



# Multi-scale computational modelling of bio-chemical systems in the condensed phase

Raphael Peltzer

Dissertation for the degree of Philosophiae Doctor

Department of Chemistry

Faculty of Mathematics and Natural Sciences

University of Oslo

Oslo, September 2018

# Contents

I.	Acknowledgements . . . . .	I
II.	List of Papers and Manuscripts . . . . .	II
III.	List of abbreviations . . . . .	III
IV.	Abstract . . . . .	V
<b>1.</b>	<b>Background and Introduction</b>	<b>1</b>
1.1.	Macroscopic vs microscopic representation . . . . .	2
1.2.	Computational estimation methods . . . . .	4
1.2.1.	The Monte Carlo Method . . . . .	4
1.2.2.	Molecular Dynamics . . . . .	5
1.2.3.	Molecular Mechanics models . . . . .	9
1.3.	Current state of the art in simulations . . . . .	10
<b>2.</b>	<b>Methods</b>	<b>15</b>
2.1.	Quantum Mechanical Methods . . . . .	15
2.1.1.	Wave function based methods . . . . .	16
2.1.2.	Density Functional Theory . . . . .	21
2.1.3.	Natural Bond Order Analysis . . . . .	26
2.2.	Basis Sets and Pseudopotentials . . . . .	26
2.3.	Monte Carlo Simulations . . . . .	29
2.3.1.	Metropolis algorithm . . . . .	29
2.4.	Molecular Dynamics Simulations . . . . .	31
2.4.1.	Born-Oppenheimer Molecular Dynamics . . . . .	31
2.4.2.	Hellman-Feynman Theorem . . . . .	34
2.5.	Thermostats . . . . .	42
2.5.1.	CSVR thermostat . . . . .	42
2.5.2.	Nosé-Hoover chain thermostat . . . . .	42

2.6.	Periodic Boundary Conditions . . . . .	44
2.7.	Enhanced Sampling Methods . . . . .	45
2.7.1.	Constrained Molecular Dynamics . . . . .	45
2.7.2.	Metadynamics . . . . .	46
2.7.3.	Simulated Annealing . . . . .	49
<b>3.</b>	<b>The Schlenk Equilibrium of the Grignard Reagents</b>	<b>51</b>
3.1.	Computational Background . . . . .	53
3.1.1.	Computational Details and System Setup . . . . .	54
3.1.2.	System Setup and Simulation Details . . . . .	55
3.2.	Results . . . . .	60
3.3.	The Schlenk equilibrium . . . . .	62
3.4.	Conclusion . . . . .	78
<b>4.</b>	<b>Direct Nucleophilic Attack Vs. Radical Propagation Mechanism in the Grignard Reaction</b>	<b>80</b>
4.1.	Introduction . . . . .	80
4.2.	Computational methods . . . . .	83
4.2.1.	Homolytic cleavage energies . . . . .	83
4.2.2.	Ab initio molecular dynamics . . . . .	85
4.3.	Results and Discussion . . . . .	87
4.3.1.	Homolytic bond cleavage . . . . .	87
4.3.2.	Nucleophilic attack . . . . .	90
4.3.3.	Nucleophilic attack vs Homolytic cleavage . . . . .	93
4.4.	Concluding remarks . . . . .	95
<b>5.</b>	<b>Modelling the self-assembly of alpha-tocopherol</b>	<b>96</b>
5.1.	Introduction . . . . .	96
5.2.	Patchy Particle Models . . . . .	99
5.3.	Computational Methods . . . . .	99
5.3.1.	System Setup . . . . .	101
5.4.	The AB-Model . . . . .	103
5.5.	The ABCD-Model . . . . .	105
5.6.	Discussion . . . . .	110

## Contents

---

<b>6. Conclusion and Future Outlook</b>	<b>116</b>
<b>7. Appendix</b>	<b>118</b>
7.1. Appendix A - Metadynamics parameters for Grignard simulations . . . . .	118
7.2. Appendix B -NBO Analysis for the Grignard simulations	121
7.3. Appendix C - Truncation of pairwise interactions . . . . .	123
7.4. Appendix D - Molecular Mechanics potentials . . . . .	124

## I. Acknowledgements

Many people have been contributed in their own way and without them I would not have been able to finish this thesis. I would like to thank those people most involved.

Foremost, I would like to thank my main supervisor Michele Cascella. I would like to thank him especially for his commitment to making my PhD as fruitful as possible, always taking the time to explain different theoretical concepts again in detail, mentoring me on improving my presentation skills, always encouraging me and giving me the self-confidence to present myself and my research proudly. It has been a privilege to be able to work with him for the last four years.

For the last year I had the privilege and joy of working with Hima Bindu Kolli, I would like to thank her for the delightful discussions and for introducing me the implementation of the Metropolis Monte Carlo Algorithm.

I would like to thank Achim Stocker at the University of Bern for his facilitating the research and his contributions on our study of self-assembly of  $\alpha$ -tocopherol transfer protein nanoparticles. I would like to thank Ainara Nova for her work on our manuscript on the Schlenk-equilibrium and especially for teaching me how to use NBO analysis and the Gaussian software package. I would like to thank Odile Eistenstein for initiating our work on the Gringard reagent and the Schlenk equilibrium. Her enthusiasm and the great discussions we had have made working together, especially on the pain-staking part of writing my first scientific paper, just the more enjoyable. I would like to thank Juergen Gauss for our collaboration on the investigation of the Grignard reaction and for hosting me in his group at the University of Mainz as part of our collaboration.

I would like to thank Elisa Rebolini and Lluis Artus Suarez for the countless discussions we have had on the work I did during my PhD studies and never hesitating to take time out of their schedules time

and time again to help me out on my projects. I would also like to thank the people of the CTTC and Hylleraas center at the University of Oslo for making my time in the group very enjoyable. At last I would like to thank my family, especially my parents and siblings, and my lovely fiancée Marianne Solhaug that have encouraged me throughout my degree.

## II. List of Papers and Manuscripts

This thesis is based on the listed papers:

- I Peltzer, R. M.; Eisenstein, O.; Nova, A.; Cascella, M. How Solvent Dynamics Controls the Schlenk Equilibrium of Grignard Reagents: A Computational Study of CH<sub>3</sub>MgCl in Tetrahydrofuran. *The Journal of Physical Chemistry B* **2017**, 121, 4226–4237,
- II Peltzer, R. M.; Kolli, H. B.; Stocker, A.; Cascella, M. Self-Assembly of -Tocopherol Transfer Protein Nanoparticles: A Patchy Protein Model. *The Journal of Physical Chemistry B* **2018**, 122, 7066–7072, PMID: 29944374.
- III Peltzer, R. M.; Eisenstein, O.; Gauss, J.; Cascella, M. Direct Nucleophilic Attack Vs. Radical Propagation Mechanism in the Grignard Reaction – a Computational Study *in preparation*

### III. List of abbreviations

**AA** All-atom

**AIMD** Ab-initio Molecular Dynamics

**$\alpha$ -tol**  $\alpha$ -tocopherol

**$\alpha$ -TTP**  $\alpha$ -tocopherol transfer protein

**BO** Born Oppenheimer

**CC** Coupled Cluster

**CG** Coarse Graining

**CI** Configuration Interaction

**CMD** Constrained Molecular Dynamics

**CPMD** Car-Parinello Molecular Dynamics

**CSVR** Canonical Sampling/Velocity Rescaling

**CV** Collective Variable

**DFT** Density Functional Theory

**DZVP** double- $\zeta$  valence polarized

**EOM** Equation of Motion

**FES** Free Energy Surface

**GGA** Generalized Gradient Approximation

**GTO** Gaussian Type Orbital

**HF theorem** Hellman Feynman Theorem

**HF** Hartree-Fock

**IS** Interaction Sites

**KS** Kohn Sham Theory

## Contents

---

- LDA** Local Density Approximation
- MA** Metropolis Algorithm
- MC** Monte Carlo
- MCMC** Markov Chain Monte Carlo
- MD** Molecular Dynaics
- MM** Molecular Mechanics
- MMC** Metropolis Monte Carlo
- MP2** second-order Møller-Plesset perturbation theory
- NBO** Natural Bond Orbital
- NLMO** Natural Localized Molecular Orbital
- NPT** constant-Pressure,constant-Temperature
- NVE** constant-Energy,constant-Volume
- NVT** constant-Volume,constant-Temperature
- PBC** Periodic boundary conditions
- PBE** Perdew-Burke-Ernzerhof
- PES** Potential Energy Surface
- PP model** Patchy particle model
- QM** Quantum Mechanics
- SA** Simulated Annealing
- SMD** Solvent Model based on Density
- STO** Slater Type Orbital
- THF** Tetrahydrofuran

## IV. Abstract

The number of bio-chemical systems and processes that are able to be modeled and understood using computational chemistry tools is ever growing and the boundaries of feasible simulations are continuously extended. This is made possible by a tandem of steadily increasing computer power and the development of new methodologies and their ever more efficient implementations in commonly available software packages. In this thesis, I employ this tandem to deepen our understandings on two processes that happen on very different size and length scales, the Grignard reaction and its associated Schlenk-Equilibrium as well as the self-assembly of a protein nanoparticle.

In chapter 1 and 2 of this thesis I revisit the fundamental, theoretical advances made in computational chemistry that allow us to model a broad range of applications and give an overview over what kind of systems one can expect to be able to model using the state-of-the-art computational chemistry as well as their associated methodology.

In Chapter 3 I investigate the Schlenk-Equilibrium, a complex reaction that governs the presence of multiple chemical species in solution of the Grignard reagents, and that has been known since the 1920s, but has not been fully characterized to this day due to its complexity. Using ab initio molecular dynamics simulations with enhanced-sampling metadynamics it was possible to determine the reaction pathway of the Schlenk-equilibrium reaction of the Grignard reagent  $\text{CH}_3\text{MgCl}$  in explicit tetrahydrofuran solvent and to characterize more precisely the chemical species at equilibrium. Furthermore, I found that the propagation of the reaction may only occur via specific solvation state of each of the different aggregates. Describing the solvent dynamics explicitly is thus key to fully understanding the mechanism of the Schlenk-Equilibrium reaction.

In Chapter 4 I investigate the competition of the radical and the nucleophilic addition pathway of the Grignard. Radical formation energies for complexes with different substrates were estimated by density

## Contents

---

functional theory after benchmarking over coupled cluster data. The nucleophilic attack mechanism was investigated by ab initio molecular dynamics in explicit tetrahydrofuran solvent. I find that the initial step of the radical reaction involves binding of the substrate to the magnesium and the subsequent release of a radical organic group. Regardless of the nature of the substrate and the solvation of the magnesium, the unpaired electron localizes on the substrate and the reaction is promoted by stabilisation of the radical, while nucleophilic addition is promoted by the fraction of the dialkyl-magnesium species produced by the Schlenk equilibrium. As a consequence, the nucleophilic addition is favored for alkyl-carbonyl species, while the radical mechanism becomes energetically favorable for extended aromatic moieties. This finding might aid the design of future metallocorganic catalysts in the homogeneous phase.

In Chapter 5 I study the self assembly of  $\alpha$ -tocopherol transfer protein, which responsible for the retention of Vitamin E in the human body, into nanoparticles. Using Monte Carlo simulations on a patchy particle model it was possible to determine the kinetics and thermodynamics of the self aggregation this protein. The oligomerization kinetics proceeds in two steps with the formation of metastable trimeric unit, and the subsequent assembly into spherical constructs. I was able to determine a direct relationship between the stability of the protein-protein interfaces formed during the oligomerisation process, the temperature of the system and the predominant oligomerisation state found in experimental studies. From this observation, I was able to establish predictions on the influence of specific protein modifications on the self-aggregation state of the protein.

# 1. Background and Introduction

Modern Computational Chemistry goes hand in hand with experiments and is able to give insight into processes that are more accurately or more efficiently described using computer simulations. Such simulations have for example been successfully applied to modeling the folding of proteins<sup>[1]</sup>, designing new materials<sup>[2]</sup> and predicting protein structures<sup>[3]</sup> as well as reactivity trends<sup>[4]</sup>.

Despite the fact that computational chemistry models have proven excellent tools to explore and predict fundamental processes, the majority of the experiments, especially in the condensed phase, are carried out on much larger scales than those accessible to computational chemistry.

To gap between experimental and computational dimensionalities can be understood by a simple Gedankenexperiment: Thinking of an simple solution containing merely 10 ml of water, the number of atoms that would have to be simulated is on the order of avogadro's number ( $10^{24}$ ). If we were able to describe a system by the position and velocity of each atom stored to 7 decimal places, we would require 6 bytes of memory storage per atom. For a single simulation frame we would thus require 6 yotabyte ( $10^{24}$  byte) of memory storage, which exceeds by orders of magnitude the data storage capability of the whole planet.

In a computational approach, one does not try to reproduce the whole experiment, but rather to build a reliable model including all the relevant microscopic features that give rise to the macroscopic properties of the system.

In this introduction I will address the following fundamental issues:  
What is the link between a macroscopic measurement and the microscopic behavior of the system?

What kind of experimental systems can be described using state-of-the-art computational chemistry methods?

### 1.1. Macroscopic vs microscopic representation

Since atomic theory became widely accepted in the 19th century, it has been known that our world is made up of atomistic particles rather than the continuum we experience. Later, through the formulation of statistical mechanics, it became evident that macroscopic observations (such as temperature and pressure) are related to the microscopic state of the system subject to microscopic variables (particle positions and momenta). The microscopic state is not steady, but it is rather constituted by an ensemble of states that continuously interconvert into each other. The introduction of statistical mechanics thus connected macroscopic quantities (such as heat capacity) to the microscopic behavior of the system. Experimental measurements are, in general,<sup>1</sup> such macroscopic observations taken over a large number of molecules over a finite time. Because the time scales of experimental measurement are much larger than the time spent in each microstate, it can be assumed that the system visits all accessible microstates during the experiment.

The exact relation between the measured, macroscopic observable and its microscopic representation is that the macroscopic observable  $\langle \mathcal{O} \rangle$  is the average obtained from weighing the microscopic observable value  $\mathcal{O}$  for each microstate with the probability  $\mathcal{P}$  that the same microstate is visited during the measurement:

---

<sup>1</sup>excluding extremely fast or extremely probed measurements

$$\langle \mathcal{O} \rangle = \int \mathcal{O}(\{\mathbf{r}\}, \{\mathbf{p}\}) \mathcal{P}(\{\mathbf{r}\}, \{\mathbf{p}\}) d\Gamma \quad (1.1.1)$$

with  $d\Gamma$  indicating integration over all microstates described by positions  $\mathbf{r}$  and momenta  $\mathbf{p}$ .  $\mathcal{P}$  is determined by the specific thermodynamics equilibrium the system is subject to. For example, for a system at thermal equilibrium with an external bath,  $\mathcal{P}$  is given by the Boltzmann distribution of states. For a system of states  $\alpha$  with energy levels  $E_\alpha$  at a given temperature  $T$  the probability of each state is given by:

$$P_\alpha = \frac{1}{Z} e^{-\beta E_\alpha}, \quad (1.1.2)$$

where  $\beta = \frac{1}{k_B T}$ ,  $k_B$  is the Boltzmann's constant, and  $Z$  is the partition function of the system defined as:

$$Z = \frac{1}{N!} \int e^{-\beta H(\{\mathbf{r}\}, \{\mathbf{p}\})} d\Gamma. \quad (1.1.3)$$

This integral in equation 1.1.1 implies a complicated mathematical integral of  $6N$  dimensions, where  $N$  is the number of particles in the system, which cannot be solved analytically. Due to the large dimensionality of the integral, simple direct numerical integration methods are also not feasible. For example, imagining numerical integration over a regular grid containing  $n$  mesh points per dimension, it would be necessary to estimate the value of the integrand function over a total of  $n^{6N}$  grid points.

In fact, the Boltzmann weight (equation 1.1.2) guarantees that only the states with minimal energy have a non-negligible weight, while all the others become exponentially marginal. It is therefore possible to exploit this condition to introduce numerical methods able to efficiently sample only the relevant microstates that contribute the macroscopic behavior of the system.

## 1.2. Computational estimation methods

Different computational tools have been developed to obtain numerical approximations to the integral in equation 1.1.1. The two most broadly applied approaches are Monte Carlo (MC) and Molecular Dynamics (MD) simulations. In the following I will in turn introduce the theory behind both approaches.

### 1.2.1. The Monte Carlo Method

The main idea of MC is to exploit the knowledge of the relative probability of each state according to the underlying statistical ensemble. The sampling of the microscopic states is achieved by making random changes to the microscopic variables of the system. The newly obtained states are weighted through a probability algorithm such that the final accumulation of microstates represents the target probability distribution. The algorithm is explained in detail in Section 2.3.

For example, under canonical conditions (constant volume, temperature and number of particles) the relative probability of two microstates of energies  $E_1$  and  $E_2$  becomes:

$$P_{rel} = e^{-\beta(E_1 - E_2)} \quad (1.2.1)$$

In MC simulations the probability to sample any microstate is thus linked to its energy. This implied that the algorithm needs the estimation of the total energy of the system at any step of the simulation to proceed.

### 1.2.2. Molecular Dynamics

Another approach to generating a probabilistically correct representation of the microstates that give rise to the macroscopic observables is founded on the Birkhoff's ergodic theorem, which states

$$\langle \mathcal{O} \rangle = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^\tau \mathcal{O}(t) dt \quad (1.2.2)$$

Thus by simulating the time evolution of the system one can similarly obtain the desired observable.

The computational tool based on this approach is Molecular Dynamics (MD). By probing the time-evolution of the system that give rise to the probability distribution of microstates, MD simulations provide not only access to the macroscopic observables of equilibrium systems, but also an in-depth dynamics description of microscopic events.

For a mechanical systems obeying the laws of classical mechanics, the time evolution of the system is determined by Newton's equation of motion:

$$\mathbf{F} = m\mathbf{a} \quad (1.2.3)$$

where  $\mathbf{F}$  is the force,  $m$  the particle mass and  $\mathbf{a}$  is the acceleration.

The forces  $F(\mathbf{R})$  acting on the particle at coordinates  $\mathbf{R}$  is related to the energy  $E_{el}(\mathbf{R})$  by

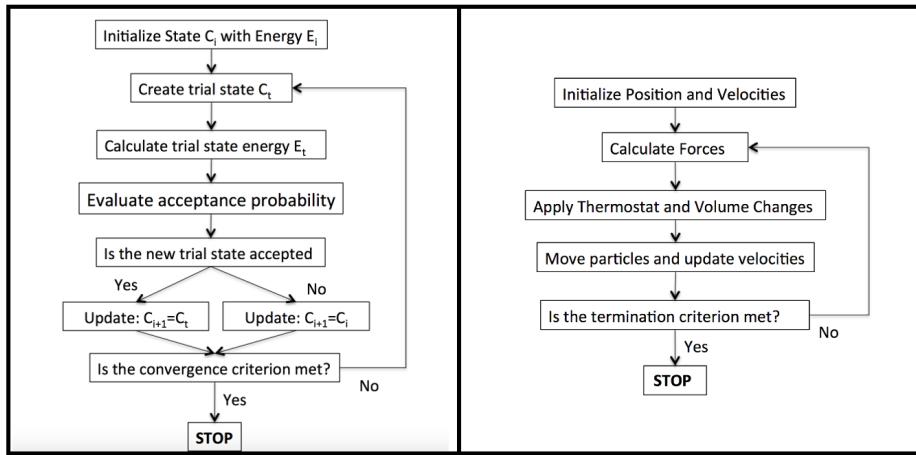
$$F(\mathbf{R}) = -\nabla E_{el}(\mathbf{R}) \quad (1.2.4)$$

For a system of  $N$  atoms, solving the Newton's EOMs imply solving a system of  $3N$  differential equations that require numerical integration through iterative algorithms. (Introduced in detail in Section 2.4)

In summary the MC and MD algorithms follow the flowcharts depicted below.

## BACKGROUND AND INTRODUCTION

---



**Figure 1.1.** – Flowcharts describing the basic algorithm for MC (left) and MD (right) simulations.

Both algorithms require the calculation of either the energy (MC) or the gradient of the energy (MD) in every iteration. Because, the number of iterations required is generally extremely large, these steps constitute the main expense of the respective computational methods.

**Evaluation of the Energy for a Molecular System** At a fundamental level any molecular system is described by quantum mechanics (QM). Accordingly, the energy of any microstate is obtained by solving the time-independent Schrödinger equation.

$$\hat{H}\Psi(\mathbf{R}, \mathbf{r}) = E\Psi(\mathbf{R}, \mathbf{r}) \quad (1.2.5)$$

where  $\hat{H}$  is the Hamiltonian operator acting on  $\Psi(\mathbf{R}, \mathbf{r})$  is the wave function representing the state of the system dependent on the nuclear coordinates  $R$  and electron coordinates  $r$ . The probability amplitude of finding the system in its corresponding state is the square modules of the wave function.

For a molecular system the Hamiltonian operator has the form:

$$\hat{H} = \hat{T}_N + \hat{V}_{ee} + \hat{V}_{Ne} + \hat{V}_{NN} = \hat{T}_N + \hat{H}_{el} \quad (1.2.6)$$

where  $T$  denotes kinetic and  $V$  potential energy terms, with  $N$  and  $e$  indicating nuclei and electrons, respectively.  $\hat{H}_{el}$  takes into account all contributions to the Hamiltonian operator except for the kinetic energy of the nuclei. The Schrödinger equation is a  $3M$ -dimensional eigenvalue problem ( $M$  being the total number of electrons and nuclei in the system) for a second order partial differential equation that is not solvable analytically for more than two particles. Therefore, this problem has to be addressed by numerical approaches.

A typical approximation to the above eigenvalue problem is based on separating adiabatically the motion of the nuclei from that of the electrons. This assumption is based on the fact that electrons are much lighter particles and therefore can respond much more rapidly to the motion of the heavier nuclei. The Born-Oppenheimer (BO) approximation, based of this assumption, allows for the factorization of the wavefunction in two, electronic and nuclear, components:

$$\Psi(\mathbf{R}, \mathbf{r}) = \Phi(\mathbf{r}; R)\chi(\mathbf{R}) \quad (1.2.7)$$

where  $\chi(\mathbf{R})$  describes the behavior of the nuclei, and  $\Phi(\mathbf{r}; R)$ , which depends only parametrically on the nuclear coordinates, describes the electronic state of the system.  $\Phi(\mathbf{r}; R)$  is obtained by solving the Schrödinger equation for fixed nuclei:

$$\hat{H}_{el}\Phi(\mathbf{r}; \mathbf{R}) = E_{el}(\mathbf{R})\Phi(\mathbf{r}; \mathbf{R}) \quad (1.2.8)$$

while the nuclei move on a potential energy surface defined by the electronic eigenvalue  $E_{el}(\mathbf{R})$ :

$$[\hat{T}_N + E_{el}(\mathbf{R})]\chi(\mathbf{R}) = E\chi(\mathbf{R}) \quad (1.2.9)$$

I will cover the BO approximation in more detail in Section 2.4.1.

**Evaluating the energy gradients** Approximating the motion of the nuclei to that of classical particles, the determination of a MD trajectory implies the calculation of the forces acting on the nuclei in the system. The QM force  $F(\mathbf{R})$  acting on a nucleus at a coordinate  $\mathbf{R}$  is obtained from the energy  $E_{el}(\mathbf{R})$  by the gradient operation:

$$F(\mathbf{R}) = -\nabla E_{el}(\mathbf{R}) \quad (1.2.10)$$

Which can be further simplified as

$$\nabla E_{el}(\mathbf{R}) = \left\langle \Phi(\mathbf{R}) \left| \nabla \hat{H}_{el}(\mathbf{R})(\lambda) \right| \Phi(\mathbf{R}) \right\rangle \quad (1.2.11)$$

thanks to the Hellman-Feynman theorem (details in next chapter).

As evident from the Hellman-Feynman theorem, when treating the system at a QM level of theory, it is necessary to compute the wave function and the electronic energy of the system at every iteration step. Because it is necessary to iterate this procedure over a large number of steps in order to obtain a reliable representation of the relevant microstates, this step becomes the computationally most crucial step and we must evaluate the many-electron problem in a computationally efficient way.

**Approximations to the evaluation of the wave function** Different theories and approximations have been developed to deal with the above many-electron problem. In this thesis I employed Density Functional Theory (DFT)<sup>[5,6]</sup> and Coupled Cluster Theory (CC)<sup>[7,8]</sup>, which will be presented in detail in the next chapter. CC is an excellent tool for high accuracy QM calculations. The CCSD(T) method is often regarded as the gold standard in quantum chemistry. However the CCSD(T) calculation costs scale as  $N^7$  with  $N$  the number of electrons in the system. This makes CC approaches not feasible for large molecular systems or for iterative procedures like MC or MD.

DFT is the most popular approach to solve the many-body problem of the Schrödinger equation in MD. The main idea of DFT is the assumption that the ground-states properties of a quantum system of  $N$  particles can be described by its density rather than its wavefunction. Therefore, these properties can be obtained from a 3-dimensional function  $\rho(\mathbf{r})$ , rather than a  $3N$  dimensional function:

$$\rho(\mathbf{r}) = N \int |\Phi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)|^2 d\mathbf{r}_2 d\mathbf{r}_3 \dots d\mathbf{r}_N \quad (1.2.12)$$

DFT based methods generally scale with  $O(N^3)$  with  $N$  the number of electrons in the system, because one has to ultimately diagonalize a Hamiltonian of dimension proportional to the number of electrons. The digitalization operation scales as  $O(N^3)$ . In reality the application of DFT includes multiple steps and comes in different implementations, which can further reduce the scaling of DFT methods. I will discuss some of those in more detail in the Methods section. In comparison to methods such as CCSD(T) the advantage is that DFT is computationally less expensive, however the results are less accurate. For various properties the accuracy of different quantum mechanical methods has been benchmarked, for example for magnetizabilities<sup>[9]</sup> and NMR shielding constants<sup>[10]</sup>

The fully quantum mechanical treatment is to date only feasible for systems up to about  $10^4$  atoms in size due to wave function based methods scaling with least  $N^3$  with  $N$  the number of electrons in the simulation.

### 1.2.3. Molecular Mechanics models

The simulation of larger systems can be achieved by introducing simplified models that profit from the definition of analytical expressions for both the energy and the forces. Such molecular mechanics (MM) models describe all intermolecular as well as intramolecular interactions

using simple harmonic-like, parameterized terms to represent molecular vibrations, and two-body terms for the intermolecular forces. A detailed explanation of the forces and their parameters is presented in Appendix D. In general the parameters of the potentials are derived by fitting them in such a way that they to reproduce accurate QM simulations of calibration systems as well as on large amounts of experimental data.<sup>[11–14]</sup>

The MM approach couples the speed-up through analytical determination of the relevant energy of force quantities to the better scalability determined by the intermolecular forces, which is in principle proportional to  $N^2$ . In fact, such scaling can be further improved by the introduction of protocols like long-range cut-off, pair-list update and reciprocal space calculations.<sup>[15]</sup> Classical MD works well under the assumption that the BO approximation is valid, and the electronic structure is not largely affected. For example, the molecular motion does not produce any chemical bond breaking or forming, or does not induce significant variations in the polarization of the system.

### 1.3. Current state of the art in simulations

Chemical and biological reactions occur at very different time scales, ranging from femto-second to seconds. Much effort has been devoted to developing robust mathematical and physical computational methods in order to simulate and understand these processes. With increasing computing power it is possible to simulate ever-larger systems with ever-higher precision methods. However, there are still practical limitations (time and resources) to what kind of detailed description is available for a system of a certain size.

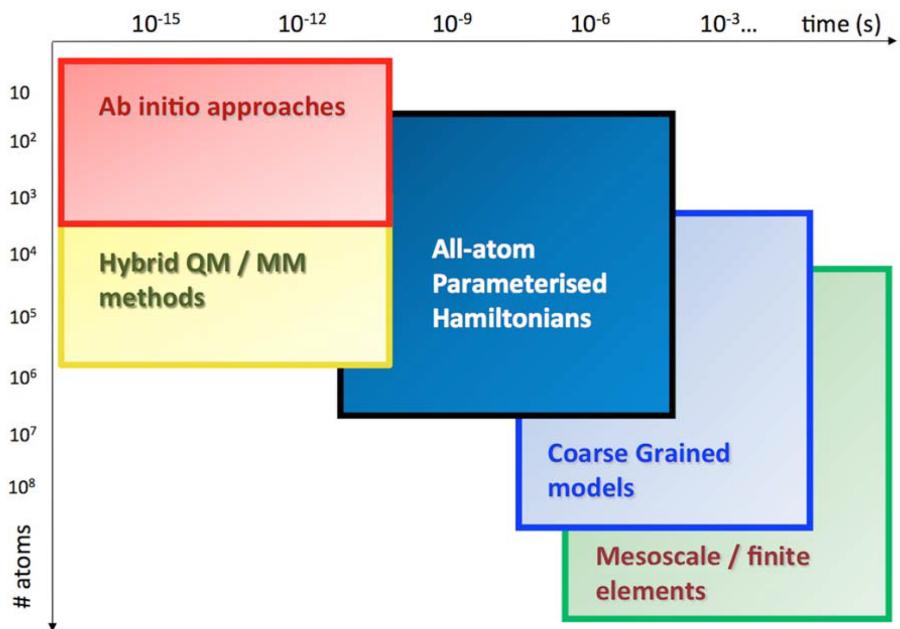
Iterative algorithm approaches still face two more bottlenecks. Firstly, the energy or the forces have to be calculated at any iteration of the

algorithm (MC or MD). Secondly, in order to be numerically stable, at each iteration the new microstate produced can only be very close the previous one. In fact, in MC algorithms any new random microstate is produced by consecutive perturbative changes, while in MD each frame of a trajectory is separated in time by just a small integration time-step  $\Delta t$ . As a matter of facts, in both algorithms, a thorough exploration of the phase-space may require iteration over  $10^{8-10}$  steps or more. The total computational time necessary for any simulations is thus the product of the time spent to compute the energy or the energy gradients with the limitation in producing sufficiently different new configurations.

Restricting this analysis to MD (even though a similar reasoning stands for MC), the time step in QM is determined by the nuclear motion and relaxation of the electrons, in MM the time step must be chosen such that the numerical integration of the equation of motion conserves energy as well as numerical stability and accuracy. In order to guarantee these criteria are met, the time step must typically be at least one order of magnitude smaller than the fastest motion in the system and depends on the properties of interested. A common choice of the time step when using MM models is 1-2 fs. Thus, a time window of 1  $\mu\text{s}$  requires integration of the equations of motion over  $10^9$  iterations.

Despite the fact, that such time intervals are becoming feasible today, characteristic times driving phenomena in large-scale macromolecular systems like biomolecular complexes in the multi-phase can easily exceed such timescales.

A good overview of which methods are applicable for certain time- and size-scales is given in Figure 1.2. It shows various bottom-up approaches, in which the approach is to describe the system of interest using computational tools based on the underlying levels of the physical hierarchy.



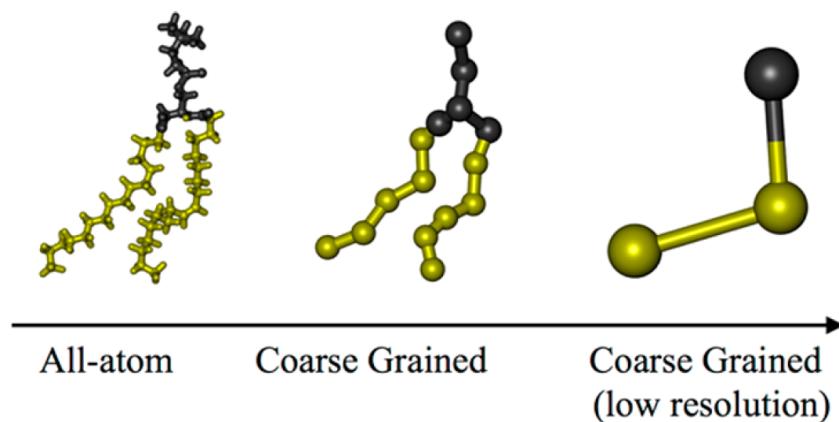
**Figure 1.2.** – Accessible time- and length-scales for state-of-the-art Bottom-up computational approaches. The figure is taken from<sup>[16]</sup>

Non-dynamics processes that require very high accuracy are explored using static calculations. The most fundamental dynamical processes are captured using ab-initio molecular dynamics approaches. The system is described using quantum mechanical methods and spans scales of up to  $\sim 10^3$  atoms and  $\sim 10^{-12}\text{--}10^{-10}$  s.<sup>[17]</sup>

In the case that the system under investigation is very large, but a QM treatment of part of the system is required, it is possible to adopt mixed quantum mechanics/molecular mechanics description of the system. In QM/MM the central idea is to treat the chemically active part of the system using QM, whereas the majority of the simulation (mostly solvent) is treated on a MM level of theory.<sup>[18]</sup> The computational restrictions for this approach are mostly determined by the same ones of the respective QM and MM regions. The QM/MM method has been particularly successful in describing enzymatic catalysis,<sup>[19]</sup> and it has been acknowledged with the Nobel prize in chemistry in 2013.

The largest number of MD simulations to date are carried out using a molecular MM description of the interatomic forces. Such approaches have been successfully applied to large molecular systems, up to  $\sim 10^6$  atoms and  $\sim 10^{-9}\text{--}10^{-4}$  s.<sup>[20,21]</sup> For example, MM models have been used to describe the HIV-1 capsid (64 million atoms)<sup>[22]</sup> or protein folding events over time lengths in the millisecond.<sup>[23]</sup>

**Coarse-grained methods** In order to break the bottlenecks present in simulations adopting all-atom models, further simplifications can be introduced. Coarse-graining (CG) is one approach that has gained a lot of popularity in the last decades. In CG, multiple atoms are mapped into one single interaction site, typically referred to as *bead* (Figure 1.3).



**Figure 1.3.** – In coarse-graining multiple atoms are summarized as a single site. The particles properties represent an average of interactions between the coarse-grained atoms and the environment. Adapted from<sup>[24]</sup>

Beads interact via effective model potentials that are built following different principles. In a flavor of CG modeling, the atoms that are mapped into one bead are chosen such that the same bead can represent well certain physical and chemical properties of the group atoms in the surrounding environment. Interaction parameters can be trained upon specific physical properties from MM models or taken directly to

## BACKGROUND AND INTRODUCTION

---

reproduce experimental data. These CG models have been successfully applied on molecular systems up to several millions of atoms on effective time scales reaching the order of 1 s.<sup>[25–37]</sup>

Other CG approaches do not aim for describing the system in its general features, rather the interest lies within describing specific aspects of the system under investigation, trying to minimize the computational complexity of the representation. Such so-called *toy models* guarantee a qualitative good estimate of the relevance of specific characteristics for the emergence of any properties of interest, rather than a realistic description of the global behavior of the system.

An example for such family of models is the one of patchy particle (PP) models, one of which was used in this thesis to simulate the self-assembly of a protein nano-sphere. From a methodological point of view PP simulation models are all based on "mesoscopic particles, [that are] characterized by a repulsive core and a discrete number of short-range and highly directional interaction sites".<sup>[38]</sup> As indicated by this broad definition patchy particle models come in various shapes of implementations for various applications have been applied successfully for example, to model double stranded DNA<sup>[39]</sup>, proteins<sup>[40–42]</sup>. A good overview of the general methodology is given in<sup>[38]</sup> and of patchy models of protein assembly in<sup>[40]</sup>.

---

The choice between the different presented tools when modeling a molecular system is a constant wager of accuracy against computational efficiency. Often, more than one method might be applicable to the chemical or biological system at hand. It requires insight into the problem under investigation to identify the relevant properties of the system to be given special attention to and thus choose the optimal computational tool for investigation.

In this thesis I will present two examples about how the use of methodologies at very different resolutions allows to address complex (bio)chemical problems at very different time and scale dimensionalities.

## 2. Methods

In this Chapter I will introduce the methodological background of this thesis. I will first describe Quantum Mechanical methods, then I will introduce Monte Carlo, and Molecular Dynamics simulations. For each of these methods I will present the detailed computational tools required for their implementation. The chapter is concluded with a summary of the enhanced sampling methods applied throughout the thesis. The specific parameter choices and computational setup used for the different applications will be detailed in their respective chapters.

### 2.1. Quantum Mechanical Methods

As presented in the Introduction, quantum mechanical methods require the solution of the electronic Hamiltonian in equation 1.2.8, which explicitly written as:

$$\hat{H}_{el} = - \sum_{i=1}^N \frac{\hbar^2}{2m} \nabla_i^2 - \frac{e^2}{4\pi\varepsilon_0} \sum_{i=1}^N \sum_{k=1}^M \frac{Z_k}{R_{ik}} + \frac{e^2}{4\pi\varepsilon_0} \sum_{i < j}^N \frac{1}{r_{ij}} + \frac{e^2}{4\pi\varepsilon_0} \sum_{k < l}^M \frac{Z_k Z_l}{R_{kl}} \quad (2.1.1)$$

where the first term is the kinetic energy operator, the second term is the nuclei-electron interaction and the last term represents the electron-electron repulsion. The first two terms are sums of single-particle operators, while the electron-electron interaction is a pair interaction.

To simplify for later use we can use:

$$\hat{H}_{el} = \sum_i \hat{h}_1(\mathbf{x}) + \frac{1}{2} \sum_{i \neq j} \hat{h}_2(\mathbf{x}_i, \mathbf{x}_j) \quad (2.1.2)$$

Because we cannot solve the corresponding eigenvalue problem for more than two electrons, we must use numerical approximation methods.

### 2.1.1. Wave function based methods

Wave function based methods make use of the variational theorem which states that the expectation value for the energy of any trial function will also be greater or equal to the ground-state energy of the true solution to the Schrödinger equation. Therefore, we can approximate the solutions by minimizing the expectation value of the energy of a trial function, dependent on a set of free parameters.

**Hartree-Fock approach** One of the earliest variational approaches is the Hartree-Fock (HF) method. Its main idea is to approximate the real wavefunction as a product of single-particle spin orbitals.

$$\Psi(1, 2, 3, \dots, N) = \chi_1(2), \chi_2(2) \dots \chi_N(N) \quad (2.1.3)$$

In order to satisfy the antisymmetry requirement for Fermi particles, the wavefunction is written using a Slater determinant, which conveniently captures all the permutations of the single-electron orbitals.

$$\Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\mathbf{x}_1) & \chi_2(\mathbf{x}_1) & \cdots & \chi_N(\mathbf{x}_1) \\ \chi_1(\mathbf{x}_2) & \chi_2(\mathbf{x}_2) & \cdots & \chi_N(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(\mathbf{x}_N) & \chi_2(\mathbf{x}_N) & \cdots & \chi_N(\mathbf{x}_N) \end{vmatrix} \quad (2.1.4)$$

The energy expression is given by:

$$\langle \Psi | \hat{H}_{el} | \Psi \rangle \quad (2.1.5)$$

Because the expression for the energy is symmetric, the variational theorem can be applied to this expression. In practice, the HF energy is minimized with respect to changes in the spin orbitals  $\chi_i$  by Lagrange's method of undetermined multipliers:

$$\mathcal{L}[\{\chi_i\}] = E_{HF}[\{\chi_i\}] - \sum_{ij} \lambda_{ij} (\langle i|j \rangle) - \delta_{ij} \quad (2.1.6)$$

where  $\lambda_{ij}$  are the undetermined Lagrange multipliers and  $\langle i|j \rangle$  is the overlap between spin orbitals  $i$  and  $j$ :

$$\langle i|j \rangle = \int \chi_i^*(\mathbf{x}) \chi_j(\mathbf{x}) d\mathbf{x} \quad (2.1.7)$$

Setting  $\delta\mathcal{L} = 0$ , and after some mathematical operations the HF equations for the energy defining the orbital is obtained as:

$$\begin{aligned} & \sum_i \lambda_{ji} \chi_i(\mathbf{x}_1) = \hat{h}_1 \chi_j(\mathbf{x}_1) \\ & + \sum_{i \neq j} \left\{ \int \chi_i^*(\mathbf{x}_2) \hat{h}_2 \chi_i(\mathbf{x}_2) \chi_j(\mathbf{x}_1) d(\mathbf{x}_2) - \int \chi_i^*(\mathbf{x}_2) \hat{h}_2 \chi_i(\mathbf{x}_2) \right. \\ & = \hat{h}_1 \chi_j(\mathbf{x}_1) + \sum_i (\hat{J}_i - \hat{K}_i) \chi_j \end{aligned} \quad (2.1.8)$$

The first integral is called the *Coulomb term* and usually, while the second term is the *exchange term*, that are commonly written as orbital dependent operators  $\hat{J}_i$  and  $\hat{K}_i$ , respectively.

Realizing that the above integrals cancel each other out for  $i = j$  and

## METHODS

---

defining the *Fock operator* as

$$\hat{F} = \hat{h}_1 + \sum_i (\hat{J}_i - \hat{K}_i) \quad (2.1.9)$$

one can rewrite eq. 2.1.8 as:

$$\hat{F}\chi_i = \sum_i \lambda_{ji}\chi_j \quad (2.1.10)$$

I will hereafter for simplicity on focus on the solutions that satisfy:

$$\lambda_{ji} = \delta_{ji}\varepsilon_i \quad (2.1.11)$$

Solving the Fock equation 2.1.10 involves the evaluation of the orbitals at every point in space. Therefore, in practice the orbitals are expanded in terms of a known basis set and variation is carried out over the coefficients  $C$  of the basis set  $\tilde{\chi}$ . I will cover the choice of basis set in detail in Section 2.1.3. Each orbital is written as:

$$\chi_i(\mathbf{x}) = \sum_{\mu=1}^{N_b} C_{\mu i} \tilde{\chi}_{\mu} \quad (2.1.12)$$

where  $N_b$  is the number of basis functions used. Introducing this into eq. 2.1.10 gives:

$$\hat{F}(\mathbf{x}) \sum_{\mu} C_{\mu i} \tilde{\chi}_{\mu}(\mathbf{x}) = \varepsilon_i \sum_{\mu} C_{\mu i} \tilde{\chi}_{\mu}(\mathbf{x}) \quad (2.1.13)$$

Multiplying by  $\tilde{\chi}_{\nu}^*(\mathbf{x})$  from the left and integrating yields:

$$\sum_{\mu} C_{\mu i} \int \tilde{\chi}_{\nu}^*(\mathbf{x}) \hat{F}(\mathbf{x}) \tilde{\chi}_{\mu}(\mathbf{x}) d\mathbf{x} = \varepsilon_i \sum_{\mu} C_{\mu i} \int \tilde{\chi}_{\nu}^*(\mathbf{x}) \tilde{\chi}_{\mu}(\mathbf{x}) d\mathbf{x} \quad (2.1.14)$$

This equation can be written in matrix form as:

$$\sum_{\mu} F_{\nu\mu} C_{\mu i} = \varepsilon_i \sum_{\mu} S_{\nu\mu} C_{\mu i} \quad (2.1.15)$$

or

$$\mathbf{F}\mathbf{C} = \mathbf{S}\mathbf{C}\varepsilon \quad (2.1.16)$$

Here  $\mathbf{F}$  is the Fock,  $\mathbf{S}$  the overlap and  $\mathbf{C}$  the coefficient matrix. Solving this matrix equation is the key step of the most common implementations of the HF and post-HF methods.

The HF approximation is a mean-field solution to the Schrödinger equation as any electron is affected by the average potential exhibited by all other electrons. Configuration Interaction (CI) and Coupled Cluster (CC) belong to the family of post-HF methods, improving on the accuracy of the method.

**Configuration Interaction** In order to reach higher accuracy than provided by the HF method, in CI the variational wave function is constructed as a linear combination of configuration state functions, which take into account that the presence of an electron modifies the probability distribution function of another electron. This is known as electron correlation. Practically, the CI wavefunction is a linear combination of Slater determinants  $|\Phi\rangle$ :

$$|\Psi\rangle = \sum_{p=0}^N C_p |\Phi_0\rangle = c_0 |\Phi_0\rangle + \sum_i^{\text{occ}} \sum_a^{\text{vir}} c_i^a |\Phi_i^a\rangle + \sum_{i < j}^{\text{occ}} \sum_{a < b}^{\text{vir}} c_{ij}^{ab} |\Phi_{ij}^{ab}\rangle \\ + \sum_{i < j < k}^{\text{occ}} \sum_{a < b < c}^{\text{vir}} c_{ijk}^{abc} |\Phi_{ijk}^{abc}\rangle + \dots \quad (2.1.17)$$

The first contribution is the HF Slater determinant, the following contributions arise from Slater determinants that describe the system

## METHODS

---

where 1,2,3 electrons are excited from any occupied orbital  $i,j,k$  to the unoccupied orbitals  $a,b,c$ . The linear coefficients are determined variationally via diagonalization of the Hamiltonian.

For any but the simplest systems, the summation in eq. 2.1.17 must be truncated, such that only a small subset of determinants has to be calculated. The disadvantage of truncated CI-methods is that they are not size extensive. This means that the energy for the system AB consisting of two independent subsystems A and B, is no longer the sum of the energies of A and B.

**Coupled Cluster** Coupled Cluster theory overcomes the self-extensivity problem by taking an exponential Ansatz to the wavefunction using a cluster operator  $T$ :<sup>[7,8]</sup>

$$|\Psi\rangle = e^T |\Phi_0\rangle \quad (2.1.18)$$

The cluster operator is given by:

$$T = T_1 + T_2 + T_3 + \dots, \quad (2.1.19)$$

where  $T_1$  is the operator of all single excitations,  $T_2$  is the operator of all double excitations and so on. In practice the order of the excitations that can be taken into account are finite and the operator is expanded as a Taylor series. For example, considering single and double excitations ( $T_1$  and  $T_2$ ), it becomes:

$$e^T = 1 + T + \frac{1}{2!} T^2 + \dots = 1 + T_1 + T_2 + \frac{1}{2} T_1^2 + T_1 T_2 + \frac{1}{2} T_2^2 + \dots \quad (2.1.20)$$

The CC wave function, that is obtained by truncating the cluster operator (e.g. CCSD, taking into account only single and double excitations), contains all determinants of such excitation in the Fock Space. Truncated CC provides therefore a systematical improvement of truncated CI, where only those excitations that arise from linear application of the cluster operator are taken into account.

In this thesis we employed CCSD(T), standing for Coupled Cluster Single and double excitation with triple excitations calculated using perturbation theory for the benchmarking purposes. It is often referenced as the gold standard of quantum chemistry.

## 2.1.2. Density Functional Theory

**Hohenberg-Kohn theorems** DFT is based on the two Hohenberg-Kohn theorems.<sup>[5]</sup> The first theorem shows that given a quantum mechanical system in its ground state, the corresponding density  $\rho(\mathbf{r})$  is unequivocally determined by one well-defined potential operator  $\hat{V}$ . As a consequence, because  $\hat{V}$  distinctly describes the Hamiltonian of the system, the ground state energy and all resulting physical observables can be expressed as functionals of the electron density.

The second theorem establishes a variational principle for the ground state density. It shows that, given any trial density  $\bar{\rho}(p) > 0$  with  $\int \bar{\rho}(\mathbf{r})d\mathbf{r} = N$  it follows that  $E[\bar{\rho}] \geq E[\rho]$ . From this theorem the ground-state density is obtainable in a variational way by rewriting the electronic Schrödinger equation 1.2.8 in terms of the density:

$$E[\rho] = T[\rho] + E_{ee} + E_{en}[\rho] = \int \rho(\mathbf{r})\nu(\mathbf{r})d\mathbf{r} + F_{HK}[\rho] \quad (2.1.21)$$

where  $\nu(\mathbf{r})$  is the potential, and  $F_{HK}[\rho] = \langle \Phi[\rho] | H_{el} | \Phi[\rho] \rangle$  the Hohenberg-Kohn functional.  $H_{el}$  consists of the kinetic energy operator and the electron-electron repulsion functional. Only the ground state density  $\rho$  fulfills the stationary principle:

$$\delta\{E[\rho] - \mu[\int \rho(\mathbf{r})d\mathbf{r} - N]\} = 0, \quad (2.1.22)$$

## METHODS

---

with  $\mu$  given by the Euler-Lagrange equation:

$$\mu = \nu(\mathbf{r}) + \frac{\delta F_{HK}[\rho]}{\delta \rho} \quad (2.1.23)$$

In this formalism  $F_{HK}$  is independent of the external potential. Though the HK theorems are rigorous, they do not yield an analytical expression. Therefore, further approximations are needed to make DFT applicable to computational chemistry.

**Kohn-Sham equations** The Hohenberg-Kohn theorems were reformulated into a computationally implementable form by Kohn and Sham.<sup>[6]</sup> The main idea of Kohn and Sham (KS) was that any system of  $N$  interacting particles can be mapped into a non-interacting system that is described by the same density. For the resulting system the density can be written as:

$$\rho(\mathbf{r}) = \sum_{i=1}^N |\varphi_i^{KS}(\mathbf{r})|^2 \quad (2.1.24)$$

where  $\varphi_i^{KS}(\mathbf{r})$  are auxiliary single-particle functions that take the name of Kohn-Sham orbitals. The kinetic energy functional has the analytical expression:

$$T_s[\rho] = \sum_{i=1}^N \left\langle \varphi_i^{KS} \left| -\frac{1}{2} \nabla^2 \right| \varphi_i^{KS} \right\rangle \quad (2.1.25)$$

The Hohenberg-Kohn functional can be re-formulated:

$$F_{HK}[\rho] = T_s[\rho] + J[\rho] + E_{xc}[\rho] \quad (2.1.26)$$

where  $J$  is a functional describing the Coulomb energy and the  $E_{xc}$  is the exchange-correlation functional:

$$E_{xc}[\rho] = T[\rho] - T_s[\rho] + E_{ee}[\rho] - J[\rho] \quad (2.1.27)$$

The name initiates from, that this term can be physically interpreted as containing the contributions of the electron correlation of and particle exchange to the system energy. The Euler-Lagrange equation 2.1.23 thus becomes:

$$\mu = \nu_{eff}(\mathbf{r}) + \frac{\delta T_s[\rho]}{\delta \rho} \quad (2.1.28)$$

$$\nu_{eff} = \nu(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + \frac{\delta E_{xc}[\rho]}{\delta \rho} \quad (2.1.29)$$

The total energy of the system is given by:

$$E = \sum_{i=1}^N \varepsilon_i - \frac{1}{2} \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' d\mathbf{r} + E_{xc} - \int \frac{\delta E_{xc}[\rho]}{\delta \rho} d\mathbf{r} \quad (2.1.30)$$

where  $\varepsilon_i$  are the KS orbital energies of the non-interacting system, which are calculated solving the KS equations self-consistently:

$$\left[ \frac{1}{2} \nabla^2 + \nu_{eff}(\mathbf{r}) \right] \varphi_i = \varepsilon_i \varphi_i \quad (2.1.31)$$

The advantage of this formulation is that the single-particle equations, in principle, exactly solve the many-body problem. The problem that remains, however, is that the  $E_{xc}$  again does not have a proper analytical expression and therefore it requires the use of approximate expressions, as presented in the next paragraph.

## Exchange-Correlation Functionals

**LDA approximation** The first successful approximation for the exchange-correlation functional was already proposed in the original paper by Hohenberg and Kohn.<sup>[5]</sup> The electron density is approximated to behave like a uniform electron gas (LDA, Local Density Approximation).

## METHODS

---

The exchange-correlation functional becomes:

$$E_{xc}^{LDA}[\rho] = -\frac{3}{2} \left( \frac{3}{4\pi} \right)^{\frac{1}{3}} \int \rho \frac{4}{3}(\mathbf{r}) d\mathbf{r} \quad (2.1.32)$$

The LDA approximation describes highly covalent systems well, but does not describe many other features, for example weaker interactions like hydrogen bonds.

**Gradient-Correlated approximation** The family of so-called Generalized-gradient approximation(GGA) functionals introduces a direct dependency on both the density, and the gradient of the density.

The energy of the exchange correlation functional  $E_{xc}$  in GGA is assumed to be separable in the sum of an exchange  $E_x$  and a correlation term  $E_c$ :

$$E_{xc}^{GGA} = E_x^{GGA} + E_c^{GGA} \quad (2.1.33)$$

The exchange part is formulated as a function of the density  $\rho$  and a dimensionless density gradient  $s$ :

$$E_x^{GGA}(\rho, \Delta\rho) = \int_{\mathbf{r}} \rho(\mathbf{r}) \varepsilon_x^{LDA}(\rho) F_x(s) \quad (2.1.34)$$

where  $\varepsilon_x^{unif}(\rho) = -3k_F/(4\pi)$  is the Slater exchange energy density in the LDA approximation,  $k_F = [3\pi^2\rho(\mathbf{r})]^{1/3}$  is the Fermi wave vector and  $F_x(s)$  is the GGC enhancement factor:

$$s = \frac{|\nabla| \rho}{2k_F \rho} \quad (2.1.35)$$

The correlation part is formulated as:

$$E_c^{GGA}(\rho, \Delta\rho) = \int_{\mathbf{r}} \rho(\mathbf{r}) \varepsilon_x^{LDA}(\rho) F_c(r_s, \zeta, t) \quad (2.1.36)$$

$\zeta$  the spin polarisation,  $r_s = [(4\pi/3)\rho(\mathbf{r})]^{1/3}$  and  $t$  another dimensionless gradient:

$$t(\mathbf{r}) = \frac{|\nabla|\rho}{2gk_s\rho} \quad (2.1.37)$$

with  $g = [(1 + \zeta)^{2/3} + (1 - \zeta)^{2/3}]/2$  and  $k_s = (4k_F/\pi)^{1/2}$ .

**Perdew-Burke-Ernzerhof functional** In this thesis I employed the Perdew-Burke-Ernzerhof (PBE) approximation.<sup>[43]</sup> The PBE functional is a parameter-free GGA. "Parameter-free" here indicates that it is not parametrized by fitting to experimental data, but rather by theoretically derived constants.

In PBE the enhancement factor of the exchange functional takes the form:

$$F_x^{PBE}(s) = 1 + \kappa = \frac{\kappa}{1 + \frac{\mu}{\kappa}s^2} \quad (2.1.38)$$

where  $\kappa = 0.804$  is set to the maximum value of the local Lieb-Oxford bound<sup>[44]</sup> and  $\mu = 0.21951$  is set to account for the linear response of the LDA approximation such that the gradient exchange part accounts for the correlation.

The correlation functional is adapted from the Perdew-Wang-91 correlation functional<sup>[45]</sup>:

$$F_c^{PBE}(r_s, \zeta, s) = 1 + \frac{H(r_s, \zeta, s)}{\varepsilon_c^{LDA}(r_s, \zeta)} \quad (2.1.39)$$

with

$$H^{PBE} = g^3 \frac{\beta^2}{2\alpha} \ln \left[ 1 + \frac{2\alpha}{\beta} \frac{t^2 + At^4}{1 + At^2 + A^2t^4} \right] \quad (2.1.40)$$

with

$$A(r_s, \zeta) = \frac{2\alpha}{\beta} \frac{1}{e^{\frac{-2\alpha\varepsilon_c^{LDA}(r_s, \zeta)}{g^3\beta^2}}} \quad (2.1.41)$$

with  $\alpha = 0.0716$  and  $\beta = 0.066725$ .

### 2.1.3. Natural Bond Order Analysis

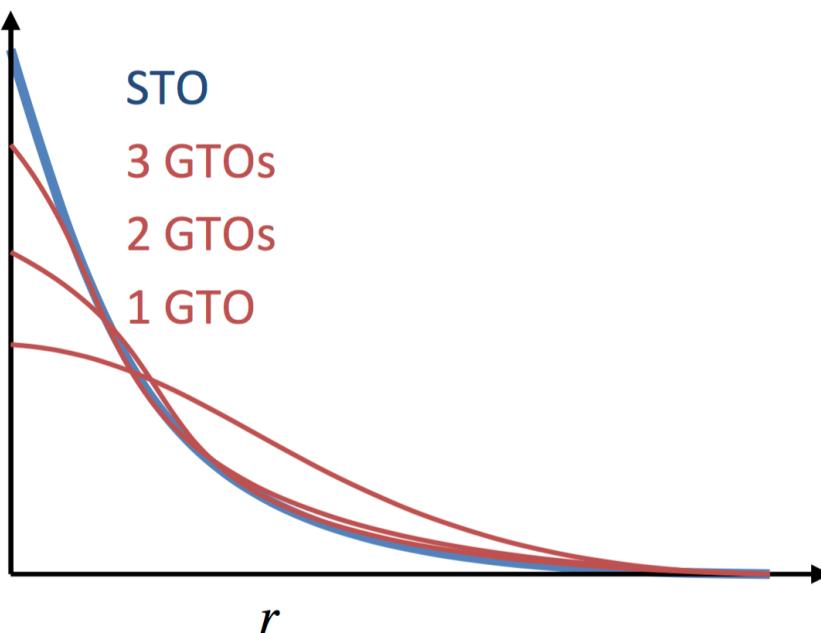
Natural Bond Order (NBO) Analysis is a tool for translating the solutions of the Schrödinger equation into chemical bondings concepts. The NBO algorithm translates atomic orbitals (AOs) into molecular orbitals (MOs) by a series of transformations, over localised ‘natural’ atomic orbitals (NAOs), hybrid orbitals (NHOs), bond orbitals (NBOs), and (semi-)localised molecular orbitals (NLMOs). Each of these forms a complete orthonormal set. The details of this transformation are covered in Ref.<sup>[46]</sup> and Ref.<sup>[47]</sup>.

## 2.2. Basis Sets and Pseudopotentials

**Basis Sets** Most of the time spent in the calculations is on numerically solving the integrals in the Fock and overlap matrices in eq. 2.1.16. Therefore, we need a smart choice for the representation of the basis functions  $\varphi_i$  in the above equations. A typical choice are the so-called Slater-type-orbitals (STOs), which have the shape of the solutions of the hydrogen atoms to first order of the Laguerre polynomials.

$$R_n(\mathbf{r}) = (2\zeta)6n + \frac{1}{2}[(2n)!]^{-1\frac{1}{2}}r^{n-1}\exp(-\zeta r) \quad (2.2.1)$$

STOs are however difficult to compute and are therefore approximated as a linear combination of more easily computable Gaussian type orbitals (GTOs) as shown in Figure 2.1. A special notation is used when GTO is a linear combination of GTOs. For example, STO-3G a minimal basis in which three Gaussian are used to represent each slater determinant. In this thesis we use the 6-31++G(d,p) basis set for Natural Bond Analysis. It uses 6 Gaussians to describe the core orbitals. For the valence orbitals, the basis set is split; three Gaussians are used to describe the contracted part of the wave function and one for the diffusive part. This split is important to allow for contracting and expansion of the valence orbitals according to the environment.



