello [2] has proven to be extremely efficient. It has successfully been applied to study of a large variety of condensed-phase systems at finite temperature. Applications to chemical reactions include the cat-ionic polymerization of 1,2,5-trioxane [27], or the acid-catalyzed hydration of formaldehyde [28].

As a first step towards the study of chemical reactions involving alcohols we present in this chapter a Car-Parrinello Molecular Dynamics (CPMD) study of the hydration of the simplest alcohol (methanol) in aqueous solution. Recent experimental work [29] has provided detailed structural information on the solvation shell. Various molecular simulation studies (e.g. Ref. [30–34] have addressed structure and dynamics of both the solute and the solvent. This experimental and numerical work has revealed that there is a distinct solvation shell around the methanol, and that the water structure is little affected by the presence of a methanol molecule. In this chapter we will address these structural properties and in addition consider the dynamics of the methanol and the water molecules in the solvation shell.

This chapter is organized as follows. First we outline the computational approach and its validation. Then we present the results for the structure and dynamics of a single solvated methanol in water. We conclude the chapter with a summary and discussion.

2.2 Methods and Validation

Electronic structure calculations are performed using the Kohn-Sham formulation [9] of DFT. [8] We employed the BLYP functional that combines the local density approximation (LDA) with a Becke gradient correction for the exchange energy [11] and a gradient correction for the correlation energy proposed by Lee, Yang and Parr [12]. Among the available functionals,

the BLYP functional has proven to give the best description of the structure and dynamics of water. [13,14] All calculations² were performed using the CPMD package [35].

The pseudopotential method is used to restrict the number of electronic states to those of the valence electrons. The interaction with the core electrons is taken into account using semi-local norm-conserving Martins-Troulier pseudopotentials. [20] The pseudopotential cutoff radius for the H was chosen 0.50 au. For O and C the radii are taken 1.11 and 1.23 a.u. for both the $l=s$ and $l=p$ term. The Kohn-Sham states are expanded in a plane-wave basis set matching the periodicity of the periodic box with waves up to a kinetic energy of 70 Ry. Test calculations showed that for this structural and energetic properties were converged within 0.01 Å and 1 kJ/mol, respectively. Frequencies are converged within 1 %, except for CO and OH stretch modes that are underestimated by 3 % and 5 % compared to basis-set limit values.

To validate the computational methods outlined above we performed a series of reference calculations of relevant gas - phase compounds with the CPMD package. Energetics and geometry were calculated for methanol, water, two mono - hydrate configurations, and the di-hydrate configuration shown in Fig. 2.1. These calculations were performed using a large periodic box of size $10 \times 10 \times 10$ Å³. The interactions among the periodic images were eliminated by a screening technique similar to that of Ref. [36]. In addition we determined for the methanol molecule both the harmonic vibrational frequencies and the frequencies at finite temperature ($T = 200$ K). The latter includes the anharmonic contributions, and were obtained from the spectrum of the velocity auto correlation function (VACF) of a 3 ps CPMD calculation at $E = 200$ K. The calculated peak positions can be compared with experimental spectra. Results of the gas-phase calculations were compared with results obtained with a state-of-the-art atomic-orbital based DFT package (ADF [37]), and with results from MP2 calculations of Ref. [38]. In the comparison of the energies zero-point energies were not taken into account.

Complexation energies and geometries of the methanol hydrates are given in Tab. 2.1 and Fig. 2.1. Deviations among CPMD and ADF are within 1 kcal/mole for the energies, smaller than 0.005 Å for the inter-molecular bonds and within 0.03 Å for the weaker intra-molecular bonds. This indicates a state-of-the art accuracy for electronic structure methods employed

²Computational resources consisted of an IBM-SP and a cluster of state-of-the-art PC's. Calculations were executed in parallel using MPI and amounted to a total of ≈ 10000 hours of CPU-time.

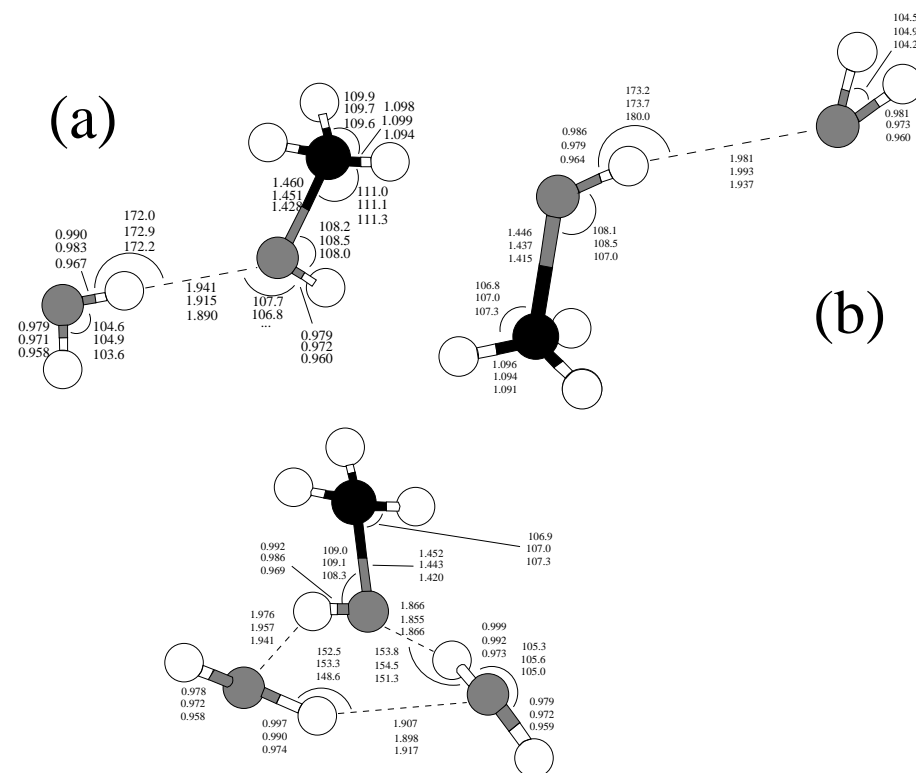


Figure 2.1: Energy-optimized geometries of two water/methanol dimers and a trimer. Distances (Å) and angles (degrees) are shown for three computational methods: CPMD-BLYP (top, present work), ADF-BLYP [37] (middle, present work) and MP2 [38] (bottom).

in CPMD. Differences among BLYP and MP2 are within acceptable limits, with BLYP complexation energies smaller by 4 kJ/mole (dimer) and 10 kJ/mole (trimer). These deviations are similar to the comparison of BLYP and MP2 for the water dimer binding energy, [13]³ where BLYP is 4 kJ/mole smaller, with the MP2 energy only 1 kJ/mol below the experimental value. Assuming similar differences for the complexation energies bonds in the methanol hydrates would suggest that BLYP underestimates the methanol-water binding energy by approximately 5 kJ/mol. Inter- and intra-molecular BLYP bond lengths are up to 0.02 and 0.06 Å longer compared to the MP2 results, respectively.

³MP2 limit estimate. See for example [39]

Table 2.1: Complexation energies (kJ/mol) of methanol hydrates shown in Fig. 2.1. Numbers are bare values without zero-point energy corrections and entropy contributions.

Complex	CPMD-BLYP	ADF-BLYP ^a	MP2 ^b
CH ₃ OH + H ₂ O (a)	20.2	20.2	24.4
CH ₃ OH + H ₂ O (b)	17.1	17.6	21.0
CH ₃ OH + 2 H ₂ O	58.3	59.6	68.8

^a Ref. [40].

^b G2(MP2) method. MP2(full)/6-311+G(d,p) optimized geometries. From Ref. [38].

Vibrational frequencies are listed in Tab. 2.2. Again comparison of CPMD and ADF is excellent, consistent with the results for the energetics and geometries. Comparing the calculated finite-temperature frequencies against the experimental values shows that BLYP tends to underestimate the frequencies of almost all modes by $\approx 10\%$. This trend is a known feature of BLYP. For example similar deviations are observed for BLYP calculation of water. [13]

Overall we conclude that the reference calculations of gas-phase provides confidence that DFT-BLYP performs with a sufficient accuracy for a quantitative study of methanol hydration.

2.3 Solvation

We performed Car-Parrinello Molecular Dynamics simulations of the solvation of a single methanol molecule. We considered two systems: one with 31 water molecules and the other with 63 water molecules, yielding methanol-water solutions with mole ratios of 1:31 and 1:63. In the following they are referred to as the small and large system, respectively. For reference we also performed a simulation of a pure water sample of 32 molecules. The molecules are placed in a periodic cubic box with edges of 9.98 Å (small solvated methanol system), 12.50 Å (large solvated methanol system), and 9.86 Å (pure water) corresponding to the experimental densities at ambient conditions. The temperature of the ions is fixed at 300 K using a Nosé-Hoover thermostat [42–44]. The fictitious mass associated with the plane-wave coefficients is chosen at 900 a.u., which allowed for a time step in the numerical integration of the equations-of-motion of 0.145 fs. The two sys-

Table 2.2: Harmonic and T=200 K vibrational frequencies of gas-phase methanol molecule.

mode	Harmonic ν (cm ⁻¹)		Anharmonic ν (cm ⁻¹)	
	CPMD-BLYP	ADF-BLYP ^a	CPMD-BLYP (T=200 K)	Exp. ^b
$\tau(\text{OH})$	280	380	280	270
$\nu(\text{CO})$	940	950	880	1034
$r(\text{CH}_3)$	1040	1050	980	1075
$r(\text{CH}_3)$	1130	1130	1070	1145
$\delta(\text{OH})$	1330	1340	1270	1340
$\delta(\text{CH}_3)$	1430	1430	1320-1430 ^c	1454
$\delta(\text{CH}_3)$	1460	1460	1320-1430 ^c	1465
$\delta(\text{CH}_3)$	1470	1470	1320-1430 ^c	1480
$\nu(\text{CH}_3)$	2940	2910	2640	2844
$\nu(\text{CH}_3)$	2990	2950	2740	2970
$\nu(\text{CH}_3)$	3060	3020	2830	2999
$\nu(\text{OH})$	3550	3590	3310	3682

^a Ref. [40].^b Ref. [41].^c Modes not separated. Broad peak with width listed.

tems were equilibrated for 1 ps from an initial configuration obtained by a force-field simulation. Subsequently we gathered statistical averages from a 10 ps trajectory of the 31+1 molecule system, from a 7 ps trajectory of the 63+1 molecule system, and from a 10 ps trajectory of the pure water system.

2.3.1 Structure

In Fig. 2.2 we have plotted the radial distribution functions (RDF) of the water oxygen atoms. The minor variations among the RDF's of the small methanol system, the large methanol system, and the pure water system is an indication that the local water structure, as measured by this RDF, is at only marginally changed by the solvation of a methanol molecule. Note, in this respect, that for the 32 molecule the first solvation shell constitutes a

significant fraction of the total number of water molecules (see below).

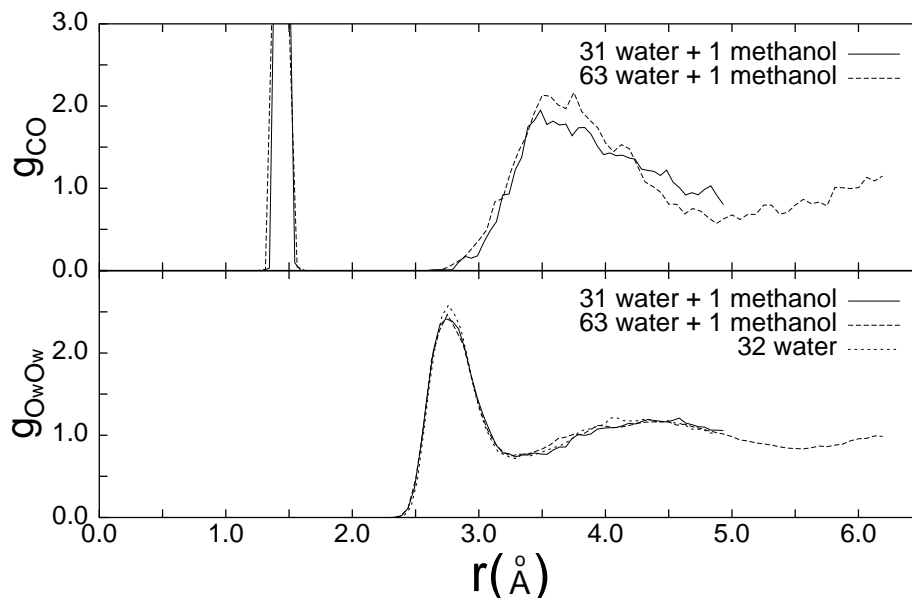


Figure 2.2: Calculated carbon-oxygen (top) and water oxygen-oxygen (bottom) radial distribution functions for the small (solid line) and large (dashed line) methanol system.

Fig. 2.2 also shows the RDF of the methanol carbon and water oxygens for the small and large methanol system. A pronounced first peak clearly indicates the existence of shell of water molecules at a distance of ≈ 3.7 Å. Comparing the RDF's of the small and large system shows a noticeable difference. This should be attributed to the limited size of the small system. It suggests that a proper description of the solvation structure of a single methanol in a cubic periodic simulation box requires at least 50 water molecules. Integrating the RDF for the large system up to the minimum at $r = 5.0$ Å yields 16 water molecules in the first solvation shell. The definite solvation shell observed in our simulations is consistent with the neutron diffraction data of Soper and Finney [29] who studied a 1:9 molar methanol-water system. Differences in molarity limits a quantitative comparison of the carbon-oxygen RDF, but a qualitative comparison learns that peak positions match with the peak values slightly more pronounced in the simulation results.

To analyze the orientational ordering of the water molecules around the methanol we computed the distribution function of the angle between the

C-O_{H2O} bond vector and the normal to the plane of the water molecules in the first solvation shell. The results show that angle distribution is relatively uniform with a small tendency towards the tangential orientation, a feature occurs for all solvation shell radii in the range of 3.7-5.0 Å. Over the range of 0°-90° the distribution gradually decays, with the value at the tangential orientation (0°) about a factor of 2 larger than at the perpendicular orientation (90°). Qualitatively, this seems consistent with data for the orientational distribution obtained from neutron-diffraction data [29]. However from this experimental data it is concluded that the water molecules prefer to lie tangential and form a cage around the methanol. Our data do not give clear evidence for a cage-like structure. However, this might be a different interpretation from similar data. Note, in this respect, also that the experimental data cannot be quantitatively compared to our data, as different orientational distribution functions are employed.

To analyze the hydrogen bonding we adopted the definition of Ref. [31]: two molecules are hydrogen bonded if simultaneously the inter-oxygen distance is less than 3.5 Å and the OHO angle is smaller than 30°. From the simulation of the large system we found that the methanol hydroxyl group donates and accepts on average 0.9 and 1.5 hydrogen bonds, respectively. For a water molecule these numbers are equal and measured to be 1.7 in the simulation of the pure water sample. These results indicate that the methanol hydroxyl group participates strongly in the hydrogen bonding network with the a donating behavior similar to water hydrogen and a accepting character somewhat smaller than a water oxygen.

2.3.2 Dynamics

The time scale (7-10 ps) of the present simulations allows for a reliable analysis of dynamical properties occurring on the picosecond time scale.

The velocity auto correlation function (VACF) of the hydrogen atoms provides an important measure of hydrogen bonding. Fig. 2.3 shows the Fourier spectrum of the calculated VACF of hydrogen atoms of the water molecules in the small and large methanol sample. The three distinct peaks correspond to the vibrational (3100 cm⁻¹), bending (1600 cm⁻¹), and librational- translational (500 cm⁻¹) modes of the water molecules. The most important observation is that mutual comparison of the two methanol samples and the comparison of these with the spectrum of the pure water sample (also plotted) shows no significant difference, not even for the small methanol sample where the solvation shell constitutes half of the water molecules in the system. This demonstrates that also the short-time

dynamics of the water molecules is hardly affected by the solvation of a methanol molecule.

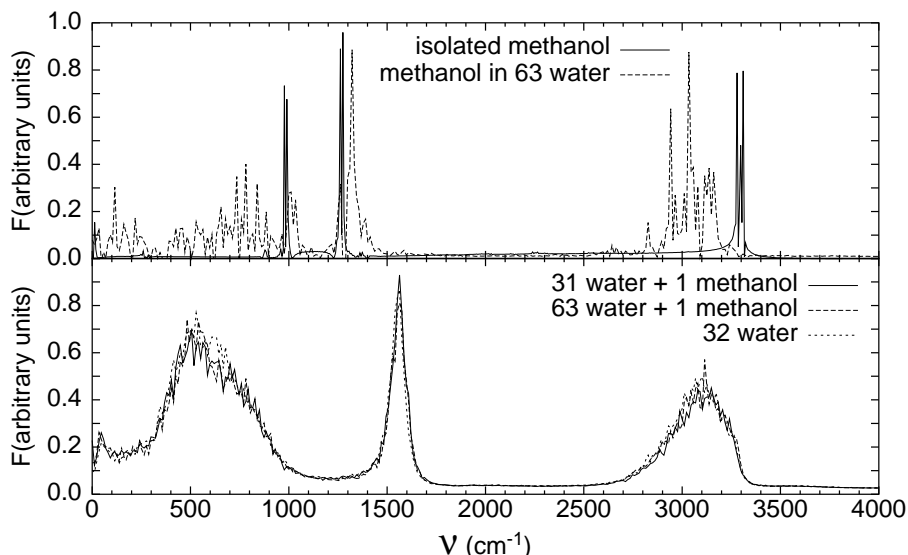


Figure 2.3: Bottom: Calculated Fourier spectrum of the velocity auto correlation function of the water hydrogens for the small methanol system (solid line), the large methanol system (dashed line), and the pure water sample (dotted line). Top: Calculated Fourier spectrum of the velocity auto correlation function of the hydrogen atom of the methanol hydroxyl group for the large methanol system (dashed line) and for an isolated methanol molecule (solid line).

An indication for the average residence time of a water molecule in the first solvation shell is obtained by monitoring the trajectories of the individual water molecules. We found that in the large methanol system over 7 ps 10 water molecules left the region within 5 Å from the methanol carbon. From this we estimate the average residence time to be of the order of a few picoseconds.

Fig. 2.3 shows the Fourier spectrum of the VACF of the hydroxyl H of methanol obtained from the trajectory of the large system. The spectrum is of limited accuracy due to the relatively short trajectories (7 ps). For comparison, the calculated spectrum for a single methanol molecule at $T = 200\text{ K}$ is also plotted. In solution the OH stretch (ν_{OH}) peak, with a calculated gas-phase position of about 3300 cm^{-1} , has shifted by $\approx 200\text{ cm}^{-1}$ to lower frequencies and has a relatively large width. The shift and width are both typical characteristics of a hydrogen bond and are also observed in the water spectrum (Fig. 2.3). In contrast to the OH stretch mode, we see that the OH-

bending mode (δ_{OH} at 1300 cm^{-1}) is blue-shifted by an amount of $50\text{-}100\text{ cm}^{-1}$. A comparison with experimental frequency shifts in infrared spectra is limited as, to our knowledge, no experimental data for dilute methanol-water solutions are reported. However, a comparison with measured shifts in liquid methanol [45] shows similar trends for the shift of infrared stretch (-354 cm^{-1}) and bend ($+78\text{ cm}^{-1}$) peaks. The torsional mode (τ_{OH}), expected to be shifted upward to around 600 cm^{-1} , is not visible in our calculated spectra due to the large statistical errors.

2.3.3 Discussion

We have studied the solvation of a single methanol molecule in water using DFT-based Car-Parrinello molecular dynamics simulation. Validation of the approach showed that energetics, structural, and dynamical properties of reference gas-phase compounds were sufficient to expect a quantitative accuracy of calculated properties.

The calculated solvation structure supports the experimental observation [29] that a shell of about 15 water molecules is formed around the methanol. Structural analysis also learns that the hydrogen bonded network of water is only minimally distorted by the presence of the methanol molecule. This confirms the proposition of Soper et al. [29] that speculations that the normal water structure is significantly enhanced by the hydrophobic alkyl group is groundless. The calculations showed that methanol OH group is strongly involved in hydrogen bonding, both as acceptor and as donor. Analysis of the dynamics learns that the average residence time of a water molecule in the first solvation shell is of the order of a few picoseconds. The vibrational spectrum of the water molecules is hardly changed by the presence of the methanol, indicating that the short-time dynamics is hardly affected by the presence of the methanol molecule. Vibrational analysis shows that methanol OH-stretch peak is a broad feature that is significantly red-shifted upon solvation, confirming its hydrogen-bonding character.

In conclusion, from comparison with available experimental data we have shown that first-principle DFT-based molecular dynamics simulation provides a reasonable accurate description of the structure and dynamics of a dilute aqueous methanol solution. This opens the way towards the study of chemistry involving methanol and larger alcohols in water.

Chapter 3

***Ab Initio* Molecular Dynamics Study of Aqueous Solvation of Ethanol and Ethylene**¹

The structure and dynamics of aqueous solvation of ethanol and ethylene are studied by density functional theory based Car-Parrinello molecular dynamics. We did not find an enhancement of the structure of the hydrogen bonded network of hydrating water molecules. Both ethanol and ethylene can easily be accommodated in the hydrogen-bonded network of water molecules without altering its structure. This supports the conclusion from recent neutron diffraction experiments that there is no hydrophobic hydration around small hydrophobic groups. Analysis of the electronic charge distribution using Wannier functions shows that the dipole moment of ethanol increases from 1.8 D to 3.1 D upon solvation, while the apolar ethylene molecule attains an average dipole moment of 0.5 D. For ethylene, we identified configurations with π -H bonded water molecules, that have rare four-fold hydrogen-bonded water coordination, yielding instantaneous dipole moments of ethylene of up to 1 D. The results provide valuable information for the improvement of empirical force fields, and point out that for an accurate description of the aqueous solvation of ethanol, and even of the apolar ethylene, polarizable force fields are required.

¹This chapter is based on: Titus S. van Erp, Evert Jan Meijer, "Ab Initio Molecular Dynamics Study of Aqueous Solvation of Ethanol and Ethylene", accepted for publication in *Journal of Chemical Physics*

3.1 Introduction

The study of the solvation of alcohols in aqueous solution is of fundamental interest in physics, biology and chemistry, but also of importance in technical applications [26, 46–55]. Among the alcohols, ethanol is one the most well studied compounds. Aqueous ethanol solutions are common in chemical research and industry applications. Ethanol can be produced by aqueous hydration of ethylene, that is readily available from natural sources. This process can be accelerated by acid catalysis. The reverse route of decomposing ethanol into water and ethylene is also of importance. For example, for developing countries who do not have a large supply of fossil fuels, dehydration of ethanol obtained from biomass is often the most economical way to produce ethylene. As is well known, solvation structures play a crucial role in aqueous solution chemistry, where reactive events often require a significant reordering of the water molecules in the solvation shell. The solvation of ethanol and ethylene are therefore crucial in the course of their (acid-catalyzed) interconversion. The aqueous solvation of these molecules will be addressed in the present paper. Elsewhere we report on an *ab initio* molecular dynamics study of the acid-catalyzed conversion [56].

Ethanol and ethylene have distinct solvation properties in aqueous solution. Ethanol is easily soluble as its polar hydroxyl group can participate in the hydrogen bonded network and the hydrophobic ethyl group is relatively small. In contrast, the apolar ethylene molecule has a much weaker interaction with water and is generally considered to be hydrophobic. Mixtures of water and ethanol have been studied extensively, both experimentally as by molecular simulation. Experimental studies employing NMR [57–61], ultrasonic absorption [62], infrared absorption spectroscopy [58, 59, 63, 64], mass spectroscopy [63, 65], X-ray diffraction measurements [63, 66], neutron diffraction [66–68], and dielectric relaxation measurements [69–72] have been performed to unravel the solvation properties of ethanol. Molecular simulation studies using empirical force fields have addressed the equation of state, thermodynamics, and structure and dynamics of solvation of aqueous ethanol solutions [73–77]. The following general picture of the aqueous solvation in dilute solutions has emerged: the hydroxyl group participates in the hydrogen-bonded network, while the hydrophobic alkyl group is accommodated in the hydrogen-bonded network of water molecules. The nature of hydration structure around the hydrophobic part of alcohols is still a controversial subject. It has been suggested [26, 46, 48, 49, 51, 54, 60, 61] that hydrophobic solutes enforces the network of hydrogen-bonded water molecules around it and decreases their mobility,

a notion referred to as hydrophobic hydration [3, 4]. However, recent experimental and computational studies have shown that for small alcohols the structure and dynamics of the water molecules in the solvation shell is almost identical to that of bulk water [67, 68, 75].

Much less is known for the solvation of ethylene in water under ambient conditions. Experimental [78] and theoretical [79, 80] work addressed the clathrate hydrates of ethylene in water. The isolated ethylene-water complex has been a subject of various studies. In the lowest energy configuration, the ethylene molecule forms a weak bond with water. An early *ab initio* study of Del Bene in 1974 [81] has characterized this interaction as a π -H hydrogen bond of a water proton with the π electrons of the C=C bond. Several experimental techniques have been applied to measure the strength of this interaction, such as the matrix isolation study of Engdahl and Nelander [82, 83], the microwave spectra study of Andrews and Kuczkowski [84] and the molecular-beam measurements of Peterson and Klemperer [85]. Recently Tarakeshwar et al. [86–88] and Dupré and Yappert [89] have performed calculations on the ethylene-water complexes with high level *ab initio* methods. The interaction is weak compared to a hydrogen bonds such as in the water dimer. The role of the π -H bond in aqueous solvation under ambient conditions is still an open question.

Molecular simulation provides an approach to study the microscopic behavior of liquids complementary to experimental studies. All molecular simulations studies of aqueous ethanol solutions up to now are based on empirical force fields that are designed to reproduce a selection of experimental data. Obviously, molecular simulations based on these potentials do not provide a picture completely independent from experiment. Moreover, the reliability of the results at conditions that are significantly different from those where the potential was designed for, may be questionable. Density functional theory (DFT) based molecular dynamics (MD) simulation, such as the Car-Parrinello molecular dynamics method [2], where the interactions are calculated by accurate electronic structure calculations, provides a route to overcome these limitations. This has been demonstrated in studies of liquid water [14, 90, 91] and aqueous solvation [92–94]. Important advantages of DFT-MD over force-field MD are that it intrinsically incorporates polarization, that it accounts for the intra-molecular motion and therefore allows for a direct comparison with spectroscopy of intra-molecular vibrations, and that it yields detailed information on the electronic properties, such as the energy levels of electronic states and the charge distribution. In a broader chemical perspective it is important to note that DFT-MD is capable to study

chemical reactions in solution, where force-field MD would fail completely as it cannot account for the change in chemical bonding.

Here, we report on a DFT-based MD simulation of the solvation of ethanol and ethylene. First we describe the simulation methods. Then we show results of geometries and energetics of relevant gas-phase complexes, that will serve as a validation of the numerical methods employed. Subsequently, results of structure, dynamics, and polarization of the solvated species will be shown. We conclude with a discussion.

3.2 Computational Methods

Electronic structure calculations are performed using the Kohn-Sham [9] formulation of DFT [8]. We employed the gradient-corrected Becke-Lee-Yang-Parr (BLYP) functional [11, 12], that has proven to give a good description of the structure and dynamics of liquid water [91]. The DFT-based MD simulations of aqueous ethanol and ethylene are performed using the Car-Parrinello method as implemented in the Car-Parrinello MD (CPMD) package [35]. Norm-conserving Martins-Troullier pseudopotentials [20] are used to restrict the number of electronic states to those of the valence electrons. Cut-off radii for H, O and C atoms were chosen to be 0.50, 1.11 and 1.23 a.u., respectively both for the $l=s$ and $l=p$ terms. The Kohn-Sham orbitals are expanded in a plane-wave basis, matching the periodicity of the periodic box with waves up to a kinetic energy of 70 Ry. With this basis set energies and geometries are converged within 0.25 kcal/mol and 0.01 Å, respectively. Vibrational frequencies are converged within 1 %, except for C-O and O-H stretch modes that are underestimated by 3 % and 5 % compared to the basis set limit values [93]. In the molecular dynamics calculations, the fictitious mass associated with the plane-wave coefficients is chosen at 900 a.u., which allowed for a time step in the numerical integration of the equations-of-motion of 0.145 fs.

To validate the computational approach we compared CPMD results for energies and structures of relevant gas-phase molecules and complexes with state-of-the-art atomic-orbital DFT-BLYP calculations performed with the Amsterdam density functional (ADF) [95–98] code, and other high-level quantum chemical results taken from literature. The gas-phase calculations with CPMD were performed in a large periodic box of 10 Å using the screening technique of Ref. [36] to eliminate the interactions among periodic images. We have not included zero-point energies in the energies of the gas-phase compounds. This also holds for computed energies taken

from literature and referred to in the present paper. This, to ensure a proper comparison between our results and those from literature.

The MD simulations of the solutions were performed for a 'small' and a 'large' system to assess the finite size effects. For the small system a simulation of 10 ps was performed in a cubic periodic box of length 10.07 Å, both for an ethanol solution of 31 waters and 1 ethanol as for an ethylene solution of 32 water and 1 ethylene. For the large system 5 ps simulations were performed using a periodic box with bcc symmetry and a volume of 1977.6 Å³. This periodic cell, a truncated octahedron, is in shape closer to a sphere than a simple cube, and therefore better suited for liquid simulation. The large ethanol solution consisted of 63 waters and 1 ethanol, while the large ethylene solution consisted of 64 waters and 1 ethylene. The box sizes for both the ethanol and ethylene solutions were set to match the experimental densities of the ethanol solutions under ambient conditions. For the ethylene system this will be slightly larger than the experimental density, as the effective volume of ethanol is a bit smaller than the combined volume of ethylene + water. However, we do not expect this to give rise to observable changes in the calculated properties. For reference we performed 10 ps MD simulations of a single ethanol and ethylene in a periodic cubic box of 10 Å, and a pure water systems of 32 water molecules in a cubic box of 9.85 Å for simulation times of 10 ps. For all simulations there was an initial equilibration trajectory of 1 ps. Temperature was controlled by a Nosé-Hoover thermostat [42–44] and fixed at 300 K.

3.3 Gas-Phase Complexes

The ethanol monomer has two stable conformers very close in energy: the symmetric trans structure and the a-symmetric gauche structure (see Fig. 3.1). The main distinction is the orientation of the OH bond with respect to the CCO plane. A microwave study [99] has shown that the trans form is slightly (0.12 kcal/mol) more stable than the gauche form. Tab. 3.1 lists the most important geometric data of the trans and gauche conformers, and compares the CPMD results with ADF calculations, B3LYP calculations [100], and experimental data [101]. CPMD and ADF bond lengths differ at most 0.01 Å and angles are within 0.5°. Comparing CPMD with B3LYP and experimental values yields differences upto 0.03 Å and 0.5°, and 0.04 Å and 1° respectively. The calculated energy difference between the two conformers is listed in Fig. 3.2. CPMD, ADF, B3LYP of Ref. [102], and fourth-order Möller-Plesset perturbation (MP4) [103] predict the trans conformer to be stable

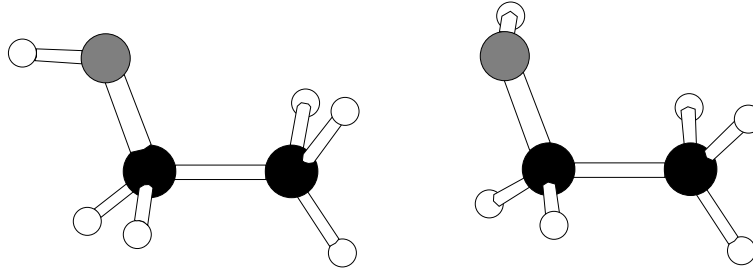
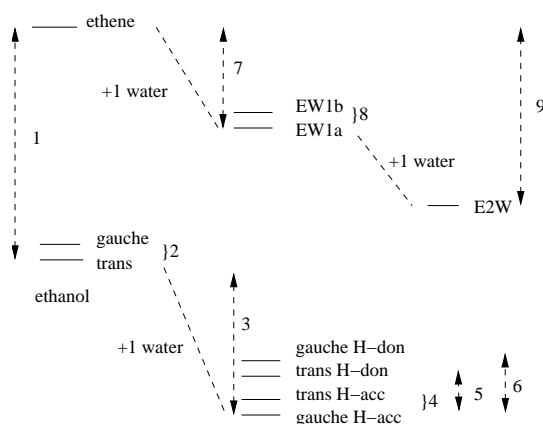


Figure 3.1: Illustration of the trans and gauche conformers of ethanol.

by 0.07-0.10 kcal/mol, in good agreement with the experimental value of 0.12 kcal/mol. In contrast, the B3LYP calculation of [100] yields the opposite, with the gauche conformer stable by 0.23 kcal/mol. Vibrational frequencies obtained by a Fourier transform of the velocity auto correlation

Table 3.1: Gas-phase complexes: ethanol gauche and trans monomer. Distances in Å, angles in degrees. Our results: CPMD-BLYP and ADF-BLYP, are compared with B3LYP/6-311G(2d,2p) [102] and experimental microwave spectroscopy data [101].

