# Structural Driving Forces and Their Spectroscopic Signatures: From Protonation Dynamics to Amyloid Formation

## Dissertation

zur Erlangung des

Doktorgrades der Naturwissenschaften (Dr. rer. nat.)

der
Naturwissenschaftlichen Fakultät II
Chemie, Physik und Mathematik
der Martin-Luther-Universität
Halle-Wittenberg

vorgelegt
von Herrn Felix Hoffmann

geb. am 19. November 1987 in Essen

**First Reviewer**

Prof. Dr. Daniel Sebastiani
MLU Halle-Wittenberg
Institut für Physikalische Chemie
Von-Dankelmann-Platz 4
06120 Halle (Saale)
Germany

**Second Reviewer**

Prof. Dr. Petra Imhof
Freie Universität Berlin
Fachbereich Physik
Arnimallee 14
14195 Berlin
Germany

Day of public defense: April 23, 2018

This dissertation has been created under the supervison of Prof. Daniel Sebastiani at the Institute of Chemistry of the Marin-Luther University Halle-Wittenberg.

This work has in parts been funded by a Kekulé Fellowship of the Fonds der Chemischen Industrie.

# Acknowledgements

I would like to acknowledge my advisor Prof. Dr. Daniel Sebastiani for his support throughout this work and for the many valuable discussions. Further, I would like to thank the members of the Sebastiani group for their support, especially in the initial stages of my work in Halle. In particular, I would like to express my gratitude to Somayeh Khazaei, Laura Scarbath-Evers, Dr. Christoph Allolio, Dr. Arne Scherrer, and Dr. Hossam Elgabarty. I am especially indebted to Dr. Gül Bekçioğlu-Neff for her great support and friendship throughout the years.

Further, I would like to thank my scientific collaborators for their interesting experiments, numerous fruitful discussions, conceptual contributions, and personal advise. For works on transient infrared experiments, Dr. Maria Ekimova and Dr. Erik T. J. Nibbering. For solid-state NMR experiments on amyloid β fibrils, Dr. Juliane Adler and Prof. Dr. Daniel Huster. For guidance and support during my time as a research scholar at The Ohio State University, Prof. Dr. Rafael Brüschweiler and Dr. Da-Wei Li.

A very special thanks goes to my family, especially to my parents and sister for giving me motivation and support, without which this work would have not been possible.

# Table of Contents

**1. Introduction** **9**

**2. Theoretical Framework** **15**
2.1. Introduction to Molecular Dynamics Simulations . . . . . . . . . . . . . 15
2.1.1. Numerical Integration of the Equations of Motion . . . . . . . . 15
2.1.2. Temperature and Pressure Coupling . . . . . . . . . . . . . . . . 17
2.1.3. Treatment of Interatomic Interactions under Periodic Boundary Conditions . . . . . . . . . . . . . . . . . . . . . . . . . . . 21
2.1.4. *Ab Initio* Molecular Dynamics Simulations . . . . . . . . . . . 25
2.1.5. Classical Molecular Dynamics Simulations . . . . . . . . . . . . 27
2.1.6. Hybrid Quantum/Classical Molecular Dynamics Simulations . . 30
2.2. Density Functional Theory . . . . . . . . . . . . . . . . . . . . . . . . 33
2.2.1. The Hohenberg-Kohn Theorems . . . . . . . . . . . . . . . . . . 33
2.2.2. The Kohn-Sham Approach . . . . . . . . . . . . . . . . . . . . . 34
2.2.3. Approximations to the Exchange-Correlation Functional . . . . 35
2.2.4. The LCAO Ansatz . . . . . . . . . . . . . . . . . . . . . . . . . 37
2.2.5. Basis Sets . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 38
2.2.6. Pseudopotentials . . . . . . . . . . . . . . . . . . . . . . . . . . 40
2.3. Spectroscopic Parameters from First Principles Calculations . . . . . . 42
2.3.1. Calculation of NMR Chemical Shifts . . . . . . . . . . . . . . . 42
2.3.2. Calculation of Electronically Excited States . . . . . . . . . . . 45
2.3.3. Calculation of Infrared Spectra . . . . . . . . . . . . . . . . . . 48

**3. Overview over the Published Papers** **51**
3.1. Protonation Dynamics Triggered by Electronic Excitation of Photoacids and -Bases in Aqueous and Methanol Solution . . . . . . . . . . . 51
3.2. Prediction of NMR Chemical Shifts of Metabolite Molecules in Aqueous Solution . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 53
3.3. Impact of the Phe$_{19}$-to-Lys$_{19}$ Mutation in Aβ(1-40) on the Fibril Structure . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 55

**4. Conclusion** **59**

**5. References** **63**

**A. Publications** **75**
A.1. Solvation-Dependent Latency of Photoacid Dissociation and Transient IR Signatures of Protonation Dynamics . . . . . . . . . . . . . . . . . 79

A.2. Combined Experimental and Theoretical Study of the Transient IR Spectroscopy of 7-Hydroxyquinoline in the First Electronically Excited Singlet State . . . . 89
A.3. Improved Quantum Chemical NMR Chemical Shift Prediction of Metabolites in Aqueous Solution toward the Validation of Unknowns . . . 103
A.4. Perturbation of the F19-L34 Contact in Amyloid β (1-40) Fibrils Induces Only Local Structural Changes but Abolishes Cytotoxicity . . . . 113

**B. Academic curriculum vitae** **121**

# 1. Introduction

The competition between hydrophobic and hydrophilic interactions is a major structural driving force in many molecular systems. Although effective at the molecular level, they account for system properties over a wide range of time and length scales, and consequently the investigation of their complex interplay is crucial for an in-depth understanding of many processes. For example on fast time scales, the protonation dynamics in aqueous solution of alcohols is governed by picosecond-fluctuations of the hydrogen bond network that competes with hydrophobic interactions between unpolar sites. Similarly in much larger biochemical systems, such as peptides or proteins, folding and aggregation are driven by hydrophobic and hydrophilic interactions between amino acid side chains of different polarity and their surrounding water molecules. One of the prime methodologies to computationally study the relations between these basic interaction patterns and the resulting structural driving forces is molecular dynamics (MD) simulations. MD simulations allow for the direct computation of thermodynamic and spectroscopic quantities *via* the laws of statistical mechanics.[1–4] In this manner, the technique enables the connection between macroscopic features of the system and their underlying atomistic and molecular structure and dynamics. Moreover, the prediction of spectroscopic fingerprints from MD simulations is crucial for validation with experimental data and offers complementary information for a further in-depth understanding. The main idea of MD simulations is to describe the time evolution of a system in a well-defined thermodynamic ensemble and to compute ensemble quantities as time averages over the resulting MD trajectory. Due to the advancement in computer hardware and algorithmic developments, MD simulations are now a well-established methodology for the description of condensed phase systems and allow for the treatment of systems with an ever larger growing number of atoms.[5–8]