**Figure 2.1.** – A slater type orbital (STO) can be approximated by Gaussian type orbitals (GTOs).

The  $++$  stands for diffusive functions that are added to the orbitals. The  $G(d,p)$  index adds polarization functions to all atoms. The latter two contributions are computationally expensive, but allow for the orbitals to accommodate different symmetry due to orbital distortion.

**Plane Wave basis set** The plane wave basis set is especially suitable for systems that use periodic boundary conditions (PBC), see Section 2.6. Using the Bloch theorem, every KS wave-function can be expanded to:

$$\varphi_i^{\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{V}} e^{i\mathbf{k} \cdot \mathbf{r}} \sum_g c_j^{\mathbf{k}}(\mathbf{g}) e^{i\mathbf{g} \cdot \mathbf{r}} \quad (2.2.2)$$

where  $V$  is the Volume of the unit cell,  $\mathbf{k}$  vectors of the first Brillouin zone,  $g$  is a reciprocal lattice vector,  $c$  the first Fourier component of the plane waves expansion and the summation is over infinite lattice vectors. One advantage of using a plane wave basis set is that convergence of the results may be systematically benchmarked over a single parameter,

## METHODS

---

the cut-off energy  $E_{cut}$  for the kinetic energy  $T_{PW}$ . Furthermore, the incomplete-basis set corrections and the basis set superposition error are eliminated by having a non-localized origin of the basis set. The additional use of Fast Fourier Transform algorithms ensures that even large systems can be handled computationally efficiently.

**Pseudopotentials** The main drawback of using plain-wave basis sets is that it is computationally too expensive to describe the core electrons accurately, which create sharp fluctuations of the wavefunction around the nuclei. One approximation can be introduced based on the fact that the core electrons are energetically much lower lying in energy than the valence electrons and do not participate in chemical properties of the system. Therefore, it could be possible to consider explicitly only the valence electrons in the electronic structure calculations. In this procedure, the core electrons are incorporated into the nucleus, producing a new central potential that takes the name of "pseudopotential". Pseudopotentials are usually derived from all electron atomic calculations and have to fulfill the following criteria:

- The valence pseudo-wave-functions should not contain any radial nodes
- The valence all electron and pseudopotential eigen-values from quantum mechanical calculation must be the same
- The all electron and pseudo radial wave functions must be the same beyond the cut-off distance
- The integrated electron density within the cut-off distance must be the same
- At the cut-off distance the pseudo-wavefunction and its first four derivatives must be continuous
- The pseudopotential should have no curvature at the origin

To accommodate these conditions, the pseudopotential wave functions

take the general form:

$$\varphi_l^{PP}(\mathbf{r}) = \begin{cases} \varphi_l^{AE}(\mathbf{r}); \mathbf{r} > r_{cut} & r^l e^{p(r)}; \mathbf{r} \leq r_{cut} \end{cases} \quad (2.2.3)$$

with

$$p(r) = c_0 + \sum_{i=1}^6 c_i r^{2i} \quad (2.2.4)$$

where the coefficients  $c$  are determined by the first three criteria.

The pseudopotential functionals take the form:

$$V_{pseudo} = V_{val} + \sum_{m,l} |Y_{l,m}\rangle V_l(\mathbf{r}) \langle Y_{l,m}| \quad (2.2.5)$$

where  $|Y_{l,m}\rangle$  are spherical harmonics.

## 2.3. Monte Carlo Simulations

Monte Carlo is a family of computational algorithms that are used to model a variety of multidimensional problems using random numbers. Here, I introduce the Markov Chain Monte Carlo (MCMC) approach to generate efficient sampling from a probability distribution, more precisely the Boltzmann distribution of states. In MCMC, a chain of states  $C_0 \rightarrow C_1 \dots C_k \rightarrow C_{k+1}$  is generated starting from an initial state  $C_0$  by a random walk in the configurational space. The probability to sample each conformation has a certain weight according to the energy of the state.

### 2.3.1. Metropolis algorithm

The practical implementation of MCMC is obtained by the Metropolis algorithm (MA). MA profits from the knowledge of the relative prob-

## METHODS

---

ability of two states (equation 1.2.1), even in the case that the total probability distribution is unknown and non-obtainable, meaning the full phase space has not been sampled and the free energy landscape is unknown. The Metropolis algorithm is gains numerical efficiency when the number of dimensions is high.

The MA aids finding the equilibrium state of the simulation more efficiently, as well as sampling the obtained equilibrium's microstates to estimate its average properties.

The MA works according to the following scheme:

- Evaluate the energy  $E_i$  of the current state value  $i$
- Create a trial state  $C_t$  and calculate  $E_t$
- If  $E_t \leq E_i$  then accept the trial state and create a new state  $C_{i+1} = C_t$
- If  $E_t > E_i$ , then
  - Calculate the Boltzmann weights  $p = \exp(-\beta(E_t - E_i))$
  - Select a random number  $\xi$
  - If  $\xi < p$  accept the trial state;  $C_{i+1} = C_t$
  - Else reject the trial state ( $C_{i+1} = C_i$ )

The whole procedure is iterated over several steps, generating a Markov chain. The MA guarantees that the collection of states sampled in the MCMC matches the target Boltzmann distribution. The desired properties can be calculated on-the-fly over the MCMC sampled state until convergence is reached. Importantly, the MA allows the systems to sample states of higher energy, thus making it possible to escape local minima. The more shallow is the minimum, the more likely the simulation overcomes the barrier.

## 2.4. Molecular Dynamics Simulations

As introduced in the first chapter, the main idea behind MD simulations is to sample the statistical ensemble of microstates from the physical time evolution of the system using numerical integration of Newton's equation of motion.

$$\mathbf{F} = m\mathbf{a} \quad (2.4.1)$$

where  $\mathbf{F}$  is the force,  $m$  the particle mass and  $\mathbf{a}$  is the acceleration.

When the evaluation of the forces is done at the QM level, the method is referred to as Ab Initio Molecular Dynamics (AIMD) simulation. AIMD in current computational chemistry is mostly implemented using either Ehrenfest Molecular Dynamics<sup>[48]</sup>, Born Oppenheimer Molecular Dynamics (BOMD) or Car Parrinello Molecular Dynamics (CPMD).<sup>[49]</sup> In this thesis I employ BOMD which will be in short introduced hereafter.

### 2.4.1. Born-Oppenheimer Molecular Dynamics

**Born-Oppenheimer Approximation** As introduced in Chapter 1, the BO approximation is based on separating adiabatically the motion of the nuclei from that of the electrons. Based of this assumption, the wavefunction is factorized in two, electronic and nuclear, components:

$$\Psi(\mathbf{R}, \mathbf{r}) = \Phi(\mathbf{r}; R)\chi(\mathbf{R}) \quad (2.4.2)$$

The Schrödinger equation (eq. 1.2.8) is rewritten as:

$$\hat{H}\Phi(\mathbf{r}; R)\chi(\mathbf{R}) = E\Phi(\mathbf{r}; R)\chi(\mathbf{R}) \quad (2.4.3)$$

using the separation of eq. 1.2.6 and the explicit term for the nuclear

## METHODS

---

potential energy term:

$$\hat{T}_N = \sum_N \frac{\hbar}{2m_N} \nabla^2 \quad (2.4.4)$$

one obtains 2.4.3 as:

$$\hat{H}_{el}\Phi(\mathbf{r}; R)\chi(\mathbf{R}) - \sum_N \frac{\hbar}{2m_N} \nabla_N^2 \Phi(\mathbf{r}; R)\chi(\mathbf{R}) = E\Phi(\mathbf{r}; R)\chi(\mathbf{R}) \quad (2.4.5)$$

Explicitly evaluating the nuclear kinetic energy term of eq. 2.4.5 by differentiating the factorized wavefunction we obtain:

$$-\sum_N \frac{\hbar}{2m_N} \nabla^2 [(\nabla_N^2 \Phi(\mathbf{r}; R))\chi(\mathbf{R}) + 2\nabla_N \Phi(\mathbf{r}; R)\nabla_N \chi(\mathbf{R}) + \Phi(\mathbf{r}; R))\nabla_N^2 \chi(\mathbf{R})] \quad (2.4.6)$$

From the approximation that the electron wavefunction is only parametrically dependent on the position of the nuclei, the first two terms on the above summation are zero. Thus we obtain for nuclear wave function (eq. 1.2.9):

$$-\left[ \sum_N \frac{\hbar}{2m_N} \nabla^2 + \hat{H}_{el}(\mathbf{R}) \right] \chi(\mathbf{R}) = E\chi(\mathbf{R}) \quad (2.4.7)$$

**BOMD Equation of Motion** The BOMD equation of motion for the ground state can be derived from the coupled Schrödinger equation of electrons and nuclei<sup>[50]</sup> and is given by:

$$m_I \frac{d^2 r_I}{dt^2} = -\nabla_I \min_{\Phi} \left\{ \left\langle \Phi \middle| \hat{H}_{el} \middle| \Phi \right\rangle \right\} \quad (2.4.8)$$

$$\hat{H}_{el}\Phi = E_{el}\Phi$$

where the indexing  $\Phi(\mathbf{r}; R)$  shall be implied hereafter. The potential energy surface for the nuclei is obtained by solving the fixed-nuclei electronic structure problem of the time-independent Schrödinger equation at every molecular dynamics step. The time dependence of the electronic structure is implicitly treated by its dependency on the time-evolution of the nuclei propagation.

For Kohn-Sham density functional theory used in this thesis, we can derive the equations of motion more specifically. With the exact wave function represented by a single slater determinant  $\Phi = 1/\sqrt{N!} \det \psi_i$  with orthonormal wave functions  $\psi_i$ , the corresponding minimization problem becomes:

$$\min_{\Phi} \left\langle \Phi | \hat{H}_{el} | \Phi \right\rangle |_{\{\langle \psi_i \psi_j \rangle = \delta_{ij}\}} \quad (2.4.9)$$

This can be reformulated as a Lagrangian formalism:

$$\mathcal{L} = - \left\langle \Phi | \hat{H}_{el} | \Phi \right\rangle + \sum_{ij} \Lambda_{ij} (\langle \psi_i \psi_j \rangle - \delta_{ij}), \quad (2.4.10)$$

where  $\Lambda_{ij}$  are the Lagrangian multipliers used to impose the orthonormality constraint. Variation of the Lagrangian  $\mathcal{L}$  with respect to the orbitals  $\psi_i$  leads to reformulation of the equation of motion as:

$$\begin{aligned} m_I \frac{d^2 r_I}{dt^2} &= -\nabla_I \min_{\Phi} \{ \langle \Phi | H_{el} | \Phi \rangle \} \\ 0 &= -\hat{H}_{el}^{KS} \psi_i + \sum_j \Lambda_{ij} \psi_j, \end{aligned} \quad (2.4.11)$$

where  $\hat{H}_{el}^{KS}$  is the effective one-particle Hamiltonian obtained from KS theory. We employ the QUICKSTEP<sup>[51]</sup> implementation of BOMD in this thesis.

### 2.4.2. Hellman-Feynman Theorem

Equation 2.4.8 requires the evaluation of the gradient of the expectation value of the electronic Hamiltonian with respect to the ground state wavefunction. According to the Hellman-Feynman theorem (HF theorem) this can be evaluated according to:

$$\nabla E_{el}(\mathbf{R}) = \left\langle \Phi(\mathbf{R}) \left| \nabla \hat{H}_{el}(\mathbf{R})(\lambda) \right| \Phi((\mathbf{R}) \right\rangle \quad (2.4.12)$$

More generally the HF theorem states that:

$$\frac{\partial E_0(\lambda)}{\partial \lambda} = \left\langle \Phi \left| \frac{\partial}{\partial \lambda} \hat{H}(\lambda) \right| \Phi \right\rangle \quad (2.4.13)$$

where we write the electronic Hamiltonian of a quantum-mechanical system as  $\hat{H}_{el} = \hat{H}_{el}(\lambda)$ , where  $\lambda$  is a general continuous parameter. The electronic problem is solved according to:

$$\hat{H}_{el}(\lambda)\Phi(\lambda; \vec{r}) = E_{el}(\lambda)\Phi(\lambda; \vec{r}) \quad (2.4.14)$$

for all  $\lambda$ s, the ground state is normalized:

$$\int |\Phi(\lambda; \vec{r})|^2 d^3r = 1 \quad (2.4.15)$$

The demonstration simply follows from the direct calculation of the derivative in the left of the former expression:

$$\frac{\partial E_0(\lambda)}{\partial \lambda} = \left\langle \Phi \left| \frac{\partial}{\partial \lambda} \hat{H}(\lambda) \right| \Phi \right\rangle + \left\langle \frac{\partial}{\partial \lambda} \Phi \left| \hat{H}(\lambda) \right| \Phi \right\rangle + \left\langle \Phi \left| \hat{H}(\lambda) \right| \frac{\partial}{\partial \lambda} \Phi \right\rangle \quad (2.4.16)$$

Because  $\hat{H}$  is a Hermitian operator, the second and third terms of the

right-hand side are equal to:

$$E_0 \left\langle \frac{\partial}{\partial \lambda} \Phi \middle| \Phi \right\rangle + E_0 \left\langle \Phi \middle| \frac{\partial}{\partial \lambda} \Phi \right\rangle = E_0 \frac{\partial}{\partial \lambda} \langle \Phi | \Phi \rangle \quad (2.4.17)$$

Thus,

$$\frac{\partial E_0(\lambda)}{\partial \lambda} = \left\langle \Phi \middle| \frac{\partial}{\partial \lambda} \hat{H}(\lambda) \middle| \Phi \right\rangle + E_0 \frac{\partial}{\partial \lambda} \langle \Phi | \Phi \rangle = \left\langle \Phi \middle| \frac{\partial}{\partial \lambda} \hat{H}(\lambda) \middle| \Phi \right\rangle \quad (2.4.18)$$

In fact,

$$\frac{\partial}{\partial \lambda} \langle \Phi | \Phi \rangle = 0 \quad (2.4.19)$$

because of the normalisation condition in eq. 2.4.15.

However, the HF Theorem is strictly valid for any *exact, normalized eigenstate* of the Hamiltonian operator. Apart from some simple model systems, in quantum mechanics we are usually incapable of obtaining any analytical formulation of such states. Nonetheless, we are able to compute approximations to  $\Psi$ .

**HF theorem in the variational regime** The way we typically solve the Hamiltonian problem is by expanding its ground state on a complete space provided by an opportune basis set, and by then truncating it over a finite number of such functions:

$$\Phi(\lambda; \vec{r}) = \sum_{n=1}^{\infty} c_n(\lambda) \phi_n(\lambda; \vec{r}) \approx \sum_{n=1}^M c_n(\lambda) \phi_n(\lambda; \vec{r}) = \Phi^{VAR}(\lambda; \vec{r}) \quad (2.4.20)$$

In the most general case, both the basis functions and the coefficients of the expansion depend on  $\lambda$ . Without losing generality, we can choose the basis functions to be orthonormal:

$$\langle \phi_n | \phi_{n'} \rangle = \delta_{nn'} \quad (2.4.21)$$

## METHODS

---

The truncated expansion in eq. 2.4.20 is a variational solution of  $\hat{H}$  if the set of coefficients  $\{c_n(\lambda)\}$  minimises the quantity:

$$\langle \Phi^{VAR} | \hat{H} | \Phi^{VAR} \rangle = E_0^{VAR}(\lambda) \geq E_0(\lambda) \quad (2.4.22)$$

If the expansion were made on a complete set of basis functions, the variational solution would converge to the exact solution and the HF theorem remain still valid, but this is not the case in the most common numerical practices.

In order to find the solution to Eq. 2.4.22, one has to project the variational solution over the selected  $M$  basis functions:

$$\begin{aligned} \hat{H} |\Phi^{VAR}\rangle &= E_0^{VAR}(\lambda) |\Phi^{VAR}\rangle \\ \hat{H} \sum_{n=1}^M c_n(\lambda) |\phi_n\rangle &= E_0^{VAR}(\lambda) \sum_{n=1}^M c_n(\lambda) |\phi_n\rangle \\ \sum_n c_n \langle \phi_{n'} | \hat{H} | \phi_n \rangle &= E_0^{VAR} \sum_n c_n \langle \phi_{n'} | \phi_n \rangle \end{aligned}$$

The solution to eq. 2.4.23 is found solving the secular problem:

$$\sum_{n=1}^M [H_{nn'}(\lambda) - E_0^{VAR}(\lambda) \delta_{nn'}] c_n(\lambda) = 0 \quad (2.4.23)$$

That is, diagonalising the  $H_{nn'}$  matrix:

$$H_{nn'} = \langle \phi_n | \hat{H} | \phi_{n'} \rangle \quad (2.4.24)$$

from which it follows:

$$\sum_n c_n H_{nn'} = E_0^{VAR} c_{n'} \quad (2.4.25)$$

Let's now compute again a generalized force over the  $\lambda$  parameter

when the solution is a variational approximation over finite basis set expansion.

$$\frac{\partial E_0^{VAR}(\lambda)}{\partial \lambda} = \left\langle \Phi^{VAR} \left| \frac{\partial}{\partial \lambda} \hat{H}(\lambda) \right| \Phi^{VAR} \right\rangle + 2\Re e \left[ \left\langle \frac{\partial}{\partial \lambda} \Phi^{VAR} \left| \hat{H}(\lambda) \right| \Phi^{VAR} \right\rangle \right] \quad (2.4.26)$$

The first term is the usual one from the HF theorem. Let's check what happens to the 2nd term when the expansion of the variational solution is made on an incomplete basis set.

$$\left\langle \frac{\partial}{\partial \lambda} \Phi^{VAR} \left| \hat{H}(\lambda) \right| \Phi^{VAR} \right\rangle = \sum_{nn'} \frac{\partial c_n^*}{\partial \lambda} c_{n'} H_{nn'} + \sum_{nn'} c_n^* c_{n'} \left\langle \frac{\partial}{\partial \lambda} \phi_n \left| \hat{H} \right| \phi_{n'} \right\rangle \quad (2.4.27)$$

Because the normalisation condition for all  $\lambda$  (eq. 2.4.15) is kept also for  $\Psi^{VAR}$  (then, for all  $\lambda$ :  $\sum_n |c_n(\lambda)|^2 = 1$ ), it follows that the real part of the first term must be zero.

Nonetheless, the second term does not vanish. These forces, coming from numerical truncation of the variational solution, must be taken into account in quantum mechanical calculations of generalised forces, otherwise spurious interactions (taking the name of *Pulay forces*) appear.

Using the definition of the unity operator:  $\mathbb{1} = \sum_n |\phi_n\rangle \langle \phi_n|$ , the second

term becomes:

$$\begin{aligned}
 \sum_{nn'} c_n^* c_{n'} \left\langle \frac{\partial}{\partial \lambda} \phi_n \left| \hat{H} \right| \phi_{n'} \right\rangle &= \sum_{nn'} c_n^* c_{n'} \left\langle \frac{\partial}{\partial \lambda} \phi_n \left| \mathbb{1} \hat{H} \right| \phi_{n'} \right\rangle \\
 &= \sum_{nn'} c_n^* c_{n'} \sum_{n''} \left\langle \frac{\partial}{\partial \lambda} \phi_n \left| \phi_{n''} \right. \right\rangle \left\langle \phi_{n''} \left| \hat{H} \right| \phi_{n'} \right\rangle \\
 &= \sum_{nn''} c_n^* \left\langle \frac{\partial}{\partial \lambda} \phi_n \left| \phi_{n''} \right. \right\rangle \sum_{n'} c_{n'} H_{n'n''} \\
 &= E_0^{VAR} \sum_{nn''} c_n^* c_{n''} \left\langle \frac{\partial}{\partial \lambda} \phi_n \left| \phi_{n''} \right. \right\rangle \quad (2.4.28)
 \end{aligned}$$

In fact, the term may still vanish in one special case, precisely, if the chosen basis set does not depend on the parameter  $\lambda$ . The most relevant example occurs if the electronic wave function is expanded on a set of plane-waves.

**Particle propagation in Molecular Dynamics** As introduced in the previous chapter, the EOMs in MD simulations have to be solved iteratively using numerical integration algorithms. These algorithms should conserve the energy of the system, be computationally as cheap as possible and maximize the applicable time step while keeping the error due to using numerical approximation as low as possible.

**Verlet algorithm** Historically the first algorithm of this kind used MD was the Verlet algorithm.<sup>[52]</sup> If the previous position of a particle is known, the summation of

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{F(t)}{m} \frac{(\Delta t)^2}{2} + \mathcal{O}(\Delta t^3) \quad (2.4.29)$$

and

$$r(t - \Delta t) = r(t) - v(t)\Delta t + \frac{F(t)}{m} \frac{(\Delta t)^2}{2} + \mathcal{O}(\Delta t^3) \quad (2.4.30)$$

yields

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + \frac{F(t)}{m} (\Delta t)^2 + \mathcal{O}(\Delta t^4) \quad (2.4.31)$$

The error becomes fourth order in  $\Delta t$  because the terms of third order cancel each other out.

A drawback of using the Verlet propagation algorithm is that velocity is not expressed explicitly. Thus they can only be estimated by:

$$v(t) = \frac{r(t + \Delta t) - r(t - \Delta t)}{2\Delta t} + \mathcal{O}(\Delta t^2) \quad (2.4.32)$$

therefore, quantities dependent on the velocities are not very accurately determined.

**Leap Frog Algorithm** However, the knowledge of two consecutive values in time for a variable yields a good estimate for its derivative at a mid-point, which is used in the Leap-Frog-algorithm.

$$v\left(t + \frac{1}{2}\Delta t\right) = \frac{1}{\Delta t}[r(t + \Delta t) - r(t)] \quad (2.4.33)$$

The velocities at  $t + \frac{1}{2}\Delta t$  can then be obtained using velocities of the previous half-step.

$$v\left(t + \frac{1}{2}\Delta t\right) = v\left(t - \frac{1}{2}\Delta t\right) + \frac{F(t)}{m}\Delta t \quad (2.4.34)$$

## METHODS

---

and the particle propagation is given by:

$$r(t + \Delta t) = r(t) + v \left( t + \frac{1}{2} \Delta t \right) \Delta t \quad (2.4.35)$$

The upside of this algorithm is that the velocities are also calculated and can be evaluated, however the time step is half as big with respect to the Verlet algorithm. More advanced algorithms such as the Velocity Verlet algorithms<sup>[53]</sup> or the Beeman algorithm<sup>[54]</sup> have also been derived. They both yield information about the time, velocities and forces at the same time, while yielding accurate description of the system.

In the Velocity Verlet algorithm, used in this thesis, the positions and velocities are simultaneously updated according to:

$$\begin{aligned} r(t + \Delta t) &= r(t) + v(t) \Delta t + \frac{a(t)}{2} \Delta t^2 \\ v(t + \Delta t) &= v(t) + \frac{a(t + \Delta t) + a(t)}{2} \Delta t \end{aligned} \quad (2.4.36)$$

In practice, the initial particle position is either obtained by an educated guess or structural data, such as X-ray crystallographic structures. The initial velocities are linked directly to the temperature of the system

$$E_{kin} = \frac{3N}{2k_B T} = \frac{1}{2} \sum_{i_1}^N m_i V_i^2 \quad (2.4.37)$$

and initialized from the Maxwell-Boltzmann distribution

$$p(v_{ix}) = \left( \frac{m_i}{2\pi k_B T} \right)^{1/2} \exp \left[ -\frac{m_i v_{ix}^2}{2k_B T} \right]. \quad (2.4.38)$$

Here  $p$  is the possibility for the particle  $i$  with the mass  $m_i$  to obtain a velocity  $v_i$ . It is crucial that the total momentum  $P$  of the particles must be zero for the simulation not to have an artificial drift throughout

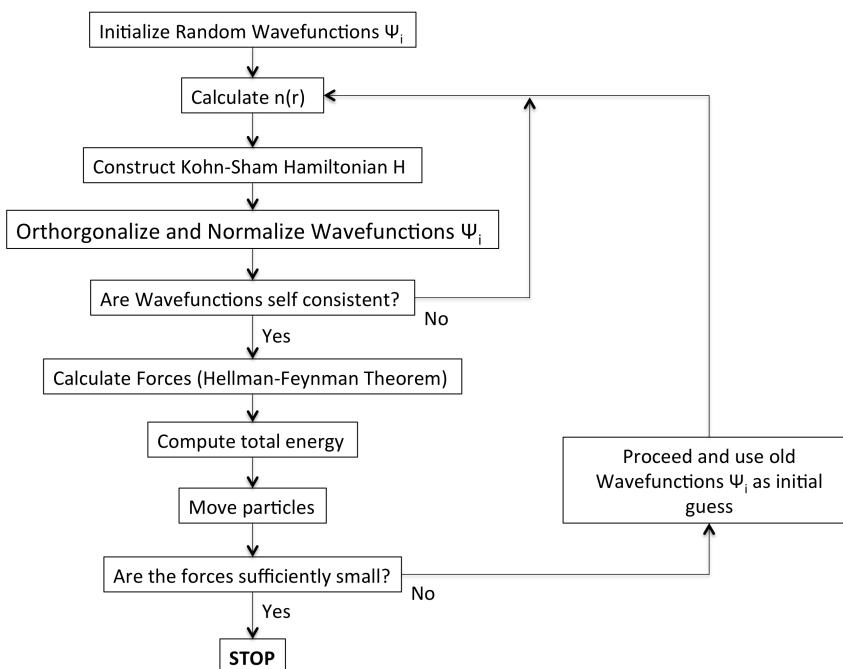
the simulation box:

$$P = \sum_{i_1}^N m_i V_i = 0 \quad (2.4.39)$$

The acceleration is derived from the potential energy according to:

$$a_i(t) = -\frac{1}{m} \frac{dE_{pot}}{dr_i(t)} \quad (2.4.40)$$

The final iteration procedure of BOMD can be summarized in following flowchart:



## 2.5. Thermostats

If the simulation is carried out in the canonical ensemble (NVT), the system must be kept at the correct temperature. In this thesis we employ the Canonical Sampling/Velocity Rescaling (CSVR) and the Nosé-Hoover chain thermostat. For any explicit deduction omitted in this section and additional choices of thermostats please refer to Frenkel and Smit.<sup>[55]</sup>

### 2.5.1. CSVR thermostat

Using the CSVR thermostat the velocities are rescaled every  $n^{th}$  step, such that the kinetic energy of the system stays close to the target temperature  $T_0$ .

$$v_{new}^N = \lambda v^N, \lambda = \sqrt{\frac{T_0}{T(t)}} \quad (2.5.1)$$

Often a temperature tolerance can be used, such that the temperatures are rescaled to the target temperature only if the system temperature differs more than a certain value from the target temperature. This choice is used to better capture statistical energy fluctuations of the canonical ensemble.

### 2.5.2. Nosé-Hoover chain thermostat

The Nosé-Hoover chain is based on the extended Lagrangian formalism. The original idea of Nosé<sup>[56]</sup> was to add two additional degrees of freedom to the system by coupling it to an imaginary heat bath. This imaginary heat bath is classified by the positions  $s$ , the momentum  $p_s$  and the effective mass  $Q$ , that has to be parameterized.

The Hamiltonian of the system is then formulated as:

$$H = \hat{V}_{kin} + \hat{V}_{pot} + \frac{p_s^2}{2Q} + k_B T (3N + 1) \ln(s) \quad (2.5.2)$$

Where the first two terms are the kinetic and potential energy terms of the original system and the latter two terms the kinetic and potential energy of the heat bath, respectively. The kinetic energy is coupled to the heat bath via:

$$\hat{V}_{kin} = \frac{1}{2} \sum_{i=0}^N m_i^{-1} |p_i|^2, p_i = m_i v_i \times s \quad (2.5.3)$$

The result is that the microcanonical simulation of the system including the heat bath yield a canonical ensemble for the original system. The advantage of the Nosé thermostat is that the dynamics of all degrees of freedom time-reversible and do not include random number generation. The drawback of this methods is that implementation of the Nosé is problematic as it introduced simulation time as a variable through the scaling factors  $s$  in eq. 2.5.3.

In order to overcome this problem Hoover developed an extended formalism of the Nosé thermostat.<sup>[57]</sup> The basic idea is to use a friction parameter to adjust the particle velocities in the Hamiltonian formulation of the system, rather than particle momentum parameter  $p_s$ .

$$H = \hat{V}_{kin} + \hat{V}_{pot} + \frac{\zeta^2 Q}{2} + 3Nk_B T \ln(s) \quad (2.5.4)$$

In the Nosé-Hoover thermostat the particle position and velocity propagation is given by:

$$\begin{aligned}\frac{\delta r_i}{\delta t} &= v_i \\ \frac{\delta v_i}{\delta t} &= -\frac{1}{m_i} \frac{\delta V_{Pot}}{\delta r_i} - \zeta v_i \\ \frac{\delta \zeta}{\delta t} &= \left( \sum_{i=0}^N m_i |v_i|^2 - 3Nk_B T \right) / Q \\ \frac{\ln(s)}{\delta t} &= \zeta\end{aligned}\tag{2.5.5}$$

The advantage of this formalism is that the variable  $\zeta$  does not change with time and that numerical integration scheme such as the Velocity Verlet method described in Chapter 1 have been developed to propagate the above equations. The Nosé-Hoover can be extended to couple to multiple heat baths to increase temperature equilibration properties. This approach is called Nosé-Hoover chain thermostat, where the chain length indicates the number of heat baths the system is coupled to.<sup>[58]</sup>

## 2.6. Periodic Boundary Conditions

Due to limited system size in simulations, we employ periodic boundary conditions (PBC). Using periodic boundary conditions the system is surrounded by infinitely many replications of itself. The use of PBC eliminates the surface and any bulk-surface interactions that otherwise dominate the system. In practice this means, that as soon as an atom leaves the unit cell its periodic image enters the unit cell through the opposite side and there is no interface. However, the drawback is that we introduce periodicity to systems that are inherently non-periodic (e.g. solutions).

In MD and MC interactions are only considered with the closest periodic image of any atom. The cutoff distance or rather the system size has to be chosen in such a way that no molecule interacts with itself through

periodic boundary conditions. The most common choice for unit cells used with PBC are orthorombic, however other tile space geometries are in principle implementable for PBC.

## 2.7. Enhanced Sampling Methods

The capability of describing specific molecular events is determined by the characteristic times of the same events compared to those accessible by the simulations.

MD allows a correct exploration of the microstates of the system whenever the conditions of the ergodic theorem, in particular that the conformational space is connected, are satisfied. However, numerous processes of interest exhibit the property of hindered ergodicity. Examples of processes and ways that ergodicity can be hindered are, for example, activated processes or processes exhibiting slow diffusivity. In activated processes the problem with achieving ergodicity arises if any of the basins of the potential energy surface are separated by a large energy barrier. In this case, if the kinetic energy of the system does not suffice, the barrier can be seemingly impossible to overcome, and the states beyond this barrier cannot be sampled.

To overcome these barriers computational chemistry uses different methods to enhance the sampling. In this thesis I employed constrained molecular dynamics, metadynamics, and simulated annealing. These methods allow the correctly weighted sampling of the states, and as a consequence, the possibility of estimating the free energy landscape of the underlying processes.

### 2.7.1. Constrained Molecular Dynamics

In Constrained Molecular Dynamics (CMD) sampling of rare events is obtained rigidly constraining the system to specific values of a prede-

## METHODS

---

terminated reaction coordinate,<sup>[59][60]</sup> while the remainder of the system is left free to move. The constrained force  $F$  is collected over multiple MD runs at different values of the reaction coordinate  $R$ . The free energy along  $R$  is obtained as the integration of the time-averaged forces  $\bar{F}$  along the reaction coordinate from initial configuration  $R_0$  to final configuration  $R_f$ :

$$\Delta G(R) = \int_{R_0}^{R_f} \bar{F}(R') dR' \quad (2.7.1)$$

For an interatomic constraint, as employed in this work, the force on the constraint is equal to the Lagrangian multiplier:

$$\bar{F}(R) = \langle \lambda(R) \rangle = - \left\langle \frac{\delta V}{\delta r_{AB}} - \frac{2k_B T}{r_{AB}} \right\rangle_{r_{AB}=R} \quad (2.7.2)$$

where the brackets indicate the ensemble average at each value of  $R$ , which is defined by the distance between atoms A and B. This integration scheme can be applied independently of the nature of the interatomic forces, and has been successful in the description of different (bio)chemical events.<sup>[61,62]</sup> In practice the value of  $R$  is varied from  $R_0$  to  $R_f$  incrementally. The size of these increments determines the computational cost and the accuracy of the simulation.

### 2.7.2. Metadynamics

Metadynamics is an enhanced sampling algorithm which uses a time dependent conformational bias based on the states previously sampled in the simulation.<sup>[17,63]</sup> Using metadynamics is especially suitable if one has a general understanding of the reactant and product states of a process, but the intermediate states are not precisely known, or there may be competitive pathways.

In Metadynamics the external bias is applied to a set of order parameters that take the name of collective variables (CVs).

An appropriate choice of the CVs is produced when the reactant, the product and the intermediate states take different values of the CVs. The number of CVs is optimally kept as low as possible to prevent too high computational costs. CVs are generally any slow varying parameter that can quantify the reaction and can be evaluated during the simulation such as coordination number, interatomic distance, torsion angle etc.

Methodologically, the Potential energy surface (PES) is projected into a Free Energy Surface (FES) in the collective variable space. The forces for exploring the FES are

$$F_i^t = -\delta \mathcal{F} / \delta \sigma^{t_i} \quad (2.7.3)$$

where  $s_i, i = 1, n$  are the collective variables and  $\mathcal{F}$  is the free energy of the system at time  $t$ . The simulation is then iterated using the following procedure:

- The system is allowed to evolve for certain number of MD steps under the
- The forces are averaged over time  $F^t = \langle F \rangle$
- The new collective variable become  $\sigma^{t+1} = \sigma^t + \delta\sigma \frac{\psi^t}{|\psi^t|}$
- A time-dependent bias potential is applied to the system every nth step.

with  $\psi$  and  $\sigma$  the scaled forces and scaled collective variables, respectively, with respect to the estimated size  $\Delta s$  of the FES projected on the CVs with  $\sigma^t = \frac{s^t}{\Delta s}$  and  $\psi^t = \frac{s^t}{\Delta s}$ .

The added bias potential is of the form:

$$V(s, t) = \omega e^{\frac{|\sigma - \sigma^t|^2}{2(\delta\sigma)^2}}$$

The external potential acting on the system in every consecutive time step  $t$  is given by:

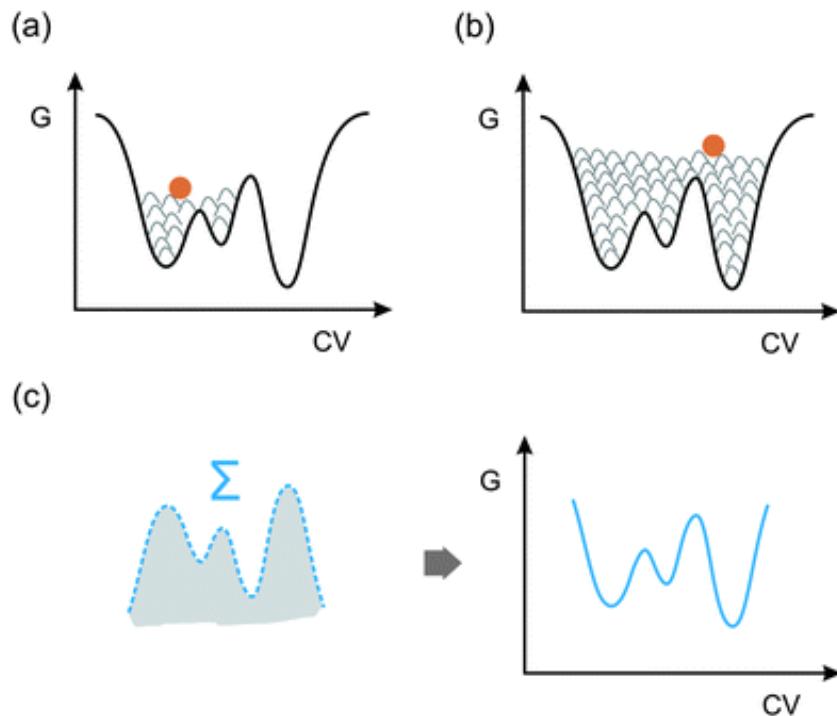
$$V(s, t) = \omega \sum_{t' \leq t} \prod_i e^{\frac{|\sigma_i - \sigma_i^{t'}|^2}{2(\delta\sigma)^2}}$$

## METHODS

---

Here  $\omega$  is the Gaussian height,  $\delta\sigma$  the Gaussian width.

By depositing these Gaussian potentials the simulation gains a "memory" of the previously sampled conformations and consecutively and systematically lowers the likelihood of visiting the same state again. Another advantage of Metadynamics is that the most relevant states, the low energy states, are sampled multiple times. Therefore, errors caused by using a more volatile CV and not directly sampling the absolute lowest energy geometry right away tend to even out over the course of the simulation.



**Figure 2.2.** – (a) The free energy landscape is explored by adding a Gaussian bias potential at the current position, orange dot, of the simulation in CV space. (b) The added bias potential makes the simulation able to overcome even large free energy barriers. (c) After convergence of the simulation, the free energy landscape is reconstructed from the sum of the added bias potentials. The graphic was taken from<sup>[64]</sup>.

The metadynamics run is converged when the added Gaussians are an inverse of the FES and the resulting FES landscape becomes entirely

flat within the limits of the Gaussian height  $\delta\omega$ . Convergence can thus be checked by the simulations achieving random diffusion over the CV-space, since there is no more bias of the FES surface on preferentially exploratory geometries. The number of Gaussian needed to fill the FES is therefore proportional to  $\sim (1/\delta\sigma)^n$ , where n is the number of CVs used. The Gaussian height and width must be varied according to the system at hand. Too large Gaussians will result in jumps between energy wells and non-flat final FES, while too small Gaussians will unnecessarily prolong the computational time needed. The width of the Gaussian must be chosen in accordance to the height selected to fit the shape of the underlying potential. For example steep wells require more conical Gaussians, whereas flat wells require flat Gaussians.

### 2.7.3. Simulated Annealing

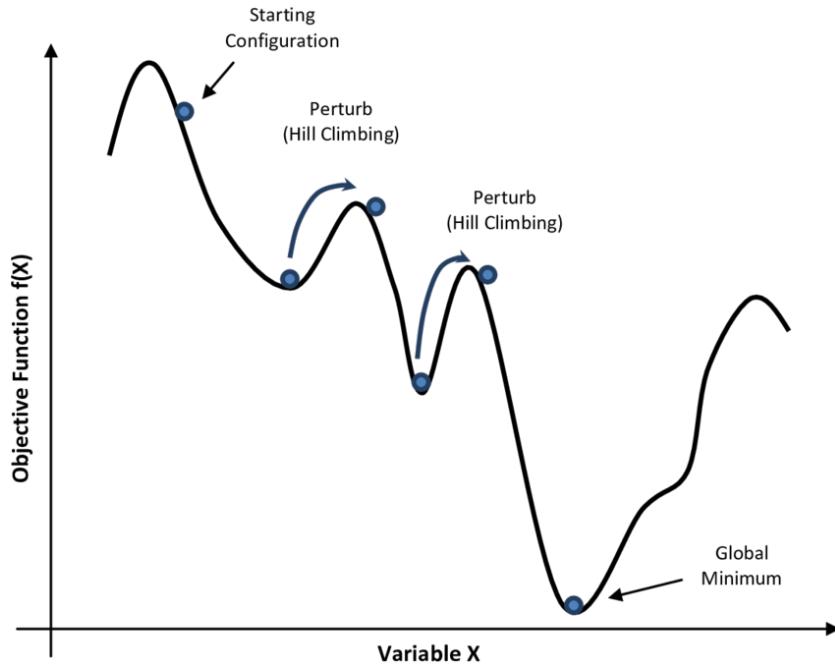
Simulated Annealing (SA) is the technique of systematically heating and cooling a simulation in order to obtain better sampling of the global energy landscape. SA gains its name from the metallurgic process in which a material is heated and cooled to systematically enlarge crystal size and reduce the amount of defects. The desired final structure of a unicrystal corresponds to the global structural (energy) minimum. Simple SA works by slowly lowering the overall simulation temperature over a series of coupled heating and cooling steps. Figure 2.3 The idea underlying SA is the one of importance sampling. Importance sampling probabilistically favors states of lower energy.

SA uses higher intermediate temperatures to sample higher-energy states that must be overcome to connect local minima and the global minimum. The basic SA algorithm can guarantee convergence to the global minimum, but the speed of convergence might be very slow, dependent on the system.<sup>[65]</sup> More specific SA algorithms, such as Generalized Simulated Annealing (GSA)<sup>[66]</sup> and Fast Simulated Annealing<sup>[67]</sup> have been derived to improve computational efficiency. GSA for example has been shown to improve efficiency for systems of

## METHODS

---

high dimensionality such as protein folding studies.<sup>[68]</sup>



**Figure 2.3.** – The simulated annealing procedure for a one-dimensional free energy surface. Figure taken from<sup>[69]</sup>.

### 3. The Schlenk Equilibrium of the Grignard Reagents

The Grignard reaction is a widely applied reaction in organic chemistry synthesis. During the Grignard reaction a carbon-carbon bond is formed between a nucleophilic organometallic halide of the formula R-Mg-X (R = organic residue, X = halogen) and an electrophilic carbon. The reaction was discovered in 1900 by Victor Grignard for which he was awarded the Nobel prize in 1912. To this day, the Grignard reaction has kept its relevancy due to its broad range of application under mild thermodynamic conditions. Moreover, the Grignard reagent (RMgX) is the stepping stone of many transmetallation reagents, which are involved in more selective and specific reactions. Prominent examples are the Stille<sup>[70]</sup>, Negishi<sup>[71]</sup> or Suzuki-Miyaura<sup>[72]</sup> coupling reactions. Nowadays organomagnesium reagents are of high interest because of their role in functionalization of phenol-based derivatives involving earth abundant metal catalysts and the development of efficient synthetic protocols allowing to work at very mild condition and with large functional tolerance.<sup>[73]</sup>

Because of its popularity, much effort has been devoted to the improvement and understanding of the Grignard reaction. Though the overall reaction equation is known, the exact intermediate and reactive states are still not entirely characterized. The reaction has multiple not fully characterized steps, including the multimerization state of the Grignard reagent, the identification of the most reactive species, and the mechanism of the final cross-coupling products.

In our first study, we tackle the problem of determining the chemical

## THE SCHLENK EQUILIBRIUM OF THE GRIGNARD REAGENTS

---

nature of the Gringard reagent dissolved in ether. It is known that the nominal Grignard reagent  $\text{RMgX}$  is, in fact, a mixture of different species at chemical equilibrium:



The first equilibrium is known as *dimerization equilibrium*, while the second equilibrium takes the name of *Schlenk equilibrium*<sup>[74]</sup>. Seyferth has described the difficulties in precisely capturing the distribution and nature of all species in scheme 1, as they vary in abundance subject both to the solvent, and the halide and organic groups bonded to Mg.<sup>[73]</sup> This has been evidenced by a multitude of different studies.<sup>[75-77]</sup> Solid-state structures found via X-ray diffraction studies have shown various coordination modes for the central Mg-atom as well as different possible nuclearities of the Grignard reagent. Ashby and Walker showed that a tetrahedral coordination is commonly observed in crystals structures obtained from ether solutions as  $\text{EtMgBr}(\text{OEt}_2)_2$ ,<sup>[78]</sup> where two molecules of solvent are bound to Mg. Instead, penta-coordination was found in trigonal bipyramidal and square pyramidal structure for  $\text{CH}_3\text{MgBrTHF}_3$  in THF solvent.<sup>[79]</sup> Further studies suggested that the underlying Schlenk equilibrium is more complex and involves structures such as  $\text{RMgCl}(\text{THF})_n$ ,  $\text{MgR}_2(\mu\text{-Cl})_2\text{Mg}(\text{THF})_{5/4}$  and  $\text{RMg}_2(\mu\text{-Cl})_3(\text{THF})_5$ .<sup>[80,81]</sup> Comparing these studies highlights how important correctly modelling the solvent is to fully capture the picture of the Grignard reagent. Due to its high reactivity Grignard reagent has to be handled in organic solvent, the most popular experimental and industrial ones are tetrahydrofuran and diethylether.

Ebullioscopic point measurement studies suggested that the solvent has a direct effect on the oligomerization of the Grignard reagents.<sup>[76]</sup> The same studies indicated that the product  $\text{MgX}_2$  and  $\text{MgR}_2$  species are mostly monomeric in THF.<sup>[76,82]</sup> Furthermore, studies using molecular weight,<sup>[83]</sup> calorimetric measurements,<sup>[84]</sup> and NMR and IR spectroscopy<sup>[77]</sup> support the existence of a multitude of different active species.

It is, at present, commonly understood that moieties in the generalized Schlenk Equilibrium can form dimers and higher order oligomers predominantly by bridging Mg-atoms through halide atoms. This causes the association between the various moieties, eventually favoring ligand exchange and in such progression of the Schlenk equilibrium.

### 3.1. Computational Background

The computational studies on the Grignard reaction available in the literature have been conducted mostly by means of second-order Møller-Plesset perturbation theory (MP2) [85,86][87] and density functional theory (DFT).<sup>[88–90][91]</sup> These studies showed that the dimerization equilibrium is strongly dependent on the solvent involved<sup>[86,88]</sup>. Furthermore, it was evidenced that the nature of the bridging ligand strongly influences the stability and the geometry of the dimer, with a general preference for two bridging ligands.<sup>[85,92]</sup> DFT studies also evidenced that the dimerisation degree decreases with increasing solvation.<sup>[90]</sup> Due to its large costs, computational studies have either taken into consideration ideal gas-phase reactions<sup>[93]</sup>, or included solvation by continuum models<sup>[94]</sup>. Typically, the solvent molecules explicitly included in the models have been restricted to the ones in the first solvation shell of the Mg atoms. In most cases, the total number of ligands to the metal center has been systematically fixed to four, which is the usual coordination number of Mg in metallo-organic compounds. Crystallographic studies<sup>[95,96]</sup> and recent computational data<sup>[89]</sup> have shown the existence of stable pentagonal and octahedral Magnesium halides with respectively three or four THF molecules in the solvation shell of the metal.

### 3.1.1. Computational Details and System Setup

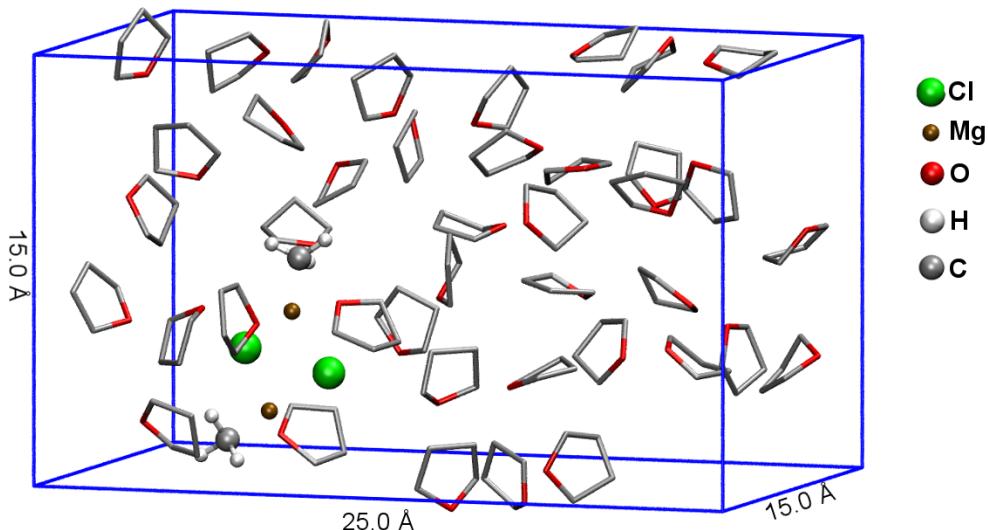
The previous crystallographic<sup>[95,97,98]</sup> and computational studies<sup>[89]</sup> computational clearly show that taking solvent into account explicitly is very important also beyond a tetrahedral geometry. In order to further investigate the dynamic and predominant species of the Schlenk-equilibrium there are multiple options to take advantage of the advancement in computational power achieved over the last decade. Firstly, one option is to cover a wide range of structures, halides and organic residues and investigating those using MP2 and DFT static calculations. The downside here lies within the cost associated when wanting to extend explicit solvation past the first solvation shell. Also the image obtained is static and transitions can only be investigated between predetermined structures. Secondly, there is investigation using ab initio molecular dynamics simulations (AIMD). The downside here lies with having our AIMD studies be limited to a specific Grignard reagent, if solvent effects past the first solvation shell shall be taken into account.

As our emphasis lies on extending the picture of the Schlenk equilibrium dynamics as well as in further investigating the role of the solvent, we chose the latter option, using DFT static calculations of selected structures, identified as Free Energy Surface (FES) minima by AIMD. AIMD has been shown to be especially appropriate to study processes in the liquid phase.<sup>[63,99]</sup> For example, this technique was used to study structural properties of various metal ions in solution, including Na,<sup>[100]</sup> Mg,<sup>[101,102]</sup> or Co and Ag.<sup>[103]</sup> Using dynamic simulation furthermore allows to explore the relevant conformations of the system rather than having to pre-select from the many different structures have been proposed for the Grignard reagent.<sup>[80,81,97,104]</sup> The system that I describe has around 500 atoms and when comparing to Figure 1.2 we find that this systems lies well within the scale of AIMD simulations. For the problem at hand it is not possible to simply sample the entire potential energy surface. Therefore, I employ the Metadynamics enhanced sampling method.

The combination of AIMD and Metadynamics has been effectively used to capture solvent effects in numerous organic<sup>[105,106]</sup> and organometallic reactions.<sup>[107,108]</sup> Larger metallo-organic systems have been treated coupling Metadynamics, AIMD and molecular mechanics schemes.<sup>[63]</sup> This approach was used to describe for example electronic solute-solvent coupling in photo-excitation dynamics of ruthenium metallorganic compounds,<sup>[109]</sup> the reaction mechanisms of hydrolysis in zinc  $\beta$ -lactamases,<sup>[110,111]</sup> the redox properties of copper in azurin,<sup>[112]</sup> the functional role of  $Mg^{2+}$  ions in ribonuclease H<sup>[113]</sup> and the design of organo-ruthenium anticancer complexes.<sup>[114]</sup> In our studies we chose the Grignard reagent MeMgCl to study with the help of AIMD and Metadynamics in THF solvent.

### 3.1.2. System Setup and Simulation Details

The Schlenk Equilibrium was investigated through characterizing the monomeric reactants, monomeric products and exploring the different dimeric species that make up the intermediates. Three different systems, containing each, one molecule of either CH<sub>3</sub>MgCl, MgCl<sub>2</sub> or Mg(CH<sub>3</sub>)<sub>2</sub> surrounded by 25 molecules of THF in a cubic, periodic box of dimensions 15.0 x 15.0 x 15.0 Å<sup>3</sup>, were built. The number of solvent molecules was set to match the experimental density of THF at room temperature.<sup>[115]</sup> The initial coordinates of THF were obtained from running a 20 ps-long equilibration in a similar box containing the pure solvent. The monomers replaced one solvent molecule each. Afterwards, all systems were relaxed for 15 ps in the microcanonical ensemble at energies matching an average temperature of 300 K. Relaxation was performed using the canonical sampling/velocity rescaling (CSVR) thermostat using a time constant of 10 fs. Equilibration was carried out until the temperature of the system oscillated around the target value of 300K. Production runs were simulated in the NVT ensemble at 300 K. We employed the Nose' -Hoover chain thermostat with a chain length of 3 and time constant 1 ps for the data production.<sup>[43,56,116]</sup>



**Figure 3.1.** – Simulation box used in this study. The atoms of Grignard reagent are represented by spheres and the THF solvent molecules by sticks in licorice and red. Hydrogen atoms of THF are not shown for clarity.

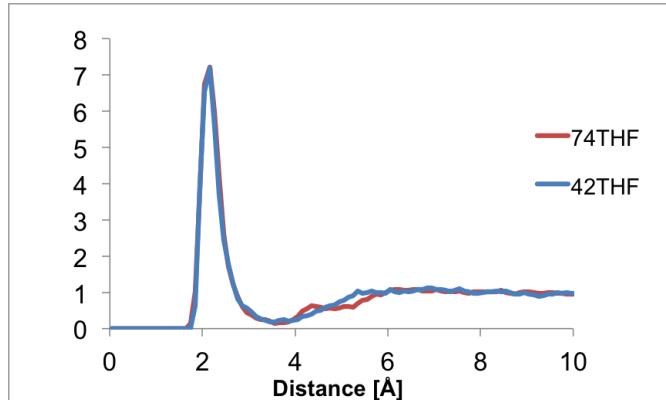
The dimeric  $(\text{MgCH}_3\text{Cl})_2$  species were simulated in an orthorhombic periodic box of dimensions  $25.2 \times 15.0 \times 15.0 \text{ \AA}^3$ , containing 42 THF molecules. The system contained initially two monomeric  $\text{MgCH}_3\text{Cl}$  molecules. Dimerisation was induced by application of a bias potential in 10 ps of NVE simulations. The systems were thermalized following the same protocol as described for the monomers. A larger box of dimensions  $25.2 \times 20.0 \times 20.0 \text{ \AA}^3$  with 74 THF molecules was also simulated to verify eventual bias due to finite-size effects. We checked three different parameters obtained from Metadynamics simulations of dimeric species in the respective boxes; the average geometry, radial distribution function of THF around the central Magnesium atoms and the THF diffusion coefficient. Firstly, the average geometry is captured in Table 3.1.

**Table 3.1.** – Geometric parameters for the system containing 42 and 74 THF molecules.

Bond distances [Å]	42-THF system	74-THF system
Mg-Cl (terminal)	2.38 (0.14)	2.36 (0.10)
Mg-Cl	2.59 (0.29)	2.56 (0.21)
Mg-CH <sub>3</sub> (terminal)	2.14 (0.10)	2.15 (0.10)
Mg-CH <sub>3</sub>	2.35 (0.20)	2.36 (0.20)
Mg-THF	2.28 (0.24)	2.30 (0.26)

Angles [°]	42-THF system	74-THF system
Mg-Cl-Mg	82.7 (6.0)	81.2 (5.9)
Mg-CH <sub>3</sub> -Mg	73.2 (6.5)	74.0 (6.4)

All differences in the geometrical parameters are within the standard deviation. Secondly, the radial distribution function is shown in Figure 3.2.



**Figure 3.2.** – Radial distribution function of THF around the two Magnesium centers for the system containing 74 THF in a  $25.2 \times 20 \times 20 \text{ \AA}^3$ , box compared to the system containing 42 THF in a  $25.2 \times 15 \times 15 \text{ \AA}^3$ , obtained from Metadynamics runs of the dimeric species. The distance plotted is the Magnesium to Oxygen of THF distance.

The agreement of the first peak and depth of the first minimum in Figure 3.2, shows how the Mg-coordination as well as the exchange rate

between solvating THF-molecules is practically unaffected by the size of the simulation box. Thirdly, The diffusion coefficient D for the THF molecules, defined as:

$$D = \frac{\delta\sigma^2(t)}{\delta t} \quad (3.1.1)$$

where  $\sigma^2$  is the mean square deviation of the center of mass of the THF molecules, is equal to  $D= 0.137 \text{ \AA}^2, \text{ ps}^{-1}$  (with a coefficient of determination for the linear fit  $R^2 = 0.976$ ) and  $D=0.153 \text{ \AA}^2, \text{ ps}^{-1}$  (with a  $R^2 = 0.91$ ), for the system with 42 and 74 THF.

These three parameters show that a simulation within the smaller simulation box is preferable as it lowers the computational cost while still correctly reproducing essential structural and dynamical features of the system.

Ab-initio Molecular Dynamics simulations The electronic problem was solved by DFT<sup>[5,6]</sup> using the Perdew-Burke-Ernzerhof (PBE) approximation to the exchange correlation functional.<sup>[43]</sup> Kohn-Sham orbitals were expanded over mixed Gaussian and plane-wave basis functions.<sup>[116]</sup> The DZVP basis set for first and second row elements and Mg, and a molecularly optimized basis set for the chlorine atoms were employed.<sup>[117]</sup> The auxiliary plane wave basis set was expanded to a 200 Ry cutoff. The core electrons were integrated out using pseudopotentials of the Goedecker-Teter-Hutter type. Dispersion forces were accounted for using the D3 Grimme approximation (GD3).<sup>[118]</sup> AIMD simulations were run over the Born-Oppenheimer surface, with a time-step of 0.25 fs, optimizing the energy gradient to a threshold of  $10^{-5}$  au.