Geometry		Gauche			
Method		CPMD	ADF	B3LYP	EXP
	r_{OH}	0.981	0.971	0.961	0.945
	r_{CO}	1.455	1.447	1.429	1.427
	r_{CC}	1.529	1.532	1.521	1.530
	$\langle r_{\text{CH}} \rangle$ (in CH ₃)	1.099	1.097	1.091	1.094
	$\langle r_{\text{CH}} \rangle$ (in CH ₂)	1.100	1.099	1.092	1.094
	α_{COH}	108.0	108.2	108.7	108.3
	α_{CCO}	113.0	113.1	113.0	112.2
Geometry		Trans			
Method		CPMD	ADF	B3LYP	EXP
	r_{OH}	0.980	0.970	0.960	0.945
	r_{CO}	1.458	1.449	1.432	1.425
	r_{CC}	1.523	1.524	1.515	1.530
	$\langle r_{\text{CH}} \rangle$ (in CH ₃)	1.098	1.096	1.090	1.094
	$\langle r_{\text{CH}} \rangle$ (in CH ₂)	1.103	1.102	1.095	1.094
	α_{COH}	108.4	108.4	109.0	108.3
	α_{CCO}	107.6	108.0	108.0	107.2



This work			literature	
1	CPMD	ADF	Computational	Exp.
1	9.09	9.35		
2	0.08	0.08	0.07 (a), 0.10 (b), -0.23 (c)	0.12(d)
3	4.75	4.60	4.4–5.4 (e)	
4	0.07	0.09		
5	0.82	0.81	-0.92 (e)	
6	1.22	1.00		
7	1.45	1.41	2.3–2.9 (f,g), 2.0–3.3 (h)	3.7 (i)
8	0.05	0.06	0.13 (f)	
9	2.27	2.25	3.8–5.6 (j), 3.8–4.7 (g)	5.5–6.1 (k)

Figure 3.2: Relative energies of ethylene-water and ethanol-water in kcal/mol. Zero-point energies (ZPE) are not included. For ethylene, the energies are relative to the separate molecules, whereas for ethanol the separate water and trans ethanol conformer is taken as the reference value. 3 and 7 indicate the binding energies of ethanol and ethylene respectively. 9 is the binding energy between ethylene and the water-dimer. The literature values are: a: MP4-(SDQT)/cc-pVTZ [103], b: DFT/B3LYP/6-311G(d,p) [102], c: DFT/B3LYP/6-311G(d,p) [100], d: microwave experiment [99] e: MP2/6-311+G(2df,2p)+BSSE [104], f: MP2/TZ2P++ [86], g: MP2/6-311+G(2d,2p) [89], h: CCSD(T)/aug-cc-pVDZ [88], i: Matrix isolation study [83], j: MP2/aug-cc-pVDZ [87], and k: Matrix isolation study [82]. For (e), (f), (g), and (h) we gave the BSSE-corrected (lowest value) and the non-BSSE corrected values (highest value). (i) and (k) are the experimental values minus the ZPE of [89].

function (VACF) of a single ethanol at 300 K yields an OH stretch frequency of 3200 cm^{-1} and a CH stretch frequencies in the range of $2600\text{--}2800\text{ cm}^{-1}$. This should be compared to the experimental values of 3653 cm^{-1} [102], and $2800\text{--}3000\text{ cm}^{-1}$ [63, 102], respectively. The tendency of BLYP to underestimate frequencies by $\approx 10\%$ is a known feature, and also observed in BLYP calculations of water [91] and methanol [93].

For the ethanol-water complex we distinguished four complexes, with both the ethanol trans and gauche conformer acting as proton donor or acceptor in the hydrogen bond with water. Tab. 3.2 lists the most important geometric data obtained with CPMD and ADF. Fig. 3.2 shows the energy differences for all complexes. As for the ethanol monomer there is excellent agreement between CPMD and ADF results, with deviations within 0.02 \AA and 2° for bond lengths and angles, and the binding energies within 0.2 kcal/mol . This indicates a state-of-the art accuracy of the electronic-structure method employed in CPMD. The CPMD-BLYP calculation predicts the complex with the water molecule donating a proton to the ethanol gauche conformer to be the most stable, with a binding energy relative to the isolated water and trans ethanol conformer of 4.75 kcal/mol . The complex with the ethanol trans conformer donating a proton is less stable by 0.82 kcal/mol . Switching to the other ethanol conformer within either complex destabilizes the complex by 0.07 and 0.4 kcal/mol , respectively. The relative stability of the ethanol donor and acceptor complexes is similar to that for the methanol-water complex where CPMD-BLYP [93] and the complete-basis-set second-order Möller-Plesset perturbation (MP2) estimate [105] yields the methanol acceptor complex stable by 0.74 and 0.35 kcal/mol , respectively. There are no experimental data for the structure and energetics of the ethanol-water dimer. The MP2 result of Ref. [104] provides the only high-level quantum mechanical study reported in literature. This study does not distinguish between the trans and gauche conformers. Therefore, a comparison with our results is somewhat limited. For the ethanol acceptor configuration the MP2 and CPMD-BLYP result for the complexation energy are similar. However, for the ethanol donor configuration the MP2 complexation energy is more than 1.3 kcal/mol larger than the CPMD-BLYP results. Consequently, the MP2 calculations yield an opposite relative stability of the two water-ethanol configurations, with an energy difference of 0.92 kcal/mol . It should be noted that in the same study the methanol-water complex with the methanol as hydrogen bond donor is found to be the most stable, in contradiction with the MP2 basis-set limit result of Ref. [105]. This suggests that in the Ref. [104] a limited basis set or other factors could

Table 3.2: Gas-Phase complexes: ethanol-water dimers. Distances in Å, angles in degrees. CPMD-BLYP and ADF-BLYP are compared. We differentiate four complexes with the ethanol can be in the gauche or trans geometry and acts as proton acceptor or proton donor in the ethanol-water hydrogen bond. H_{nb} is the hydrogen of the water that is not involved in a hydrogen bond, C_1 is the hydroxyl carbon, C_2 is the methyl carbon and O_w is the water oxygen.

Geometry	Gauche			
H-bond	H-acc		H-don	
Method	CPMD	ADF	CPMD	ADF
r_{OH} in H ₂ O	0.994	0.981	0.981	0.972
$r_{OH_{nb}}$ in H ₂ O	0.981	0.970	0.981	0.971
r_{OH} in eth.	0.981	0.971	0.987	0.978
r_{OH} H-bond	1.934	1.912	1.988	2.011
r_{CC}	1.526	1.528	1.532	1.533
r_{CO}	1.469	1.457	1.452	1.440
r_{OO}	2.924	2.884	2.978	2.983
$r_{O_w C_1}$	3.653	3.626	3.773	3.793
$r_{O_w C_2}$	3.932	3.920	4.128	4.033
α HOH in H ₂ O	104.2	105.0	104.6	104.9
α COH	107.9	108.4	108.5	108.7
α OH-O	172.6	170.2	172.4	172.8
Geometry	Trans			
H-bond	H-acc		H-don	
Method	CPMD	ADF	CPMD	ADF
r_{OH} in H ₂ O	0.993	0.982	0.981	0.972
$r_{OH_{nb}}$ in H ₂ O	0.978	0.969	0.982	0.971
r_{OH} in eth.	0.982	0.970	0.988	0.977
r_{OH} H-bond	1.928	1.929	1.960	1.993
r_{CC}	1.521	1.522	1.523	1.526
r_{CO}	1.474	1.460	1.454	1.442
r_{OO}	2.908	2.989	2.988	2.968
$r_{O_w C_1}$	3.641	3.641	3.815	3.764
$r_{O_w C_2}$	4.013	4.008	5.229	5.189
α HOH in H ₂ O	104.6	105.0	104.2	104.9
α COH	107.9	108.2	108.9	108.5
α OH-O	168.1	168.4	175.0	174.1

have lead to spurious reversal of the relative stability of the two water-ethanol configurations. The CPMD-BLYP hydrogen bond interaction energy for the ethanol-water dimer is of the same order of magnitude as the CPMD-BLYP result for the water-water [91], water-methanol [93], and methanol-methanol [106] dimer. Comparison of these three dimers against high-level quantum chemical calculations and experimental values indicates that BLYP underestimates the binding energy by approximately 1 kcal/mol. For the ethanol-water dimer we could expect a similar difference. Here we should add that going from water, via methanol, to ethanol the dispersion forces become increasingly important. These are not accounted for in gradient-corrected functionals such as BLYP. Correlated methods such MP2 incorporate, to a good approximation, dispersion forces. In Ref. [106] we estimated, in a comparison of BLYP and high-level MP2 calculations [107] for the methanol-dimer in non-hydrogen bonded configurations, that the absence of dispersion interaction in BLYP amounted to an underestimate of the binding energy of ≈ 1 kcal/mol. For the ethanol-water dimer this number could serve as an underestimate. Yet, although by far not insignificant, the magnitude of the deviation is much smaller than the hydrogen-bond interaction. Therefore, it can be argued that for a study of aqueous ethanol neglecting the dispersion interaction is acceptable.

Next we discuss the ethylene-water complexes. We will consider complexes with a single water and a water dimer. The dominant interaction of water with ethylene is a π -H bond, where a proton of the water molecule binds to π electrons of the double C=C bond. There exist two stable complexes between ethylene and a single water molecule, indicated as the EW1a and EW1b geometry (see Fig. 3.3). Both geometries have Cs symmetry with

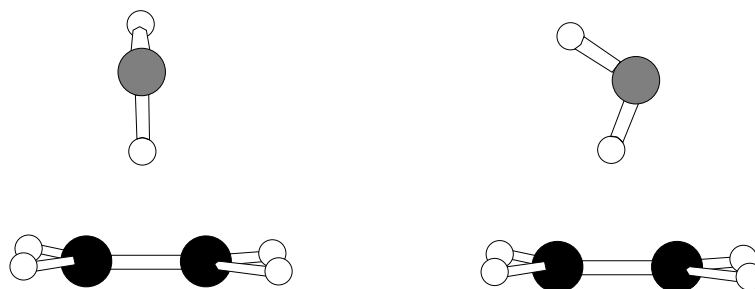


Figure 3.3: Illustration of the EW1a and EW1b ethylene-water complexes.

the plane of the water molecule orthogonal to that of ethylene molecule. In

the EW1a structure the water-plane is parallel to the C=C bond, whereas in EW1b it is orthogonal. In Tab. 3.3 we compare the CPMD-BLYP geometries with ADF-BLYP and MP2 [86,89] and coupled-cluster singles and doubles (CCSD) [88] results. Fig. 3.2 shows the energy differences, and compares these to values reported in literature. Again, the CPMD and ADF results are in excellent agreement, with energies smaller than 0.02 kcal/mol, and geometries within 0.01 Å and 0.3°, except for the non bonded OH_{nb} distance that is related to water orientation, a coordinate along which the energy surface is relatively flat. The BLYP result shows a significant π -H binding energy of ≈ 1.4 kcal/mol. However, in MP2 [86,88,89] and CCSD [88] calculations, binding energies are ≈ 1 -2 kcal/mol stronger, accompanied by a shorter π -H bond with differences of 0.12 Å and 0.14 Å for EW1a and EW1b, respectively. The comparison with the experimental value gives a similar picture with the BLYP underestimating the binding energy by ≈ 2 kcal/mol.

Table 3.3: Gas-Phase complexes: ethylene-water, EW1a, EW1b structures (see Fig. 3.3). Distances in Å, angles in degrees. CPMD-BLYP, ADF-BLYP and MP2-TZ2P++ [86] calculations are compared. H is the water-hydrogen involved in the π -H bond; H_{nb} is the other hydrogen of the water molecule. CM is the midpoint of the two carbons of the ethylene molecule.

Geometry	EW1a			EW1b		
Method	CPMD	ADF	MP2	CPMD	ADF	MP2
r_{CC}	1.334	1.336		1.335	1.336	
r_{H-CM}	2.483	2.485	2.363	2.570	2.524	2.383
r_{O-CM}	3.452	3.460	3.301	3.527	3.497	3.337
$r_{H_{nb}-CM}$	3.909	3.833	3.816	3.957	3.881	3.793
r_{OH}	0.984	0.975	0.962	0.978	0.974	0.962
$r_{OH_{nb}}$	0.980	0.970	0.958	0.976	0.970	0.958
α_{HOH}	104.2	104.5	104.7	104.4	104.5	104.7

Matrix isolation studies [82,83] have revealed that the complex of ethylene with two water molecules consists of a water dimer that has one water molecule π -H bonded to the C=C double bond. Recent high-level MP2 calculations have indicated that the presence of the second water molecule enhances the strength of the π -H bond [87,89]. From Fig. 3.2 we see that BLYP result are qualitatively in line with this observation, with a binding energy of the (water-dimer)-ethylene complex of 2.25 kcal/mol, up from 1.41 kcal/mol for the single-water binding. However, quantitatively MP2 and BLYP compare less well, with BLYP underestimating both the total bin-

ding energy as well as the increase from the single-water binding. Again, the experimental value [82, 83], that is closer to the MP2 result, shows a significantly stronger binding than BLYP. Both for the water-ethylene and for the (water-dimer)-ethylene complex the differences of BLYP with MP2 and experiment are significant relative to the total binding energy. This absence of dispersion interaction in BLYP will be an important factor contribution to this discrepancy.

Overall, we conclude that the reference calculations provide confidence that BLYP is capable of a quantitative study of a the aqueous solvation of a single ethanol molecule where interactions are dominated by relatively strong hydrogen bonds. BLYP qualitatively accounts for the weaker π -H binding in water-ethylene systems. However, comparison with MP2 and experimental data suggest that its strength is significantly underestimated. Yet, for a single solvated ethylene the hydrogen bonds among the solvating water molecules will be the dominant interaction. Hence, we believe that BLYP will be able to quantitatively describe dilute aqueous ethylene solutions. Both for ethanol and ethylene the absence of dispersion attraction in BLYP will have some impact, in particular for the coordination around the CH_2 and CH_3 groups, where BLYP only accounts for the steric repulsion. Except for the notion of the absence of the dispersion interaction, a better understanding of the performance of BLYP, its agreement and differences with MP2 and experimental results, is desirable. This falls outside the scope of the present work but would be an important topic of future work.

3.4 Solvation Structure

3.4.1 Ethanol Solvation

Fig. 3.4 shows ethanol-water radial distribution functions (RDFs) of the small and large ethanol solutions. The pronounced structure in the hydrogen bonding RDFs (HH, OH, HO, OO) are a clear indication of the presence of hydrogen bonds. All RDFs show that the small system gives a good description of the first solvation shell, while the large system also includes the second solvation shell. The peak positions of the force-field simulations of Fidler and Rodger [75], indicated by crosses, are close to our results. Note that the good agreement for the CO RDFs, that are potentially sensitive to a proper description of the dispersion attraction, suggests that the absence of the dispersion interaction in our CPMD-BLYP simulation is of limited importance for a proper description of the aqueous solvation of the

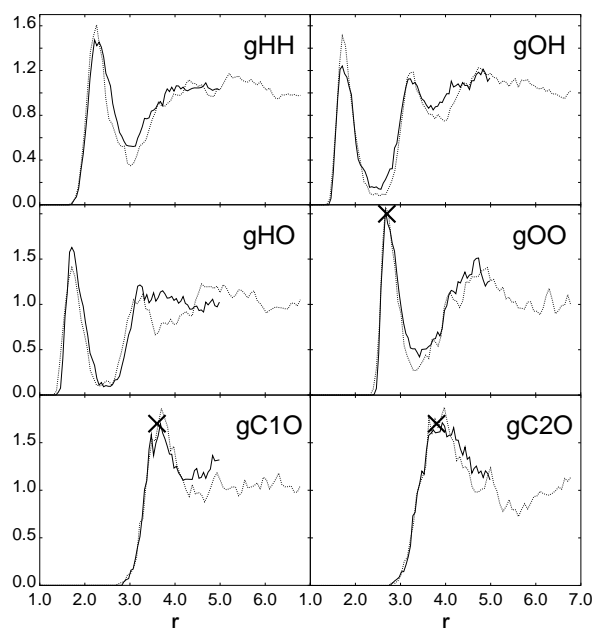


Figure 3.4: Ethanol-water RDFs of the ethanol solutions. All insets show both the results of the small (solid line) and the large system (dashed line). For all graphs gAB indicates that the first atom A belongs to ethanol and the second atom B belongs to water. If the first is a hydrogen, $A=H$, then always the hydroxyl hydrogen of the ethanol is meant. C_1 is the carbon bonded to the hydroxyl group, C_2 is the carbon of the methyl group.

hydrophobic group of ethanol. The position of the first peak of the OH- and HO-RDF indicates that the average hydrogen-bond length is 1.7 \AA for both the ethanol-donor and -acceptor bond. Integration of the OO-RDF up to the first minimum $r = 3.3 \text{ \AA}$ yields on average 3 water molecules in the first solvation shell of the hydroxyl group, in good agreement with the neutron diffraction value of ≈ 3 [68]. Integrating the methyl-oxygen (C_2O) RDF up to $r = 5.7 \text{ \AA}$ indicates that the first solvation shell of the methyl group consists of approximately 21 water molecules. These coordination numbers are of the same order as the experimental estimation of Petrillo et al. [66], who found that there are 18 ± 2 water molecules within 4 \AA from the center-of-mass of an ethanol in aqueous solution.

3.4.2 Ethylene Solvation

In Fig. 3.5 we show the ethylene-water RDFs of the small and large ethylene solution. Both the carbon-oxygen and the intermolecular carbon-hydrogen

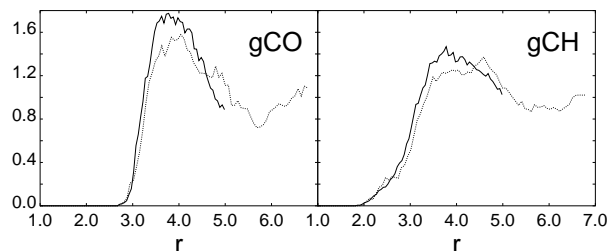


Figure 3.5: Ethylene-water RDFs of the ethylene solutions. The result for small system and large system are given by the solid and dashed line, respectively. H denotes the water hydrogens only.

RDFs are shown. Experimental or simulation data for ethylene in aqueous solution at room temperature were not available. The comparison between the small and large system shows a larger deviation in the first solvation shell than was found for the ethanol solutions. Apparently, the small ethylene system is not able to accommodate properly the first solvation shell. Integration the CO-RDF of the large system up to the first minimum at 5.7 Å yields a hydration shell of 23 water molecules. More spatial information can be extracted from Fig. 3.6 where, for the large system, the distribution of water H-atoms around the ethylene C=C axis is shown. X and Y are components of the vector joining a water hydrogen and the midpoint of the C=C bond. Here, X is the distance orthogonal to the C=C axis and Y the parallel distance. Note that we used the four-fold symmetry to improve the statistical accuracy yielding four identical quadrants. The ethylene hydrogens are shown to visualize the size of the ethylene molecule, but do not indicate any angle dependence around the C=C axis. The figure shows a well defined elliptic region with no hydrogens present except for two weak, but clearly visible, peaks in the mid plane $Y = 0$ at $X = \pm 2.5$. These should be attributed to the presence of $\pi - H$ bonded configurations.

3.4.3 Water Structure

In Fig. 3.7 the calculated water oxygen-oxygen RDFs of both the ethanol and the ethylene solutions are compared to those calculated for pure water. Note

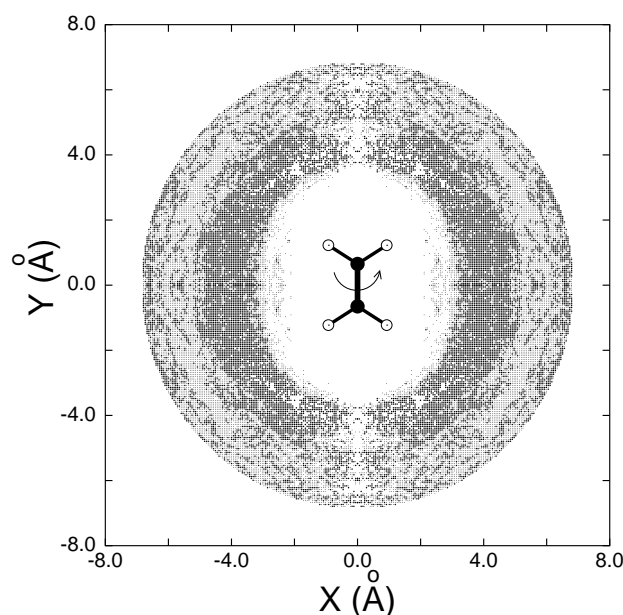


Figure 3.6: Time and rotational averaged normalized density distribution $\rho(r)/\rho_0$ (with ρ_0 is the average hydrogen density) of the water-hydrogens around the solvated ethylene molecule obtained from the large system simulation. The rotational average is about the C=C axis. The four-fold symmetry is imposed so that all quadrants contain the same information. The density regions $1.0 < \rho(r)/\rho_0 < 2.0$ and $\rho(r)/\rho_0 > 2.0$ are illustrated by small and large pixels, respectively.

that the first solvation shell contains a large fraction of the total number of water molecules, especially for the small system. Structural changes due to the solute molecule should therefore be detectable by comparing these systems with the pure water system. For the ethanol solution we see a small drop of the first peak accompanied by a slight increase at the first minimum. This indicates some decrease in the hydrogen-bond structure, when compared to the pure water. As the small ethylene-water solution is too small to accommodate a fully relaxed water solvation shell, Fig. 3.7 only shows the RDFs of the large ethylene system. The RDFs show a similar behavior as for the ethanol solvation, with a slight decrease of the first peak and a small increase of the RDF in the region around the first minimum. The small changes in the structure of the solvating water shell around ethanol and ethylene indicate that the hydrogen-bonded water network is very flexible and can easily open up to accommodate small apolar solute groups with-

out significantly changing its structure. Our findings are consistent with the neutron diffraction experiments of Turner and Soper [67] and force-field molecular dynamics simulation of Fidler and Rodger [75] who also did not find any evidence of structural enhancement in the hydration shell of ethanol. Turner and Soper found that only for larger hydrophobic solutes this effect was experimentally detectable, but even then very small. The same trend was found in a CPMD-BLYP molecular dynamics study of the solvation of methanol [93].

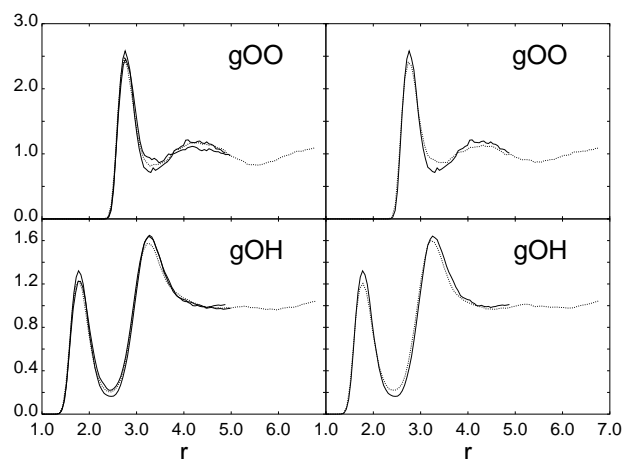


Figure 3.7: Water-water RDFs. Left: the two ethanol solution systems are compared with pure water. Right: the large ethylene solution system is compared with pure water. Thin solid lines indicate the pure water, thick solid lines the small system, and thick dashed lines the large system.

3.4.4 Hydrogen Bonds

To examine the hydrogen bond statistics in the ethanol solution we adopted the definition of Ref. [31]: two molecules are hydrogen bonded if simultaneously the inter-oxygen distance is less than 3.4 \AA and the OHO angle is smaller than 30° . With this definition we found that, for the large system, ethanol oxygen donates on average 0.9 hydrogen bonds and accepts 1.7. For the small system we found 1.0 and 1.6, respectively. This is consistent with the fact that approximately three water molecules occupy the first solvation shell of the hydroxyl group. For comparison, in a CPMD-BLYP simulation we found that methanol in a dilute aqueous solution donates on average 0.9

hydrogen bonds and accepts 1.5 [93]. From the pure water simulation these numbers were measured to be 1.7.

For the ethylene molecule a well defined definition for the π -H hydrogen bond does not exist. To investigate the influence of the π -H hydrogen bond on the solvation structure we looked at the water hydrogen positions relative to the C=C axis (see Fig 3.6). Inside the elliptic region around the C=C bond that is depleted of hydrogens we clearly detect near the center-of-mass of the ethylene molecule at a distance of ≈ 2.5 Å a region with an increased amount of water hydrogens. Integration over this region with $0 < X < 2.6$ and $0 < Y < 1$ yields a value of 0.42. This implies that approximately 40% of the time a water molecule is oriented towards the double bond, forming a π -H bond.

3.5 Solute Dynamics

The time scale of the present simulations (5-10 ps) allows for an analysis of the short-time dynamics. As we mentioned above BLYP underestimates most vibrational frequencies by more than 10 %. However, the frequency shifts upon solvation still provide valuable information and allow for a direct comparison to experiments. Fig. 3.8 shows the spectrum of the velocity auto correlation functions (VACFs) of the ethanol hydroxyl hydrogen for the large ethanol solution and for a single isolated ethanol. The spectrum is of limited accuracy due to the relative short trajectory. However, the calculations show that, upon solvation, there is a clear red shift of about 200 cm^{-1} of the OH stretch frequency. Experimental data for the OH frequency shift of ethanol in dilute aqueous solution is not available. However, the OH red shift is a typical characteristic of a hydrogen bonded liquid. A comparison with measured shifts in liquid ethanol, from 3676 cm^{-1} [108] to 3330 cm^{-1} [109], shows a similar trend. In a CPMD-BLYP simulation of a dilute aqueous solution of methanol we found a similar red shift for the methanol OH frequency of about 200 cm^{-1} .

The vibrational spectrum of the VACF of the ethylene molecule is not shown. As for the ethanol system there was a limited accuracy due to the relatively short calculated trajectory. We observed that upon solvation the CH and CC peaks do broaden. However, an estimate for peak shift falls outside the accuracy of the calculated spectra. The matrix isolation studies of Engdahl and Nelander [82,83] show minor changes in frequencies when isolated ethylene is compared with the ethylene-water complex, with the largest shift being a blue shift of 12 cm^{-1} of the out-of-plane bending mode

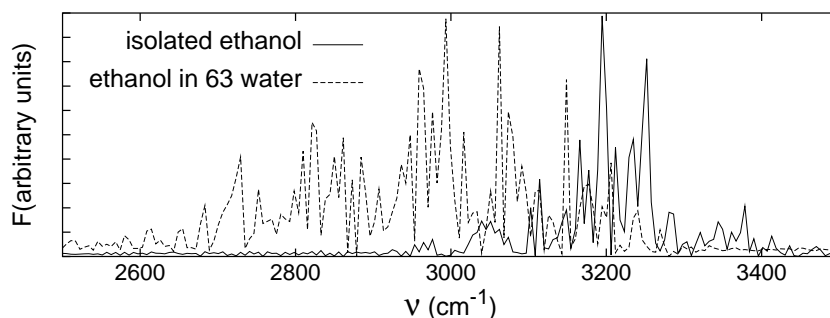


Figure 3.8: Spectrum of the VACF of the hydroxyl ethanol hydrogens in the gas-phase and in the large aqueous system.

($\nu_7 \approx 947 \text{ cm}^{-1}$).

3.6 Polarization

As the electronic structure is an intrinsic part of a CPMD simulation, detailed information about the electronic charge distribution can be obtained. To quantify the charge distribution we used the method of maximally localized Wannier functions that transforms the Kohn-Sham orbitals into Wannier functions, whose centers (WFC) can be assigned a chemical meaning such as being associated with an electron bonding- or lone-pair [110, 111]. It also provides a unambiguous route to assign dipole moments to individual molecules in a condensed phase by assuming the electronic charge to be distributed as point charges located on the WFCs. Obviously, for this procedure to work it is required that the charge distributions of different molecule can be clearly distinguished. This has been demonstrated to hold for liquid water in Ref. [112] where charge overlap of neighboring molecules was rather small. Although we did not perform a similar analysis for the systems of the present study it can be expected that for ethanol-water pairs, that have a very similar hydrogen bonding, the charge distribution overlap will be equally small. The ethylene-water π -H bond is much weaker and therefore expected to show even less charge overlap. We therefore believe that the WFCs provide a meaningful picture of the charge distribution of the systems considered in the present study.

Tab. 3.4 list the calculated dipole moments for the gas-phase systems and the average dipole moment for the solvated systems. The latter were

Table 3.4: Dipole moments obtained by Wannier-function analysis. The liquid water value (a) was taken from Silvestrelli et al. [112, 113]. In the last column the total dipole moments of the gas-phase complexes are given. The first value for water in the E2W complex is the water closest to the ethylene, the second value is the dipole of the other water molecule. The solvated ethanol and ethylene dipole moments are obtained by taking the average of 28 independent configurations from the calculated trajectories.

complex	ethanol/ethylene	water	total
single-water	-	1.82	1.82
liquid water	-	3.0 (a)	
ethanol gauche	1.83	-	1.83
ethanol trans	1.66	-	1.66
gauche P-acc	2.32	2.16	2.76
gauche P-don	1.91	1.91	2.62
trans P-acc	2.13	2.13	2.47
trans P-don	1.91	2.12	2.29
solvated ethanol	3.08	2.97	
ethylene	0.00	-	0.00
EW1a	0.37	1.95	2.10
EW1b	0.33	1.90	2.02
E2W	0.51	2.25 , 2.17	2.41
solvated ethylene	0.51	3.00	

obtained from 28 independent configurations of the small solvated systems. The distributions of these dipole moments are shown in Fig. 3.9. For ethanol we observe an significant increase in going from the gas-phase molecule via the water-ethanol cluster to the fully solvated system. This trend is very similar to what is found for water [112, 113] and methanol [106] and may be considered typical feature of a strongly hydrogen bonding molecule. The calculated distribution of dipole moments of solvated ethanol shows that thermally driven fluctuations give rise to a significant variation ranging from 2.0 D to 4.0 D.

Also for ethylene we observed a significant change of the dipole moment upon solvation. Being apolar in the gas phase, the average dipole moment increases up to 0.5 D when complexed with the water dimer and in solution. In solution it exhibits rare fluctuations where the dipole moment reaches values of up to 1.0 D. Fig. 3.10 shows a typical snapshot where such an extreme high value of the dipole moment is reached. The two WFCs of the

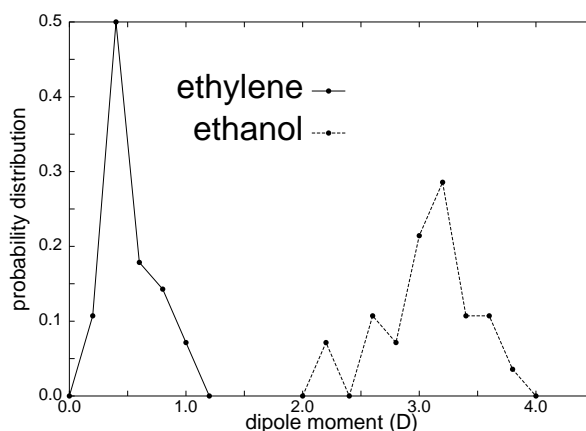


Figure 3.9: Dipole moment distribution of the solvated ethylene and the solvated ethanol molecule.

ethylene double bond are located near the middle of the C=C bond, just below and above the plane of the molecule. In the figure we see clearly that the upper π -WFC is acting as proton acceptor. The π -H bond shifts this WFC further out of plane inducing an ethylene dipole moment orthogonal to the plane of the molecule. The snapshot suggests an enhancement of the ethylene polarization by the fact that the π -H bonded water molecule has in turn a rare four-fold hydrogen-bonded water coordination, with three of the waters donating a proton. This should be seen as a manifestation of the strengthening of the π -H bond upon complexing water molecules with the water molecule that is π -H bonded to the ethylene, a feature found in gas-phase MP2 calculations [87, 89] and also in the present BLYP calculations, and discussed above in section 3.3.