In the current work, two different flavors of MD simulations have been employed, namely *ab initio* (AIMD) and force field MD simulations. In both cases, the nuclei are treated as classical point particles whose time evolution is governed by Newton's equation of motion. They differ, however, in the treatment of the potential acting on the nuclei due the electronic degrees of freedom. In the case of AIMD simulations, this potential is evaluated "on-the-fly" at every integration time step by approximately solving the electronic structure problem. Conversely in force field MD simulations, the potential is computed from analytical potential functions, so-called force fields, parametrized *a priori* for atoms in a given environment, *e.g.* for carbonyl oxygens as part of an amino acid. Consequently, AIMD simulations allow for a more general and accurate description of interatomic interactions on a quantum-mechanical footing at the costs of an increased numerical complexity. There are many cases where a quantum-mechanical treatment of electrons is crucial. This is most obviously the case if chemical reactions occur, for example, during proton transfer (PT) in aqueous solution.[9–12] Due to their predefined functional form, standard protein force fields[13–15] fail to describe the breaking and formation of covalent bonds. Moreover, other properties, like charge transfer and electronic polarization are only implicitly considered in the parameter sets. By contrast, AIMD simulations allow for changes of chemical bonds and perform well for the description of electronic polarization which in turn is needed for a realistic representation of hydrogen bonding. This is especially crucial for PT reactions because proton transport in aqueous media is governed by complex hydrogen bonding rearrangements in the solvation shell of the proton transferring intermediates, such as the Eigen and Zundel complexes.[9,11,12,16–20] However, AIMD simulations quickly become prohibitive for biological systems, which easily involve hundreds of thousands of atoms with relevant time scales in the nanosecond regime and beyond. As a result, there has been a long-standing effort to parametrize accurate force field potentials for biological systems, such as peptides and proteins. To date, force fields, like Amber ff99SB[13] and CHARMM36,[14] are commonly used in computational biology for globular proteins and have been demonstrated to be in good agreement with experimental structural and dynamics data.[21]

This thesis covers a series of computational investigations that address the complex interplay between hydrophobic and hydrophilic interactions and their spectroscopic signatures. In particular, this work encompasses the following three projects:

I. The elucidation of protonation dynamics in water and water-methanol mixtures triggered by electronic excitation of photoacids and -bases

II. The impact of the $\mathrm{Phe}_{19}$-to-$\mathrm{Lys}_{19}$ mutation in $\mathrm{A}\beta$(1-40) on the fibril structure.

III. The efficient prediction of NMR chemical shifts of metabolite molecules in aqueous solution.

The first project deals with the investigation of PT reactions in water and water-methanol mixtures. PT reactions are of paramount importance for a large number of chemical and biological processes, such as the autoionization of water,[11] the von Grotthuss mechanism,[9,22–26] the proton conduction through cell membranes[27] as well as for technical applications, like fuel cells.[28–30] To date, still many details of PT reactions remain elusive. One reason is that even for a seemingly simple system like water, PT is fairly complicated. For instance, it has been known for a long time that, depending on the pH value of the solution, one either observes the migration of a positive, *i.e.* a hydronium ion, or a negative charge defect, *i.e.* an hydroxide ion, where the diffusion constants are much larger than those of ions with comparable size.[31,32] The advent of AIMD simulations has been a milestone to the field, since the method has enabled the elucidation of PT mechanisms at the atomic level. One of the key results of these early AIMD studies was that both step-wise and concerted transfer mechanisms coexist for PT, challenging a sequential Grotthuss picture.[33] Clearly, the situation gets more involved if co-solvents, like ions, alcohols, acids, or bases are added to the solution. A promising means to investigate these systems are molecules with photoacidic and -basic functionalities.[17–20,34–36] An example of such a system is the bifunctional chromophore 7-hydroxyquinoline (7HQ), shown in Figure 1.1.[37–39] If the molecule is electronically excited by an UV pulse to the first singlet state ($S_1$),

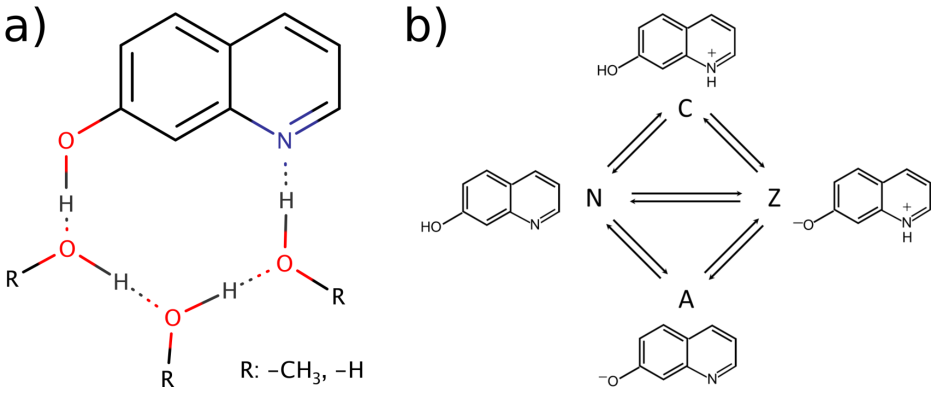


**Figure 1.1.** – a) 7-Hydroxyquinoline stabilizing a chain of three water/methanol molecules. b) Various protonation states of 7HQ: Neutral (N), cationic (C), anionic (N) and zwitterionic (Z) state.

the equilibrium in Figure 1.1b) shifts from the neutral (N) to the zwitterionic (Z) state, where the reaction can proceed either *via* the cationic or anionic intermediate. Experimentally, this tautomerization can be traced by recording infrared fingerprint absorptions of N and Z as a function of the pump-probe delay in transient infrared spectroscopy. Complementary, molecular solvent configurations are accessible from AIMD simulations. For instance, 7HQ can stabilize hydrogen bonded configurations of three water molecules over times one order of magnitude longer than the time for a concerted reaction from N to Z.[40] The reminiscence of these configurations to water structure in proton channels which regulate PT through cell membranes, makes it an interesting model system to study ultrafast PT reactions.

The second project focusses on the structural response of amyloid β (1-40) peptide fibrils to the replacement of a hydrophobic with a charged amino acid in its hydrophobic core. Amyloid fibrils are the hallmark of several severe human diseases associated with the misfolding of peptides, where the list includes diseases like Huntington's and Alzheimer's disease as well as type II diabetes.[41] A common characteristic of the disease causing peptides is their aggregation into so-called cross-β structures.[42] As shown in Figure 1.2, this structural motive exhibits lateral β-sheets and a dense side chain packing, where two or three cross-β structures in turn assemble to dimers or trimers to form several micrometer long unbranched fibrils. Astonishingly, peptides with no apparent sequence similarity and function are able to adopt this aggregation state. Based on this structural similarity it seems obvious that the formation of amyloid fibrils follows generic principles and is governed by distinct inter- and intramolecular forces with apparent relations to synthetic polymers.[43] For a further understanding of these principles, amyloid fibrils with slightly different amino acid sequences have been previously compared regarding their local structure and dynamics.[44] However, a detailed atomistic elucidation of these structures is inherently difficult as high-resolution techniques such as X-ray crystallography cannot be applied due to a lack of long-range order. In addition, fibrils exhibit polymorphism, *i.e.* their structures vary depending on particular growing conditions, which is a consequence of their rough free energy landscape.[45] Nevertheless, structural models of some peptides