**Free-energy calculations** The exploration of the conformational and reactive landscape, and determination of the corresponding free-energy surface, was performed by coupling AIMD to metadynamics simulations.<sup>[63,119]</sup> The solvation of different chemical species observed during the AIMD runs, as well as the transmetallation reaction, were investigated by independent metadynamics runs. All collective variables (CVs) employed in this study were defined as the coordination-number

of specific ligand species to individual Mg atoms. The coordination number of a species X around Mg at a given time t ( $CN[X](t)$ ) was evaluated according to the formula shown in equation 1, as defined in previous works:

$$CN[X](t) = \sum_i^{N_x} \frac{1 - \left(\frac{d_i(t)}{d_0}\right)^p}{1 - \left(\frac{d_i(t)}{d_0}\right)^q} \quad (3.1.2)$$

Where  $N_x$  is the number of species  $X$  present in the system,  $d_i$  is the distance of the  $i^{th}$  atom  $X$  from  $Mg$ , and  $d_0$ ,  $p$  and  $q$  are free parameters specific to each CV(see Chapter 7 Section 7.1). The time-dependent bias potential was formed by sets of Gaussians of 0.25 kcal mol<sup>-1</sup> height and 0.04 width for coordination variables and added every 50 steps of AIMD for the di-chloride bridged system. Additional details of the Metadynamics parameters used as well as the associated errors is given in Table 7.1 and Table 7.2 of Section 7.1. The metadynamics simulations convergence was confirmed by checking that the calculations reached the diffusion limit within their wall constraints; statistical errors, computed according to references,<sup>[120,121]</sup> are within 1-3 kT (see Table 7.2). AIMD runs were computed using the QUICKSTEP<sup>[17,117]</sup> module of the CP2K 2.5.1 package. Trajectory analysis was performed using the tools available in the VMD 1.9.2 package.<sup>[122]</sup>

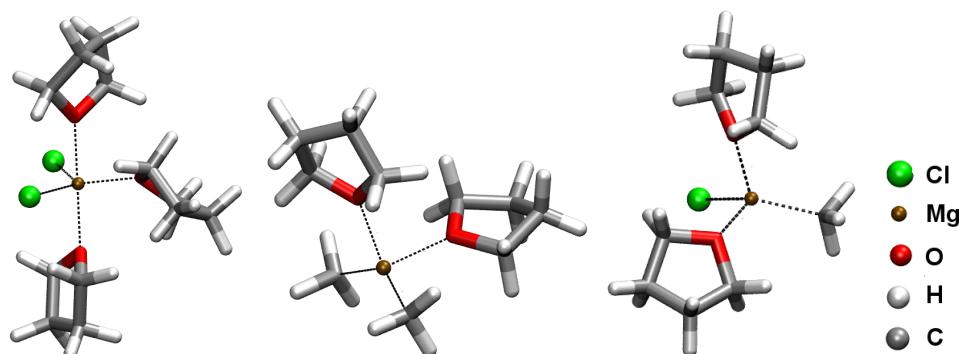
**Static calculations and electronic structure analysis** Chemically relevant geometries sampled by AIMD were fully optimized at the DFT(PBE+GD3) level by using the Gaussian09 software package.<sup>[123]</sup> Mg, C, H and O were described with the all-electron double- $\zeta$  6-31+G\*\* basis set.<sup>[124-126]</sup> Vibrational frequencies were computed analytically to verify that the stationary points found were energy minima. In addition to the solvent molecules bound to Mg, implicit solvation was modeled by using the density based solvation model (SMD).<sup>[127]</sup> In selected cases, Gibbs energies were obtained for T = 298.15 K and p = 1 atm. In the bimolecular steps, these energies were corrected for

the 1M standard state ( $T = 298.15\text{ K}$  and  $p = 24.465\text{ atm}$ ). Donor-acceptor interactions were explored by means of natural bond orbital calculations (NBO6 version).<sup>[128,129]</sup> The nature of these interactions was determined by computing the associated natural localized molecular orbitals (NLMO).<sup>[130]</sup>

## 3.2. Results

The investigation was restricted to monomers and di-nuclear species and we have excluded possible higher order aggregates as previous studies have evidenced that average aggregation state in THF is very low, with monomeric species being dominant.<sup>[76]</sup>

**Solvation states of the Monomeric structures** During 40 ps of AIMD simulations at 300K, both  $\text{CH}_3\text{MgCl}$  and  $(\text{CH}_3)_2\text{Mg}$  assumed solely a tetrahedral conformation, with two THF molecules in the solvation sphere of the Mg atom.



**Figure 3.3.** – Optimized structures of  $\text{MgCl}_2$  (left) and  $\text{MgMe}_2$  (middle) and  $\text{MgMeCl}$  (right) solvated in THF in Licorice representation. The dotted black lines represent coordination of the ligands to the Mg-center.

On the contrary, both tetrameric- and pentameric-coordination were observed for  $\text{MgCl}_2$ . The nature of the solvation of  $\text{MgCl}_2$  was further

investigated by metadynamics simulations, using the coordination number of oxygen atoms of THF to Mg as collective variable (CV). From our calculations, we estimate that the most likely solvation state is  $\text{MgCl}_2(\text{THF})_3$  arranged in a trigonal-bipyramidal structure (Figure 3.3 ).

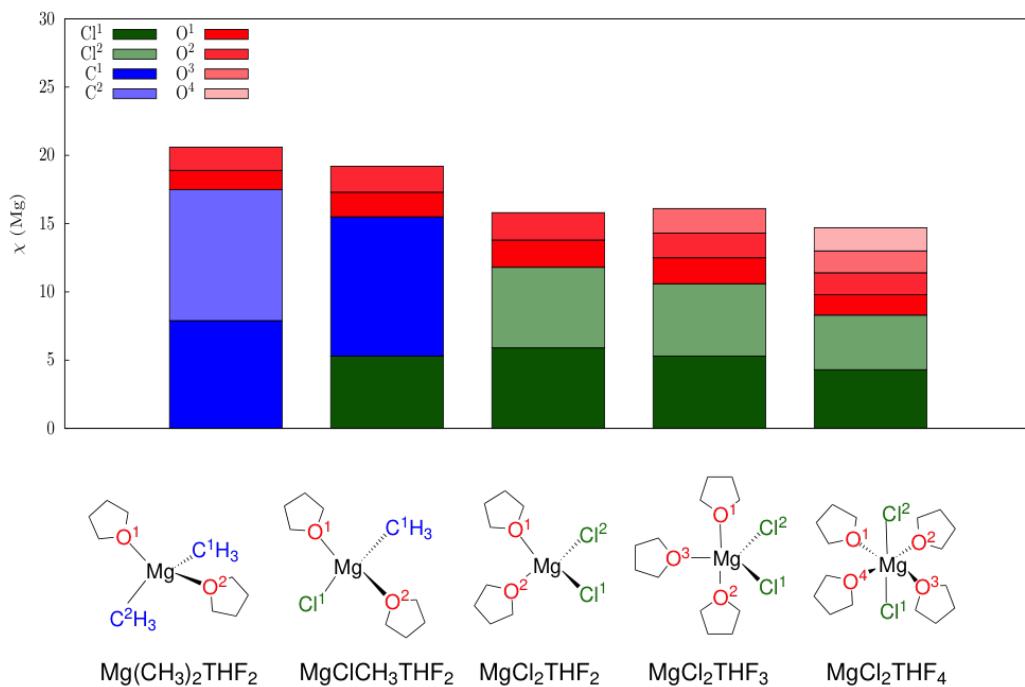
In fact the  $\text{MgCl}_2(\text{THF})_3$  trigonal-bipyramidal structure is about 1.6 kcal mol<sup>-1</sup> and 2.8 kcal mol<sup>-1</sup> more stable than the tetrahedral or octahedral complexes involving two and four THF molecules, respectively. The activation energy barriers separating the different solvation states are both less than 7 kcal mol<sup>-1</sup>. This allows for constant exchange between different solvation states at room temperature. The moieties are present in an equilibrium the ratio of 0.07:1:0.01 for the di-, tri- and tetra-solvated complexes, respectively. The average solvation state for  $\text{MgCl}_2(\text{THF})_x$  is given for x=2.9.

We further re-optimized the minima on the FES obtained from Metadynamics using single point DFT calculations. All solvation structures found on the FES also showed to be distinct minima on the PES. To obtain insight on the topological difference between  $\text{CH}_3\text{MgCl}$  and  $(\text{CH}_3)_2\text{Mg}$ , and  $\text{MgCl}_2(\text{THF})_x$  we used Natural Bond Orbital (NBO) analysis on the optimized structures. The Mg contribution on the lone pairs NLMO of  $\text{CH}_3$  and  $\text{Cl}, \zeta^{Mg}$ , is shown in Figure 3.4. This analysis showed that Cl had lower electron donation to Mg than  $\text{CH}_3$ . This is best evidenced by the difference in Mg contribution on the lone pairs NLMOs of  $\text{CH}_3$  and  $\text{Cl}, \zeta^{Mg}$ , which are 10.2% and 5.3%, respectively. Therefore, the  $\text{MgCl}_2(\text{THF})$  system is more likely to gain additional THF-coordination to improve screening of the Magnesium charge.

Our NBO Analysis reveals another reason for the stability of  $\text{MgCl}_2(\text{THF})_3$ . Adding one molecule of THF solvent to  $\text{MgCl}_2(\text{THF})_2$  increases the overall electron donation, i.e. stabilization of the positive charge on the central Mg-atom, whereas a second solvent molecule decreases the total electron donation, as the geometry becomes enlarged and the total electron donation from both THF-solvent as well as Cl-ligands decreases.

## THE SCHLENK EQUILIBRIUM OF THE GRIGNARD REAGENTS

---



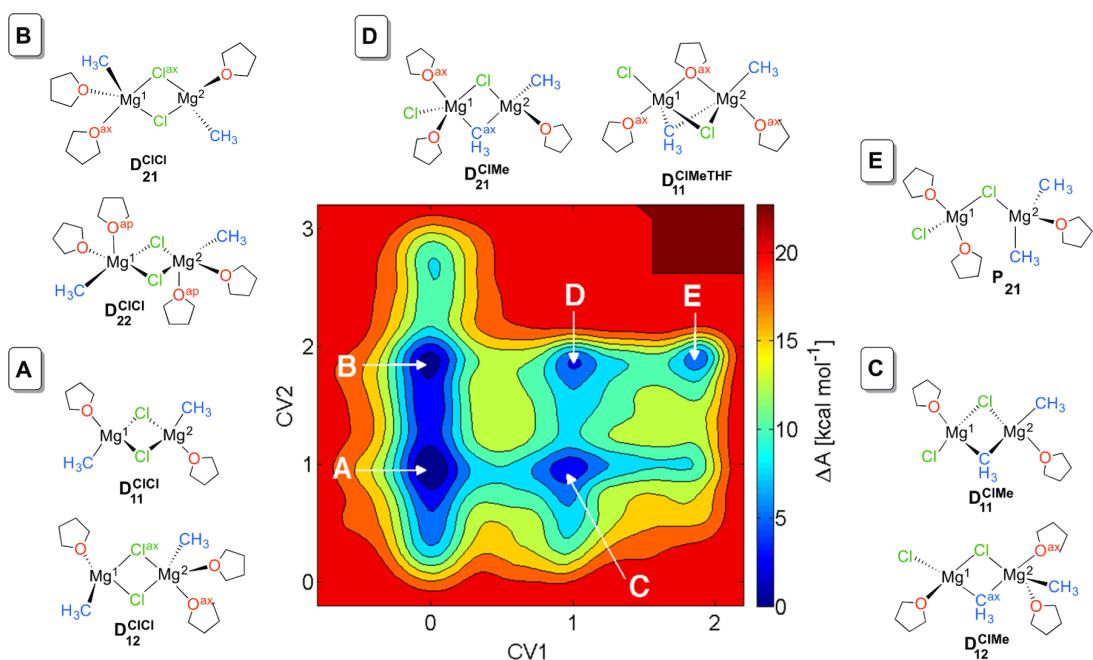
**Figure 3.4.** – Lone pair delocalization into empty orbitals of Mg. ( $\chi(Mg)$ )

**Solvation states of the Dimeric structures** The initial structure for investigation of the Dimeric structure was obtained by placing two molecules of Grignard reagent,  $\text{CH}_3\text{MgCl}$ , 9.5 Å apart in an orthorhombic box of  $25.2 \times 15.0 \times 15.0 \text{ \AA}^3$  dimensions. To accelerate dimerization a bias restraining potential was placed on the central Magnesium atoms. The dimeric structures bridged by two Chloride atoms, was formed as the distance between both Magnesium atoms approached 5.5 Å.

### 3.3. The Schlenk equilibrium

The Schlenk equilibrium is achieved via exchange of Chloride/organic group. In our case this is a methyl group that has to be exchanged to with a chloride atom to complete the Schlenk equilibrium. The dynamic processes of this exchange had been elusive and not much is known

about via which path of intermediates this exchange process occurs. Therefore, we started the investigation from the dichloride-bridged structure, that had been reported in multiple previous studies,<sup>[104,131]</sup> and used Metadynamics to explore the further FES landscape. We used as CVs the THF solvation number on one arbitrary Mg (CV1) and the difference in methyl coordination between the two Mg atoms (CV2). These CVs were chosen to project the free energy change as a function of methyl/chloride exchange and the local solvation of Mg.



**Figure 3.5.** – Free-energy surface of the Schlenk-Equilibrium. The collective variables for this representation are the difference in Mg-CH<sub>3</sub> coordination number between Mg<sup>2</sup> and Mg<sup>1</sup> (CV1) and the THF-coordination number to Mg<sup>1</sup> (CV2). The chemical structures drawn in the figure depict the most representative species obtained for wells A, B, C, D and E.

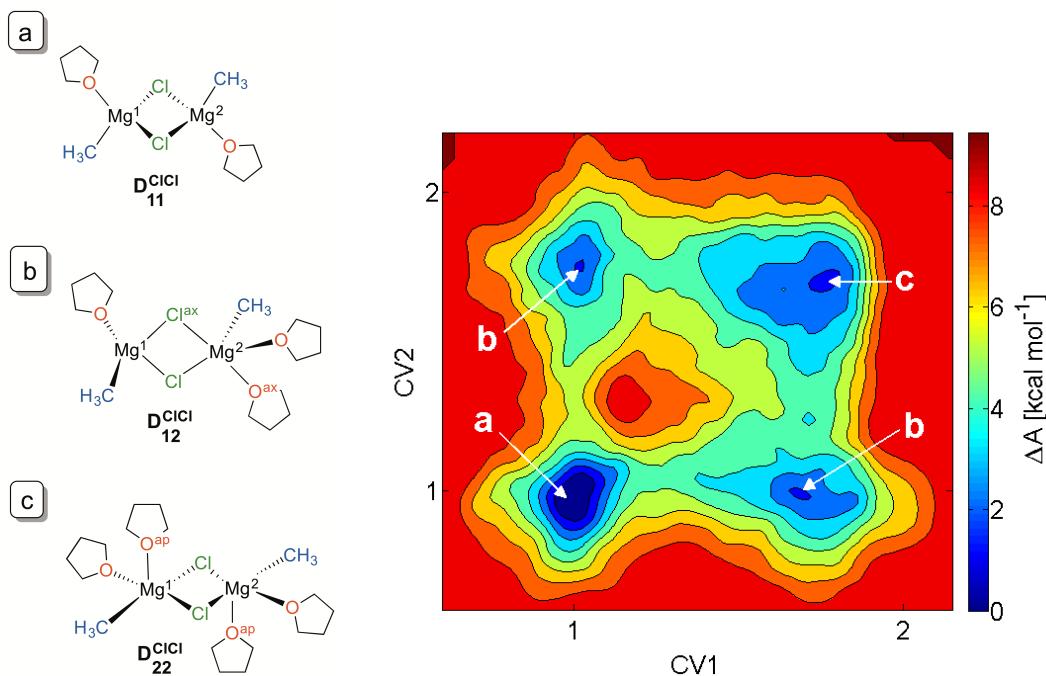
The resulting FES (Figure 3.5) shows five separated wells (A,B,C,D and E) that can be identified according to the nature of their Mg bridging groups. The most abundant structures classifying each well are shown alongside the FES. Wells A and B correspond to dichloro-bridged species, C and D correspond to methyl-chloride-bridged structures,

while well E represents monochloride bridged structures of the formula  $\text{ClMg}(\mu\text{-Cl})\text{MgMe}_2$ . The structures in well E are a shallow minimum that occurs prior to dissociation of the dimer into the product monomers  $\text{MgCl}_2$  and  $\text{Mg}(\text{CH}_3)_2$ .

The dimers shall henceforth be referred to via the notation  $D_{ij}^{XY}$ , where D stands for dimer,  $i$  and  $j$  are the number of THF molecules coordinated to  $\text{Mg}^1$  and  $\text{Mg}^2$ , respectively, while  $X$  and  $Y$  are the bridging ligands.  $D^{XY}$  describes the ensemble of all solvation structures with the same bridging ligands. From Figure 3.5 we can deduct, that ligand exchange process occurs via the path B-A-C-D-E. The transformation from Well B to D is statistically unlikely, because the energy barriers for the A-C and C-D transition are  $8 \text{ kcal mol}^{-1}$  and  $6 \text{ kcal mol}^{-1}$ , respectively, while the A-B transition is almost barrier-less. The B-D transition state, however, lies  $13 \text{ kcal mol}^{-1}$  higher than well A. The preference for this path was further investigated by closer examination of the solvation structures of  $D^{ClCl}$  and  $D^{ClMe}$ .

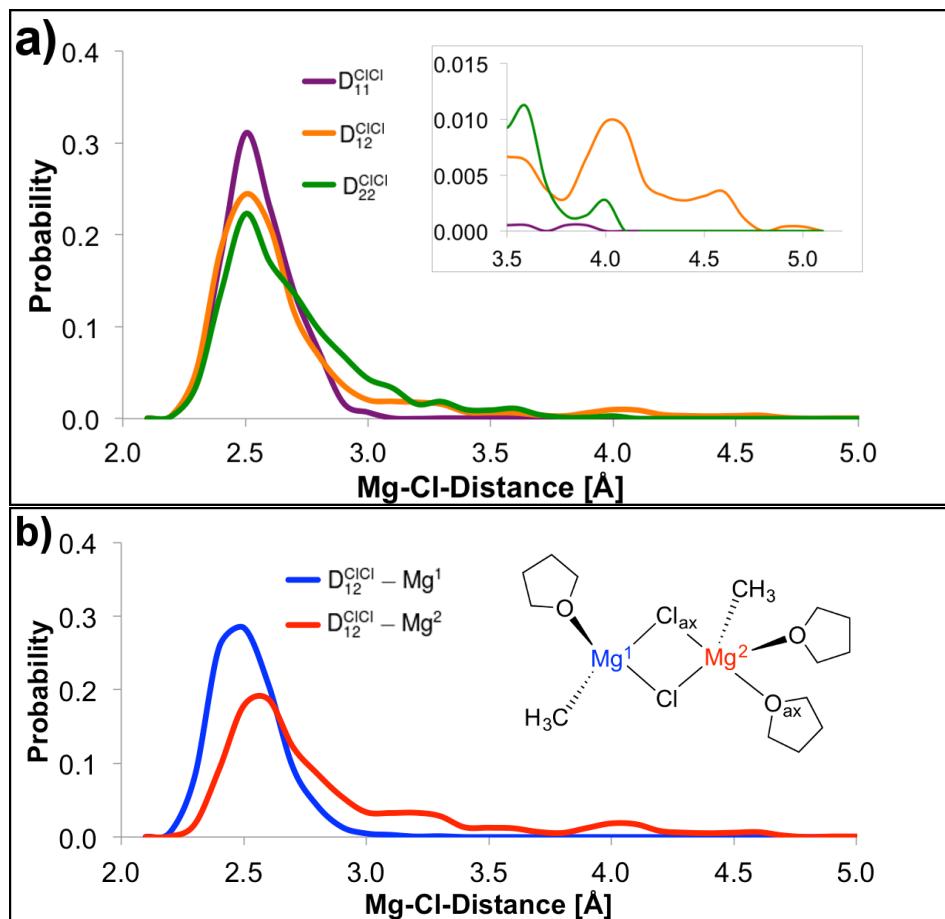
**The solvation structures of  $D^{ClCl}$**  In order to identify the individual structures captured in well A and B more closely (Figure 3.5), we used metadynamics using the coordination number of the oxygen atoms of THF to each of the Mg atoms as CVs. This ensures that all different solvation structures can be captured as individual FES wells.

The simulation revealed that there are three different solvation states, the resulting FES is shown in Figure 3.6. The lowest energy state is the symmetric dimeric structure in which each Mg atom is tetra-coordinated  $D_{11}^{ClCl}$ . This state is  $1.3 \text{ kcal mol}^{-1}$  and  $2.7 \text{ kcal mol}^{-1}$  more stable than  $D_{22}^{ClCl}$  and  $D_{12}^{ClCl}$ , respectively.  $D_{12}^{ClCl}$  and  $D_{21}^{ClCl}$  are degenerate states. Geometries with higher coordination of THF molecules could be observed during simulation, but they are more than  $5 \text{ kcal mol}^{-1}$  higher in energy than  $D_{11}^{ClCl}$  and therefore expected not to be present in relevant quantities at room temperature.



**Figure 3.6.** – Solvated  $D^{ClCl}$  structures found by metadynamics simulations (left) and corresponding FES (right). CV1 and CV2 are defined as the coordination numbers of THF at  $Mg^1$  and  $Mg^2$ , respectively, following Equation 3.1.2

The Grignard reaction is usually carried out at mild temperatures, therefore our investigation is restricted to simulations at 300K. In  $D_{22}^{ClCl}$  Mg is penta-coordinated, while in  $D_{12}^{ClCl}$  the structure is asymmetric. The activation energy between all those species is lower than 5 kcal mol<sup>-1</sup>, such that species are inter-converted rapidly at room temperature and we can assume equilibrium. Analysis of the interconversion from  $D_{11}^{ClCl}$  to  $D_{12}^{ClCl}$  reveals that the addition of THF strictly occurs anti to one of the chloride atoms, giving rise to a trigonal-bipyramidal geometry. Here the bridging Chloride atoms occupy an axial and an equatorial position each. The  $Cl_{ax}$ -Mg bond was significantly longer than the  $Cl_{eq}$  bond, resulting in transient, reversible cleavage of the  $Cl_{ax}$ -Mg. The resulting singly chloride-bridged structure could be observed during our AIMD trajectory. Statistical analysis of said AIMD trajectory is presented in Figure 3.7

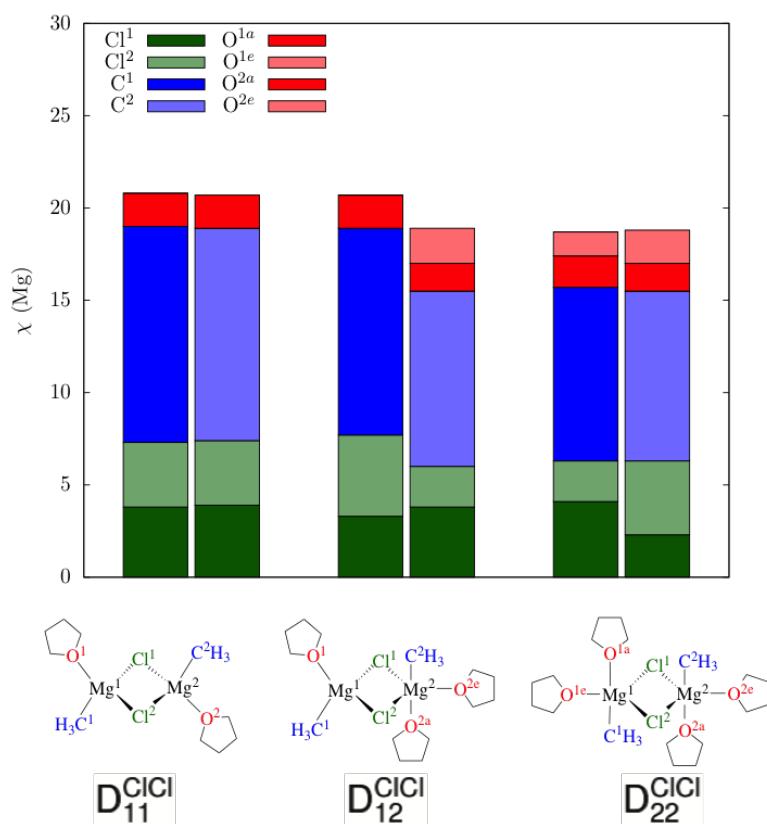


**Figure 3.7.** – (a) Mg-Cl distance distributions in  $D^{ClCl}$  dimeric structures. (b)  $Mg^1$ -Cl (blue) and  $Mg^2$ -Cl (red) distance distributions in  $D_{12}^{ClCl}$ . Mg-Cl bond cleavage is observed when the Mg-Cl distance is larger than 3.7 Å.

Expectedly, the increase in solvation is linked to an increase in average Mg-Cl bond distance. Interestingly, Mg-Cl bond cleavage is only observed for the asymmetric  $D_{12}^{ClCl}$  structure and not for the even more solvated  $D_{22}^{ClCl}$  structure. In confirmation with visual inspection of the AIMD trajectory bond cleavage is present for Mg-Cl bond distances above 3.7 Å. Bond distances lower than this can be attributed to molecular vibrations or geometrical rearrangement. The number of structures with bond distances above 3.7 Å for  $D_{11}^{ClCl}$  and  $D_{22}^{ClCl}$  are vanishingly small. Figure 3.7 b) shows that cleavage of the Mg-Cl bond also strictly occurs on the penta-coordinated Mg atom. Trajectory analysis confirms

that this bond breakage occurs solely with the  $\text{Cl}_{ax}$  bridging atom. Though the asymmetric structure is expected to be present only in 1% of the room-temperature samples, these monochloride bridged, unsymmetrically solvated structures are key to progression of the Schlenk equilibrium. We proposed that bond breaking constitutes the first step of ligand exchange.

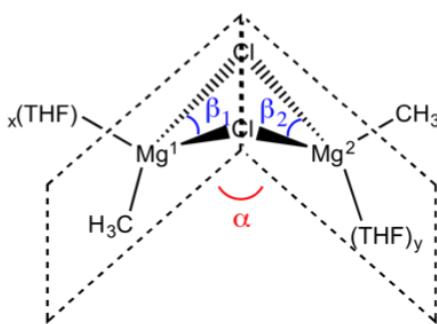
Due to the importance of this step we employed further NBO analysis of the  $D^{ClCl}$  structures and confirmed that pentacoordinated Mg atoms receive the least electron donation from Cl (Mg 6%), compared to (Mg 7.5%) in the tetrahedral case. Overall, as can be seen from Figure 3.8, the penta-solvated Mg atoms experience total electron donation.



**Figure 3.8.** – Lone pair delocalization into empty orbitals of Mg. ( $\chi(Mg)$ )

Especially the axial Chloride had very low contribution in the Mg lone orbitals. (Mg 2%). This in turn implies overpolarization of the  $\text{Cl}_{ax}\text{-Mg}^2$  bond and induces opening of the four-membered ring. The difference in electron between  $\text{Cl}_{ax}$  and  $\text{Cl}_{eq}$  also explains the selective  $\text{Cl}_{ax}\text{-Mg}^2$  bond cleavage.

Our DFT calculations predict that for a penta-coordinated Mg atom the interaction between Mg and the axial Chloride atom is weaker than with the equatorial Chloride atom. From NBO, the Mg- $\text{Cl}_{ax}$  in both  $D_{12}^{ClCl}$  and  $D_{22}^{ClCl}$  are predicted to be equally weak. We do not observe this behavior in our AIMD simulations. Instead of the trigonal-bipyramidal geometry observed for the penta-coordinated Mg atom in  $D_{12}^{ClCl}$ , the geometry of  $D_{22}^{ClCl}$  was observed to be square pyramidal in our AIMD simulations. Here a THF molecule occupies the apical position. Herein lies a discrepancy within the structures obtained from AIMD and the PES minima obtained from DFT optimization. While the geometrical features for  $D_{11}^{ClCl}$  and  $D_{12}^{ClCl}$  are reproduced by DFT optimizations, the geometries for  $D_{22}^{ClCl}$  differ. AIMD shows only distorted square-pyramidal geometries for the Mg-centers, while DFT predicts trigonal-bipyramidal geometries. This discrepancy on the solvent coordination is due two factors. On the one hand the explicit consideration of solvent beyond the first solvation shell has structural implications on the inner solvation shell. These effects become significant for the geometry of the more solvated and more flexible  $D_{22}^{ClCl}$ . On the other hand dynamics show that solvent molecules in  $D_{22}^{ClCl}$  are interchanged through an associative mechanism via a transient octahedral structure. The dissociation of said octahedral structure naturally results into a square pyramidal geometry.

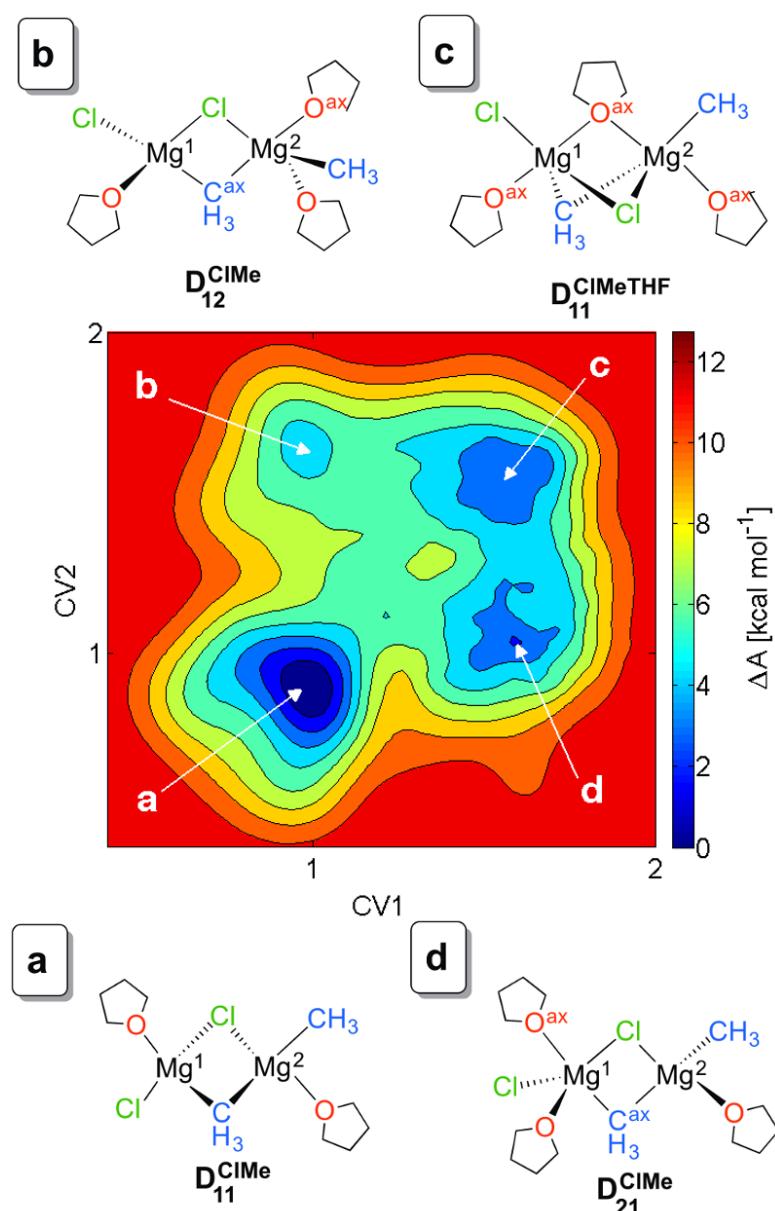


	$\alpha^{dyn}$	$\alpha^{st}$	$\beta_1^{dyn}$	$\beta_1^{st}$	$\beta_2^{dyn}$	$\beta_2^{st}$
D <sub>11</sub> <sup>ClCl</sup>	167(8)	169.7	89(5)	91.3	89(5)	91.6
D <sub>12</sub> <sup>ClCl</sup>	166(9)	165.4	91(7)	92.1	82(6)	83.9
D <sub>22</sub> <sup>ClCl</sup>	165(11)	160.6	94(6)	84.1	94(6)	84.2

**Figure 3.9.** – Average angles ( $\alpha^{dyn}$ ,  $\beta_1^{dyn}$  and  $\beta_2^{dyn}$ , in degrees) with its standard deviation (in parenthesis) obtained from cluster analysis of the metadynamics trajectory, and optimized angles from DFT calculations with implicit solvent ( $\alpha^{st}$ ,  $\beta_1^{st}$  and  $\beta_2^{st}$ , in degrees).

Geometry analysis show, that for D<sub>22</sub><sup>ClCl</sup>, D<sub>12</sub><sup>ClCl</sup> and D<sub>22</sub><sup>ClCl</sup> the equilibrium values for the Cl-Mg-Cl bond angle are close to 90° (Table 1), even though the coordination geometries of the Mg atoms in the structures are different from each other. As an additional structural feature, it is observed that, during AIMD simulations, the four atoms of the Mg( $\mu$ -Cl)<sub>2</sub>Mg moiety do not lie in the same plane. The average angles between the two MgCl<sub>2</sub> planes are reported in Figure 3.9.

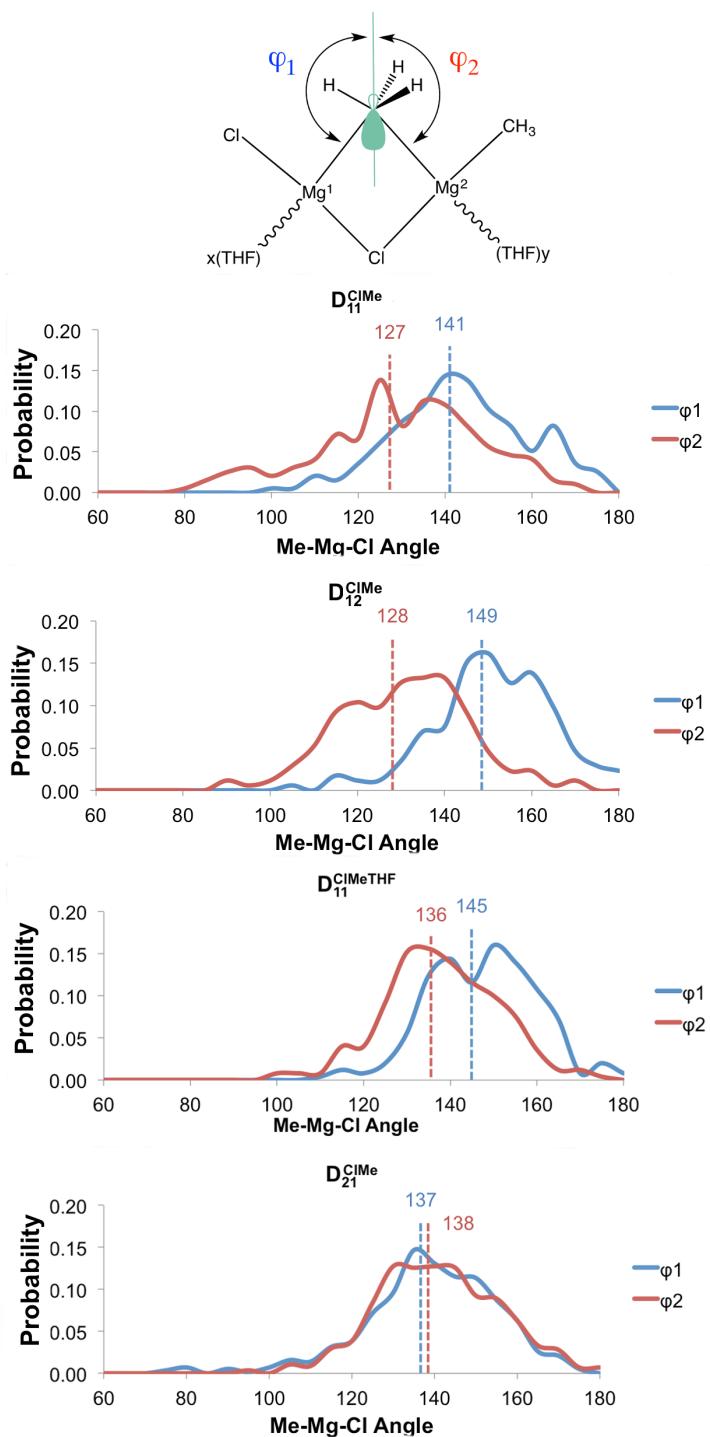
**The solvation structures of D<sup>ClCM<sub>e</sub></sup>** In order to identify the individual structures captured in Well C and D more closely (Figure 3.5), we used metadynamics using the coordination number of the oxygen atoms of THF to each of the Mg atoms as CVs. This ensures that all different solvation structures of the chloride/methyl-bridged structures can be captured as individual FES wells. Three low-energy solvation structures were observed on the FES, which is shown in Figure 3.10.



**Figure 3.10.** – Solvated  $D^{ClCl}$  structures found by metadynamics simulations (left) and corresponding FES (right). CV1 and CV2 are defined as the coordination numbers of THF at  $Mg^1$  and  $Mg^2$ , respectively, following Equation 3.1.2.

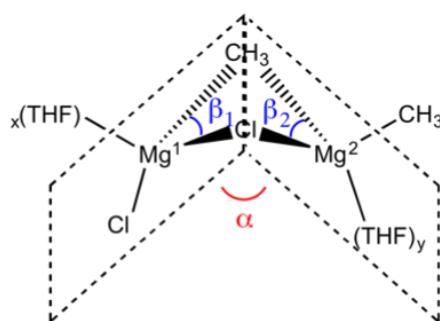
The lowest energy solvation structure is  $D_{11}^{ClMe}$  with both Mg centers being tetrahedrally coordinated. The other two low-energy structures

are  $D_{21}^{ClMe}$  and  $D_{11}^{ClMeTHF}$ , which are about 3 kcal mol<sup>-1</sup> higher in energy than the  $D_{11}^{ClMe}$ . The energy difference between those two structures is only 0.5 kcal mol<sup>-1</sup> and the activation energy for the transition between both states is about 2 kT.  $D_{11}^{ClTHFMe}$  is bridged not only by Cl and Me, but also by a THF molecule. A fourth structure  $D_{21}^{ClMe}$  was observed, but it lies about 5 kcal mol<sup>-1</sup> higher than  $D_{11}^{ClMe}$  and is therefore only present at trace amounts at room temperature. The barriers between  $D_{21}^{ClMe}$ ,  $D_{11}^{ClTHFMe}$  and  $D_{12}^{ClMe}$  are all less than 4 kcal mol<sup>-1</sup>, allowing for rapid exchange of the structures. The most  $D_{11}^{ClMe}$  is separated from these structures by a barrier of about 6 kcal mol<sup>-1</sup>. For the  $D^{ClMe}$  the tetra-solvated Mg-centers occupy a tetrahedral geometry, whereas the penta-solvated all assume trigonal-bipyramidal geometry. Since the structures are no longer point-symmetric about its center we sought after the implications of this asymmetry, which had shown to be quite essential for the dichloro-bridged structures. Firstly, during trajectory analysis, we noticed that the bridging methyl interacts differently with both Mg centers. Secondly, the FES in Figure 3.10 shows different accessible solvation states for both Mg-centers. Mg<sup>1</sup> having a terminal chloride, Mg<sup>2</sup> having a terminal methyl group. In order to quantify the way the bridging methyl group interacts with the Mg-centers, we measured the direction of the  $p_z$  orbital axis relative to the two C-Mg directions (Figure 3.11).



**Figure 3.11.** – Orientation of the methyl group in  $\text{DClMe}$  as a function of the solvation state, represented by  $\psi_1$  and  $\psi_2$ . A larger  $\psi$  angle is indicative of a stronger  $\text{Mg}-\text{CH}_3$  interaction.

In both  $D_{11}^{ClMe}$  and  $D_{12}^{ClMe}$ , the  $p_z$  orbital is oriented toward  $Mg^1\text{-Cl}$  moiety, with average angles for  $\psi_1$  of approximately  $141^\circ$  and  $149^\circ$ , respectively. In these two structures  $\psi_2$  has an average value of  $127^\circ$  and  $128^\circ$  indicating a poorer interaction of the bridging methyl with  $Mg^2$   $p_z$  orbital. In  $D_{11}^{ClTHFMe}$  the  $\psi_1$  and  $\psi_2$  average values are  $145^\circ$  and  $136^\circ$ , respectively. For  $D_{21}^{ClMe}$  the bridging methyl is almost equally shared between the two Mg center, with average  $\psi_1$  and  $\psi_2$  angles are  $137^\circ$ . The difference between the  $\psi$  values rises as THF-solvation is increased on  $Mg^2$  and decreases with increasing solvation on  $Mg_1$ . In order for the transformation from  $\text{CH}_3\text{MgCl}$  to  $\text{MgCl}_2$  to  $\text{Mg}(\text{CH}_3)_2$  to be complete the bridging methyl group has to be entirely transferred to  $Mg^2$ . As additional solvent increases the relative interaction with  $Mg^2$  it facilitate the transfer.



	$\alpha^{dyn}$	$\alpha^{st}$	$\beta_1^{dyn}$	$\beta_1^{st}$	$\beta_2^{dyn}$	$\beta_2^{st}$
$D_{11}^{ClMe}$	167(11)	175.6	101(10)	103.7	94(7)	100.0
$D_{12}^{ClMe}$	166(8)	179.4	102(7)	110.6	89(7)	93.2
$D_{11}^{ClTHFMe}$	165(6)	154.4	103(7)	102.0	96(7)	99.6
$D_{21}^{ClMe}$	161(7)	172.3	94(7)	100.4	101(7)	102.5

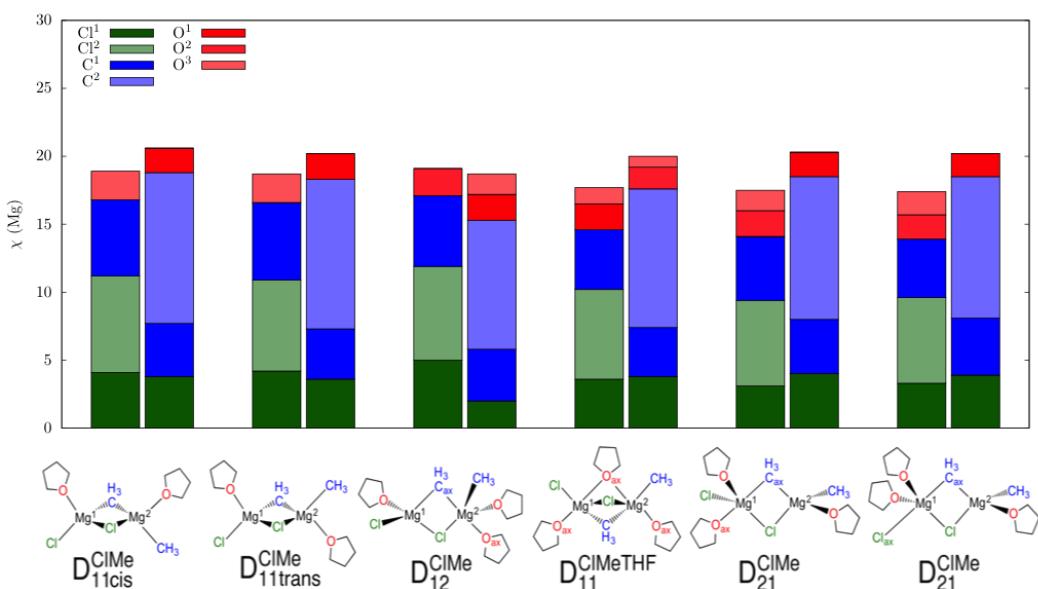
**Figure 3.12.** – Average angles ( $\alpha^{dyn}$ ,  $\beta_1^{dyn}$  and  $\beta_2^{dyn}$ , in degrees) with its standard deviation (in parenthesis) obtained from cluster analysis of the metadynamics trajectory, and optimized angles from DFT calculations with implicit solvent ( $\alpha^{st}$ ,  $\beta_1^{st}$  and  $\beta_2^{st}$ , in degrees).

In order to better understand the reason between the unequal sharing of the Methyl group and the role of the solvent, we undertook single point geometry optimizations on the PES using DFT methods for all 4 minima

## THE SCHLENK EQUILIBRIUM OF THE GRIGNARD REAGENTS

---

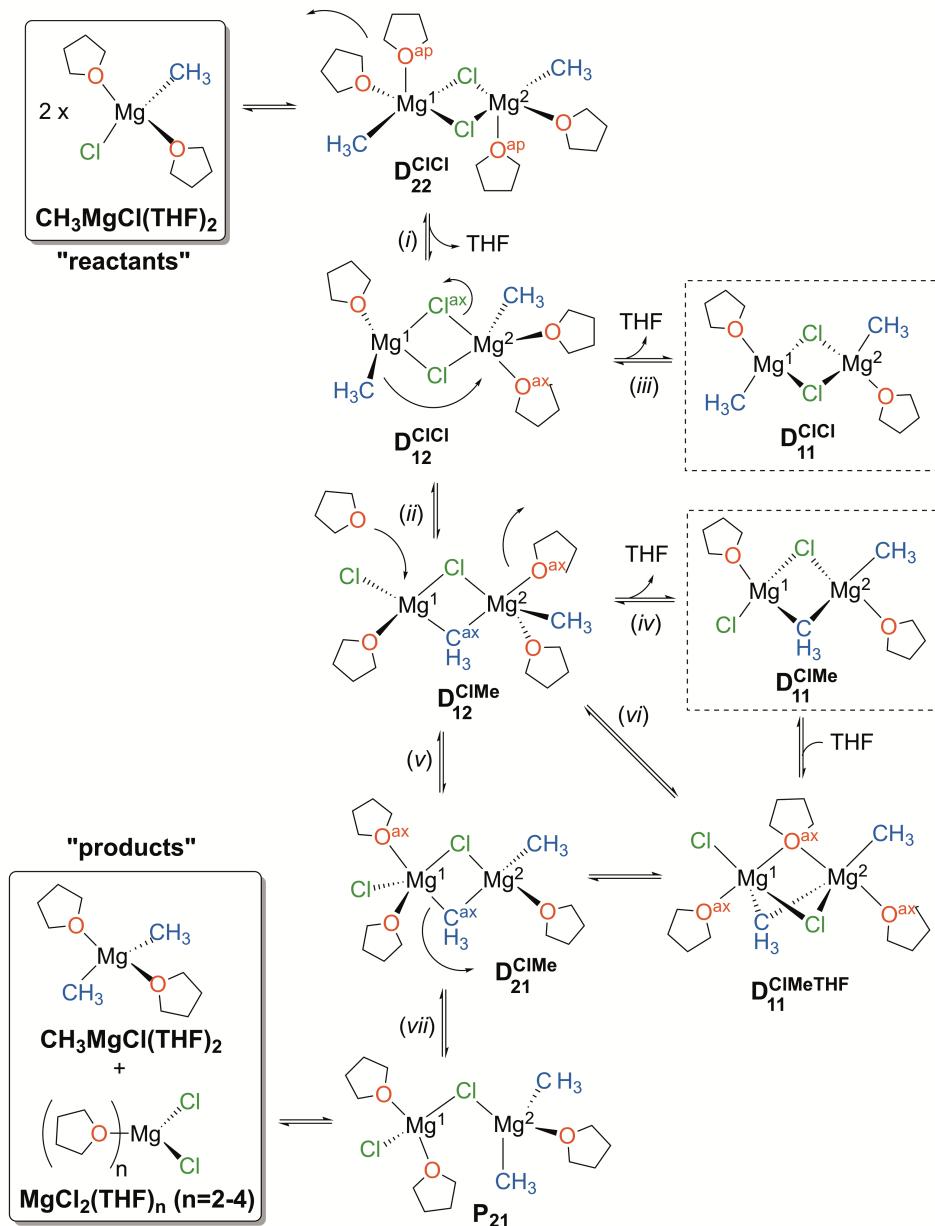
on the FES;  $D_{11}^{ClMe}$ ,  $D_{12}^{ClMe}$ ,  $D_{11}^{ClTHFMe}$  and  $D_{21}^{ClMe}$ . In agreement with the AIMD calculations all geometries were obtained as tetrahedral and trigonal bipyramids (Figure 3.12) for tetra-coordinated and penta-coordinated center. NBO analysis shows that the bridging  $\text{CH}_3$  group donates more strongly to  $\text{Mg}^1$  ( $\psi^{Mg1} = 5.7, 5.2$  and  $4.6\%$ ) than to  $\text{Mg}^2$  ( $\psi^{Mg2} = 3.7, 3.8$  and  $3.6\%$ ) for  $D_{11}^{ClMe}$ ,  $D_{12}^{ClMe}$ ,  $D_{11}^{ClTHFMe}$ . For  $D_{21}^{ClMe}$  the donation becomes equal among the two Mg centers ( $\psi^{Mg1} = 4.3\%$ ,  $\psi^{Mg2} = 4.2\%$ ). AIMD and DFT both predict  $D_{21}^{ClMe}$  to be the species most likely to yield complete transfer of the bridging methyl group to  $\text{Mg}^2$ . DFT calculations, shown Figure 3.13, furthermore predict that the geometry of the other ligands has an influence on the donation of the bridging Me-group, with an axial THF group providing optimal conditions for propagation of the Schlenk-equilibrium.



**Figure 3.13.** – Lone pair delocalization into empty orbitals of Mg. ( $\chi(\text{Mg})$ )

**The reaction pathway of the Schlenk equilibrium** Summarizing the findings of our 3 different metadynamics studies, the Schlenk equilibrium reaction pathway can be reconstructed. The pathway is summarized in Scheme 3.1, going from the Grignard reactant via the

dimeric intermediates to the product  $\text{MgCl}_2$  and  $\text{Mg}(\text{CH}_3)_2$ . All geometries shown in Scheme 3.1 are minima on the FES. From the scheme it is notable that the  $\text{Cl}/\text{CH}_3$  exchange only occurs via specific solvation states.



**Scheme 3.1** – Orientation of the methyl group in  $\text{DClMe}$  as a function of the solvation state, represented by  $\psi_1$  and  $\psi_2$ . A larger  $\psi$  angle is indicative of a stronger  $\text{Mg}-\text{CH}_3$  interaction.

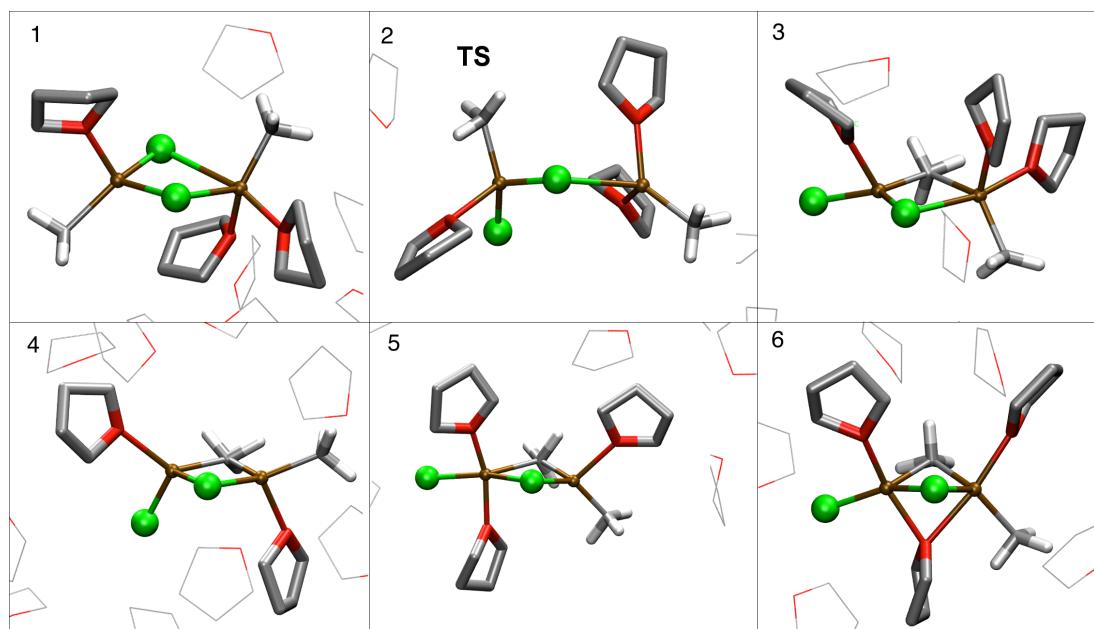
## THE SCHLENK EQUILIBRIUM OF THE GRIGNARD REAGENTS

---

The propagation of the equilibrium starts with the dimerization of the monomeric Grignard reagent  $\text{CH}_3\text{MgCl}(\text{THF}_2)$  into a dimeric, dichloride-bridged aggregate  $\text{D}_{22}^{\text{ClCl}}$ . We estimate that this reaction is endergonic by 4.9 kcal mol<sup>-1</sup> by static calculations. This observation is in alignment with previous experimental findings.<sup>[76]</sup>

The doubly penta-coordinated structure is energetically unfavored and desolvates twice to the most stable, dichloro-bridged structure,  $\text{D}_{11}^{\text{ClCl}}$ . After loss of the first solvating THF molecule an asymmetrical dimer  $\text{D}_{12}^{\text{ClCl}}$  is obtained. This structure lies about 2.7 kcal mol<sup>-1</sup> higher in energy than  $\text{D}_{21}^{\text{ClMe}}$  and is therefore also thermally accessible at room temperature. The asymmetry in solvation, with one Mg-atom being tetra-coordinated and the other Mg-atom being penta-coordinated yields interesting electronic properties.

Investigation by AIMD showed that the Mg-Cl bridging bond in  $\text{D}_{11}^{\text{ClCl}}$  is only transiently broken in the  $\text{D}_{12}^{\text{ClCl}}$  solvation state. NBO analysis confirms that here the Mg-Cl bond interaction is significantly weakened for the Mg atom solvated by 2 THF molecules. The higher solvation on  $\text{Mg}^2$  weakens the  $\text{Mg}^2\text{-Cl}$  bond, while the lower solvation on  $\text{Mg}^1$  makes it tolerant of an extra anionic ligand. Analysis of the AIMD trajectory further showed that during every bond Mg-Cl bond breakage of the  $\text{D}_{11}^{\text{ClCl}}$  system, the axial Mg-Cl bond of the penta-coordinated Mg-center in  $\text{D}_{12}^{\text{ClCl}}$  is broken. The bond cleavage is captured in snapshot 2 of Figure 3.14 and leads to ligand exchange and formation of the  $\text{D}_{12}^{\text{ClMe}}$  dimer.

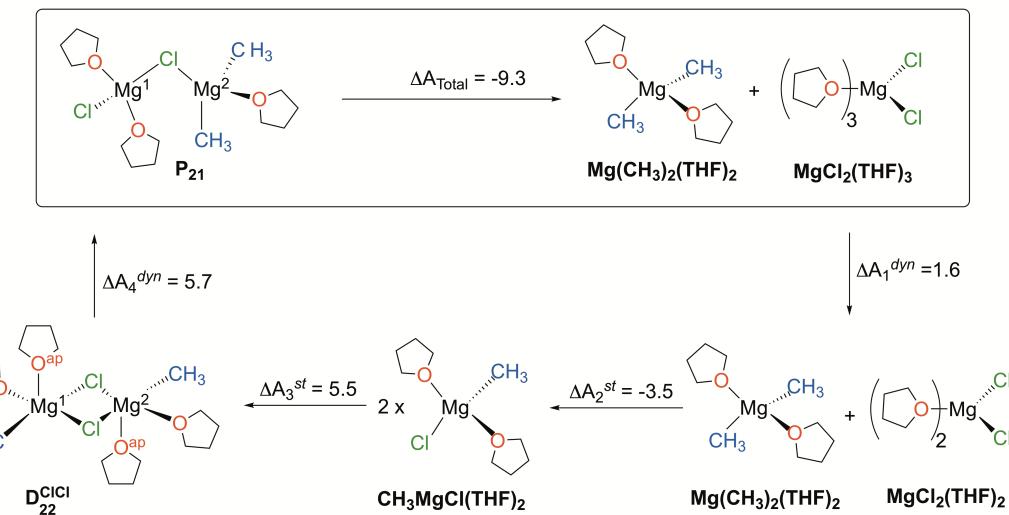


**Figure 3.14.** – Snapshots for the methyl transfer reaction in  $D_{12}^{ClCl}$  ( $Mg^1$  on the left hand side and  $Mg^2$  on the right for all snapshots): 1) initial  $D_{12}^{ClCl}$  structure, 2) Transition State of the transmetalation reaction, 3) formation of  $D_{12}^{ClMe}$ , 4) solvent loss to form  $D_{11}^{ClMe}$ , 5) solvent addition to form  $D_{21}^{ClMe}$ , and 6)  $D_{11}^{ClTHFMe}$ . The atoms for the Grignard reagent and the coordinating THF molecules are depicted in balls and/or sticks and colored according to standard color codes. Selected neighboring solvent molecules are drawn in thin lines.

A terminal  $CH_3$  group is exchanged with a bridging chloride atom In the process. (Snapshot 3 of Figure 3.14) The resulting  $D_{12}^{ClMe}$  is the least stable  $D^{ClMe}$  geometry and rapidly desolvates to the more stable  $D_{11}^{ClMe}$ ,  $D_{21}^{ClMe}$  and  $D_{11}^{ClTHFMe}$  solvation states.  $D_{11}^{ClMe}$  is the most stable solvation state, while  $D_{21}^{ClMe}$  and  $D_{11}^{ClTHFMe}$  are only about 3 kcal mol<sup>-1</sup> higher in energy and therefore thermally accessible at room temperature. In order for the transmetalation to complete and the products to be formed, the  $Mg^2$ -Cl and  $Mg^1$ - $CH_3$  bonds need to be cleaved. As the Mg-Cl bonds are stronger than the Mg- $CH_3$  bonds, the  $Mg^1$ - $CH_3$  has to be cleaved first. Analysis of the orientation of the bridging methyl-group and NBO analysis of the bridging lone pair revealed that increase in solvation is necessary for the methyl-group to be oriented toward away from  $Mg^1$ . The most favorable solvation state for this process is

## THE SCHLENK EQUILIBRIUM OF THE GRIGNARD REAGENTS

$D_{21}^{ClMe}$ , in which the bridging methyl group is equally shared. In this solvation structure the  $Mg^2\text{-CH}_3^{ax}$  is preferentially broken leading to the pre-product state,  $P_{12}$  of the Schlenk-equilbirum. This pre-product state evolves to the products  $MgCl_2$  and  $Mg(CH_3)_2$ , completing the Schlenk equilibrium reaction pathway. We estimate this part of the reaction to be very fast with a dissociation energy of about -8.7 kcal mol<sup>-1</sup>. The dissociation energy was estimated using the thermodynamic cycle depicted in scheme Scheme 3.2.



**Scheme 3.2** – Free energy differences  $\Delta A_1^{dyn}$  and  $\Delta A_4^{dyn}$  were determine from metadynamics runs described in the main text. Free energy differences  $\Delta A_2^{st}$  and  $\Delta A_3^{st}$  were determined from static calculations as described in the computational details.  $\Delta A_{Total}$  is obtained as :  $\Delta A_{Total} = \Delta A_1^{dyn} + \Delta A_2^{st} + \Delta A_3^{st} + \Delta A_4^{dyn}$ .

## 3.4. Conclusion

The Schlenk equilibrium proceeds from  $CH_3MgCl$  to  $MgCl_2$  and  $Mg(CH_3)_2$ . The reaction pathway proceeds via various dimeric adducts. We identified that the pathway does go via the previously identified, most stable tetra-coordinated Mg species. However, crucial for the propagation of the Schlenk equilibrium are asymmetrically solvated structures, as they induce bond-breaking between the Mg-centers and the bridging ligands.

Thus, the two-step Cl/CH<sub>3</sub> ligand exchange is only initiated by D<sub>12</sub><sup>ClCl</sup> and D<sub>21</sub><sup>ClMe</sup>. Our studies show that these intermediate structures are present at low abundance at room temperature and therefore crucial for understanding an efficient propagation of the Schlenk equilibrium. We further show that explicit consideration of the solvent is important to understand the dynamics of the Schlenk-Equilbirium.