3.7 Conclusions

We have studied the solvation of ethanol and ethylene in water by DFT-based (Car-Parrinello) Molecular Dynamics. Validation of the computational approach by comparing structure and energetics of relevant gas-phase complexes against experimental results and state of the art quantum chemical calculations showed that CPMD employing the DFT-BLYP is capable of qualitatively describing the aqueous solvation of a single ethanol or ethylene molecule.

The structural properties of the ethanol solvation were in good agree-

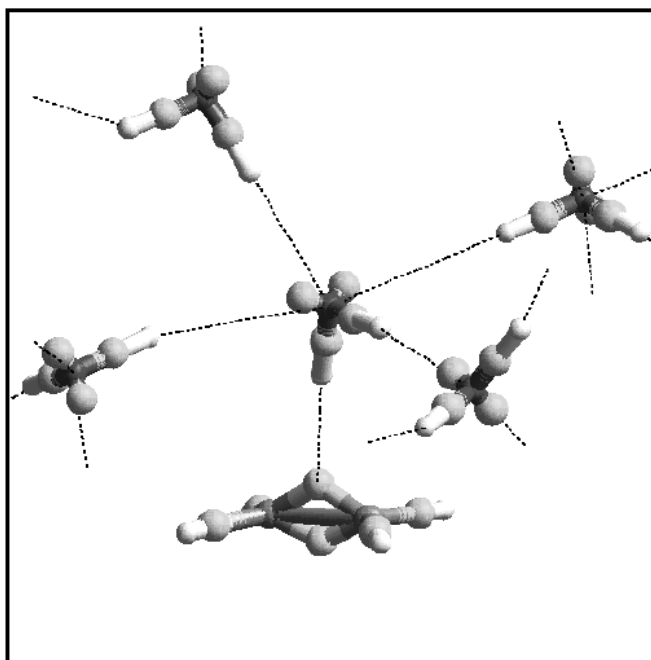


Figure 3.10: A snapshot of ethylene solvated in water. The figure shows a rare configuration where the ethylene molecule has a large dipole moment of ≈ 1 D. For clarity, only the ethylene molecule and five water molecules are shown. Carbon and oxygens are dark grey. Hydrogens are light grey. Besides the atoms also the WFCs are shown (middle grey). The water molecules have WFCs in both OH bonds and two aside the oxygen corresponding to the lone pairs. Ethylene has four WFCs along the CH bonds and two out-of-plane WFCs between the carbons corresponding to the π bond. The dashed lines indicate the hydrogen bonds. One the protons of the central water molecules points towards the 'upper' π bond WFC. This is the manifestation of the π -H bond in this picture.

ment with both neutron diffraction data [67] and force-field MD simulations [75]. We found that in aqueous solution ethanol accepts on average 1.7 hydrogen bonds and donates on average 0.9 hydrogen bonds. For ethylene we found it has approximately 0.4 π -H bonds with a water molecule. Both for ethanol and ethylene the simulations provide no structural evidence for hydrophobic hydration: the structure of the hydrating water shell was not enhanced compared to that of pure water. The calculation even indicated a slight decrease in the structure. For aqueous ethanol the calculated red shift of the hydroxyl vibration upon solvation was consistent with experimental

findings.

Analysis of the electronic charge distribution by means of Wannier functions showed that the interaction with water can significantly increase the dipole moment of the ethanol and the ethylene molecule. The average dipole moment of ethanol increases from 1.8 D in the gas phase to 3.1 D in aqueous solution. Ethylene, that is apolar in the gas phase, attains an average dipole moment of 0.5 D in solution. We identified configurations, with a π -H bonded water molecule that has a rare four-fold hydrogen-bonded water coordination, where the instantaneous dipole moment of ethylene takes values of up to 1 D. Such configurations with large solute dipole moments may also play an important role in activating chemical reactions involving the solutes as we have seen in a CPMD-BLYP study of the acid-catalyzed hydration of ethene [56]. The strong polarization effect raises questions towards the common consideration of thinking ethylene as a apolar and hydrophobic molecule. The electronic charge analysis also points out the necessity of polarizable force fields for both ethanol and ethylene when dissolved in water. The present results may be considered valuable for the design of such force-fields.

Chapter 4

Hydration/Dehydration Reaction between Ethanol and Ethene¹

We performed a DFT-based based Molecular Dynamics simulation of the acid-catalyzed hydration of ethylene in aqueous solution under ambient conditions. As this reaction can be considered as a rare event we applied a constrained method. This method yields the free-energy barrier and can give clues how spontaneous events occur. Our simulation confirms the Ad_E3 mechanism at which two bonds are formed simultaneously causing a sudden change in its electronic structure and a flip from negative to positive constraint force. The specific structure of the hydrogen bonded network plays a crucial role in both the hydration reaction as in the reverse reaction. As it is not included in the imposed reaction coordinate it leads to hysteresis in a small window of the reaction coordinate.

4.1 Introduction

Hydration of ethene (also called ethylene) has been used in industry for large scale production of ethanol since several decades. In the industrial process the hydration of ethylene proceed in a gas-phase pressure recycling operation over solid phosphoric acid catalysts, manufactured by impregnating silica carriers [114, 115]. It may sound surprising that the reverse reaction, the dehydration of ethanol, is also used in industry for the synthesis

¹This chapter is based on work in progress, to be published as a letter and as a full article: Titus S. van Erp, Evert Jan Meijer, "Ethene Hydration in Acid Solution. Revealing the Solvation Structure of Reacting Pathways" and Titus S. van Erp, Evert Jan Meijer, "Ab Initio Study: Hydration/Dehydration Reaction between Ethanol and Ethene"

for ethylene. Although this technology is now only of very limited importance because of the current availability of ethylene from more economical favorable petroleum based sources. However, the situation can be different for developing countries, that have scant petroleum resources, but do have substantial amounts of fermentation ethanol available derived from agricultural waste products. For them, a vapor-phase dehydration of ethanol is often the most efficient and economical way to produce ethylene [116]. The presence of the acid is of great importance for both industrial processes. It allows for a different reaction mechanism, which lowers the activation energy and increases the reaction rates yielding an energetically and economically favorable conversion. Baliga and Whalley [117] have measured the hydration of ethylene in dilute aqueous perchloric acid in the range of 170-190 C° at 100 bars and estimated a barrier height of 33.3 ± 1.0 kcal/mol. Perchloric and sulfuric acids are known to be better catalysts than phosphoric acid in terms of conversion. However, industry uses phosphoric acid instead as side reactions are more suppressed for this catalyst. Still, the basic mechanism is roughly same for all acids. The important catalytic group of the acid is the positively charged proton that can easily split off forming a hydronium ion H_3O^+ when solvated in water. Therefore, if one wants to study the general mechanism of the acid catalyzed hydration in a molecular simulation, it is sufficient, to a good approximation, to limit the presence of the acid by the hydronium only.

If the catalyst is absent, the reaction barrier is about twice as high. MP2 calculations show energy barriers of 57.7 kcal/mol for the hydration and 67.1 kcal/mol for the dehydration reaction [118]. Consequently, this reaction will only occur under extreme conditions. The non-catalyzed dehydration reaction is experimentally observed for an ethanol molecule that is strongly vibrational excited via a chemical activation [118, 119] or by a pulsed HF [120] or CO_2 laser [121]. Theoretical calculations of rate constants were performed in the high temperature regime (700 K- 2500 K) [122] and for the laser induced decomposition of ethanol [123]. These measurements and calculations show that, for these conditions, the dehydration process yielding ethylene and water is the most favorable unimolecular decomposition reaction of ethanol.

The presence of an acid changes the reactive pathway and enhances the rate. If, in addition, the reaction evolves in a solvent, the reaction mechanism can be very complex as solvent molecules may participate in the reaction process. However, also the gas-phase association reaction between H_3O^+ and ethylene, and its reverse, the gas-phase dissociation of proto-

nated ethanol, have attracted scientists interest. The gas-phase association reaction is, for example, of importance for astronomical physics. It is believed that this process is responsible for the presence of ethanol observed in interstellar clouds [124]. The precise mechanism of this association reaction is still matter of scientific debates. There is evidence that the gas-phase formation of protonated ethanol $\text{C}_2\text{H}_5\text{OH}_2^+$ has a π -complex $\text{C}_2\text{H}_4\cdot\text{H}_3\text{O}^+$ as an intermediate. In this complex the proton of the hydronium is bonded to the π -electrons of the double bond of the ethylene. For a single proton, computational studies to proton affinities have shown that the bridged π -bonded structure is more stable than the open ethylene cation [125–128]. Likewise, for the hydronium-ethylene complex, experimental [129–133] and theoretical [128, 132, 134] studies have shown the stability of the π -bonded ethylene-hydronium ion molecular associate $\text{C}_2\text{H}_4\cdot\text{H}_3\text{O}^+$. Therefore, it is believed that the mechanism for the gas-phase reaction between ethylene and the hydronium ion implies an indirect carbon protonation. First, the proton of the hydronium will approach the middle of the ethylene forming a bond to the π -electrons of the double bond. In a second stage, this proton will move to a carbon site and the H_2O group to the other site forming the open structure of the protonated ethanol.

In general, one has to be cautious to make hard conclusions on the reaction mechanism in a solution based on the knowledge of the gas-phase results. An early experimental study to the acid catalyzed hydration of alkenes suggested that this π - complex also existed as an intermediate in the condensed phase [135]. Indeed, also for other alkenes than ethylene, the bridged structure seems to be more stable than the open cation structure [136, 137]. However, when solvation is taken into account energetics can change considerably. Jorgensen and Munroe have shown that the open ethylene cation is getting more and more favored to the bridged structure upon increasing solvation in HCl clusters [127]. Moreover, measurements of ethylene hydration in concentrated sulfuric acid advocate the proton attachment to the alkene directly through the formation of a carbocation [138]. Therefore, the general opinion is nowadays that for the hydration of olefins in the condensed phase the formation of an intermediate π -complex is not necessary.

Reactions in liquid water are of special interest as the role of the hydrogen bonded network of liquid water in chemical processes is presently subject of many discussions. In a broader perspective, this reaction can be related to other reactions in solvents in which proton transfer plays an important role. In this context we can refer to many other Car-Parrinello stud-

ies (see e.g. [28,139–142]). Of great importance is also the pioneering work of Hynes and co-workers (see e.g. Ref. [143–148]) in which they deal with the solvent influence in proton transfer reactions by using a reaction coordinate that incorporates the degrees of freedom of the solution. More specifically related to this work is the *ab initio* study of Mohr et al [55], in which they showed that the hydrogen bonded network around an ethylene-cation radical can stimulate or prevent the hydration reaction depending on its precise structure. Recently, we showed that ethylene, usually considered as a hydrophobic molecule, can be induced to instantaneous dipole moments of ≈ 1 . D due to the polarization by the surrounding water molecules [149]. This points out that the presence of the solvent introduces many new effects that can significantly alter the reaction mechanism in comparison with the gas-phase reaction.

Molecular simulation provides an approach to study the microscopic behavior of liquids complementary to experimental studies. Standard molecular simulation techniques are based on empirical force fields that are designed to reproduce a selection of experimental data. Obviously, molecular simulations based on these potentials do not provide a picture completely independent from experiment. Moreover, the reliability of the results at conditions that are significantly different from those where the potential was designed for, may be questionable. Density functional theory (DFT) based molecular dynamics (MD) simulation, such as the Car-Parrinello molecular dynamics method [2], where the interactions are calculated by accurate electronic structure calculations, provides a route to overcome these limitations. Important advantages of DFT-MD over force-field MD are that it intrinsically incorporates polarization, that it accounts for the intra-molecular motion and therefore allows for a direct comparison with spectroscopy of intra-molecular vibrations, and that it yields detailed information on the electronic properties, such as the energy levels of electronic states and the charge distribution. The most important advantage is that DFT-MD, in contrary to force-field MD, is capable of studying chemical reactions. The making and breaking of chemical bonds are accompanied with huge changes in the electronic density. The forces in intermediate stages, during the chemical event, cannot be known a-priori, which makes the design of accurate force fields for these situations an almost impossible task.

We studied this reaction in both directions under ambient conditions in the presence of an aqueous solution. We limited the catalytic contribution to the addition of a single H_3O^+ to the solution. The accuracy of the used *ab initio* method for our system is supported by earlier work, such as the studies

of liquid water [14,90,91] and aqueous solvation [92–94,149]. This chapter is organized as follows: First we will discuss the computational method in section 4.2. In section 4.3 we will show the specific stages of the reaction when the system is forced over the barrier and estimate the free-energy barrier in section 4.4. In section 4.5 we will look in more detail to structural changes as function of our chosen reaction coordinate and as function of time when the reaction coordinate is fixed. Then in section 4.6 we will apply the Wannier analysis to obtain more insight in the electronic changes in the system. In section 4.7 we will look at the specific hydrogen bonded network structure to determine the topology conditions under which the reaction can take place. The conclusions are summarized finally in section 4.8.

4.2 Computational Methods

We performed MD simulations using the CPMD package [35]. This program combines MD motion of the ions with electronic structure calculations using the Kohn-Sham [9] formulation of Density Functional Theory (DFT) [8] in a efficient way using the Car-Parrinello algorithm [2]. This algorithm treats the basis set coefficients of the Kohn-Sham orbitals as dynamical variables and avoids in this way the expensive self-consistent field calculation, that have to be solved iteratively each time step for ordinary Born-Oppenheimer dynamics. Although the electronic wave function is never in its ground state during the simulation, it stays sufficient close so that the calculated forces on the ions are close to the true forces and deviations are not systematic and cancel out [21]. For the exchange correlation energy we chose the BLYP functional that includes Becke's [11] gradient correction for the exchange energy of the uniform electron gas and the correlation functional of Lee, Yang and Parr [12]. This functional has proven to give a good description of liquid water [13,14], for proton transfer [28,139,141,142], for the aqueous solvation of alcohols [93,149], and the aqueous solvation of ethylene [149]. CPMD uses a plane wave basis set, matching the periodicity of the periodic box with waves up to a kinetic energy of 70 Ry. An uniformly distributed negative background charge was added to compensate the positive charge of the proton. Pseudopotentials were used to limit the number of electronic states to the valence electrons. Interaction with the core electrons was described by semi-local norm-conserving Martins-Troullier pseudopotentials [20]. Cut-off radii for H, O and C atoms were chosen to be 0.50 a.u. , 1.11 and 1.23 a.u. respectively for both $l=s$ and $l=p$ terms. The simulations were performed in a periodic box of length 10.19 Å , that matches to the experimental density

of the ethanol solution under ambient conditions. Starting from the ethylene side of the reaction, the box contained 31 water molecules, one H_3O^+ for the acidity, and one ethylene molecule. Temperature was controlled by a Nosé-Hoover [42–44] thermostat and fixed at 300 K. The fictitious mass associated with the plane-wave coefficients is chosen at 900 a.u., allowing a time step for the MD ions motion of 0.145 fs.

Ab Initio MD simulations are computationally demanding and typical simulation times are consequently limited to ≈ 10 pico seconds. In view of the large reaction barrier, the chance that within this short period a spontaneous reaction occurs is negligible. To enforce the reaction we use the constrained dynamics method combined with thermodynamic integration (see e.g. [23]). This method requires a definition of a reaction coordinate that connects reactant and product side. With the reaction coordinate fixed by constrained MD, the time average constraint force can be measured. Proper integration of this mean force over the reaction coordinate yields the reversible work or the difference in free-energy. By making a series of simulations, each with a different value of reaction coordinate in the intermediate range between reactant and product state, one can numerically approximate this integration and obtain the free-energy profile and thus the activation and reaction energy. Recently the corrections to the mean force to obtain the exact free-energy have been outlined [24, 25]. However it has been shown that this kind of corrections to similar systems are of minor importance [28, 150]. We have chosen a reaction coordinate Q similar to [28], that controls the transfer of the proton to the ethylene carbon. This reaction coordinate Q is defined as the difference between two distances: $Q = |\text{O-H}| - |\text{H-C}|$ where O and H are the oxygen and hydrogen of the hydronium ion and the C is one of the carbons from the ethylene molecule. This constraint controls the proton transfer from the hydronium ion to the ethylene molecule. This is the reaction step with the highest energy barrier and is the only part of the reaction, that we force by our constraint. A negative value of Q implies that the proton H is close to the oxygen O forming a hydronium ion H_3O^+ , while positive value means that the proton is closer to the carbon C. Note that this definition of Q allows a lot of flexibility. When Q is fixed, the $|\text{O-H}|$ and $|\text{H-C}|$ distances can still fluctuate in time. This allows the reaction to evolve in a more free way, than if we would have chosen for a more simple reaction coordinate such as the $|\text{C-H}|$ distance. In total we performed 21 simulations in which Q was fixed at different values in the range $[-1.59 \text{ \AA} : 1.59 \text{ \AA}]$. The values close to the transition state around $Q = 0 \text{ \AA}$ were performed over simulation times of approximately 10 ps. In this range hysteresis oc-

curs. Therefore, performed in this range simulation series with increasing and decreasing Q , in the following indicated by "(hydration)" and "(dehydration)". For the values $|Q| > .212 \text{ \AA}$ shorter simulation times of 5 ps were performed. In this range there is no hysteresis. All simulation data were sampled after 0.7 ps of equilibration time.

4.3 Mechanism

We started with a simulation without any constraint at the ethylene side. The extra proton in solution can be freely transferred along hydrogen bonds between water molecules. We waited until a hydronium was formed close to the ethylene and then we applied the constraint. Starting point of the series of simulations was a constraint value of $Q = -1.06 \text{ \AA}$ at which a simulation of 5 ps was performed. In this simulation the selected proton is on average 2.07 \AA away from the carbon atom and 1.01 \AA from the oxygen. In panel A) of Fig. 4.1 we see a particular snapshot of this simulation. One of the other non-constrained hydrogens of the hydronium ion has moved towards one of the neighboring water molecules. We observed that the proton transfers of these hydrogens for this Q value was limited to the first neighboring water molecules and that these hydrogens always returned to the constrained-oxygen. Thus, the positively charged proton was not lost in the solution but remained close to the ethylene. Therefore, extra constraints on the hydronium ion were not needed.

The $Q = -1.06 \text{ \AA}$ simulation was followed by a series of subsequent simulations where we moved the proton towards the ethene-carbon. Panel B) of Fig. 4.1 shows a snapshot of the simulation $Q = 0.0 \text{ \AA}$ (hydration), where the proton H is forced to stay exactly in between the oxygen and the carbon. Proton transfers of the other non-constrained hydrogens to neighboring water molecules were no longer observed, indicating that the molecular bonds between the oxygen and these hydrogens are stabilized. The $|\text{C-H}|$ distance is strongly fluctuating in the range $[1.24 \text{ \AA} : 1.58 \text{ \AA}]$. Bonds between atoms in this figure are drawn according to distance dependent bond definitions, that assigns a CH bond when the interatomic distance is less than 1.31 \AA and a OH bond when it is less than 1.27 \AA . The simulation movie for this constraint value shows a constant appearance and disappearance of the two bonds. In section 4.6 we try to give a more elucidated definition of a bond not only dependent on the distance.

Panel C) of Fig. 4.1 is a snapshot for the constraint value of $Q = +0.106 \text{ \AA}$ (hydration). The average $|\text{C-H}|$ distance is decreased compared to pre-

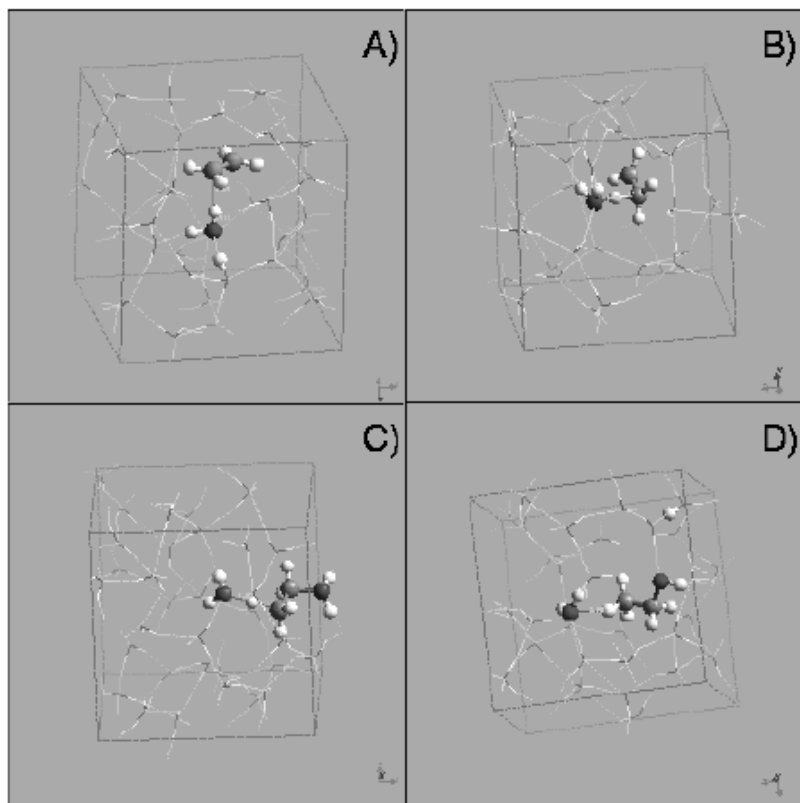


Figure 4.1: Four snapshots from four different simulations, each with a different constraint value for Q . The important atoms and bonds are visualized by ball stick representation, the others by sticks. Dark grey indicate oxygen, white hydrogen and middle grey carbon atoms. The white dashed lines indicate the hydrogen bonds between the waters and there is a dashed line that connects the three atoms that determine Q . The constraint values are, A): $Q = -1.06 \text{ \AA}$, B): $Q = 0.0 \text{ \AA}$ (hydration), C): $Q = 0.106 \text{ \AA}$ (hydration) and D): $Q = 1.06 \text{ \AA}$. Each picture shows one typical time shot of four simulations.

vious $Q = 0.0 \text{ \AA}$ (hydration) simulation. However, in this typical snapshot the instantaneous distance is a bit higher than in panel B). The interatomic distances are beyond the distance of what one would normally assign for a bond. Therefore, the constrained-proton looks isolated not forming a bond with either the carbon or the oxygen atom. In section 4.6 we show however that this picture is misleading. Contrary to the distance dependent

bond definitions, but based on electronic density, one does can speak of a stable CH bond in panel C) and not in panel B). At the other side of the ethylene, one of the waters out of the solution has formed a bond to the other carbon. This reaction step happens spontaneously when the reaction coordinate Q is fixed at this value.

The last panel D): $Q = +1.06 \text{ \AA}$, shows that the proton has moved closer to the carbon almost at the ideal CH bond length of 1.10 \AA . At the other carbon the CO bond has remained stable for all $Q > 0.106 \text{ \AA}$ and protonated ethanol has been formed. Subsequently, the protonated ethanol loses one proton at the oxygen site. Apparently the barrier for deprotonation of the protonated ethanol is small or absent. This is consistent with a recent study of hydrated methanol clusters [151] that show that for the largest cluster considered ($\text{H}^+\text{CH}_3\text{OH}(\text{H}_2\text{O})_6$) energy differences between configurations with H^+ attached to either methanol, water, or the H^+ in between methanol and water are very small. The issue of deprotonation of aqueous protonated alcohols is presently under investigation [152].

4.4 Energetics

The time-averaged constraint force is plotted in Fig. 4.2. We see that, for large negative and positive Q values, the force is approaching zero. For the transition region $Q \approx 0 \text{ \AA}$ we detect sudden jumps in the mean force. From left to right the force changes from negative to positive sign. This is directly related to the formation of the CO bond at the other carbon. When we start from a configuration where this CO bond is established and decrease the reaction coordinate again, this CO bond is not directly broken. The solvation shell has to rearrange before the reaction will occur backwards, which gives rise to hysteresis. Similar effects were found in the S_N2 reaction between CH_3Cl and Cl^- [150] and in the acid catalyzed addition of water to formaldehyde [28] in aqueous solution.

For the reaction coordinate $Q = 0.053 \text{ \AA}$ we found two almost stable configurations, $Q = 0.053 \text{ \AA}$ (hydration) and $Q = 0.053 \text{ \AA}$ (dehydration). The $Q = 0.053 \text{ \AA}$ (hydration) configuration was stable during the complete 10 ps simulation. For the $Q = 0.053 \text{ \AA}$ (dehydration) configuration the CO bond stayed intact for about 8 ps when suddenly this bond was broken and the constraint force swapped from a positive to negative value. Also in the $Q = 0.106 \text{ \AA}$ (hydration) and $Q = 0 \text{ \AA}$ (dehydration) simulations, a sudden reactive event was observed after 2 and 3 ps respectively.

The mean force was integrated by taking the stable point values and

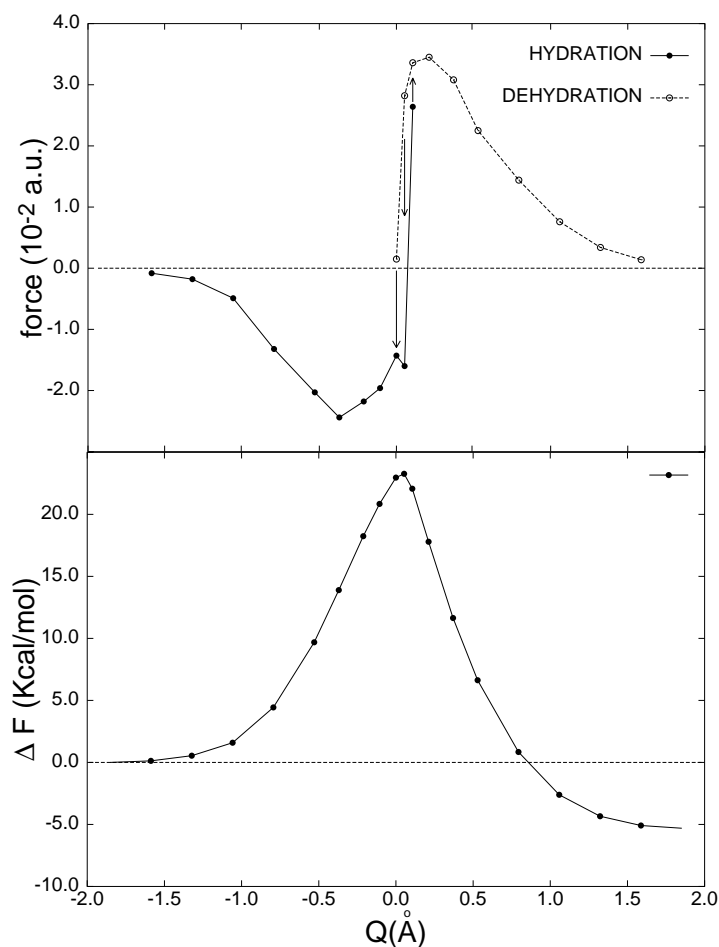


Figure 4.2: Top: the average mean force as function of reaction coordinate Q . Values for $|Q| > .2$ Å were obtained after averaging over a MD run of at least 5 ps, for $|Q| < .2$ Å more than 10 ps simulation runs were applied. Each run started with .7 ps of equilibration for which no data was collected. Both direction (hydration) and (dehydration) are indicated by solid and dashed lines respectively. There is hysteresis in a small window with a strong change in the mean force. The arrows indicate that during the simulation run a flip in the constraint force was observed in the direction of the arrow. Bottom: The integrated curve yielding a free-energy barrier of 23.3 kcal/mol and a reaction energy of 5.3 kcal/mol.

the weighted average for the two $Q = 0.053$ Å cases. The obtained free-energy barrier is 23.3 kcal/mol, while the reaction energy is 5.3 kcal/mol as

is shown in the lower panel of Fig. 4.2. The barrier is 10 kcal/mol lower than the experimental value of Baliga and Whalley [117], who found a reaction barrier of 33.3 ± 1.0 kcal/mol. However, static BLYP calculations with perchloric acid in the gas-phase shows a similar barrier of 24.7 kcal/mol [153]. It is a known feature that DFT tends to underestimate barrier heights to some extent (see for example Ref. [150]). The comparison between our result and the gas-phase BLYP calculation of Ref. [153] is complicated by two factors. First of all the mechanism is different, where in the gas-phase reaction the hydrating water molecule also provides the proton, whereas in the our solution-calculation the proton providing water and the hydrating water are different molecules approaching the ethylene molecule from opposite side. Secondly, it should be noted the approximate reaction coordinates we have chosen gives rise to a underestimate of the true reaction barrier. To arrive at a better insight of the effect of the solvent we performed some calculations of protonated hydrated ethylene configurations mimicing the mechanism observed in solution. Preliminary calculations show that for a cluster with a single H_3O^+ providing the proton and a single hydrating water molecule at the opposite side of the ethylene molecule the barrier is absent. A small barrier of ≈ 4 kcal/mol appears of both the H_3O^+ as the hydrated water molecule are solvated by two water molecules. This suggests that the solvation effects play a significant role increasing the barrier in the order of 20 kcal/mol. Note that these processes are highly unlikely in the gas-phase as they require simultaneous collisions between three or more molecules.

Obviously, the catalytic effect is significant as a decrease of 30 kcal/mol implies an increase to the reaction rate in orders of magnitude. We can make an additional comparison to acid catalyzed hydrations of other alkenes. An early measurement of Lucas and Liu [154] showed a reaction barrier of 18.9 kcal/mol for the acid-catalyzed hydration of trimethylethylene. This lower activation energy is consistent with the view that larger molecules are better able to delocalize the charge allowing for a more stable transition state [155].

The calculated reaction energy in solution of 5.3 kcal/mol is lower than the MP2 value of 9.4 kcal/mol [118,119] and the BLYP-CPMD value of 9.09 kcal/mol [149]. However, the latter values are just energy differences and do not take the change in entropy into account. The entropy is higher at the ethylene side, as here is one extra water molecule in solution that can freely move. This lowers the free-energy at that side and thus decreases the free-energy difference between reactant and product side, or equivalently the reaction energy.

4.5 Structural Properties

In Fig. 4.3 we showed two relevant time averaged interatomic distances as function of the reaction coordinate. The distance $|CH|$ is the distance between the carbon and the hydrogen, that is partly controlled by the constraint Q . The other distance $|CO|$ is the distance between the other carbon and the closest water-oxygen. Note that this is not always the same oxygen as water molecules move to and away from the ethylene molecule during the simulation. Surprisingly, in the reactive region around $Q = 0 \text{ \AA}$ there is no noticeable decrease in the $|CH|$ distance, whereas the mean constraint force shows a sudden change from a negative to a positive value. The depicted $|CO|$ distance instead does show a clear and sudden decrease as function of Q in this region. The hysteresis effect is also clearly noticeable here, while for the $|CH|$ curve the (hydration) and (dehydration) values almost completely coincide. Furthermore, when looking to the $|CH|$ curve, we see that it is gradually decreasing and it is difficult to pinpoint where a real molecular bond is formed. Finally, for $Q > .5 \text{ \AA}$ the distance seems to be converged to its normal bond length of 1.10 \AA .

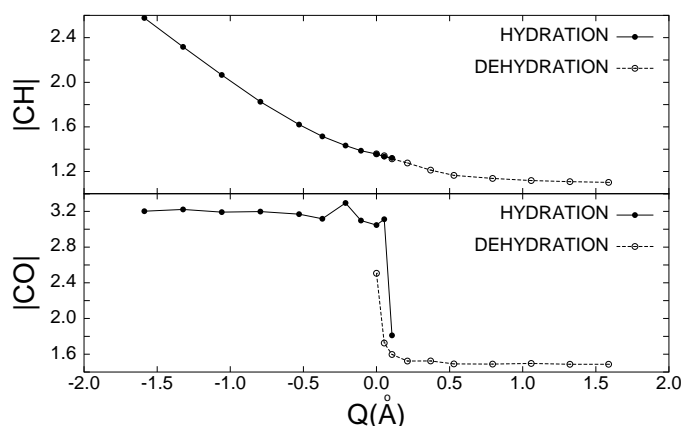


Figure 4.3: Time averaged distances as function of the constrained value Q . Top: $|CH|$ is the distance between the constrained hydrogen atom and carbon atom. Bottom: $|CO|$ is the distance between the other carbon and the closest oxygen atom from the liquid water. Note that this oxygen by this definition does not necessarily belong to the same water molecule. The (hydration) and (dehydration) indicate that the simulation started from the ethylene or ethanol side respectively

In Fig. 4.4 we show the time evolution of these two distances for two

specific 'reactive' simulations: $Q = 0.053 \text{ \AA}$ (dehydration) and $Q = 0.106 \text{ \AA}$ (hydration). Clearly the $|\text{CO}|$ distance and the constraint force are strongly correlated, with the two graphs almost as if they were mirror reflections. The $|\text{CH}|$ distance shows almost no correlation with the constraint force. For the $Q = 0.106 \text{ \AA}$ (hydration) we can detect only a bit more fluctuative behavior in the reactive time domain. As the constraint force acts on these atoms and a fixed reaction coordinate Q allows changes in the $|\text{CH}|$ distance, one would expect a stronger effect. Fig. 4.4 shows, however, that the constraint force changes sign, indicating that the proton is now attracted to the carbon instead of to the water oxygen, without influencing the interatomic $|\text{CH}|$ distance.