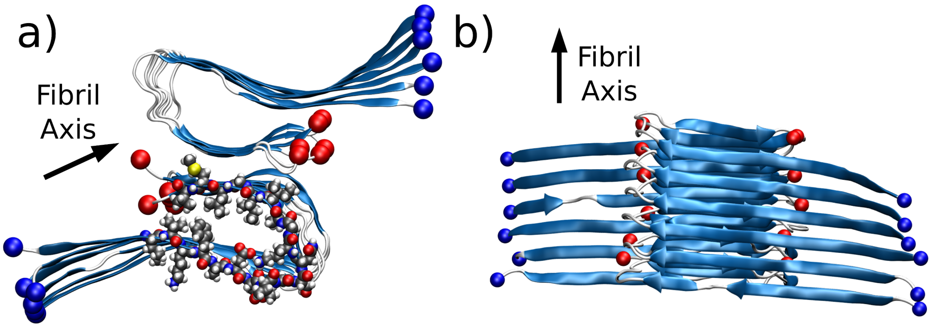


**Figure 1.2.** – Schematic depiction of two cross-β structures in a F19K amyloid β (1-40) fibril, viewed from a) side and b) top. N- and C-termini are marked with blue and red spheres, respectively. β-Sheets are represented light blue and side chains in the hydrophobic core are indicated by small spheres.

are nowadays available from solid-state NMR studies or cryo-electron microscopy in conjunction with molecular modeling.[46–53] Certainly the by far best studied system in this context is amyloid β (Aβ). By now, many studies have revealed structural and dynamics information for fibrils and transient oligomeric and protofibrillar species.[54,55] The underlying physicochemical principles governing the fibrillation process can be systematically investigated by analyzing the effects of non-physiological point mutations.[44,56] For instance, the replacement of hydrophobic residues with polar or charged amino acids and the accompanying weakening of hydrophobic interactions within the fibril core allows for the assessment of critical interaction patterns as well as for the investigation of the structural response. In this context, Adler *et al.* recently examined the role played by the hydrophobic contact between phenylalanine 19 and leucine 34 for fibril structure and dynamics in a mutagenesis study of eight different Aβ(1-40) variants.[44] The surprising result was that all but one mutation lead to the formation of fibrils with only very local effects for secondary structure and backbone dynamics. Even more striking was the observation of Das *et al.* that all mutations completely abolished the toxicity of the fibrils, highlighting the crucial role played by the hydrophobic contact $\text{Phe}_{19}$-$\text{Leu}_{34}$ for the mechanism that eventually leads to cell death.[57] It is therefore compelling to further understand the structural changes in the fibril that result from these point mutations. In this thesis, the $\text{Phe}_{19}$-to-$\text{Lys}_{19}$ mutation in amyloid β (1-40) was studied. For that purpose, a number of atomistic models of the amyloid fibril have been created using homology modeling and force field MD simulations. For verification, structural key features of the models have been compared to solid-state NMR measurements.

The third project deals with the efficient prediction of NMR chemical shifts of metabolite molecules in aqueous solution. The analysis of complex metabolic mixtures by means of NMR spectroscopy has become increasingly popular over the past decade.[58–61] This is in part due to the ability of NMR spectroscopy to provide information on multiple metabolites in the same sample and at the same time to allow for large sample cohorts that are amenable to statistical analysis. In this way, new insights into the state of biological systems as well as into metabolite pathways are possible.[62–65] However, the identification of unknown metabolites, *i.e.* those that have never been identified before and are therefore not part of any database, remains still a major challenge. In this regard, semi-empirical chemical shift predictors are useful to efficiently compare chemical shift information with a large number of candidate structures, but their accuracy is determined by the quality and size of the underlying database.[66,67] As a consequence, one often encounters excellent prediction results for structures which are part of the database, whereas structures that have not been considered in the fitting procedure are usually predicted with a significantly lower accuracy. Therefore, predictors that are less dependent on, or in the best case independent of databases potentially allow for a more balanced prediction accuracy in terms of known and unknown structures. In this context, quantum-chemical based methods are a promising alternative. However, their computational cost is in most cases several orders of magnitude larger and requires the use of high performance computing resources, when a large number of predictions needs to be carried out. As illustrated in Figure 1.3, a widely-used approach to keep the computational costs within a limit is to calculate NMR shielding constants only for a relatively small number of structural minima and to account for solvent effects implicitly by means of a polarizable continuum model. Regarding the quantum-chemical method, density functional theory (DFT) has been proven to be particularly useful in providing both sufficient accuracy and numerical efficiency, allowing for NMR shielding calculations on a routine basis.[68–74] The calculated shielding constants can be converted to chemical shifts in three ways: Firstly, by subtracting them from the shielding constant of an internal reference, such as tetramethylsilane (TMS) or 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS).

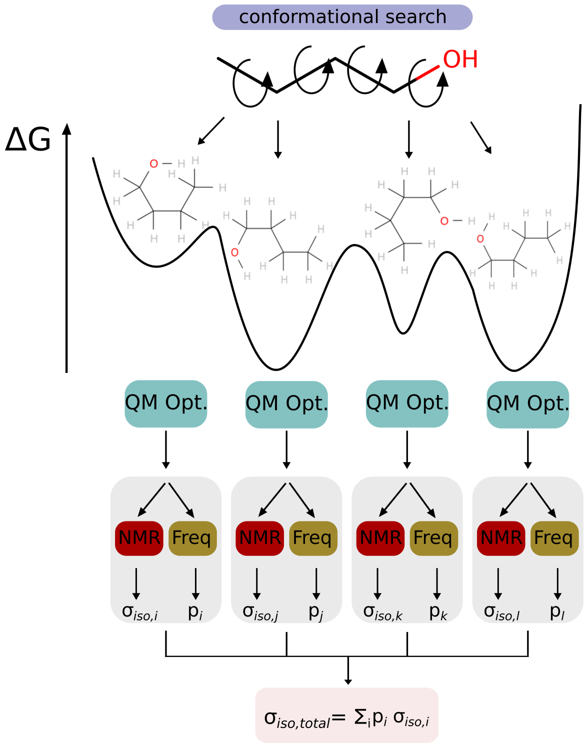


**Figure 1.3.** – Flow chart illustrating the calculation of NMR shielding constants from a limited number of low energy conformers.

Secondly, by intermediate references or multi-standards, or, thirdly, by means of (linear) regression. Whereas the first two approaches usually rely on a fortuitous error cancellation between shielding constants calculated for the query and reference atom, (linear) regression allows to correct for systematic errors as a function of the value of isotropic shielding constant, $\sigma_{\mathrm{iso}}$. Several studies report excellent agreements with experimental shifts when using DFT together with linear regression and report fitting parameters for several exchange-correlation/basis set combinations.[70,75–77] However, most of them were conducted in organic solvents, such as chloroform on a relatively small number of molecules. While these results are helpful in the field of general organic chemistry, they are somewhat less practical for metabolomics studies, where NMR chemical shifts are usually measured in aqueous solution. Naturally, water is posing additional challenges to the prediction of NMR chemical shifts of metabolites especially when combined with implicit solvent models. For example, chemical shifts of atoms in close proximity to hydrogen bond donating and/or accepting sites will likely have different systematic errors than shifts of those atoms that are further away.[78] Moreover, molecules that allow for intramolecular hydrogen bonding often exhibit large chemical shift deviations, if implicit solvent models are employed. For this reason, conformational search and geometry optimization in implicit solvent favor geometries that are stabilized by direct hydrogen bonds which often lead to biased conformational ensembles. As a result, atoms in different solvent environments will have different systematic errors. Here, a molecular motif-specific linear regression approach was employed to convert shielding values into chemical shifts. The comparison with experimental data demonstrated a very good performance of our method, where it performed significantly better than global regression schemes and empirical chemical shift predictors.

This cumulative thesis is structured as follows: In the second chapter a brief introduction to the methodological background will be presented. Subsequently, an overview over the published papers will be given, followed by a conclusion highlighting the key results of this thesis. The scientific papers are given in the appendix along with details on the author's contribution.