# 4. Direct Nucleophilic Attack Vs. Radical Propagation Mechanism in the Grignard Reaction

## 4.1. Introduction

The Grignard reaction is a fundamental, text-book, process used in organic synthesis to promote the formation of carbon-carbon bonds.<sup>[132]</sup> The reaction involves the initial synthesis of the Grignard reagent, a magnesium-organic compound of general formula RMgX where R is a organic residue and X is a halogenous element (typically, Cl, or Br), and the subsequent reaction with an electrophilic species (most prominently, carbonyl groups: R'C=OR") to form adducts of general formula RR'R"CO-MgX:



The reaction occurs strictly in polar-aprotic solvents. Addition of water is used to hydrolyze the product, yielding the corresponding alcohol.<sup>[133]</sup>

Despite the fact that the Grignard reaction has been known since more than one hundred years, its characterization at a molecular level has remained elusive so far. Difficulties arise from the fact that the same molecular structure of the Grignard reagent is complex, with

several organometallic species co-existing in solution according to the *generalized Schlenk equilibrium*<sup>[74]</sup>:



As a matter of facts, the experimental characterization of such species, as well as the identification of the most reactive species for the Grignard equilibrium has resulted so far problematic.<sup>[73]</sup>

The relative abundance of the various chemical species can be modulated by varying the concentration, the organic residue or the halide, and the temperature. Several computational and experimental data available in the literature indicate that the equilibrium in equation 4.1.2 involves different intermediate states for the dimer dependent on the choice of the organic compound, the halide, and the solvent.<sup>[90]</sup>

The general understanding of the Grignard reaction (equation 4.1.1) points at a nucleophilic addition to an electrophilic substrate by a metal-activated nucleophilic carbon.<sup>[82]</sup> In fact, experiment with substrates characterized by aryl-conjugated carbonyl moieties yielded products that are not fully consistent with a nucleophilic addition, while they could be interpreted as the outcome of a radical-chain propagation mechanism. This was evidenced by the formation of pinacol structures in the reaction of benzophenone or quinones with Grignard reagent.<sup>[75,134–136]</sup> A series of studies by Ashby and co-workers further evidenced radical intermediates and further pointed at a possible competition between the two mechanisms, influenced by several factors among which the nature of the solvent, and the chemical characteristics of both the substrate and the residue. This competition between radical and non-radical pathway for different organomagnesium cross-coupling reactions could be demonstrated experimentally using chiral Grignard reagents.<sup>[137]</sup> However, it has not been possible to fully characterize the competing pathways on a molecular level.

Computer modeling is an excellent strategy to complement experimental data, providing clear insights about the molecular details of

## DIRECT NUCLEOPHILIC ATTACK VS. RADICAL PROPAGATION MECHANISM IN THE GRIGNARD REACTION

---

chemical processes in very highly controlled conditions. The constitution and reactions of Grignard reagents have been investigated in the past years by different approaches including by second-order Møller-Plesset perturbation theory (MP2) [85,86] and density functional theory (DFT). [88–90,131]

Computer modeling has also been successfully applied<sup>[87]</sup> to the study of the formation of the Grignard reagent, for which a radical pathway has been established.<sup>[138,139]</sup> Furthermore, computational studies have been used to investigate the character of the Grignard reaction on metallic surfaces using cluster models<sup>[140]</sup> as well as using implicit solvation by dimethylether solvent<sup>[131]</sup> using DFT calculations.

Recently, we have characterized in detail the Schlenk equilibrium for a model CH<sub>3</sub>MgCl Grignard reagent dissolved in tetrahydrofuran (THF). We determined that multiple chemical species, including, CH<sub>3</sub>MgCl, Mg(CH<sub>3</sub>)<sub>2</sub>, MgCl<sub>2</sub>, and higher-order aggregates, co-exists at room temperature. Moreover, we observed that, in the different compounds, the magnesium centers can accept a variable number of THF in their shell of first ligands, and that local changes in solvation are directly connected to the interchange between the chemical species at equilibrium. The accurate determination of the chemical species present in a solution of Grignard reagent provides an excellent starting point to investigate the Grignard reaction, and shed light on the conditions at which the two nucleophilic or radical mechanisms may be preferred or co-exist.

Here, we employ quantum mechanical calculations and ab initio molecular dynamics to investigate the likelihood of different reactive paths for the Grignard reaction in the presence of different substrates. We find that for highly conjugated systems such as fluorenone, that have the ability to stabilize a radical, the radical pathway is preferred. For non-conjugated systems, such as acetaldehyde, a non-radical pathway via nucleophilic addition is preferred. We furthermore find that the nucleophilic addition reaction shows a clear trend between the different species of the Schlenk-equilibrium. We find that the most reactive species in THF solvent are of the kind Mg(CH<sub>3</sub>)<sub>2</sub> and Mg(CH<sub>3</sub>)<sub>2</sub> ·

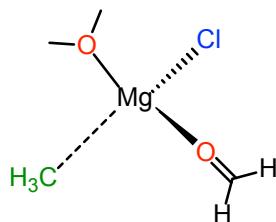
$\text{MgCl}_2$ .

## 4.2. Computational methods

### 4.2.1. Homolytic cleavage energies

**Benchmarking studies** Due to the large size of the systems under consideration in our study, we employ DFT to determine the electronic structure of different chemical species. Different exchange correlation (xc) functionals have been shown to yield discrepancies in the estimation thermochemistry and kinetics of chemical processes<sup>[141]</sup> and binding energies.<sup>[142]</sup> Therefore, we benchmarked the performance of different xc functionals against CCSD(T) data.

We benchmarked the quality of DFT calculations using PBE<sup>[43]</sup>, PBE0<sup>[143]</sup>, B3LYP<sup>[144,145]</sup>, M062X<sup>[146]</sup>, B97X<sup>[147]</sup> and TPSSH<sup>[148,149]</sup> exchange-correlation functionals (xc) in reproducing the homolytic C-Mg bond cleavage energy with respect to coupled-cluster (CC) data. The benchmark energies were computed over geometries optimized using the PBE0 functional for the system shown in Figure 4.1, considering both the whole system, and the  $\text{CH}_3\cdot$  and  $\text{RR}'\text{ClMg}\cdot$  fragments.



**Figure 4.1.** – Homolytic cleavage reaction that was used for calibration of the Exchange correlation method takes place along the dashed line.

The reference CC energies were obtained using single, double and perturbative triple excitations (CCSD(T)). CCSD(T) calculations were performed using the PTVZ basis set and frozen core approximation,

---

DIRECT NUCLEOPHILIC ATTACK VS. RADICAL  
PROPAGATION MECHANISM IN THE GRIGNARD REACTION

---

with an energy convergence cutoff to  $10^{-9}$  Hartree. The calculations were carried out using the CFOUR package.<sup>[150]</sup> The DFT energies were obtained using the 6-31+G(d,p) basis set using the Gaussian09 software package.<sup>[151]</sup>

Table Table 4.1 reports the homolytic dissociation energies of a model Grignard reagent bound to formaldehyde. The M062X xc functional is the best in reproducing CC data, in particular, showing a discrepancy not higher than 1.2 kcal mol<sup>-1</sup> in the dissociation energy. Re-optimization of the structures at the M062X level and consecutively obtaining the energy heightens the barrier by 1.7 kcal mol<sup>-1</sup>. The GGA PBE functional, whose reduced costs are necessary for the *ab initio* MD simulations, still show a reasonable performance, even though it tends to underestimate dissociation energies by 4.2 kcal mol<sup>-1</sup>.

**Table 4.1.** – Homolytic dissociation energies of a model Grignard reagent bound to Formaldehyde evaluated using different levels of theory according to Figure 4.1.

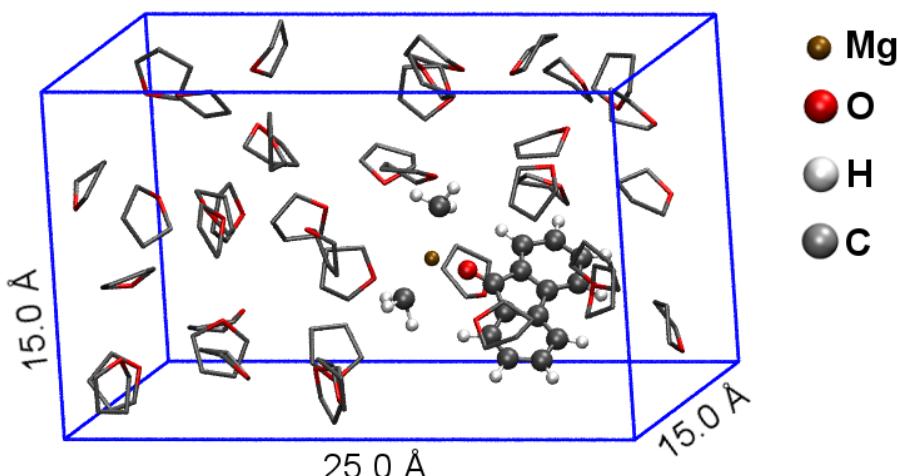
Functional	Basis Set	$\Delta E$ [kcal mol <sup>-1</sup> ]
CCSD(T)	DZ	43.64
CCSD(T)	TV	41.44
PBE0	6-31(d,p)	34.59
B3LYP	6-31(d,p)	32.55
M062X	6-31(d,p)	42.69
B97X	6-31(d,p)	39.39
TPSSh	6-31(d,p)	32.48
PBE	6-31(d,p)	37.22

If not explicitly stated otherwise, all electronic structures following in this paper have been optimized using the M062X functional and the 6-31 basis set. Dispersion forces were accounted for using the D3 Grimme approximation.<sup>[118]</sup> The calculations were performed in implicit THF solvent using the SMD solvent model.<sup>[127]</sup>

### 4.2.2. Ab initio molecular dynamics

**Monomeric species** We built two different monomeric systems, containing each one molecule of either  $\text{CH}_3\text{MgCl}$  or  $\text{Mg}(\text{CH}_3)_2$  surrounded by 41 molecules of tetrahydrofuran (THF) and one molecule of acetaldehyde in a periodic box of dimensions  $25.2 \times 15.0 \times 15.0 \text{ \AA}^3$ . The initial coordinates of THF were generated by 20 ps of ab-initio MD at the experimental density at room temperature.<sup>[115]</sup>

**Dimeric species** The  $(\text{MgCH}_3\text{Cl})_2$  species were simulated in an orthorhombic periodic box of dimensions  $25.2 \times 15.0 \times 15.0 \text{ \AA}^3$ , containing 41 THF molecules and one molecule of Acetaldehyde. Two systems that were obtained as equilibrium structures from our previous investigation<sup>[152]</sup> were used as starting geometries.



**Figure 4.2.** – Example simulation Box. Grignard reagent and Fluorenone in CPK, solvent in Licorice representation, the solvent hydrogen atoms are not shown in visualization.

**Simulation parameters** The electronic problem was solved by Density Functional Theory<sup>[5,6,116]</sup> using the Perdew-Burke-Ernzerhof exchange correlation functional (PBE)<sup>[43]</sup>. Kohn-Sham orbitals were

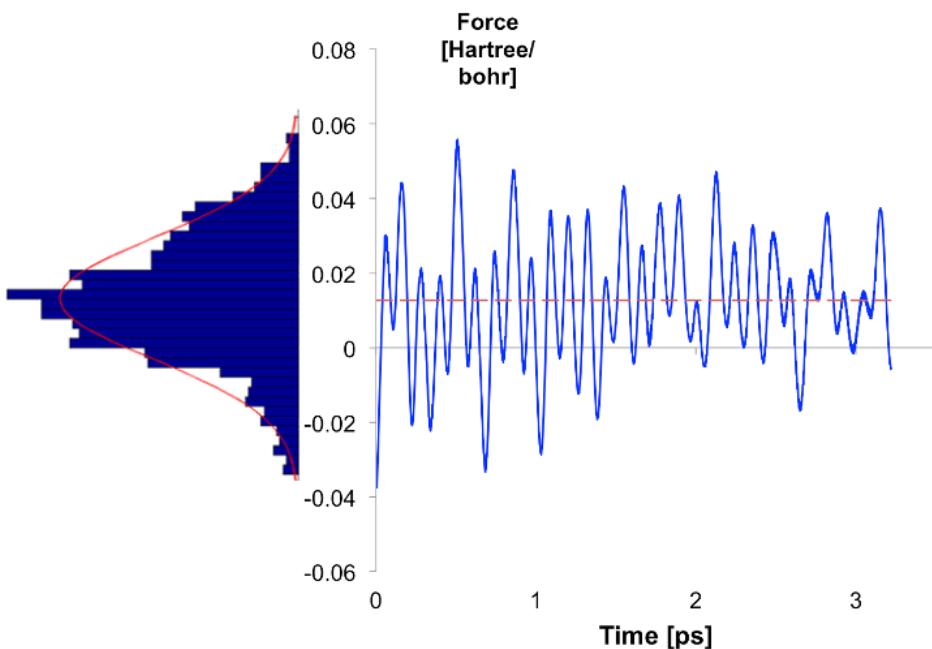
## DIRECT NUCLEOPHILIC ATTACK VS. RADICAL PROPAGATION MECHANISM IN THE GRIGNARD REACTION

---

expanded over mixed Gaussian and plane-wave basis functions. The DZVP basis set for first and second row elements, and a molecularly optimised (MOLOPT) basis set for the Chlorine atoms were employed<sup>[153]</sup>. The auxiliary plane-wave basis set was expanded to a 200 Ry cut-off. The core electrons were integrated out using pseudopotentials of the Goedecker-Teter-Hutter type (GTH).<sup>[154]</sup> Dispersion forces was accounted for using the D3 Grimme approximation.<sup>[118]</sup> Ab initio Molecular Dynamics simulations (AIMD) were run over the ground state potential energy surface, with a time-step of 0.25 fs, optimising the energy gradient to a threshold of  $10^{-5}$  au. All systems were first relaxed for 15 ps in the microcanonical ensemble at a average temperature of 300 K. Then, production runs were simulated in the NVT ensemble at 300 K. Relaxation at the target temperature was first performed using canonical sampling/velocity rescaling (CSVR) thermostat with a time constant of 10 fs until the temperature of the system oscillated around the target value. Then, a Nosé-Hoover chain thermostat with a chain length of 3 and time constant 1 ps was used for data production.<sup>[56–58]</sup>

Trajectory analysis was performed using the tools available in the VMD 1.9.2 package.<sup>[122]</sup>

**Constrained Ab initio Molecular Dynamics** In Constrained Molecular Dynamics (CMD) the reaction coordinate was constrained using the Shake algorithm. The Lagrangian Multipliers of the Shake Algorithm were collected every MD Step after equilibration and converged to a Normal distribution with a steady mean value within 1 ps of simulations. (Figure 4.3) The activation energy was computed by trapezoidal integration over the average Lagrangian multiplier from reactant to transition state, identified as the point along the reaction coordinate at which the average constraint force is zero. The reaction coordinate used for the investigation of the nucleophilic attack is the methyl carbon-carbonyl carbon distance.



**Figure 4.3.** – The Lagrangian Multipliers converged to a Normal distribution with a steady mean value within 1ps of simulations.

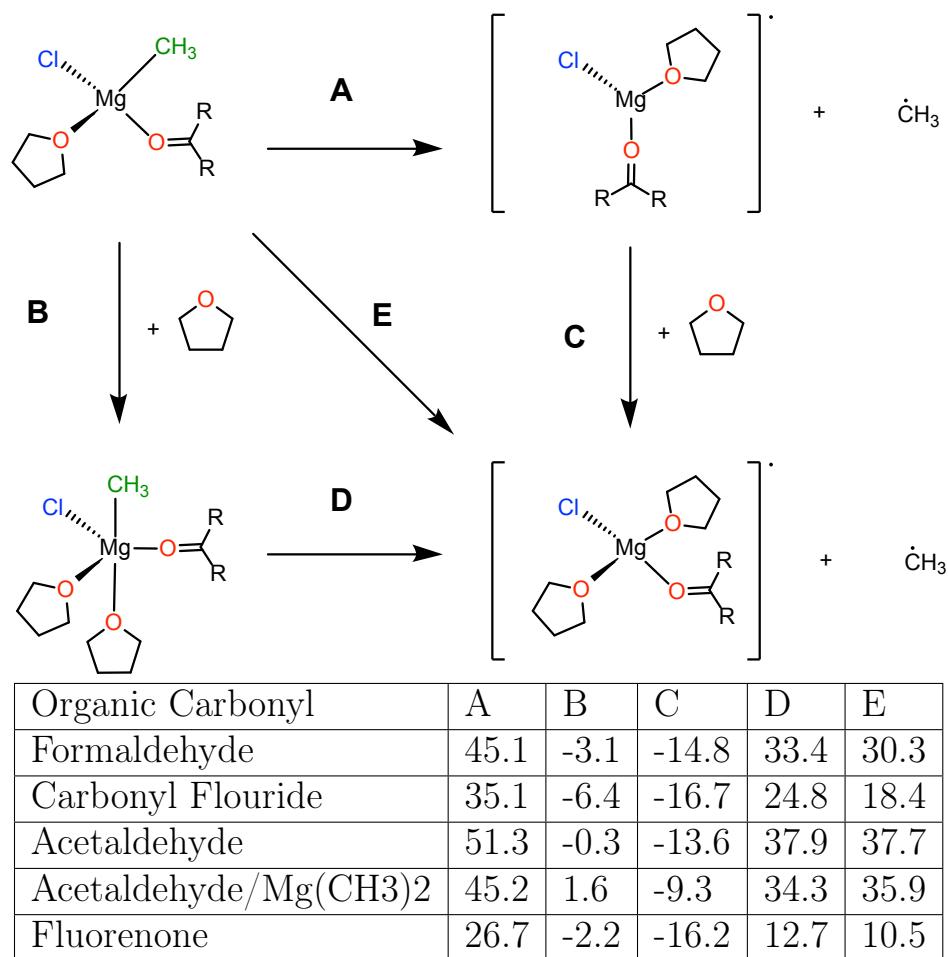
## 4.3. Results and Discussion

### 4.3.1. Homolytic bond cleavage

**Reactivity trends for different substrates** We investigated the change in the homolytic CH<sub>3</sub>-Mg cleavage energy when different substrates are bound to the Grignard reagent. Specifically, we compared formaldehyde, acetaldehyde, carbonyl flouride and fluorenone.

DIRECT NUCLEOPHILIC ATTACK VS. RADICAL PROPAGATION MECHANISM IN THE GRIGNARD REACTION

---

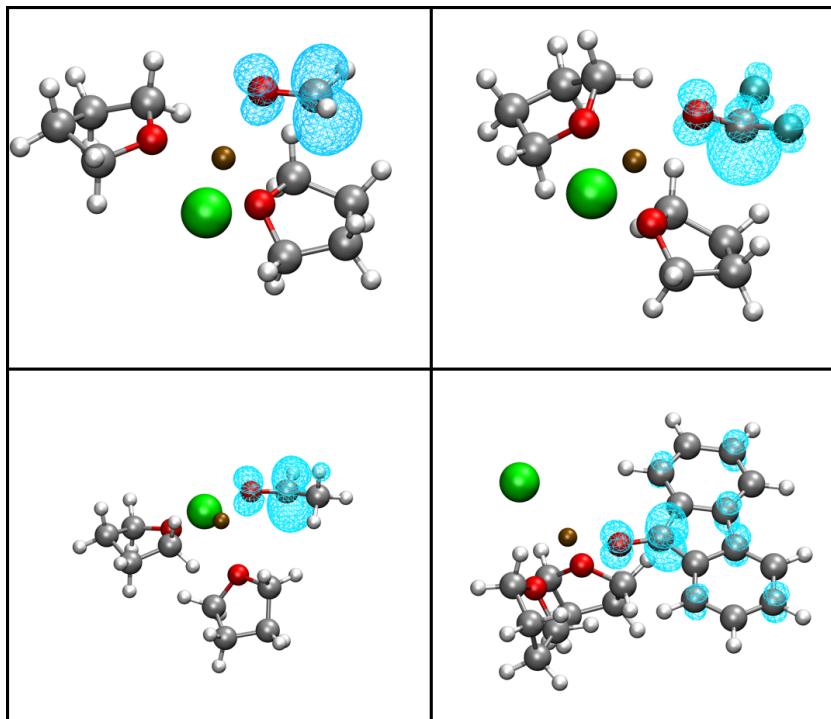


**Figure 4.4.** – Free energies ( $\Delta G$ ) for the monomeric Grignard reagent calculated according to the reaction scheme above, reported are in [kcal mol<sup>-1</sup>]. In the Acetaldehyde/Mg(CH<sub>3</sub>)<sub>2</sub> system the chloride is replaced with a methyl group.

Figure 4.4 reports the free energies for the homolytic cleavage reaction for the various systems computed as the difference between the change in the electronic energy and standard entropic corrections. In all cases, our data show that the complex between the Grignard reagent and the substrate coexists at room temperature with different solvation structures, yielding both a tetrahedral or pentahedral coordination for the Mg atom. On the contrary, the radical product exists strictly in a tetrahedral form, with the pentahedral coordination at least 10 kcal

$\text{mol}^{-1}$  higher in free energy for all species.

The presence of progressively more electron-withdrawing groups attached to the carbonyl ( $\text{CH}_3 < \text{H} < \text{F}$ ) facilitates the release of the  $\text{CH}_3\cdot$  radical from Mg. The reaction is even more facilitated by coupling the carbonyl to an aromatic moiety like in fluorenone.

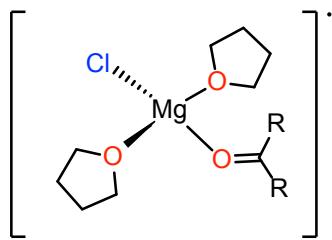


**Figure 4.5.** – Spin density of the homolytic cleavage product for (*Top Left*) Formaldehyde, (*Top Right*) Carbonyl Fluoride, (*Bottom Left*) Acetaldehyde and (*Bottom Right*) Fluorenone.

Figure 4.5 reports the localization of the spin density on the carbonyl moieties for the four different substrates. The presence of electron-releasing substituents forces a higher spin density localization on the carbonyl moiety; on the contrary, fluorine atoms facilitate its delocalization. This effect is prominent in fluorenone, where the unpaired electron is spread over the extended aromatic system.

## DIRECT NUCLEOPHILIC ATTACK VS. RADICAL PROPAGATION MECHANISM IN THE GRIGNARD REACTION

---

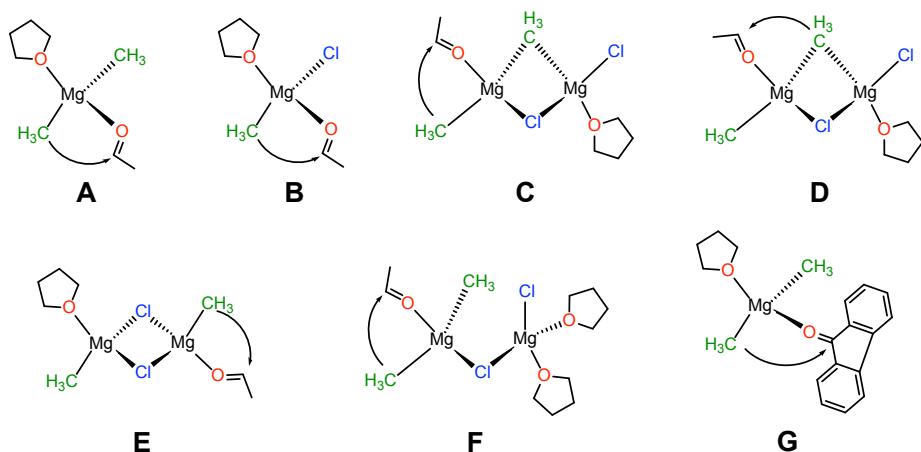


Organic Carbonyl	Carbonyl
Formaldehyde	1.04
Carbonyl Flouride	1.04
Acetaldehyde	1.15
Fluorenone	0.60

**Figure 4.6.** – Spin densities on the carbonyl carbon.

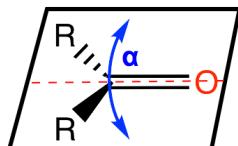
### 4.3.2. Nucleophilic attack

The energies reported in Figure 4.4 indicate that formation of a radical species is very unlikely for non-aromatic substrates. On the contrary, binding to fluorenone makes a homolytic cleavage of the C-Mg bond feasible for a relatively low energy. To determine the feasibility of the radical mechanism, we computed the activation energy for the competing nucleophilic attack pathway. The reacting complexes were built from the structures of the Grignard reagent obtained in a previous study<sup>[152]</sup>, and are shown in Figure 4.7.



**Figure 4.7.** – Structures investigated using Constrained Molecular Dynamics. The arrows indicate the path of nucleophilic attack investigated. The activation energy for the nucleophilic attack in each of the structures is detailed in Figure 4.8.

Figure 4.8 reports the activation free energies computed by CMD for the different complexes. We find that the more electron donating methyl ligands there are on the magnesium center, the lower the activation energy is for the nucleophilic attack. This trend in activation energy persists for monomeric as well as dimeric Grignard reagent species. Furthermore, the nucleophilic attack from a terminal ligand is preferred over the one from a bridging ligand (Figure 4.7). This behavior confirms our previous study where it was shown that the free electron pair of the bridging methyl group is shared between the magnesium centers and points away from the remaining ligands.<sup>[152]</sup>



System	A	B	C	D	E	F	G
Activation Energy [kcal mol <sup>-1</sup> ]	6.6	8.7	12.9	22.6	13.1	6.4	14.6
TS distance [Å]	2.57	2.54	2.38	2.22	2.37	2.63	2.48
Distortion angle [°]	10.8	12.8	14.2	18.3	14.5	10.7	14.3

**Figure 4.8.** – Dissociation energy, estimated TS distance and distortion angle obtained for the nucleophilic attack for the structures in Figure 4.7. The transition state distance was estimated by linear extrapolation of the forces obtained from CMD taken at intervals of 0.1 Å. The distortion angle is with respect to the ideal geometry of a  $sp^2$  hybridized system.

Analysis of the TS geometry shows that as the more reactive species have a more reactant-like transition state, with a longer C-C distance and a more planar geometry of electrophilic carbonyl carbon. Comparing the same Grignard reagent structure, comparing different substrates (for example, systems A - G) we find that for the most reactive acetaldehyde, the distortion angle is smaller than for the bulkier fluorenone.

The nucleophilic attack is most probably catalyzed by monomeric Grig-

## DIRECT NUCLEOPHILIC ATTACK VS. RADICAL PROPAGATION MECHANISM IN THE GRIGNARD REACTION

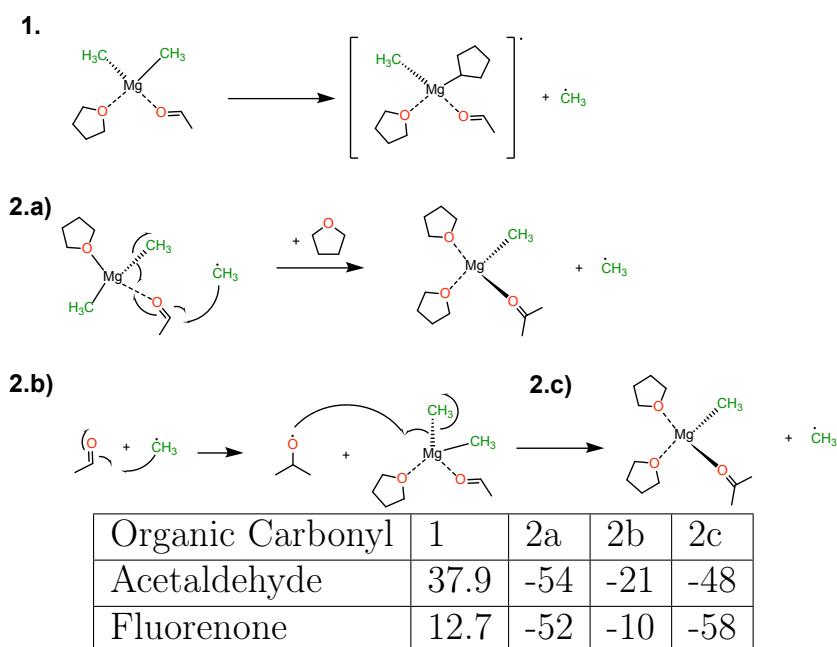
---

nard reagents, as these do not only show the lowest activation energy but also constitute the most abundant species in THF-solvent.<sup>[152]</sup> For example, structure F ( Figure 4.7), which shows a comparable energy of activation with respect to dimethylmagnesium, was found in our previous study to be energetically unfavored by about 8 kcal mol<sup>-1</sup> with respect to monomeric geometries, and it is therefore significantly less abundant in solution.

### 4.3.3. Nucleophilic attack vs Homolytic cleavage

The determination of the preference for the nucleophilic attack over a radical mechanism for the Grignard reactions depends on the relative height of the respective activation energies.

The homolytic bond cleavage energy represents the highest possible activation energy for a radical reaction producing a free methyl radical in solution. In that case, this highly energetic species is expected to recombine either to a free substrate, or to a substrate bound to a Grignard reagent. In both cases, we find that the overall reaction is highly exothermic by  $\sim -50$  kcal mol $^{-1}$ . (Figure 4.9):

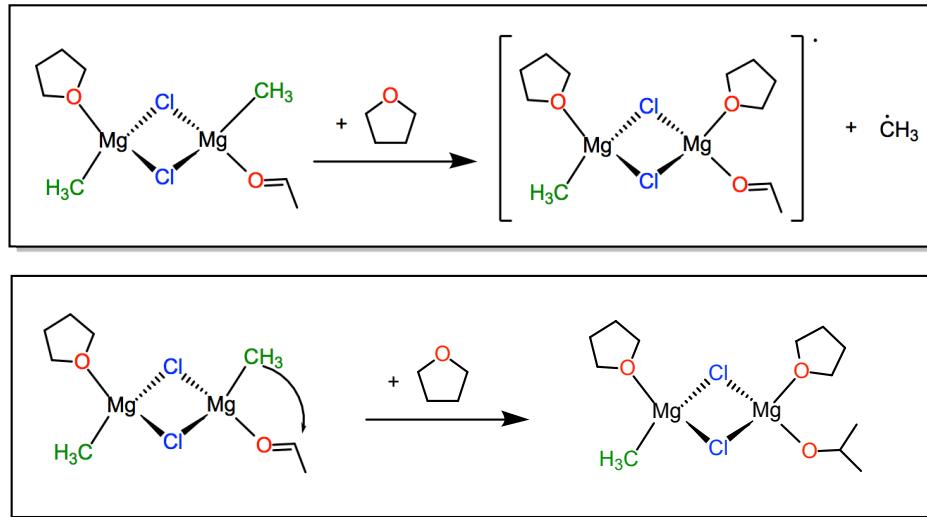


**Figure 4.9.** – Energies ( $\Delta E$ ) of the radical propagation pathway in [kcal mol $^{-1}$ ].

In fact, a radical mechanism may require even lower activation barriers if the initial formation of the radical is followed by a rapid recombination with the substrate within the same solvation cage, avoiding the release of the relatively unstable free methyl radical.

## DIRECT NUCLEOPHILIC ATTACK VS. RADICAL PROPAGATION MECHANISM IN THE GRIGNARD REACTION

System	A	B	C	E	G
Radical dissociation energy [kcal mol $^{-1}$ ]	35.9	37.7	63.2	34.5	12.7
Non-radical activation energy [kcal mol $^{-1}$ ]	6.6	8.7	12.9	13.1	14.6



**Figure 4.10.** – (Top) Dissociation and activation energy obtained for the radical and non-radical pathway for the structures shown in Figure 4.7, exemplified in (Bottom) for structure E.

Regardless of the different Grignard reagent structures, the radical mechanism for an alkyl carbonyl like acetaldehyde seems extremely unlikely. In fact, the activation energy for the radical pathway is in all cases above 30 kcal mol $^{-1}$ . This contrasts the very favorable activation energy for the nucleophilic addition, which can be as small as 6.6 kcal mol $^{-1}$ . We remark that the activation free energies for the nucleophilic attack were computed by AIMD using the PBE functional, while the radical formation energies were computed with the M062X one. For a better consistency between the data, we estimated the total energy difference between the reactant and TS during the nucleophilic attack using the two functionals, finding that, although the geometries are very similar, PBE underestimates the barrier by  $\sim 3.3$  kcal mol $^{-1}$  compared to M062X. This discrepancy is well below the energy differences found between the two pathways, and thus it does not change the qualitative picture.

For fluorenone, we find that the radical pathway is favored, even though the difference between the relative barriers is less pronounced. The radical formation is in particular stabilised by the large delocalization of the unpaired electron over the conjugated  $\pi$  system, while the nucleophilic attack is disfavoured by the necessary distortion of the geometry of the carbonyl that needs to break the same conjugation.

#### 4.4. Concluding remarks

Our study confirm the hypothesis that the Grignard reaction may proceed both by a nucleophilic addition and by a radical mechanism. We found that the preferences of either of the two mechanisms is associated to the nature of the substrate. In particular, extended aromatic moieties like in fluorenone are crucial for the initial stabilization of an unpaired electron that can initiate the radical cascade. In particular, they are in agreement with previous studies that have observed radical formation in the Grignard reaction of aromatic ketones such as Benzophenone<sup>[77]</sup> and Fluorenone.<sup>[155]</sup> For non-aromatic carbonyls where the radical can be localized solely on the carbonyl carbon, strongly negative inductive effects, for example by electronegative substituents, are needed to lower enough the energy of the species. We furthermore found that the nucleophilic addition reaction shows a clear trend between the different species present in the Schlenk-equilibrium. Specifically, the most reactive species in THF solvent are  $Mg(CH_3)_2$ ,  $CH_3MgCl$  and  $Mg(CH_3)_2 \cdot MgCl_2$ , respectively.

The presented data shed light on some complex mechanistic aspects of the magnesium chemistry in solution, and can provide solid ground for future studies on catalytic metallocorganic chemistry.

# 5. Modelling the self-assembly of alpha-tocopherol

## 5.1. Introduction

Vitamin E is known as an important part of a healthy diet. It acts as an antioxidant preventing oxidative stress in mitochondrial membranes and is capable of quenching singlet oxygen as well as capturing other radical species.<sup>[156–158]</sup> Vitamin E deficiency can lead to, for example, impairment of the immune response and neurological problems, such as Ataxia.<sup>[159–161]</sup>  $\alpha$ -tocopherol transfer protein ( $\alpha$ -TTP) is the liver factor responsible for the retention of RRR- $\alpha$ -tocopherol ( $\alpha$ -tol), which is the active isoform of vitamin E, in the human body.<sup>[162–164]</sup> The mechanism by which retention works is  $\alpha$ -TTP solubilizing  $\alpha$ -tol from the external leaflet of maturing endosomal compartments, promoting its release into the blood.

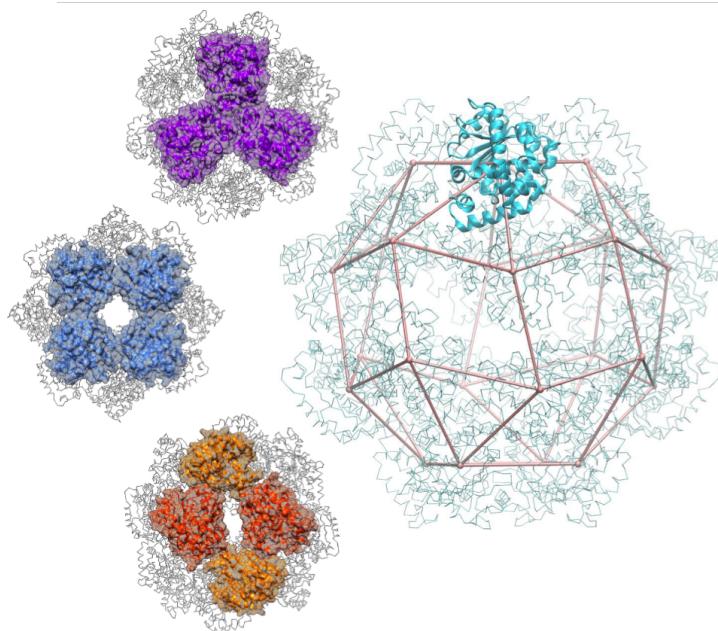
Structural studies<sup>[165,166]</sup> indicate that  $\alpha$ -TTP is active as a monomer, alike other transporters of its family.<sup>[167–169]</sup> However, the mechanism or cascade of mechanism by which  $\alpha$ -tol is secreted into the blood, and consecutively absorbed into the target tissues, is currently not well understood. A proposed possible pathways for the secretion and blood transport of  $\alpha$ -tol has been suggested to follow the enrichment into the leaflets of the plasma membrane by a lipid-exchange mechanism,<sup>[170,171]</sup> and consecutive transport into the blood by aggregating to very-low

density lipoprotein vesicles.<sup>[163]</sup>

The involvement of  $\alpha$ -TTP into  $\alpha$ -tol trafficking has not been clearly defined. In a recent work, Arai and coworkers have suggested that transfer of  $\alpha$ -tol to the plasma membrane is coupled to the extraction of phosphatidylinositolphosphates from the same membrane by  $\alpha$ -TTP.<sup>[164]</sup> They also suggested that lipid-exchange at the plasma membrane may involve higher order aggregates of  $\alpha$ -TTP than the monomers.<sup>[164]</sup>

Very recently, Aeschimann and coworkers have shown that binding of  $\alpha$ -tol facilitates  $\alpha$ -TTP oligomerization.<sup>[172]</sup> They further showed that aggregation of  $\alpha$ -TTP can be enhanced by well-defined conditions like the presence of negatively charged lipids or higher salt concentrations. The oligomerisation product they observe are thermodynamically stable, regular spherical 24-mers of  $\alpha$ -TTP (henceforth  $\alpha$ -TTP<sub>s</sub>). They characterized the spheres using cryo-EM and X-ray diffraction. (PDB: 5MUE and 5MUG).<sup>[172]</sup> They further showed that oxidative conditions enhance thermodynamic stability of  $\alpha$ -TTP<sub>s</sub> by formation of twelve disulfide-bonds that cross-linked different  $\alpha$ -TTP units. Their assembly follows the placement of the monomers at the vertices of a slightly distorted twice cantellated cube. The edges of the cantellated cube are precisely located on the axis along which interactions between neighboring proteins are formed. Every  $\alpha$ -TTP molecule has 4 different interacting interfaces that form two different types of contacts with its neighboring proteins.

The first interface is formed along the tetrameric  $C_4$  symmetry axis of the assembly, and consists of a patch of surface amino acids that are exposed to the solvent in the native monomeric folding of  $\alpha$ -TTP. The second interface is responsible for the assembly of  $\alpha$ -TTP around the trimeric  $C_3$  axis. This interface is located on the surface of the proper SEC-14-like binding domain. This means that in native  $\alpha$ -TTP this surface is buried within the protein. Only unfolding of the first 47 residues of the N-terminus (first N-terminal  $\alpha$ -helix) exposes this interface to the solvent and makes formation of the trimeric interface possible. The



**Figure 5.1.** – *Left:* (From top to bottom) The trimeric, tetrameric and dimeric interface that constitute  $\alpha$ -TTP<sub>S</sub>. *Right:* The single  $\alpha$ -TTP proteins are located on the vertices of a twice cantellated, distorted cube in the assembly of  $\alpha$ -TTP<sub>S</sub>. Graphic taken from Ref.<sup>[172]</sup>

orientation of the unfolded residues could not be determined due to their absence in the x-ray structure. Chromatographic data evidences that when the monomeric form is the most stable aggregation state for  $\alpha$ -TTP, it is in equilibrium with a small amount of low-weight dimeric or tetrameric aggregates. If the environmental conditions allow for the formation of  $\alpha$ -TTP<sub>S</sub> on the other hand, only monomeric  $\alpha$ -TTP or regular  $\alpha$ -TTP<sub>S</sub> are observed, while no other low- or middle-weight assemblies are detected.<sup>[172]</sup> The authors show  $\alpha$ -TTP<sub>S</sub> is selectively and efficiently transported through *in vitro* models of endothelial barriers,<sup>[172]</sup> making it a viable candidate for the physiological route of vitamin E delivery into the brain.<sup>[173,174]</sup>

Despite these recent discoveries, the mechanism of  $\alpha$ -TTP oligomerisation is unprobed and remains unknown and might be crucial for its transfecting properties. To explain these findings we developed a model to investigate the self-assembly of  $\alpha$ -TTP<sub>S</sub> based on the X-ray

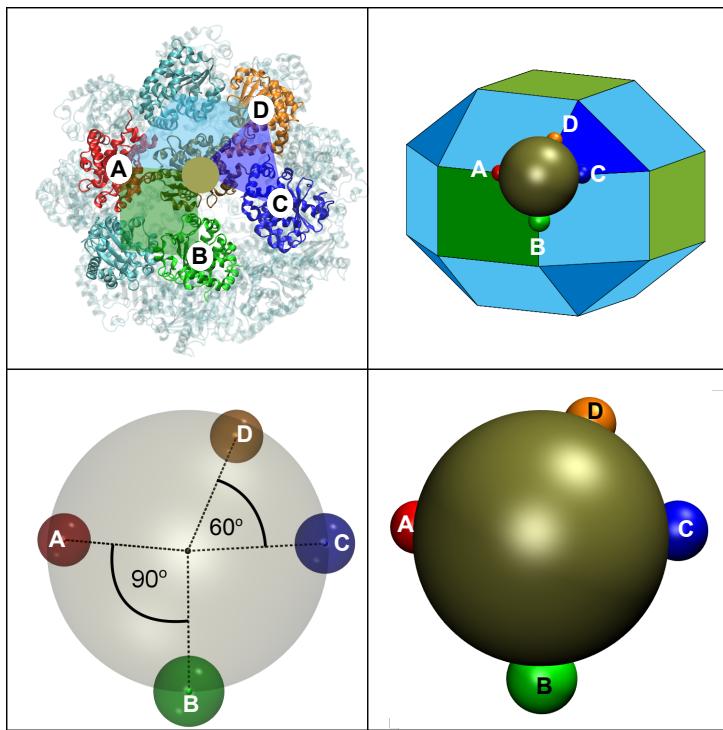
crystallographic and chromatographic data. We chose the number of proteins included in the simulation should allow for the formation of multiple aggregates to yield meaningful statistics. Here, the number of proteins should also not be too high, such that simulation time does not become excessive. We employ a patchy particle toy model as it allows for the simulation to be kept as simple as possible, while sufficiently capturing the dominant effects responsible for the self-assembly process.

A number of studies on the self-assembly of patchy colloids, fictionalized nanoparticles and biomolecules have shown molecular simulations using patchy models to be an effective modeling technique.<sup>[40,41,175,176]</sup> If the interactions beyond the patches of the toy-particles are highly directional and anisotropic, protein assembly is especially well captured.<sup>[177–182]</sup>

## 5.2. Patchy Particle Models

## 5.3. Computational Methods

The patchy model for the  $\alpha$ -TTP monomers is based on the geometry and interactions of the biological protein in the X-ray crystallographic structure of  $\alpha$ -TTP<sub>S</sub>. Thus, the patchy model for the monomer is made up of one hard sphere on which four interaction sites (IS). The relative orientation of the IS is chosen according to the geometry of  $\alpha$ -TTP<sub>S</sub>. The crystallographic structure revealed that only interactions between IS<sub>A</sub> and IS<sub>B</sub> as well as IS<sub>C</sub> and IS<sub>D</sub> are formed.<sup>[172]</sup>



**Figure 5.2.** – Coarse Grained model of  $\alpha$ -TTP<sub>S</sub>. *Top:* (left) In the native  $\alpha$ -TTP<sub>S</sub>, any monomeric  $\alpha$ -TTP (gold spot) is in contact with four other proteins along the edges of a cantellated cube (top right). *Bottom:*  $\alpha$ -TTP is described as a sphere with four interaction sites corresponding to the protein-protein contacts in  $\alpha$ -TTP<sub>S</sub>.

At the start of our investigation, the interaction for the two IS pairs were described by the simple potential wells of depth  $E_{A/B} = u$ ,  $E_{C/D} = 3u$  ( $u$  being an arbitrary unit of energy):

$$E_{A/B}(r) = \begin{cases} 0 & r > r_{cut} \\ E_{A/B} & r \leq r_{cut} \end{cases} \quad (5.3.1)$$

$$E_{C/D}(r) = \begin{cases} 0 & r > r_{cut} \\ E_{C/D} & r \leq r_{cut} \end{cases} \quad (5.3.2)$$

where  $r$  is the IS-IS distance,  $r_{cut} = 0.2 R$  is the maximum range of the interaction.

The initial 1:3 ratio between  $E_{A/B}$  and  $E_{C/D}$  was calibrated on an estimate of the dimerization free-energies from atomistic models using a standard thermodynamic cycle,<sup>[183]</sup> computing the solvation free energy of individual and dimeric structures solving the linearized Poisson-Boltzmann equation using the APBS software<sup>[183]</sup>, and the binding energy in vacuo using the Amber force field.<sup>[184]</sup> Protein dimers were extrapolated from the x-ray structure of  $\alpha$ -TTP<sub>S</sub> (PDB:5MUE).<sup>[172]</sup>

The four IS are identified by four vectors with origin in the center of the hard sphere, and ends in:

$$\begin{aligned} \text{IS}_A &= (0, R, 0) \\ \text{IS}_B &= (-R, 0, 0) \\ \text{IS}_C &= \left(0, \frac{R}{\sqrt{2}}, -\frac{R}{\sqrt{2}}\right) \\ \text{IS}_D &= \left(\frac{R}{\sqrt{2}}, 0, -\frac{R}{\sqrt{2}}\right) \end{aligned} \quad (5.3.3)$$

where  $R = 2$  nm is the radius of a hard sphere with its center in  $O(0, 0, 0)$ .

### 5.3.1. System Setup

We simulated a system having  $N = 216$  particles at thermal equilibrium. The protein particles were initiated on a uniform grid in a periodic cubic simulation box of edge  $24.625 R$  (where  $R$  is the radius of the protein), corresponding to roughly the experimental concentration at which  $\alpha$ -TTP aggregation is observed.<sup>[172]</sup> The accessible conformational space in the canonical NVT ensemble was explored using a Metropolis Monte Carlo (MC) algorithm, described in Section 2.3.1.<sup>[185]</sup>

For the protein self-assembly we included 4 different moves: the rotation or translation of the single particles or whole clusters of bound particles.<sup>[55,186,187]</sup> Two particles were considered to belong to same

cluster if they were connected by a chain of bonds.<sup>[188]</sup> A bond was defined if there was non-zero interaction between two particles, i.e. the distance between the corresponding interaction sites was less than  $0.2 R$ . Rotational moves made use of quaternion representation of the particle's orientation, which was modified by a smaller random orientation and then renormalized.<sup>[55,186]</sup>.

All results here are presented in reduced units:  $\mathbf{U}^* = U/u$  and  $\mathbf{T}^* = k_B T/u$  for the inner energy and the temperature of the system. We tested for interactions over a temperatures ranging from  $\mathbf{T}^*=0.014$  to  $\mathbf{T}^*=0.5$ . For none of the simulations run above  $\mathbf{T}^*=0.3$  could any stable interaction be observed, thus only data in the temperature range  $0.001 < \mathbf{T}^* < 0.03$  is presented in this thesis.

Simulations were organized in cycles, each cycle consisting of a number of attempted particle moves. Typical equilibration runs consisted of  $6 \times 10^6 - 9 \times 10^6$  MC cycles and were followed by a production run of additional  $3 \times 10^5$  MC cycles, during which averages of energy and cluster abundance were calculated. Convergence of the results was tested by evaluating both the convergence of the expectation value of the energy, and comparing the variance to the typical short time (1000 steps) energy fluctuations at a given temperature over the last  $6 \times 10^5$  MC cycles.

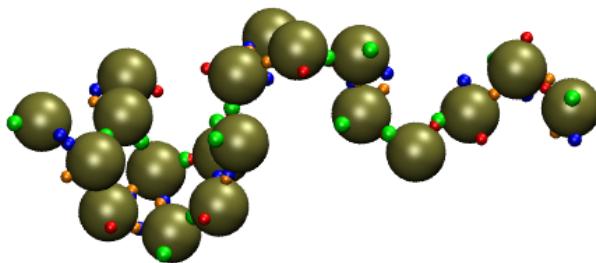
The acceptance probability for the single particle moves like translation or rotation was evaluated according to the scheme presented in Section 2.3.1. The cluster moves were implemented following the early rejection scheme.<sup>[55]</sup> During production runs, the overall acceptance ratio was adjusted to be 30% - 40% by a suitable choice of the maximum displacement and rotation parameters, and these values were never altered during the run.

Trajectory analysis was performed using the tools available in the VMD 1.9.2 package.<sup>[122]</sup>

## 5.4. The AB-Model

In the AB-Model we investigated the oligomerisation under the condition that only the  $IS_{A/B}$  interface is active. This setup mimics the experimental conditions at which the N-terminal region is folded, and the  $IS_{C/D}$  is not exposed to the solvent. We observe only monomeric species at high temperatures, while below around  $T^* = 0.09$ , we observed the appearance of low weight aggregates (dimers and tetramers). At low temperatures,  $T^* < 0.07$  high weight aggregates in form of polymeric chains could be observed.

For the lowest temperatures we would, from experiment, expect tetrameric aggregates rather polymeric chains, with all monomers lying in plane. In the polymeric chains the interfaces along the  $IS_{AB}$  are twisted by 90 degrees out of plane. The same behavior was also reported for the system with  $IS_{CD}$  active as well. A typical polymer chain obtained from such unconstrained simulations is shown in Figure 5.3.



**Figure 5.3.** – If no directionality is applied to the  $IS_{AB}$  or  $IS_{CD}$ , at low temperatures the protein aggregates into polymeric chains.

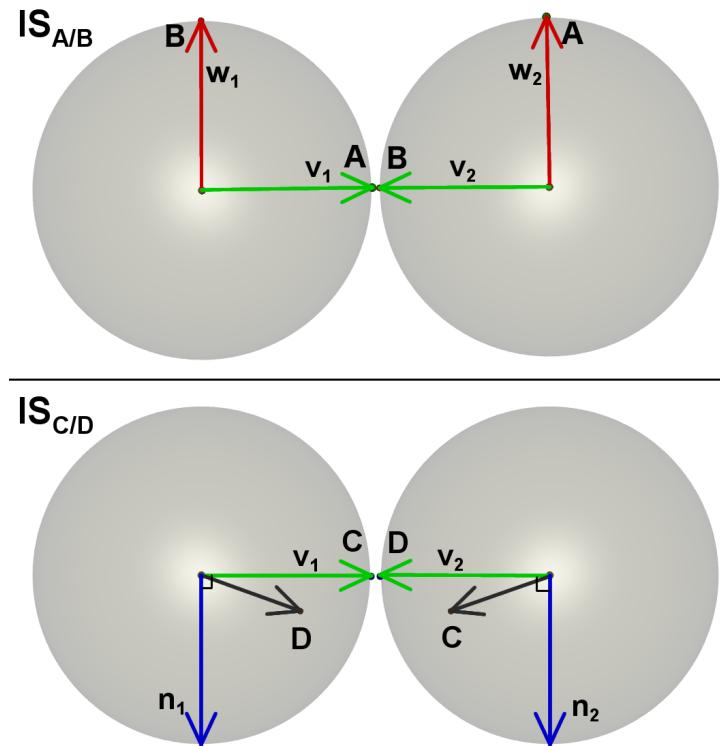
The protein interfaces however are chiral and interaction is only possible in a certain orientation of both proteins and happens over an extended surface, tightly interlocking.

Since our current model allows for rotation as in a joint an angular term had to be added to more precisely mimic biological conditions and properly describe the self-assembly.

The new interaction potential becomes:

$$E_{A/B}(r, \psi, \phi) = \begin{cases} 0 & r > r_{cut} \\ E_{A/B} \cos \psi \cos \phi & r \leq r_{cut} \end{cases} \quad (5.4.1)$$

where  $\cos \psi, \cos \phi, \cos \theta$  are defined from the scalar multiplication of the normalized vectors as in figure Figure 5.4.



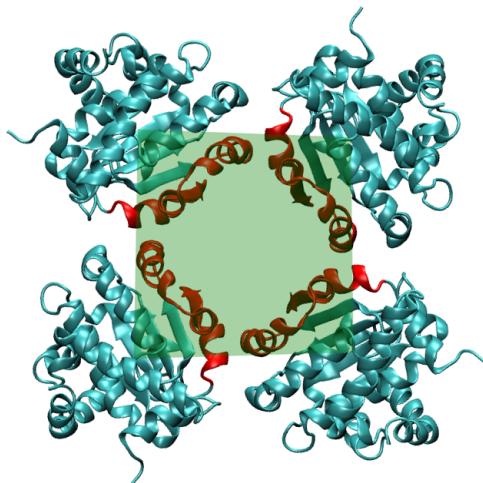
**Figure 5.4.** – Normalized vectors used to define the angular dependency of the interaction energy.  $\cos \psi = \mathbf{v}_1 \cdot \mathbf{v}_2$ , for any A/B or C/D interaction (green arrows);  $\cos \phi = \mathbf{w}_1 \cdot \mathbf{w}_2$  for any A/B interaction (top panel, red arrows);  $\cos \theta = \mathbf{n}_1 \cdot \mathbf{n}_2$ , for any C/D interaction (bottom panel, blue arrows)

The angular term was chosen as a cosines function, restricting tightly the formed interaction geometry while allowing for a broader interaction radius while  $IS_A$  and  $IS_B$  are in proximity. Thus, the number of states from which interaction of  $IS_A$  and  $IS_B$  can be initiated is statistically increased. Implementation of this angular term shows suppression of

polymeric chains. More simplified, computationally inexpensive angular terms were tried, but they partially produced biologically non-relevant pentamers.

Now, at around  $T^* = 0.09$ , we observed the appearance of low weight aggregates, mostly tetramers, with a small presence of trimers and dimers. The relative abundance of  $\alpha$ -TTP aggregates with respect to thermal energy is shown in the top panel of figure Figure 5.6. With lowering temperature the predominance of the tetramer is increased over other aggregates, which highlights that the most stable state for the system when only  $IS_{AB}$  is active is indeed the aggregation state of four  $\alpha$ -TTP proteins around the  $C_4$  Symmetry axis in  $\alpha$ -TTP<sub>S</sub>.

The resulting tetrameric interface is visualized in Figure Figure 5.5.



**Figure 5.5.** – The Tetrameric interface in  $\alpha$ -TTP<sub>S</sub> is formed by 4 monomers interlocked via multiple  $IS_{AB}$ , topologically located at the ligand binding site region of  $\alpha$ -TTP.

## 5.5. The ABCD-Model

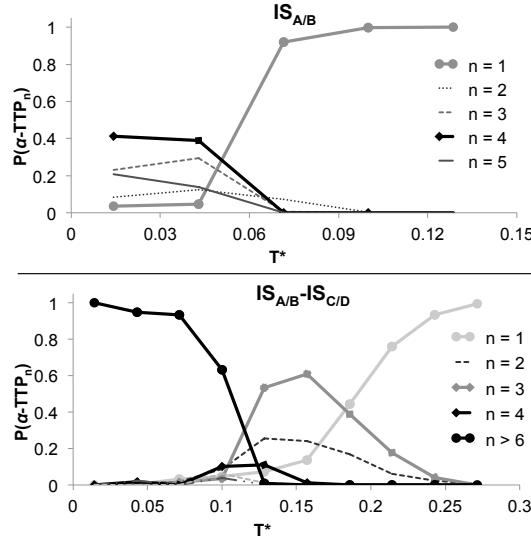
In a second set of MC simulations, both the  $IS_{A/B}$ ,  $IS_{C/D}$  interactions were active. This setup mimics  $\alpha$ -TTP with an unfolded N-terminus.

For the  $IS_{C/D}$  the angular term must be modified from the  $IS_{A/B}$  term, as the  $IS_{C/D}$  sites are oriented toward each other in a 60 degree angle. Therefore the angle between the normal vectors perpendicular to the plane had to be restricted more carefully, for the system not to twist out of plane. The energy is evaluated according to:

$$E_{C/D}(r, \psi, \theta) = \begin{cases} 0 & r > r_{cut} \\ E_{C/D} \cos \psi \cos 2\theta & r \leq r_{cut} \end{cases} \quad (5.5.1)$$

with the angles visualized in Figure 5.4.

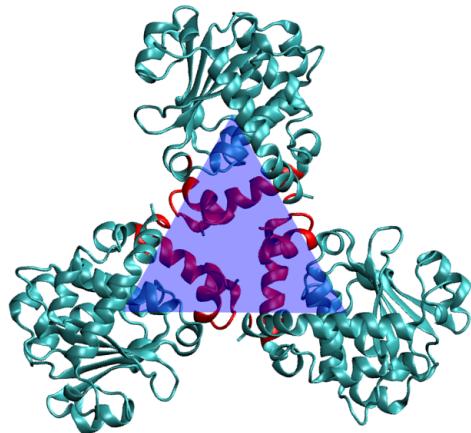
With all IS active, there are three temperature regimes with distinct aggregation states (Figure Figure 5.6 (Bottom)). In the high temperature range ( $T^* > 0.25$ ), only  $\alpha$ -TTP monomers are present. In the narrow ( $0.10 < T^* < 0.25$ ) region, trimeric species ( $\alpha$ -TTP<sub>3</sub>) are dominant, while for ( $T^* < 0.12$ ) high-weight aggregates ( $\alpha$ -TTP<sub>S</sub>) (Figure Figure 5.6) show the highest relative abundance.



**Figure 5.6.** – Relative abundance of  $\alpha$ -TTP oligomers ( $\alpha$ -TTP<sub>n</sub>) as a function of the reduced temperature  $T^*$ , when (top panel) only the  $IS_{A/B}$  interface is active, or (bottom panel) when both the  $IS_{A/B}$  and  $IS_{C/D}$  interfaces are active.