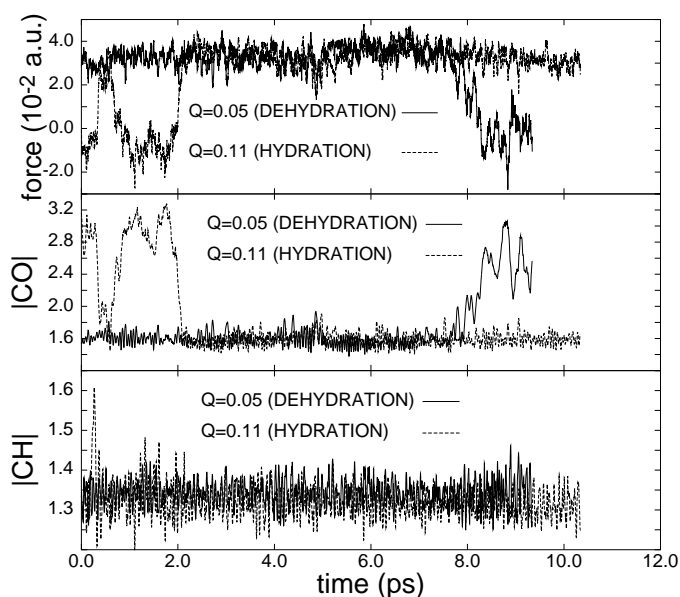


Figure 4.4: constraining force, distances $|\text{CO}|$ and $|\text{CH}|$ as function of time for two 'reactive' simulations: $Q = 0.053 \text{ \AA}$ (hydration) and $Q = 0.106 \text{ \AA}$ (dehydration). Average values of each these two simulation runs are depicted in Fig. 4.2 and Fig. 4.3 as one point each.

4.6 Electronic Structure

From the previous we can conclude that, although the constraint force suddenly changes, the structural changes occur at the other side of the molecule.

Somehow the structural changes at that side of the molecule must change the affinity of this proton to form a bond with the carbon or with the oxygen. A change in the electronic structure must be the cause of this phenomenon. To quantify the electronic density distribution we used the method of maximally localized Wannier functions whose centers (WFC) can be assigned with a chemical meaning such as being associated with an electron bonding or lone-pair [111] and can also be used to calculate individual dipole moments of molecules in solution [106, 112, 113, 149]. The dipole moment of a single molecule can then be approximated as if it consists of point charges located at the atom positions and on the WFC positions, at atom positions a positive charge equal to its atom number minus its core electrons and at the WFCs a negative charge of $-2 e$. The total number of WFCs is the total number of valence electrons divided by two and its positions can give a clear insight in the electronic changes of the system during a reaction. From our series of simulations we took a few snapshots at important points just before and after a reaction. On these configurations we applied the Wannier function analysis to detect the electronic changes. We did not make enough calculations to obtain accurate statistical averages, but our aim was to see on a more qualitative level how the electronic structure is changing in the reactive process. Ethylene has in total six Wannier orbitals, four at the CH bonds and two, forming the double bond in the middle of the carbons just below and above the plane of the ethylene molecule. The water molecules have in total four WFCs tetrahedrally oriented around the oxygen, two of them forming the OH bonds and the other two indicating the lone pairs. The situation for hydronium is more or less the same, but then three form an OH bond and only one WFC is left at the oxygen site. Fig. 4.5 gives in a cartoon illustration of the observed structural and electronic changes of the reaction.

Panel 1) is the typical situation for Q around -0.5 \AA . The hydronium OH bond is directed towards the middle of the CC-bond. The WFC below the ethylene plane is attracted by the positive charge of the hydronium and is shifted slightly towards this proton. This cause an increase of the angle α_1 from 27° to 35° and induces a dipole moment orthogonal to the plane of the ethylene molecule with a strength of approximately 1.9 D. This is a significant dipole moment, approximately equal to a water molecule in the gas-phase [112, 113]. The interaction between the hydronium OH bond and the WFC of the double CC bond can be considered as a kind of hydrogen bond. In Ref. [149] we showed that even in a neutral aqueous solution this effect is also significant and can induce instantaneous dipole moments of

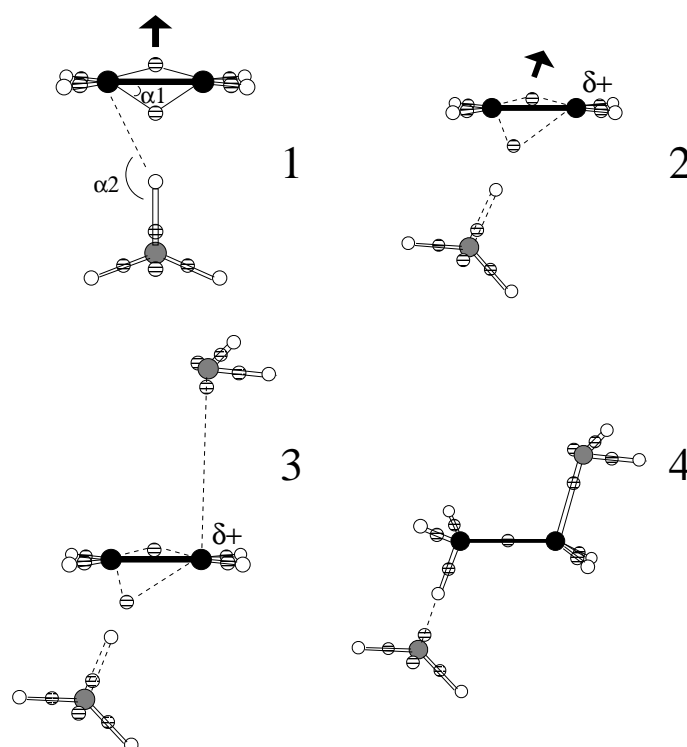


Figure 4.5: Cartoon representation of the change of atomic configuration and WFCs along the reaction path. The solid dark circles are the carbon atoms, the gray circles the oxygen and white circles the hydrogen atoms. The striped circles depict the WFC positions.

1.0 D to a solvated ethylene molecule. It is not surprising that the effect of the hydronium ion is even stronger. The upper Wannier orbital has also moved a bit downwards to the middle of the CC bond, but this shift is much smaller.

Step 2) is the situation for Q around zero. The constrained proton is in the middle of the carbon and the oxygen. The induced dipole moment is no longer exactly orthogonal as the lower WFC has moved away from the middle and is now closer to the left carbon. α_1 has become approximately 50° and α_2 is approximately 120° . Clearly, the proton attack is directing to the carbon site and is not forming a bridged structure. This observation is consistent with the experiments of Nowlan and Tidwell [138] and the calculations of Jorgensen and Munroe [127], that indicate that the protonation in

a solution is approaching directly to the carbon site and that the formation of an intermediate π -bonded complex is not necessary.

Step 3) is a situation for $Q = 0.106 \text{ \AA}$ (hydration). α_1 has increased till $50\text{-}80^\circ$, while α_2 is still around $120\text{-}130^\circ$. The lower WFC is further away from the right carbon, which make this side more positively charged. It is a general feature of protonation, that the charge donation to the proton comes primarily from the terminal atom rather than the atom bonded to the proton [126]. Incidentally, water molecules at the other side approach the other carbon at distances of 2.6 \AA , but usually are rejected back unless this coincides with a ideal solvation structure.

This brings us to step 4), the same Q value, but later in time. The incoming water molecule and the solvation structure around it have found a proper configuration for a nucleophilic attack. The CO distance rapidly decreases and forms a bond. Fig. 4.4 shows that there is a chance that directly thereafter this water is rejected back to the solution. The specific water configuration has decreased the reaction barrier, but this is also true for the backward reaction. This can induce a sequence of hydration and dehydration reactions, but at some moment the situation stabilizes and the water gets definitely attached to the carbon. The $|\text{CH}|$ distance in this reaction process does not change that much, but the lower WFC flips over $30\text{-}40^\circ$ more in line to the CH bond, yielding $\alpha_1 \approx 100^\circ$ and $\alpha_2 \approx 170^\circ$. Whereas first, this WFC was playing part in both the bonding between the two carbons and between hydrogen and the carbon, now its function is only the CH bond. The increase of α_1 over more than 90° indicates that this WFC does no longer take part in the CC bond. The constraint force on the proton H is first repulsive with respect to the carbon, while after the transformation it becomes suddenly attractive. The fixed constraint Q prevents however that the CH bond will fully relax to its rest value. The carbon, the WFC, and the proton are almost in line according to $\alpha_2 \approx 170^\circ$. This gives a clear indication that the CH bond is formed. Moreover, the fact that now only one WFC is positioned in the middle of the two carbons followed by the out-of-plane bending of the ethylene-hydrogens shows clearly the change from sp^2 to sp^3 hybridizing of the CC bond.

The Wannier analysis shows that both molecular bonds, the CH and the CO bond, are formed simultaneously. This indicates that the process occurs following an $\text{Ad}_{\text{E}}3$ -type mechanism. Conversely, the dehydration reaction follows the E2 mechanism, being the inverse of the $\text{Ad}_{\text{E}}3$. The alternative $\text{Ad}_{\text{E}}2$ hydration and E1 dehydration mechanisms imply a stable ethylene cation as intermediate. Although, several undergraduate text books use

the E1 mechanism for the description of the acid catalyzed dehydration of ethanol (and, similarly, an $\text{Ad}_{\text{E}2}$ mechanism for the acid catalyzed hydration of ethylene), the current opinion is that a concerted E2-type mechanism more probably governs the hydration of primary alcohols [5,6]. Consistent with this view are the experiments of Baliga and Whalley that indicate a transition state for the hydration of ethylene with at least one firmly bound water molecule [117].

4.7 Hydrogen Bonds

In this section we will discuss the role of the hydrogen bonded network of the liquid water for this reaction. The constraint values for which we observed a spontaneous reaction are: $+0.106 \text{ \AA}$ (hydration) and 0.0 \AA (dehydration), $+0.05 \text{ \AA}$ (dehydration). The spontaneous reaction step that is not controlled by the constraint is the formation or breaking of the CO bond. From our simulation with $Q = +0.05 \text{ \AA}$ (dehydration) we saw that the protonated ethanol and thus the CO bond was stable for about 8 ps, when suddenly the OH_2 group was split off. We estimate a distance between the carbon and the oxygen of 1.8 \AA that we call a critical dehydration distance. Incidentally, due to fluctuation in the molecule the CO distance stretches beyond this critical distance as shown by Fig. 4.4. As the vibrational energy itself is not large enough to break this bond, it must be a specific hydrogen bonded structure that, at a certain moment, pushes the reaction just over the barrier. This happens only when such an extreme fluctuation beyond 1.8 \AA coincides with a specific solvation structure. To examine this we compared a few cases where this critical situation was reached and looked why in one case the reaction was successful and in the other not. The top panels of Fig. 4.6 show two typical cases that are very similar in structure from the same $Q = 0.053 \text{ \AA}$ (dehydration) simulation, but at different times. Only the important molecules and WFCs are visualized. The left situation at $t = 3.03 \text{ ps}$ shows an unsuccessful and the right one at $t = 8.14 \text{ ps}$ shows a successful dehydration configuration. The $|\text{CO}|$ distances are 1.805 \AA and 1.786 \AA respectively and the $|\text{CH}|$ distances are 1.32 and 1.35 \AA . The two situations are almost identical except that in the second case the protonated ethanol forms three hydrogen bonds, donating two and accepting one, while in the first configuration there are only two donating hydrogen bonds. We checked this for the whole simulation run up to the moment of reaction, and it was found that OH_2 group of the protonated ethanol was constantly bounded by two hydrogen bonds, in which the OH_2 group acted as proton donor.

The times that it also accepted a proton via a hydrogen bond were rare and there was no overlap with the fluctuations larger than 1.8 \AA , until just before the reaction point at $t = 8.14 \text{ ps}$.

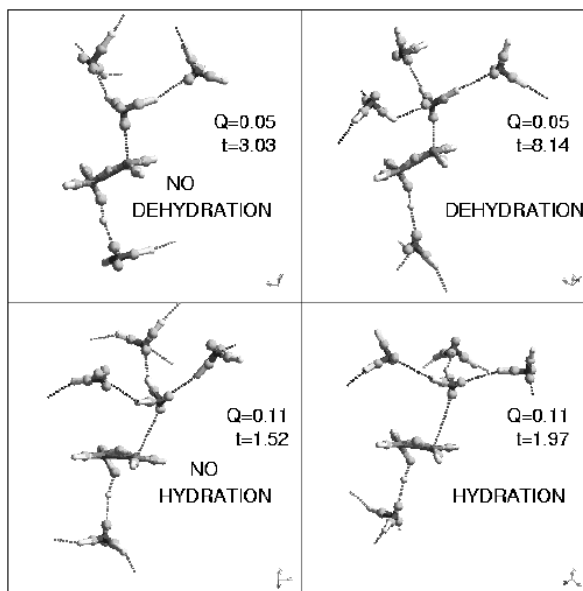


Figure 4.6: Four structures obtained from two simulations. Dark grey indicate the oxygens, white the hydrogen, middle gray the carbons and light grey the WFCs. The dashed lines indicate hydrogen bonds and other weak attracting forces between WFCs and atoms. Top left panel: $Q = 0.0529 \text{ \AA}$ (dehydration) at 3.03 ps, top right panel: $Q = 0.0529 \text{ \AA}$ (dehydration) at 8.14 ps. Bottom left panel: $Q = 0.106 \text{ \AA}$ (hydration) at 1.52 ps, Bottom right panel: $Q = 0.106 \text{ \AA}$ (hydration) at 1.97 ps.

For the hydration reaction case $Q = 0.106 \text{ \AA}$ (hydration) we estimated the critical hydration distance for the incoming water to be 2.6 \AA and postulated that the reaction can only be successful when the water approaches from the other side of the ethylene plane with respect to the hydronium ion. Again we looked closely at similar configurations. At the bottom panels we see two snapshots of the $Q = 0.106 \text{ \AA}$ (hydration) simulation, one at 1.52 ps and the other at 1.97 ps. The $|\text{CO}|$ distances are 2.63 \AA and 2.64 \AA respectively and the $|\text{CH}|$ distances are 1.39 and 1.37 \AA . Although, in both situations the attacking water molecule is three times hydrogen bonded, only the last one was successful. We must conclude that the possibility of a reaction is not simply depending on the number of hydrogen bonds, but depends

in a more subtle way on the precise hydrogen bonded structure. Further examination of the two panels gives a clue what could be the crucial difference between the two configurations. If we relate the axis between the carbon and the middle water molecule with the hydrogen bonds around this water, we can detect a kind of umbrella structure. We think that the right panel is more favorable as it is much more similar to a tetrahedral structure. In this case the three outer water molecules position the middle one in an ideal orientation with its free WFC in a perfect line between its oxygen and the (positively charged) carbon of the ethylene. The other situation deviates significantly from the ideal tetrahedral configuration. To restore a more tetrahedral configuration the central water molecule has to move upwards, removing it from the ethylene molecule and thus leading to a non reactive event. More quantitative, the COO angles in the left panel are from left to right: 112.0° , 89.4° and 149.3° and in the right panel they are: 105.7° , 98.1° and 118.4° . The latter case is indeed much closer to the tetrahedral structure with angles of 109° .

4.8 Conclusions

We performed a DFT-based based Molecular Dynamics simulation of the acid-catalyzed hydration of ethylene in aqueous solution under ambient conditions. To overcome the reaction barrier, we used the constrained dynamics method in combination with thermodynamic integration to obtain the free-energy profile along a chosen reaction coordinate. We found a reaction barrier of 23.3 kcal/mol and a reaction energy of 5.3 kcal/mol. This is much lower than the experimental barrier of 33.3 ± 1.0 kcal/mol [117]. However, the BLYP functional underestimates the reaction barrier to some extent and the approximate reaction coordinate gives probably an additional underestimation of the barrier. A DFT-BLYP calculation of the gas-phase reaction showed a similar barrier of 24.7 kcal/mol [153]. However, the comparison is somewhat misplaced as the mechanism for the reaction is different, where in the gas-phase reaction the hydrating water molecule also provides the proton, whereas in the our solution-calculation the proton providing water and the hydrating water are different molecules. Primarily calculations suggest that the latter mechanism in the gas-phase, involving the $\text{H}_3\text{O}^+ \cdot \text{C}_2\text{H}_4 \cdot \text{H}_2\text{O}$ complex, has no barrier at all. In that sense, one could say that the solvent effects increase the barrier by about 20 kcal/mol. Though, this process is not so likely in the gas-phase as it requires a simultaneous collision of three molecules.

The protonation of the ethylene is directly pointed towards the carbon site without the formation of a π -bonded bridged structure as an intermediate. Moreover, our simulations confirm the assumption that the reaction follows the $\text{Ad}_{\text{E}}3$ mechanism for which simultaneously two bonds are formed. The $|\text{CH}|$ distance is minimally changing after the reaction has been established, but the constraint force has a sudden change from repelling to attracting the proton towards the carbon. We studied this effect by the Wannier analysis of the electronic structure. The WFC participating in the double bond between the carbons has a sudden change and moves away from the ethylene center more in alignment with the CH axis. This indicates a change from sp^2 to sp^3 hybridization and the formation of a CH bond. The Wannier analysis gives in this way clear information on how and when molecular bonds are formed, which is not possible to achieve from the atomic positions only. Furthermore, we looked at the specific hydrogen bonded network that enables a reaction to occur when the reaction coordinate is fixed close to the top of the free-energy barrier. The increasing constraint value transfers the proton towards the carbon and polarizes the ethylene molecule, yielding a positively charged carbonium ion at the other side. This enables the spontaneous formation of a CO bond between that carbon and a water molecule from the solution. For the dehydration reaction of the protonated ethanol we found that when due to spontaneous fluctuations the CO bond is over stretched beyond 1.8 \AA and simultaneous the OH_2 group of the protonated ethanol has more than two hydrogen bonds, the reaction is likely to occur. For the hydration reaction we found that the attacking water molecule should approach at a distance of 2.6 \AA . Mostly, this water will be rejected back into the solution. However, when this water has three hydrogen bonds and the COO angles are close to the tetrahedral value of 109° the hydration can occur and protonated ethanol will be formed. This proves once more the solvation structure participates strongly in the mechanism and is, therefore, part of the true reaction coordinate [143–148]. This study shows that the occurrence of a reactive event depends on the solvation structure in a very subtle way. Our simulations suggest that the reaction is initiated by a number of effects, in which both rare fluctuations of the solute molecule as the occurrence of properly structured water groups play a crucial role.

Chapter 5

Transition Interface Sampling

A Novel Path Sampling Method for the Calculation of Rate Constants ¹

We derive a novel efficient scheme to measure the rate constant of transitions between stable states separated by high free energy barriers in a complex environment within the framework of transition path sampling. The method is based on directly and simultaneously measuring the fluxes through many phase space interfaces and increases the efficiency by at least a factor two with respect to existing transition path sampling rate constant algorithms. The new algorithm is illustrated on the isomerization of a diatomic molecule immersed in a simple fluid.

5.1 Introduction

The calculation of rate constants of activated processes dominated by rare events, chemical reactions being a prime example, is still one of the major computational challenges. As transition rates depend exponentially on the activation barrier height, the expectation time for an event can exceed current computer capabilities by many orders of magnitude. As a result most chemical reactions cannot be simulated by direct molecular dynamics (MD) methods, except those with very low activation energies. The conventional way to tackle this time scale problem is based on transition state the-

¹This chapter is based on: Titus S. van Erp, Daniele Moroni, and Peter G. Bolhuis, "A Novel Pathsampling Method for the Calculation of Rate Constants", accepted for publication in *Journal of Chemical Physics*. Supplemented to this article are the appendixes (5.5.1) and (5.5.3).

ory (TST) and separates the problem in two steps [156–160]. The first step is the calculation of the free energy barrier as function of a reaction coordinate, the second stage is the calculation of the transmission coefficient by sampling fleeting trajectories departing from the top of the barrier. If the reaction coordinate is well chosen, the top of the free energy barrier corresponds to points in phase space close to the true transition state, and the transmission coefficient will have a reasonable value. However, in high dimensional complex systems the choice of reaction coordinate can be extremely difficult and usually requires detailed *a priori* knowledge of the transition mechanism. Consequently, an intuitively chosen but wrong reaction coordinate can result in a very low transmission coefficient, and hence a statistically inaccurate or immeasurable rate constant.

Chandler and collaborators [161–165] devised a method for which no prior knowledge of the system is needed. This method, called transition path sampling (TPS), gathers a collection of trajectories connecting the reactant to the product region by employing a Monte Carlo (MC) algorithm. The resulting path ensemble can be used to elucidate reaction mechanisms, transition states and reaction coordinates. The TPS method has been successfully used on such diverse systems as cluster isomerization, auto-dissociation of water, ion pair dissociation and on isomerization of a dipeptide, as well as reactions in aqueous solution (see Ref. [164] for an overview). Just as in the conventional case mentioned above, an additional second simulation is needed to determine the rate constant within TPS. This simulation combines the path sampling method with the umbrella sampling technique to estimate the probability to reach the product state from the initial reactant state. The final macroscopic rate constant is given by a plateau in the time derivative of a correlation function [163]. In case of two distinct stable states this plateau region should always exist at times longer than the typical molecular relaxation time. However, when reaction pathways are complex and exhibit multiple recrossings, these typical molecular relaxation timescales can be relatively long. In that case the TPS rate constant calculation is computationally expensive, as the path length must exceed these timescales.

In this paper we improve the efficiency of the TPS rate constant calculation on several points by introducing an alternative scheme for calculating reaction rates, named transition interface sampling (TIS). The first of these improvements is allowing the path length to vary, so that by a well chosen definition of the stable states we can limit the length of each path to the strict minimum. Secondly, the new method is based on the effective positive flux through dividing surfaces or interfaces and is consequently much

less sensitive to multiple recrossings or diffusive barrier crossings. Thirdly, the number of different types of Monte Carlo moves is reduced, making the implementation of the algorithm conceptually simpler.

This paper is organized as follows: In section 5.2 we briefly describe the existing algorithms and present the theoretical derivation for the TIS rate constant expression. The implementation of the algorithm is discussed in section 5.3. We illustrate the algorithm on a diatomic molecule in a fluid of repulsive particles and make a quantitative comparison to the original TPS calculation in section 5.4. We end with concluding remarks in section 5.5.

5.2 Theory

5.2.1 Transition State Theory and the Calculation of Rate Constants

Consider a dynamical system in which transitions can take place between two stable states A and B . If the barrier between A and B is sufficiently high, the system will show exponential relaxation for which the forward and backward rate constants k_{AB} and k_{BA} are well defined and can be expressed in terms of microscopic properties. Measuring these rate constants by computer simulation is traditionally done by the two stage Bennett-Chandler (BC) procedure based on the principles of TST [159,160]. The first step is the calculation of the reversible work or free energy to bring the system from stable state A to the transition state. This free energy $F(\lambda)$ has to be calculated as a function of a suitably chosen reaction coordinate λ . This λ can be a complex function of all particle coordinates r and momenta p : $\lambda = \lambda(x)$, with $x \equiv \{r, p\}$. The maximum in $F(\lambda)$ defines the transition state dividing surface λ^* [23,166]. By convention, the system is in A if $\lambda(x) < \lambda^*$ and in B if $\lambda(x) > \lambda^*$.

The main assumption in TST is that any trajectory coming from A and crossing the transition state dividing surface $\lambda(x) = \lambda^*$ will remain at the B side of the dividing surface for a long time. The reaction rate can therefore be expressed as the positive flux through the multidimensional dividing surface λ^* . Two equivalent (see Appendix 5.5.1) expressions for this flux are

$$\begin{aligned} k_{AB}^{TST} &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \frac{\langle \theta(\lambda^* - \lambda(x_0)) \theta(\lambda(x_{\Delta t}) - \lambda^*) \rangle}{\langle \theta(\lambda^* - \lambda) \rangle} \\ &= \frac{\langle \dot{\lambda}(x_0) \delta(\lambda(x_0) - \lambda^*) \theta(\dot{\lambda}(x_0)) \rangle}{\langle \theta(\lambda^* - \lambda) \rangle} \end{aligned}$$

$$= \left\langle \dot{\lambda}(x_0) \theta(\dot{\lambda}(x_0)) \right\rangle_{\lambda^*} \frac{e^{-\beta F(\lambda^*)}}{\int_{-\infty}^{\lambda^*} e^{-\beta F(\lambda)} d\lambda}, \quad (5.1)$$

where x_t specifies the set of coordinates and momenta of the system at time t , the dots denote derivatives with respect to time t , the brackets $\langle \dots \rangle$ denote equilibrium ensemble averages and $\theta(x)$ and $\delta(x)$ are the Heaviside step-function and the Dirac delta function respectively. In the last equality of Eq. (5.1) the connection to the reversible work $F(\lambda)$ is made, and $\beta = 1/k_B T$, where k_B is Boltzmann's constant and T is the temperature. The subscript λ^* to the ensemble brackets, indicates that the ensemble is constrained to the top of the barrier λ^* .

We consider the system to be completely deterministic and thus we can write $x_t = f(x_{t'}, t - t') = f(x_0, t)$, in which f is the time-propagator function. Evaluation of the function $f(x, t)$ requires integrating the equations of motion over the time interval t starting with configuration x . Nevertheless, the equations derived in this paper are still valid when applied to stochastic dynamics.

Even when the TST assumption is accurate, it can be extremely difficult to find a proper reaction coordinate for which recrossings do not occur. As a result a wrong choice for the reaction coordinate will give a much lower free energy barrier than the real activation free energy and will correspondingly overestimate the rate constant. Fig. 5.1 illustrates that Eq. (5.1) overcounts trajectories. One can correct for this overcounting by multiplying the TST rate constant with the transmission coefficient $\kappa(t)$ to obtain the true rate constant

$$k_{AB}(t) = k_{AB}^{TST} \kappa(t). \quad (5.2)$$

The calculation of the time dependent transmission coefficient $\kappa(t)$ constitutes the second part of the two stage BC procedure [159, 160]. $\kappa(t)$ belongs to the approximate dividing surface λ^* [156, 158–160] and can be determined by taking an ensemble average of many short trajectories starting from the dividing surface:

$$\kappa(t) = \frac{1}{\left\langle \dot{\lambda}(x_0) \theta(\dot{\lambda}(x_0)) \right\rangle_{\lambda^*}} \left\langle \dot{\lambda}(x_0) \theta(\lambda(x_t) - \lambda^*) \right\rangle_{\lambda^*}. \quad (5.3)$$

After a short molecular time t_{mol} the trajectories are committed to a stable state and $\kappa(t)$, and hence $k_{AB}(t)$, become constant: the transmission coefficient κ , and the rate constant k_{AB} , respectively. It is however important to start sufficiently close to the true transition state dividing surface. Otherwise the transmission coefficient will be extremely low, making an accurate

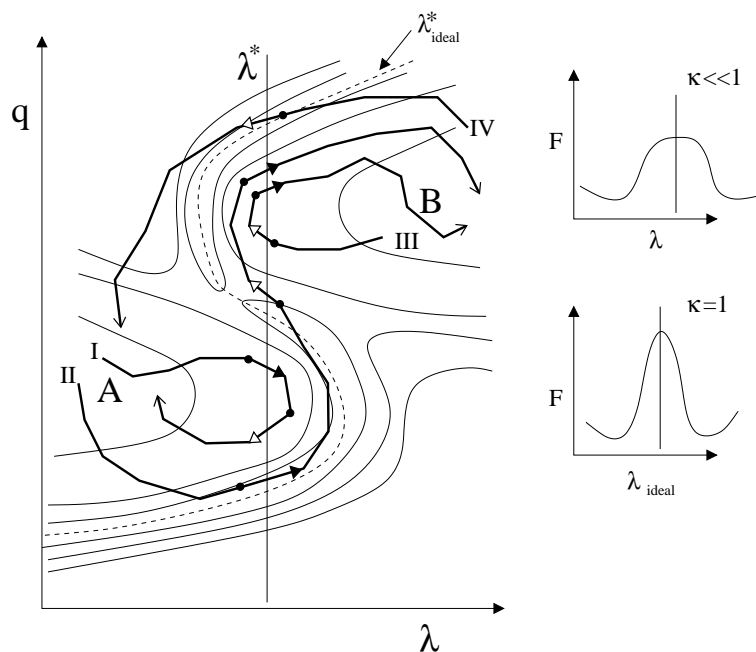


Figure 5.1: The thin solid curve show the two dimensional free energy landscape in contour plot. λ is the chosen reaction coordinate, q represents all other degrees of freedom. A and B denote the state regions. The vertical line at λ^* corresponds to the maximum in the free energy function $F(\lambda)$ as is shown at the right upper side. The free energy as function of the ideal reaction coordinate is also shown at the right lower side. This reaction coordinate is a complex function of all degrees of freedom $\lambda_{\text{ideal}} = \lambda_{\text{ideal}}(\lambda, q)$ and the corresponding free energy function has its maximum at the true transition state dividing surface $\lambda_{\text{ideal}} = \lambda_{\text{ideal}}^*$. This true dividing surface is the dashed curved line. The corresponding free energy barrier is much more narrow and higher than the artificial barrier due to the incorrect reaction coordinate. Four possible trajectories are shown. The black solid arrows indicate a positive flux through the surface λ^* and the white solid arrows indicate the negative fluxes. TST rate expression (5.1) counts all positive fluxes of trajectories I, II and III. Consequently, non-true reactive events like I and III have a artificial contribution to the rate constant and also trajectory II is overcounted one time. To correct for this, one can calculate the transmission coefficient κ . In the TPS equation (5.4), if $\lambda_A = \lambda_B = \lambda^*$, trajectories III and IV are not counted because of the $h_A(x_0)$ term. Trajectories I and II are correctly counted in the final summation due to the cancellation of positive and negative flux terms.

estimate of the rate constant problematic or even impossible. In many cases, in particular for complex condensed matter systems, a sufficiently close re-

action coordinate is difficult to find and requires considerable a-priori knowledge about the system.

5.2.2 Transition Path Sampling

Transition path sampling (TPS) is developed to overcome the difficulties mentioned above [161–165]. Its main advantage is that no prior knowledge of the transition state is needed. The rate constant in TPS is expressed as the time derivative of a general time correlation function.

$$k_{AB}^{TPS}(t) = \frac{d}{dt}C(t), \quad C(t) = \frac{\langle h_A(x_0)h_B(x_t) \rangle}{\langle h_A(x_0) \rangle}, \quad (5.4)$$

in which $h_A(x)$ and $h_B(x)$ are the characteristic functions defined by:

$$\begin{aligned} h_A(x) &= 1, & \text{if } x \in A, & \quad \text{else } h_A(x) = 0 \\ h_B(x) &= 1, & \text{if } x \in B, & \quad \text{else } h_B(x) = 0. \end{aligned} \quad (5.5)$$

In case of a single order parameter $\lambda(x_t)$ describing the transition, the phase space regions A and B are defined by λ_A and λ_B : $x_t \in A$ if $\lambda(x_t) < \lambda_A$ and $x_t \in B$ if $\lambda(x_t) > \lambda_B$. Knowledge of the precise location of the dividing surface λ^* , $\lambda_A < \lambda^* < \lambda_B$, is not required in TPS. Therefore, the order parameter λ does usually not correspond to the reaction coordinate.

The microscopic expression for the rate constant in Eq. (5.4) is time dependent, while the phenomenological rate constant is not. However, just as the transmission coefficient $\kappa(t)$ becomes a constant, the time dependent function $k_{AB}^{TPS}(t)$ reaches a plateau after a molecular timescale t_{mol} . The phenomenological rate constant is equal to the plateau value: $k_{AB} = k_{AB}^{TPS}(T)$. This plateau region should always exist for times T between the molecular timescale and the characteristic reaction time: $t_{\text{mol}} < T \ll t_{\text{rxn}}$. In other words, T is larger than the timescale to commit to one of the stable states, but much shorter than the expectation time t_{rxn} of a completely new reactive event. If we take $\lambda_A = \lambda_B = \lambda^*$ and the limit $t \rightarrow 0+$, Eq. (5.4) transforms into the expression for the positive reactive flux or, equivalently, the TST equation (5.1). For $t > 0$, however, the reactive flux measured by Eq. (5.4) no longer consists of purely positive contributions. The final rate constant is a sum of positive and negative fluxes, and thus the overcounting of trajectories in Eq. (5.1) is circumvented. (See Fig.(5.1)).