# 2. Theoretical Framework

## 2.1. Introduction to Molecular Dynamics Simulations

Molecular dynamics simulations is a method to calculate the time evolution of a collection of interacting particles by numerically solving their equations of motions. Macroscopic observables, such as as temperature, pressure, and spectroscopic parameters can then be readily computed from time averages over the particle's trajectories. This way, relations between the microscopic structure of matter and its macroscopic properties become accessible. Molecular dynamics simulations are applied to a wide range of systems, such as crystals, amorphous solids, and liquids, and played a central role for the studies outlined in this thesis. In this chapter, a brief overview on some of the numerical details of molecular dynamics simulations will be given. To this end, firstly a number of schemes that can be used to integrate the classical equations of motion will be introduced, and, next, approaches will be presented to conduct simulations in the canonical (NVT) and isothermal-isobaric (NPT) ensembles. At the end of this chapter, two flavors of simulations, namely classical and *ab initio* molecular dynamics simulations, will be described in more detail.

Since molecular dynamics simulations calculate the propagation of a system in time, it is natural to start with the equations of motion. If the quantum nature of the particles can be ignored, as it is the case for almost all nuclei at ambient conditions, these equations are simply given by Newton's second law

$$m_I \cdot \frac{\partial^2 \mathbf{R}_I(t)}{\partial t^2} = -\nabla_I V(\mathbf{R}_1(t), \ldots, \mathbf{R}_I(t), \ldots, \mathbf{R}_N(t)) \ . \tag{2.1}$$

Here, $m_I$ and $\mathbf{R}_I(t)$ denote the mass of particle $I$ and its cartesian coordinate at time $t$, respectively, and $V$ the potential energy as a function of the positions of all $N$ particles in the system. In classical and *ab initio* molecular dynamics simulations, the particles that are propagated are the nuclei, where the fast motion of the electrons has been 'averaged-out' and is only implicitly considered in the potential energy function $V$. However, this potential is obtained differently in the two methods: In classical simulations, the potential energy function is approximated *a priori* by a set of classical potential functions, whose parameters have been obtained from fits to experimental and quantum-chemistry data, whereas in *ab initio* molecular dynamics simulations the potential acting on the nuclei is calculated 'on-the-fly' by quantum-chemical methods, usually at the density functional theory level.

### 2.1.1. Numerical Integration of the Equations of Motion

The standard procedure to solve a set of coupled second-order partial differential equations, like Equation (2.1), is by finite-difference integration. As a matter of course, there exists a plethora of algorithms in numerical mathematics that could in principle be used.[79] However, not all algorithms are equally-well suited.[1,3] This is because some of them fail to fulfill a number of fundamental laws of physics, i.e. they are not time-reversible and/or they do not conform to the conservation of energy and momentum. Moreover, from the practical point of view, a 'good' integrator should allow for large time-steps and not require more than one force evaluation per step, which is usually the most expensive part, in terms of computational effort, in the algorithm. On the other hand, it is hopeless to expect that any numerical integration scheme will exactly reproduce the true classical trajectory for an extended period of simulation time. In fact, any two initially very close trajectories will eventually diverge exponentially with time and lead to uncorrelated mechanic quantities already within the first hundred time steps when using standard integration schemes.[1]

This in the first place disheartening observation turns out not to be crucial at all because all that is required for the accurate evaluation of thermodynamic quantities is that the simulation reproduces the thermodynamic ensemble of interest. In the case of the microcanonical ensemble, i.e. when particle number, volume and energy are fixed, this means that a good integrator must conserve the energy (and momentum) over the whole simulation time to guarantee meaningful results, i.e. it must be long-time stable. Note that short-term fluctuations of the total energy are usually acceptable as long as the integrator solves the equations of motions with high precision over the correlation times of interest.[1] The integration of the equations of motion was already a fairly standard task in the advent of molecular dynamics simulations and the algorithms that are nowadays used are essentially unchanged. The two most widely employed approaches are the leap-frog[80] and velocity Verlet[81] algorithm, although other algorithms like predictor-corrector algorithms of the Gear family are more accurate at the cost of larger memory requirements.[82,83]

### The Störmer-Verlet Method

One of the most well-known integrator was firstly developed by Störmer[83] and later adopted by Verlet.[84] It is readily derived from the addition of two Taylor expansions of the positions to third order around their initial values in forward and backward time direction, respectively. The resulting equation for propagating the position of particle $I$ in time then reads

$$\mathbf{r}_I(t+\delta t) = 2\mathbf{r}_I(t) - \mathbf{r}_I(t-\delta t) + \delta t^2 \mathbf{a}_I(t) \ . \tag{2.2}$$

Due to the fact that odd orders in $\delta t$ cancel each other in the derivation of Equation (2.2), the error in the positions is only of order $\delta t^4$. The Störmer-Verlet method does explicitly include the calculation of velocities, however, e.g. for estimating the kinetic energy at time $t$, they may be calculated as

$$\mathbf{v}_I(t) = \frac{\mathbf{r}_I(t+\delta t) - \mathbf{r}_I(t-\delta t)}{2\delta t} \ . \tag{2.3}$$

The error of the velocities as calculated from Equation (2.3) is of order $\delta t^2$ but more accurate estimates can be obtained by considering a larger number of position steps, which naturally raises the memory requirements.

### The Leap-Frog Method

A numerical disadvantage of the Störmer-Verlet method is that a small value ($\mathcal{O}(\delta t^2)$) is added to a difference of large values ($\mathcal{O}(\delta t^0)$). This may result in numerical inaccuracies that can readily be avoided by the leap-frog method,[80] which takes the form

$$\mathbf{r}_I(t+\delta t) = \mathbf{r}_I(t) + \delta t \mathbf{v}_I\left(t+\frac{1}{2}\delta t\right) \tag{2.4}$$

$$\mathbf{v}_I\left(t+\frac{1}{2}\delta t\right) = \mathbf{v}_I\left(t-\frac{1}{2}\delta t\right) + \delta t \mathbf{a}_I(t) \ . \tag{2.5}$$

The leap-frog algorithm is algebraically equivalent to the Störmer-Verlet method but does not require the summation of large and small values and thus is numerically advantageous.

### The Velocity Verlet Method

One may argue that the leap-frog algorithm still has the disadvantage that positions and velocities are not computed at the same time step. This is achieved by a slightly different algebraic formulation that leads to the velocity Verlet algorithm:[81]

$$\mathbf{r}_I(t+\delta t) = \mathbf{r}_I(t) + \delta t \mathbf{v}_I(t) + \frac{1}{2}\delta t^2 \mathbf{a}_I(t) \tag{2.6}$$

$$\mathbf{v}_I(t+\delta t) = \mathbf{v}_I(t) + \frac{1}{2}\delta t \left[\mathbf{a}_I(t) + \mathbf{a}_I(t+\delta t)\right] \ . \tag{2.7}$$

This algorithm does require two steps: First, the velocities at time $t+\frac{1}{2}\delta t$ are computed from the current velocities and accelerations. Next, the accelerations at time $t+\delta t$ are computed and the velocity propagation is completed:

$$\mathbf{v}_I(t+\delta t) = \mathbf{v}_I\left(t+\frac{1}{2}\delta t\right) + \frac{1}{2}\delta t \mathbf{a}_I(t+\delta t) \tag{2.8}$$

To conclude, all algorithms outlined above lead to the same global errors and produce essentially identical position trajectories. Moreover, these methods have the merit of being less sensitive to the use of a larger time step in terms of the global error of the total energy than higher order predictor-corrector schemes[1,82] and are therefore nowadays used almost exclusively in molecular dynamics simulations.