The trimeric phase is populated by monomers, dimers and predomi-

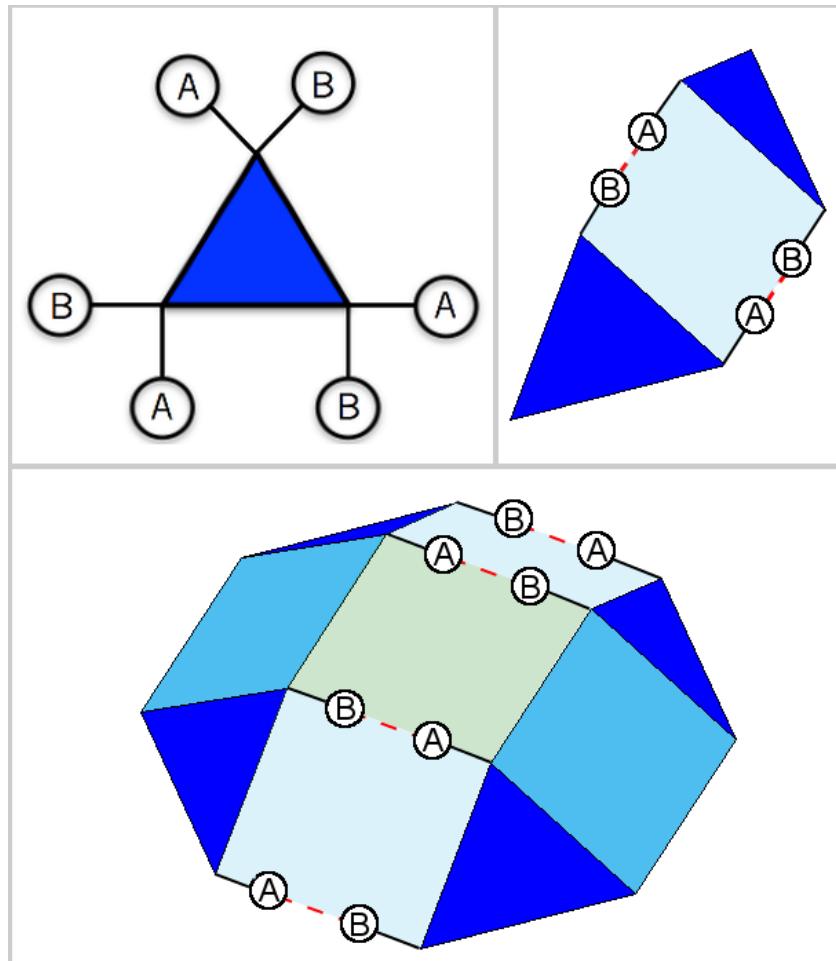
nantly by trimeric aggregates built by connection of three  $\text{IS}_{\text{C/D}}$  interfaces. The trimeric aggregates  $\alpha\text{-TTP}_3$ 's are stable in a region where the thermodynamic energy of the  $\text{IS}_{\text{C/D}}$  is sufficient to account for the loss of entropy caused by multimerisation, however the temperatures is too high to allow for formation of  $\text{IS}_{\text{A/B}}$  contacts. The  $\alpha\text{-TTP}_3$  interface is formed by the assembly of three  $\alpha\text{-TTP}$  around the  $C_3$  axis of  $\alpha\text{-TTP}_S$ . The resulting trimeric interface is shown in Figure Figure 5.7.



**Figure 5.7.** – The Trimeric interface in  $\alpha\text{-TTP}_S$  is formed by 3 monomers interlocked via multiple  $\text{IS}_{\text{CD}}$ , topologically located at the N-terminal region of  $\alpha\text{-TTP}$ . The interaction surface is not exposed to the solvent in folded  $\alpha\text{-TTP}$  monomers.

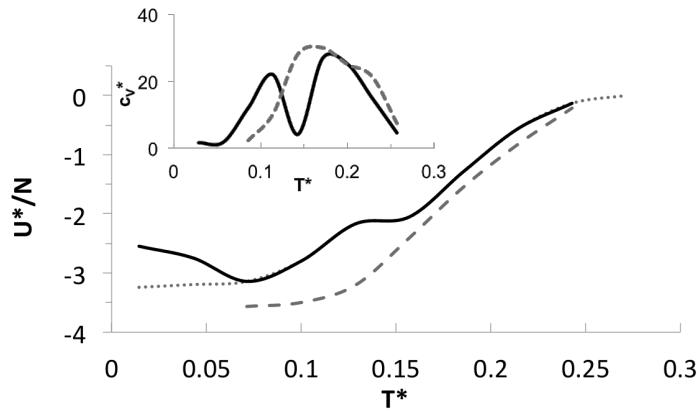
The last phase ( $T^* < 0.12$ ) is mostly populated by high-weight oligomers, up to  $\alpha\text{-TTP}_S$ . For formation of these high weight oligomers both  $\text{IS}_{\text{A/B}}$  contacts and  $\text{IS}_{\text{C/D}}$  contacts have to be established. Interestingly, by comparison of both graphs in Figure Figure 5.6 it becomes evident that oligomerisation to  $\alpha\text{-TTP}_S$  begins at a higher temperature ( $T^* = 0.12$ ) than the one characterizing  $\alpha\text{-TTP}$  aggregation when only  $\text{IS}_{\text{A/B}}$  are active ( $T^* = 0.07$ ). The shift is not caused by a change in the  $\text{IS}_{\text{A/B}}$  contacts, rather previous assembly of  $\alpha\text{-TTP}_S$  along the  $\text{IS}_{\text{C/D}}$  contacts enables the simultaneous, directed formation of two  $\text{IS}_{\text{A/B}}$  contacts at the same time. The result of this interaction is the formation a

hexameric structure along the  $C_2$  symmetry axis of  $\alpha\text{-TPP}_S$ , which is geometrically aided by the two  $\text{IS}_{A/B}$  contacts already being aligned in the pre-formed  $\alpha\text{-TPP}_3$ . Figure 5.8



**Figure 5.8.** – Oligomerization of  $\alpha\text{-TPP}_3$  (Blue triangle). Each oligomerization step involves the formation of at least two  $\text{IS}_{A/B}$  contacts. Newly formed contacts are represented by dashed lines.

Consecutively, more  $\text{IS}_{A/B}$  contacts are formed as  $\alpha\text{-TPP}_6$  and higher order aggregates further oligomerize. In this way the cooperative effect of creating multiple  $\text{IS}_{A/B}$  contacts, while forming a single  $\text{IS}_{A/B}$  contact is entropically unfavored causes the absence of structures of intermediate size between  $\alpha\text{-TPP}_3$  and  $\alpha\text{-TPP}_S$ .



**Figure 5.9.** – Inner energy per protein  $U^*/N$ , and specific heat  $c_V^*$  (inset) as a function of the reduced temperature for the system with both active  $IS_{A/B}$  and  $IS_{C/D}$  (continuous line). The dotted line was obtained by simulated annealing starting from converged data at  $T^*=0.07$ . The dashed lines report the same data for the system with  $E_{A/B}=1.5\text{ }u$ .

We further evidence the existence of three distinct phases by analyzing the transition between each phase. Figure Figure 5.9 shows the energy per protein versus temperature. The energy profile is described by two step functions whose transition is thermodynamically damped, characterized by two separate peaks for the heat capacity with respect to temperature. Thus, we identify two first-order phase transitions for the ABCD-model.

The location of these phase transitions is governed by the total energy of the  $IS_{A/B}-C/D$  contacts as well as their relative energy. The total energy with respect to temperature determines at what temperature the enthalpy gain from assembly outweighs the loss in entropy. In the current model, two  $IS_{A/B}$  contacts are formed at the same time in a hexameric interface (Figure 5.8) while exerting the same entropy loss of creating a single  $IS_{A/B}$  contact. This feature allows for the sphere formation below any temperature at which this interaction becomes favorable. We propose that, if  $2 * E_{A/B} \leq E_{C/D}$  one phase transition and with it the trimeric phase vanishes. To investigate this proposition we ran an additional simulation with  $E_{A/B} = 1.5\text{ }u = 0.5E_{C/D}$ . The result is shown as the dotted line in Figure Figure 5.9. As proposed,

we only observed one sigmoid in the energy vs temperature plot and only one peak in the heat capacity graph. The two phases merged into one. Visual inspection of the final geometries further confirms direct aggregation from  $\alpha$ -TTP monomers to  $\alpha$ -TTP<sub>S</sub>.

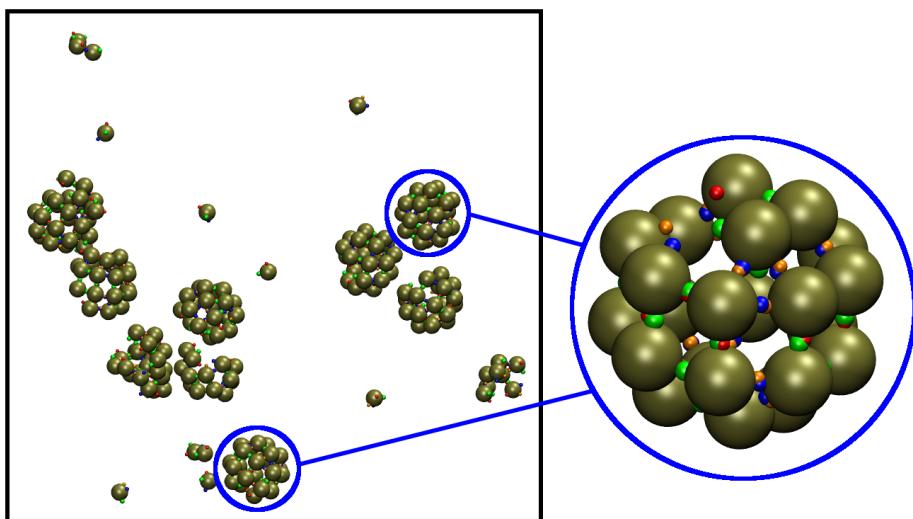
## 5.6. Discussion

Although our data shows the tendency for aggregation from  $\alpha$ -TTP to  $\alpha$ -TTP<sub>S</sub>, we did not observe perfect spheres of  $\alpha$ -TTP<sub>24</sub> in our simulations. Rather, in our simulations we could observe the formation of spherical particles with single point defects. The existence of such defects is evident from the discrepancy between the average potential energy reported in Figure Figure 5.9 and the target value of  $-4u$ , that would be reached at  $T^* = 0.00$  for all monomers forming perfect spheres. In fact, the average aggregation number observed is 20.2.

We suggested three reasons why no perfect spheres were observed in our simulations. Firstly, the concentration of free monomers in order to fill defects at later stages of the simulations was too low. Secondly, since the self-assembly was not directed more spheres were initiated by self-assembly than particles were in the simulation to complete the assembly. Thirdly, at low temperatures we observed kinetically trapped states. These states are dominant in MC runs at values of  $T^*u \ll E_{A/B}$ , which yielded poorly aggregated structures, as it is difficult for the simulation to escape local minima and the self-assembly of the structure is irregular. The total energy of the system after convergence was also observed to be higher than those observed for higher temperatures of the same aggregation phase. Improvement of the structures at  $T^* \leq 0.06$  was obtained by applying 30 cycles of simulated annealing, between  $T^* = 0.07$  and the target temperature. We consequently observed the formation more regular  $\alpha$ -TTP<sub>S</sub> structures with aggregation number  $\approx 23$ .

This observation is very similar to observations made during crystal

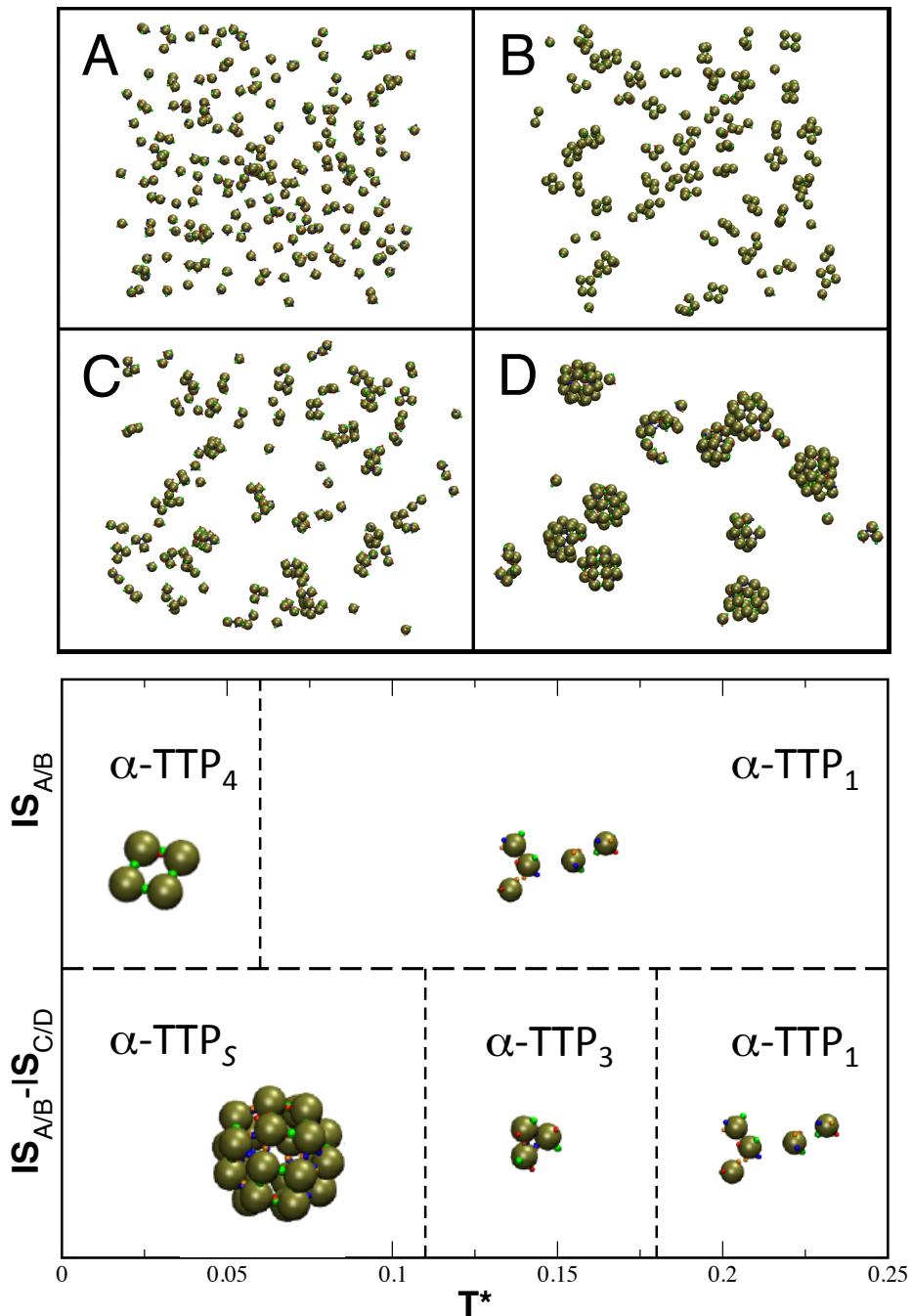
growth. Large crystals are usually only formed at intermediate solute concentrations, whereas at very high solute concentrations fast growth of many small crystals is observed. In order to tackle these three reasons we started 2 different simulations. Firstly, to tackle the lack of monomers, we initiated a simulation starting from a converged structure at  $T^* = 0.07$ . A new layer of monomers ( $6 \times 6 \times 6 \rightarrow 7 \times 7 \times 7$  particles) was initiated around the previous simulation box and the cubic simulation box extended to match the experimental density. (box length  $24.625R \rightarrow 28.73$ ). The new monomers were only allowed to interact with particles of the previous grid to suppress formation of new spheres. We chose this setup to simulate constant supply of monomers during self-assembly in the human body. Results were hardly improved structures. We did not observe 24-mers again, as mostly lower-weight aggregates were improved by the new monomers, and matching the last missing protein in  $\alpha\text{-TTP}_S$  is statistically unlikely. This observation further evidences the importance of the cooperative effect between previously formed  $\alpha\text{-TTP}_3$  when forming the sphere in solution.



**Figure 5.10.** – (Left) The simulation box at  $T^* = 0.07$  close to convergence. Two of the spheres, highlighted with blue circles only possess a single defect (Right). The number of free monomers at this point of the simulation is very low, making it statistically improbable to accommodate the defect in the near-perfect spheres.

Studies of our experimental colleagues also showed that if the environmental conditions were different from the one, where full-sphere formation was observed, spherical particles with slight defects made up a majority of the species. This, along with our observations on the self-assembly might suggest that other co-factors in the human body aid the self-assembly.

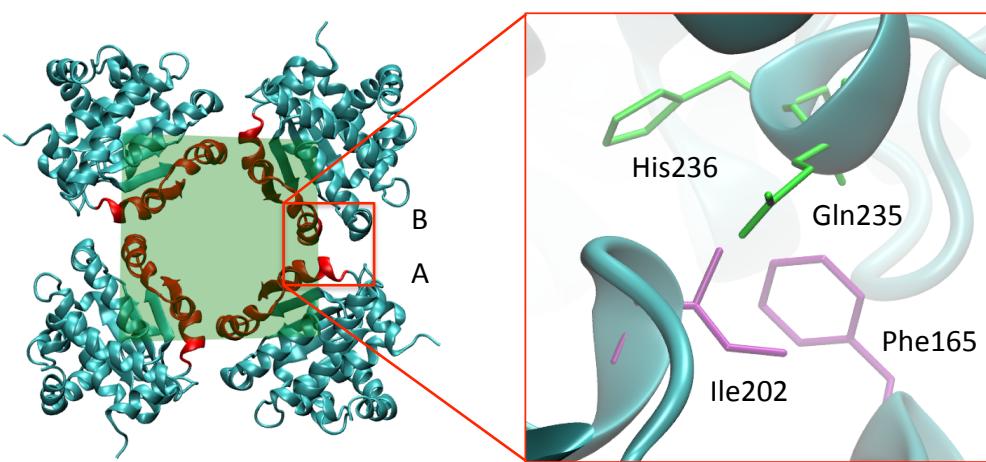
Our data are in optimal agreement with the native gel electrophoresis experiment reported in ref<sup>[172]</sup>. Natively folded  $\alpha$ -TTP, where the IS<sub>C/D</sub> are inactive, showed the predominance of a monomeric form, with residual presence of low-weight aggregates (dimer, tetramer, figure Figure 5.6). After initiation of unfolding of the N-terminus, the proteins assembled into stable  $\alpha$ -TTP<sub>S</sub>. The spheres showed no tendency to disassemble to lighter oligomers in further incubation tests over a time-window of 24 hours.<sup>[172]</sup> These experimental observations on the biological system correspond to the region of  $T^* \approx 0.06-0.08$  in our toy system, as can be determined by comparison with Figure Figure 5.11



**Figure 5.11.** – Top: Aggregation states of  $\alpha$ -TTP at different conditions. (A) Monomeric dispersion at  $T^* = 0.27$ ; (B) aggregation of low-weight oligomers at  $T^* = 0.04$ , for the system with only active  $S_{A/B}$ ; (C) system with all active  $S$ 's at intermediate  $T^* = 0.13$ , where trimers begin to form; (D) same system at  $T^* = 0.07$ , characterized by formation of  $\alpha\text{-TTP}_S$ . Bottom: phase diagram with dominating species at different conditions of temperature and folded state. The region between 0.06 and 0.11  $T^*$  corresponds to the experimentally observed behavior, with either properly folded monomers, or assembled  $\alpha\text{-TTP}_S$ .

$T^* \approx 0.06$ - $0.08$  is also the region for which we observed the most complete sphere aggregation and  $\alpha$ -TTP<sub>S</sub> was formed by the cooperative effect shown in Figure Figure 5.8, while the formation of single IS<sub>A/B</sub> contacts was thermodynamically unstable. We observe mostly monomers for only IS<sub>A/B</sub> active, while with IS<sub>C/D</sub> active as well, we observe aggregation of  $\alpha$ -TTP<sub>S</sub>. The third phase of stability, dominated by  $\alpha$ -TTP<sub>3</sub>, was not observed by their initial experiments. We predict that the region of thermodynamic stability for this phase corresponds to temperature range of  $T^* \approx 0.11$  -  $0.18$  in the toy system. At experimental conditions with an unfolded N-terminus, the system evolves to  $\alpha$ -TTP<sub>S</sub>, showing no lower weight aggregates, corresponding to a lower temperature of the toy system. Because proteins are only stable in a narrow temperature region, the proposed trimeric phase cannot be explored by increasing the temperature of the real system. However, we predict from our simulations that the region of thermodynamic stability of the trimeric phase can be moved by changing the relative interaction strength of IS<sub>A/B</sub> and IS<sub>C/D</sub>, as shown in our data. We propose, that if the interaction strength of IS<sub>A/B</sub> is lowered, the trimeric phase is extended to lower temperatures and  $\alpha$ -TTP<sub>3</sub> should be observable at experimental conditions.

Aeschimann and coworkers<sup>[172]</sup> showed, that mostly hydrophobic contacts at the protein-protein interfaces are responsible for the strength of the interaction on the  $C_3$  and  $C_4$  symmetric axis. We predict, that single point mutations of key residues could weaken the interaction strength of IS<sub>A/B</sub> and thus make the trimeric phase most stable at experimental temperatures. A such mutation would be the F165R mutation in which the hydrophobic patch is disrupted by a positively charged arginine residue. Experiments to validate these predictions are currently planned by our collaborators.



**Figure 5.12.** – The most relevant residues responsible for the binding at the  $C_4$  symmetric interface are shown in licorice. The green and purple colors refer to amino acids belonging to two different  $\alpha$ -TTP units.

Our study shows how patchy-particle models can help with understanding the self-assembly of protein systems. Also, understanding the self-assembly process could be relevant to design of drugs that can be incorporated in the hollow  $\alpha$ -TTP<sub>S</sub> nanosphere. Furthermore, being able to make predictions on what kind of effects point-mutations will have on the system is key in order to steer experiments more efficiently. Our study also sheds light on how the cooperative effect between the  $C_3$  symmetric interface and the IS<sub>A/B</sub> contacts aids to avoid defects in the assembly of  $\alpha$ -TTP<sub>S</sub>. Two potential biological benefits of this effect include, firstly that sphere aggregation can be triggered rapidly under the right conditions while maintaining the functionality of native, monomeric species else-wise. Secondly, the cooperative effect might help aggregation to more ordered and to obtain perfect spheres at a higher rate.

## 6. Conclusion and Future Outlook

The results of paper I and II showed that the solvent is crucial for the correct characterization of the different species in and the propagation of the Schlenk-equilibrium. However, we also determined that the most abundant species in THF solvent are monomeric Grignard reagents in line with previous experimental findings.<sup>[76]</sup> In continuation of this study we propose the investigation of the Schlenk-Equilibrium of in diethyl ether, that shows a higher degree of association for the Grignard reagent and whose equilibrium is suggested to be governed by more complex higher order aggregates.<sup>[76]</sup> Enhanced stability of higher order aggregates might furthermore have implications on the reactivity trends determined for competing mechanism of the Grignard reaction determined in paper II. We further suggest additional investigation on the specific influence of the inductive effect over a variety of substrates on the stability of the radical pathway of the Grignard reaction.

In regards to our investigation on the self-assembly of the  $\alpha$ -tocopherol transfer protein ( $\alpha$ -ttp) experiments to validate our predictions on the influence of specific protein modifications on the self-aggregation state of the protein. Our studies shown how the formation of a the trimeric interface of  $\alpha$ -ttps is key to the formation of the protein nanocage. However, the surface making up the trimeric interface is not exposed in the native form of  $\alpha$ -ttp. Experimentally it was possible to show that the unfolding of the N-terminus of the protein and exposure of the trimeric interface is facilitated by presence of negatively charged lipids or higher salt concentrations. The mechanism of the unfolding

---

however remains to be investigated computationally in future work. Another continuation of the research could involve the experimental investigation of the conservation of these aggregation properties in other members of the SEC14-like family, for example CRALBP and SPF.

# 7. Appendix

## 7.1. Appendix A - Metadynamics parameters for Grignard simulations

**Table 7.1.** –  $d_0$ ,  $p$ ,  $q$  parameters for the definition of the coordination number of THF or CH<sub>3</sub> to Mg used as Collective Variables 3.1.2.

	d0	p	q
THF	2.75	12	24
CH3	3.0	12	24

A quadratic potential wall was used to restrain CV1 in the Schlenk equilibrium reaction (difference between of the coordination of Methyl groups to the Mg atoms) to values larger than -0.25, using a force constant of 0.1 Hartree. This bias was introduced in order to sample only the Methyl transfer from Mg<sup>1</sup> to Mg<sup>2</sup>, and not vice versa. This hinders the simulation from having to explore a twice as big accessible FES, with very limited additional information gain. For the different metadynamics simulations different sets of time-dependent bias potentials were used for the different sets of surfaces and collective variables, given in Table S2. The error is estimated according to the formula:

$$\bar{\varepsilon} = \frac{S^2 \omega T}{D\tau_G} \left( \frac{\delta s}{S} \right)^d (2\pi)^{d/2} \sum^k \frac{1}{\pi^2 k^2} e^{\left( \frac{-\pi^2 k^2}{2} \left( \frac{\delta s}{S} \right)^2 \right)} \quad (7.1.1)$$

from<sup>[120]</sup>

---

## Appendix A - Metadynamics parameters for Grignard simulations

---

There,  $\delta s$  is the Gaussian with,  $\omega$  the Gaussian height,  $\tau_G$  the Gaussian deposition time,  $S$  is the size of the respective CV-space to be explored,  $D$  is the diffusion coefficient and  $T$  the temperature and  $d$  the dimensionality of the system.

**Table 7.2.** – Metadynamics parameters and error estimate. The error estimate is calculated according to equation 7.1.1.

System	Collective variable	diffusion coefficient [1/fs]	S	Temperature [K]	Gaussian height [kcal/mol]	Gaussian width	deposition time [fs]	error approx-imate [kcal/mol]
$D^{ClCl}$	THF coordination	2.87E-6	1.25	300	0.35	0.04	12.5	0.54
$D^{ClMe}$	THF coordination	1.02E-5	1.25	300	0.5	0.1	12.5	1.36
Schlenk Equilibrium	Difference in Mg-CH <sub>3</sub> coordination	5.16E-5	2.5	300	0.25	0.1	12.5	0.30
	THF-coordination to Mg <sup>1</sup>	3.01E-5	1.25	300	0.25	0.1	12.5	0.56

## 7.2. Appendix B -NBO Analysis for the Grignard simulations

**Table 7.3.** – Lone pair delocalization into empty orbitals of Mg for the monomeric species.

	Cl1	Cl2	O1	O2	O3	O4	C1	C2
Mg(CH <sub>3</sub> ) <sub>2</sub> (THF) <sub>2</sub>			1.4	1.7			7.9	9.6
MgClCH <sub>3</sub> (THF) <sub>2</sub>	5.3		1.8	1.9			10.2	
MgCl <sub>2</sub> (THF) <sub>2</sub>	5.9	5.9	2	2				
MgCl <sub>2</sub> (THF) <sub>3</sub>	5.3	5.3	1.9	1.8	1.8			
MgCl <sub>2</sub> (THF) <sub>4</sub>	4.3	4.0	1.5	1.6	1.6	1.7		

**Table 7.4.** – Lone pair delocalization into empty orbitals of Mg for the D<sup>ClCl</sup> species.

		Cl1	Cl2	O1a	O1e	O2a	O2e	C1	C2
D <sup>ClCl</sup> <sub>11</sub>	Mg <sup>1</sup>	4.1	2.2	1.7	1.3			9.4	
	Mg <sup>2</sup>	2.3	4			1.5	1.8		9.2
D <sup>ClCl</sup> <sub>12</sub>	Mg <sup>1</sup>	3.3	4.4	1.8		0	0	11.2	
	Mg <sup>2</sup>	3.8	2.2			1.5	1.9		9.5
D <sup>ClCl</sup> <sub>22</sub>	Mg <sup>1</sup>	3.8	3.5	1.8				11.7	
	Mg <sup>2</sup>	3.9	3.5			1.8			11.5

## APPENDIX

---

**Table 7.5.** – Lone pair delocalization into empty orbitals of Mg for the  $D^{ClMe}$  species.

		C11	Cl2	O1a	O1e	O2a	O2e	C1	C2
$D_{11cis}^{ClMe}$	$Mg^1$	4.1	7.1				2.1	5.6	
	$Mg^2$	3.8		1.8				3.9	11.1
$D_{11trans}^{ClMe}$	$Mg^1$	4.2	6.7				2.1	5.7	
	$Mg^2$	3.6		1.9				3.7	11.0
$D_{12}^{ClMe}$	$Mg^1$	5	6.9		2.0			5.2	
	$Mg^2$	2		1.9			1.5	3.8	9.5
$D_{11}^{ClMeTHF}$	$Mg^1$	4.1	6.6	1.9			1.2	4.6	
	$Mg^2$	3.8				1.6	0.8	3.6	10.2
$D_{21}^{ClMe}$ (Cl axial)	$Mg^1$	3.1	6.3			1.9	1.5	4.7	
	$Mg^2$	4		1.8				4.0	10.5
$D_{21}^{ClMe}$ (THF axial)	$Mg^1$	3.3	6.3			1.8	1.7	4.3	
	$Mg^2$	3.9		1.7				4.2	10.4

### 7.3. Appendix C - Truncation of pairwise interactions

Corrections for the cut interactions of truncating pairwise interactions of the Lennard-Jones (LJ) potential. For the LJ-potential the correction term for the potential energy and for the truncated, but non-shifted potential is given by:

$$V_{corr} = \frac{8\pi N^2}{3V} \varepsilon \sigma^3 \left[ \frac{1}{3} \left( \frac{\sigma}{r_c} \right)^{-9} - \left( \frac{\sigma}{r_c} \right)^{-3} \right] \quad (7.3.1)$$

The correction to the potential energy for the truncated and shifted potential is:

$$V_{corr} = \frac{2\pi N^2}{3V} r_c^3 V(r_c) + \frac{2\pi N^2}{V} \int_{r_c}^{\infty} r^2 V(r) dr \quad (7.3.2)$$

The correction to the pressure for the truncated and shifted as well as non-shifted potential is:

$$V_{corr} = \frac{16\pi N^2}{3V^2} \varepsilon \sigma^3 \left[ \frac{2}{3} \left( \frac{\sigma}{r_c} \right)^{-9} - \left( \frac{\sigma}{r_c} \right)^{-3} \right] \quad (7.3.3)$$

## 7.4. Appendix D - Molecular Mechanics potentials

The major difference between AIMD and Molecular Mechanics (MM) arises from the basis on which the forces are evaluated. In AIMD these stem from the potential energy surface according to the solutions of the time-independent Schroedinger equation. In MM the motion is determined by evaluating Newton's EOMs. Moving from quantum to classical mechanics these two expressions are the respective equivalents of each other. The forces in Newton's equation of motion are obtained from the derivative of the potential energy function.

The potential energy function,  $V_{MM}$  used in MM is shown below. We will in turn look at the different interactions that make up the potential energy equation

$$\begin{aligned} V_{MM} = & \sum_b k_b(r - r_0)^2 + \sum_\sigma k_\sigma(\sigma - \sigma_0)^2 + \sum_{phi} k_\phi[1 + \cos(n\phi - \delta)] \\ & + 4 \sum_{i \in A} \sum_{i \in B} \varepsilon \left[ \left( \frac{\sigma}{r_{ij}} \right)^{12} - \left( \frac{\sigma}{r_{ij}} \right)^6 \right] + \frac{1}{4\pi\varepsilon_0} \sum_{i \in A} \sum_{i \in B} \frac{q_i q_j}{r_{ij}} \end{aligned} \quad (7.4.1)$$

Firstly, the intramolecular interactions are the forces between neighboring atoms in the same molecule. They major contributions arise from bond stretching, bond angle bending and bond torsions. Bond stretching is accurately described using the Morse Potential:

$$V(r) = D_e \left[ 1 - e^{-k_b(r-r_0)} \right]^2 \quad (7.4.2)$$

where  $r$  is the length of the bond,  $k_b$  a constant,  $r_0$  the equilibrium bond length, and  $D_e$  the well depth minimum. This term requires

three parameters per bond and is somewhat expensive to compute due to the exponential term. Because the energy of bond stretching is relatively high most bonds only deviate slightly from their equilibrium geometry. We can therefore use a second-order Taylor expansion around the equilibrium bond length to describe bond stretching.

$$V(r) = k_b(r - r_0)^2 \quad (7.4.3)$$

The bond angle bending is similarly described by a second-order Taylor expansion of the angle  $\sigma$  around the equilibrium bond angle  $\sigma_0$ :

$$V(\sigma) = k_\sigma(\sigma - \sigma_0)^2 \quad (7.4.4)$$

The deviation from the equilibrium geometry for bond torsion are energetically much more easily accessible and are well approximated using a cosine expansion:

$$V(\phi) = \sum_{n=0}^N c_n \cos(n\phi)^n \quad (7.4.5)$$

with  $\phi$  the torsional angle, summed over  $N$  terms with coefficient  $k_n$ . In practice for faster convergence, the following mathematically equivalent expression is used:

$$V(\phi) = \sum_{n=0}^N c_n [1 + \cos(n\phi) - \delta] \quad (7.4.6)$$

Secondly, the intermolecular interactions stem from the interactions of all atoms in the simulation with each other. Even though the intermolecular forces are dependent on different multi-body interactions the expression, its potential energy expression is in simulation truncated at the second order and the neglected interactions are incorporated

using parameterization of the one- and two-body interactions. The reason for this approximation is that the scaling of the method is determined by the treatment of the intermolecular forces. Including those interactions to the third or forth order results in  $N^3$  or  $N^4$  scaling of the method, making it much less applicable to larger systems.

The pair-wise interactions can be dissected into electrostatic interactions, van-der-Waals attractions and excluded volume repulsions. The potential energy term for the electrostatic interactions follows Coulomb's law,

$$V_{coulomb} = \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} \quad (7.4.7)$$

where  $q_i$  and  $q_j$  are the partial charged of atoms  $i$  and  $j$ ,  $r_{ij}$  is the interatomic distance and  $\epsilon_0$  is the free space permittivity. The van der Waals attractions and Excluded Volume repulsions can be combined into a single expression. The resulting expression is the potential proposed by Lennard-Jones:<sup>[189]</sup>

$$V_{LJ} = 4\epsilon \left[ \left( \frac{\sigma}{r_{ij}} \right)^{12} - \left( \frac{\sigma}{r_{ij}} \right)^6 \right] \quad (7.4.8)$$

where  $\epsilon$  and  $\sigma$  are constant dependent on the types of atoms  $i$  and  $j$ . In summation the three intramolecular and two intermolecular terms describe the potential energy, summarized in Equation 7.4.1.

The  $N^2$  scaling of MM, determined by the intermolecular interactions, can be further reduced with the use of cutoff radii, periodic pair-list updates and the particle mesh Ewald (PME) method. These methods will be further examined in the Methods section. Classical MD works well under the pretense that the BO approximation is valid, the electronic structure is not of interest, there is no bond breaking or forming and the electrons are highly localized.

# Bibliography

- [1] Sugita, Y.; Okamoto, Y. *Chemical Physics Letters* **1999**, *314*, 141 – 151.
- [2] Morris, A. J.; Grey, C. P.; Pickard, C. J. *Phys. Rev. B* **2014**, *90*, 054111.
- [3] Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *Journal of Computational Chemistry*, *24*, 1999–2012.
- [4] Nørskov, J. K.; Abild-Pedersen, F.; Studt, F.; Bligaard, T. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 937–943.
- [5] Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864—B871.
- [6] Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133–A1138.
- [7] Čížek, J. *The Journal of Chemical Physics* **1966**, *45*, 4256–4266.
- [8] Kümmel, H. *Int. J. Mod. Phys. B* **2003**, *17*, 5311–5325.
- [9] Lutnæs, O. B.; Teale, A. M.; Helgaker, T.; Tozer, D. J.; Ruud, K.; Gauss, J. *The Journal of Chemical Physics* **2009**, *131*, 144104.
- [10] Teale, A. M.; Lutnæs, O. B.; Helgaker, T.; Tozer, D. J.; Gauss, J. *The Journal of Chemical Physics* **2013**, *138*, 024111.
- [11] Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *Journal of Computational Chemistry*, *4*, 187–217.

## Bibliography

---

- [12] Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. *The Journal of Physical Chemistry A* **1999**, *103*, 3596–3607.
- [13] Brooks, B. R.; et al. *J. Comput. Chem.* **2009**, *30*, 1545–614.
- [14] Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *Journal of the American Chemical Society* **1996**, *118*, 11225–11236.
- [15] Darden, T.; York, D.; Pedersen, L. *The Journal of Chemical Physics* **1993**, *98*, 10089–10092.
- [16] Cascella, M.; Vanni, S. *Toward accurate coarse-graining approaches for protein and membrane simulations*, 2016; Vol. 12.
- [17] Krack, M.; Parrinello, M.; *QUICKSTEP: Make the Atoms Dance*; Tech. Rep.; 2004. <http://www.fz-juelich.de/nic-series/volume25/volume25.html>.
- [18] Levitt, M.; Warshel, A. *Nature* **1975**, *253*, 694.
- [19] Warshel, A. *Annual Review of Biophysics and Biomolecular Structure* **2003**, *32*, 425–443; PMID: 12574064.
- [20] Pierce, L. C.; Salomon-Ferrer, R.; Augusto F. de Oliveira, C.; McCammon, J. A.; Walker, R. C. *Journal of Chemical Theory and Computation* **2012**, *8*, 2997–3002; PMID: 22984356.
- [21] Stone, J. E.; Phillips, J. C.; Freddolino, P. L.; Hardy, D. J.; Trabuco, L. G.; Schulten, K. *Journal of Computational Chemistry*, **28**, 2618–2640.
- [22] Zhao, G.; Perilla, J. R.; Yufenyuy, E. L.; Meng, X.; Chen, B.; Ning, J.; Ahn, J.; Gronenborn, A. M.; Schulten, K.; Aiken, C.; Zhang, P. *Nature* **2013**, *497*, 643.
- [23] Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. *Science* **2010**, *330*, 341–346.

- [24] Soares, T. A.; Vanni, S.; Milano, G.; Cascella, M. *The Journal of Physical Chemistry Letters* **2017**, *8*, 3586–3594; PMID: 28707901.
- [25] Periole, X.; Kneppe, A. M.; Sakmar, T. P.; Marrink, S. J.; Huber, T. *Journal of the American Chemical Society* **2012**, *134*, 10959–10965; PMID: 22679925.
- [26] Chandler, D. E.; Strümpfer, J.; Sener, M.; Scheuring, S.; Schulten, K. *Biophys. J.* **2014**, *106*, 2503–2510.
- [27] Reddy, T.; Shorthouse, D.; Parton, D.; Jefferys, E.; Fowler, P.; Chavent, M.; Baaden, M.; Sansom, M. *Struct. Engng.* **1993**, *2015*, *23*, 584–597.
- [28] Spiga, E.; Alemani, D.; Degiacomi, M. T.; Cascella, M.; Dal Peraro, M. *Journal of Chemical Theory and Computation* **2013**, *9*, 3515–3526; PMID: 26584108.
- [29] Groot, R. D.; Warren, P. B. *The Journal of Chemical Physics* **1997**, *107*, 4423–4435.
- [30] Liwo, A.; Khalili, M.; Scheraga, H. A. *Proceedings of the National Academy of Sciences* **2005**, *102*, 2362–2367.
- [31] Shinoda, W.; DeVane, R.; Klein, M. L. *Current Opinion in Structural Biology* **2012**, *22*, 175 – 186; Theory and simulation/Macromolecular assemblages.
- [32] Noid, W. G.; Chu, J.-W.; Ayton, G. S.; Krishna, V.; Izvekov, S.; Voth, G. A.; Das, A.; Andersen, H. C. *The Journal of Chemical Physics* **2008**, *128*, 244114.
- [33] Shelley, J. C.; Shelley, M. Y.; Reeder, R. C.; Bandyopadhyay, S.; Klein, M. L. *The Journal of Physical Chemistry B* **2001**, *105*, 4464–4470.
- [34] Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *The Journal of Physical Chemistry B* **2007**, *111*, 7812–7824; PMID: 17569554.

## Bibliography

---

- [35] Marrink, S. J.; de Vries, A. H.; Mark, A. E. *The Journal of Physical Chemistry B* **2004**, *108*, 750–760.
- [36] Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. *Chemical Reviews* **2016**, *116*, 7898–7936; PMID: 27333362.
- [37] Dans, P. D.; Walther, J.; Gómez, H.; Orozco, M. *Current Opinion in Structural Biology* **2016**, *37*, 29 – 45; Theory and simulation • Macromolecular machines.
- [38] Rovigatti, L.; Russo, J.; Romano, F. *The European Physical Journal E* **2018**, *41*, 59.
- [39] De Michele, C.; Bellini, T.; Sciortino, F. *Macromolecules* **2012**, *45*, 1090–1106.
- [40] Zhang, Z.; Glotzer, S. C. *Nano Lett.* **2004**, *4*, 1407–1413.
- [41] Sciortino, F.; Giacometti, A.; Pastore, G. *Phys. Rev. Lett.* **2009**, *103*, 237801.
- [42] Cai, J.; Townsend, J. P.; Dodson, T. C.; Heiney, P. A.; Sweeney, A. M. *Science* **2017**, *357*, 564–569.
- [43] Perdew, J. P.; Burke, K.; Ernzerhof, M.; of Physics, D.; Quantum Theory Group Tulane University, N. O. L. . J. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- [44] Lieb, E. H.; Oxford, S. *International Journal of Quantum Chemistry*, *19*, 427–439.
- [45] Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, *45*, 13244–13249.
- [46] Weinhold, F.; Landis, C. R. In *Discov. Chem. with Nat. Bond Orbitals*; John Wiley & Sons, Inc., 2012; pp 307–308.
- [47] Weinhold, F.; Landis, C.; Glendening, E. *International Reviews in Physical Chemistry* **2016**, *35*, 399–440.
- [48] Ehrenfest, P. *Zeitschrift für Physik* **1927**, *45*, 455–457.

- [49] Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471–2474.
- [50] Marx, D.; Hutter, J. *Ab initio molecular dynamics: basic theory and advanced methods*; Cambridge University Press: Cambridge, 2009.
- [51] Vandevondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. *Comput. Phys. Commun.* **2005**, *167*, 103–128.
- [52] Verlet, L. *Phys. Rev.* **1967**, *159*, 98–103.
- [53] Swope, W. C. *J. Chem. Phys.* **1982**, *76*, 637.
- [54] Beeman, D. *Journal of Computational Physics* **1976**, *20*, 130 – 139.
- [55] Frenkel, D.; Smit, B. *Understanding Molecular Simulation*, 2nd ed.; Academic Press, Inc.: Orlando, FL, USA, 2001.
- [56] Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511–519.
- [57] Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.
- [58] Martyna, G.; Klein, M.; Tuckerman, M. *Journal of Chemical Physics* **1992**, *97*, 2635–2643.
- [59] Carter, E.; Ciccotti, G.; Hynes, J. T.; Kapral, R. *Chemical Physics Letters* **1989**, *156*, 472 – 477.
- [60] Sprik, M. **1998**, 437–445.
- [61] Tilocca, A.; Selloni, A. *The Journal of Chemical Physics* **2003**, *119*, 7445–7450.
- [62] Carloni, P.; Sprik, M.; Andreoni, W. *The Journal of Physical Chemistry B* **2000**, *104*, 823–835.
- [63] Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562–12566.

## Bibliography

---

- [64] Van Speybroeck, V.; De Wispelaere, K.; Van der Mynsbrugge, J.; Vandichel, M.; Hemelsoet, K.; Waroquier, M. *Chem. Soc. Rev.* **2014**, *43*, 7326–7357.
- [65] Hajek, B. *Math. Oper. Res.* **1988**, *13*, 311–329.
- [66] Tsallis, C.; Stariolo, D. A. *Physica A: Statistical Mechanics and its Applications* **1996**, *233*, 395 – 406.
- [67] Szu, H.; Hartley, R. *Physics Letters A* **1987**, *122*, 157 – 162.
- [68] Melo, M.; Bernardi, R.; Pascutti, P. *Biophys. J.* **2013**, *104*, 228; cited By 1.
- [69] Ghasemalizadeh, O.; Khaleghian, S.; Taheri., S. *Int. J. Adv. Res.* **2016**, *4*, 1668–1686.
- [70] Stille, J. K. *Angew. Chemie* **1986**, *25*, 508–524.
- [71] Negishi, E. *Acc. Chem. Res.* **1982**, *15*, 340–348.
- [72] Miyaura, N.; Suzuki, A.; January, W. *Chem. Rev.* **1995**, *95*, 2457–2483.
- [73] Seyferth, D. *Organometallics* **2009**, *28*, 1598–1605.
- [74] Schlenk, W.; Schlenk, W. *Aus d. Chem. Inst. d. Univ. Berlin.* **1929**, *62*, 920–924.
- [75] Ashby, E. C.; Nackashi, J.; Parris, G. E. *J. Am. Chem. Soc.* **1975**, *97*, 3162–3171.
- [76] Walker, F. W.; Ashby, E. C. *J. Am. Chem. Soc.* **1969**, *91*, 3845–3850.
- [77] Ashby, E. C. *Pure Appl. Chem.* **1980**, *52*, 545–569.
- [78] Guggenberger, L. J.; Rundle, R. E. *J. Am. Chem. Soc.* **1968**, *90*, 5375–5378.
- [79] Vallino, M. *J. Organomet. Chem.* **1969**, *20*, 1–10.

- [80] Sobota, P.; Duda, B. *J. Organomet. Chem.* **1987**, *332*, 239–245.
- [81] Sakamoto, S.; Imamoto, T.; Yamaguchi, K. **2001**, *10*–12.
- [82] Silverman, G. S.; Rakita, P. E. *Handbook of Grignard reagents*; CRC Press: New York, 1996.
- [83] Smith, M.; Becker, W. *Tetrahedron* **1967**, *23*, 4215–4227.
- [84] Smith, M. B.; Becker, W. E. *Tetrahedron Lett.* **1965**, *43*, 3843–3847.
- [85] Mori, T.; Kato, S. *J. Phys. Chem. A* **2009**, *113*, 6158–6165.
- [86] Chmgci, R.; Schlenk, C. T.; Axtен, J.; Troy, J.; Jiang, P.; Trachtman, M.; Bock, C. W. *Struct. Chem.* **1994**, *5*, 99–108.
- [87] Chen, Z.-N.; Fu, G.; Xu, X. *Org. Biomol. Chem.* **2012**, *10*, 9491–9500.
- [88] Tammiku, J.; Burk, P.; Tuulmets, A. *J. Phys. Chem. A* **2001**, *105*, 8554–8561.
- [89] Tammiku-Taul, J.; Burk, P.; Tuulmets, A. *J. Phys. Chem. A* **2004**, *108*, 133–139.
- [90] Ehlers, A. W.; van Klink, G. P. M.; van Eis, M. J.; Bickelhaupt, F.; Nederkoorn, P. H. J.; Lammertsma, K. *J. Mol. Model.* **2000**, *6*, 186–194.
- [91] Porsev, V. V.; Barsukov, Y. V.; Tulub, A. V. *Comput. Theor. Chem.* **2012**, *995*, 55–65.
- [92] Lioe, H.; White, J. M.; O’Hair, R. A. J. *J. Mol. Model.* **2011**, *17*, 1325–1334.
- [93] Jiménez-Halla, J. O. C.; Bickelhaupt, F. M.; Solà, M. J. *Organomet. Chem.* **2011**, *696*, 4104–4111.
- [94] Henriques, A. M.; Barbosa, A. G. H. *J. Phys. Chem. A* **2011**, *115*, 12259–12270.

## Bibliography

---

- [95] Ramirez, F.; Sarma, R.; Chaw, F.; Mccaffrey, T. M. **1977**, *2*, 5285–5289.
- [96] Pirinen, S.; Koshevoy, I. O.; Denifl, P.; Pakkanen, T. T. *Organometallics* **2013**, *32*, 4208–4213.
- [97] Pirinen, S.; Koshevoy, I. O.; Denifl, P.; Pakkanen, T. T. *Organometallics* **2013**, *32*, 4208–4213.
- [98] Ashby, E. C.; Becker, W. E. *J. Am. Chem. Soc.* **1963**, *85*, 118–119.
- [99] Iannuzzi, M.; Laio, A.; Parrinello, M. **2003**, 23–26.
- [100] Vuilleumier, R.; Sprik, M. *J. Chem. Phys.* **2001**, *115*, 3454–3468.
- [101] Lightstone, F. C.; Schwegler, E.; Hood, R. Q.; Gygi, F.; Galli, G. *Chem. Phys. Lett.* **2001**, *343*, 549–555.
- [102] Bernasconi, L.; Baerends, E. J.; Sprik, M. *J. Phys. Chem. B* **2006**, *110*, 11444–11453.
- [103] Blumberger, J.; Bernasconi, L.; Tavernelli, I.; Vuilleumier, R.; Sprik, M. *J. Am. Chem. Soc.* **2004**, *126*, 3928–3938.
- [104] Tammiku-Taul, J.; Burk, P.; Tuulmets, A. *J. Phys. Chem. A* **2004**, *108*, 133–139.
- [105] Guido, C. A.; Pietrucci, F.; Gallet, G. A.; Andreoni, W. *J. Chem. Theory Comput.* **2013**, *9*, 28–32.
- [106] Boero, M.; Ikeshoji, T.; Liew, C. C.; Terakura, K.; Parrinello, M.; Boero, M.; Ikeshoji, T.; Liew, C. C.; Terakura, K. **2004**, *137*, 6280–6286.
- [107] Vidossich, P.; Lledós, A.; Ujaque, G. In *Comput. Stud. Organomet. Chem.*; Macgregor, S. A.; Eisenstein, O., Eds.; Springer International Publishing: Cham, 2016; pp 81–106.
- [108] Vidossich, P.; Lledós, A.; Ujaque, G. *Acc. Chem. Res.* **2016**, *49*, 1271–1278.

- [109] Moret, M.-E.; Tavernelli, I.; Chergui, M.; Rothlisberger, U. *Chem. – A Eur. J.* **2010**, *16*, 5889–5894.
- [110] Dal Peraro, M.; Ruggerone, P.; Raugei, S.; Gervasio, F. L.; Carloni, P. *Curr. Opin. Struct. Biol.* **2007**, *17*, 149–156.
- [111] Dal Peraro, M.; Llarrull, L. I.; Rothlisberger, U.; Vila, A. J.; Carloni, P. *J. Am. Chem. Soc.* **2004**, *126*, 12661–12668.
- [112] Cascella, M.; Magistrato, A.; Tavernelli, I.; Carloni, P.; Rothlisberger, U. *Proc. Natl. Acad. Sci.* **2006**, *103*, 19641–19646.
- [113] De Vivo, M.; Dal Peraro, M.; Klein, M. L. *J. Am. Chem. Soc.* **2008**, *130*, 10955–10962.
- [114] Gossens, C.; Tavernelli, I.; Rothlisberger, U. *Chimia (Aarau)*. **2005**, *59*, 81–84.
- [115] Metz, D. J.; Glines, A. *J. Phys. Chem.* **1967**, *71*, 1158–1158.
- [116] Lippert, G.; Hutter, J.; Parrinello, M. *Theor. Chem. Acc.* **1999**, *103*, 124–140.
- [117] VandeVondele, J.; Hutter, J. *J. Chem. Phys.* **2007**, *127*, 114105.
- [118] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.
- [119] Hutter, J.; Iannuzzi, M.; Schiffmann, F.; VandeVondele, J. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 15–25.
- [120] Laio, A.; Gervasio, F. L. *Reports on Progress in Physics* **2008**, *71*, 126601.
- [121] Laio, A.; Rodriguez-Forte, A.; Gervasio, F. L.; Ceccarelli, M.; Parrinello, M. *The Journal of Physical Chemistry B* **2005**, *109*, 6714–6721; PMID: 16851755.
- [122] Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33–38.

## Bibliography

---

- [123] Frisch, M. J.; et al.; *Gaussian 09*; 2009.
- [124] Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257–2261.
- [125] Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. R. *J. Comput. Chem.* **1983**, *4*, 294–301.
- [126] Frisch, M. J.; Pople, J. A.; Binkley, J. S. *J. Chem. Phys.* **1984**, *80*, 3265.
- [127] Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.
- [128] Glendening, E. D.; Landis, C. R.; Weinhold, F. *J. Comput. Chem.* **2013**, *34*, 1429–1437.
- [129] *display*.
- [130] Weinhold, F.; Landis, C. R. *Discovering Chemistry with Natural Bond Orbitals*; Wiley, 2012.
- [131] Yamazaki, S.; Yamabe, S. *J. Org. Chem.* **2002**, *67*, 9346–9353.
- [132] Vollhardt, K.; Schore, N. *Organic Chemistry: Structure and Function*; W. H. Freeman, 2014.
- [133] Grignard, V. C. *ompt. rend. Hebd. Séances Acad. Sci.* **1900**, *130*, 1322–1324.
- [134] Fauvarque, J.; Rouget, E. *Hebd. Séances Acad. Sci., Ser. C* **1968**, *257*, 1355; cited By 5.
- [135] Blomberg, C.; Grootveld, H.; Gerner, T.; Bickelhaupt, F. *Journal of Organometallic Chemistry* **1970**, *24*, 549 – 553.
- [136] Ashby, E. C.; Lopp, I. G.; Buhler, J. D. *Journal of the American Chemical Society* **1975**, *97*, 1964–1966.
- [137] *Chemical Society reviews* **2003**, *32*, 225–230.

- [138] Garst, J. F.; Soriaga, M. P. *Coord. Chem. Rev.* **2004**, *248*, 623–652.
- [139] Shao, Y.; Liu, Z.; Huang, P.; Liu, B. *Phys. Chem. Chem. Phys.* **2018**, *20*, 11100–11108.
- [140] Ignatov, S. K.; Panteleev, S. V.; Maslennikov, S. V.; Spirina, I. V. *Russian Journal of General Chemistry* **2012**, *82*, 1954–1961.
- [141] Goerigk, L.; Grimme, S. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670–6688.
- [142] Zhao, Y.; Truhlar, D. G. *Journal of Chemical Theory and Computation* **2005**, *1*, 415–432; PMID: 26641508.
- [143] Adamo, C.; Barone, V. *The Journal of Chemical Physics* **1999**, *110*, 6158–6170.
- [144] Becke, A. D. *The Journal of Chemical Physics* **1993**, *98*, 5648–5652.
- [145] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
- [146] Zhao, Y.; Truhlar, D. G. *Theoretical Chemistry Accounts* **2008**, *120*, 215–241.
- [147] Chai, J.-D.; Head-Gordon, M. *The Journal of Chemical Physics* **2008**, *128*, 084106.
- [148] Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.
- [149] Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. *The Journal of Chemical Physics* **2003**, *119*, 12129–12137.
- [150] *CFOUR, a quantum chemical program package written by J.F. Stanton, J. Gauss, L. Cheng, M.E. Harding, D.A. Matthews, P.G. Szalay with contributions from A.A. Auer, R.J. Bartlett, U. Benedikt, C. Berger, D.E. Bernholdt, Y.J. Bomble, O. Christiansen, F. Enge.*

## Bibliography

---

- [151] Frisch, M. J.; et al.; *Gaussian?09 Revision {D}.01*.
- [152] Peltzer, R. M.; Eisenstein, O.; Nova, A.; Cascella, M. *The Journal of Physical Chemistry B* **2017**, *121*, 4226–4237; PMID: 28358509.
- [153] VandeVondele, J.; Hutter, J. *J. Chem. Phys.* **2007**, *127*, 114105.
- [154] Goedecker, S.; Teter, M.; Hutter, J. *Phys. Rev. B* **1996**, *54*, 1703–1710.
- [155] Lund, T.; Pedersen, M. L.; Frandsen, L. A. *Tetrahedron Letters* **1994**, *35*, 9225 – 9226.
- [156] Tappel, A. L. *Vitam. Horm.* **1962**, *20*, 493–510.
- [157] Packer, L.; Weber, S. U.; Rimbach, G. *The Journal of Nutrition* **2001**, *131*, 369S–373S.
- [158] Kamal-Eldin, A.; Appelqvist, L.-Å. *Lipids* **1996**, *31*, 671–701.
- [159] Brigelius-Flohe, R.; Traber, M. G. *Faseb J* **1999**, *13*, 1145–1155.
- [160] Kowdley, K. V.; Mason, J. B.; Meydani, S. N.; Cornwall, S.; Grand, R. J. *Gastroenterology* **1992**, *102*, 2139 – 2142.
- [161] Di Donato, I.; Bianchi, S.; Federico, A. *Neurological Sciences* **2010**, *31*, 511–515.
- [162] Traber, M. G.; Sokol, R. J.; Burton, G. W.; Ingold, K. U.; Papas, A. M.; Huffaker, J. E.; Kayden, H. J. *J. Clin. Invest.* **1990**, *85*, 397–407.
- [163] Traber, M. G.; Burton, G. W.; Hamilton, R. L. *Ann. N. Y. Acad. Sci.* **2004**, *1031*, 1–12.
- [164] Kono, N.; Ohto, U.; Hiramatsu, T.; Urabe, M.; Uchida, Y.; Satow, Y.; Arai, H. *Science* **2013**, *340*, 1106–1110.
- [165] Meier, R.; Tomizaki, T.; Schulze-Briese, C.; Baumann, U.; Stocker, A. *J. Mol. Biol.* **2003**, *331*, 725–734.

- [166] Min, K. C.; Kovall, R. a.; Hendrickson, W. a. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 14713–14718.
- [167] Sha, B.; Phillips, S. E.; Bankaitis, V. A.; Luo, M. *Nature* **1998**, *391*, 506–510.
- [168] Stocker, A.; Tomizaki, T.; Schulze-Briese, C.; Baumann, U. *Structure* **2002**, *10*, 1533–1540.
- [169] He, X.; Lobsiger, J.; Stocker, A. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 18545–18550.
- [170] Stocker, A. *Ann. N. Y. Acad. Sci.* **2004**, *1031*, 44–59.
- [171] Chung, S.; Ghelfi, M.; Atkinson, J.; Parker, R.; Qian, J.; Carlin, C.; Manor, D. *J. Biol. Chem.* **2016**, *291*, 17028–17039.
- [172] Aeschimann, W.; Staats, S.; Kammer, S.; Olieric, N.; Jeckelmann, J. M.; Fotiadis, D.; Netscher, T.; Rimbach, G.; Cascella, M.; Stocker, A. *Sci. Rep.* **2017**, *7*, 1–13.
- [173] Saraiva, C.; Praça, C.; Ferreira, R.; Santos, T.; Ferreira, L.; Bernardino, L. *J. Control. Release* **2016**, *235*, 34–47.
- [174] Deli, M. A.; Ábrahám, C. S.; Kataoka, Y.; Niwa, M. *Cell. Mol. Neurobiol.* **2005**, *25*, 59–127.
- [175] Kraft, D. J.; Ni, R.; Smallenburg, F.; Hermes, M.; Yoon, K.; Weitz, D. A.; van Blaaderen, A.; Groenewold, J.; Dijkstra, M.; Kegel, W. K. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 10787–10792.
- [176] Wilber, A. W.; Doye, J. P. K.; Louis, A. A.; Noya, E. G.; Miller, M. A.; Wong, P. *J. Chem. Phys.* **2007**, *127*, 085106.
- [177] Carlsson, F.; Linse, P.; Malmsten, M. *J. Phys. Chem. B* **2001**, *105*, 9040–9049.
- [178] Fantoni, R.; Gazzillo, D.; Giacometti, A.; Miller, M. A.; Pastore, G. *J. Chem. Phys.* **2007**, *127*, 234507.