We can rewrite the time dependent rate constant of Eq. (5.4) into [163]:

$$k_{AB}^{TPS}(t) = \frac{\langle \dot{h}_B(t) \rangle_{A, H_B(T)}}{\langle h_B(t') \rangle_{A, H_B(T)}} \cdot C(t'), \quad (5.6)$$

where $H_B(T) = \max_{0 < t < T} h_B(x_t)$ and $\langle \dots \rangle_{A, H_B(T)}$ denotes an average on the ensemble of paths of fixed length T starting in A and entering B at least once [163]. These ensemble averages are evaluated using a Monte Carlo procedure employing the *shooting* and *shifting* moves [162]. The two factors in Eq. (5.6) have to be evaluated separately. First, a path sampling simulation is performed to compute $\langle h_B(t) \rangle_{A, H_B(T)}$ in the interval $[0, T]$. The path length T must be long enough for the time derivative to display a plateau. Subsequently, one chooses a t' in interval $[0, T]$ and computes $C(t')$ using the path sampling in combination with an umbrella sampling technique [163]. A drawback of the TPS rate constant calculation is that the function $k_{AB}^{TPS}(t)$ can be strongly oscillatory because of recrossings and will reach a plateau only after a relatively long time. The path length in TPS must exceed the typical timescale of these oscillations, and consequently, in that case TPS is computationally costly.

5.2.3 Transition Interface Sampling

Just as the BC and the TPS rate constant algorithms, the TIS method is based on a flux calculation. In contrast to these schemes, however, TIS measures the effective positive flux², instead of a conditional general flux as in Eq. (5.4) or Eq. (5.1). This implies that only positive terms contribute to the rate, allowing for faster numerical convergence. A flux is normally defined through a hypersurface in phase space defined by an order parameter, the reaction coordinate. But, similar to the TPS case, we do not want to suffer from a bad choice of reaction coordinate. Therefore, instead of using a single dividing surface, we introduce a series of interfaces through which we measure this flux. We then derive an expression that relates the flux through a certain interface to the flux through an interface which is closer to A to replace the expensive TPS umbrella sampling procedure.

In order to formulate a proper flux, we have to divide the entire phase space into two complementary regions called *overall* states \mathcal{A} and \mathcal{B} . These states do not only depend on the position at the time of consideration but also on its past behavior. Overall state \mathcal{A} covers all phase space points lying inside stable region A , which constitutes the largest part, but also all phase space points that visit A , before reaching B when the equations of motion are integrated backward in time. Similarly, state \mathcal{B} comprises stable state B and all phase points, coming directly from this state in the past, i.e. without having been in A . It is useful to generalize the characteristic functions in

²Here, effective means that the recrossings through the interfaces are not being counted

Eq.(5.5) for an arbitrary phase space region Ω

$$h_{\Omega}(x) = 1, \quad \text{if } x \in \Omega, \quad \text{else } h_{\Omega}(x) = 0. \quad (5.7)$$

For each phase point x and each phase space region Ω we can determine the minimum (first entrance) times $t_{\Omega}^b(x)$ and $t_{\Omega}^f(x)$ needed to reach Ω starting from configuration x by integrating the equations of motion backward and forward in time, respectively:

$$\begin{aligned} t_{\Omega}^b(x) &\equiv -\max [\{t | h_{\Omega}(f(x, t)) = 1 \wedge t \leq 0\}] \\ t_{\Omega}^f(x) &\equiv +\min [\{t | h_{\Omega}(f(x, t)) = 1 \wedge t \geq 0\}], \end{aligned} \quad (5.8)$$

where the min and max function return respectively the lowest and highest value of their arguments. In addition, it is useful to define for each phase point x and each set of two non-overlapping phase space regions $\{\Omega_1, \Omega_2\}$ the following characteristic functions:

$$\begin{aligned} \bar{h}_{\Omega_1, \Omega_2}^b(x) &= 1 \quad \text{if } h_{\Omega_1}(f(x, -t_{\Omega_1 \cup \Omega_2}^b(x))) = 1, \\ 0 &\quad \text{otherwise} \end{aligned} \quad (5.9)$$

$$\bar{h}_{\Omega_1, \Omega_2}^f(x) = 1 \quad \text{if } h_{\Omega_1}(f(x, +t_{\Omega_1 \cup \Omega_2}^f(x))) = 1, \quad (5.10)$$

$$0 \quad \text{otherwise} \quad (5.11)$$

In words, these functions measure whether a trajectory reaches Ω_1 before Ω_2 or not. As the system is ergodic, each phase space region will be visited in finite time and thus $\bar{h}_{\Omega_1, \Omega_2}^b(x) + \bar{h}_{\Omega_2, \Omega_1}^b(x) = \bar{h}_{\Omega_1, \Omega_2}^f(x) + \bar{h}_{\Omega_2, \Omega_1}^f(x) = 1$ for any x . Using these definitions the characteristic functions for the overall states \mathcal{A} and \mathcal{B} are given by

$$h_{\mathcal{A}}(x) = \bar{h}_{A, B}^b(x), \quad h_{\mathcal{B}}(x) = \bar{h}_{B, A}^b(x). \quad (5.12)$$

These states together span the complete phase space, as the system can never stay in the intermediate region between A and B forever. The overall states \mathcal{A} and \mathcal{B} do not sensitively depend on the definition of stable state A and B as long as it is reasonably. Of course, the stable regions should not overlap, each trajectory between the stable states must be a true rare event for the reaction we are interested in. In addition, the probability that after this event the reverse reaction occurs shortly thereafter must be as unlikely as an entirely new event. In other words, the system must be committed to the stable states. Therefore, a reasonable definition of A and B requires

that they should lie completely inside the basin of attraction of the respective two states³ (see also Ref. [165]). Special care has to be taken with this condition for processes which show many recrossings between state A and B before settling down. Such processes can occur in solution or in dilute gasses. For instance, for organic reactions in aqueous solution, a rare specific hydrogen bonded network can lower the bond-breaking barrier and initiate the reaction. If the lifetime of those rare solvation structures is high, a sudden reverse reaction can occur as the barrier for the backward reaction is also lowered by the same amount [167–169]. A similar phenomenon can happen in dilute gasses for which rare spontaneous fluctuations in the kinetic energy are the main driving force. A particle moving from one state to another due to a very high kinetic energy as result of sequence of collisions can cross the potential energy barrier several times before it will dissipate its energy by a new collision and relax into one of the stable states (see e.g. Refs. [170, 171]). These problems can in principle be solved by an adequate choice of the stable state definitions. For instance, the definition can depend explicitly on the presence of certain hydrogen bonds or on kinetic energy terms.

With our definition of overall states \mathcal{A} and \mathcal{B} we can write down our rate equation in the spirit of Eq. (5.4):

$$k_{AB} = \frac{\langle h_{\mathcal{A}}(x_0) \dot{h}_{\mathcal{B}}(x_0) \rangle}{\langle h_{\mathcal{A}}(x_0) \rangle}, \quad (5.13)$$

where the dot denotes the time derivative taken at $t = 0$. This rate expression does not depend on time although the evaluation of the characteristic functions still requires integration of the equations of motion. The transition from \mathcal{A} into \mathcal{B} takes place when the system coming from A will cross the interface λ_B for the first time (see Fig. 5.2). After this event the system will stay in \mathcal{B} . Eq. (5.13) counts therefore only the first crossing through interface λ_B and is hence equivalent to the effective positive flux expression

$$\begin{aligned} k_{AB} &= \frac{\langle h_{\mathcal{A}}(x_0) \dot{\lambda}(x_0) \theta(\dot{\lambda}(x_0)) \delta(\lambda(x_0) - \lambda_B) \rangle}{\langle h_{\mathcal{A}}(x_0) \rangle} \\ &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \frac{\langle h_{\mathcal{A}}(x_0) \theta(\lambda_B - \lambda(x_0)) \theta(\lambda(x_{\Delta t}) - \lambda_B) \rangle}{\langle h_{\mathcal{A}}(x_0) \rangle}. \end{aligned} \quad (5.14)$$

³In general, A and B are defined in phase space, but for most practical cases configuration space might be enough.

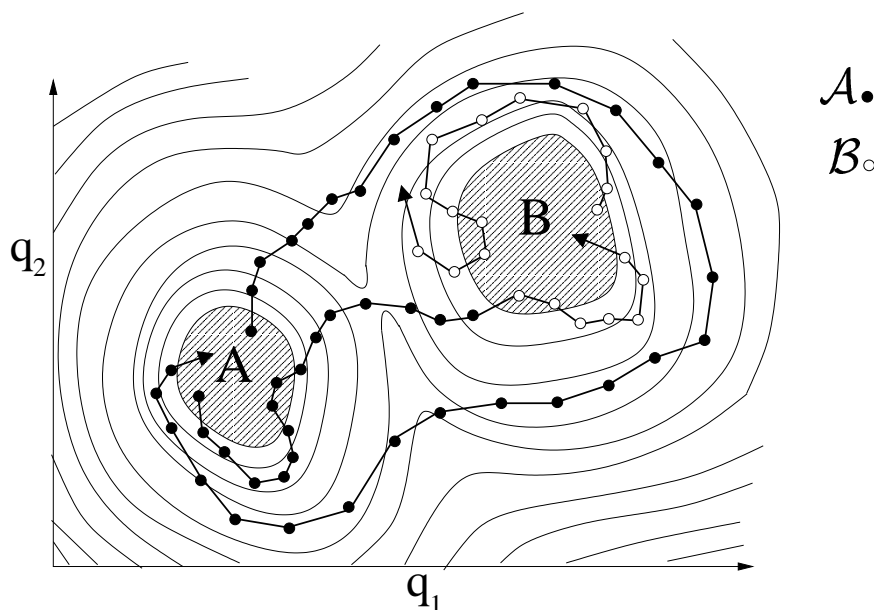


Figure 5.2: Example of phase space regions in TIS. Thin solid curves denote the free energy contour lines. q_1 and q_2 are two arbitrary projections of the degrees of freedom. A and B are the two stable states. The dots on the three shown trajectories indicate the positions of the system at successive time steps. The overall state \mathcal{A} and \mathcal{B} are indicated by black and white dots respectively. Only one trajectory starts in A and ends in B and is therefore a true reactive event. The system changes from state A into B when it enters region B for the first time. It can leave stable region B shortly thereafter, but never go back to A in a short time. The stable regions have to be chosen to fulfill that condition.

Note the similarity with Eq. (5.1). Strictly speaking $\theta(\lambda_B - \lambda(x_0))$ is redundant in Eq. (5.14) as $h_A(x_0) = 0$ if $\theta(\lambda_B - \lambda(x_0)) \neq 1$. The last expression in Eq. (5.14) is most suitable for a numerical approach with Δt as the time step in a molecular dynamics simulation. Evaluation of Eq. (5.14) requires counting all phase space points which at $t = 0$ are just about to cross interface λ_B in one time step *and* will enter region A before B when integrating backward in time starting from x_0 . Unfortunately, Eq. (5.14) is not very efficient from a computational point of view because only a very small fraction of phase points close to interface λ_B actually belong to \mathcal{A} , leading to poor statistics. We can enhance the statistical accuracy by relating the

flux through λ_B to the flux through an interface closer to A . We therefore introduce a set of n non-intersecting interfaces $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n$, each interface λ_i closer to A than the next interface λ_{i+1} (see Fig. 5.3). We define the corresponding phase space regions $\Omega_{\lambda_i} \equiv \{x | \lambda(x) > \lambda_i\}$. In this way Ω_{λ_B} is equivalent to our stable state B , while Ω_{λ_A} is the phase space outside stable state A . By introducing the following definition

$$\Phi_{A, \lambda_i}(x_0) \equiv \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \bar{h}_{A, \Omega_{\lambda_i}}^b(x_0) \theta(\lambda_i - \lambda(x_0)) \theta(\lambda(x_{\Delta t}) - \lambda_i), \quad (5.15)$$

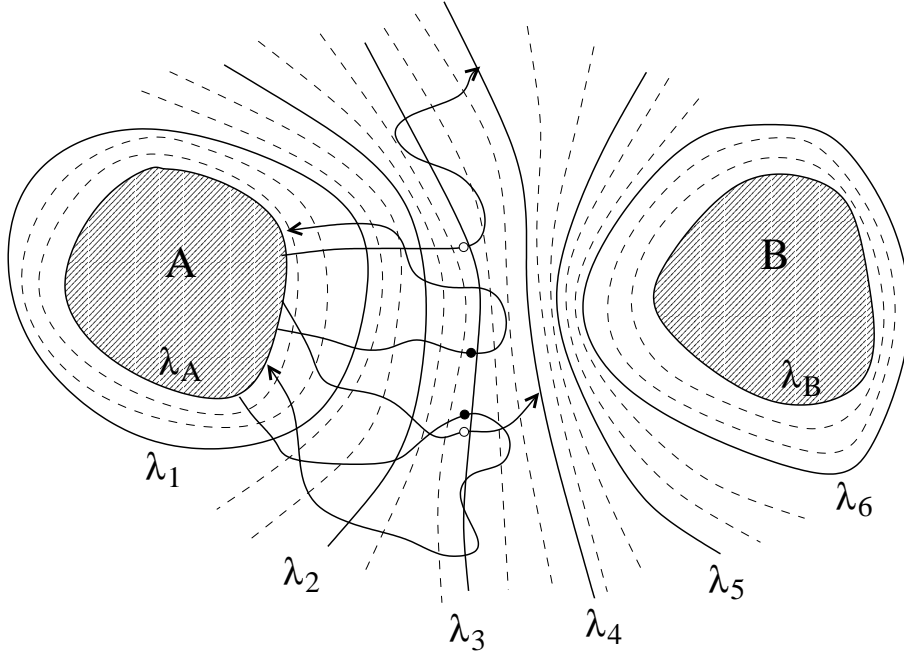


Figure 5.3: Example of the division of the phase space by interfaces. A and B are the stable state regions with interfaces λ_A and λ_B . The interfaces $\lambda_1 \dots \lambda_6$ correspond to a calculation of Eq. (5.18) with $n = 6$. The dashed lines are the sub interfaces in between. Four trajectories are shown corresponding to a $\mathcal{P}(\lambda_4 | \lambda_3)$ ensemble calculation. On each trajectory the x_0 time slice is indicated with a circle. Black circles correspond to $\bar{h}_{A, \Omega_{\lambda_4}}^f(x_0) = 0$ and white circles correspond to $\bar{h}_{A, \Omega_{\lambda_4}}^f(x_0) = 1$.

Eq. (5.14) reduces to

$$k_{AB} = \langle \Phi_{A, \lambda_B} \rangle / \langle h_A \rangle. \quad (5.16)$$

where $\langle \Phi_{A,\lambda_i} \rangle$ denotes the effective positive flux through interface λ_i . The rate constant is thus equal to the effective positive flux through interface λ_B with the condition the trajectories came directly from A . Note again that $\langle \Phi_{A,\lambda_A} \rangle / \langle h_A \rangle$ is equal to the TST rate expression in Eq. (5.1) in case $\lambda^* = \lambda_A = \lambda_B$. The effective flux $\langle \Phi_{A,\lambda_i} \rangle$ can now be related to the effective flux $\langle \Phi_{A,\lambda_{i-1}} \rangle$ through an interface λ_{i-1} closer to A by (see Appendix 5.5.2)

$$\langle \Phi_{A,\lambda_i}(x_0) \rangle = \left\langle \bar{h}_{\Omega_{\lambda_i},A}^f(x_0) \right\rangle_{\Phi_{A,\lambda_{i-1}}} \times \langle \Phi_{A,\lambda_{i-1}}(x_0) \rangle, \quad (5.17)$$

where $\langle \dots \rangle_{\Phi_{A,\lambda_{i-1}}}$ denotes the ensemble average over all phase space points x_0 for which $\Phi_{A,\lambda_{i-1}}(x_0) \neq 0$. The factor $\langle \bar{h}_{\Omega_{\lambda_i},A}^f(x_0) \rangle_{\Phi_{A,\lambda_{i-1}}} \equiv \mathcal{P}(\lambda_i | \lambda_{i-1})$ is the conditional probability that a trajectory, coming from A , passes λ_i , given the fact that it has passed the interface λ_{i-1} at an earlier time. By recursively substituting Eq. (5.17) into Eq. (5.16) the rate constant can be expressed as a product of conditional probabilities:

$$\begin{aligned} k_{AB} &= \frac{\langle \Phi_{A,\lambda_1} \rangle}{\langle h_A \rangle} \prod_{i=1}^{n-1} \left\langle \bar{h}_{\Omega_{\lambda_{i+1}},A}^f \right\rangle_{\Phi_{A,\lambda_i}} \left\langle \bar{h}_{B,A}^f \right\rangle_{\Phi_{A,\lambda_n}} \\ &\equiv \frac{\langle \Phi_{A,\lambda_1} \rangle}{\langle h_A \rangle} \prod_{i=1}^{n-1} \mathcal{P}(\lambda_{i+1} | \lambda_i) \mathcal{P}(\lambda_B | \lambda_n) \\ &= \frac{\langle \Phi_{A,\lambda_1} \rangle}{\langle h_A \rangle} \left\langle \bar{h}_{B,A}^f \right\rangle_{\Phi_{A,\lambda_1}} \equiv \frac{\langle \Phi_{A,\lambda_1} \rangle}{\langle h_A \rangle} \mathcal{P}(\lambda_B | \lambda_1). \end{aligned} \quad (5.18)$$

This expression is the central equation for TIS (In Appendix 5.5.3 we show how Eq. 5.17 can also be used to derive an alternative expression for the transmission coefficient allowing a more efficient computation than Eq. 5.3). Instead of just calculating the individual terms in the product of Eq. (5.18) we can equivalently determine a continuous crossing probability function $\mathcal{P}(\lambda | \lambda_1)$ for λ between λ_1 and λ_B . This is reminiscent of umbrella sampling where a free energy difference is usually estimated as a function of a continuous parameter λ [23]. When calculating the ensemble average for $\mathcal{P}(\lambda_i | \lambda_{i-1})$ we can also evaluate $\mathcal{P}(\lambda | \lambda_{i-1})$ for interfaces λ between λ_{i-1} and λ_i by dividing the phase space into a finer grid of sub interfaces (see Fig. (5.3)). In this way we acquire useful information without significant extra cost, and, in addition, a measure for the convergence of the ensemble averages. The final monotonically decreasing crossing probability function $\mathcal{P}(\lambda | \lambda_1)$ can be obtained by matching the histograms from the different ensemble simulations. Techniques commonly applied in umbrella sampling

such as overlapping windows between two successive ensemble averages and the use of biasing functions can also be employed here.

5.3 The Transition Interface Sampling Algorithm

Inspection of Eq. (5.18) clearly shows that the TIS rate constant calculation is also a two step procedure. The first step, the effective flux $\langle \Phi_{A,\lambda_1} \rangle / \langle h_A \rangle$ can be computed by simply running a MD simulation starting with a configuration in A and counting the number of effective crossings. For an interface λ_1 close enough to stable state A one can obtain a statistically accurate value.

The second part of the calculation consists of evaluating the product of the $\mathcal{P}(\lambda_{i+1}|\lambda_i)$ ensemble averages for the different interfaces λ_i in Eq. (5.18). Here we need to sample all paths from region A to either A or $\Omega_{\lambda_{i+1}}$ that exhibit at least one crossing with interface λ_i . The Monte Carlo moves in TIS are very similar to the shooting move used in the TPS algorithm. The main difference is that the backward and forward integration is abandoned as soon as the edge of either A or $\Omega_{\lambda_{i+1}}$ is reached. If the new path is accepted there is only one phase point x along this path for which $\Phi_{A,\lambda_i}(x) \neq 0$, defining phase space point x_0 . The shifting moves that were required in the original TPS implementation to enable proper sampling and improve statistical accuracy are here unnecessary.

To bootstrap the sampling procedure we first generate an initial path that starts in A , then crosses the interface λ_i and finally ends in either A or $\Omega_{\lambda_{i+1}}$ (see for more details on initial path generation Ref. [165]). The phase space point x_0 is then defined as the first crossing point of this path with interface λ_i . Further, let $\tau = \text{int}(t/\Delta t)$ be the discrete time slice index, and $\tau^b \equiv \text{int}(t_A^b(x_0)/\Delta t)$ and $\tau^f \equiv \text{int}(t_{A \cup \Omega_{\lambda_{i+1}}}^f(x_0)/\Delta t)$ the forward and backward terminal time slice indices, respectively. Including x_0 , the initial path then consists of $N^{(o)} = \tau^b + \tau^f + 1$ time slices. With these definitions in mind is the TIS algorithm as follows:

1. From the current path with length $N^{(o)}$ choose a random time slice τ , with $-\tau^b \leq \tau \leq \tau^f$.
2. Change all momenta of $x_{\tau\Delta t}$ by adding small random displacements δp from a Gaussian distribution. Make sure the total momentum is conserved [165]
3. In case of a constant energy (NVE) simulation, rescale the new momenta to the old energy value and continue with step 4. In case of

constant temperature (NVT) accept the new momenta (else reject the whole TIS move) with a probability [23]:

$$\min \left[1, \exp \left(\beta (E(x_{\tau\Delta t}^{(o)}) - E(x_{\tau\Delta t}^{(n)})) \right) \right].$$

Here, $E(x)$ is the total energy of the system at phase space point x .

4. Integrate equations of motion backward in time by reversing the momenta at time slice τ , until reaching either A or $\Omega_{\lambda_{i+1}}$. Reject in case of $\Omega_{\lambda_{i+1}}$ else continue with the next step.
5. Integrate from time slice τ forward until reaching either A or $\Omega_{\lambda_{i+1}}$. Reject if the entire trial path does not cross interface λ_i , else continue with the next step.
6. Accept the trial path with a probability

$$\min \left[1, \frac{N^{(o)}}{N^{(n)}} \right],$$

where $N^{(n)}$ is the length of the new path. If accepted, replace the old path with the new one.

7. Reassign x_0 to be the first crossing point with λ_i and sample the value of $\bar{h}_{\Omega_{\lambda_{i+1}}, A}^f(x_0)$ to measure $\mathcal{P}(\lambda_{i+1} | \lambda_i)$.
8. Repeat from step 1.

As usual in Monte Carlo schemes, any rejection along this route implies counting the old path again in the ensemble average. The acceptance probabilities at step 3 and step 6 are required to satisfy the detailed balance condition (see e.g. Ref [23]).

Instead of generating a complete path and then accepting or rejecting accordingly to the probability at step 6, it is more efficient to determine a maximum path length in advance. Before embarking on the time consuming fourth and fifth step, we first take a uniform random number α between 0 and 1 and determine the maximum allowed path length by:

$$N_{\max}^{(n)} = \text{int}(N^{(o)}/\alpha). \quad (5.19)$$

In this way we can directly stop the integration and reject the TIS move as soon the path length $N^{(n)}$ exceeds the maximum $N_{\max}^{(n)}$. In the course of the

TIS simulation the path-length fluctuates. This also means that the average path length becomes automatically shorter when changing from ensemble average $\mathcal{P}(\lambda_{i+1}|\lambda_i)$ to ensemble average $\mathcal{P}(\lambda_i|\lambda_{i-1})$ closer to A .

The algorithm presented here does not require shifting moves because there is only one unique x_0 phase point along each pathway. However, one could consider the use of path-reversal moves as they have negligible computational cost and can sometimes facilitate ergodic sampling [165].

5.4 Numerical Results

5.4.1 The Model

We tested the TIS algorithm on a simple diatomic bistable molecule immersed in a fluid of purely repulsive particles. Such a system has previously been used in illustrating TPS rate constant calculations [163] and is therefore a good starting point for a comparison between the two methods. The system consists of N particles in 2 dimensions with interactions given by a pairwise Lennard-Jones (LJ) potential truncated and shifted at the minimum, often referred to as the Weeks-Chandler-Andersen (WCA) potential [172]

$$V_{WCA}(r) = 4\epsilon[(r/\sigma)^{-12} - (r/\sigma)^{-6}] + \epsilon \quad \text{if } r \leq r_0 \quad (5.20)$$

$$0 \quad \text{if } r > r_0, \quad (5.21)$$

where r is the interatomic distance, and $r_0 \equiv 2^{1/6}\sigma$. Throughout this section reduced units are used so that ϵ and σ , respectively the LJ energy and length parameters, as well as the mass of the particles are equal to unity. The LJ unit of time $(m\sigma^2/\epsilon)^{1/2}$ is therefore also unity. In addition, two of the N particles are interacting through a double well potential

$$V_{dw}(r) = h \left[1 - \frac{(r - r_0 - w)^2}{w^2} \right]^2. \quad (5.22)$$

This function has two minima separated by a barrier of height h corresponding to the two stable states of the molecule: a compact state for $r = r_0$ and extended state for $r = r_0 + 2w$. For a high enough barrier, transitions between the states become rare and the rate constant is well defined. Hence, this system provides a useful test case for the TPS and TIS algorithms.

The system is simulated at a constant energy E in a simulation square box with periodic boundary conditions. The total linear momentum is conserved and is set zero for all trajectories. The equations of motion are integrated using the velocity Verlet algorithm with a time step $\Delta t = 0.002$. As

in Ref. [163] we focus here on the computation of the rate constant for the isomerization reaction of the dimer from the compact state to the extended state. In the following section we describe general simulation details. In section 5.4.3 we discuss the results for a system with a high enough barrier to avoid recrossings. Subsequently, we reproduce the simulations from Ref. [163] in section 5.4.4. These results do show recrossings, and we discuss the consequences for TPS and TIS.

5.4.2 Methodology

The TPS rate constant calculation evaluates the two factors in Eq. (5.6) separately as explained in section 5.2.2. The first term in Eq. (5.6) is the ratio between the plateau value of the reactive flux correlation function $\langle \dot{h}_B(T) \rangle_{A, H_B(T)}$ and the correction $\langle h_B(t') \rangle_{A, H_B(T)}$. The second term $C(t')$ requires an umbrella sampling simulation in the form of a series of window calculations. An order parameter is chosen to define the characteristic functions of the stable states and to partition phase space in windows for the umbrella sampling. Besides shooting and shifting Monte Carlo moves to generate new paths in the transition path sampling we also employ a *diffusion* move that shifts the path by one time slice in arbitrary direction. This move is computationally very cheap but increases the statistics of the correlation functions. In all our simulations we therefore set the percentages for shooting, shifting and diffusion to 5%, 10% and 85%, respectively. The parameters involved are always gaged such that the acceptance ratio is around 40% for shooting and shifting moves, ensuring an optimum efficiency of the sampling [163].

The TIS method involves a direct determination of the flux and the calculation of the crossing probability functions $\mathcal{P}(\lambda_i | \lambda_{i-1})$ between a series of successive interfaces as given by Eq. (5.18). The flux term in Eq. (5.18) is computed by means of a straightforward MD simulation starting in state A and counting the number of effective positive crossings through interface λ_1 , i.e. when the trajectory is directly coming from A . The second term in Eq. (5.18) is computed using the TIS algorithm of section 5.3. The basic requirement is a definition of a set of interfaces partitioning the phase space. Between these interfaces we defined a finer grid of sub-interfaces to construct the crossing probability function $\mathcal{P}(\lambda | \lambda_1)$. As in the TPS calculation we adjusted the momentum displacement for the shooting move to give an acceptance of about 40%.

Many parameters are involved in the two methods and to compare the relative efficiency we measured the CPU-time required for an arbitrary fixed error of 2.5% for each step in both the TPS and TIS calculations under the

same computational conditions (1Ghz AMD Athlon). In both methods the final rate constant consists of a product of factors which have to be calculated independently. For each factor we performed M simulation blocks of N Monte Carlo cycles and adjusted N such that after M block averages the relative standard deviation of each term in Eq. (5.4) and (5.13) was 2.5%. The total CPU-time is given by summing the individual 2.5% error CPU-times for each factor. The final error in the rate constants is obtained by the standard propagation rules using all simulation results (i.e. not only the ones for the 2.5% error CPU time calculation).

5.4.3 System with High Energy Barrier

This system had a total number of particles $N = 25$, and a total energy $E = 25$. The square simulation box was adjusted to give a number density of 0.7. The barrier height was $h = 15$ and the width-parameter $w = 0.5$, so that the minima of $V_{dw}(r)$ were located at $r \simeq 1.12$ and $r \simeq 2.12$ while the top of the barrier was at $r \simeq 1.62$. In the TPS rate calculation we defined stable states A and B as $r < r_A = 1.5$ and $r > r_B = 1.74$, respectively. We computed the correlation function $\langle h_B(t) \rangle_{A, H_B(T)}$ using TPS with a fixed path length $T = 2.0$. The correlation function is shown in Fig. 5.4 together with its time derivative, the reactive flux. The latter function clearly displays a plateau. Next, we chose four different $t' = 0.1, 0.3, 1.0, 2.0$ and performed umbrella sampling simulations using 8 windows to calculate $C(t')$. In each window we measured the probability to find the path's end point $r(t')$ at a certain value of r . These probability histograms were rematched and normalized. The final probability functions are shown in Fig. 5.5. Integration of the area under the histogram belonging to region B leads to $C(t')$ and finally to the rate constant [163]. In Tab. 5.1 we give the values of the different contributions to the rate constant given by Eq. 5.6, together with the rate constant. We report the average relative computation time needed to reach the 2.5% error (see section 5.4.2) in Tab. 5.2.

For the TIS calculations we use the same order parameter r and the same definition for region B , i.e. interface λ_B is set at $r = 1.74$. Stable state A was defined by setting $\lambda_A = \lambda_1$ at $r = 1.24$. This interface is closer to the basin of attraction than the TPS stable state definition but yields a higher flux term $\langle \Phi_{A, \lambda_1} \rangle / \langle h_A \rangle$ and gives better statistics. Note that the different definition of stable state A does not change the final rate constant, as the overall state \mathcal{A} does not sensitively depends on this definition. The flux term is calculated by straightforward NVE MD. As λ_A is equal to λ_1 every positive crossing of this interface is counted in the flux because all trajectories must by default

come directly from A. The conditional crossing probabilities $\mathcal{P}(\lambda_{i+1}|\lambda_i)$ in Eq. (5.18) are calculated for $n = 5$ interfaces between the stable states (see Fig. 5.6). Between these interfaces we impose a finer grid to obtain the entire crossing probability function. The results for each stage and the final rate constant are shown in Tab. 5.1. The rate constants of both methods agree within the statistical accuracy, showing that the TIS method is correct.

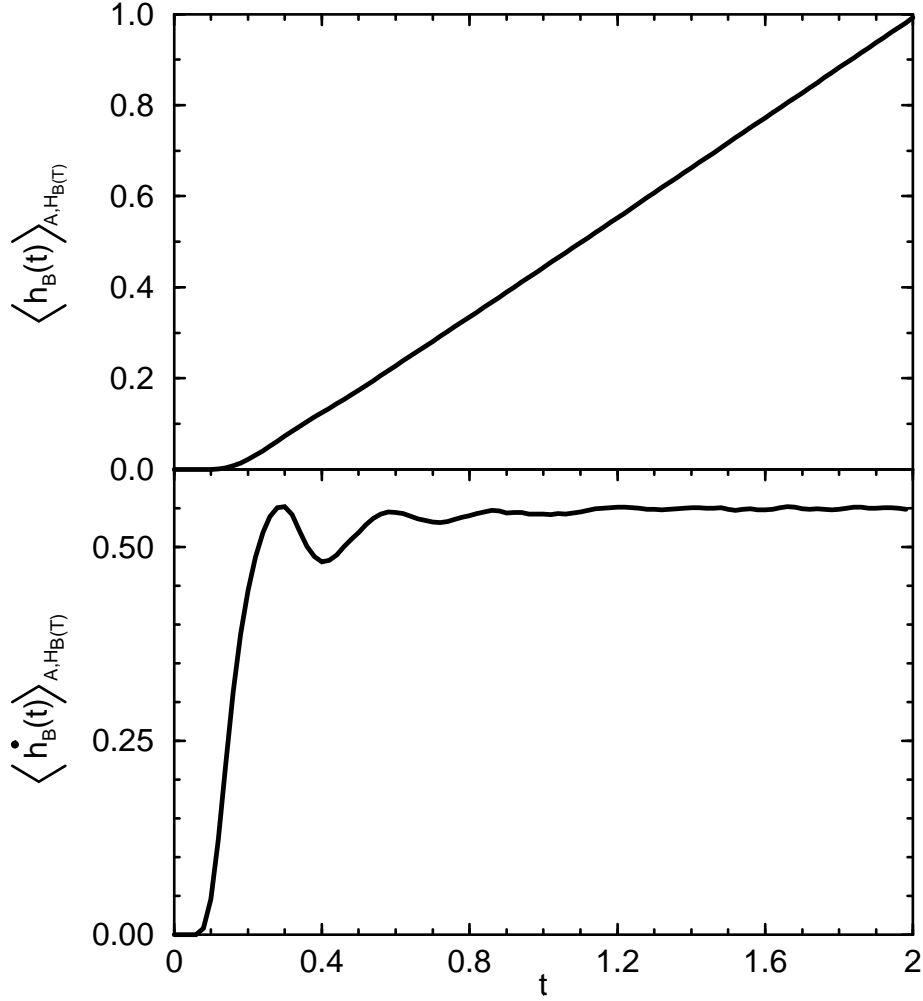


Figure 5.4: TPS correlation function $\langle h_B(t) \rangle_{A, H_B(T)}$ (top) and its time derivative (bottom) for the system with high energy barrier. The error is comparable to line thickness.