## 2.1.2. Temperature and Pressure Coupling

The numerical integration of the classical equations of motion (ideally) conserves the total Hamiltonian and therefore leads to the microcanonical ensemble in which the particle number, volume and total energy are conserved (NVE). In many cases, however, one might wish to compare simulated results to thermodynamic quantities that are associated to other ensembles. For instance, many experiments are carried out in the isobaric-isothermal (NPT) ensemble, which requires the coupling of the simulation system to an external temperature bath and piston to maintain constant temperature and pressure, respectively. Several schemes for temperature and pressure coupling have been suggested in the literature[85–92] and have been extensively reviewed.[1–3] Here, only a brief introduction to the methods that have been used in the molecular dynamics simulations, which lead to the results outlined in this thesis, is presented. It should also be mentioned that the maintenance of temperature and pressure during a molecular dynamics simulation leads to somewhat non-physical dynamics, especially in the case of a strong coupling to the bath. Strictly speaking, in molecular dynamics simulations thermodynamic quantities are sampled, which are associated to a given thermodynamic ensemble without reproducing the system's 'exact' time evolution.

### Canonical Sampling Velocity-Rescaling Thermostat

Certainly one of the most elegant ways to maintain a constant temperature during a molecular dynamics simulation is by the *canonical sampling through velocity rescaling* (CSVR) method introduced by Bussi and coworkers.[90] Their approach is based on the idea of a global scaling of the velocities by a factor $\alpha$ at each time step. If this factor is obtained from a target kinetic energy, $K_t$, drawn from the canonical equilibrium distribution of the kinetic energy, and the instantaneous kinetic energy $K$ according to

$$\alpha = \sqrt{\frac{K_T}{K}}\ , \tag{2.9}$$

the resulting dynamics will sample a canonical (NVT) ensemble.

In fact, in the CSVR approach this scaling is 'distributed' over several time steps so as to avoid large disturbances of the dynamics that would be introduced by drawing $K_t$ directly from the canonical probability distribution. To this end, $K_t$ itself is smoothly incremented by $dK$, which depends on the previous value of $K$. The full expression for $dK$ takes the form

$$dK = (\overline{K} - K)\frac{dt}{\tau} + 2\sqrt{\frac{K\overline{K}}{N_f}\frac{dW}{\sqrt{\tau}}}\ , \tag{2.10}$$

where $\overline{K}$ denotes the kinetic energy that corresponds to the target temperature, $dt$ is the integration time step, $\tau$ is a coupling parameter in units of time, which determines how fast the target temperature will be reached. Further, $N_f$ stands for the number of degrees of freedom and $dW$ denotes Wiener noise.

There are a few important aspects to note about Equation (2.10). Firstly, the parameter $\tau$, which refers to the coupling strength, controls how much the dynamics is affected by the thermostat: In the limit $\tau \to \infty$, the system is instantaneously thermalized leading to discontinuous dynamics, whereas setting $\tau = 0$ removes the coupling to the bath and retains the microcanonical ensemble. Secondly, if the system is far from the target temperature, the difference $\overline{K} - K$ will become very large leading to an exponential temperature equilibration. On the other hand, if the system is close to equilibrium, the second term in Equation (2.10) will ensure that the system evolves on the canonical hypersurface by adding a stochastic term to the kinetic energy increment.

### The Nosé-Hoover Chain Thermostat

A closely related scheme to the CSVR approach is the Nosé-Hoover chain thermostat.[3,87–89,93] In this method the coupling to the heat bath is realized by extending the system's Hamiltonian by a generalized friction term. For a discussion it is most instructive to first consider the case of a chain length of 1. In this case, the equations of motion become

$$\frac{d^2\mathbf{r}_I}{dt^2} = \frac{\mathbf{F}_I}{m_I} - \frac{p_\xi}{Q}\frac{d\mathbf{r}_I}{dt} \ , \tag{2.11}$$

where $\mathbf{F}_I$ is the conservative force equalling the negative gradient of the potential energy function. The heat bath parameter $p_\xi$ is associated with its own equation of motion reading

$$\frac{dp_\xi}{dt} = (T - T_0) \ . \tag{2.12}$$

Therefore, the change of $p_\xi$ is equal to the difference of the instantaneous temperature $T$ and the target temperature $T_0$. Since the coupling strength parameter $Q$ also depends on $T_0$, it is more convenient to express $Q$ through the period $\tau_T$ of the oscillations of the kinetic energy between the system and the reservoir:

$$Q = \frac{\tau_T^2 T_0}{4\pi^2} \ . \tag{2.13}$$

It is important to note the difference between $\tau_T$ and the parameter $\tau$ in the CSVR method: The Nosé-Hoover method leads to a oscillatory relaxation, whereas $\tau$ in the CSVR method leads to a strongly damped exponential relaxation. Because the relaxation time of the Nosé-Hoover thermostat is much longer than the period of oscillations, $\tau_T$ is typically chosen 4-5 times larger than $\tau$ in the CSVR method.[93]

There are certain systems that are only 'weakly' or even non-ergodic causing the Nosé-Hoover thermostat of chain length 1 to fail to sample all the relevant parts of the phase space, even in the limit of infinite simulation time. This prohibits the calculation of thermodynamic quantities from time averages over the simulation trajectory. A famous example of such a system is a collection of harmonic oscillators. To retain ergodicity in this case, an infinite chain of Nosé-Hoover thermostats is required, where each thermostat is again thermostatted by another one. Luckily, for most practical applications a chain length of 8-10 is sufficient for an ergodic sampling. The equations of motion for a chain length of $N$ take the form[3,89]

$$\frac{d^2\mathbf{r}_I}{dt^2} = \frac{\mathbf{F}_I}{m_I} - \frac{p_{\xi_1}}{Q_1}\frac{d\mathbf{r}_I}{dt} \tag{2.14}$$

$$\frac{dp_{\xi_1}}{dt} = (T - T_0) - p_{\xi_1}\frac{p_{\xi_2}}{Q_2} \tag{2.15}$$

$$\frac{dp_{\xi_{i=2\ldots N}}}{dt} = \left(\frac{p_{\xi_{i-1}}^2}{Q_{i-1}} - kT\right) - p_{\xi_i}\frac{p_{\xi_{i+1}}}{Q_{i+1}} \tag{2.16}$$

$$\frac{dp_{\xi_N}}{dt} = \left(\frac{p_{\xi_{N-1}}^2}{Q_{N-1}} - kT\right) \ . \tag{2.17}$$

### The Parrinello-Rahman Barostat

Analogously to the realization of a canonical ensemble, the system can be coupled to an external 'pressure bath' by extending the Hamiltonian of the system. There exists a number of schemes,[85,86,91] but perhaps the most widely used method today is the approach developed by Parrinello and Rahman in the early 1980s.[94] For a brief introduction to their approach, consider a system with $N$ particles, where the position $\mathbf{r}_I$ of particle $I$ in a box with with cell vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ can be written as

$$r_I = \mathbf{h}\mathbf{s}_I = \xi_I \mathbf{a} + \eta_I \mathbf{b} + \zeta_I \mathbf{c}\ . \tag{2.18}$$