## Bibliography

---

- [179] Villar, G.; Wilber, A. W.; Williamson, A. J.; Thiara, P.; Doye, J. P.; Louis, A. A.; Jochum, M. N.; Lewis, A. C.; Levy, E. D. *Phys. Rev. Lett.* **2009**, *102*, 118106.
- [180] Liu, H.; Kumar, S. K.; Douglas, J. F. *Phys. Rev. Lett.* **2009**, *103*, 018101.
- [181] Li, Y.; Shi, T.; An, L.; Huang, Q. *J. Phys. Chem. B* **2012**, *116*, 3045–3053.
- [182] Li, W.; Persson, B. A.; Morin, M.; Behrens, M. A.; Lund, M.; Zackrisson Oskolkova, M. *J. Phys. Chem. B* **2015**, *119*, 503–508.
- [183] Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037–10041.
- [184] Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C.; Brook, S.; Brook, S.; Brook, S. *Proteins* **2006**, *65*, 712–725.
- [185] Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087–1092.
- [186] Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: New York, NY, USA, 1987.
- [187] Wu, D.; Chandler, D.; Smit, B. *J. Phys. Chem.* **1992**, *96*, 4077–4083.
- [188] Miller, M. A.; Amon, L. M.; Reinhardt, W. P. *Chem. Phys. Lett.* **2000**, *331*, 278–284.
- [189] Jones, J. E. *Proceedings of the Royal Society of London Series A* **1924**, *106*, 463–477.

# Paper I

# How solvent dynamics controls the Schlenk equilibrium of Grignard reagents: a computational study of CH<sub>3</sub>MgCl in tetrahydrofuran.

Raphael M. Peltzer,<sup>a</sup> Odile Eisenstein,<sup>a,b</sup> Ainara Nova,<sup>\*a</sup> Michele Cascella<sup>\*a</sup>

<sup>a</sup> Department of Chemistry and Centre for Theoretical and Computational Chemistry  
(CTCC), University of Oslo, Postbox 1033 Blindern 0315 Oslo, Norway

<sup>b</sup> Institut Charles Gerhardt, UMR 5253 CNRS-Université de Montpellier, cc 1501, Place E.  
Bataillon, Université de Montpellier 34095 Montpellier, France.

\* email to: [ainara.nova@kjemi.uio.no](mailto:ainara.nova@kjemi.uio.no), [michele.cascella@kjemi.uio.no](mailto:michele.cascella@kjemi.uio.no)

**Keywords:** Transmetallation mechanism, metadynamics, di-alkyl magnesium, dichloride magnesium.

## ABSTRACT

The Schlenk equilibrium is a complex reaction governing the presence of multiple chemical species in solution of Grignard reagents. The full characterization at the molecular level of the transformation of  $\text{CH}_3\text{MgCl}$  into  $\text{MgCl}_2$  and  $\text{Mg}(\text{CH}_3)_2$  in tetrahydrofuran (THF) by means of *ab initio* molecular dynamics simulations with enhanced-sampling metadynamics is presented. The reaction occurs via formation of dinuclear species bridged by chlorine atoms. At room temperature, the different chemical species involved in the reaction accept multiple solvation structures, with two to four THF molecules that can coordinate the Mg atoms. The energy difference between all dinuclear solvated structures is lower than 5 kcal mol<sup>-1</sup>. The solvent is shown to be a direct key player driving the Schlenk mechanism. In particular, this study illustrates how the most stable symmetrically solvated dinuclear species,  $(\text{THF})\text{CH}_3\text{Mg}(\mu\text{-Cl})_2\text{MgCH}_3(\text{THF})$  and  $(\text{THF})\text{CH}_3\text{Mg}(\mu\text{-Cl})(\mu\text{-CH}_3)\text{MgCl}(\text{THF})$ , need to evolve to *less* stable asymmetrically solvated species,  $(\text{THF})\text{CH}_3\text{Mg}(\mu\text{-Cl})_2\text{MgCH}_3(\text{THF})_2$  and  $(\text{THF})\text{CH}_3\text{Mg}(\mu\text{-Cl})(\mu\text{-CH}_3)\text{MgCl}(\text{THF})_2$ , in order to yield ligand exchange or product dissociation. In addition the transferred ligands are always departing from an axial position of a penta-coordinated Mg atom. Thus, the solvent dynamic is key to successive Mg-Cl and Mg-CH<sub>3</sub> bond cleavages since cleavage of bond occurs at the most solvated Mg and formation of bonds at the least solvated one. The dynamics of solvent also contributes to keep relatively flat the free energy profile of the Schlenk equilibrium. These results shed light to one of the most used organometallic reagents whose structures in solvent remains experimentally unresolved. These results may also help to developing more efficient catalyst for reactions involving these species.

## INTRODUCTION

In 1900, the presentation by Victor Grignard of a new compound, completely soluble in ether and formed upon reaction of magnesium with alkyl halides,<sup>1</sup> paved the way to the establishment of one of the most widely used organometallic reagents in organic synthesis, worth the Noble prize in 1912.<sup>2</sup> The so-called *Grignard reagent*, with nominal formula RMgX, is used in nucleophilic addition or substitution reactions, taking advantage of the enriched electron density in the R group coordinated to the Mg atom.

Initial limitations to the direct use of this reagent were found due to its relatively modest tolerance to different functional groups compared to that of corresponding boron and zinc metals. However, the use of Ni catalysts on coupling reactions using the Grignard reagent, independently developed by Corriu and Massé, and Kumada and Tamao, increased the relevance of this reagent.<sup>3-4</sup> Nowadays, the interest in the functionalization of phenol-based derivatives involving earth abundant metal catalysts and the development of efficient synthetic protocols allowing to work at very mild condition and with large functional tolerance make organomagnesium reagents very attractive.<sup>5-11,12</sup>

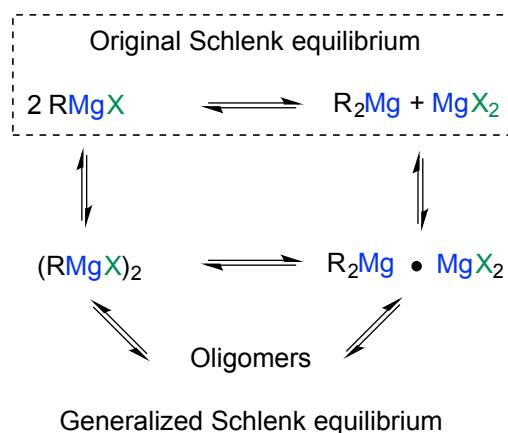
Despite the use of Grignard reagents in organic synthesis, their molecular structure in solution has remained elusive so far, even if considerable research effort has been spent on this topic. In an illuminating essay, Seydel described the difficulties in characterizing the structures of Grignard reagents, which can be affected by both the solvent, and the halide and organic groups bonded to Mg.<sup>13</sup> Solid state structures determined by X-ray diffraction studies revealed a diversity of coordination modes for Mg and nuclearity of the molecular system. A tetrahedral coordination is commonly observed in crystals structures obtained from ether solutions as EtMgBr(OEt<sub>2</sub>)<sub>2</sub>,<sup>14</sup> where two molecules of solvent are bound to Mg. Instead, both trigonal bipyramidal and square pyramidal structures were found for CH<sub>3</sub>MgBrTHF<sub>3</sub> when working in THF solvent.<sup>15</sup> X-ray data of aggregates of two EtMg<sub>2</sub>Cl<sub>3</sub> moieties showed a

greater than four coordination for the Mg centers.<sup>16,17</sup> Due to the lattice packing restrictions, the structures of Grignard reagents observed in the crystalline solid state may not be necessarily representative of the stable species present in solution. Investigations using molecular weight,<sup>18</sup> calorimetric measurements,<sup>19</sup> and NMR and IR spectroscopy<sup>20</sup> support in fact the existence of structures other than those detected by X-ray crystallography.

The determination of the structures of Grignard reagents in solution is particularly complex because of the presence of multiple chemical species at thermodynamic equilibrium. This complication has been evident since 1929, when Schlenk father and son<sup>21</sup> proposed that redistribution of ligands yielding  $MgR_2$  and  $MgX_2$  from  $RMgX$  could take place (original Schlenk equilibrium in Scheme 1). At present, it is commonly recognized that the halide atoms in  $RMgX$  and  $MgX_2$  tend to form bridges between magnesium atoms to yield dimers and oligomeric structures, which are in equilibrium with monomeric species (generalized Schlenk equilibrium in Scheme 1). This causes the association between the various moieties, eventually favoring ligand exchange. The degree of association among the Mg centers is dependent on the solvent, as shown by Ashby and Walker, who found a degree of association around ~1 in THF and between 1 and 4, depending on the nature of R and X, in diethyl ether.<sup>22</sup> Further studies suggested that some complex equilibrium between  $RMgCl(THF)_n$ ,  $MgR_2(\mu-Cl)_2Mg(THF)_{5/4}$  and  $RMg_2(\mu-Cl)_3(THF)_5$  better describes the traditional Schlenk equilibrium.<sup>23-24</sup>

Given the complexity of the problem, computational modeling offers an excellent tool to determine in detail the chemical nature of the species involved in the Schlenk equilibrium. In the past, examination of the Cambridge Crystallographic Data base and Density Functional Theory (DFT)<sup>25-26,27</sup> calculations for isolated molecules led to establish the overwhelming prevalence of structures with two over three bridging ligands, and a preference for halide over alkyl ligands in that position. Moreover, they observed that the maximum stabilizing effect of

the coordinating solvent is obtained for dimeric moieties.<sup>28</sup> Including the solvent as a continuum model in a systematic way and combining Generalized Valence Bond, second order Møller-Plesset perturbation theory (MP2) and DFT calculations indicated that radical and charged species may be present at the equilibrium, although in small concentrations.<sup>29</sup> Explicit solvent was limited to the first coordination sphere of the Mg atom and in many cases the total number of ligands at Mg including the solvent was fixed to four. However, both crystallographic<sup>30-32</sup> and computational studies<sup>33</sup> showed that Mg could expand its coordination sphere over more than four ligands, up to hexacoordination. This limitation is probably even more inappropriate when trying to understand the dynamic processes involved in the Schlenk equilibrium (Scheme 1).



**Scheme 1:** Schlenk equilibria.

In consideration of the longstanding and recently renewed importance of Grignard in organic synthesis,<sup>34-36</sup> it is important to reach a better understanding of its structure and dynamic behavior in solvent. In this respect, dynamic modeling approaches may constitute the optimal choice for an accurate description of the Grignard reagent. In the past decades, *ab initio* molecular dynamics simulations (AIMD) have been shown to be especially appropriate to study processes in the liquid phase.<sup>37-38</sup> For example, this technique was used to study

structural properties of various metal ions in solution, including Na,<sup>39</sup> Mg<sup>40-41</sup>, or Co and Ag.<sup>42</sup> In addition, AIMD with enhanced sampling methods by means of metadynamics were effectively used to provide an accurate description of solvent effects in various organic<sup>43-44</sup> and organometallic reactions.<sup>45-46</sup> Larger metallo-organic systems have been treated coupling AIMD to molecular mechanics schemes.<sup>47</sup> This approach was used to describe for example electronic solute-solvent coupling in photo-excitation dynamics of ruthenium metallorganic compounds,<sup>48</sup> the reaction mechanisms of hydrolysis in zinc  $\beta$ -lactamases<sup>49-50</sup>, the redox properties of copper in azurin,<sup>51</sup> the functional role of Mg<sup>2+</sup> ions in ribonuclease H<sup>52</sup> and the design of organo-ruthenium anticancer complexes.<sup>53</sup>

In this work, this methodology was used to investigate the equilibria that take place when CH<sub>3</sub>MgCl, a model of Grignard reagent, is dissolved in tetrahydrofuran (THF). Our study was also complemented by DFT static calculations of selected structures identified as minima by the ab initio molecular dynamics. The data presented in this work show that multiple structures with different solvation patterns are in equilibrium, and evidence how the dynamic exchange of solvent molecules in the Mg coordination sphere is key to the evolution from the reactant to product in the Schlenk equilibrium.

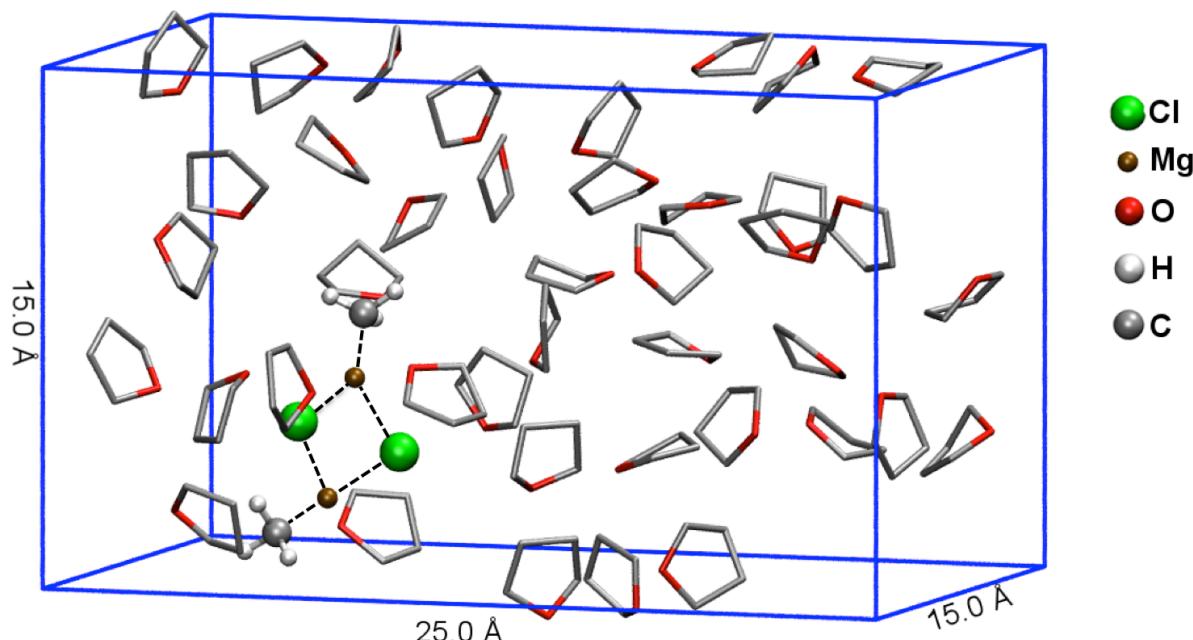
## COMPUTATIONAL METHODS

**System Setup** The Schlenk equilibrium was investigated through a set of simulations describing monomeric reactants, monomeric products, and different intermediate dimeric states.

*Monomeric species* We built three different systems, containing each, one molecule of either CH<sub>3</sub>MgCl, MgCl<sub>2</sub> or Mg(CH<sub>3</sub>)<sub>2</sub> surrounded by 25 molecules of THF in a periodic box of dimensions 15.0 x 15.0 x 15.0 Å<sup>3</sup>. The initial coordinates of THF were obtained from a previous 20 ps-long equilibration run in a similar box containing the pure solvent. The number of solvent molecules was set to be consistent with the experimental density of THF at room temperature.<sup>54</sup> All systems were first relaxed for 15 ps in the microcanonical ensemble

at energies matching an average temperature of 300 K. Relaxation at the target temperature was performed using canonical sampling/velocity rescaling (CSVR) thermostat with a time constant of 10 fs until the temperature of the system oscillated around the target value. Production runs were simulated in the NVT ensemble at 300 K. The Nosé -Hoover chain thermostat with a chain length of 3 and time constant 1 ps was used for data production.<sup>55-57</sup>

*Dimeric species* The  $(\text{MgCH}_3\text{Cl})_2$  species were simulated in an orthorhombic periodic box of dimensions  $25.2 \times 15.0 \times 15.0 \text{ \AA}^3$ , containing 42 THF molecules. A larger box of dimensions  $25.2 \times 20.0 \times 20.0 \text{ \AA}^3$  with 74 THF molecules was also simulated to verify that there is no significant bias due to finite-size effects (see Supporting Information). The systems were thermalized following the same protocol as described for the monomers.



**Figure 1.** Simulation box used in this study. The atoms of Grignard reagent are represented by spheres and the THF solvent molecules by sticks in licorice and red. Hydrogen atoms of THF are not shown for clarity.

**Ab-initio Molecular Dynamics simulations** The electronic problem was solved by DFT<sup>25-26</sup> using the Perdew-Burke-Ernzerhof approximation to the exchange correlation functional.<sup>58</sup> Kohn-Sham orbitals were expanded over mixed Gaussian and plane-wave basis functions.<sup>59</sup> The DZVP basis set for first and second row elements and Mg, and a molecularly optimized basis set for the chlorine atoms were employed.<sup>60</sup> The auxiliary plane wave basis set was expanded to a 200 Ry cutoff. The core electrons were integrated out

using pseudopotentials of the Goedecker-Teter-Hutter type.<sup>61</sup> Dispersion forces were accounted for using the D3 Grimme approximation.<sup>62</sup> AIMD simulations were run over the Born-Oppenheimer surface, with a time-step of 0.25 fs, optimizing the energy gradient to a threshold of  $10^{-5}$  au.

**Free-energy calculations** The exploration of the conformational and reactive landscape, and determination of the corresponding free-energy surface, was performed by coupling AIMD to metadynamics simulations.<sup>37, 63</sup> The solvation of different chemical species observed during the AIMD runs, as well as the transmetallation reaction, were investigated by independent metadynamics runs. All collective variables (CVs) employed in this study were defined as the coordination-number of specific ligand species to individual Mg atoms. The coordination number of a species X around Mg at a given time  $t$  ( $CN[X](t)$ ) was evaluated according to the formula shown in equation 1, as defined in previous works:  
38

$$CN[X](t) = \sum_i^{N_X} \frac{1 - \left(\frac{d_i(t)}{d_0}\right)^p}{1 - \left(\frac{d_i(t)}{d_0}\right)^q} \quad (1)$$

Where  $N_X$  is the number of species X present in the system,  $d_i$  is the distance of the  $i^{th}$  atom X from Mg, and  $d_0$ ,  $p$  and  $q$  are free parameters (see Supporting Information).

The time-dependent bias potential was formed by sets of Gaussians of 0.25 kcal mol<sup>-1</sup> height and 0.04 width for coordination variables and added every 50 steps of AIMD for the di-chloride bridged system. Additional details of the Metadynamics parameters used as well as the associated errors is given in the Table S1 of the Supporting Information.

The metadynamics simulations convergence was confirmed by checking that the calculations reached the diffusion limit within their wall constraints; statistical errors, computed according to references<sup>64-65</sup>, are within 1-3 kT (see Supporting Information).

AIMD runs were computed using the QUICKSTEP<sup>66, 67</sup> module of the CP2K 2.5.1 package. Trajectory analysis was performed using the tools available in the VMD 1.9.2 package.<sup>68</sup>

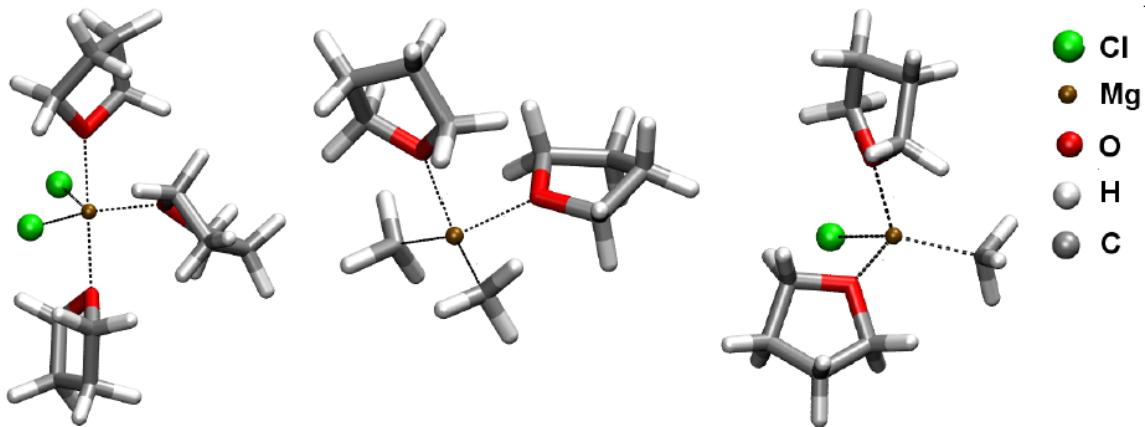
**Static calculations and electronic structure analysis** Chemically relevant geometries sampled by AIMD were fully optimized at the DFT(PBE+GD3) level by using the Gaussian09 software package.<sup>69</sup> Mg, C, H and O were described with the all-electron double-

$\zeta$  6-31+G\*\* basis set.<sup>70-72</sup> Vibrational frequencies were computed analytically to verify that the stationary points found were energy minima. In addition to the solvent molecules bound to Mg, implicit solvation was modeled by using the SMD solvation model.<sup>73</sup> In selected cases, Gibbs energies were obtained for T = 298.15 K and p = 1 atm. In the bimolecular steps, these energies were corrected for the 1M standard state (T = 298.15 K and p = 24.465 atm). Donor-acceptor interactions were explored by means of natural bond orbital calculations (NBO6 version).<sup>74</sup> The nature of these interactions was determined by computing the associated natural localized molecular orbitals (NLMO).<sup>75</sup>

## RESULTS

We restricted our investigation to monomers and di-nuclear species, excluding higher aggregation states, because experimental studies showed that in THF the aggregation is very low, mainly consisting of monomeric species.<sup>22</sup>

**Monomeric structures** The solvation of CH<sub>3</sub>MgCl, Mg(CH<sub>3</sub>)<sub>2</sub> and MgCl<sub>2</sub> species were investigated over 40 ps of AIMD simulations at 300 K. CH<sub>3</sub>MgCl and Mg(CH<sub>3</sub>)<sub>2</sub> were only found with two THF molecules in the solvation sphere of Mg and a tetrahedral coordination at this atom (Figure 2). On the contrary, both tetrameric- and pentameric coordination were found for MgCl<sub>2</sub>. The nature of the solvation of MgCl<sub>2</sub> was then further investigated by metadynamics simulations, using the coordination number of oxygen atoms of THF to Mg as the collective variable (CV). These simulations showed that the most likely solvation state is MgCl<sub>2</sub>(THF)<sub>3</sub> arranged in a trigonal-bipyramidal coordination at Mg. MgCl<sub>2</sub>(THF)<sub>3</sub> is ~1.6 kcal mol<sup>-1</sup> and ~2.8 kcal mol<sup>-1</sup> more stable than the tetrahedral or octahedral complexes with two and four coordinating THF molecules, respectively. Thus, at room temperature the different solvation states are in equilibrium in solution in a 0.07:1.0:0.01 ratio for the di-, tri- and tetra-solvated complexes, yielding an average MgCl<sub>2</sub>(THF)<sub>x</sub> solvation with value x= 2.9. DFT optimization of the structures located as minima on the free energy surface (FES) showed that they are also minima on the potential energy surface (PES) with similar structural features (these structures are shown in Figure 2). Natural bond orbital analysis (NBO) on these optimized structures showed a higher electron donation of CH<sub>3</sub> to Mg compared to Cl. This is shown by the Mg contribution on the lone pairs NLMOs of CH<sub>3</sub> and Cl ( $\chi^{Mg}$ ), which are 10.2% and 5.3%, respectively, in CH<sub>3</sub>MgCl (see Figure S1 and Table S1 for further details). This may explain the need of additional coordinating THF in MgCl<sub>2</sub> to improve the screening of the Mg charge.



**Figure 2.** Most likely solvation structures for  $\text{MgCl}_2$  (left),  $\text{Mg}(\text{CH}_3)_2$  (middle) and  $(\text{CH}_3)\text{MgCl}$  (right). The dotted black lines represent the coordination of the ligands to the Mg center.

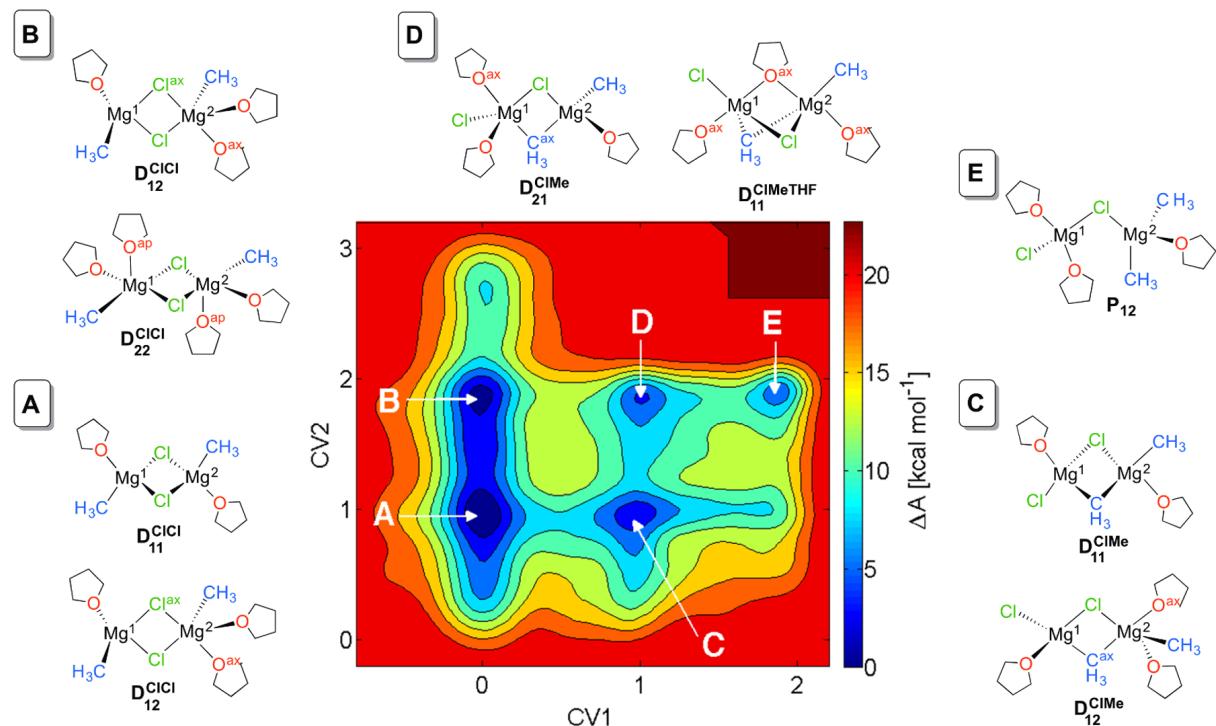
**Dinuclear structures** The formation of dinuclear species was first studied by performing an AIMD simulation at 300 K starting from two molecules of  $\text{CH}_3\text{MgCl}$  placed with their respective Mg atoms at a distance of 9.5 Å. In order to accelerate the dimerization, a bias restraining potential to induce the approach of the two molecules was used (see Computational Methods Section). A dimeric structure with two bridging chlorine atoms was formed as soon as the two Mg atoms were at a distance of 5.5 Å.

**Cl/CH<sub>3</sub> exchange process** Chloride/methyl exchange between the two Mg centers was investigated combining AIMD and metadynamics simulations, using as CVs the THF solvation number on one arbitrary Mg (CV1) and the difference in methyl coordination between the two Mg atoms (CV2). The CVs where chosen to monitor the free energy change upon methyl transfer between the Mg atoms as a function of the local solvation of Mg.

The resulting FES (Figure 3) shows five separated wells (**A**, **B**, **C**, **D** and **E**) that can be classified according to the nature of the Mg bridging groups. The most representative structures captured in each well are shown in Figure 3. Wells **A** and **B** correspond to dichloride-bridged dinuclear species, **C** and **D** correspond to the mixed Me and Cl bridged species, and **E** encloses monochloride bridged structures of formula  $\text{ClMg}(\mu\text{-Cl})\text{MgMe}_2$ . While wells **A** and **B** are representing the structures of the reactants, well **E** represents the pre-product state prior to final dissociation into  $\text{MgCl}_2$  and  $\text{Mg}(\text{CH}_3)_2$ .

We introduce a notation for describing the nature of the dimer **D**. The notation  $D_{ij}^{XY}$  stands for dimer (**D**) where  $i$  and  $j$  are the numbers of THF on  $\text{Mg}^1$  and  $\text{Mg}^2$ , respectively and X and

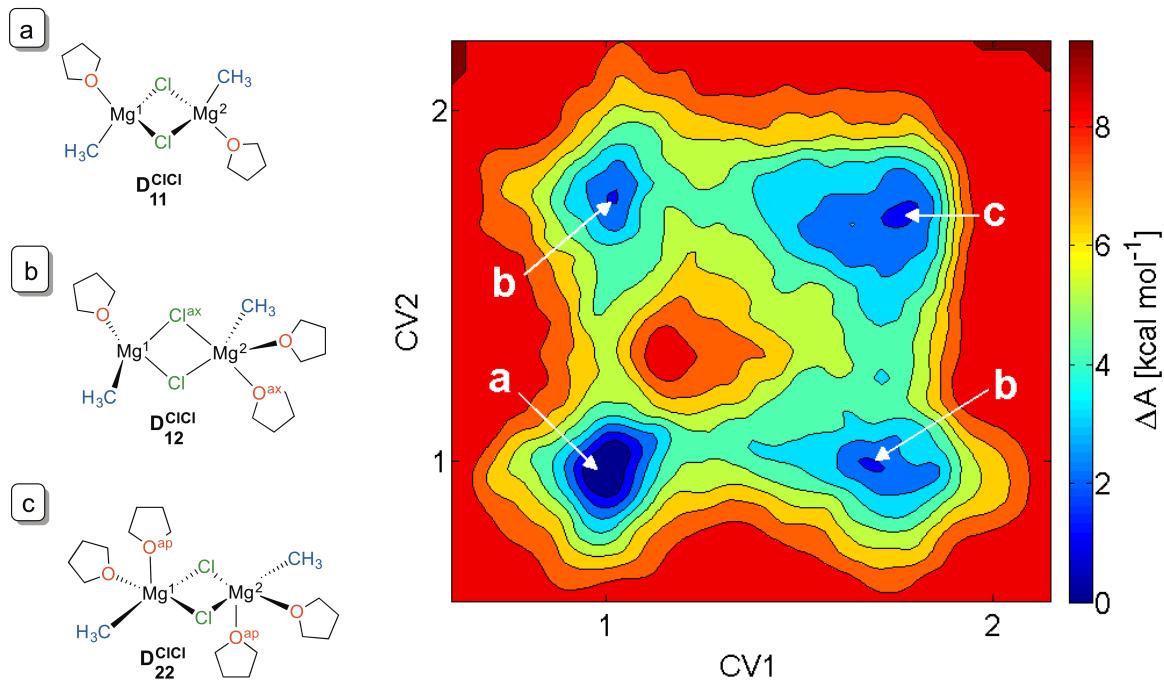
$\text{Y}$  describe the nature of the bridging ligands.  $\text{D}^{\text{XY}}$  describes the entire family of  $\text{X}$ , and  $\text{Y}$  bridging species regardless of the number of THF on either Mg. The FES plot in Figure 3 shows that the ensemble of  $\text{D}^{\text{ClCl}}$  structures is more stable by  $\sim 3.5$  kcal mol $^{-1}$  than the one for  $\text{D}^{\text{CIMe}}$ . In addition, the FES shows a clear path connecting the **A** - **E** wells, thus giving information on the ligand exchange reaction. The transformation from wells **A**-**B** to **D** goes preferentially via well **C**, while exchange connecting directly **B** to **D** does not constitute a similarly viable path. The energy barriers for the **A**-**C** and **C**-**D** transition are  $\sim 8$  kcal mol $^{-1}$  and  $\sim 6$  kcal mol $^{-1}$ , respectively, while the free energy barrier between **B** and **D** is 13 kcal mol $^{-1}$  higher than well **A**. The preference for this path was analyzed by studying more closely the solvation structures of  $\text{D}^{\text{ClCl}}$  and  $\text{D}^{\text{CIMe}}$ .



**Figure 3.** Free-energy surface of the Schlenk-Equilibrium. The collective variables for this representation are the difference in Mg-CH<sub>3</sub> coordination number between Mg<sup>2</sup> and Mg<sup>1</sup> (CV1) and the THF-coordination number to Mg<sup>1</sup> (CV2). The chemical structures drawn in the figure depict the most representative species obtained for wells **A**, **B**, **C**, **D** and **E**.

**Solvation of dichloro-bridged structures** The accessible solvation states of  $\text{D}^{\text{ClCl}}$  in the **A** and **B** wells (Figure 3) were investigated by metadynamics simulations using the coordination number of the oxygen atoms of THF to each of the Mg atoms as CVs. The simulations revealed the existence of three solvation states, corresponding to distinct free-energy minima reported in Figure 4. The solvation state with the lowest free energy corresponds to a

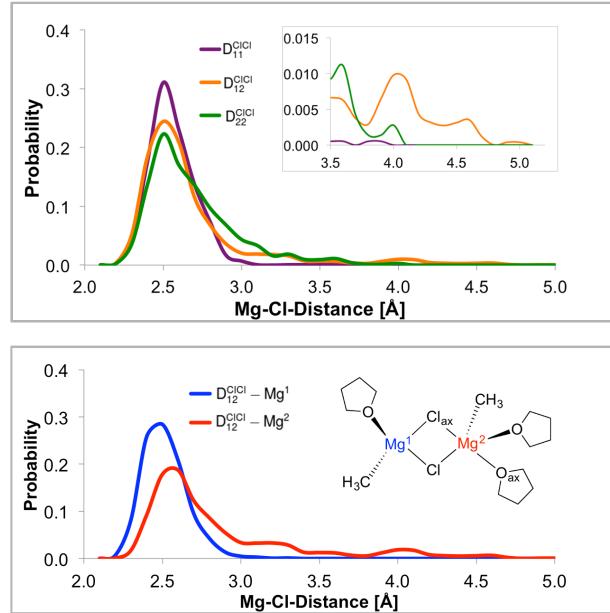
symmetric dimeric structure in which each Mg atom is solvated by a single THF ( $\mathbf{D}^{\text{ClCl}}_{11}$ ). The second most stable solvation state, only 1.3 kcal mol<sup>-1</sup> higher in energy, corresponds to another symmetric adduct with two THF molecules coordinating each Mg ( $\mathbf{D}^{\text{ClCl}}_{22}$ ). In the later, each Mg is five coordinated. The interconversion between these two structures occurs through a metastable asymmetric solvation state  $\mathbf{D}^{\text{ClCl}}_{12}$  with one THF solvent on one Mg and two on the other. This asymmetric solvation state  $\mathbf{D}^{\text{ClCl}}_{12}$ , which is 2.7 kcal mol<sup>-1</sup> higher in energy than  $\mathbf{D}^{\text{ClCl}}_{11}$ , is expected to be statistically present in 1% of room-temperature samples, also taking into account degenerate states. Higher solvation states with three THF in the first solvation shell of a single Mg-center are more than 5 kcal mol<sup>-1</sup> higher in energy than  $\mathbf{D}^{\text{ClCl}}_{11}$ , and therefore are expected to be present only in minor amounts in room-temperature samples. Both the relative stability and activation energies for the interchange between  $\mathbf{D}^{\text{ClCl}}_{11}$ ,  $\mathbf{D}^{\text{ClCl}}_{12}$  and  $\mathbf{D}^{\text{ClCl}}_{22}$  are low (less than 5 kcal mol<sup>-1</sup>), implying fast interconversion between the different structures at room temperature.



**Figure 4.** Solvated  $\mathbf{D}^{\text{ClCl}}$  structures found by metadynamics simulations (left) and corresponding FES (right). CV1 and CV2 are defined as the coordination numbers of THF at  $\text{Mg}^1$  and  $\text{Mg}^2$ , respectively, following equation 1.

Analysis of the structural changes from  $\mathbf{D}^{\text{ClCl}}_{11}$  to  $\mathbf{D}^{\text{ClCl}}_{12}$  showed that the addition of THF to  $\text{Mg}^2$  (Figure 4) occurs always anti to one of the chloride atoms, producing a local trigonal-bipyramidal coordination with the bridging Cl's in axial and equatorial positions. AIMD

simulations indicated significant lengthening of the Mg-Cl<sub>ax</sub> bond relative to the Mg-Cl<sub>eq</sub> one. Reversible cleavage of the axial Cl-Mg bond, and consequent formation of transient single chloride-bridged structures were observed during the analysis of AIMD trajectories. In order to analyze the influence of the solvent on the Mg-Cl cleavage, the statistical distribution of the Mg-Cl bond lengths in  $\mathbf{D}^{\text{ClCl}}_{11}$ ,  $\mathbf{D}^{\text{ClCl}}_{12}$  and  $\mathbf{D}^{\text{ClCl}}_{22}$  was computed (Figure 5). As expected, an increase in solvation is associated with a statistical elongation of the Mg-Cl bonds.



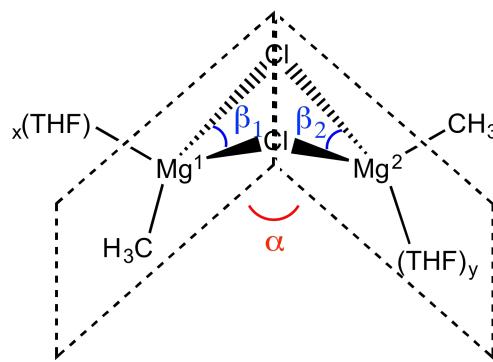
**Figure 5.** (a) Mg-Cl distance distributions in  $\mathbf{D}^{\text{ClCl}}$  dimeric structures. (b) Mg<sup>1</sup>-Cl (blue) and Mg<sup>2</sup>-Cl (red) distance distributions in  $\mathbf{D}^{\text{ClCl}}_{12}$ . Mg-Cl bond cleavage is observed when the Mg-Cl distance is larger than 3.7 Å.

Strikingly, Mg-Cl bond cleavage was observed only in the asymmetric  $\mathbf{D}^{\text{ClCl}}_{12}$  structure with statistically significant probability of finding Mg-Cl distances with longer values than 3.7 Å, as shown in the inset of Figure 5a. On the contrary, both the Mg-Cl bond length distributions of the least and the most solvated  $\mathbf{D}^{\text{ClCl}}_{11}$  and  $\mathbf{D}^{\text{ClCl}}_{22}$  species vanish above 3.7 Å, indicating statistical irrelevance of Mg-Cl bond cleaved structures. Analysis of the AIMD trajectory evidenced that in all cases the cleavage of the Mg-Cl bond occurs between the pentacoordinated Mg<sup>2</sup> and the Cl<sub>ax</sub> bridged atom in axial position in the trigonal bipyramidal, as shown in Figure 5b. The presence in thermal samples of the monochlorine bridged, unsymmetrically solvated structures appear to be an essential ingredient in the ligand exchange at each Mg.

The activation of the Mg-Cl bond in axial position is also consistent with the NBO analysis of  $\mathbf{D}^{\text{ClClI}}_{12}$ , which shows that the pentacoordinated Mg atoms in both  $\mathbf{D}^{\text{ClClI}}_{12}$  and  $\mathbf{D}^{\text{ClClI}}_{22}$  receive the smallest electron donation from Cl ( $\chi^{\text{Mg}} \sim 6\%$ ). In particular, the bridging chloride, axially coordinated to  $\text{Mg}^2$ ,  $\text{Cl}_{\text{ax}}$ , has the lowest contribution to this magnesium ( $\chi^{\text{Mg}} \sim 2\%$ ). Such a small donation implies an overpolarization of the  $\text{Cl}_{\text{ax}}\text{-Mg}^2$  bond, which in turn facilitates the opening of the di-chloride-bridged moiety of  $\mathbf{D}^{\text{ClClI}}_{12}$  as observed during AIMD. The asymmetry in the donation between the axial and the equatorial Cl explains also the selective opening of the ring at the  $\text{Cl}_{\text{ax}}\text{-Mg}^2$  bond.

The trigonal-bipyramidal geometry on  $\text{Mg}^2$  in  $\mathbf{D}^{\text{ClClI}}_{12}$  was not found in  $\mathbf{D}^{\text{ClClI}}_{22}$ . Instead, the coordination geometry of Mg atoms took a distorted square pyramidal geometry, with an apical THF, and equatorial Cl atoms, methyl group and remaining THF. In all  $\mathbf{D}^{\text{ClClI}}_{11}$ ,  $\mathbf{D}^{\text{ClClI}}_{12}$  and  $\mathbf{D}^{\text{ClClI}}_{22}$  the equilibrium values for the Cl-Mg-Cl bond angle are close to  $90^\circ$  (Table 1), even though the coordination geometries of the Mg atoms in the structures are different from each other. As an additional structural feature, it is observed that, during AIMD simulations, the four atoms of the  $\text{Mg}(\mu\text{-Cl})_2\text{Mg}$  moiety do not lie in the same plane. The average angles between the two  $\text{MgCl}_2$  planes are reported in Table 1.

**Table 1.** Average angles ( $\alpha^{\text{dyn}}$ ,  $\beta_1^{\text{dyn}}$  and  $\beta_2^{\text{dyn}}$ , in degrees) with associated standard deviation (in parenthesis) obtained from cluster analysis of the metadynamics trajectory, and angles from DFT optimization with implicit solvent ( $\alpha^{\text{st}}$ ,  $\beta_1^{\text{st}}$  and  $\beta_2^{\text{st}}$ , in degrees).

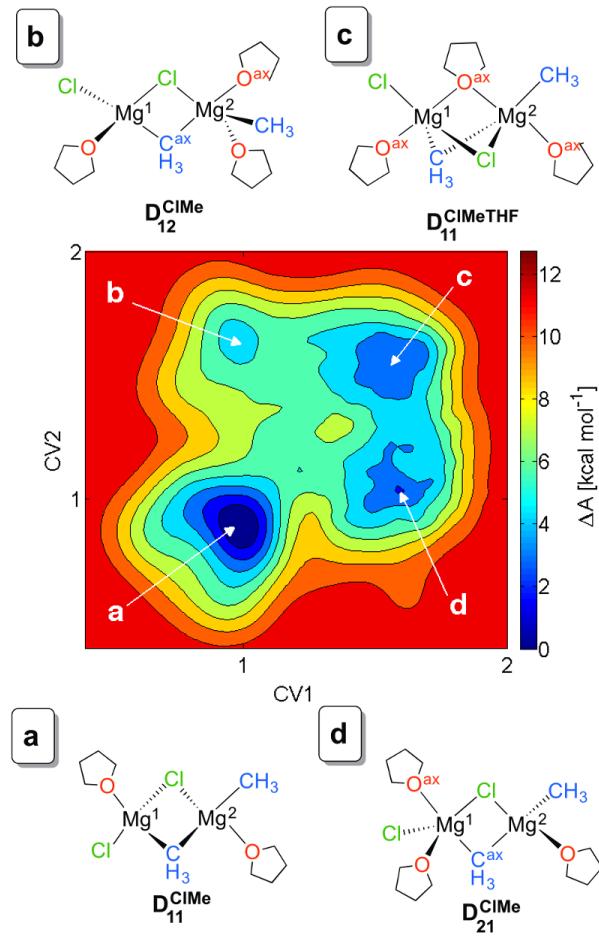


	$\alpha^{\text{dyn}}$	$\alpha^{\text{st}}$	$\beta_1^{\text{dyn}}$	$\beta_1^{\text{st}}$	$\beta_2^{\text{dyn}}$	$\beta_2^{\text{st}}$
$\mathbf{D}^{\text{ClClI}}_{11}$	167(8)	169.7	89(5)	91.3	89(5)	91.6
$\mathbf{D}^{\text{ClClI}}_{12}$	166(9)	165.4	91(7)	92.1	82(6)	83.9
$\mathbf{D}^{\text{ClClI}}_{22}$	165(11)	160.6	94(6)	84.1	94(6)	84.2

DFT optimization of  $\mathbf{D}^{\text{ClCl}}_{11}$ ,  $\mathbf{D}^{\text{ClCl}}_{12}$  and  $\mathbf{D}^{\text{ClCl}}_{22}$  confirmed that the minima on the FES correspond to minima on the PES. However, structural discrepancies are obtained for the species with the largest number of THF in the coordination sphere of the Mg centers. Thus, the Mg centers in  $\mathbf{D}^{\text{ClCl}}_{22}$  are trigonal-bipyramidal on the PES, while they are distorted square pyramidal on the FES. This indicates that the solvent cage has an increasing influence on the structure and dynamics of the more flexible, solvated species. AIMD simulations also revealed the existence of interchange of THF molecules in the Mg solvation shell of  $\mathbf{D}^{\text{ClCl}}_{22}$  taking place through an associative mechanism. In this case, a short-lived octahedral structure was formed. The statistical abundance of these structures is however marginal.

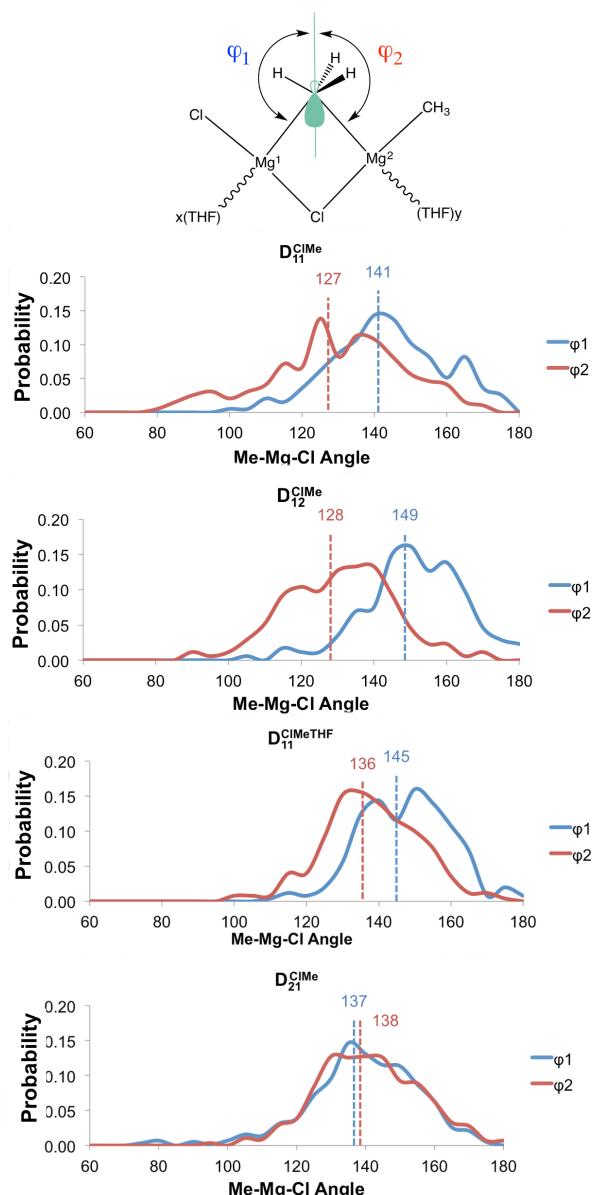
### Solvation of methyl chloride bridged structures

The accessible solvation states of the  $\mathbf{D}^{\text{ClMe}}$  species in **C** and **D** wells (Figure 3) were analyzed by metadynamics simulations using the coordination number of the oxygen atoms of THF to each of the Mg atoms as CVs (Figure 6). Three low-energy solvation structures were obtained. The lowest in energy,  $\mathbf{D}^{\text{ClMe}}_{11}$ , has two tetrahedral-coordinated Mg centers bridged by a Cl-atom and a methyl group. This structure is favored over the other ones by about 3 kcal mol<sup>-1</sup>. The two other major structures are dinuclear species with three coordinated solvent molecules,  $\mathbf{D}^{\text{ClMe}}_{21}$  and  $\mathbf{D}^{\text{ClMeTHF}}_{11}$ . In the former species, the additional THF is bound to the Mg-Cl moiety, while in the latter THF is bridging the two Mg centers (Figure 6). The free energies of triply-bridged species  $\mathbf{D}^{\text{ClMeTHF}}_{11}$  and the asymmetrically solvated  $\mathbf{D}^{\text{ClMe}}_{21}$  differ by only 0.5 kcal mol<sup>-1</sup> separated by a barrier of the order of 2 kT. A well corresponding to the asymmetric structure with two THF molecules coordinated to the Mg-CH<sub>3</sub> moiety ( $\mathbf{D}^{\text{ClMe}}_{12}$ ) was also obtained about 5 kcal mol<sup>-1</sup> above  $\mathbf{D}^{\text{ClMe}}_{11}$ . Overall, fast interchange between  $\mathbf{D}^{\text{ClMe}}_{21}$ ,  $\mathbf{D}^{\text{ClMeTHF}}_{11}$  and  $\mathbf{D}^{\text{ClMe}}_{12}$  is observed with low activation energy barriers of around 4 kcal mol<sup>-1</sup>.



**Figure 6.** FES for the methyl-bridged dimer  $\mathbf{D}^{\text{ClMe}}$  equilibria using the THF-coordination number to  $\text{Mg}^1$  (CV1) and the THF-coordination number to  $\text{Mg}^2$  (CV2) as variables, together with the most representative species obtained for wells **a**, **b**, **c** and **d**.

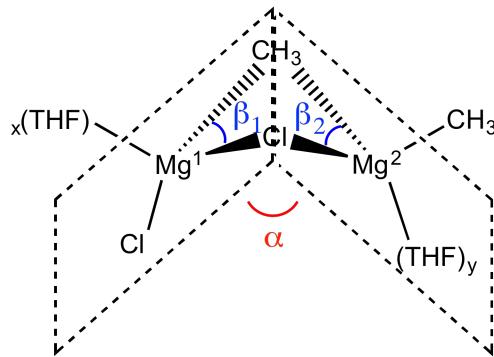
All  $\mathbf{D}^{\text{ClMe}}$  structures show tetrahedral and trigonal-bipyramidal geometries for the Mg-centers solvated by one and two THF, respectively. The bridging methyl ( $\mu\text{-CH}_3$ ) interacts differently with the two Mg atoms in the  $\text{ClMg}(\mu\text{-CH}_3)(\mu\text{-Cl})\text{MgCH}_3$  dimer. This is due the asymmetric distribution of the ligands at the two Mg atoms since  $\text{Mg}^1$  has a terminal chloride while  $\text{Mg}^2$  has a terminal methyl group. Moreover, the two Mg can have different solvation states at room temperature, as shown by the FES in Figure 6. In order to quantify the way the bridging  $\text{CH}_3$  group interacts with the two Mg centers in the different structures, the direction of the bridging  $p_z$  orbital axis relative to the two C-Mg directions was monitored (Figure 7).



**Figure 7.** Orientation of the methyl group in  $D^{ClMe}$  as a function of the solvation state, represented by  $\varphi_1$  and  $\varphi_2$ . A larger  $\varphi$  angle is indicative of a stronger  $Mg-CH_3$  interaction.

In both  $D^{ClMe}_{11}$  and  $D^{ClMe}_{12}$ , the  $p_z$  orbital is oriented toward  $Mg^1$ -Cl moiety, with average angles  $\varphi_1$  of approximately  $141^\circ$  and  $149^\circ$ , respectively. In these two structures  $\varphi_2$  has an average value of  $127^\circ$  and  $128^\circ$  indicating a poorer interaction of the bridging methyl with  $Mg^2$   $p_z$  orbital. In  $D^{ClMeTHF}_{11}$  the  $\mu$ - $CH_3$  bridging group is more equally shared between the two Mg centers as indicated by the values of  $\varphi_1$  and  $\varphi_2$  that oscillate around  $145^\circ$  and  $136^\circ$ , respectively. A larger solvation for the Mg bound to the terminal Cl than the other Mg atom ( $D^{ClMe}_{21}$ ) results in a practically equal sharing of the bridging methyl between the two Mg centers. In this case, both angles  $\varphi_1$  and  $\varphi_2$  oscillate around an average value of  $\sim 137^\circ$ .

**Table 2:** Average angles ( $\alpha^{\text{dyn}}$ ,  $\beta_1^{\text{dyn}}$  and  $\beta_2^{\text{dyn}}$ , in degrees) with its standard deviation (in parenthesis) obtained from cluster analysis of the metadynamics trajectory, and optimized angles from DFT calculations with implicit solvent ( $\alpha^{\text{st}}$ ,  $\beta_1^{\text{st}}$  and  $\beta_2^{\text{st}}$ , in degrees).



	$\alpha^{\text{dyn}}$	$\alpha^{\text{st}}$	$\beta_1^{\text{dyn}}$	$\beta_1^{\text{st}}$	$\beta_2^{\text{dyn}}$	$\beta_2^{\text{st}}$
<b>D<sup>ClMe</sup><sub>11</sub></b>	167(11)	175.6	101(10)	103.7	94(7)	100.0
<b>D<sup>ClMe</sup><sub>12</sub></b>	166(8)	179.4	102(7)	110.6	89(7)	93.2
<b>D<sup>ClMeTHF</sup><sub>11</sub></b>	165(6)	154.4	103(7)	102.0	96(7)	99.6
<b>D<sup>ClMe</sup><sub>21</sub></b>	161(7)	172.3	94(7)	100.4	101(7)	102.5

The orientation of the bridging methyl group is thus highly sensitive to the chemical groups and the solvation at each Mg center. For equal solvation, the methyl interacts more with the more electron deficient Mg<sup>1</sup> center, *i.e.*, the one with the terminal chloride. Increasing solvation of Mg<sup>1</sup> by either bridging or terminal THF results in an increase interaction of the bridging methyl group with Mg<sup>2</sup> bearing the terminal methyl group. In this way, the solvent helps the bridging methyl group to weaken its interaction with Mg<sup>1</sup> and increase that to Mg<sup>2</sup>, assisting the transformation of CH<sub>3</sub>MgCl into MgCl<sub>2</sub> and Mg(CH<sub>3</sub>)<sub>2</sub>.

Geometry optimizations on the potential energy surface with DFT methods of the minima **D<sup>ClMe</sup><sub>11</sub>**, **D<sup>ClMe</sup><sub>12</sub>**, **D<sup>ClMeTHF</sup><sub>11</sub>** and **D<sup>ClMe</sup><sub>21</sub>** on the FES gave minima with similar structures as indicated in Table 2. In all cases, trigonal-bipyramidal geometries were found on the penta-coordinated Mg atoms with the bridging CH<sub>3</sub> and a terminal THF in the axial positions. NBO analysis on the optimized structures showed a slightly higher electron donation of the bridged CH<sub>3</sub> lone pair to the Mg of the Mg-Cl moiety ( $\chi^{\text{Mg}1} = 5.7$ , 5.2 and 4.6 %) compared to the one of the Mg-CH<sub>3</sub> moiety ( $\chi^{\text{Mg}2} = 3.7$ , 3.8 and 3.6 %) for **D<sup>ClMe</sup><sub>11</sub>**, **D<sup>ClMe</sup><sub>12</sub>** and **D<sup>ClMeTHF</sup><sub>11</sub>**, respectively. With increasing solvation of Mg<sup>1</sup>, as in **D<sup>ClMe</sup><sub>21</sub>** the donation of the bridging methyl group to the two Mg centers becomes equivalent ( $\chi^{\text{Mg}1} = 4.3$  %,  $\chi^{\text{Mg}2} = 4.2$  %), consistent with the distribution of the  $\varphi_1$  and  $\varphi_2$  angles during AIMD simulations (Figure 8).

Although small, the increased electron donation of the Me group to Mg<sup>2+</sup>, which already bears the other methyl group, is in agreement with  $\mathbf{D}^{\text{ClMe}}_{21}$  being the most prone of the methyl-chloride bridged dimers to yield the final products MgCl<sub>2</sub> and MgCH<sub>3</sub>)<sub>2</sub>.

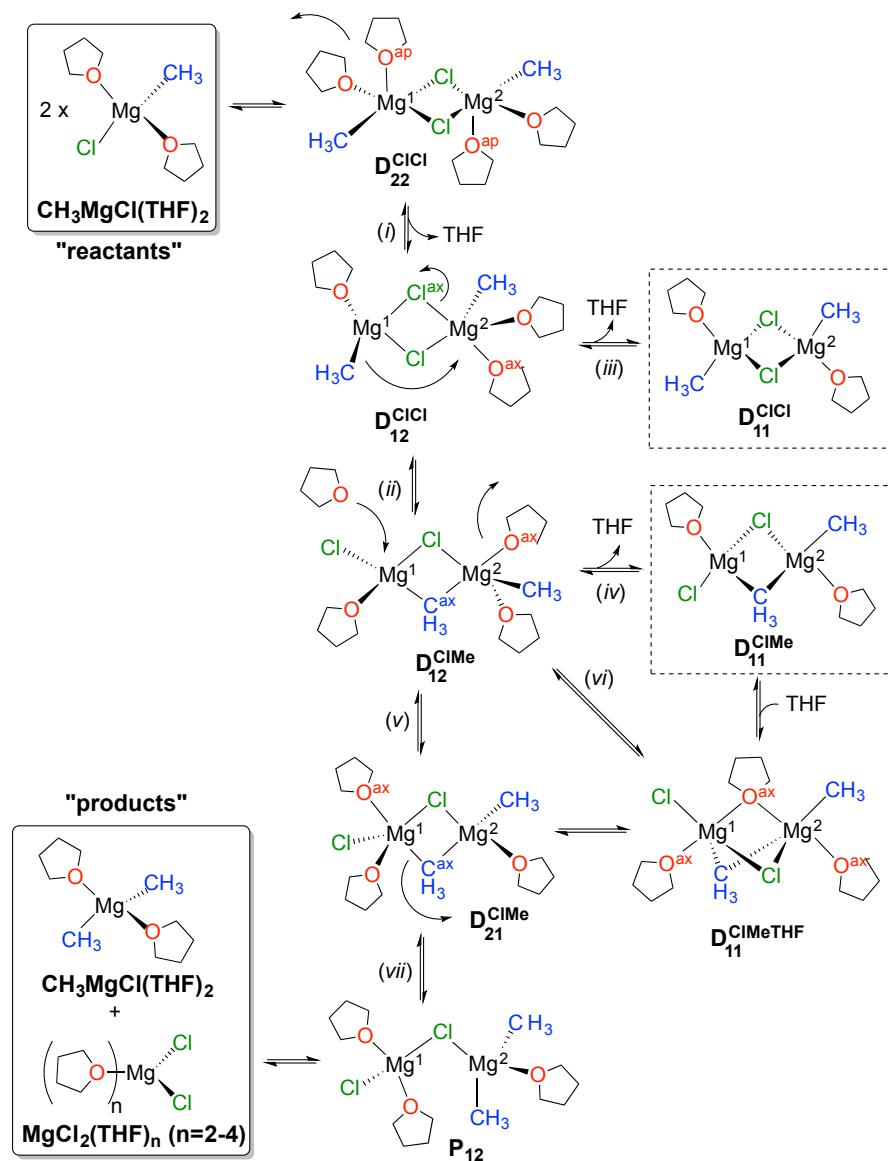
### The reaction pathway of the Schlenk equilibrium

The computational results presented above allow the reconstruction of the full reaction pathway of the Schlenk equilibrium in a THF solution. Figure 8 presents the set of structures, localized as minima on the FES, that are involved in the transformation of the reactants CH<sub>3</sub>MgCl(THF)<sub>2</sub> to the products Mg(CH<sub>3</sub>)<sub>2</sub>(THF)<sub>2</sub> and MgCl<sub>2</sub>(THF)<sub>n</sub> (n = 2 - 4). The scheme also highlights the key role played by the solvent in assisting the Cl/CH<sub>3</sub> ligand exchange.