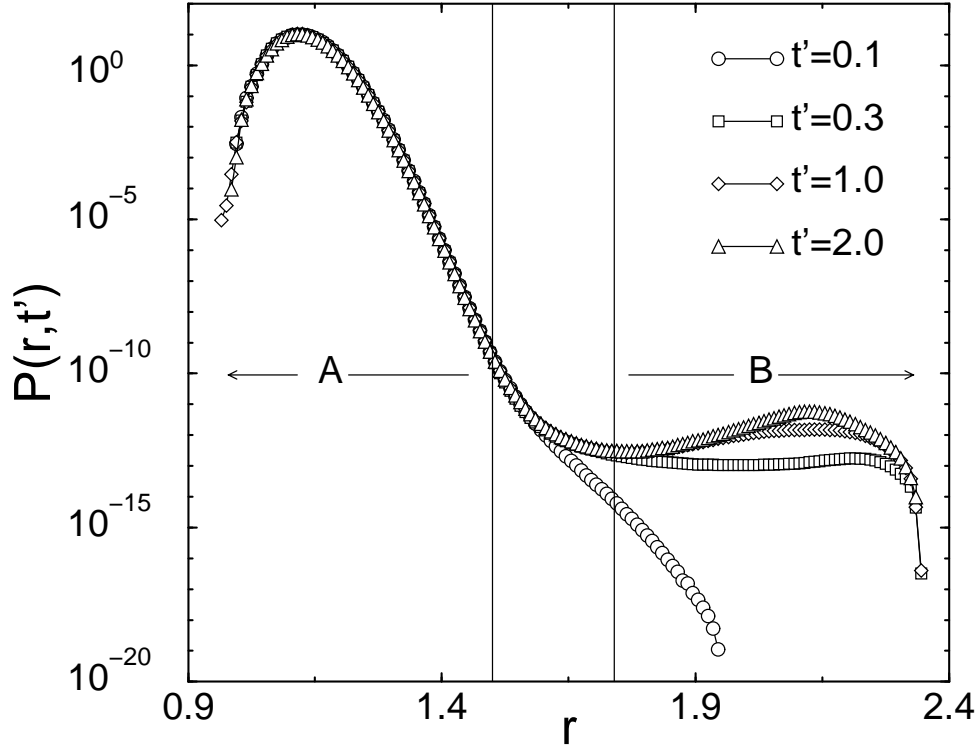


Figure 5.5: TPS probability distributions $P(r, t')$ for four $t' = 0.1, 0.3, 1.0, 2.0$ for the high energy barrier. The probability $P(r, t')$ is the chance that a path of length t' and starting in A will have the end point conformation with a diatomic distance r . The graph is the result of the matching of eight window calculations. These eight windows are defined as $r < 1.19$, $1.18 < r < 1.28$, $1.27 < r < 1.35$, $1.34 < r < 1.40$, $1.39 < r < 1.47$, $1.46 < r < 1.54$, $1.53 < r < 1.75$, $r > 1.74$. The errors on the histogram points are within the symbol size.

In Tab. 5.2 we give the relative computation time to reach the 2.5% error for each term.

In comparing both methods we have to recall that the efficiency of TPS depends strongly on the choice of t' . On the one hand the umbrella calculation of $C(t')$ is much faster for low values of t' . But on the other hand the error in the correction term $\langle h_B(t') \rangle_{A, H_B(T)}$ increases for lower t' . As a result there is an optimum t' for the error/CPU-time ratio, in this case approximately at $t' = 0.3$. Even for this optimized situation the TIS calculation is about two times faster. One could object that the correlation function in Fig.

TPS			
t'	$\frac{\langle h_B(T) \rangle_{AB}}{\langle h_B(t') \rangle_{AB}}$	$C(t')/10^{-13}$	$k_{A \rightarrow B}/10^{-13}$
0.1	3300 ± 100	0.0018 ± 0.0001	6.0 ± 0.5
0.3	7.54 ± 0.03	0.76 ± 0.02	5.8 ± 0.1
1.0	1.236 ± 0.005	4.8 ± 0.3	5.9 ± 0.4
2.0	0.553 ± 0.002	11.4 ± 0.9	6.3 ± 0.5
TIS			
$\langle \Phi_{A, \lambda_1} \rangle / \langle h_A \rangle$	$\mathcal{P}(\lambda_B \lambda_1)/10^{-13}$	$k_{A \rightarrow B}/10^{-13}$	
0.1196 ± 0.0005	49 ± 1	5.9 ± 0.2	

Table 5.1: Comparison of rate constants for the high energy barrier, computed with TPS at different t' and TIS. Contributing factors from Eq. (5.6) and Eq. (5.18) are also given.

5.4 has reached a plateau for $t = 1.5$ already, reducing the TPS computation time by a factor 3/4. But the choice for a path length $T = 1.5$ cannot be taken without a-priori knowledge. The first term in Eq. (5.6) implicitly depends on the path length T . Changing T would alter the ensemble and might result in a different shape of the flux correlation function. We did not check this in detail, but we believe that T cannot be chosen much smaller without introducing systematic errors. Furthermore, we emphasize here that we

TPS										
t'	$\frac{\langle h_B(T) \rangle_{AB}}{\langle h_B(t') \rangle_{AB}}$	W1	W2	W3	W4	W5	W6	W7	W8	Total
0.1	11.0	0.01	0.05	0.1	0.04	0.23	0.27	1.3	0.01	13.01
0.3	0.2	0.01	0.14	0.28	0.13	0.58	0.43	0.19	0.02	1.98
1.0	0.1	1.7	1.7	0.9	0.6	3.0	2.6	6.4	0.2	17.2
2.0	0.1	0.03	1.8	4.5	4.4	15.3	8.0	20.3	0.6	55.03
TIS										
$\langle \Phi_{A, \lambda_1} \rangle / \langle h_A \rangle$	Int λ_1	Int λ_2	Int λ_3	Int λ_4	Int λ_5	Total time				
0.07	0.265	0.09	0.15	0.21	0.215	1				

Table 5.2: Comparison of CPU-times required for the 2.5% error at each stage for the system with high energy barrier. The times are renormalized to the TIS total computation time. W1 to W8 denote the different windows used in the calculation, Int λ_1 to Int λ_5 denote the interface ensemble calculations.

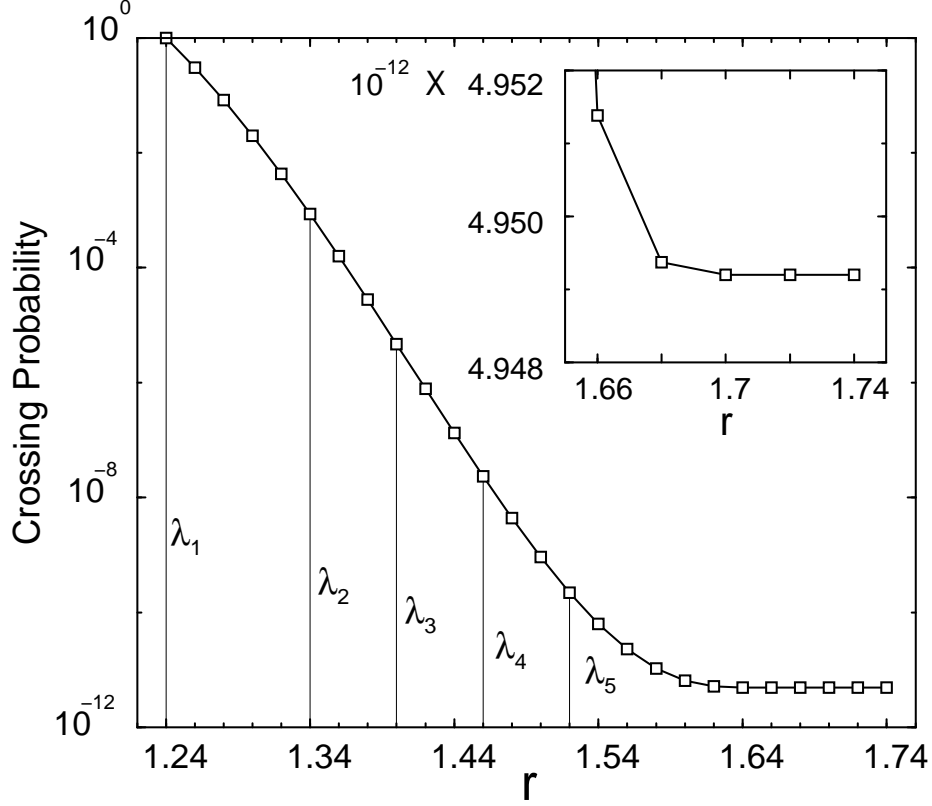


Figure 5.6: TIS crossing probability $\mathcal{P}(\lambda|\lambda_1) = \langle \bar{h}_{\Omega_{\lambda},A}^f \rangle_{\Phi_{A,\lambda_1}}$ as function of $\lambda = r$ for the system with a high energy barrier. The function is computed by matching the five interface ensemble calculations. These interfaces were chosen at: $\lambda_1 = 1.24, \lambda_2 = 1.34, \lambda_3 = 1.40, \lambda_4 = 1.46$ and $\lambda_5 = 1.52$. The error on the points is within symbol size. The inset is an enlargement in linear scale of the last part of the function. We clearly detect a horizontal plateau when approaching λ_B .

put much more effort in optimizing the TPS algorithm by tuning t' , the windows, the ratio between shooting, shifting and diffusion moves than we did for TIS.

Figure 5.7 shows the histograms of path lengths for each TIS ensemble calculation and shows why TIS is faster. Sampling paths of fixed length with TPS results in spending unnecessary computation time inside the initial and final stable regions A and B. In the TIS algorithm instead every path is adapted to its minimum length. Bringing the interface in closer to A reduces these transition times. TIS optimizes itself during the simulation.

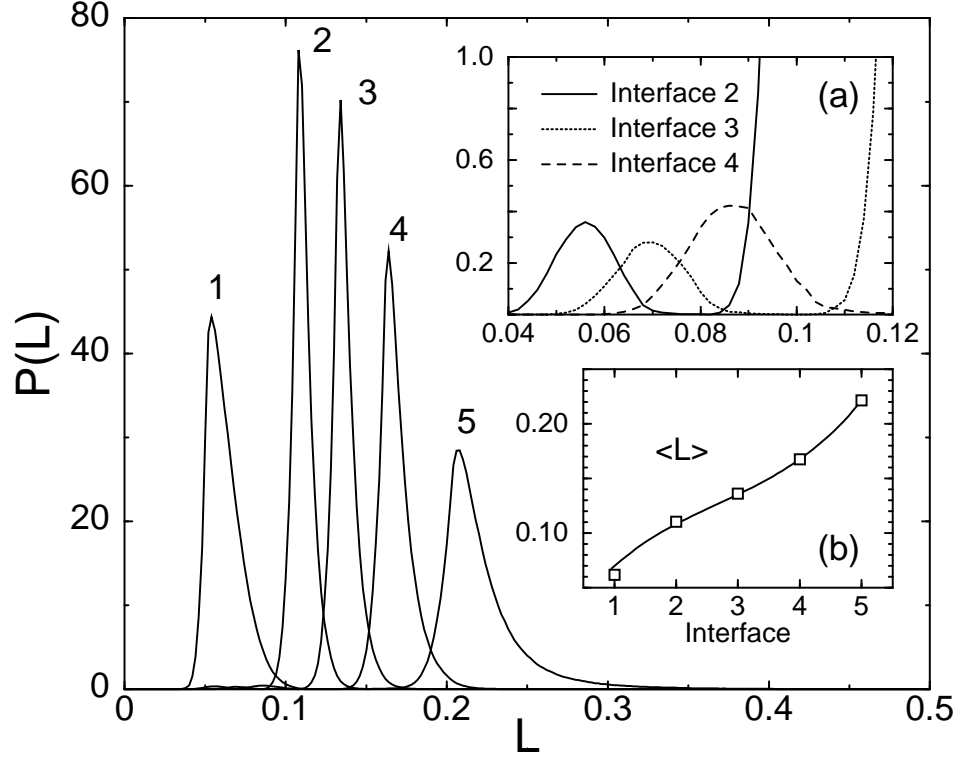


Figure 5.7: Histograms $P(L)$ of path length L for each ensemble, computed for the system with the high energy barrier. Inset (a) is an enlargement of the bottom left area, where windows 2,3,4 display a second peak. They represent that small fraction of paths that are able to cross all the interfaces up to the rightmost interface and do not have to return to A (cf. the trajectories with the white circle in Fig.3). Inset (b): average path length in each window. At variance with TPS the TIS algorithm adapts the path length to the ensemble. In going from interface 5 to interface 1 one gets closer to state A and the path length shortens accordingly.

5.4.4 System with Low Energy Barrier

In order to compare with previous results, we adopted the parameters from Ref. [163]. The total number of particles was $N = 9$, the total energy was $E = 9$ and the square simulation box was adjusted for a number density of 0.6. The barrier height is $h = 6$ and the width-parameter is $w = 0.25$. Minima are at $r \simeq 1.12$ and $r \simeq 1.62$, while the top of the barrier is at $r \simeq 1.37$. This barrier is much lower than in the previous section resulting

in more frequent transitions. An approximate rate constant could even be achieved by straightforward MD simulations.

For the TPS calculations we defined the stable states A and B by $r < r_A = 1.30$ and $r > r_B = 1.45$, respectively [163]. Using standard TPS simulation we computed the correlation function $\langle h_B(t) \rangle_{A, H_B(T)}$ with a total path length $T = 2$ (shown in Fig. 5.8). Next, we measured the probability

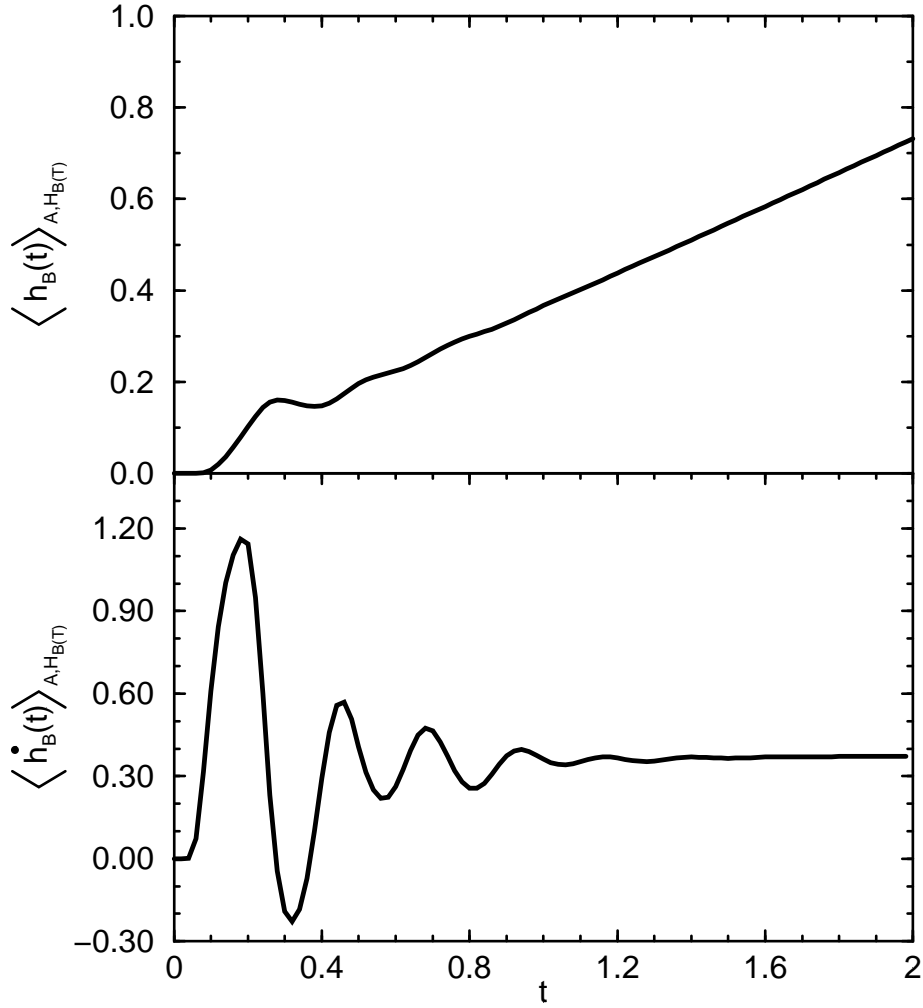


Figure 5.8: TPS correlation function $\langle h_B(t) \rangle_{A, H_B(T)}$ (top) and its time derivative (bottom) for the system with low energy barrier. The error is comparable to line thickness.

histograms to find the paths end point at a certain order parameter value r for four different times $t' = 0.1, 0.4, 0.8, 2.0$, using five windows [163] (see Fig. 5.9). As described in the previous section, matching the probability his-

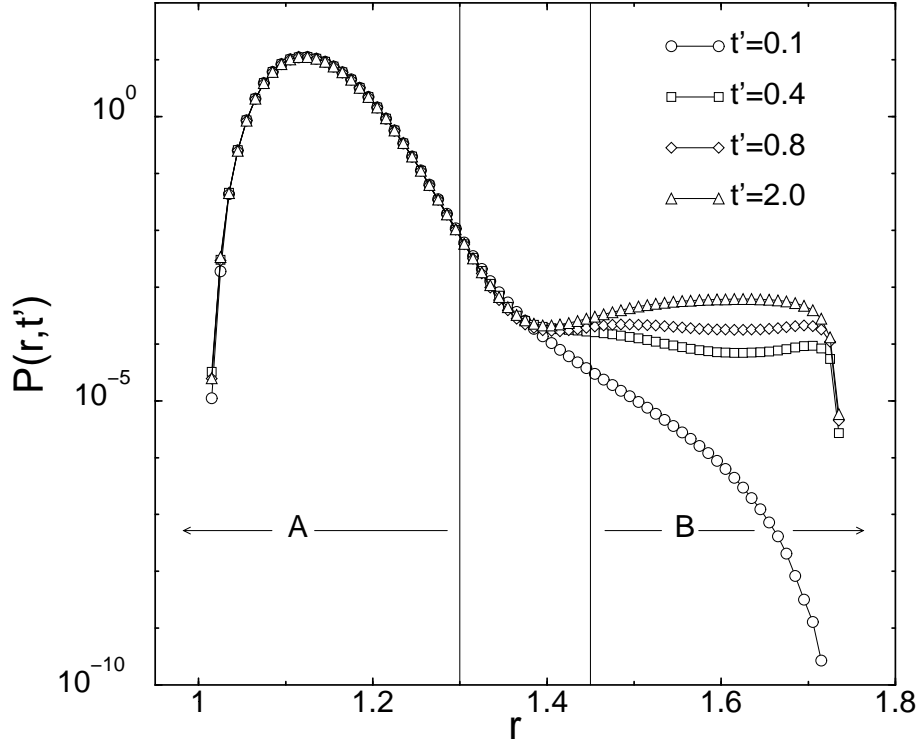


Figure 5.9: TPS probability distributions $P(r, t')$ for four $t' = 0.1, 0.4, 0.8, 2.0$ for the system with low energy barrier $P(r, t')$ as in Fig. 5.5. The graph is the result of the matching of five window calculations. These five window calculations are defined as $r < 1.22$, $1.21 < r < 1.26$, $1.25 < r < 1.30$, $1.29 < r < 1.46$, $r > 1.45$. The errors on the histogram points are within the symbol size.

tograms and subsequent integration leads to $C(t')$. The resulting final rate constants, shown in Tab. 5.3, are comparable with the results of Ref. [163], but more accurate. We will discuss these values after giving the results of TIS.

Fig. 5.10 shows that fast recrossings can occur for a low barrier, implying that r alone is not sufficient as an order parameter to define the stable states in the simulations. Apparently, this does not effect the TPS results much, but it is very important for TIS because of the assumption that stable region

TPS			
t'	$\frac{\langle h_B(T) \rangle_{AB}}{\langle h_B(t') \rangle_{AB}}$	$C(t')/10^{-5}$	$k_{A \rightarrow B}/10^{-5}$
0.1	47.3 ± 0.2	1.408 ± 0.007	6.67 ± 0.04
0.4	2.505 ± 0.007	2.67 ± 0.01	6.68 ± 0.03
0.8	1.240 ± 0.003	5.42 ± 0.05	6.72 ± 0.07
2.0	0.507 ± 0.001	13.9 ± 0.2	7.03 ± 0.09
TIS			
	$\frac{\langle \Phi_{A,\lambda_1} \rangle}{\langle h_A \rangle}$	$\langle \bar{h}_{B,A}^f \rangle_{\Phi_{A,\lambda_1}}/10^{-5}$	$k_{A \rightarrow B}/10^{-5}$
	0.2334 ± 0.0003	29.6 ± 0.2	6.90 ± 0.06

Table 5.3: Comparison of rate constants for the low energy barrier computed with TPS at different t' and with TIS, including the contributing factors from Eq. (5.6) and Eq. (5.18), respectively. Computation times are reported in units of the TIS CPU-time.

B is really stable and recrossings do not take place. To ensure the stability of the TIS stable states we chose a new order parameter that not only depends on the inter-atomic distance r in the dimer but also on a kinetic term, given by \dot{r} . The stable states can then be defined by

$$\begin{aligned}
 E_d(r, \dot{r}) &\equiv \frac{\dot{r}}{4} + V_{dw}(r) \\
 x \in A &\text{ if } r < 1.37 \text{ and } E_d(r, \dot{r}) \leq 1.5 \\
 x \in B &\text{ if } r > 1.37 \text{ and } E_d(r, \dot{r}) \leq 1.5,
 \end{aligned} \tag{5.23}$$

where E_d is the sum of the kinetic and potential energy of the dimer that has a reduced mass of $1/2$. In the $\{r, \dot{r}\}$ plane these stable states form a D-shape and an inverse D-shape regions for A and B respectively (see Fig. 5.11). Crossing the interface λ_A or λ_B implies that the vibrational energy is decreased below the threshold, $E_d = 1.5$. This threshold is made low enough to make fast recrossings to the other state unlikely. However, if we would have chosen it too low the paths would have become very long. We evaluated the crossing probability function in Eq. (5.18) for $n = 3$ interfaces. The entire crossing probability function was obtained by partitioning the phase space in sub-interfaces of the form $r = \lambda$ and $E_d(r, \dot{r}) = \lambda$ as shown in Fig. 5.11. Note that in TIS multidimensional or multiple order parameters can be used in one simulation without a problem. This is more difficult in TPS, where a proper mapping of the complete phase space is required.

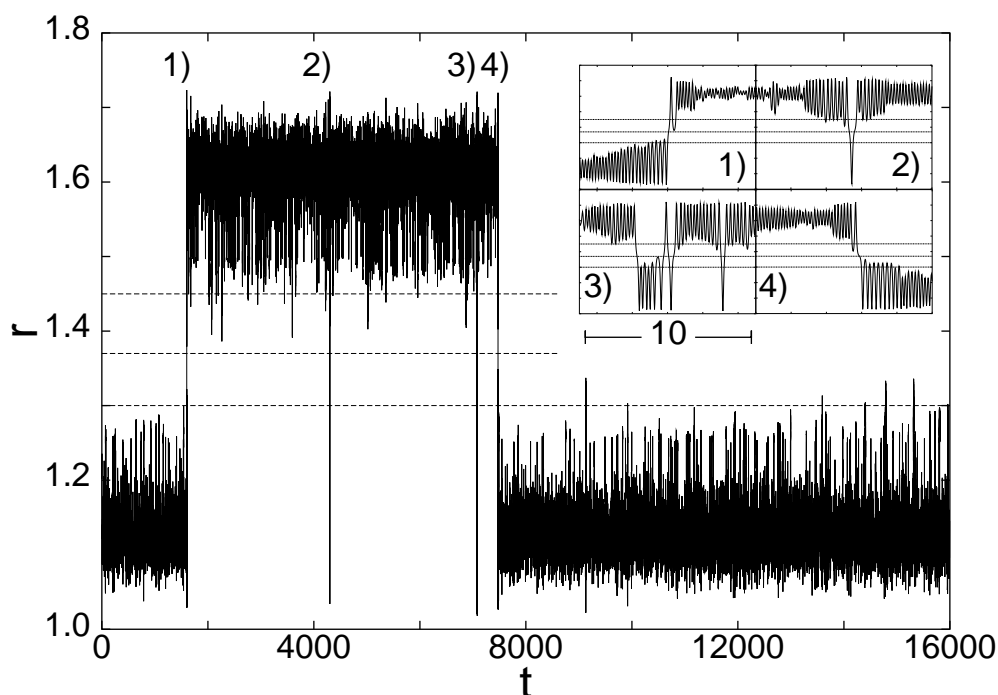


Figure 5.10: Intra-molecular distance of the dimer as function of time from a straightforward MD simulation for the system with the low energy barrier. Horizontal dashed line at 1.37 corresponds to the top of the potential barrier. Horizontal dashed lines at 1.3 and 1.45 correspond to the TPS state definitions of Ref. [163]. Insets are enlargements of four typical events on a scale of 10. 1) and 4) correspond to true reactive events, $A \rightarrow B$ and $B \rightarrow A$ respectively while 2) and 3) are non-true, fast recrossing events. In particular, event 3) shows capricious behavior with many crossings of the barrier. The figure shows a clear separation of timescales, $t_{\text{mol}} \sim 1$ and $t_{\text{rxn}} \sim 1000$.

Fig. 5.12 shows the final rematched crossing probability. The monotonically decreasing function tends to reach a plateau on approaching the last interface. The last two values are not exactly equal but differ by 0.03%, indicating that a small fraction of the paths crossing the one but last sub-interface still succeed to return to A without crossing λ_B . This difference is comparable with the chance of a new independent transition (given by the rate constant). Note that without the kinetic energy definition for the stable states Eq. (5.23), the final crossing probability and thus the rate constant would have been overestimated by a factor $5/4$.

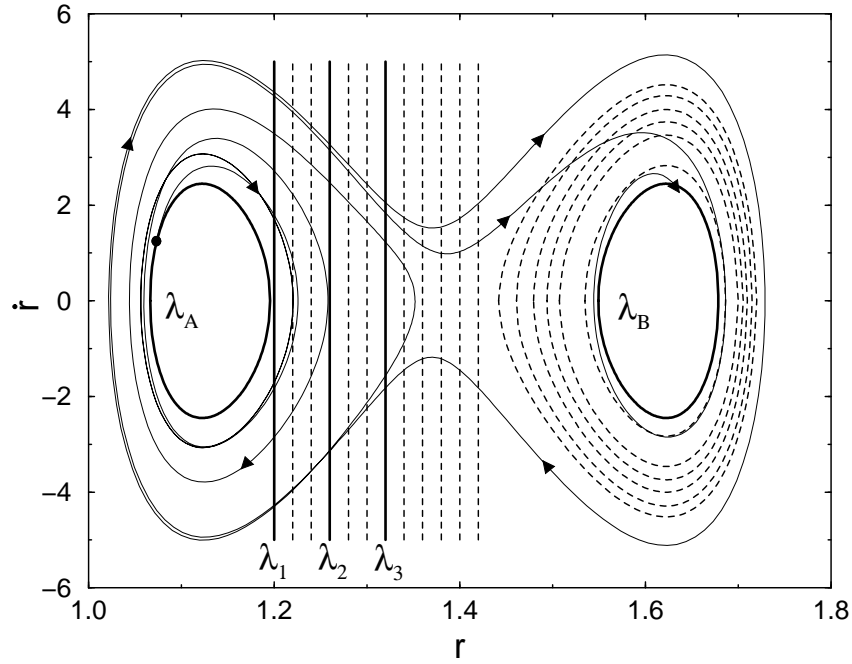


Figure 5.11: One calculated path of the low energy barrier system shown in the $\{r, \dot{r}\}$ plane. The vertical solid lines are the interface λ_1, λ_2 and λ_3 . The curves λ_A and λ_B are the boundaries of the TIS stable states. The dashed lines are the sub-interfaces. The path starts at the dot on λ_A and crosses the barrier three times before dissipating its energy and relaxing into state B .

For the effective flux $\langle \Phi_{A, \lambda_1} \rangle / \langle h_A \rangle$ calculation we performed MD simulations as described in section 5.4.2. In contrast to the high barrier case, λ_1 is not equal λ_A , and not all positive crossings with λ_1 are effective crossings. We counted only the first crossing when the system left region A and waited until the system fell back to region A before counting a new crossing. As the MD trajectory sometimes displayed a spontaneous transition to region B , we stopped the simulation and started again by replacing the system in a randomized configuration of A . Tab. 5.3 shows the final values and the corresponding errors of these calculations. The relative computation time for each term is detailed in Tab. 5.4.

If we compare the final results of Tab. 5.3 we see that the efficiency of TIS is more than nine times better than the TPS efficiency for $t' = 2$, and more than two times better than TPS value for $t' = 0.8$. But the TPS $t' = 0.1$ and

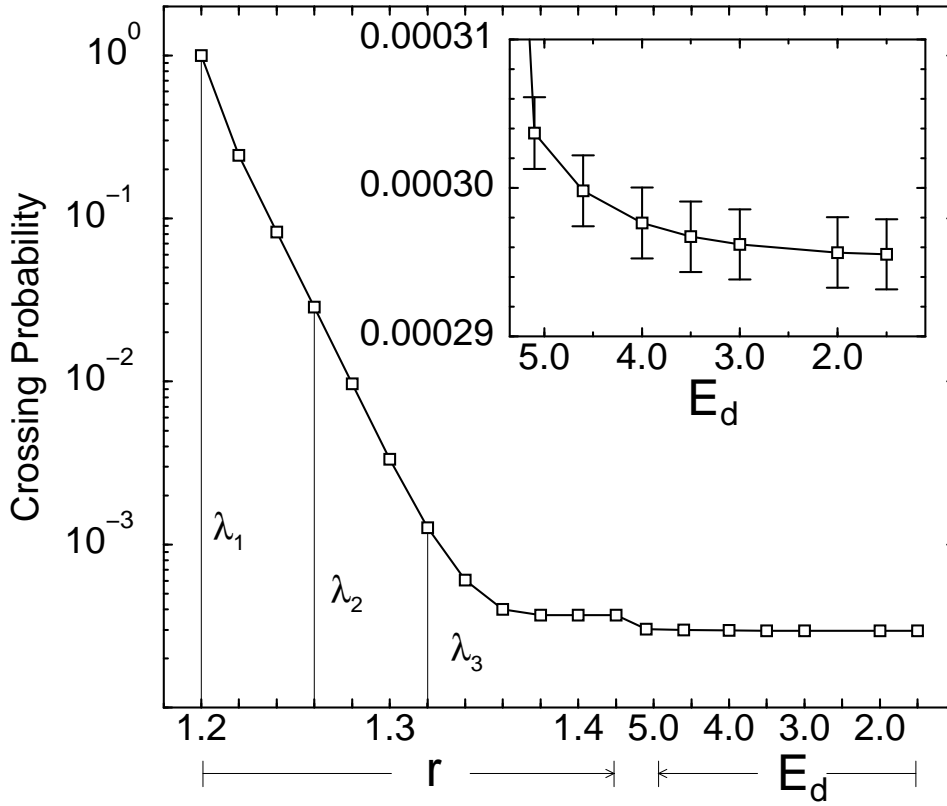


Figure 5.12: The crossing probability $\mathcal{P}(\lambda|\lambda_1)$ for the system with the low energy barrier. The function is computed by matching ensemble calculations with interfaces λ_1 at $r = 1.20$, λ_2 at $r = 1.26$ and λ_3 at $r = 1.32$. The inset is an enlargement of the final part. The function is converging to a plateau but has not yet reached it. The different values of the last points are due to the presence of fast recrossings. The error is inside the symbol size.