The cell matrix $\mathbf{h}$ is formed by arranging $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ into the columns of a $3 \times 3$ matrix. The cell volume $\Omega$ is then simply calculated as $\det(\mathbf{h})$. Accordingly, $\xi_I$, $\eta_I$ and $\zeta_I$ are the components of a column vector $\mathbf{s}_I$, whose entries take values between 0 and 1. To derive the equations of motion of a pressurized system, consider the following Lagrange function:[92,94]

$$\begin{aligned}\mathcal{L} &= E_{kin} - V \qquad (2.19)\\ &= \frac{1}{2}\sum_{I=1}^{N} m_I \dot{\mathbf{s}}_I^T \mathbf{G}\dot{\mathbf{s}}_I \\ &\quad - \sum_{I=1}^{N}\sum_{J>I}^{N} \phi(r_{IJ}) + \frac{1}{2}W\sum_{I,J}^{N} \dot{h}_{IJ}^2 - P_0\Omega\ . \qquad (2.20)\end{aligned}$$

The first term in Equation (2.20) is the kinetic energy of the particles as a function of the scaled velocities $\dot{\mathbf{r}}_I = \mathbf{h}\dot{\mathbf{s}}_I$, where the transformation matrix $\mathbf{G}$ is defined as $\mathbf{h}^T\mathbf{h}$. The second term is the potential energy function evaluated at positions $\mathbf{r}_I$, where a pair-wise potential $\phi$ was adopted for simplicity. The third term represents the kinetic energy of the cell vectors and incorporates a coupling parameter $W$ that has the unit of a mass. The last term is the potential energy resulting from the external reference pressure $P_0$ times the cell volume. Following the rules of classical mechanics, the equations of motion for the positions and cell vectors can be straightforwardly derived from the Lagrangian as

$$m_I\ddot{\mathbf{s}}_I = \mathbf{h}^{-1}\mathbf{F}_I - m_I\mathbf{G}^{-1}\dot{\mathbf{G}}\dot{\mathbf{s}}_I \tag{2.21}$$

$$W\ddot{\mathbf{h}} = \Omega(\mathbf{\Pi} - \mathbf{P_0})(\mathbf{h}^T)^{-1}\ . \tag{2.22}$$

Equation (2.21) is very similar to Equation (2.11) of the Nosé-Hoover thermostat, which extends the equations of motion of the particle positions by an additional friction term. However, it is important to note that in the case of the Parrinello-Rahman barostat the equation also affects the positions of the particle through the cell matrix, which is propagated according to Equation (2.22). As a result the simulation cell changes its shape as the response to the imbalance of the internal stress $\mathbf{\Pi}$ and the external pressure $\mathbf{P_0}$,[92] which in the isotropic case reduces to $\mathbf{P_0} = p_0\mathbb{I}$. Therefore, it is possible to investigate phase transitions e.g. of a crystal during a molecular dynamics simulation.

Similarly to the Nosé-Hoover thermostat, the coupling strength $W$ is most conveniently expressed through the time period of the pressure oscillations $\tau_p$ between system and bath:[93]

$$W = \frac{3\tau_p^2 L}{4\pi^2 \beta} , \tag{2.23}$$

with $L$ being the largest box matrix element and $\beta$ the isotropic, isothermal compressibility of the system. For practical applications, the Parrinello-Rahman barostat is most commonly combined with a thermostat, such as CSVR or Nosé-Hoover chains, to generate the NPT ensemble. However, this method has the disadvantage that, if the system is far from equilibrium, the barostat will lead to large fluctuations of the box vectors, which might even lead to an abortion of the run. Therefore, it might be necessary to employ a different barostat, such as the Berendsen barostat,[85] for equilibration and switch to the Parrinello-Rahman method for production.[93]

### 2.1.3. Treatment of Interatomic Interactions under Periodic Boundary Conditions

Molecular dynamics simulations of bulk systems are usually carried out under periodic boundary conditions to avoid artefacts from molecules which are located at the system boundaries, and which therefore experience different forces than molecules in the bulk. The use of periodic boundary conditions means that the system of interest is replicated in each spacial dimension to form an infinite lattice. Consequently, the 'original' particles move exactly the same way as the 'image' particles. For instance, if a particle crosses the boundary at one side, it will enter through the opposite face. Since the 'image' particles interact with the original particles the interatomic potential in principle has to be evaluated for an infinite number of particles. However, depending on the range of interactions, various approximative schemes are at hand allowing for a numerically efficient force calculation without a significant loss of accuracy.

#### Short-Range Interactions

For short-range interactions, i.e. interactions that decay faster than $1/r^3$, the potential is usually truncated beyond a cutoff radius $r_c$, i.e.

$$V_{\text{short}}(r_{IJ}) = \begin{cases} V(r_{IJ}), \text{ if } r_{IJ} \leq r_c \\ 0, \text{ else} \end{cases} \tag{2.24}$$

for a pairwise potential $V(r_{IJ})$ between particles $I$ and $J$. However, Equation (2.24) is problematic because the potential is discontinuous at $r_{IJ} = r_c$. As a result, the total energy of a particle crossing $r_c$ will not be conserved.[1] To circumvent this, the potential can be shifted, such that it is zero at the cutoff radius, i.e. by $-v_c = -V(r_c)$:

$$V_{\text{short}}(r_{IJ}) = \begin{cases} V(r_{IJ}) - v_c, \text{ if } r_{IJ} \leq r_c \\ 0, \text{ else} \end{cases} \tag{2.25}$$

Since $v_c$ is a constant, it does not affect the forces. Nevertheless, the force is still discontinuous at $r_c$, which might lead to numerical instabilities. Therefore, an additional correction terms can be introduced, so that the potential derivative, that is the force, is zero at the cutoff distance. For instance, in the simulation package GROMACS[95] the potential is modified according to

$$V_{\text{short}}(r_{IJ}) = \begin{cases} V(r_{IJ}), \text{ if } r_{IJ} < r_1 \\ V_s(r_{IJ}) = V(r_{IJ}) + \int_{r_{IJ}}^{r_c} S(x)\ dx, \text{ if } r_1 \leq r_{IJ} < r_c \\ 0, \text{ else} \end{cases} \tag{2.26}$$

where, $S(x)$ is a third order polynomial function that ensures a continuous force (and force derivative) in the switching region $r_1 \leq r_{IJ} < r_c$.[93]

### Long-Range Dispersion Correction

One peculiarity about cutting off short-range interactions is in simulations in the isothermal-isobaric (NPT) ensemble. In contrast to the error in the potential energy, the error in the pressure due to the use of a finite cutoff is significant and, hence, if accurate pressures are required, an additional correction term needs to be taken into account.[93] For a derivation, consider the following van der Waals potential

$$V_{\text{vdW}}(r_{IJ}) = \frac{C_{12}}{r_{IJ}^{12}} - \frac{C_6}{r_{IJ}^{6}} \ , \tag{2.27}$$

with parameters $C_{12}$ and $C_6$. Since the first term falls off with $1/r^{12}$, only the second term proportional to $1/r^6$ needs to be considered for the correction. The isotropic pressure $P$ in a simulation box with volume $\Omega$ is defined as

$$P = \frac{2}{3\Omega}(E_{\text{kin}} - \Xi) \ , \tag{2.28}$$

with the kinetic energy $E_{\text{kin}}$ and the virial

$$\begin{aligned} \Xi &= -\frac{1}{2}\mathbf{r}_{IJ} \cdot \mathbf{F}_{IJ} \\ &= 3C_6 r_{IJ}^{-6} \ , \end{aligned} \tag{2.29}$$

where $\mathbf{F}_{IJ}$ is the force due the last term in Equation (2.27). Splitting the pressure into a contribution $P_c$ with cutoff and a long-range correction $P_{\text{lr}}$ and using Equations (2.28) and (2.29) yields

$$\begin{aligned} P \approx P_c + P_{\text{lr}} &= P_c - \frac{2}{3\Omega}\Xi_{\text{lr}} \\ &= P_c - \frac{2}{3\Omega}\frac{1}{2}N\rho \int_{r_c}^{\infty} 4\pi \cdot 3C_6 r_{IJ}^{-6}\ dr \ . \end{aligned} \tag{2.30}$$