The Schlenk equilibrium starts by a dimerization of (CH<sub>3</sub>)MgCl(THF)<sub>2</sub> via the two chloride atoms to form the dichloride bridged species  $\mathbf{D}^{\text{ClCl}}_{22}$ . Static calculations show that this reaction is slightly endoergic by 4.9 kcal mol<sup>-1</sup>, consistent with experimental results.<sup>22, 76</sup>

The coordination geometry of the pentacoordinated Mg atoms in  $\mathbf{D}^{\text{ClCl}}_{22}$  was found to be distorted square-pyramidal when exploring the FES but trigonal-bipyramidal when performing ab initio geometry optimizations on the PES. Such discrepancy evidences a strong influence of the surrounded solvent molecules on the structural geometry of  $\mathbf{D}^{\text{ClCl}}_{22}$ . THF exchange on the Mg atom may occur through addition of one THF to form an octahedral species.<sup>77-79</sup> These last transient structures are very short-lived, and thus have not been included in Figure 8.

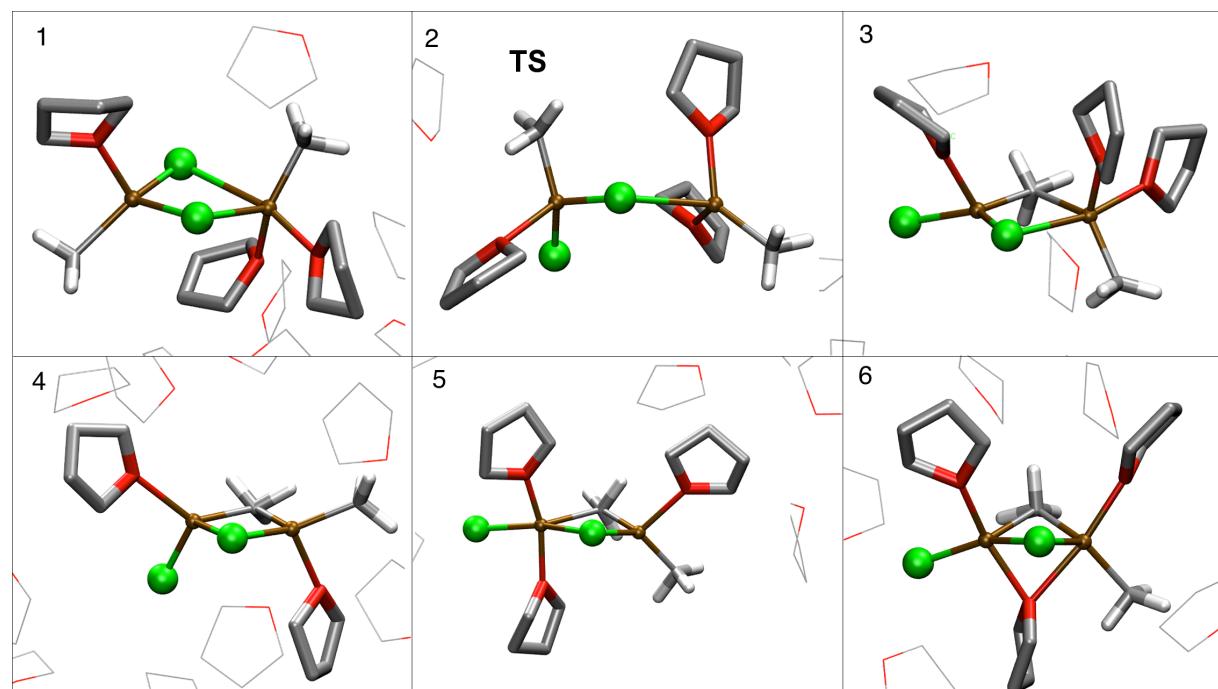
Desolvation of  $\mathbf{D}^{\text{ClCl}}_{22}$  by one THF molecule yields an asymmetrical dinuclear complex ( $\mathbf{D}^{\text{ClCl}}_{12}$ ) with tetrahedral and pyramidal-trigonal geometries for Mg<sup>1</sup> and Mg<sup>2</sup>, respectively. This tri-solvated species can lose one THF molecule to produce the di-solvated ( $\mathbf{D}^{\text{ClCl}}_{11}$ ), which was found to be the most stable species in solution. The tetrahedral coordination of Mg is also the one preferred in the solid state as shown by reported crystallographic structures of related Mg<sub>2</sub>X<sub>2</sub>R<sub>2</sub> species.<sup>28, 80</sup> The difference in energy between  $\mathbf{D}^{\text{ClCl}}_{11}$  and the least stable intermediate  $\mathbf{D}^{\text{ClCl}}_{12}$  is 2.7 kcal/mol with an activation energy barrier of less than 5 kcal mol<sup>-1</sup> (Figure 4). Therefore, all  $\mathbf{D}^{\text{ClCl}}_{11}$ ,  $\mathbf{D}^{\text{ClCl}}_{12}$  and  $\mathbf{D}^{\text{ClCl}}_{22}$  species are expected to coexist and undergo interconversion at room temperature.



**Figure 8.** Intermediates involved in the Schlenk equilibrium according to dynamic simulations. Arrows indicate the chemical transformations along the main pathway leading from monomeric reactants to products (inside solid squares). The most stable dichloride and methyl-chloride bridged dinuclear species are inside dashed squares.

Interestingly, not all species favor the Mg-Cl bond cleavage required for the Cl/Me ligand exchange. Analysis of the Mg-Cl bond distances in  $\text{D}^{\text{ClCl}}_{11}$ ,  $\text{D}^{\text{ClCl}}_{22}$  and  $\text{D}^{\text{ClCl}}_{12}$  (Figure 5a) shows that the formation of mono-chloride-bridged species takes place preferentially from  $\text{D}^{\text{ClCl}}_{12}$ . NBO analysis indicates that the asymmetric solvation of  $\text{D}^{\text{ClCl}}_{12}$  favors the  $\text{Mg}^2\text{-Cl}^{\text{ax}}$  bond cleavage (step (ii) in Figure 8). In this complex the higher solvation on  $\text{Mg}^2$  weakens the  $\text{Mg}^2\text{-Cl}$  bond, while the lower solvation on  $\text{Mg}^1$  makes it prone to accept an extra anionic ligand during the Cl/CH<sub>3</sub> exchange process. Analysis on several trajectories, as the one

represented in Figure 9, reaffirms that the chloride involved in the cleavage is the one located in axial position of  $Mg^2$  (Figure 5b). The mono-bridged species generated by the Mg-Cl bond cleavage (snapshot 2 in Figure 9) has a short-lived reaction time and rapidly evolves into methyl-chloride bridged structures ( $D^{ClMe}$ ). The methyl group occupying the bridging position comes from  $Mg^1$ , as shown in snapshot 3, this is consistent with the higher electron density on  $MgCl(THF)(\mu-Cl)$  fragment compared to  $Mg(THF)_2(\mu-Cl)$  calculated for the structure shown in snapshot 2.



**Figure 9.** Snapshots for the methyl transfer reaction in  $D^{ClCl}_{12}$  ( $Mg^1$  on the left hand side and  $Mg^2$  on the right for all snapshots): 1) initial  $D^{ClCl}_{12}$  structure, 2) Transition State of the transmetalation reaction, 3) formation of  $D^{ClMe}_{12}$ , 4) solvent loss to form  $D^{ClMe}_{11}$ , 5) solvent addition to form  $D^{ClMe}_{21}$ , and 6)  $D^{ClMeTHF}_{11}$ . The atoms for the Grignard reagent and the coordinating THF molecules are depicted in balls and/or sticks and colored according to standard color codes. Selected neighboring solvent molecules are drawn in thin lines.

The first  $D^{ClMe}$  species generated after the Mg-Cl bond cleavage is the  $D^{ClMe}_{12}$  complex (snapshot 3 in Figure 9). This species is however the least stable of the four solvation states observed by AIMD, with a free energy difference of +5 kcal mol<sup>-1</sup> (Figure 4) over the most stable solvation structure. As in the case of  $D^{ClCl}$  species, the most stable  $D^{ClMe}$  complex is the di-solvated  $D^{ClMe}_{11}$ , in which both Mg atoms have tetrahedral coordination geometry.

However,  $\mathbf{D}^{\text{ClMe}}$  incorporates other tri-solvated species,  $\mathbf{D}^{\text{ClMeTHF}_{11}}$  and  $\mathbf{D}^{\text{ClMe}_{21}}$ , which are both 3 kcal mol<sup>-1</sup> above  $\mathbf{D}^{\text{ClMe}_{11}}$ . Exchange between these species has low activation energy barriers (smaller than 4 kcal mol<sup>-1</sup>) and involves either the addition of an external THF at the bridging position ( $\mathbf{D}^{\text{ClMe}_{11}} \rightarrow \mathbf{D}^{\text{ClMeTHF}_{11}}$ ) or at the terminal position of Mg<sup>1</sup> ( $\mathbf{D}^{\text{ClMe}_{11}} \rightarrow \mathbf{D}^{\text{ClMe}_{21}}$ ). Internal rearrangement of THF from bridging to terminal is also observed ( $\mathbf{D}^{\text{ClMeTHF}_{11}} \rightarrow \mathbf{D}^{\text{ClMe}_{21}}$ ) with even lower energy barrier (<0.5 kcal mol<sup>-1</sup>). The higher stability of  $\mathbf{D}^{\text{ClMe}_{21}}$  compared to  $\mathbf{D}^{\text{ClMe}_{12}}$  is consistent with the higher solvation of the most electrophilic Mg atom (Mg<sup>1</sup>), *i.e.* the one with terminal chloride.

In order to form the final products, both Mg<sup>2</sup>-Cl and Mg<sup>1</sup>-CH<sub>3</sub> bonds need to be cleaved. This however does not take place simultaneously but consecutively, and solvation plays an important role. Since Cl is a much better ligand than CH<sub>3</sub>,<sup>28</sup> the Mg-CH<sub>3</sub> bond is the first to be broken. Analysis on the orientation of the formal lone pair in the bridging methyl ligand shows that at equal solvation of Mg atoms in  $\mathbf{D}^{\text{ClMe}}$ ,  $\mu\text{-CH}_3$  is more strongly bonded to Mg<sup>1</sup> than to Mg<sup>2</sup> because Mg<sup>1</sup> is more electrophilic (Figure 7). This preference is however modified by increasing the solvation of Mg<sup>1</sup> relative to Mg<sup>2</sup>. Thus, in  $\mathbf{D}^{\text{ClMe}_{21}}$  the methyl is equally bonded to Mg<sup>1</sup> and Mg<sup>2</sup>. In addition, in this species, the axial position occupied by the bridged methyl group in the coordination sphere of Mg<sup>1</sup> also favors the Mg<sup>1</sup>-C<sup>ax</sup>H<sub>3</sub> bond cleavage. Increase in solvation of Mg<sup>1</sup> thus yields the pre-product of the Schlenk equilibrium (**P**<sub>12</sub> and Minimum **E** in Figure 3), consisting of a Mg<sup>1</sup>Cl<sub>2</sub> species bridged by way of a single chlorine group to Mg<sup>2</sup>(CH<sub>3</sub>)<sub>2</sub>. Increase in solvation of both Mg centers is expected to favor the final release of the mononuclear MgCl<sub>2</sub> and Mg(CH<sub>3</sub>)<sub>2</sub> products. The dissociation energy from **P**<sub>12</sub> to CH<sub>3</sub>MgCl(THF)<sub>2</sub> and the most stable MgCl<sub>2</sub>(THF)<sub>3</sub> has been estimated to be -8.7 kcal mol<sup>-1</sup> in this study. (see Figure S4 for further details).

Despite their relative low abundance in solution, as also experimentally observed,<sup>22</sup> formation of dimeric adducts is key to the evolution of the Schlenk equilibrium reaction. The reaction pathway identified in this study highlights the crucial role of the solvent in assisting the Cl/Me exchange in the Schlenk equilibrium. Most importantly, it shows that the tetra-coordinated Mg species proposed in most computational studies<sup>28, 81-82</sup> are indeed the most stable structures but are not on the reactive pathway. Instead, asymmetric solvation on Mg atoms is needed to promote the Mg-Cl and Mg-CH<sub>3</sub> bond cleavage to go from MgCH<sub>3</sub>Cl to Mg(CH<sub>3</sub>)<sub>2</sub> and MgCl<sub>2</sub>. These tetra/pentacoordinated Mg dimers ( $\mathbf{D}^{\text{ClCl}_{12}}$  and  $\mathbf{D}^{\text{ClMe}_{21}}$ ) are transient intermediates that interchange with the most stable solvation structures at room temperature. In addition, the bridging ligand involved in the bond breaking process is located always in the axial position of the penta-coordinated Mg atom. This detailed information on

the Schlenk equilibrium mechanism will be useful for a better understanding of the reactivity of the Grignard reagents. In addition, this study may help to better understand other transmetallation processes, such as those involved in the Kumada and Negishi cross-coupling reactions,<sup>4, 83-84 85-86</sup> which are also assisted by solvent.

## Conclusion

The ab initio molecular dynamics study of the Schlenk equilibrium reaction, in which two molecules of  $\text{CH}_3\text{MgCl}$  exchange methyl and chloride groups to yield  $\text{Mg}(\text{CH}_3)_2$  and  $\text{MgCl}_2$ , showed how the ether solvent (THF here) has a crucial role in assisting the reaction. Although coordination of the solvent is needed for stabilizing the various mono and dinuclear Mg complexes, the stabilizing effect is not the only factor at work.

The reaction goes via the formation of a dinuclear  $(\text{CH}_3)\text{Mg}(\mu\text{-Cl})_2\text{Mg}(\text{CH}_3)$  intermediate whose most stable solvated form (with a single THF at each Mg) is unreactive. The chloride/methyl exchange is promoted by making the two Mg electronically different. The calculations show that these differences are created by different solvations of the two Mg centers. The cleavage of the Mg-Cl bond and associated shift of the methyl group from terminal to bridging position is assisted by increasing the solvation at the Mg involved into the bond cleavage, while keeping the other Mg less solvated.

A similar process occurs at the dinuclear  $(\text{CH}_3)\text{Mg}(\mu\text{-CH}_3)(\mu\text{-Cl})\text{MgCl}$ : the cleavage of the bond between Mg and the bridging  $\text{CH}_3$  group to form solvated  $\text{Mg}(\text{CH}_3)_2$  and  $\text{MgCl}_2$  requires the chloride rich Mg atom to be more solvated than the methyl rich Mg one. Increasing solvation at one Mg favors bond cleavage while decreasing solvation favors the formation of new terminal bonds.

The species selected by the dynamics to be on the reaction pathway are not necessarily minima on the potential energy surface, which emphasize the need for the former method. Our findings highlight the need of including explicit solvent dynamics in the modelisation of such reactions, since this is crucial in allowing the Cl/ $\text{CH}_3$  groups exchange to occur with a low energy barrier.

## Associated Content

## **Supporting Information**

The Supporting Information is available free of charge on the ACS Publication web site at

Metadynamics parameters, Natural Bond Orbital Analysis, DFT optimized geometries, Solvation properties for different sizes of the simulation box.

## **Notes**

The authors declare no competing financial interest.

## **Acknowledgement**

This work was supported by the Research Council of Norway (RCN) through the CoE Centre for Theoretical and Computational Chemistry (CTCC) Grant No. 179568/V30 and 171185/V30, and by the Norwegian Supercomputing Program (NOTUR) (Grant No. NN4654K). A.N. thanks the RCN for the grants 221801/F20 and 250044/F20. Authors thank Elisa Rebolini for enlightening discussion.

## **References**

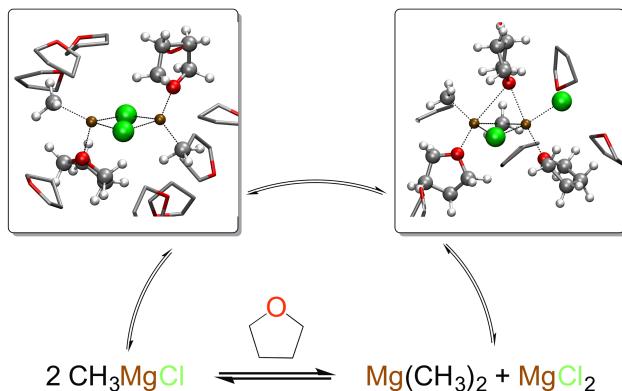
1. Grignard, V. *C. r. d. l'Acad. des Sciences* **1900**, *130*, 1322-1324.
2. Grignard, V. *Chemistry* **1921**, *1966*, 234-246.
3. Corriu, R. J. P.; Massé, J. P. *J. Chem. Soc., Chem. Commun.* **1972**, 144a-144a.
4. Tamao, K.; Sumitani, K.; Kumada, M. *J. Am. Chem. Soc.* **1972**, *94*, 4374-4376.
5. Fürstner, A.; Leitner, A.; Méndez, M.; Krause, H. *J. Am. Chem. Soc.* **2002**, *124*, 13856-13863.
6. Frisch, A. C.; Beller, M. *Angew. Chem. Int. Ed.* **2005**, *44*, 674-688.
7. Terao, J.; Kato, Y.; Kambe, N. *Chem. Asian J.* **2008**, *3*, 1472-1478.
8. Vechorkin, O.; Barmaz, D.; Proust, V.; Hu, X. *J. Am. Chem. Soc.* **2009**, *131*, 12078-12079.
9. Adrio, J.; Carretero, J. C. *ChemCatChem* **2010**, *2*, 1384-1386.
10. Jana, R.; Pathak, T. P.; Sigman, M. S. *Chem. Rev.* **2011**, *111*, 1417-1492.
11. Cong, X.; Tang, H.; Zeng, X. *J. Am. Chem. Soc.* **2015**, *137*, 14367-14372.
12. Neufeld, R.; Teuteberg, T. L.; Herbst-Irmer, R.; Mata, R. A.; Stalke, D. *J. Am. Chem. Soc.* **2016**, *138*, 4796-4806.
13. Seydel, D. *Organometallics* **2009**, *28*, 1598-1605.
14. Guggenberger, L. J.; Rundle, R. E. *J. Am. Chem. Soc.* **1968**, *90*, 5375-5378.
15. Vallino, M. *J. Organomet. Chem.* **1969**, *20*, 1-10.
16. Toney, J.; Stucky, G. D. *J. Organomet. Chem.* **1971**, *28*, 5-20.
17. Blasberg, F.; Bolte, M.; Wagner, M.; Lerner, H.-W. *Organometallics* **2012**, *31*, 1001-1005.

18. Smith, M. B.; Becker, W. E. *Tetrahedron* **1967**, *23*, 4215-4227.
19. Smith, M. B.; Becker, W. E. *Tetrahedron Lett.* **1965**, *43*, 3843-3847.
20. Ashby, E. C.; Nackashi, J.; Parris, G. E. *J. Am. Chem. Soc.* **1975**, *97*, 3162-3171.
21. Schlenk, W.; Schlenk, W. *Ber. dtsch. Chem. Ges. A/B* **1929**, *62*, 920-924.
22. Walker, F. W.; Ashby, E. C. *J. Am. Chem. Soc.* **1969**, *91*, 3845-3850.
23. Sobota, P.; Duda, B. *J. Organomet. Chem.* **1987**, *332*, 239-245.
24. Sakamoto, S.; Imamoto, T.; Yamaguchi, K. *Org. Lett.* **2001**, *3*, 1793-1795.
25. Hohenberg, P.; Kohn, W. *Phys. Rev. B* **1964**, *136*, 864-871.
26. Kohn, W.; Sham, L. *J. Phys. Rev. A* **1965**, *140*, 1133-1138.
27. Jiménez-Halla, J. O. C.; Bickelhaupt, F. M.; Solà, M. *J. Organomet. Chem.* **2011**, *696*, 4104-4111.
28. Lioe, H.; White, J. M.; O'Hair, R. A. *J. J. Mol. Model.* **2011**, *17*, 1325-1334.
29. Henriques, A. M.; Barbosa, A. G. H. *J. Phys. Chem. A* **2011**, *115*, 12259-12270.
30. Ramirez, F.; Sarma, R.; Chaw, F.; McCaffrey, T. M. *J. Am. Chem. Soc.* **1977**, *99*, 5285-5289.
31. Pirinen, S.; Koshevoy, I. O.; Denifl, P.; Pakkanen, T. T. *Organometallics* **2013**, *32*, 4208-4213.
32. Ashby, E. C.; Becker, W. E. *J. Am. Chem. Soc.* **1963**, *85*, 118-119.
33. Tammiku-Taul, J.; Burk, P.; Tuulmets, A. *J. Phys. Chem. A* **2004**, *108*, 133-139.
34. Tobisu, M.; Chatani, N. *Acc. Chem. Res.* **2015**, *48*, 1717-1726.
35. Cahiez, G.; Moyeux, A.; Cossy, J. *Adv. Synth. Catal.* **2015**, *357*, 1983-1989.
36. Tasker, S. Z.; Standley, E. A.; Jamison, T. F. *Nature* **2014**, *509*, 299-309.
37. Laio, A.; Parrinello, M. *Proc. Nat. Acad. Sci. USA* **2002**, *99*, 12562-12566.
38. Iannuzzi, M.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2003**, *23*-26.
39. Vuilleumier, R.; Sprik, M. *J. Chem. Phys.* **2001**, *115*, 3454-3468.
40. Lightstone, F. C.; Schwegler, E.; Hood, R. Q.; Gygi, F.; Galli, G. *Chem. Phys. Lett.* **2001**, *343*, 549-555.
41. Bernasconi, L.; Baerends, E. J.; Sprik, M. *J. Phys. Chem. B* **2006**, *110*, 11444-11453.
42. Blumberger, J.; Bernasconi, L.; Tavernelli, I.; Vuilleumier, R.; Sprik, M. *J. Am. Chem. Soc.* **2004**, *126*, 3928-3938.
43. Guido, C. A.; Pietrucci, F.; Gallet, G. A.; Andreoni, W. *J. Chem. Theory. Comput.* **2013**, *9*, 28-32.
44. Boero, M.; Ikeshoji, T.; Liew, C. C.; Terakura, K.; Parrinello, M.; Boero, M.; Ikeshoji, T.; Liew, C. C.; Terakura, K. *J. Am. Chem. Soc.* **2004**, *137*, 6280-6286.
45. Vidossich, P.; Lledós, A.; Ujaque, G. In *Computational Studies in Organometallic Chemistry*, **2016**; *167*, pp 81-106 Macgregor, S. A.; Eisenstein, O., Eds. Structure and Bonding Springer International Publishing (*Berlin, Ger.*).
46. Vidossich, P.; Lledós, A.; Ujaque, G. *Acc. Chem. Res.* **2016**, *49*, 1271-1278.
47. Laio, A.; VandeVondele, J.; Rothlisberger, U. *J. Chem. Phys.* **2002**, *116*, 6941-6947.
48. Moret, M.-E.; Tavernelli, I.; Chergui, M.; Rothlisberger, U. *Chem. Eur. J.* **2010**, *16*, 5889-5894.
49. Dal Peraro, M.; Llarrull, L. I.; Rothlisberger, U.; Vila, A. J.; Carloni, P. *J. Am. Chem. Soc.* **2004**, *126*, 12661-12668.
50. Dal Peraro, M.; Vila, A. J.; Carloni, P.; Klein, M. L. *J. Am. Chem. Soc.* **2007**, *129*, 2808-2816.
51. Cascella, M.; Magistrato, A.; Tavernelli, I.; Carloni, P.; Rothlisberger, U. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 19641-19646.
52. De Vivo, M.; Dal Peraro, M.; Klein, M. L. *J. Am. Chem. Soc.* **2008**, *130*, 10955-10962.
53. Gossens, C.; Tavernelli, I.; Rothlisberger, U. *Chimia* **2005**, *59*, 81-84.

54. Metz, D. J.; Glines, A. *J. Phys. Chem.* **1967**, *71*, 1158-1158.
55. Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695-1697.
56. Martyna, G. J.; Klein, M. L.; Tuckerman, M. *J. Chem. Phys.* **1992**, *97*, 2635-2643.
57. Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511-519.
58. Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865-3868.
59. Lippert, G.; Hutter, J.; Parrinello, M. *Theor. Chem. Acc.* **1999**, *103*, 124-140.
60. VandeVondele, J.; Hutter, J. *J. Chem. Phys.* **2007**, *127*, 114105.
61. Goedecker, S.; Teter, M.; Hutter, J. *Phys. Rev. B* **1996**, *54*, 1703-1710.
62. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.
63. Hutter, J.; Iannuzzi, M.; Schiffmann, F.; VandeVondele, J. *WIREs* **2014**, *4*, 15-25.
64. Laio, A.; Gervasio, F. L. *Rep. Prog. Phys.* **2008**, *71*, 126601.
65. Laio, A.; Rodriguez-Fortea, A.; Gervasio, F. L.; Ceccarelli, M.; Parrinello, M. *The Journal of Physical Chemistry B* **2005**, *109*, 6714-6721.
66. Vandevondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. r. *Comput. Phys. Commun.* **2005**, *167*, 103-128.
67. Krack, M.; Parrinello, M. In *QUICKSTEP: Make the Atoms Dance*, Forschungszentrum Jülich, NIC Series, 2004; p 29.
68. Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33-38.
69. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Gaussian, Inc.: Wallingford, CT, USA, 2009.
70. Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257-2261.
71. Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. R. *J. Comput. Chem.* **1983**, *4*, 294-301.
72. Frisch, M. J.; Pople, J. A.; Binkley, J. S. *J. Chem. Phys.* **1984**, *80*, 3265.
73. Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, *113*, 6378-6396.
74. (a) Glendening, E. D.; Landis, C. R.; Weinhold, F. *J. Comput. Chem.* **2013**, *34*, 1429-1437. (b) **NBO 6.0**. E. D. Glendening, J. K. Badenhoop, A. E. Reed, J. E. Carpenter, J. A. Bohmann, C. M. Morales, C. R. Landis, and F. Weinhold, Theoretical Chemistry Institute, University of Wisconsin, Madison (2013)
75. Weinhold, F.; Landis, C. R. *Discovering Chemistry with Natural Bond Orbitals* Wiley **2012**.
76. Silverman, G. S.; Rakita, P. E. *Handbook of Grignard reagents* CRC Press New York, **1996**.
77. Vestergren, M.; Eriksson, J.; Håkansson, M. *Chem. Eur. J.* **2003**, *9*, 4678-4686.
78. Vestergren, M.; Eriksson, J.; Håkansson, M. *J. Organomet. Chem.* **2003**, *681*, 215-224.
79. Vestergren, M.; Gustafsson, B.; Davidsson, Ö.; Håkansson, M. *Angew. Chem. Int. Ed.* **2000**, *39*, 3435-3437.
80. Harrison-Marchand, A.; Mongin, F. *Chem. Rev.* **2013**, *113*, 7470-7562.

81. Yamazaki, S.; Yamabe, S. *J. Org. Chem.* **2002**, *67*, 9346-9353.
82. Mori, T.; Kato, S. *J. Phys. Chem. A* **2009**, *113*, 6158-6165.
83. Hölzer, B.; Hoffmann, R. W. *Chem. Commun.* **2003**, *2*, 732-733.
84. King, A. O.; Okukado, N.; Negishi, E.-i. *J. Chem. Soc., Chem. Commun.* **1977**, 683-684.
85. Negishi, E.-i *Acc. Chem. Res.* **1982**, *15*, 340-348.
86. García-Melchor, M.; Fuentes, B.; Lledós, A.; Casares, J. A.; Ujaque, G.; Espinet, P. *J. Am. Chem. Soc.* **2011**, *133*, 13519-13526.

## Table of Contents



## Paper II

# **Direct Nucleophilic Attack Vs. Radical Propagation Mechanism in the Grignard Reaction – a Computational Study**

Raphael Mathias Peltzer,<sup>†</sup> Odile Eisenstein,<sup>†,‡</sup> Jürgen Gauss,<sup>¶</sup> and Michele Cascella<sup>\*,†</sup>

<sup>†</sup>*Department of Chemistry, and Hylleraas Centre for Quantum Molecular Sciences, University of Oslo, PO Box 1033 Blindern, 0315 Oslo, Norway*

<sup>‡</sup>*Institut Charles Gerhardt, UMR 5253 CNRS-Universit de Montpellier, Universit de Montpellier, cc 1501, Place E. Bataillon, 34095 Montpellier, France*

<sup>¶</sup>*Institut für Physicalische Chemie, Johannes Gutenberg-Universität Mainz, Staudingerweg 9, 55128 Mainz, Germany*

E-mail: michele.cascella@kjemi.uio.no

## **Abstract**

We investigated the competition between the nucleophilic addition and the radical pathway for the Grignard reaction by computational modelling. Radical formation energies for complexes with different substrates were estimated by density functional theory after benchmarking over coupled cluster data. The nucleophilic attack mechanism was investigated by ab initio molecular dynamics in explicit tetrahydrofuran solvent. The initial step of the radical reaction involves binding of the substrate to the magnesium and the subsequent release of a radical organic group. Regardless of the nature of the substrate and the solvation of the magnesium, the unpaired electron localizes on the substrate. We found that the nucleophilic addition is favored

for alkyl-carbonyl species, while the radical mechanism becomes energetically favorable for extended aromatic moieties. We further determined that the nucleophilic addition is promoted by the fraction of the dialkyl-magnesium species produced by the Schlenk equilibrium. The identification of competitive mechanism that are selected by the nature of the substrate can help the design of future metalloorganic reactions and catalysts in the homogeneous phase.

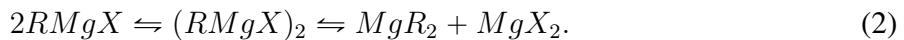
## Introduction

The Grignard reaction is a fundamental, text-book, process used in organic synthesis to promote the formation of carbon-carbon bonds.<sup>1</sup> The reaction involves the initial synthesis of the Grignard reagent, a magnesium-organic compound of general formula RMgX where R is a organic residue and X is a halogenous element (typically, Cl, or Br), and the subsequent reaction with an electrophilic species (most prominently, carbonyl groups: R'C=OR") to form adducts of general formula RR'R"CO-MgX:



The reaction occurs strictly in polar-aprotic solvents. Addition of water is used to hydrolyze the product, yielding the corresponding alcohol.<sup>2</sup>

Despite the fact that the Grignard reaction has been known since more than one hundred years, its characterization at a molecular level has remained elusive so far. Difficulties arise from the fact that the same molecular structure of the Grignard reagent is complex, with several organometallic species co-existing in solution according to the *generalized Schlenk equilibrium*:<sup>3</sup>



As a matter of facts, the experimental characterization of such species, as well as the identification of the most reactive species for the Grignard equilibrium has resulted so far problematic.<sup>4</sup>

The relative abundance of the various chemical species can be modulated by varying the concentration, the organic residue or the halide, and the temperature. Several computational and experimental data available in the literature indicate that the equilibrium in equation 2 involves different intermediate states for the dimer dependent on the choice of the organic compound, the halide, and the solvent.<sup>5</sup>

The general understanding of the Grignard reaction (equation 1) points at a nucleophilic addition to an electrophilic substrate by a metal-activated nucleophilic carbon.<sup>6</sup> In fact, experiment with substrates characterized by aryl-conjugated carbonyl moieties yielded products that are not fully consistent with a nucleophilic addition, while they could be interpreted as the outcome of a radical-chain propagation mechanism. This was evidenced by the formation of pinacol structures in the reaction of benzophenone or quinones with Grignard reagent.<sup>7–10</sup> A series of studies by Ashby and co-workers further evidenced radical intermediates and further pointed at a possible competition between the two mechanisms, influenced by several factors among which the nature of the solvent, and the chemical characteristics of both the substrate and the residue. This competition between radical and non-radical pathway for different organomagnesium cross-coupling reactions could be demonstrated experimentally using chiral Grignard reagents.<sup>11</sup> However, it has not been possible to fully characterize the competing pathways on a molecular level.

Computer modeling is an excellent strategy to complement experimental data, providing clear insights about the molecular details of chemical processes in very highly controlled conditions. The constitution and reactions of Grignard reagents have been investigated in the past years by different approaches including by second-order Møller-Plesset perturbation theory (MP2)<sup>12,13</sup> and density functional theory (DFT).<sup>5,14–16</sup>

Computer modeling has also been successfully applied<sup>17</sup> to the study of the formation of the Grignard reagent, for which a radical pathway has been established.<sup>18,19</sup> Furthermore, computational studies have been used to investigate the character of the Grignard reaction on metallic surfaces using cluster models<sup>20</sup> as well as using implicit solvation by dimethylether solvent<sup>16</sup> using DFT calculations.

Recently, we have characterized in detail the Schlenk equilibrium for a model  $\text{CH}_3\text{MgCl}$  Grignard reagent dissolved in tetrahydrofuran (THF). We determined that multiple chemical species, including,  $\text{CH}_3\text{MgCl}$ ,  $\text{Mg}(\text{CH}_3)_2$ ,  $\text{MgCl}_2$ , and higher-order aggregates, co-exists at room temperature. Moreover, we observed that, in the different compounds, the magnesium centers can accept a variable number of THF in their shell of first ligands, and that local changes in solvation are directly connected to the interchange between the chemical species at equilibrium. The accurate determination of the chemical species present in a solution of Grignard reagent provides an excellent starting point to investigate the Grignard reaction, and shed light on the conditions at which the two nucleophilic or radical mechanisms may be preferred or co-exist.

Here, we employ quantum mechanical calculations and ab initio molecular dynamics to investigate the likelihood of different reactive paths for the Grignard reaction in the presence of different substrates. We find that for highly conjugated systems such as fluorenone, that have the ability to stabilize a radical, the radical pathway is preferred. For non-conjugated systems, such as acetaldehyde, a non-radical pathway via nucleophilic addition is preferred. We furthermore find that the nucleophilic addition reaction shows a clear trend between the different species of the Schlenk-equilibrium. We find that the most reactive species in THF solvent are of the kind  $\text{Mg}(\text{CH}_3)_2$  and  $\text{Mg}(\text{CH}_3)_2 \cdot \text{MgCl}_2$ .

## Computational methods

### Homolytic cleavage energies

**Benchmarking studies** Due to the large size of the systems under consideration in our study, we employ DFT to determine the electronic structure of different chemical species. Different exchange correlation (xc) functionals have been shown to yield discrepancies in the estimation thermochemistry and kinetics of chemical processes<sup>21</sup> and binding energies.<sup>22</sup> Therefore, we benchmarked the performance of different xc functionals against CCSD(T) data.

We benchmarked the quality of DFT calculations using PBE,<sup>23</sup> PBE0,<sup>24</sup> B3LYP,<sup>25,26</sup> M062X,<sup>27</sup>

B97X<sup>28</sup> and TPSSH<sup>29,30</sup> exchange-correlation functionals (xc) in reproducing the homolytic C-Mg bond cleavage energy with respect to coupled-cluster (CC) data. The benchmark energies were computed over geometries optimized using the PBE0 functional for the system shown in Figure 1, considering both the whole system, and the CH<sub>3</sub><sup>·</sup> and RR'ClMg<sup>·</sup> fragments.

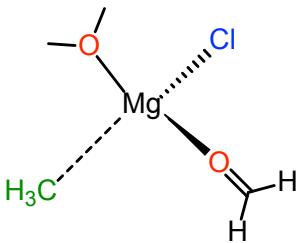


Figure 1: Homolytic cleavage reaction that was used for calibration of the Exchange correlation method takes place along the dashed line.

The reference CC energies were obtained using single, double and perturbative triple excitations (CCSD(T)). CCSD(T) calculations were performed using the PTVZ basis set and frozen core approximation, with an energy convergence cutoff to 10<sup>-9</sup> Hartree. The calculations were carried out using the CFOUR package.<sup>31</sup> The DFT energies were obtained using the 6-31+G(d,p) basis set using the Gaussian09 software package.<sup>32</sup>

Table 1 reports the homolytic dissociation energies of a model Grignard reagent bound to formaldehyde. The M062X xc functional is the best in reproducing CC data, in particular, showing a discrepancy not higher than 1.2 kcal mol<sup>-1</sup> in the dissociation energy. Re-optimization of the structures at the M062X level and consecutively obtaining the energy heightens the barrier by 1.7 kcal mol<sup>-1</sup>. The GGA PBE functional, whose reduced costs are necessary for the *ab initio* MD simulations, still show a reasonable performance, even though it tends to underestimate dissociation energies by 4.2 kcal mol<sup>-1</sup>.

Table 1: Homolytic dissociation energies of a model Grignard reagent bound to Formaldehyde evaluated using different levels of theory according to Figure 1.

Functional	Basis Set	$\Delta E$ [kcal mol $^{-1}$ ]
CCSD(T)	DZ	43.64
CCSD(T)	TV	41.44
PBE0	6-31(d,p)	34.59
B3LYP	6-31(d,p)	32.55
M062X	6-31(d,p)	42.69
B97X	6-31(d,p)	39.39
TPSSh	6-31(d,p)	32.48
PBE	6-31(d,p)	37.22

If not explicitly stated otherwise, all electronic structures following in this paper have been optimized using the M062X functional and the 6-31 basis set. Dispersion forces were accounted for using the D3 Grimme approximation.<sup>33</sup> The calculations were performed in implicit THF solvent using the SMD solvent model.<sup>34</sup>

## Ab initio molecular dynamics

**Monomeric species** We built two different monomeric systems, containing each one molecule of either CH<sub>3</sub>MgCl or Mg(CH<sub>3</sub>)<sub>2</sub> surrounded by 41 molecules of tetrahydrofuran (THF) and one molecule of acetaldehyde in a periodic box of dimensions 25.2 x 15.0 x 15.0 Å<sup>3</sup>. The initial coordinates of THF were generated by 20 ps of ab-initio MD at the experimental density at room temperature.<sup>35</sup>

**Dimeric species** The (MgCH<sub>3</sub>Cl)<sub>2</sub> species were simulated in an orthorhombic periodic box of dimensions 25.2 x 15.0 x 15.0 Å<sup>3</sup>, containing 41 THF molecules and one molecule of Acetaldehyde. Two systems that were obtained as equilibrium structures from our previous investigation<sup>36</sup> were used as starting geometries.

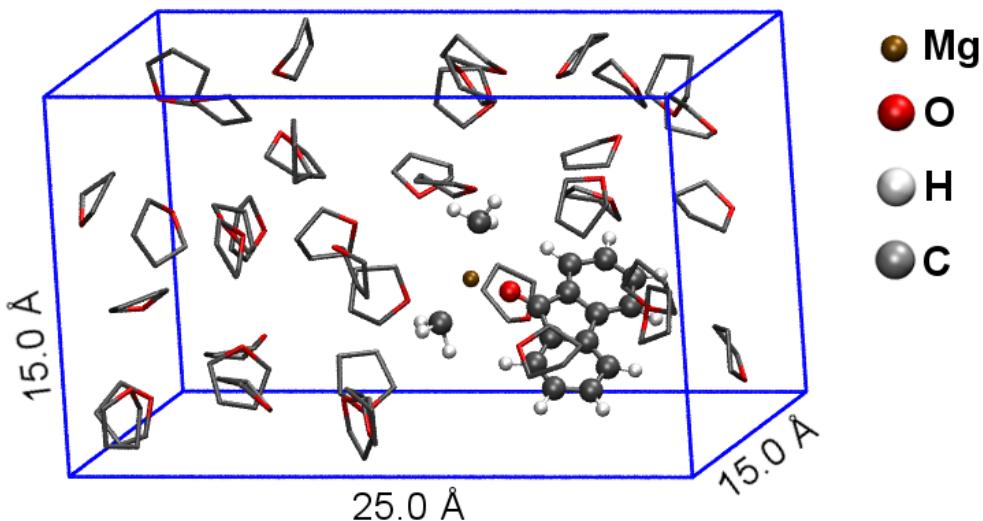


Figure 2: Example simulation Box. Grignard reagent and Fluorenone in CPK, solvent in Licorice representation, the solvent hydrogen atoms are not shown in visualization.

**Simulation parameters** The electronic problem was solved by Density Functional Theory <sup>37–39</sup> using the Perdew-Burke-Ernzerhof exchange correlation functional (PBE).<sup>23</sup> Kohn-Sham orbitals were expanded over mixed Gaussian and plane-wave basis functions. The DZVP basis set for first and second row elements, and a molecularly optimised (MOLOPT) basis set for the Chlorine atoms were employed.<sup>40</sup> The auxiliary plane-wave basis set was expanded to a 200 Ry cutoff. The core electrons were integrated out using pseudopotentials of the Goedecker-Teter-Hutter type (GTH).<sup>41</sup> Dispersion forces was accounted for using the D3 Grimme approximation.<sup>33</sup> Ab initio Molecular Dynamics simulations (AIMD) were run over the ground state potential energy surface, with a time-step of 0.25 fs, optimising the energy gradient to a threshold of  $10^{-5}$  au. All systems were first relaxed for 15 ps in the microcanonical ensemble at a average temperature of 300 K. Then, production runs were simulated in the NVT ensemble at 300 K. Relaxation at the target temperature was first performed using canonical sampling/velocity rescaling (CSVR) thermostat with a time constant of 10 fs until the temperature of the system oscillated around the target value. Then, a Nosé-Hoover chain thermostat with a chain length of 3 and time constant 1 ps was used for data production.<sup>42–44</sup>

Trajectory analysis was performed using the tools available in the VMD 1.9.2 package.<sup>45</sup>

**Constrained Ab initio Molecular Dynamics** In Constrained Molecular Dynamics (CMD) the reaction coordinate was constrained using the Shake algorithm. The Lagrangian Multipliers of the Shake Algorithm were collected every MD Step after equilibration and converged to a Normal distribution with a steady mean value within 1ps of simulations. The activation energy was computed by trapezoidal integration over the average Lagrangian multiplier from reactant to transition state, identified as the point along the reaction coordinate at which the average constraint force is zero. The reaction coordinate used for the investigation of the nucleophilic attack is the methyl carbon-carbonyl carbon distance.

## Results and Discussion

### Homolytic bond cleavage

**Reactivity trends for different substrates** We investigated the change in the homolytic CH<sub>3</sub>-Mg cleavage energy when different substrates are bound to the Grignard reagent. Specifically, we compared formaldehyde, acetaldehyde, carbonyl flouride and fluorenone.

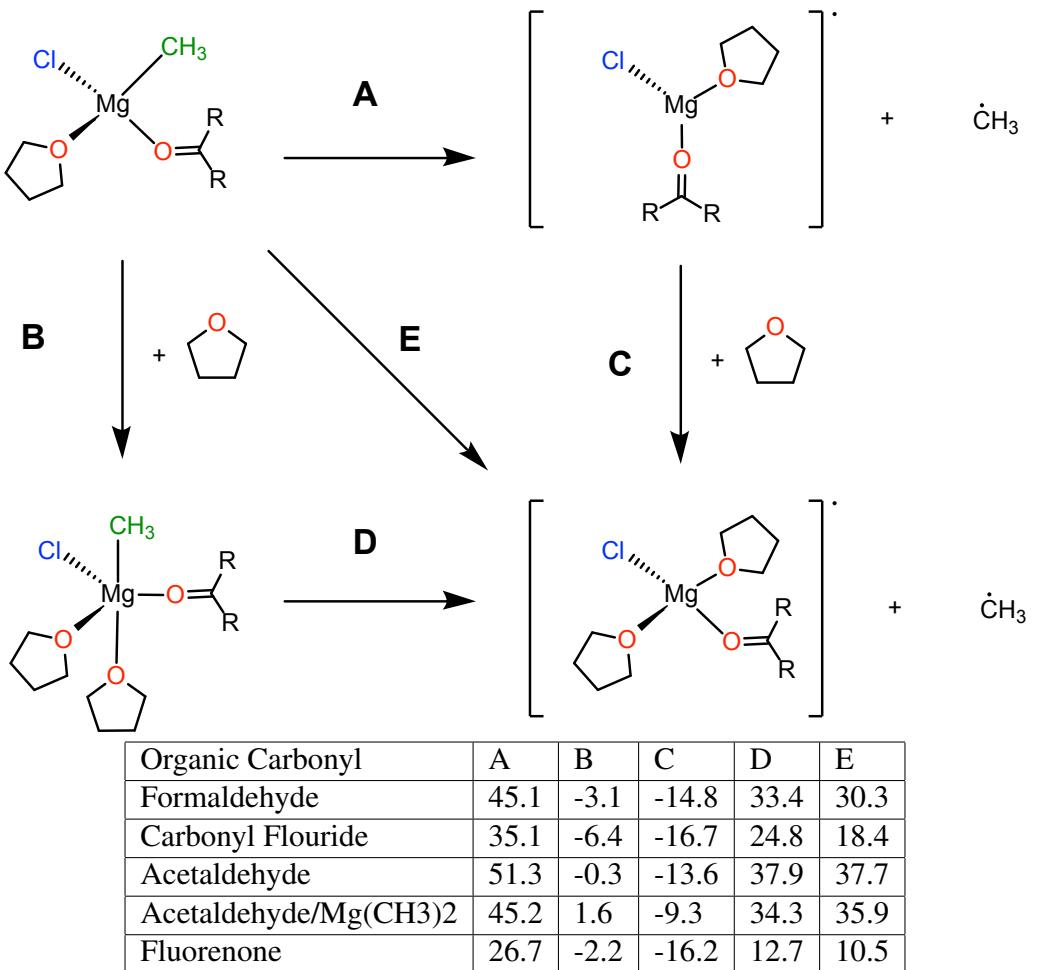


Figure 3: Free energies ( $\Delta G$ ) for the monomeric Grignard reagent calculated according to the reaction scheme above, reported are in [ $\text{kcal mol}^{-1}$ ]. In the Acetaldehyde/ $\text{Mg}(\text{CH}_3)_2$  system the chloride is replaced with a methyl group.

Figure 3 reports the free energies for the homolytic cleavage reaction for the various systems computed as the difference between the change in the electronic energy and standard entropic corrections. In all cases, our data show that the complex between the Grignard reagent and the substrate coexists at room temperature with different solvation structures, yielding both a tetrahedral or pentahedral coordination for the Mg atom. On the contrary, the radical product exists strictly in a tetrahedral form, with the pentahedral coordination at least 10  $\text{kcal mol}^{-1}$  higher in free energy for all species.

The presence of progressively more electron-withdrawing groups attached to the carbonyl

$(CH_3 < H < F)$  facilitates the release of the  $CH_3\cdot$  radical from Mg. The reaction is even more facilitated by coupling the carbonyl to an aromatic moiety like in fluorenone.

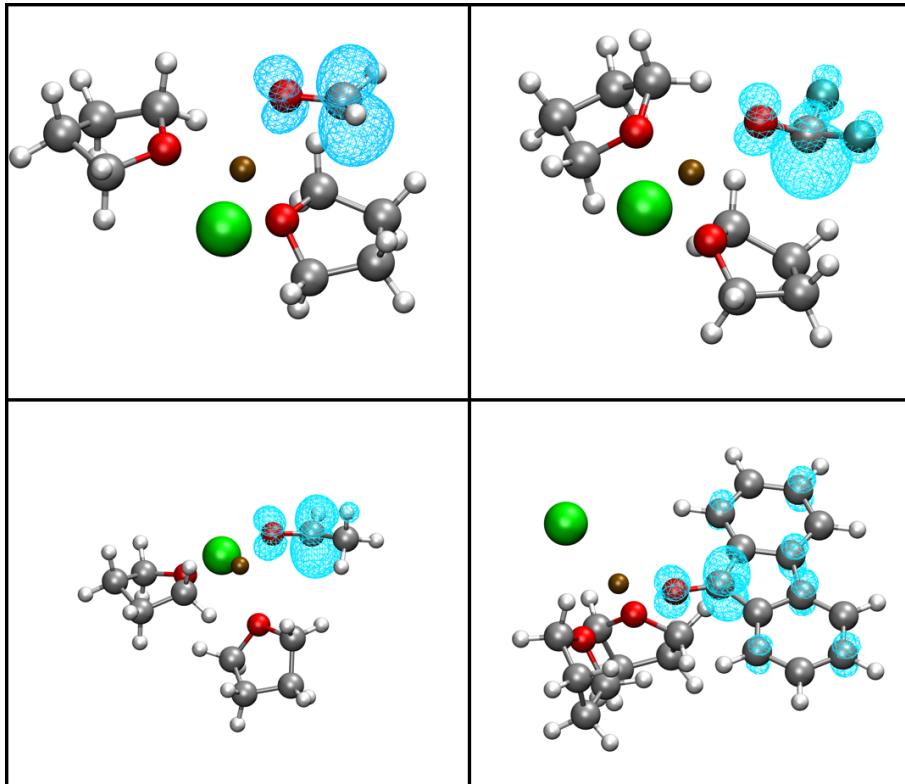


Figure 4: Spin density of the homolytic cleavage product for (*Top Left*) Formaldehyde, (*Top Right*) Carbonyl Fluoride, (*Bottom Left*) Acetaldehyde and (*Bottom Right*) Fluorenone.

Figure 4 reports the localization of the spin density on the carbonyl moieties for the four different substrates. The presence of electron-releasing substituents forces a higher spin density localization on the carbonyl moiety; on the contrary, fluorine atoms facilitate its delocalization. This effect is prominent in fluorenone, where the unpaired electron is spread over the extended aromatic system.

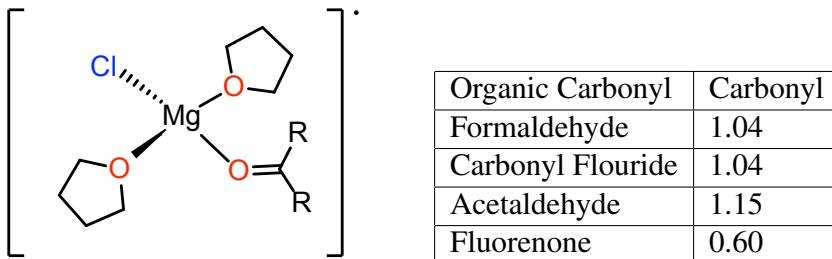


Figure 5: Spin densities on the carbonyl carbon.

## Nucleophilic attack

The energies reported in Figure 3 indicate that formation of a radical species is very unlikely for non-aromatic substrates. On the contrary, binding to fluorenone makes a homolytic cleavage of the C-Mg bond feasible for a relatively low energy. To determine the feasibility of the radical mechanism, we computed the activation energy for the competing nucleophilic attack pathway. The reacting complexes were built from the structures of the Grignard reagent obtained in a previous study,<sup>36</sup> and are shown in Figure 6.

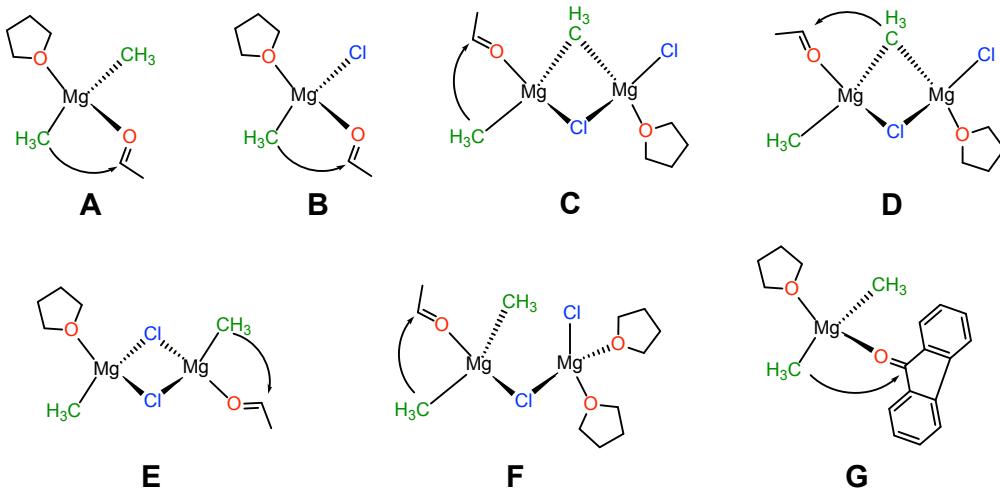
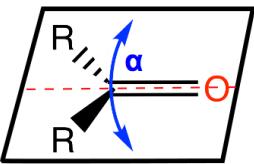


Figure 6: Structures investigated using Constrained Molecular Dynamics. The arrows indicate the path of nucleophilic attack investigated. The activation energy for the nucleophilic attack in each of the structures is detailed in Figure 7.

Figure 7 reports the activation free energies computed by CMD for the different complexes. We find that the more electron donating methyl ligands there are on the magnesium center, the

lower the activation energy is for the nucleophilic attack. This trend in activation energy persists for monomeric as well as dimeric Grignard reagent species. Furthermore, the nucleophilic attack from a terminal ligand is preferred over the one from a bridging ligand (Figure 6). This behavior confirms our previous study where it was shown that the free electron pair of the bridging methyl group is shared between the magnesium centers and points away from the remaining ligands.<sup>36</sup>



System	A	B	C	D	E	F	G
Activation Energy [kcal mol <sup>-1</sup> ]	6.6	8.7	12.9	22.6	13.1	6.4	14.6
TS distance [Å]	2.57	2.54	2.38	2.22	2.37	2.63	2.48
Distortion angle [°]	10.8	12.8	14.2	18.3	14.5	10.7	14.3

Figure 7: Dissociation energy, estimated TS distance and distortion angle obtained for the nucleophilic attack for the structures in 6. The transition state distance was estimated by linear extrapolation of the forces obtained from CMD taken at intervals of 0.1 Å. The distortion angle is with respect to the ideal geometry of a  $sp^2$  hybridized system.

Analysis of the TS geometry shows that as the more reactive species have a more reactant-like transition state, with a longer C-C distance and a more planar geometry of electrophilic carbonyl carbon. Comparing the same Grignard reagent structure, comparing different substrates (for example, systems A - G) we find that for the most reactive acetaldehyde, the distortion angle is smaller than for the bulkier fluorenone.

The nucleophilic attack is most probably catalyzed by monomeric Grignard reagents, as these do not only show the lowest activation energy but also constitute the most abundant species in THF-solvent.<sup>36</sup> For example, structure F (Figure 6), which shows a comparable energy of activation with respect to dimethylmagnesium, was found in our previous study to be energetically unfavored by about 8 kcal mol <sup>-1</sup> with respect to monomeric geometries, and it is therefore significantly less abundant in solution.

## Nucleophilic attack vs Homolytic cleavage

The determination of the preference for the nucleophilic attack over a radical mechanism for the Grignard reactions depends on the relative height of the respective activation energies.

The homolytic bond cleavage energy represents the highest possible activation energy for a radical reaction producing a free methyl radical in solution. In that case, this highly energetic species is expected to recombine either to a free substrate, or to a substrate bound to a Grignard reagent. In both cases, we find that the overall reaction is highly exothermic by  $\sim -50 \text{ kcal mol}^{-1}$ . (Figure 8):

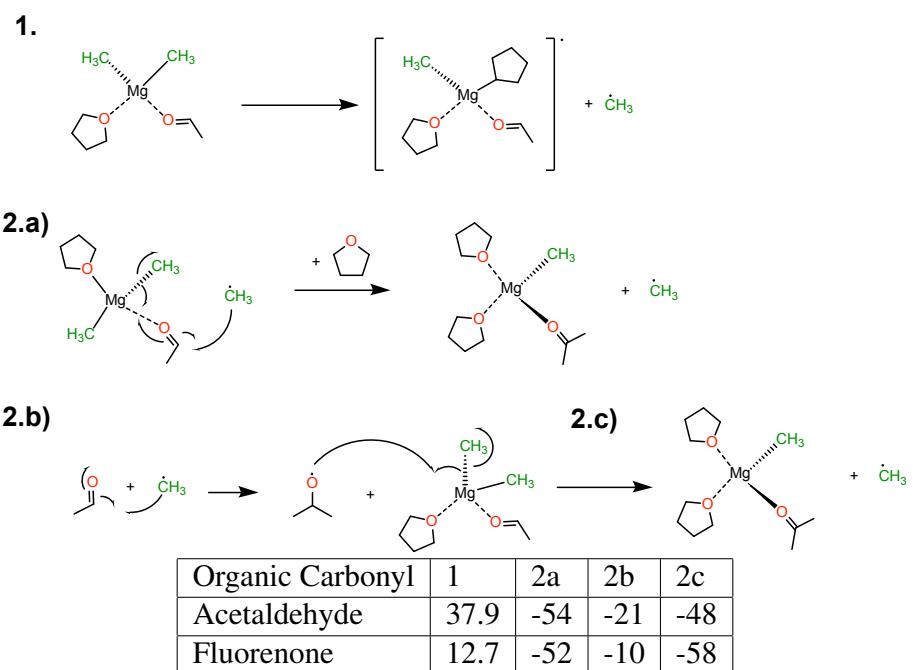


Figure 8: Energies ( $\Delta E$ ) of the radical propagation pathway in [ $\text{kcal mol}^{-1}$ ].

In fact, a radical mechanism may require even lower activation barriers if the initial formation of the radical is followed by a rapid recombination with the substrate within the same solvation cage, avoiding the release of the relatively unstable free methyl radical.