$t' = 0.4$ efficiencies are about 20 % better than TIS. When we compare the rate constants, however, we notice that the TPS results for different t' do not agree. Among the TPS rate constants only the $t' = 2$ case is consistent with the TIS result. We believe that the $t' = 0.1$ and $t' = 0.4$ results suffer from systematic errors. For instance, for the shorter paths the TPS simulations might not be completely ergodic. Another explanation might be that a path length of $T = 2$ is too short to allow convergence of the reactive flux. In the TIS calculation the average path length in the three interface simulations, from the closest to B to the closest to A , is, respectively, 7.4, 4.3, and 0.63;

TPS							
t'	$\frac{\langle h_B(T) \rangle_{AB}}{\langle h_B(t') \rangle_{AB}}$	W1	W2	W3	W4	W5	Total
0.1	0.68	0.03	0.009	0.01	0.1	0.001	0.83
0.4	0.4	0.09	0.03	0.04	0.25	0.01	0.82
0.8	0.28	0.21	0.07	0.11	1.5	0.04	2.21
2.0	0.35	0.28	0.38	0.93	7.27	0.14	9.35
TIS							
	$\frac{\langle \Phi_{A,\lambda_1} \rangle}{\langle h_A \rangle}$	Int λ_1	Int λ_2	Int λ_3	Total		
	0.015	0.085	0.45	0.45	1		

Table 5.4: Comparison of CPU-times required for the 2.5% error at each stage for the system with the low energy barrier. The times are renormalized to the TIS total computation time.

much longer than the TPS path length (see Fig.5.13). It is therefore surprising that the TPS approach with the simple stable state definition and very short paths still gives approximately the right rate constant. And indeed, when we computed the TPS correlation function with the TIS state definitions Eq. (5.23), we found that the path length had to be at least $T = 20$ to see a plateau. We think that TPS works even with the simple state definitions and the short paths because both positive flux and negative flux terms contribute to Eq. (5.6). The TPS algorithm collects many paths of which some are not real transitions, but fast recrossings. The cancellation of positive and negative terms of these fast recrossing paths ensure the (almost) correct final outcome. In TIS each path must be true transition event and contributes as a positive term in rate equation (5.18), enhancing the convergence. This explains that the CPU time for the TIS calculation despite the much longer paths is still comparable with TPS one for low t' . We note that the path ensemble using the more strict stable state definition is of course more useful in the analysis of the reaction mechanism.

For a more accurate comparison of the computation time we must keep the systematic errors lower than the statistical errors. In other words, we have to make sure that the results are converged. To test the convergence of the flux correlation function in TPS we can derive the following equality from Eq. (5.6) :

$$\frac{\langle h_B(t') \rangle_{A,H_B(T)}}{\langle h_B(t'') \rangle_{A,H_B(T)}} = \frac{C(t')}{C(t'')}. \quad (5.24)$$

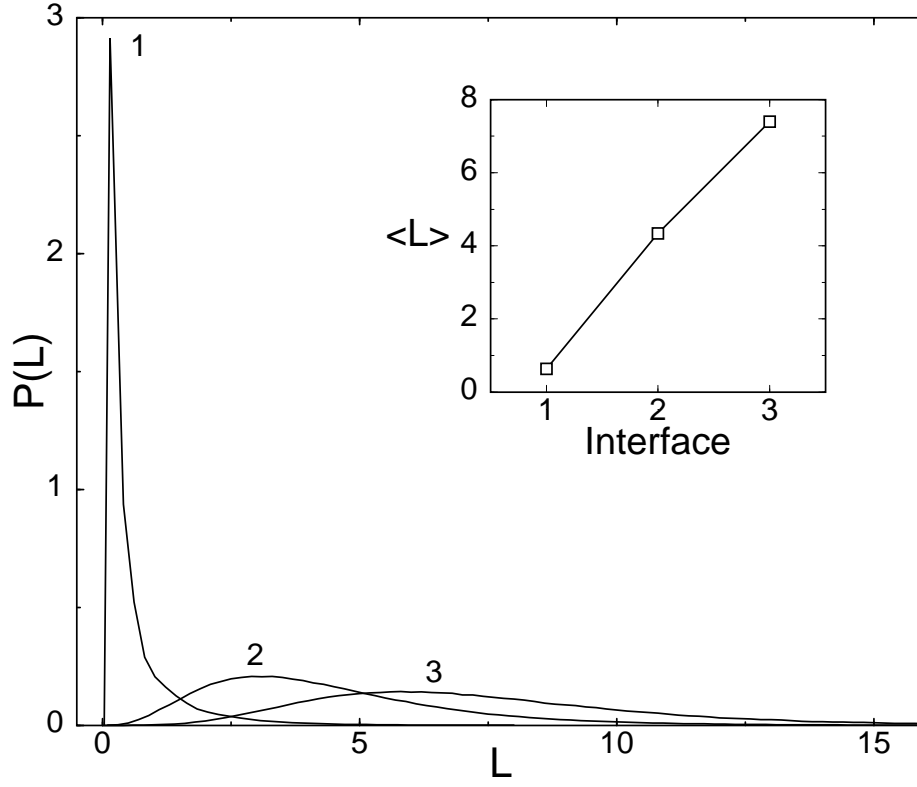


Figure 5.13: Path length distribution $P(L)$ for each interface ensemble in the low energy barrier system. The inset shows the average path length in each ensemble.

This equation is valid for any $t', t'' < T$ if T is large enough. We found that the equality does not hold for the system with the low barrier, indicating that T is too low in the TPS calculation. Further examination of the flux correlation function $\langle h_B(t') \rangle_{A, H_B(T)}$ reveals that the apparent plateau has in fact a small positive slope. Calculations for higher values of T suggest that one has to increase the path length at least to $T = 8$ to converge to a plateau. With this in mind we think that the TIS computation is about a factor five more efficient than the TPS algorithm for the model system with the low barrier.

5.5 Conclusion

We developed a novel method, named transition interface sampling, for the calculation of rate constants based on transition path sampling concepts. Just as the original transition path sampling, the new method enables the calculation of rate constants of transitions between stable states separated by high free energy barriers without prior knowledge of the reaction coordinate. The new algorithm is different in spirit from the rate constant calculation that was introduced in Refs [162, 163]. In TPS the time correlation function $\langle h_A(x_0)h_B(x_t) \rangle / \langle h_A \rangle$ is determined for a single time using an umbrella sampling scheme followed by a calculation of the reactive flux prefactor in a separate path sampling simulation. The path length used in this simulation has to be long enough for the plateau to be reached. The TIS method advocated here calculates the flux correlation directly by measuring the fluxes through a number of different interfaces and relating the flux through one interface to the next one. The big advantage of a flux instead of a correlation function is that trajectories going through the interfaces all contribute to the rate whereas in TPS there are recrossings to be counted. In addition, the new method improves the original TPS method on several other points. Once the interface is reached, the integration of motion can be stopped instead of going all the way to region B . In this way the TIS algorithm adapts itself to the optimal path length. One does not have to optimize the new method as much as TPS, where one has to find the optimal t' value and a proper balance between shooting and shifting. Besides being faster, the concept of calculating a flux comes natural with the rate constant definition, and implementation of the algorithm is hence simpler. Also, multidimensional or even discrete order parameters can easily be implemented in TIS. In the illustrative example we showed that we can obtain an increase in efficiency of at least a factor of two to five with respect to the TPS method used in Ref. [163].

However, one has to be more careful in the definitions of the stable states, meaning that stable states have to be really stable. In TPS the choice of stable states is a bit more flexible as the final rate constant consists of cancellation of positive and negative terms. In section 5.4.4 we showed how this problem for TIS can be solved by defining stable regions that explicitly depend on kinetic energy terms.

In summary, we believe that the TIS algorithm can make the rate constant calculation of many processes feasible that were hitherto difficult to obtain. For instance, chemical reactions in solution, isomerization of clusters and conformational transitions in biomolecules. In a future publication

we will report on these, more complex, applications.

5.5.1 Appendix A: Proof of Flux Equivalence

Here we prove that $\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \theta(\lambda^* - \lambda) \theta(\lambda(\Delta t) - \lambda^*) = \dot{\lambda} \delta(\lambda - \lambda^*) \theta(\dot{\lambda})$ by showing that

$$\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int d\dot{\lambda} \int d\lambda \theta(\lambda^* - \lambda) \theta(\lambda(\Delta t) - \lambda^*) a(\lambda, \dot{\lambda}) = \int d\dot{\lambda} \int d\lambda \dot{\lambda} \delta(\lambda - \lambda^*) \theta(\dot{\lambda}) a(\lambda, \dot{\lambda}) \quad (5.25)$$

for any arbitrary function $a(\lambda, \dot{\lambda})$. Here we used the short hand notation $\lambda = \lambda(x_0)$, $\dot{\lambda} = \dot{\lambda}(x_0)$, and $\lambda(\Delta t) = \lambda(x_{\Delta t})$.

In first order of Δt we can write

$$\theta(\lambda^* - \lambda) \theta(\lambda(\Delta t) - \lambda^*) = \theta(\lambda^* - \lambda) \theta(\lambda + \Delta t \dot{\lambda} - \lambda^*), \quad (5.26)$$

which is only non-zero for $\dot{\lambda} > 0$ and $\lambda^* - \Delta t \dot{\lambda} < \lambda < \lambda^*$. Thus, we can rewrite the first part of Eq. 5.25 as:

$$\begin{aligned} \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_0^\infty d\dot{\lambda} \int_{\lambda^* - \Delta t \dot{\lambda}}^{\lambda^*} d\lambda a(\lambda, \dot{\lambda}) &= \\ \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_0^\infty d\dot{\lambda} \int_{\lambda^* - \Delta t \dot{\lambda}}^{\lambda^*} d\lambda \left[a(\lambda^*, \dot{\lambda}) + \sum_{m=1}^\infty \frac{(\lambda - \lambda^*)^m}{m!} \frac{\partial^m a(\lambda, \dot{\lambda})}{\partial \lambda^m} \Big|_{\lambda=\lambda^*} \right] &= \\ = \int_0^\infty d\dot{\lambda} a(\lambda^*, \dot{\lambda}) \dot{\lambda} + \mathcal{O}(\Delta t) &= \int d\dot{\lambda} \int d\lambda \dot{\lambda} \delta(\lambda - \lambda^*) \theta(\dot{\lambda}) a(\lambda, \dot{\lambda}) \end{aligned} \quad (5.27)$$

In the second line, we have used an infinite Taylor expansion and in the last line we have dropped all terms of order $\mathcal{O}(\Delta t)$ and higher. The equality derived here justifies the change between $\frac{1}{\Delta t} \theta(\lambda^* - \lambda) \theta(\lambda(\Delta t) - \lambda^*)$ and $\dot{\lambda} \delta(\lambda - \lambda^*) \theta(\dot{\lambda})$, which is used in Eqs. 5.1 and 5.14.

5.5.2 Appendix B: Flux Relation

In this appendix we show how the effective flux Φ_{A, λ_i} can be related to the effective flux $\Phi_{A, \lambda_{i-1}}$ through an interface λ_{i-1} closer to A . If at time $t = 0$ a trajectory passes interface λ_i while having started in A some time earlier, there must always be a unique time when it passed interface λ_{i-1} for the first time. Therefore we can write:

$$\Phi_{A, \lambda_i}(x_0) = \Phi_{A, \lambda_i}(x_0) \int_0^{t_{A \cup \Omega_{\lambda_i}}^{(x_0)}} dt \Phi_{A, \lambda_{i-1}}(x_{-t}) \quad (5.28)$$

and hence,

$$\begin{aligned}
\langle \Phi_{A,\lambda_i}(x_0) \rangle &= \\
&= \int_0^\infty dt \langle \Phi_{A,\lambda_{i-1}}(x_{-t}) \Phi_{A,\lambda_i}(x_0) \theta(t_{A \cup \Omega_{\lambda_i}}^b(x_0) - t) \rangle \\
&= \int_0^\infty dt \langle \Phi_{A,\lambda_{i-1}}(x_0) \Phi_{A,\lambda_i}(x_t) \theta(t_{A \cup \Omega_{\lambda_i}}^b(x_t) - t) \rangle \\
&= \langle \Phi_{A,\lambda_{i-1}}(x_0) \int_0^\infty dt \Phi_{A,\lambda_i}(x_t) \theta(t_{A \cup \Omega_{\lambda_i}}^b(x_t) - t) \rangle \\
&= \left\langle \Phi_{A,\lambda_{i-1}}(x_0) \int_0^{t_{A \cup \Omega_{\lambda_i}}^f(x_0)} dt \Phi_{A,\lambda_i}(x_t) \right\rangle \\
&= \langle \Phi_{A,\lambda_{i-1}}(x_0) \bar{h}_{\Omega_{\lambda_i},A}^f(x_0) \rangle. \tag{5.29}
\end{aligned}$$

The one but last equation follows because for each phase point x and phase space region Ω it can be shown that $t > t_\Omega^f(x) \Rightarrow t_\Omega^b(f(x, t)) \leq t \Rightarrow \theta(t_\Omega^b(f(x, t)) - t) = 0$. We rewrite the last expression of Eq. (5.29) as a different ensemble average:

$$\begin{aligned}
&\langle \Phi_{A,\lambda_{i-1}}(x_0) \bar{h}_{\Omega_{\lambda_i},A}^f(x_0) \rangle = \\
&= \frac{\langle \Phi_{A,\lambda_{i-1}}(x_0) \bar{h}_{\Omega_{\lambda_i},A}^f(x_0) \rangle}{\langle \Phi_{A,\lambda_{i-1}}(x_0) \rangle} \times \langle \Phi_{A,\lambda_{i-1}}(x_0) \rangle \\
&\equiv \langle \bar{h}_{\Omega_{\lambda_i},A}^f(x_0) \rangle_{\Phi_{A,\lambda_{i-1}}} \times \langle \Phi_{A,\lambda_{i-1}}(x_0) \rangle, \tag{5.30}
\end{aligned}$$

where $\langle \dots \rangle_{\Phi_{A,\lambda_{i-1}}}$ denotes the ensemble average over all phase points x_0 for which $\Phi_{A,\lambda_{i-1}}(x_0) \neq 0$. The last equality gives rise to Eq. (5.17).

5.5.3 Appendix C: Alternative Transmission Coefficient

For some systems, the free energy calculation combined with the calculation of a transmission coefficient will be more efficient than path sampling. For instance, if one wants to know the free energy profile as function of a certain reaction coordinate, it is more efficient to use this information instead of calculating the crossing probability function, which would require a comparable additional computational effort. Of course, an accurate rate constant can only be obtained for systems for which the transmission coefficient is not too small. Else, the free energy approach will fail and the path sampling technique is required.

If one uses the free energy calculation in combination with a transmission coefficient, then usually the free energy calculation is the computationally most expensive part. However, if the barrier crossings are diffusive with many recrossings of the transition dividing surface, then the Bennet-Chandler expression of the transmission coefficient (5.3) can be very inefficient. For those cases, Ruiz-Montero et al. [173] designed an alternative transmission expression to improve the efficiency.

The effective positive flux formalism derived in this chapter, allows also an alternative expression for the transmission coefficient. Starting with Eq. 5.16 and with the use of Eq. 5.29 we can write:

$$\begin{aligned}
 k_{AB} &= \frac{\langle \Phi_{A,\lambda_B} \rangle}{\langle h_A \rangle} = \frac{\langle \Phi_{A,\lambda^*} \bar{h}_{\Omega_{B,A}}^f \rangle}{\langle h_A \rangle} = \\
 &= \frac{\langle \bar{h}_{A,\Omega_{\lambda^*}}^b \delta(\lambda - \lambda^*) \theta(\dot{\lambda}) \dot{\lambda} \bar{h}_{\Omega_{B,A}}^f \rangle}{\langle h_A \rangle} = \frac{\langle \bar{h}_{A,\Omega_{\lambda^*}}^b \delta(\lambda - \lambda^*) \theta(\dot{\lambda}) \dot{\lambda} \bar{h}_{\Omega_{B,A}}^f \rangle}{\langle \delta(\lambda - \lambda^*) \rangle} \times \\
 &\times \frac{\langle \delta(\lambda - \lambda^*) \rangle}{\langle \dot{\lambda} \delta(\lambda - \lambda^*) \theta(\dot{\lambda}) \rangle} \times \frac{\langle \theta(\lambda^* - \lambda) \rangle}{\langle h_A \rangle} \times \frac{\langle \dot{\lambda} \delta(\lambda - \lambda^*) \theta(\dot{\lambda}) \rangle}{\langle \theta(\lambda^* - \lambda) \rangle} = \\
 &= \frac{\langle \bar{h}_{A,\Omega_{\lambda^*}}^b \theta(\dot{\lambda}) \dot{\lambda} \bar{h}_{\Omega_{B,A}}^f \rangle_{\lambda^*}}{\langle \dot{\lambda} \theta(\dot{\lambda}) \rangle_{\lambda^*}} \times \frac{\langle \theta(\lambda^* - \lambda) \rangle}{\langle h_A \rangle} \times k_{AB}^{TST}. \quad (5.31)
 \end{aligned}$$

Here we have used Eqs. (5.15) and (5.25) in the second line and the substitution of the TST rate constant (5.1) in the last line. Using the fact that $\langle \theta(\lambda^* - \lambda) \rangle \approx \langle h_A \rangle$ we can write for the transmission coefficient:

$$\kappa^{TIS} = \frac{\langle \bar{h}_{A,\Omega_{\lambda^*}}^b \theta(\dot{\lambda}) \dot{\lambda} \bar{h}_{\Omega_{B,A}}^f \rangle_{\lambda^*}}{\langle \dot{\lambda} \theta(\dot{\lambda}) \rangle_{\lambda^*}}. \quad (5.32)$$

This expression is more efficient than Eq. (5.3) as it only exists of positive terms and therefore enhances the convergence. It also limits the number of acquired integration steps. The calculation is performed by taking a set of independent configurations with $\lambda = \lambda^*$ obtained from the free energy calculation. For each configuration, we consider a set of random velocities of all particles in the system taken from the appropriate Maxwell-Boltzmann distribution for constant temperature NVT or constant energy NVE. Then, only if $\dot{\lambda} > 0$ we start integrating the equations of motion backward in time.

Now, if the crossing is diffusive, there is big change that this backward integration will rapidly cross λ^* . As for that case $\bar{h}_{A,\Omega_{\lambda^*}}^b = 0$, we can directly stop the integration. This decreases strongly the total number of integration steps compared to Eq. (5.3), where we need to perform the integration till reaching either reactant or product side for both $\dot{\lambda} > 0$ as $\dot{\lambda} < 0$. Besides, if the backward integration does end up in A without recrossing λ^* , then this trajectory is likely to have a high kinetic energy and thus also the forward integration will probably not be trapped on the barrier. It would therefore be interesting to compare this method with that of Ref. [173] that was precisely designed for diffusive barrier problems.

Chapter 6

Non-Catalyzed Hydration of Ethylene

A Combination of Born-Oppenheimer Dynamics and Transition Interface Sampling ¹

We applied the new method of transition interface sampling in combination with Born-Oppenheimer dynamics on the direct gas-phase hydration of ethylene. This method is an improvement of the original transition path sampling, that is designed for the calculation of rate constants for systems for which transition state theory does not work. The direct hydration of ethylene implies the breaking of the ethylene double bond and the water OH bond making the reaction barrier high. The sampled paths show very different behavior and seldomly cross the barrier via the saddle point of the potential energy surface.

6.1 Introduction

The direct hydration/dehydration reaction between ethylene and ethanol without a catalyst can occur under extreme conditions. Experimental measurements and theoretical calculations show that ethanol dehydrates into ethylene and water, when it is strongly chemically activated [118, 119] or vibrational excited by a laser [120, 121, 123]. Park et al have calculated rate constants in the temperature range of 700-2500 K by very accurate *ab initio* methods within the context of transition state theory [122]. However, if the

¹This chapter is based on work in progress, to be published as a letter: Titus S. van Erp, Evert Jan Meijer, "Non-Catalyzed Hydration of Ethylene. A Combination of Born-Oppenheimer Dynamics and Transition Interface Sampling"

system has many saddle points and temperatures are high, it is very uncertain that the transition state theory is accurate. To overcome these problems Chandler and co-workers proposed the transition path sampling method (TPS) [164]. Transition interface sampling (TIS) improves strongly the efficiency of TPS and it is conceptually more transparent as it needs only one type of Monte-Carlo (MC) move [174]. The TPS method uses two types of MC moves, the shooting and shifting move. The TIS method uses only an adapted shooting moves at which the path length is variable. This allows to reduce the number of integration steps to the strict minimum. Shifting moves are redundant in the TIS sampling method. Besides a more efficient calculation of the rate constant, the analysis of the sampled paths of the TIS algorithm, yields more useful information about the reaction mechanism than TPS [174]. Geissler et al successfully applied the path sampling technique in combination with the Car-Parrinello (CP) method for the proton transfer in a small water cluster [141] and for the autoionization in liquid water [142]. With the same techniques, Ensing and Baerends have studied reaction paths for the chemical reaction between iron(II) and hydrogen peroxide in aqueous solution [175]. Here we apply the TIS sampling technique to the hydration reaction of ethylene in combination with Born-Oppenheimer Molecular Dynamics (MD).

6.2 Theory and Methods

The basic concept of TIS theory is that besides stable definitions for A and B also overall states \mathcal{A} and \mathcal{B} are introduced as indication of the reactant and product state. A particle is no longer assigned to one state on basis of its instantaneous position, but on its position in the past. Using these state definitions the TIS rate constant no longer depends on time. This means in practice that the path length of the sampled paths can be limited to the strict minimum. The TIS expression for the rate constant is:

$$k_{AB} = \frac{\langle h_{\mathcal{A}}(0) \dot{h}_{\mathcal{B}}(0) \rangle}{\langle h_{\mathcal{A}} \rangle} = \frac{\langle \Phi_{A,\lambda_0} \rangle}{\langle h_{\mathcal{A}} \rangle} \prod_{i=1}^n \langle \bar{h}_{\Omega_{\lambda_i}, A}^f \rangle_{\Phi_{A,\lambda_{i-1}}} \quad (6.1)$$

The last part of Eq. 6.1 allows the computational calculation for rare events in a similar way as in umbrella sampling. The expression consists of the flux $\langle \Phi_{A,\lambda_0} \rangle / \langle h_{\mathcal{A}} \rangle$ out of stable region A times the product of the conditional probabilities $\langle \bar{h}_{\Omega_{\lambda_i}, A}^f \rangle_{\Phi_{A,\lambda_{i-1}}}$. These are the probabilities for a particle coming from A to cross interface λ_i given that it has crossed interface λ_{i-1} .

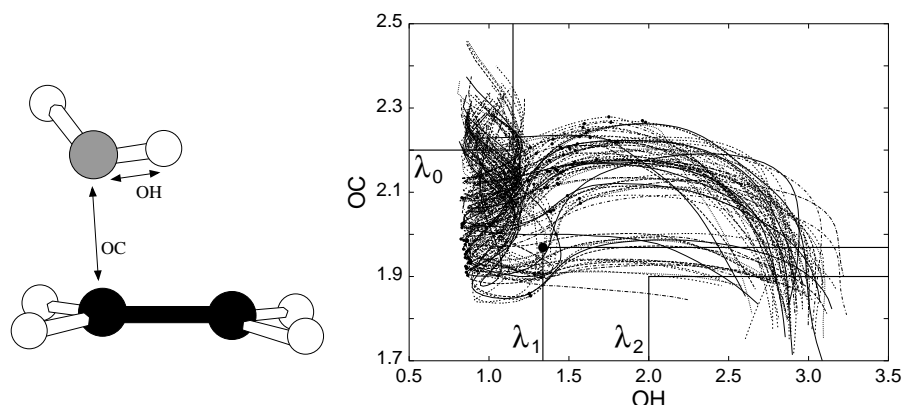


Figure 6.1: Left: illustration of the two order parameters OC and OH. Right: all 89 sampled paths projected in the OC-OH square. λ_1 is the interface, λ_0 and λ_2 are the boundaries of stable state A: (ethylene+water) and B: (ethanol) respectively. The large black dot at the corner of interface λ_1 is the true transition state, the small black dots represent the highest potential energy points (the barrier crossing points) for these paths.

before. We calculated this conditional probability for one interface between the stable states. The definitions of the stable states and interface, depending on two order parameters (the distances O-C and O-H), are given in Fig. 6.1.

As the dynamics involves large fluctuations in the electron density due to the bond breaking and forming, we used the BO dynamics method of the CPMD package [35]. We did not use the 'on the fly' Car-Parrinello (CP) algorithm, that was not stable under these conditions. The time step was 0.73 fs and after each 5 time steps the algorithm checked if the system has reached either *A* or *B*. Each path was accepted according to the Metropolis MC algorithm. Temperature of the Boltzmann-weights was 1000 K. After each shooting move we continued the BO dynamics at a constant energy along the path.

6.3 Results and Conclusion

From a total of 387 shooting moves we collected 89 different paths with the condition of starting in *A* and crossing the interface λ_1 at least once. We found a value 82 % for the conditional probability to reach *B*. In Fig. 6.1 these 89 paths are projected in the {O-C, O-H} plane. The small dots indicate

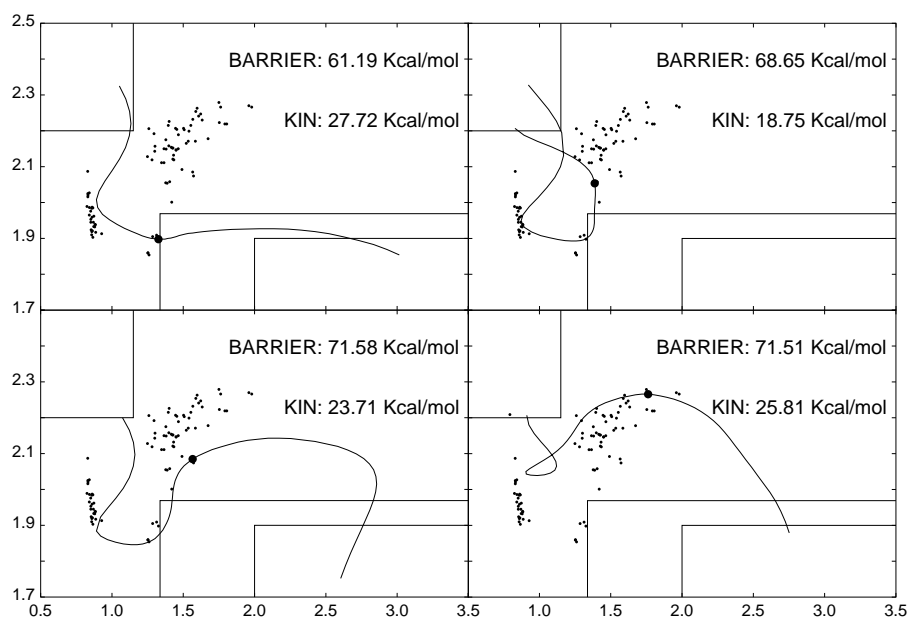


Figure 6.2: Four very different pathways taken from Fig. 6.1. The small dots are the same as in Fig. 6.1. The large dot indicates the highest potential energy point of the given path. The potential and kinetic energy at this barrier crossing point are also given for these four paths.

the potential maximum barrier height for each path (for a better view see Fig 6.2 where these points are also depicted). As we see, most path cross the barrier at a point not closely to the transition state (the large dot at the corner of the interface λ_1), that is the saddle point (60.1 kcal/mol). Fig. 6.2 shows four typical, but very different pathways from Fig. 6.1. The big dot indicates the point of barrier crossing of the corresponding path. The barrier height and the kinetic energy at this point are also given in Fig. 6.2.

Additional interface ensemble calculations with interfaces closer to region *A* are needed complete the rate constant calculation with the TIS method. However, we can already draw the conclusion that the pathways of this system and with this temperature have very different character and seldomly reaches their highest potential energy close to the transition state. This implies that rate constants are not reliable when calculated within the context of transition state theory.

Bibliography

- [1] P. Ball, *Life's matrix: A biography of water* (Farrar, Straus, and Giroux, New York, 2000).
- [2] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
- [3] H. S. Frank and M. W. Evans, J. Chem. Phys. **13**, 507 (1945).
- [4] A. Ben-Naim, *Hydrophobic Interactions* (Plenum Press, New York, 1980).
- [5] X. Xu, C. De Almeida, and M. J. Antal, J. Supercrit. Fluids **3**, 228 (1990).
- [6] X. Xu, C. P. De Almeida, and M. J. Antal, Ind. Eng. Chem. Res. **30**, 1478 (1991).
- [7] G. C. Groenenboom *et al.*, PRL **84**, 4072 (2000).
- [8] P. Hohenberg and W. Kohn, Phys. Rev. B **136**, 864 (1964).
- [9] W. Kohn and L. J. Sham, Phys. Rev. **140**, 1133 (1965).
- [10] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).
- [11] A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
- [12] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
- [13] M. Sprik, J. Hutter, and M. Parrinello, J. Chem. Phys. **105**, 1142 (1996).
- [14] P. L. Silvestrelli, M. Bernasconi, and M. Parrinello, Chem. Phys. Lett. **277**, 478 (1997).
- [15] P. G. Dacosta, O. H. Neilsen, and C. Kunc, J. Phys. C. **19**, 3163 (1986).
- [16] G. Pastore, E. Smargiassi, and F. Buda, Phys. Rev. A **44**, 6334 (1991).

- [17] J. C. Philips, Phys. Rev. **112**, 685 (1958).
- [18] M. H. Cohen and V. Heine, Solid State Physics **24**, 37 (1970).
- [19] M. T. Yin and M. L. Cohen, Phys. Rev. B **26**, 3259 (1982).
- [20] N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).
- [21] D. Marx and J. Hutter, in *Modern Methods in Algorithms of Quantum Chemistry*, Vol. 1 of *NIC Series*, edited by J. Grotendorst (John von Neumann Institute for Computing, Jülich, 2000), pp. 301–449.
- [22] M. Fuchs and M. Scheffler, Comput. Phys. Commun. **119**, 67 (1999).
- [23] D. Frenkel and B. Smit, *Understanding molecular simulation*, 2nd ed. (Academic Press, San Diego, CA, 2002).
- [24] W. den Otter and W. Briels, J. Chem. Phys. **109**, 4139 (1998).
- [25] M. Sprik and G. Ciccotti, J. Chem. Phys. **109**, 7737 (1998).
- [26] F. Franks, in *Water: A comprehensive treatise*, edited by F. Franks (Plenum Press, New York, 1973), Vol. 2.
- [27] A. Curioni *et al.*, J. Am. Chem. Soc. **116**, 11251 (1994).
- [28] E. J. Meijer and M. Sprik, J. Am. Chem. Soc. **120**, 6345 (1998).
- [29] A. K. Soper and J. L. Finney, Phys. Rev. Lett. **71**, 4346 (1993).
- [30] W. L. Jorgensen and J. D. Madura, J. Am. Chem. Soc. **105**, 1407 (1983).
- [31] M. Ferrario, M. Haughney, I. R. McDonald, and M. L. Klein, J. Chem. Phys. **93**, 5156 (1990).
- [32] G. Pálkás, I. Bakó, K. Heinzinger, and P. Bopp, Mol. Phys. **73**, 897 (1991).
- [33] H. Tanaka and K. E. Gubbins, J. Chem. Phys. **97**, 2626 (1992).
- [34] A. Laaksonen, P. G. Kusalik, and I. M. Svishchev, J. Phys. Chem. A **101**, 5910 (1997).
- [35] CPMD, version 3.0f, developed by J. Hutter, P. Ballone, M. Bernasconi, P. Focher, E. Fois, S. Goedecker, M. Parrinello, and M. Tuckermann, at MPI für Festkörperforschung and IBM Zurich Research Laboratory (1990-1997).