In the last step, the spherical integration was carried out by assuming a homogeneous particle density $\rho$ beyond $r_c$. It can be easily seen that the long-range pressure correction then takes the form

$$P_{\mathrm{lr}} = -\frac{4}{3}\rho^2 \pi C_6 r_c^{-3} \tag{2.31}$$

In the case of homogeneous mixtures an average dispersion constant $\langle C_6 \rangle$ can be used instead of separately calculating the contribution for each particle pair.[93]

### Electrostatic Interactions

As opposed to van der Waals interactions, electrostatic interactions cannot be treated by a cutoff scheme due to the slower decay of the potential of $1/r$. Instead, the full electrostatic potential between point charges $q_I$ and $q_J$ that reside in a periodic box of length $L$

$$V(\mathbf{r}_1, ..., \mathbf{r}_N) = \sum_{\mathbf{s}} \sum_{I>J} \frac{q_I q_J}{r_{IJ,\mathbf{s}}} \ , \tag{2.32}$$

has to be evaluated differently by the so-called Ewald summation.[96] In the above equation, $\mathbf{r}_1, ..., \mathbf{r}_N$ are the charge positions within the unit cell and $r_{IJ,\mathbf{s}} = |\mathbf{r}_I - \mathbf{r}_J + \mathbf{s}|$ their mutual, periodic distances. The vector $\mathbf{s}$ is given by $L \cdot \mathbf{m}$ with a vector of integers $\mathbf{m}$. As illustrated in Figure 2.1, the summation over the periodic images $\mathbf{s}$ corresponds to the successive addition of spherical symmetric shells.[1] Unfortunately, the evaluation of Equation (2.32) under periodic boundary conditions is problematic: If more and more particle layers are added, $V$ will increasingly oscillate depending on whether the added particle layer has a positive or negative net charge, i.e. the series is only conditionally convergent. However, it can be reformulated so as to achieve absolute convergence. In the Ewald method,[96] this is done by splitting the electrostatic potential $V$ into two terms:

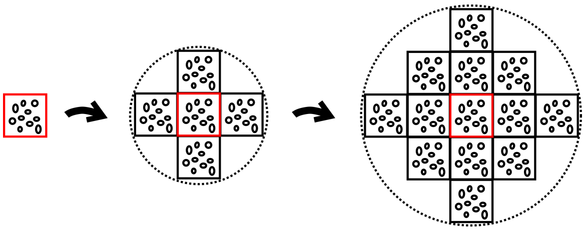


**Figure 2.1.** – Two-Dimensional representation of the successive addition of spherical layers as implied by the summation over s in Equation (2.32). The Figure was adapted from Ref.[1]

$$V(\mathbf{r}_1, ..., \mathbf{r}_N) = V_{\mathrm{short}}(\mathbf{r}_1, ..., \mathbf{r}_N) + V_{\mathrm{long}}(\mathbf{r}_1, ..., \mathbf{r}_n), \tag{2.33}$$

where $V_{\text{short}}$ is short and $V_{\text{long}}$ long ranged. This is realized by introducing an auxiliary Gaussian function, which leads to the following transformation of $1/r$:[3]

$$\begin{aligned}\frac{1}{r} &= \frac{2}{\sqrt{\pi}}\frac{\int_0^\infty e^{-t^2}dt}{r} \\ &= \frac{2}{\sqrt{\pi}}\frac{\int_0^x e^{-t^2}dt}{r} + \frac{2}{\sqrt{\pi}}\frac{\int_x^\infty e^{-t^2}dt}{r} = \frac{\operatorname{erf}(x)}{r} + \frac{\operatorname{erfc}(x)}{r} \ .\end{aligned} \tag{2.34}$$

Thus, the short-ranged part and long-ranged part read

$$V_{\text{short}}(\mathbf{r}_1, ..., \mathbf{r}_N) = \sum_{\mathbf{s}}\sum_{I>J} \frac{q_I q_J \ \operatorname{erfc}(\alpha r_{IJ,\mathbf{s}})}{r_{IJ,\mathbf{s}}} \tag{2.35}$$

$$V_{\text{long}}(\mathbf{r}_1, ..., \mathbf{r}_N) = \sum_{\mathbf{s}}\sum_{I>J} \frac{q_I q_J \ \operatorname{erf}(\alpha r_{IJ,\mathbf{s}})}{r_{IJ,\mathbf{s}}} \ , \tag{2.36}$$

respectively. The parameter $\alpha$ in the error function $\operatorname{erf}(\alpha r_{IJ,\mathbf{s}})$ and the complementary error function $\operatorname{erfc}(\alpha r_{IJ,\mathbf{s}})$ controls the range over which $V_{\text{short}}$ is not negligible. For the numerical evaluation it is convenient to introduce an additional cutoff radius $r_c$, beyond which $V_{\text{short}}$ is set to zero. Consequently, the parameters $\alpha$ and $r_c$ cannot be chosen independently from each other and have to be determined by convergence tests.

Concerning the long-ranged part of the potential, a transformation into Fourier space seems natural because $V_{\text{long}}$ is periodic and long ranged. Transforming Equation (2.36) yields

$$V_{\text{long}}(\mathbf{r}_1, ..., \mathbf{r}_N) = \sum_{I>J} q_I q_J \frac{1}{\Omega} \sum_{\mathbf{g}} C_{\mathbf{g}} e^{i\mathbf{g}\cdot(\mathbf{r}_I - \mathbf{r}_J)} \ , \tag{2.37}$$

with Fourier coefficients

$$C_{\mathbf{g}} = \frac{4\pi}{|\mathbf{g}|^2} e^{-|\mathbf{g}|^2/4\alpha^2} \ . \tag{2.38}$$

The summation is carried out over reciprocal cell vectors $\mathbf{g} = 2\pi\mathbf{m}/L$ with $\mathbf{m} \in \mathbb{N}^3$. Note, the Fourier coefficients are not defined for $\mathbf{m} = \mathbf{0}$ and are therefore set to zero in this case. It can be shown, that this choice corresponds to the solution of the Possion equation, where the infinite periodic system is surrounded by an ideal conductor. A modification is necessary when the surrounding medium should be dielectric.[1] The expression in Equation (2.37) can be further simplified by taking into account the symmetry of $e^{-|\mathbf{g}|^2/4\alpha^2}$. As a result, the sum can be restricted to the hemisphere $\mathcal{S}$ yielding

$$V_{\text{long}}(\mathbf{r}_1, ..., \mathbf{r}_n) = \frac{1}{\Omega} \sum_{I \neq J} q_I q_J \sum_{\mathbf{g} \in \mathcal{S}} \frac{4\pi}{|\mathbf{g}|^2} e^{-|\mathbf{g}|^2/4\alpha^2} e^{i\mathbf{g}\cdot(\mathbf{r}_I - \mathbf{r}_J)} \ . \tag{2.39}$$