System	A	B	C	E	G
Radical dissociation energy [kcal mol <sup>-1</sup> ]	35.9	37.7	63.2	34.5	12.7
Non-radical activation energy [kcal mol <sup>-1</sup> ]	6.6	8.7	12.9	13.1	14.6

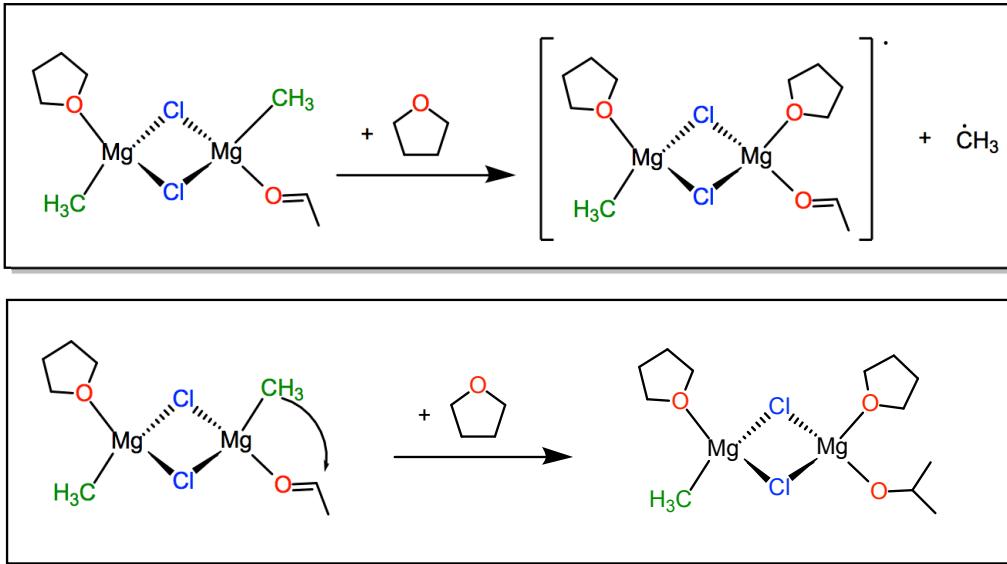


Figure 9: (*Top*) Dissociation and activation energy obtained for the radical and non-radical pathway for the structures shown in Figure 6, exemplified in (*Bottom*) for structure E.

Regardless of the different Grignard reagent structures, the radical mechanism for an alkyl carbonyl like acetaldehyde seems extremely unlikely. In fact, the activation energy for the radical pathway is in all cases above 30 kcal mol<sup>-1</sup>. This contrasts the very favorable activation energy for the nucleophilic addition, which can be as small as 6.6 kcal mol<sup>-1</sup>. We remark that the activation free energies for the nucleophilic attack were computed by AIMD using the PBE functional, while the radical formation energies were computed with the M062X one. For a better consistency between the data, we estimated the total energy difference between the reactant and TS during the nucleophilic attack using the two functionals, finding that, although the geometries are very similar, PBE underestimates the barrier by ~3.3 kcal mol<sup>-1</sup> compared to M062X. This discrepancy is well below the energy differences found between the two pathways, and thus it does not change the qualitative picture.

For fluorenone, we find that the radical pathway is favored, even though the difference between the relative barriers is less pronounced. The radical formation is in particular stabilised by the large

delocalization of the unpaired electron over the conjugated  $\pi$  system, while the nucleophilic attack is disfavoured by the necessary distortion of the geometry of the carbonyl that needs to break the same conjugation.

## Concluding remarks

Our study confirm the hypothesis that the Grignard reaction may proceed both by a nucleophilic addition and by a radical mechanism. We found that the preferences of either of the two mechanisms is associated to the nature of the substrate. In particular, extended aromatic moieties like in fluorenone are crucial for the initial stabilization of an unpaired electron that can initiate the radical cascade. In particular, they are in agreement with previous studies that have observed radical formation in the Grignard reaction of aromatic ketones such as Benzophenone<sup>46</sup> and Fluorenone.<sup>47</sup> For non-aromatic carbonyls where the radical can be localized solely on the carbonyl carbon, strongly negative inductive effects, for example by electronegative substituents, are needed to lower enough the energy of the species. We furthermore found that the nucleophilic addition reaction shows a clear trend between the different species present in the Schlenk-equilibrium. Specifically, the most reactive species in THF solvent are  $\text{Mg}(\text{CH}_3)_2$ ,  $\text{CH}_3\text{MgCl}$  and  $\text{Mg}(\text{CH}_3)_2 \cdot \text{MgCl}_2$ , respectively.

The presented data shed light on some complex mechanistic aspects of the magnesium chemistry in solution, and can provide solid ground for future studies on catalytic metallocorganic chemistry.

## Associated Content

The authors declare no competing financial interest.

## Acknowledgement

This work was supported by the Research Council of Norway (RCN) through the CoE Hylleraas Centre for Quantum Molecular Sciences Grant No. 262695, and by the Norwegian Supercomputing Program (NOTUR) (Grant No. NN4654K).

## References

- (1) Vollhardt, K.; Schore, N. *Organic Chemistry: Structure and Function*; W. H. Freeman, 2014.
- (2) Grignard, V. C. Sur quelques nouvelles combinaisons organométalliques du magnésium et leur application à des synthèses d'alcools et d'hydrocarbures. *ompt. rend. Hebd. Séances Acad. Sci.* **1900**, *130*, 1322–24.
- (3) Schlenk, W.; Schlenk, W. Über die Konstitution der Grignardschen Magnesiumverbindungen. *Berichte der deutschen chemischen Gesellschaft (A and B Series)* **1929**, *62*, 920–924.
- (4) Seyferth, D. The Grignard Reagents. *Organometallics* **2009**, *28*, 1598–1605.
- (5) Ehlers, A. W.; van Klink, G. P. M.; van Eis, M. J.; Bickelhaupt, F.; Nederkoorn, P. H. J.; Lammertsma, K. Density-Functional Study of (Solvated) Grignard Complexes. *Journal of Molecular Modeling* **2000**, *6*, 186–194.
- (6) Silverman, G. S.; Rakita, P. E. *Handbook of Grignard reagents*; CRC Press: New York, 1996.
- (7) Fauvarque, J.; Rouget, E. *Hebd. Séances Acad. Sci., Ser. C* **1968**, *257*, 1355, cited By 5.
- (8) Blomberg, C.; Grootveld, H.; Gerner, T.; Bickelhaupt, F. Radical formation during reactions of Grignard reagents with quinones. *Journal of Organometallic Chemistry* **1970**, *24*, 549 – 553.
- (9) Ashby, E. C.; Nackashi, J.; Parris, G. E. Composition of Grignard compounds. X. NMR, ir, and molecular association studies of some methylmagnesium alkoxides in diethyl ether,

tetrahydrofuran, and benzene. *Journal of the American Chemical Society* **1975**, *97*, 3162–3171.

- (10) Ashby, E. C.; Lopp, I. G.; Buhler, J. D. Mechanisms of Grignard reactions with ketones. Polar vs. single electron transfer pathways. *Journal of the American Chemical Society* **1975**, *97*, 1964–1966.
- (11) The quest for chiral Grignard reagents. *Chemical Society reviews* **2003**, *32*, 225–230.
- (12) Mori, T.; Kato, S. Grignard reagents in solution: Theoretical study of the equilibria and the reaction with a carbonyl compound in diethyl ether solvent. *Journal of Physical Chemistry A* **2009**, *113*, 6158–6165.
- (13) Chmgci, R.; Schlenk, C. T.; Axten, J.; Troy, J.; Jiang, P.; Trachtman, M.; Bock, C. W. An ab initio Molecular Orbital Study of the Grignard. *Structural Chemistry* **1994**, *5*, 99–108.
- (14) Tammiku, J.; Burk, P.; Tuulmets, A. 1,10-phenanthroline and its complexes with magnesium compounds. Disproportionation equilibria. *Journal of Physical Chemistry A* **2001**, *105*, 8554–8561.
- (15) Tammiku-Taul, J.; Burk, P.; Tuulmets, A. Theoretical study of magnesium compounds: The Schlenk equilibrium in the gas phase and in the presence of Et<sub>2</sub>O and THF molecules. *Journal of Physical Chemistry A* **2004**, *108*, 133–139.
- (16) Yamazaki, S.; Yamabe, S. A computational study on addition of grignard reagents to carbonyl compounds. *Journal of Organic Chemistry* **2002**, *67*, 9346–9353.
- (17) Chen, Z.-N.; Fu, G.; Xu, X. Theoretical studies on Grignard reagent formation: radical mechanism versus non-radical mechanism. *Organic & Biomolecular Chemistry* **2012**, 9491–9500.
- (18) Garst, J. F.; Soriaga, M. P. Grignard reagent formation. *Coordination Chemistry Reviews* **2004**, *248*, 623–652.

- (19) Shao, Y.; Liu, Z.; Huang, P.; Liu, B. A unified model of Grignard reagent formation. *Phys. Chem. Chem. Phys.* **2018**, *20*, 11100–11108.
- (20) Ignatov, S. K.; Panteleev, S. V.; Maslennikov, S. V.; Spirina, I. V. Mechanism of the Grignard reaction in terms of the cluster model of reaction center. A quantum-chemical study. *Russian Journal of General Chemistry* **2012**, *82*, 1954–1961.
- (21) Goerigk, L.; Grimme, S. A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670–6688.
- (22) Zhao, Y.; Truhlar, D. G. Benchmark Databases for Nonbonded Interactions and Their Use To Test Density Functional Theory. *Journal of Chemical Theory and Computation* **2005**, *1*, 415–432, PMID: 26641508.
- (23) Perdew, J. P.; Burke, K.; Ernzerhof, M.; of Physics, D.; Quantum Theory Group Tulane University, N. O. L. . J. Generalized Gradient Approximation Made Simple. *Physical Review Letters* **1996**, *77*, 3865–3868.
- (24) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *The Journal of Chemical Physics* **1999**, *110*, 6158–6170.
- (25) Becke, A. D. Densityfunctional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics* **1993**, *98*, 5648–5652.
- (26) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.
- (27) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theoretical Chemistry Accounts* **2008**, *120*, 215–241.

- (28) Chai, J.-D.; Head-Gordon, M. Systematic optimization of long-range corrected hybrid density functionals. *The Journal of Chemical Physics* **2008**, *128*, 084106.
- (29) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional Ladder: Nonempirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003**, *91*, 146401.
- (30) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes. *The Journal of Chemical Physics* **2003**, *119*, 12129–12137.
- (31) CFOUR, a quantum chemical program package written by J.F. Stanton, J. Gauss, L. Cheng, M.E. Harding, D.A. Matthews, P.G. Szalay with contributions from A.A. Auer, R.J. Bartlett, U. Benedikt, C. Berger, D.E. Bernholdt, Y.J. Bomble, O. Christiansen, F. Enge.
- (32) Frisch, M. J. et al. Gaussian09 Revision B.01. 2009; Gaussian Inc. Wallingford CT.
- (33) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics* **2010**, *132*, 154104.
- (34) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *The Journal of Physical Chemistry B* **2009**, *113*, 6378–6396, PMID: 19366259.
- (35) Metz, D. J.; Glines, A. Density, viscosity, and dielectric constant of tetrahydrofuran between -78 and 30.degree. *The Journal of Physical Chemistry* **1967**, *71*, 1158–1158.
- (36) Peltzer, R. M.; Eisenstein, O.; Nova, A.; Cascella, M. How Solvent Dynamics Controls the Schlenk Equilibrium of Grignard Reagents: A Computational Study of CH<sub>3</sub>MgCl in Tetrahydrofuran. *The Journal of Physical Chemistry B* **2017**, *121*, 4226–4237, PMID: 28358509.

- (37) Lippert, G.; Hutter, J.; Parrinello, M. The Gaussian and augmented-plane-wave density functional method for ab initio molecular dynamics simulations. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)* **1999**, *103*, 124–140.
- (38) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864–B871.
- (39) Kohn, W.; Sham, L. J. Self-consistent equations including exchange and correlation effects. 1965.
- (40) VandeVondele, J.; Hutter, J. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *The Journal of Chemical Physics* **2007**, *127*, 114105.
- (41) Goedecker, S.; Teter, M.; Hutter, J. Separable dual-space Gaussian pseudopotentials. *Physical Review B* **1996**, *54*, 1703–1710.
- (42) Martyna, G. J.; Klein, M. L.; Tuckerman, M. NoséHoover chains: The canonical ensemble via continuous dynamics. *The Journal of Chemical Physics* **1992**, *97*, 2635–2643.
- (43) Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *Journal of Chemical Physics* **1984**, *81*, 511–519.
- (44) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A* **1985**, *31*, 1695–1697.
- (45) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **1996**, *14*, 33–38.
- (46) Ashby, E. C. A detailed description of the mechanism of reaction of Grignard reagents with ketones. *Pure and Applied Chemistry* **1980**, *52*, 545–569.
- (47) Lund, T.; Pedersen, M. L.; Frandsen, L. A. Does the reaction between fluorenone and grignard reagents involve free fluorenone anion radicals? *Tetrahedron Letters* **1994**, *35*, 9225 – 9226.

## Paper III

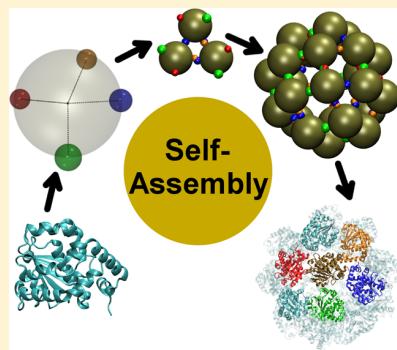
# Self-Assembly of $\alpha$ -Tocopherol Transfer Protein Nanoparticles: A Patchy Protein Model

Raphael Mathias Peltzer,<sup>†</sup> Hima Bindu Kolli,<sup>†</sup> Achim Stocker,<sup>\*,‡</sup> and Michele Casella<sup>\*,†</sup>

<sup>†</sup>Department of Chemistry, and Hylleraas Centre for Quantum Molecular Sciences, University of Oslo, P.O. Box 1033, Blindern, 0315 Oslo, Norway

<sup>‡</sup>Department of Chemistry and Biochemistry, University of Bern, Freiestrasse 3, 3012 Bern, Switzerland

**ABSTRACT:** We describe the mechanism of self-aggregation of  $\alpha$ -tocopherol transfer protein into a spherical nanocage employing Monte Carlo simulations. The protein is modeled by a patchy coarse-grained representation, where the protein–protein interfaces, determined in the past by X-ray diffraction, are represented by simplified two-body interaction potentials. Our results show that the oligomerization kinetics proceeds in two steps, with the formation of metastable trimeric units and the subsequent assembly into the spherical aggregates. Data are in agreement with experimental observations regarding the prevalence of different aggregation states at specific ambient conditions. Finally, our results indicate a route for the experimental stabilization of the trimer, crucial for the understanding of the physiological role of such aggregates in vitamin E body trafficking.



## INTRODUCTION

$\alpha$ -Tocopherol transfer protein ( $\alpha$ -TTP hereafter) is the liver factor responsible for the retention of RRR- $\alpha$ -tocopherol ( $\alpha$ -tol), the active isoform of vitamin E, in the human body.<sup>1–3</sup>  $\alpha$ -TTP solubilizes  $\alpha$ -tol from the external leaflet of maturing endosomal compartments, promoting its release into the blood. Structural studies over the years<sup>4,5</sup> indicate that  $\alpha$ -TTP is active as a monomer, similarly to other transporters of its family.<sup>6–8</sup> Unlike for retention, the mechanism(s) by which  $\alpha$ -tol is secreted into the blood and then absorbed into the target tissues is at present not well understood. Possible pathways for the secretion and blood transport of  $\alpha$ -tol include enrichment into the leaflets of the plasma membrane by a lipid-exchange mechanism<sup>9,10</sup> and transport into the blood by aggregating to very-low-density lipoprotein vesicles.<sup>2</sup> The absorption from the blood into the target tissues is even less understood, but it must imply some mechanism of recognition to bypass endothelial barriers, like the blood–brain barrier or the placenta.

The involvement of  $\alpha$ -TTP into  $\alpha$ -tol trafficking has not been clearly defined. In a recent work, Arai and co-workers have suggested that transfer of  $\alpha$ -tol to the plasma membrane is coupled to the extraction of phosphatidylinositol phosphates (PIPs) from the same membrane by  $\alpha$ -TTP.<sup>3</sup> Interestingly, they also suggested that lipid exchange at the plasma membrane may involve higher order aggregates of  $\alpha$ -TTP than the monomers.<sup>3</sup>

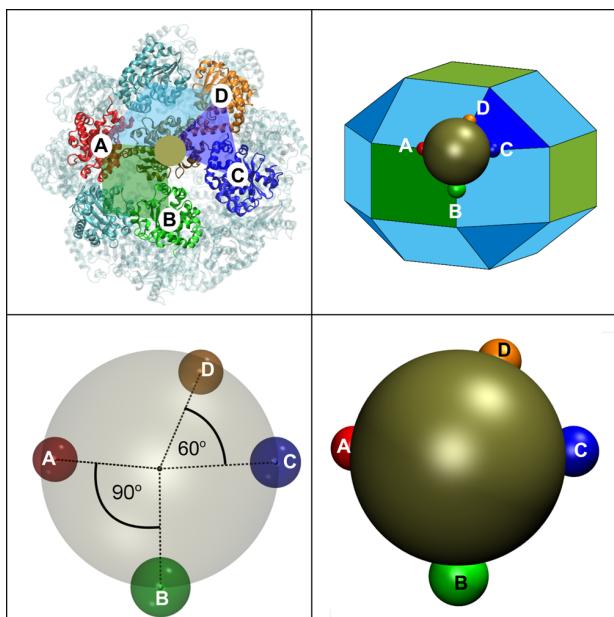
Very recently, we provided structural evidence that upon binding to  $\alpha$ -tol,  $\alpha$ -TTP acquires the tendency to oligomerize.<sup>11</sup> The oligomerized proteins form stable, regular spherical nanoparticles composed of 24  $\alpha$ -TTP units ( $\alpha$ -TTP<sub>S</sub>), which could be characterized by a series of methods,

including cryo-EM and X-ray diffraction (PDB: 5MUE and 5MUG).<sup>11</sup> Thermal analysis demonstrated that  $\alpha$ -TTP<sub>S</sub> is thermodynamically stable; furthermore, oxidative conditions enhance its stability by promoting the formation of 12 disulfide bonds cross-linking different  $\alpha$ -TTP units. X-ray crystallography data of  $\alpha$ -TTP<sub>S</sub> revealed a regular assembly of 24 monomers organized in a cubic symmetry. Each  $\alpha$ -TTP unit is located on one vertex of a cantellated cube, and it is involved in two kinds of molecular contacts with four neighboring proteins (Figure 1). The first interface builds around the C<sub>4</sub> symmetry axis of the assembly, and it is constituted by a patch of surface amino acids that are exposed to the solvent in the native monomeric folding of  $\alpha$ -TTP. The second interface is responsible for the assembly of  $\alpha$ -TTP around the trimeric C<sub>3</sub> axis. This interface is located on the surface of the proper SEC-14-like binding domain; in  $\alpha$ -TTP, it is screened from the solvent by the N-terminal helical domain. In  $\alpha$ -TTP<sub>S</sub>, the interfaces are accessible to the partner proteins thanks to the unfolding of the first N-terminal helix, which is not detectable in the corresponding X-ray structure. The partial unfolding of the N-terminal helix is triggered by external conditions, including binding to  $\alpha$ -tol or interaction with negatively charged lipids.<sup>11</sup>

$\alpha$ -TTP<sub>S</sub> shows selective and efficient transport properties through in vitro models of endothelial barriers,<sup>11</sup> making it a potential candidate as one of the physiological route for the delivery of vitamin E into the brain.<sup>12,13</sup> As much as oligomerization of  $\alpha$ -TTP is crucial for its transfecting properties, its mechanistic aspects remain obscure. In

Received: June 21, 2018

Published: June 26, 2018



**Figure 1.** Coarse-grained model of  $\alpha$ -TTP<sub>S</sub>. Top: (left) In the native  $\alpha$ -TTP<sub>S</sub>, any monomeric  $\alpha$ -TTP (gold spot) is in contact with four other proteins along the edges of a cantellated cube (top right). Bottom:  $\alpha$ -TTP is described as a sphere with four interaction sites (IS) corresponding to the protein–protein contacts in  $\alpha$ -TTP<sub>S</sub>.

particular, chromatographic data showed that when the monomeric form is the most stable aggregation state for  $\alpha$ -TTP, this is at equilibrium with a small presence of low-weight dimeric or tetrameric aggregates. On the contrary, when the aggregation into the high weight oligomer is triggered, the only species present in the solution is monomeric  $\alpha$ -TTP or regular  $\alpha$ -TTP<sub>S</sub> constructs, whereas no other low-, middle-weight assemblies coexist in detectable concentration.<sup>11</sup>

In the present study, we investigated the formation of  $\alpha$ -TTP<sub>S</sub> using Monte Carlo (MC) simulations of a toy patchy model of  $\alpha$ -TTP. Molecular simulations using patchy models are proven to be very effective in understanding the nature of self-assembly in systems like patchy colloids, soft functionalized nanoparticles, and biomolecules.<sup>14–17</sup> In particular, models with anisotropic and highly directional interacting patches are particularly suited to describe protein assembly.<sup>18–23</sup>

## COMPUTATIONAL METHODS

**Computational Model.** The patchy model for one  $\alpha$ -TTP consisted of one hard sphere, with four interaction sites (IS) located on its surface (Figure 1). The relative orientation of the IS was chosen to mimic the geometric organization of  $\alpha$ -TTP<sub>S</sub>. Interactions were imposed between IS<sub>A</sub> and IS<sub>B</sub> and between IS<sub>C</sub> and IS<sub>D</sub> types, consistently with the experimental structure of  $\alpha$ -TTP<sub>S</sub>.

The interaction potentials for the two IS pairs were described by toy potential wells of depth  $E_{A/B} = u$ ,  $E_{C/D} = 3u$  ( $u$  being an arbitrary unit of energy), dependent on both the distance between the IS and the relative orientation of the proteins. The initial 1:3 ratio between  $E_{A/B}$  and  $E_{C/D}$  was calibrated on an estimate of the dimerization free energies from atomistic models using a standard thermodynamic cycle,<sup>24</sup> computing the solvation free energy of individual and dimeric

structures solving the linearized Poisson–Boltzmann equation using the APBS software<sup>24</sup> and the binding energy in vacuo using the Amber force field.<sup>25</sup> Protein dimers were extrapolated from the X-ray structure of  $\alpha$ -TTP<sub>S</sub> (PDB: SMUE).<sup>11</sup>

The four IS are identified by four vectors with origin in the center of the hard sphere and ends in

$$\text{IS}_A = (0, R, 0)$$

$$\text{IS}_B = (-R, 0, 0)$$

$$\text{IS}_C = \left(0, \frac{R}{\sqrt{2}}, -\frac{R}{\sqrt{2}}\right)$$

$$\text{IS}_D = \left(\frac{R}{\sqrt{2}}, 0, -\frac{R}{\sqrt{2}}\right)$$

(1)

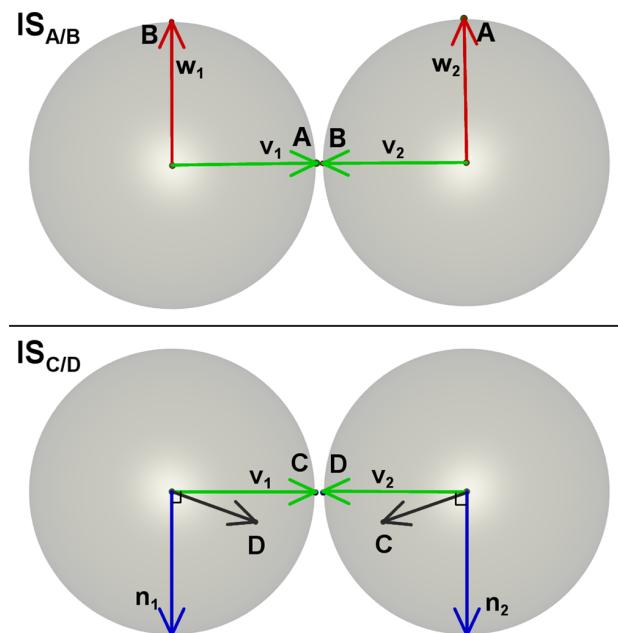
where  $R = 2$  nm is the radius of a hard sphere with its center in  $O(0,0,0)$ .

The potential energies between the IS<sub>A</sub>–IS<sub>B</sub> and IS<sub>C</sub>–IS<sub>D</sub> couples are described by the following well potentials

$$E_{A/B}(r, \psi, \phi) = \begin{cases} 0 & r > r_{\text{cut}} \\ E_{A/B} \cos \psi \cos \phi & r \leq r_{\text{cut}} \end{cases} \quad (2)$$

$$E_{C/D}(r, \psi, \theta) = \begin{cases} 0 & r > r_{\text{cut}} \\ E_{C/D} \cos \psi \cos 2\theta & r \leq r_{\text{cut}} \end{cases} \quad (3)$$

where  $r$  is the IS–IS distance,  $r_{\text{cut}} = 0.2 R$  is the maximum range of the interaction.  $\cos \psi$ ,  $\cos \phi$ , and  $\cos \theta$  are defined from the scalar multiplication of the normalized vectors as in Figure 2. The angular dependency is necessary to model both the chirality of the protein and that protein binding occurs over



**Figure 2.** Normalized vectors used to define the angular dependency of the interaction energy.  $\cos \psi = \mathbf{v}_1 \cdot \mathbf{v}_2$ , for any A/B or C/D interaction (green arrows);  $\cos \phi = \mathbf{w}_1 \cdot \mathbf{w}_2$  for any A/B interaction (top panel, red arrows);  $\cos \theta = \mathbf{n}_1 \cdot \mathbf{n}_2$ , for any C/D interaction (bottom panel, blue arrows).

an extended surface that requires a well-defined orientation of the two partners.

**System Setup.** We simulated a system having  $N = 216$  particles at thermal equilibrium. The protein particles were initially distributed uniformly in a periodic cubic simulation box of edge  $24.625 R$  (where  $R$  is the radius of the protein), corresponding to roughly the experimental concentration at which  $\alpha$ -TTP aggregation is observed.<sup>11</sup> The accessible conformational space in the canonical NVT ensemble was explored using a Metropolis Monte Carlo (MC) algorithm.<sup>26</sup>

Random moves included the rotation or translation of the single particles or whole clusters of bound particles.<sup>27–29</sup> A bond between two particles was assumed to exist if the distance between the corresponding interaction sites was less than  $0.2 R$ . Two particles were considered to belong to same cluster if they were connected by a chain of bonds.<sup>30</sup> Rotational moves made use of quaternion representation of the particle's orientation, which was modified by a smaller random orientation and then renormalized.<sup>27,28</sup>

The canonical ensemble was sampled at different values of temperature to determine thermodynamic regimes at which different  $\alpha$ -TTP aggregates exist. All of the simulations at different temperatures started from a random configuration of the proteins in the box. Simulations were organized in cycles, each cycle consisting of a number of attempted particle moves. Typical equilibration runs consisted of  $(6–9) \times 10^6$  MC cycles and were followed by a production run of additional  $3 \times 10^5$  MC cycles, during which averages of energy and cluster abundance were calculated. Convergence of the results was tested by evaluating both the convergence of the expectation value of the energy and comparing the variance to the typical short time (1000 steps) energy fluctuations at a given temperature over the last  $6 \times 10^5$  MC cycles.

The acceptance probability for the single particle moves like translation or rotation was evaluated according to

$$\left\{ \begin{array}{l} \text{if } E_t \leq E_i, \quad P(\text{acc}) = 1 \\ \text{if } E_t > E_i, \quad P(\text{acc}) = \exp(-\beta(E_t - E_i)) \end{array} \right\} \quad (4)$$

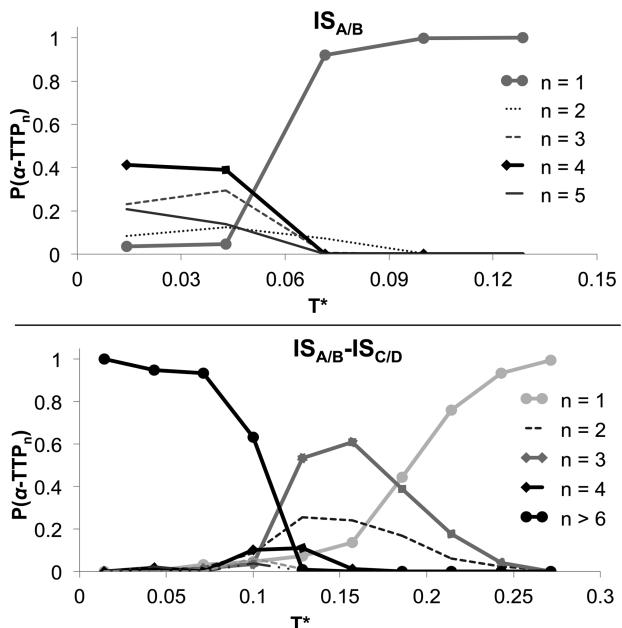
Here,  $P$  is the acceptance probability,  $E_i$  is the initial energy, and  $E_t$  is the energy after the test step.  $\beta$  is the reciprocal thermodynamic temperature of the system. The cluster moves were implemented following the early rejection scheme.<sup>27</sup>

All results here are presented in reduced units:  $U^* = U/u$  and  $T^* = k_B T/u$  for the inner energy and the temperature of the system.

Trajectory analysis was performed using the tools available in the VMD 1.9.2 package.<sup>31</sup>

## RESULTS AND DISCUSSION

In a first set of MC runs, we investigated the existence of  $\alpha$ -TTP aggregates when only the  $IS_{A/B}$  interface is active. This setup mimics the experimental conditions at which the N-terminal region is folded and the  $IS_{C/D}$  is not exposed to the solvent. The top panel of Figure 3 shows the relative abundance of  $\alpha$ -TTP aggregates as a function of the thermal energy. At high temperatures, only monomeric species are present. At around  $T^*u = 0.09$ , we observed the appearance of low-weight aggregates, mostly tetramers, with also a non-negligible presence of trimers and dimers. Lowering the temperature stabilizes the tetrameric packing, which corresponds to the aggregation state of four  $\alpha$ -TTP proteins around the  $C_4$  symmetry axis in  $\alpha$ -TTP<sub>S</sub> (Figure 1). Higher-



**Figure 3.** Aggregation states of  $\alpha$ -TTP ( $\alpha$ -TTP<sub>n</sub>) as a function of the reduced temperature  $T^*$ : (top panel) when only the  $IS_{A/B}$  interface is active or (bottom panel) when both the  $IS_{A/B}$  and  $IS_{C/D}$  interfaces are active.

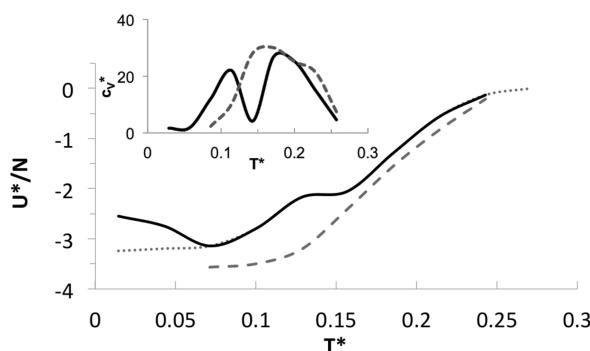
molecular-weight structures, for example, linear chainlike structures along sequences of  $IS_{A/B}$  contacts were not observed, as they are forbidden by the chirality condition on the interaction energy, which only allows the formation of ringlike tetramers.

In a second set of MC simulations, both the  $IS_{A/B}$  and  $IS_{C/D}$  interactions were active. This setup mimics  $\alpha$ -TTP with an unfolded N-terminus. In this case, we observed three temperature regimes at which distinct aggregation states appear (Figure 3 (bottom)). In the high-temperature range ( $T^* > 0.25$ ), only  $\alpha$ -TTP monomers were present. In the narrow ( $0.10 < T^* < 0.25$ ) region, we detected the formation of trimeric species ( $\alpha$ -TTP<sub>3</sub>), whereas for ( $T^* < 0.12$ ), the systems rapidly evolved in high-weight aggregates ( $\alpha$ -TTP<sub>S</sub>, Figure 6).

The trimeric phase includes aggregates built along the  $IS_{C/D}$  interface. In fact,  $\alpha$ -TTP<sub>3</sub>'s are stable in a temperature range at which the thermal energy is too high to allow the formation of A/B contacts. The structure of  $\alpha$ -TTP<sub>3</sub> corresponds to the assembly of three  $\alpha$ -TTP proteins around the  $C_3$  axis of  $\alpha$ -TTP<sub>S</sub>. The high-weight aggregates appearing for ( $T^* < 0.12$ ) are constituted by oligomerization of  $\alpha$ -TTP<sub>3</sub>, up to  $\alpha$ -TTP<sub>S</sub> by formation of  $IS_{A/B}$  contacts.

During the MC runs, we observed a marginal degree of polydispersion especially near the transition temperatures (Figure 3). Nonetheless, the profile of the energy as a function of the temperature shows two clear sigmoidal jumps, indicating the presence of two distinct first-order phase transitions from  $\alpha$ -TTP to  $\alpha$ -TTP<sub>3</sub> to  $\alpha$ -TTP<sub>S</sub> (Figure 4).

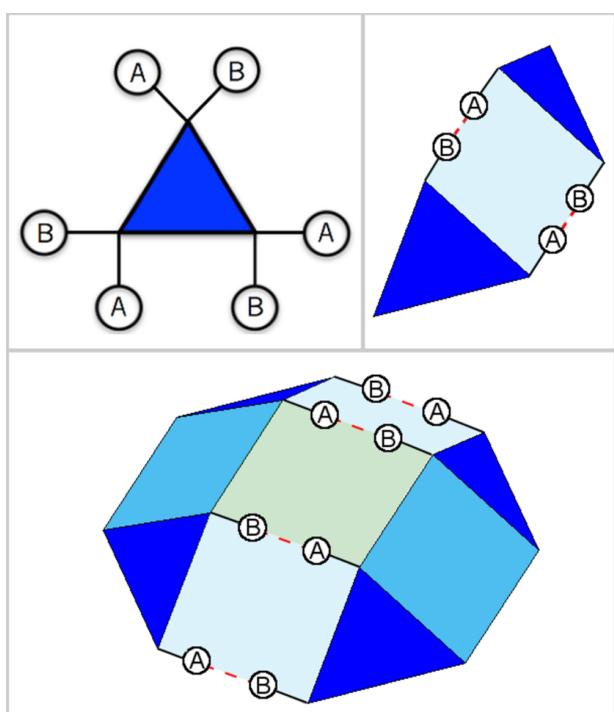
Oligomerization of  $\alpha$ -TTP<sub>S</sub> begins at a higher temperature ( $T^* = 0.12$ ) than the one characterizing  $\alpha$ -TTP aggregation when only  $IS_{A/B}$  is active ( $T^* = 0.07$ ). In fact,  $\alpha$ -TTP<sub>3</sub> dimerization involves binding over two  $IS_{A/B}$  contacts, producing a hexameric structure centered around a  $C_2$  symmetry axis corresponding to one of the  $C_2$  axes of  $\alpha$ -



**Figure 4.** Inner energy per protein  $U^*/N$  and specific heat  $c_V^*$  (inset) as a function of the reduced temperature for the system with both active  $IS_{A/B}$  and  $IS_{C/D}$  (continuous line). The dotted line was obtained by simulated annealing starting from converged data at  $T^* = 0.07$ . The dashed lines report the same data for the system with  $E_{A/B} = 1.5 \mu$ .

$\alpha\text{-TTP}_S$ . Simultaneous formation of two  $IS_{A/B}$  interactions is facilitated by the preorganization of the interaction sites along the edges of the rigid  $\alpha\text{-TTP}_3$ .

The cooperative effect of the  $IS_{A/B}$  onto the binding of  $\alpha\text{-TTP}_3$  is responsible for the absence of intermediate weight aggregates between  $\alpha\text{-TTP}_3$  and  $\alpha\text{-TTP}_S$ . Practically, the assembly of  $\alpha\text{-TTP}_S$  may be schematically seen as the progressive dimerization of  $\alpha\text{-TTP}_3$ ,  $\alpha\text{-TTP}_6$ , and  $\alpha\text{-TTP}_{12}$  over two, four, and eight  $IS_{A/B}$  contacts (Figure 5). Below the critical temperature that allows the first dimerization of  $\alpha\text{-TTP}_3$ , further assemblies involve increasingly larger number of  $IS_{A/B}$  interactions, yielding  $\alpha\text{-TTP}_S$ .



**Figure 5.** Oligomerization of  $\alpha\text{-TTP}_3$  (blue triangle). Each oligomerization step involves the formation of at least two  $IS_{A/B}$  contacts. Newly formed contacts are represented by dashed lines.

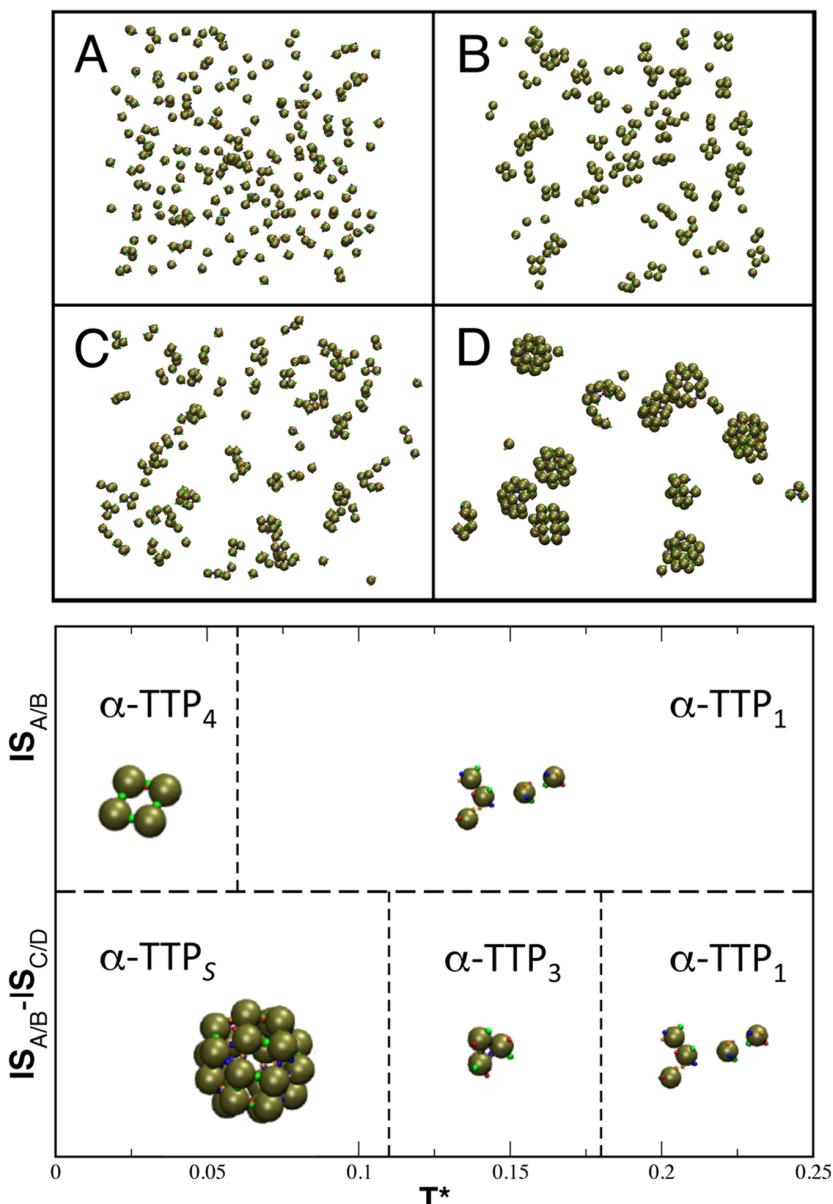
The existence of a region of thermodynamic stability for  $\alpha\text{-TTP}_3$  depends on the relative magnitude of  $E_{A/B}$  versus  $E_{C/D}$ . To verify that, we ran one additional set of MC simulations on a system where  $E_{A/B} = 1.5 \mu$ . In this case, we expected the critical temperature for the formation of  $\alpha\text{-TTP}_3$  to be very similar to that of  $\alpha\text{-TTP}_3$  dimerization. In fact, we observed only one sigmoidal profile of the  $U^*$  vs  $T^*$  plot, indicating the coalescence of the two phase transitions into one (Figure 4) and a direct aggregation from  $\alpha\text{-TTP}$  to  $\alpha\text{-TTP}_S$ .

Although our MC runs depict the clear tendency of the system to form  $\alpha\text{-TTP}_S$ , statistically, we obtained the formation of imperfect spherical complexes with an average aggregation number of 20.2. The presence of defects is visible from the average energy per protein reported in Figure 4, which is always larger than the ideal value of  $-4\mu$  even for low values of the thermal energy. The formation of defectuous  $\alpha\text{-TTP}_S$  assemblies is due to convergence issues related to the appearance of kinetically trapped states at lower temperatures. These states are dominant in MC runs at values of  $T^*\mu \ll E_{A/B}$ , which yielded poorly aggregated structures, with energies consistently higher than the best organized  $\alpha\text{-TTP}_S$ -like structures found at higher values of the temperature. Improvement of the sampled structures at  $T^* \leq 0.06$  was obtained by applying 30 cycles of simulated annealing, between  $T^* = 0.07$  and the target temperature. In this case, we could observe the formation of more regular  $\alpha\text{-TTP}_S$  structures with aggregation number  $\approx 23$ .

Overall, our data are in optimal agreement with the native gel electrophoresis experiment reported in ref 11. In particular, natively folded  $\alpha\text{-TTP}$ , which can oligomerize only through the A/B interface, showed the predominance of a monomeric form, with residual presence of low-weight aggregates (dimer, tetramer, Figure 3). On the contrary, after triggering aggregation by unfolding of the N-terminus, the proteins assembled into stable  $\alpha\text{-TTP}_S$ , which showed no tendency to disaggregate back into lighter oligomers in further incubation tests over a time window of 24 h.<sup>11</sup> Thus, the experimental condition of the real system would correspond to the region of  $T^* \approx 0.06\text{--}0.08$  in our toy system, where the folded state is mostly monomeric, whereas the partially unfolded state yields almost pure  $\alpha\text{-TTP}_S$  (Figure 6).

Apart from  $\alpha\text{-TTP}$  and  $\alpha\text{-TTP}_S$ ,  $\alpha\text{-TTP}_3$  is another oligomerization state for which, when  $IS_{C/D}$  is active, there exists a region of thermodynamical stability. The nature of  $\alpha\text{-TTP}_3$  as a true thermodynamic stable aggregate is confirmed by a diverging specific heat in correspondence of the boundary transition temperatures  $T^* \approx 0.11, 0.18$  (Figure 4, inset), which is a clear indication of the presence of two separate phase transitions. Experimentally, native  $\alpha\text{-TTP}$  rapidly evolves into  $\alpha\text{-TTP}_S$ , indicating that ambient thermodynamic conditions fall in the region of stability of the phase diagram for  $\alpha\text{-TTP}_S$ ; nonetheless,  $\alpha\text{-TTP}_3$  should be the dominating species in an intermediate region at higher temperature. The narrowness of such a region depends on the relative strength of  $IS_{A/B}$  and  $IS_{C/D}$ .

According to our results, point mutations at the surface of the protein that either weaken the  $IS_{A/B}$  interaction or, alternatively, strengthen the  $IS_{C/D}$  interaction would both yield an enlargement of the region of stability for  $\alpha\text{-TTP}_3$ . In our previous study, the analysis of the  $\alpha\text{-TTP}_S$  protein–protein interfaces at the 4-fold symmetry revealed hydrophobic contact areas that are mostly responsible for binding (Figure 7).<sup>11</sup> Introducing specific point mutations of these key residues can

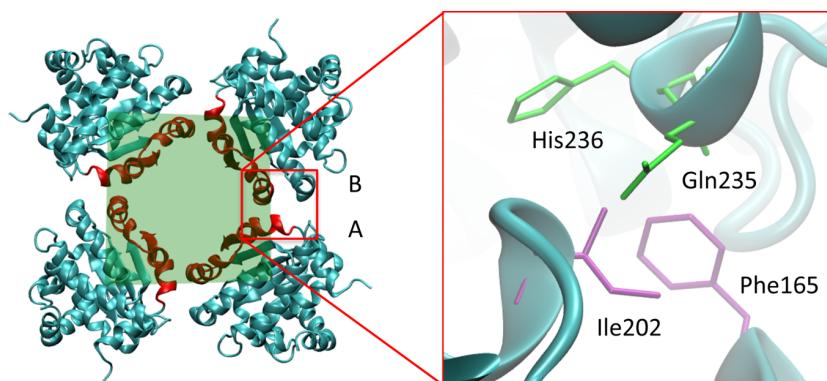


**Figure 6.** Top: Aggregation states of  $\alpha$ -TTP at different conditions. (A) Monomeric dispersion at  $T^* = 0.27$ ; (B) aggregation of low-weight oligomers at  $T^* = 0.04$ , for the system with only active  $IS_{A/B}$ ; (C) system with all active IS's at intermediate  $T^* = 0.13$ , where trimers begin to form; (D) same system at  $T^* = 0.07$ , characterized by formation of  $\alpha$ -TTP<sub>5</sub>. Bottom: phase diagram with dominating species at different conditions of temperature and folded state. The region between 0.06 and 0.11  $T^*$  corresponds to the experimentally observed behavior, with either properly folded monomers, or assembled  $\alpha$ -TTP<sub>5</sub>.

have stark impact on particle assembly. Any disruptive mutation, for example F165R, where a positively charged residue is introduced into the hydrophobic patch through site-directed mutagenesis, should weaken the cooperative effects of the  $IS_{A/B}$ . In this way, it should be possible to inactivate progressive dimerization, favoring instead  $\alpha$ -TTP<sub>3</sub> as the dominating species.

Interestingly, partial  $\alpha$ -TTP aggregation is not strictly bound to the unfolding of the N-terminus. Rather, low-weight aggregates of folded  $\alpha$ -TTP can be formed by binding through the natively solvent-exposed A/B interface. The transition from monomers to A/B dimers or tetramers is determined by the balance between the A/B binding energy and the dimerization

entropy loss. In solution, this balance is in favor of the monomeric species. Nonetheless, external factors like pre-organization of the monomers on a surface may favor the formation of such oligomers. Data by Arai and co-workers<sup>3</sup> reported that a mixture of  $\alpha$ -TTP,  $\alpha$ -tol, and lipid fractions containing different PIPs, such as PI(3,4)P2 or PI(4,5)P2, induced the formation of  $\alpha$ -TTP tetramers. In this study, also crystals of such ternary mixtures were analyzed by X-ray crystallography at 2.6 or 2.0 Å resolution. Superposition of the open (PDB: 1OIZ), closed (PDB: 1OIP), and PI(4,5)P2-bound (PDB: 3W68) structures revealed a semiopened mobile gate conformation in the ternary structure of  $\alpha$ -TTP.<sup>3</sup> It was also shown that such ternary complexes possess intermem-



**Figure 7.**  $\alpha$ -TTP interaction at the left: structure of the assembly of  $\alpha$ -TTP around the  $C_4$  symmetry axis of  $\alpha$ -TTP<sub>S</sub>. The structure interlocked via multiple IS<sub>A/B</sub>, topologically located at the ligand binding site region of  $\alpha$ -TTP. Right: the most relevant residues responsible for the binding at this interface are shown in licorice. The green and purple colors refer to amino acids belonging to two different  $\alpha$ -TTP units.

brane transfer activity *in vitro* when using donor or acceptor liposomes doped with PIPs. Neither  $\alpha$ -TTP<sub>3</sub> nor  $\alpha$ -TTP<sub>S</sub> aggregates were reported in this study. This is in accordance with our previous observations that aggregation into spherical particles occurs only when  $\alpha$ -tol is bound to  $\alpha$ -TTP with the mobile gate being in its fully closed state and subsequent unfolding of the N-terminus has unmasked the trimeric interaction interface of  $\alpha$ -TTP.

## CONCLUDING REMARKS

Our model provides a description of the thermodynamically stable aggregation states of  $\alpha$ -TTP that is consistent with experimental data. We report the existence of a metastable low-weight oligomerization state ( $\alpha$ -TTP<sub>3</sub>) that is key to the fast and regular assembly of  $\alpha$ -TTP<sub>S</sub>.

Expression of functional mutants with different assembling properties should be feasible by minimal modifications of the native sequence. Studies by Kortemme et al.<sup>32</sup> have shown in other systems that in general a single mutation is sufficient to redesign functional protein–protein interfaces and thus alter specificity. Interestingly, self-assembly into a similar spherical homo-multimer structure composed of 24 monomers has been reported in ferritin, an evolutionarily unrelated protein than  $\alpha$ -TTP.<sup>33</sup> In a very recent study, Dmochowski and co-workers showed that a homo-dimeric state represents a common intermediate during protein cage assembly of the 24-meric ferritin and that the dimer/24-mer balance can be experimentally altered by introducing single positive charges at sites along the dimer–dimer interface.<sup>34</sup> It is important to notice that even small variations in the binding affinity can have a very large impact on the assembling process, due to the cooperativity effects taking place during the assembling, as evidenced in the present study.

Functional low-weight oligomers, such as  $\alpha$ -TTP<sub>3</sub>, may play a crucial role in transcytosis through endothelial membranes. Our former transfection studies<sup>11</sup> showed that the transcytotic flux does not follow a diffusive regime, with larger  $\alpha$ -TTP<sub>S</sub> transfecting at a faster rate than smaller monomeric  $\alpha$ -TTP. In fact, the delay observed in the  $\alpha$ -TTP flux may imply that aggregation of a minimal unit larger than the monomer is required to activate the transport. Further studies introducing disruptive mutations at the trimeric or tetrameric interfaces could help the understanding of the assembly kinetics and thermodynamics at different physiological conditions. Most importantly, they may lead to the identification of the minimal

biologically active units that are active for the transfection through the endothelium, a crucial step toward the engineering of these protein for targeted drug delivery.

## AUTHOR INFORMATION

### Corresponding Authors

\*E-mail: achim.stocker@dcb.unibe.ch (A.S.).  
\*E-mail: michele.cascella@kjemi.uio.no (M.C.).

### ORCID

Michele Cascella: 0000-0003-2266-5399

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was supported by the Research Council of Norway (RCN) through the CoE Hylleraas Centre for Quantum Molecular Sciences Grant No. 262695 and by the Norwegian Supercomputing Program (NOTUR) (Grant No. NN4654K). H.B.K. received funding from the European Union Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement No. 704491.

## REFERENCES

- Traber, M. G.; Sokol, R. J.; Burton, G. W.; Ingold, K. U.; Papas, A. M.; Huffaker, J. E.; Kayden, H. J. Impaired ability of patients with familial isolated vitamin E deficiency to incorporate  $\alpha$ -tocopherol into lipoproteins secreted by the liver. *J. Clin. Invest.* **1990**, *85*, 397–407.
- Traber, M. G.; Burton, G. W.; Hamilton, R. L. Vitamin E trafficking. *Ann. N. Y. Acad. Sci.* **2004**, *1031*, 1–12.
- Kono, N.; Ohto, U.; Hiramatsu, T.; Urabe, M.; Uchida, Y.; Satow, Y.; Arai, H. Impaired  $\alpha$ -TTP-PIP<sub>S</sub> interaction underlies familial vitamin E deficiency. *Science* **2013**, *340*, 1106–1110.
- Meier, R.; Tomizaki, T.; Schulze-Briese, C.; Baumann, U.; Stocker, A. The molecular basis of vitamin E retention: Structure of human  $\alpha$ -tocopherol transfer protein. *J. Mol. Biol.* **2003**, *331*, 725–734.
- Min, K. C.; Kovall, R. A.; Hendrickson, W. A. Crystal structure of human alpha-tocopherol transfer protein bound to its ligand: implications for ataxia with vitamin E deficiency. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 14713–14718.
- Sha, B.; Phillips, S. E.; Bankaitis, V. A.; Luo, M. Crystal structure of the *Saccharomyces cerevisiae* phosphatidylinositol-transfer protein. *Nature* **1998**, *391*, 506–510.
- Stocker, A.; Tomizaki, T.; Schulze-Briese, C.; Baumann, U. Crystal structure of the human supernatant protein factor. *Structure* **2002**, *10*, 1533–1540.

- (8) He, X.; Lobsiger, J.; Stocker, A. Bothnia dystrophy is caused by domino-like rearrangements in cellular retinaldehyde-binding protein mutant R234W. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 18545–18550.
- (9) Stocker, A. Molecular mechanisms of vitamin E transport. *Ann. N. Y. Acad. Sci.* **2004**, *1031*, 44–59.
- (10) Chung, S.; Ghelfi, M.; Atkinson, J.; Parker, R.; Qian, J.; Carlin, C.; Manor, D. Vitamin E and phosphoinositides regulate the intracellular localization of the hepatic  $\alpha$ -tocopherol transfer protein. *J. Biol. Chem.* **2016**, *291*, 17028–17039.
- (11) Aeschimann, W.; Staats, S.; Kammer, S.; Olieric, N.; Jeckelmann, J. M.; Fotiadis, D.; Netscher, T.; Rimbach, G.; Cascella, M.; Stocker, A. Self-assembled  $\alpha$ -Tocopherol Transfer Protein Nanoparticles Promote Vitamin E Delivery Across an Endothelial Barrier. *Sci. Rep.* **2017**, *7*, No. 4970.
- (12) Saraiva, C.; Praça, C.; Ferreira, R.; Santos, T.; Ferreira, L.; Bernardino, L. Nanoparticle-mediated brain drug delivery: Overcoming blood-brain barrier to treat neurodegenerative diseases. *J. Controlled Release* **2016**, *235*, 34–47.
- (13) Deli, M. A.; Ábrahám, C. S.; Kataoka, Y.; Niwa, M. Permeability studies on *in vitro* blood-brain barrier models: Physiology, pathology, and pharmacology. *Cell. Mol. Neurobiol.* **2005**, *25*, 59–127.
- (14) Zhang, Z.; Glotzer, S. C. Self-assembly of patchy particles. *Nano Lett.* **2004**, *4*, 1407–1413.
- (15) Sciotino, F.; Giacometti, A.; Pastore, G. Phase Diagram of Janus Particles. *Phys. Rev. Lett.* **2009**, *103*, No. 237801.
- (16) Kraft, D. J.; Ni, R.; Smalenburg, F.; Hermes, M.; Yoon, K.; Weitz, D. A.; van Blaaderen, A.; Groenewold, J.; Dijkstra, M.; Kegel, W. K. Surface roughness directed self-assembly of patchy particles into colloidal micelles. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 10787–10792.
- (17) Wilber, A. W.; Doye, J. P. K.; Louis, A. A.; Noya, E. G.; Miller, M. A.; Wong, P. Reversible self-assembly of patchy particles into monodisperse icosahedral clusters. *J. Chem. Phys.* **2007**, *127*, No. 085106.
- (18) Carlsson, F.; Linse, P.; Malmsten, M. Monte Carlo simulations of polyelectrolyte-protein complexation. *J. Phys. Chem. B* **2001**, *105*, 9040–9049.
- (19) Fantoni, R.; Gazzillo, D.; Giacometti, A.; Miller, M. A.; Pastore, G. Patchy sticky hard spheres: Analytical study and Monte Carlo simulations. *J. Chem. Phys.* **2007**, *127*, No. 234507.
- (20) Villar, G.; Wilber, A. W.; Williamson, A. J.; Thiara, P.; Doye, J. P.; Louis, A. A.; Jochum, M. N.; Lewis, A. C.; Levy, E. D. Self-assembly and evolution of homomeric protein complexes. *Phys. Rev. Lett.* **2009**, *102*, No. 118106.
- (21) Liu, H.; Kumar, S. K.; Douglas, J. F. Self-Assembly-Induced Protein Crystallization. *Phys. Rev. Lett.* **2009**, *103*, No. 018101.
- (22) Li, Y.; Shi, T.; An, L.; Huang, Q. Monte Carlo Simulation on Complex Formation of Proteins and Polysaccharides. *J. Phys. Chem. B* **2012**, *116*, 3045–3053.
- (23) Li, W.; Persson, B. A.; Morin, M.; Behrens, M. A.; Lund, M.; Zackrisson Oskolkova, M. Charge-induced patchy attractions between proteins. *J. Phys. Chem. B* **2015**, *119*, 503–508.
- (24) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037–10041.
- (25) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple AMBER force fields and development of improved protein backbone parameters. *Proteins* **2006**, *65*, 712–725.
- (26) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **1953**, *21*, 1087–1092.
- (27) Frenkel, D.; Smit, B. *Understanding Molecular Simulation*, 2nd ed.; Academic Press, Inc.: Orlando, 2001.
- (28) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: New York, 1987.
- (29) Wu, D.; Chandler, D.; Smit, B. Electrostatic analogy for surfactant assemblies. *J. Phys. Chem.* **1992**, *96*, 4077–4083.
- (30) Miller, M. A.; Amon, L. M.; Reinhardt, W. P. Should one adjust the maximum step size in a Metropolis Monte Carlo simulation? *Chem. Phys. Lett.* **2000**, *331*, 278–284.
- (31) Humphrey, W.; Dalke, A.; Schulz, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (32) Kortemme, T.; Joachimiak, L. A.; Bullock, A. N.; Schuler, A. D.; Stoddard, B. L.; Baker, D. Computational redesign of protein-protein interaction specificity. *Nat. Struct. Mol. Biol.* **2004**, *11*, 371–379.
- (33) Lawson, D. M.; Artymiuk, P. J.; Yewdall, S. J.; Smith, J. M. A.; Livingstone, J. C.; Treffry, A.; Luzzago, A.; Levi, S.; Arosio, P.; Cesareni, G.; Thomas, C. D.; Shaw, W. V.; Harrison, P. M. Solving the structure of human H ferritin by genetically engineering intermolecular crystal contacts. *Nature* **1991**, *349*, 541–544.
- (34) Pulsipher, K. W.; Villegas, J. A.; Roose, B. W.; Hicks, T. L.; Yoon, J.; Saven, J. G.; Dmochowski, I. J. Thermophilic ferritin 24mer assembly and nanoparticle encapsulation modulated by interdimer electrostatic repulsion. *Biochemistry* **2017**, *56*, 3596–3606.