- [36] R. N. Barnett and U. Landman, *Phys. Rev. B* **48**, 2081 (1993).
- [37] ADF 2.3, see Refs. [176–178] Theoretical Chemistry, Vrije Universiteit, Amsterdam.
- [38] L. González, O. Mó, and M. Yáñez, *J. Chem. Phys.* **109**, 139 (1998).
- [39] M. Schütz *et al.*, *J. Chem. Phys.* **107**, 4597 (1997).
- [40] Kohn-Sham orbitals are expanded in an even-tempered, all-electron Slater type basis set augmented with 2p and 3d polarization functions for H and 3d and 4f polarization functions for C and O. The core states are kept frozen.
- [41] W. A. Benedict, N. Gailer, and E. K. Plyler, *J. Chem. Phys.* **24**, 1139 (1956).
- [42] S. Nosé, *J. Chem. Phys.* **81**, 511 (1984).
- [43] S. Nosé, *Mol. Phys.* **52**, 255 (1984).
- [44] W. G. Hoover, *Phys. Rev. A* **31**, 1695 (1985).
- [45] T. Shimanouchi, *Tables of molecular vibrational frequencies consolidated*, Volume I, National Bureau of Standards (1972).
- [46] F. Franks and D. J. Ives, *Rev. Chem. Soc.* **20**, 1 (1966).
- [47] D. Eisenberg and W. Kauzmann, *The Structure and Properties of Water* (Clarendon Press, Oxford, U.K., 1969).
- [48] F. Franks and D. S. Reid, in *Water: A comprehensive treatise*, edited by F. Franks (Plenum Press, New York, 1973), Vol. 2.
- [49] M. Blandamer, in *Water: A comprehensive treatise*, edited by F. Franks (Plenum Press, New York, 1973), Vol. 2.
- [50] *Structure of Water and Aqueous Solutions*, edited by W. A. P. Luck (Verlag Chemie-Physik, Weinheim, Germany, 1974).
- [51] F. Franks, in *Water: A comprehensive treatise*, edited by F. Franks (Plenum Press, New York, 1975), Vol. 4.
- [52] *L'eau et les systèmes biologiques; (Water and biological systems)*, edited by A. Alfsen and A. J. Berteaud (Editions due CNRS, Paris, 1976).

- [53] F. Franks and S. Mathias, *Biophysics of Water* (Wiley, New York, 1982).
- [54] F. Franks and J. E. Desnoyers, in *Water science reviews*, edited by F. Franks (Cambridge Univ. Press, Cambridge, 1985), Vol. 2.
- [55] M. Mohr, D. Marx, M. Parrinello, and H. Zipse, *Chem. Eur. J.* **6**, 4009 (2000).
- [56] T. S. van Erp and E. J. Meijer, (to be published).
- [57] A. Coccia, P. L. Indovina, F. Pado, and V. Viti, *Chem. Phys.* **7**, 30 (1975).
- [58] K. R. Harris, P. J. Newitt, and Z. J. Derlacki, *J. Chem. Soc. Faraday Trans.* **94**, 1963 (1998).
- [59] K. Mizuno, Y. Miyashita, Y. Shindo, and H. Ogawa, *J. Phys. Chem.* **99**, 3225 (1995).
- [60] R. Ludwig, *Chem. Phys.* **195**, 329 (1995).
- [61] R. Lamanna and S. Cannistraro, *Chem. Phys.* **213**, 95 (1996).
- [62] M. Brai and U. Kaatze, *J. Phys. Chem.* **96**, 8946 (1992).
- [63] N. Nishi *et al.*, *J. Phys. Chem.* **99**, 462 (1995).
- [64] G. Onori and A. Santucci, *J. Mol. Liq.* **69**, 1661 (1996).
- [65] N. Nishi *et al.*, *J. Am. Chem. Soc.* **110**, 5246 (1988).
- [66] C. Petrillo, G. Onori, and F. Sacchetti, *Mol. Phys.* **67**, 697 (1989).
- [67] J. Turner and A. K. Soper, *J. Chem. Phys.* **101**, 6116 (1994).
- [68] K. S. Sidhu, J. M. Goodfellow, and J. Z. Turner, *J. Chem. Phys.* **110**, 7943 (1999).
- [69] S. Mashimo, T. Umehara, and H. Redlin, *J. Chem. Phys.* **95**, 6257 (1991).
- [70] J. Z. Bao, M. L. Swicord, and C. C. Davis, *J. Chem. Phys.* **104**, 4441 (1996).
- [71] T. Sato, A. Chiba, and R. Nozaki, *J. Chem. Phys.* **110**, 2508 (1999).
- [72] P. Petong, R. Pottel, and U. Kaatze, *J. Phys. Chem. A* **104**, 7420 (2000).

- [73] F. Müller-Plathe, *Mol. Sim.* **18**, 133 (1996).
- [74] V. N. Levchuk, I. I. Sheykhet, and B. Y. Simkin, *Chem. Phys. Lett.* **185**, 339 (1991).
- [75] J. Fidler and P. M. Rodger, *J. Phys. Chem. B* **103**, 7695 (1999).
- [76] M. Tarek, D. J. Tobias, and M. L. Klein, *Chem. Soc., Faraday Trans.* **92**, 559 (1996).
- [77] M. Tarek, D. J. Tobias, and M. L. Klein, *Physica A* **231**, 117 (1996).
- [78] T. Sugahara, K. Morita, and K. Ohgaki, *Chem. Eng. Science* **55**, 6015 (2000).
- [79] H. Tanaka, *Fluid Phase Equilibria* **144**, 361 (1998).
- [80] B. Kvamme and H. Tanaka, *J. Phys. Chem.* **99**, 7114 (1995).
- [81] J. E. Del Bene, *Chem. Phys. Lett.* **24**, 203 (1974).
- [82] A. Engdahl and B. Nelander, *Chem. Phys. Lett.* **113**, 49 (1985).
- [83] A. Engdahl and B. Nelander, *J. Phys. Chem.* **90**, 4982 (1986).
- [84] A. M. Andrews and R. L. Kuczkowski, *J. Chem. Phys.* **98**, 791 (1993).
- [85] K. I. Peterson and W. Klemperer, *J. Chem. Phys.* **85**, 725 (1986).
- [86] P. Tarakeshwar *et al.*, *J. Chem. Phys.* **111**, 5838 (1999).
- [87] P. Tarakeshwar, K. S. Kim, and B. Brutschy, *J. Chem. Phys.* **112**, 1769 (2000).
- [88] P. Tarakeshwar, H. S. Choi, and K. S. Kim, *J. Am. Chem. Soc.* **123**, 3323 (2001).
- [89] D. B. DuPré and M. C. Yappert, *J. Phys. Chem. A* **106**, 567 (2002).
- [90] K. Laasonen, M. Sprik, M. Parrinello, and R. Car, *J. Chem. Phys.* **99**, 9080 (1993).
- [91] M. Sprik, J. Hutter, and M. Parrinello, *J. Chem. Phys.* **105**, 1142 (1996).
- [92] D. Marx, M. Sprik, and M. Parrinello, *Chem. Phys. Lett.* **273**, 360 (1997).

- [93] T. S. van Erp and E. J. Meijer, Chem. Phys. Lett. **333**, 290 (2001), (by mistake the labels (a) and (b) in figure 1 are reversed).
- [94] S. Raugei and M. L. Klein, J. Chem. Phys. **116**, 196 (2002).
- [95] ADF 2002.01, SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands, <http://www.scm.com>.
- [96] G. te Velde *et al.*, J. Comput. Chem. **22**, 931 (2001).
- [97] C. F. Guerra, J. G. Snijders, G. te Velde, and E. J. Baerends, Theor. Chem. Acc. **99**, 391 (1998).
- [98] Kohn-Sham orbitals are expanded in an even-tempered, all-electron Slater type basis set augmented with 2p and 3d polarization functions for H and 3d and 4f polarization functions for C and O.
- [99] R. K. Kakar and C. R. Quade, J. Chem. Phys. **72**, 4300 (1980).
- [100] P. Borowski, Janowski, and Woliński, Mol. Phys. **98**, 1331 (2000).
- [101] Y. Sasada, M. Takano, and T. Satoh, Mol. Spec. **38**, 33 (1971).
- [102] S. Coussan, Y. Bouteiller, J. P. Perchard, and W. Q. Zheng, J. Phys. Chem. A **102**, 5789 (1998).
- [103] M. L. Senent, Y. G. Smeyers, R. Domínguez-Gómez, and M. Villa, J. Chem. Phys. **112**, 5809 (2000).
- [104] M. Masella and J. P. Flament, J. Chem. Phys. **108**, 7141 (1998).
- [105] K. N. Kirschner and R. J. Woods, J. Phys. Chem. A **105**, 4150 (2001).
- [106] J. W. Handgraaf, T. S. van Erp, and E. J. Meijer, Chem. Phys. Lett. **367**, 617 (2003).
- [107] W. T. M. Mooij, F. B. van Duijneveldt, J. G. C. M. van Duijneveldt-van de Rijdt, and B. P. van Eijck, J. Phys. Chem. A **103**, 9872 (1999).
- [108] J. P. Perchard and M.-L. Josien, J. Chim. Phys. Phys.-Chim. Biol. **105**, 1238 (1986).
- [109] S. W. A. J. Lock and H. J. Bakker, J. Phys. Chem. A. **105**, 1238 (2001).
- [110] N. Marzari and D. Vanderbilt, Phys. Rev. B **56**, 12847 (1997).

- [111] P. L. Silvestrelli, N. Marzari, D. Vanderbilt, and M. Parrinello, *Solid State Commun.* **107**, 7 (1998).
- [112] P. L. Silvestrelli and M. Parrinello, *J. Chem. Phys.* **111**, 3572 (1999).
- [113] P. L. Silvestrelli and M. Parrinello, *Phys. Rev. Lett.* **82**, 3308 (1999).
- [114] A. F. Millidge, in *Ethylene and its Industrial Derivatives*, edited by S. A. Miller (Ernest Benn Ltd, London, 1969).
- [115] C. M. Fougret, M. P. Atkins, and W. F. Hölderich, *Appl. Cat. A* **181**, 145 (1999).
- [116] O. Winter and E. Ming-Teck, *Hydrocarbon Process.* 125 (1976).
- [117] B. T. Baliga and E. Whalley, *Can. J. Chem* **43**, 2453 (1965).
- [118] N. I. Butkovskaya, Y. Zhao, and D. W. Setser, *J. Phys. Chem.* **98**, 10779 (1994).
- [119] N. I. Butkovskaya and D. W. Setser, *J. Chem. Phys.* **105**, 8064 (1996).
- [120] L. Selwyn, R. A. Back, and C. Willis, *Chem. Phys.* **36**, 323 (1978).
- [121] W. C. Danen, *J. Am. Chem. Soc.* **101**, 1188 (1979).
- [122] J. Park, R. S. Zhu, and M. C. Lin, *J. Chem. Phys.* **117**, 3224 (2002).
- [123] T. Yamabe, M. Koizumi, K. Yamashita, and A. Tachibana, *J. Am. Chem. Soc.* **106**, 2255 (1984).
- [124] M. Guelin, in *Laboratory and Astronomical High Resolution Spectra, Series 81*, edited by A. J. Sauval, R. Blomme, and N. Grevesse (Astronomical Society of the Pacific Conference, San Francisco, 1995), p. 438.
- [125] J. E. Del Bene, K. Raghavachari, and J. Pople, *J. Phys. Chem.* **86**, 1529 (1982).
- [126] R. L. DeKock and C. P. Jasperse, *Inorg. Chem.* **22**, 3839 (1983).
- [127] W. L. Jorgensen and J. E. Munroe, *Tetrahedron Lett.* **6**, 581 (1977).
- [128] B. S. Jursic, *J. Mol. Struct. (Theochem)* **487**, 163 (1999).
- [129] M. F. Jarrold, N. J. Kirchner, S. Liu, and M. T. Bowers, *J. Phys. Chem.* **90**, 78 (90).

-
- [130] C. Wesdemiotis, A. Fura, and F. W. McLafferty, *J. Am. Soc. Spectrom.* **2**, 459 (1991).
- [131] K. K. Matthews, N. G. Adams, and N. D. Fisher, *J. Phys. Chem. A* **101**, 2841 (1997).
- [132] D. A. Fairley *et al.*, *J. Phys. Chem. A* **101**, 2848 (97).
- [133] R. S. Mason, *J. Phys. Chem. A* **102**, 10090 (1998).
- [134] G. Bouchoux and Y. Hoppilliard, *J. Am. Chem. Soc.* **112**, 9110 (1990).
- [135] E. L. Purlee and R. W. Taft, *J. Am. Chem. Soc.* **78**, 5807 (1956).
- [136] J. W. de M. Carneiro, P. v. R. Schleyer, W. Koch, and K. Raghavachari, *J. Am. Chem. Soc.* **112**, 4064 (1990).
- [137] K. Kloppe and W. Kutzelnigg, *J. Phys. Chem.* **94**, 5625 (1990).
- [138] V. J. Nowlan and T. T. Tidwell, *Acc. Chem. Res.* **10**, 252 (1977).
- [139] E. J. Meijer and M. Sprik, *J. Phys. Chem. A* **102**, 2893 (1998).
- [140] D. E. Sagnella, K. Laasonen, and M. L. Klein, *Biophys. J.* **71**, 1172 (1996).
- [141] P. L. Geissler *et al.*, *Chem. Phys. Lett.* **321**, 225 (2000).
- [142] P. L. Geissler *et al.*, *science* **291**, 2121 (2001).
- [143] J. J. i Timoneda and J. T. Hynes, *J. Phys. Chem.* **95**, 10431 (1991).
- [144] J. T. Hynes, H. J. Kim, J. R. Mathis, and J. J. i Timoneda, *J. Mol. Liq.* **57**, 53 (1993).
- [145] A. Staib, D. Borgis, and J. T. Hynes, *J. Chem. Phys.* **102**, 2487 (1995).
- [146] K. Ando and J. T. Hynes, *J. Phys. Chem. B* **101**, 10464 (1997).
- [147] K. Ando and J. T. Hynes, *J. Phys. Chem. A* **103**, 10398 (1999).
- [148] P. M. Kiefer and J. T. Hynes, *J. Phys. Chem. A* **106**, 1834 (2002).
- [149] T. S. van Erp and E. J. Meijer, (accepted for publication in JCP).
- [150] B. Ensing, E. J. Meijer, P. E. Blöchl, and E. J. Baerends, *J. Phys. Chem. A* **105**, 3300 (2001).

- [151] C. C. Wu *et al.*, J. Chem. Phys. **112**, 176 (2000).
- [152] J. Heuft and E. J. Meijer, (to be published).
- [153] Y. Nakagawa, N. Tajima, and K. Hirao, J. Comp. Chem. **21**, 1292 (2000).
- [154] H. J. Lucas and Y.-P. Liu, J. Am. Chem. Soc. **56**, 2138 (1934).
- [155] T. G. Solomons, *Organic chemistry* (Wiley, New York, 1976).
- [156] J. Keck, Discuss. Faraday Soc. **33**, 173 (1962).
- [157] J. Keck, Adv. Chem. Phys. **13**, 85 (1967).
- [158] J. Anderson, J. Chem. Phys. **58**, 4684 (1973).
- [159] C. Bennett, in *Algorithms for Chemical Computations, ACS Symposium Series No. 46*, edited by R. Christofferson (American Chemical Society, Washington, D.C., 1977).
- [160] D. Chandler, J. Chem. Phys. **68**, 2959 (1978).
- [161] C. Dellago, P. G. Bolhuis, F. Csajka, and D. Chandler, J. Chem. Phys. **108**, 1964 (1998).
- [162] P. Bolhuis, C. Dellago, and D. Chandler, Faraday Discuss. **110**, 421 (1998).
- [163] C. Dellago, P. Bolhuis, and D. Chandler, J. Chem. Phys. **110**, 6617 (1999).
- [164] P. Bolhuis, D. Chandler, C. Dellago, and P. Geissler, Annu. Rev. Phys. Chem. **53**, 291 (2002).
- [165] C. Dellago, P. Bolhuis, and P. Geissler, Adv. Chem. Phys. **123**, 1 (2002).
- [166] D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford University, New York, 1987).
- [167] M. Strnad *et al.*, J. Chem. Phys. **106**, 3643 (1997).
- [168] Y. Okuno, J. Chem. Phys. **105**, 5817 (1996).
- [169] T. S. van Erp and E. J. Meijer, (to be published).
- [170] Y. Cho, S. R. V. de Linde, L. Zhu, and W. Hase, J. Chem. Phys. **96**, 8275 (1992).

- [171] L. Sun, W. L. Hase, and K. Song, *J. Am. Chem. Soc.* **123**, 5753 (2001).
- [172] J. D. Weeks, D. Chandler, and H. C. Andersen, *J. Chem. Phys.* **54**, 5237 (1971).
- [173] M. J. Ruiz-Montero, D. Frenkel, and J. J. Brey, *Mol. Phys.* **90**, 925 (1997).
- [174] T. S. van Erp, D. Moroni, and P. G. Bolhuis, accepted for publication in *J. Chem. Phys.* .
- [175] B. Ensing and E.-J. Baerends, *J. Phys. Chem. A* **106**, 7902 (2002).
- [176] E. J. Baerends, D. E. Ellis, and P. Ros, *Chem. Phys.* **2**, 42 (1973).
- [177] G. te Velde and E. J. Baerends, *J. Comput. Phys.* **99**, 84 (1992).
- [178] C. F. Guerra, J. G. Snijders, G. te Velde, and E. J. Baerends, *Theor. Chem. Acc.* **99**, 391 (1998).

Samenvatting

Dit proefschrift gaat enerzijds over het bestuderen van chemische reacties via computer simulaties en anderzijds draagt het ook bij aan het verbeteren van deze methodes. Het probleem van computer simulaties is dat deze over het algemeen 'duur' zijn. Hiermee wordt bedoeld dat de berekeningen zeer lange rekestijden in beslag kunnen nemen. Zelfs op supercomputers kan een naïeve implementatie van een scheikundige berekening jaren rekestijd vergen. Dit is zeker het geval voor zogenaamde *ab initio* simulaties, waarbij de krachten op iedere tijdstap via een quantum mechanische berekening worden verkregen. *Ab initio* moleculaire dynamica (MD) onderscheidt zich hiermee van de standaard *force-field* MD simulaties, waarin de krachten op de nucleï of op de moleculen worden berekend via een empirische potentiaal. Dit is gewoonlijk een functie van de relatieve afstanden en oriëntatie tussen atomen en moleculen. Er wordt hierbij vanuit gegaan dat de kracht op een deeltje (een atoom of een molecule) beschreven kan worden als de som van alle paar-interacties tussen dit deeltje en de andere naburige deeltjes in het systeem. Deze *force-fields* worden gewoonlijk gevonden door een standaard-functie te gebruiken met een paar vrije variabelen die dan weer gefit kunnen worden aan experimentele data of aan de uitkomsten van theoretische berekeningen aan kleine moleculaire clusters. Nadat deze empirische potentialen zijn ontwikkeld, worden alle krachten in het systeem volledig gedetermineerd door de posities van de atomen en moleculen. Dit heeft het voordeel dat, als deze empirische potentialen eenmaal bekend zijn, er geen tijdrovende quantum mechanische berekeningen meer hoeven worden uitgevoerd. De aanname dat de kracht afgeleid kan worden uit de onderlinge positie binnen paren van deeltjes, leidt meestal tot een vrij nauwkeurig resultaat. Er zijn echter situaties waarin dit minder het geval is. Zo kan de onderlinge kracht tussen twee watermoleculen sterk beïnvloed worden door de aanwezigheid van een derde watermolecule. Een nauwkeurige beschrijving van water vereist derhalve ook complexere *force-*

fields, zoals *polarizable force-fields* en *three-body potentials* of zelfs *four - body potentials*. *Ab initio* methoden rekenen de krachten direct goed uit doordat ze gebaseerd zijn op een volledige berekening van de elektronenstructuur. Er moet echter wel een prijs voor betaald worden, aangezien *ab initio* MD aanzienlijk duurder is dan *force-field* MD. Echter, als men het proces van een chemische reactie wil bestuderen, is *ab initio* MD bijna altijd noodzakelijk, omdat *force-field* MD hier meestal volledig faalt. Bij het vormen en breken van chemische bindingen vinden er namelijk grote en plotselinge veranderingen plaats in de elektronische structuur. Men kan derhalve onmogelijk a priori voorspellen hoe de krachten tussen ionen veranderen gedurende dit proces. Dit maakt het ontwerpen van accurate empirische potentialen, die ook werken tijdens de daadwerkelijke scheikundige transformatie, een haast onmogelijke taak. Daarnaast is de informatie over de elektronische toestand ook zeer welkom, aangezien het zeer veel inzicht verschaft in het ontstaan van moleculaire bindingen. Tot voor kort was het echter niet goed mogelijk complexe systemen aan te pakken met *ab initio* methoden vanwege te grote computationele kosten. Alleen berekeningen aan reacties met een klein aantal moleculen in de gasfase behoorden daardoor tot de mogelijkheden. De meeste chemische reacties, in de industrie, in de natuur en in laboratoriumexperimenten, vinden echter niet plaats in de gasfase, maar in een oplosmiddel, zoals bijvoorbeeld water. Het is een bekend gegeven dat chemische reacties geheel anders kunnen verlopen in oplossing dan in de gasfase, maar voor een goede beschrijving van bijvoorbeeld vloeibaar water zijn op zijn minst zo'n twintigtal watermoleculen vereist, hetgeen een zeer zware berekening vergt. In 1985 kwam echter een doorbraak. Car en Parrinello bedachten een nieuw algoritme dat deze *ab initio* berekeningen vele malen deed versnellen. Dit bracht de mogelijkheid om veel grotere systemen te bestuderen, waarin veel meer deeltjes interageren. De huidige techniek is in staat om de dynamica van systemen met ongeveer 30 watermoleculen te simuleren over een tijdspanne van zo'n tien pico ($=10^{-12}$) seconden. Dit is net genoeg om chemische reacties in water goed te kunnen bestuderen.

De hoofdvraag van het proefschrift was dan ook het onderzoeken van het effect van water als oplosmiddel op het chemische proces en in het bijzonder op chemische processen met alcoholen. Water heeft een zeer unieke structuur doordat het waterstofbruggen kan vormen. Het verkrijgt daardoor een structuur die sterk doet denken aan een kristal. Het is een soort van tetraëdrische netwerk waarbij ieder watermolecule gemiddeld met vier ander watermoleculen verbonden is via een waterstofbrug. Alcohol heeft

dezelfde eigenschap als water, aangezien het een hydroxylgroep heeft, waardoor het ook in staat is om waterstofbruggen te vormen. Dit maakt dat de kleine alcoholen uitstekend oplosbaar zijn in water. Naast de hydroxylgroep bestaan alcoholen echter nog uit een zogenaamde alkylgroep, die minder interactie heeft met water en daarom ook wel *hydrophoob* genoemd wordt. Door deze tegenstrijdige haat-liefde verhouding tussen alcohol en water, hebben mengsels tussen alcoholen en water een grote wetenschappelijke belangstelling verkregen. Naar gelang de alcoholconcentratie en het type alcohol, bestaat er een grote verscheidenheid aan moleculaire structuren. In ons onderzoek hebben we gekeken naar zeer verdunde oplossingen van alcoholen, waarbij een enkel alcohol wordt omgeven door een zogenaamde oplossschil van watermoleculen. De structuur van de eerste oplossschil van watermoleculen rondom het alcohol is van cruciaal belang voor het bestuderen van chemische reacties. Het initiëren van een reactie vereist namelijk vaak een gehele herordening van de watermoleculen in deze oplossschil.

In hoofdstuk 2 hebben we de structuur en de dynamica van deze oplos-eigenschappen voor het kleinste alcohol, het methanol molecule, bestudeerd. Uit onze bevindingen blijkt dat het methanol molecule niet in staat is om de typische waterstructuur te veranderen. De hydroxylgroep van het methanol gedraagt zich niet veel anders dan een watermolecule, doordat het waterstofbruggen kan maken met de naburige watermoleculen. De methylgroep blijkt klein genoeg en wordt gemakkelijk opgenomen in de open ruimtes van dit waterstofbrug-netwerk.

Hoofdstuk 3 gaat over de oplossingsstructuur van ethanol en ethyleen. Aangezien de hydratatie-reactie van ethyleen naar ethanol een belangrijk industrieel proces is, kan dit hoofdstuk ook gezien worden als een belangrijke voorstudie voor deze reactie. Al zijn ethanol en ethyleen een stuk groter dan methanol, ook hier lijkt de waterstructuur nauwelijks te veranderen. Een belangrijke conclusie is derhalve dat de tetraëdrische structuur heel flexibel is en dat het water *hydrophobe* groepen gemakkelijk opneemt zonder sterk van structuur te veranderen. Daarbij blijkt water een sterk polariserende invloed te hebben op de oplosmoleculen. Zelfs het neutrale ethyleen verkrijgt een aanzienlijk dipool moment, met zeldzame uitschieters van maar liefst 1 Debye. Deze zeldzame uitschieters worden veroorzaakt door structuren waarbij meerdere watermoleculen samenwerken met als gevolg een zeer sterke polariserende invloed op het ethyleenmolecule.

Hoofdstuk 4 gaat vervolgens over hydratatie/dehydratatie-reactie tussen ethyleen en ethanol. We gebruiken hier de methode van geconstraineerde dynamica om de reactie te forceren. Hierdoor wordt het systeem als het ware

stapje voor stapje over de reactie- barrière heen getild, terwijl ondertussen de hoogte van deze barrière gemeten wordt. Een nadeel van deze methode is dat de informatie over het spontane reactie-verloop gedeeltelijk verloren gaat. Aangezien echter slechts een stukje van de reactie wordt geforceerd en sommige delen van de reactie spontaan verlopen, kan er toch veel over het reactieproces geconcludeerd worden. Zo vinden wij dat het reactieproces via het $\text{Ad}_{\text{E}3}$ mechanisme verloopt, waarbij twee bindingen tegelijkertijd gevormd worden. De precieze vorming van een binding is ook niet altijd even duidelijk. Vaak wordt er een binding toegekend louter op basis van de afstand tussen twee atomen. We laten hier zien dat dit niet altijd opgaat en dat er structuren zijn van dicht op elkaar gesitueerde atomen waarbij er geen sprake is van een binding, terwijl tussen twee verder van elkaar verwijderde atomen juist wel een binding blijkt te zijn. Specifieke analyse van de elektronenstructuur door middel van zogenaamde Wannier analyse laat dit zien. Bovendien blijken de gevonden waterstructuren, zoals ze zijn gevonden in hoofdstuk 3, ook hier een zeer belangrijke invloed te hebben op de spontane chemische processen die niet door onze constraint worden geforceerd.

Hoofdstuk 5 was eigenlijk een niet gepland stuk onderzoek en is dus eigenlijk een typisch geval van spin-off. Omdat de constraint methode de reactie toch op een onnatuurlijke wijze forceert, wilde ik de *transition path sampling* methode toepassen, die meer natuurlijke reactie-paden genereert. Tijdens de bestudering van de methode kwam ik echter op nieuwe ideeën, ter verbetering van deze methode. Dit hebben we vervolgens uitgewerkt tot een geheel nieuwe methode, die we tot *transition interface sampling* hebben gedoopt. Deze methode blijkt, afhankelijk van het te bestuderen systeem, enkele malen sneller te zijn dan het originele *transition path sampling* algoritme.

Tenslotte laat hoofdstuk 6 de resultaten zien waarbij *ab initio* MD met *transition interface sampling* wordt gecombineerd om de directe (zonder katalysator) gasfase-hydratatie van ethyleen te bestuderen. De meeste berekeningen aan deze reactie gaan er vanuit dat *transition state theory* een goede benadering voor deze reactie is. Dit houdt in dat er vanuit gegaan wordt dat bijna alle paden over het laagste punt van de barrière gaan, het zogenaamde zadelpunt of de transitie-toestand. De reactie-paden die ik met de *path sampling* methode heb gegenereerd zijn echter zeer verschillend van elkaar, laten een zeer grillig karakter zien en gaan zelden via de transitie-toestand naar de andere kant van de barrière. Hieruit volgt dat de in het verleden berekende reactieconstanten voor dit systeem, die gebruik maak-

ten van deze transitie-toestand aanname, zeer twijfelachtig zijn.

Concluderend laat dit proefschrift zien dat de invloed van het oplosmiddel op de chemische reactie zeer belangrijk is en dat zeldzame structuren, van waterstofbrug-gebonden watermoleculen, een sterk polariserende invloed kunnen hebben, hetgeen een reactie sterk kan beïnvloeden. De nieuwe methode van *transition interface sampling* verbetert de oude methode van *transition path sampling* aanzienlijk en kan helpen om deze reacties verder te onderzoeken.

*eh...Sorry, ...maybe I missed it,
...but what does this contribute
to the unification theory?*

Publications

- [1] T.S. van Erp, A. Fasolino, O. Radulescu and T. Janssen, "Pinning and phonon localization in Frenkel-Kontorova models on quasiperiodic substrates", *Physical Review B* **60**, 6522-6528, (1999).
- [2] Titus S. van Erp, Evert Jan Meijer, "Hydration of methanol in water. A DFT-based molecular dynamics study", *Chemical Physics Letters* **333**, 290-296, (2001).
- [3] T.S. van Erp, A. Fasolino and T. Janssen, "Structural Transitions and Phonon Localization in Frenkel Kontorova Models with Quasi-Periodic Potentials", *Ferroelectrics* **250**, 421-424, (2001).
- [4] T.S. van Erp and A. Fasolino, "Aubry transition studied by direct evaluation of the modulation functions of infinite incommensurate systems", *Europhysics Letters* **59**, 330-336 (2002).
- [5] Jan-Willem Handgraaf, Titus S. van Erp, and Evert Jan Meijer. "Ab Initio molecular dynamics study of liquid methanol", *Chemical Physics Letters* **367**, 617-624, (2003).
- [6] Titus S. van Erp, Daniele Moroni, and Peter G. Bolhuis, "A Novel Path-sampling Method for the Calculation of Rate Constants", accepted for publication in *Journal of Chemical Physics*.
- [7] Titus S. van Erp, Evert Jan Meijer, "Ab Initio Molecular Dynamics Study of Aqueous Solvation of Ethanol and Ethylene", accepted for publication in *Journal of Chemical Physics*.
- [8] Titus S. van Erp, Evert Jan Meijer, "Ethene Hydration in Acid Solution. Revealing the Solvation Structure of Reacting Pathways", in progress.
- [9] Titus S. van Erp, Evert Jan Meijer, "Ab Initio Study: Hydration / Dehydration Reaction between Ethanol and Ethene", in progress.

- [10] Titus S. van Erp, Evert Jan Meijer, "Non-Catalyzed Hydration of Ethylene. A Combination of Born-Oppenheimer Dynamics and Transition Interface Sampling", in progress.

Curriculum Vitae

Titus van Erp was born in Tilburg, the Netherlands, on September 22, 1974. After finishing elementary school and high school, he started his physics study in 1993 in Nijmegen. After some trivial stuff such as finishing exams, he finally could start doing some real research for the project of the last year of his study at theoretical physics group under supervision of Annalisa Fasolino and Ted Janssen. There he learned about solid state physics, non-linearity, chaos theory, incommensurability, quasi-crystals and more. His research was specified to an extension of a widely studied model, the Frenkel-Kontorova model, that has inside all the characteristics mentioned above. In 1999 Titus completed his study by finishing his master thesis with the title: *"Frenkel-Kontorova models on quasiperiodic substrate potentials"* and became a Drs¹. Directly thereafter, he started his PhD at the department of chemical engineering of the university of Amsterdam (UvA) under supervision of Evert Jan Meijer and Berend Smit. Of course, the subject: reactions with **alcohols** in water' was directly appealing to him. Besides the practice, Titus learned the use of highly advanced molecular modeling programs, such as Car-Parrinello Molecular Dynamics. However, as a side project not related to this PhD work, Titus kept on working on nonlinear science, resulting in an important discovery that shows that a twenty year old hypothesis concerning the standard Frenkel Kontorova model is probably wrong. For his PhD research, a thorough study to the hydration of ethylene in water, a process of importance for the industrial synthesis of ethanol, was the final major achievement. The awareness that rare but spontaneous occurring subtle structures between water molecules, preferably well ordered tetrahedral structures, are important for these kind of reactions was the most important result of that project. Unexpectedly, Titus was also fortunate to be able to contribute to the methodology of molecular modeling. Studying the method

¹This is a typical Dutch title, abbreviated from, "Doctorandus", which is Latin for "He who has to become a Doctor." So, unfortunately this does not mean multiple-Doctor, a common mistake made by people of foreign countries.

of Transition Path Sampling with only the intention to apply it, Titus found ways to simplify the method, gaining besides an increase of efficiency a lot of other advantages. This method is now christened 'Transition Interface Sampling' and could be of great importance for future research in theoretical chemistry. The 25th of April Titus hopes to lose his 's' and to turn from a Drs. into a Dr. Meanwhile, Titus already started a post-doc. After moving from Tilburg to Nijmegen and from Nijmegen to Amsterdam it was time to go abroad, something almost inevitable when continuing in science. Hosted at the Centre Européen de Calcul Atomique et Moléculaire (CECAM), he will work with Michel Peyrard of the Ecole Normale Supérieure (ENS) in Lyon, France. There he will be able to use his knowledge of both nonlinear dynamics and computational chemistry. In particular, he will study the occurrence and the importance of some strange nonlinear phenomena, called discrete breathers, for surface catalysis... mais pour maintenant ça sera tout.