Finally, by introducing the electrostatic structure factor $S(\mathbf{g}) = \sum_I q_I e^{i\mathbf{g}\cdot\mathbf{r}_I}$ the total electrostatic potential can be expressed as

$$
\begin{aligned}
V_{\text{Ewald}}(\mathbf{r}_1, ..., \mathbf{r}_n) = \frac{1}{2} \sum_{\mathbf{s}} \sum_{I \neq J} & \frac{q_I q_J \ \text{erfc}(\alpha r_{IJ,\mathbf{s}})}{r_{IJ,\mathbf{s}}} \\
& + \frac{1}{\Omega} \sum_{\mathbf{g} \in \mathcal{S}} \frac{4\pi}{|\mathbf{g}|^2} e^{-|\mathbf{g}|^2/4\alpha^2} |S(\mathbf{g})|^2 \\
& - \frac{\alpha}{\sqrt{\pi}} \sum_I q_I^2 \ ,
\end{aligned}
\tag{2.40}
$$

where the first term in $V_{\text{Ewald}}$ is the short-ranged and the second the long-ranged contribution to the potential. The last term is often referred to as the self-interaction correction as it corrects the interaction of a charge with itself.

As the evaluation of the electrostatic structure factor for an increasing number of charges quickly becomes demanding, modern molecular dynamics simulations software employ further approximations. One of the most commonly used ones is the smooth particle mesh Ewald (SPME) scheme.[97] Here, the electronic structure factor is projected on a grid such that it can be efficiently Fourier transformed, and spline interpolation is used to calculate the interactions between charges that are located between grid points.

## 2.1.4. Ab Initio Molecular Dynamics Simulations

Many questions in chemistry involve chemical reactions, *i.e.* the breaking and formation of covalent bonds and associated rearrangements of electron density. In these cases, the electronic potential needs to be explicitly taken into account by the molecular dynamics simulation method. This is realized in *ab initio* molecular dynamics simulations. In this approach, the electronic potential is directly calculated from first-principles, *i.e.* without any adjustable parameters. In Born-Oppenheimer molecular dynamics, which is one the most commonly used flavors of *ab initio* molecular dynamics, the electronic potential is obtained by solving the electronic structure problem at every time step, *i.e.* for fixed nuclear coordinates.

### Born Oppenheimer Molecular Dynamics

The theoretical foundation of the Born Oppenheimer molecular dynamics method is laid by a separation of nuclear and electronic degrees of freedom, and subsequently by the adoption of a classical description for the nuclei.[4] The starting point for that is the time-dependent Schrödinger equation of a system consisting of $N$ nuclei and $n$ electrons,

$$
i\frac{\partial}{\partial t}\Phi(\mathbf{r}, \mathbf{R}, t) = \hat{H} \ \Phi(\mathbf{r}, \mathbf{R}, t) \ , \tag{2.41}
$$

where the Hamiltonoperator $\hat{H}$ (in atomic units) for the total system is defined as

$$\begin{aligned}\hat{H} = &-\sum_I \frac{1}{2M_I}\nabla_I^2 - \sum_i \frac{1}{2}\nabla_i^2 \\ &+\sum_{i<j}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{I,i}\frac{Z_I}{|\mathbf{R}_I - \mathbf{r}_i|} + \sum_{I<J}\frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} \\ = &-\sum_I \frac{1}{2M_I}\nabla_I^2 + \hat{H}_{\text{elec}}(\mathbf{r},\mathbf{R}) \ ,\end{aligned} \tag{2.42}$$

with nuclear and electronic coordinates $\mathbf{R} = \{\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_N\}$ and $\mathbf{r} = \{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n\}$, respectively. The electronic Hamiltonian $\hat{H}_{\text{elec}}$ incorporates all terms, but the kinetic energy of the nuclei. Since the eigenfunctions of $\hat{H}$ form a complete basis, the total wave function can be expressed as

$$\Phi(\mathbf{r},\mathbf{R},t) = \sum_{l=0}^{\infty} \Psi_l(\mathbf{r},\mathbf{R}) \cdot \Xi_l(\mathbf{R},t) \ , \tag{2.43}$$

where it was assumed that $\{\Psi_i(\mathbf{r},\mathbf{R})\}$ is a discrete set of functions. Note that the expansion coefficients, $\{\Xi_i(\mathbf{R},t)\}$, solely depend on the nuclear coordinates and time. By inserting Equation (2.43) into Equation (2.41) and integrating over the electronic degrees of freedom, one obtains

$$\left[-\sum_I \frac{1}{2M_I}\nabla_I^2 + E_n(\mathbf{R})\right]\Xi_n + \sum_l C_{nl}\Xi_l = i\frac{\partial}{\partial t}\Xi_n \ , \tag{2.44}$$

with coupling matrix elements

$$\begin{aligned}C_{nl} = &\int \Psi_n^* \left[-\sum_I \frac{1}{2M_I}\nabla_I^2\right] \Psi_l \, d\mathbf{r} \\ &+ \frac{1}{M_I}\sum_I \left\{\int \Psi_n^* \left[-i\nabla_I\right] \Psi_l \, d\mathbf{r}\right\} \left[-i\nabla_I\right] \ .\end{aligned} \tag{2.45}$$

It should be noted that the coupling matrix $\mathbf{C}$ drives transitions between nuclear wave functions corresponding to different electronic quantum states. In many situations, however, it can be safely assumed that the nuclei only move on one specific electronic hypersurface, *e.g.* on the electronic ground surface, which means that only diagonal elements of $\mathbf{C}$ contribute. This is often called the 'adiabatic' approximation. This leads to a decoupling of electronic and nuclear motions according to

$$\Phi(\mathbf{r},\mathbf{R},t) \approx \Psi_n(\mathbf{r};\mathbf{R}) \cdot \Xi_n(\mathbf{R},t) \tag{2.46}$$

The Born-Oppenheimer approximation further neglects the diagonal elements of the coupling matrix which results in the time-dependent Schrödinger equation

$$\left[-\sum_I \frac{1}{2M_I}\nabla_I^2 + E_n(\mathbf{R})\right]\Xi_n = i\frac{\partial}{\partial t}\Xi_n \ . \tag{2.47}$$

Setting out from Equation (2.47) it can be rigorously shown[4] that, in the limit of classical nuclear motion, the equations of motion are given by

$$M_I\ddot{R}_I = -\nabla_I E_n(\mathbf{R}) \ , \tag{2.48}$$

which forms the basis of Born-Oppenheimer molecular dynamics simulations.

Clearly for most applications, molecular dynamics simulations are carried out in the electronic ground state ($n = 0$). In this case, any electronic structure method that yields accurate ground state energies and forces can principally be used for the computation of the electronic potential. However, molecular dynamics simulations often require the potential to be calculated at many tens of thousands of time steps, and many liquid state simulations at the same time require the consideration of a large number of atoms. Despite some recent developments of wave function based approaches, such as RI-MP2, which have resulted in a significantly increased numerical effiency, these methods do not yet qualify for routine applications.[98] By contrast, density functional theory (DFT) has been proven to be numerically efficient and at the same time reasonably accurate for many research problems. For this reason, modern *ab initio* simulations are almost exclusively based on DFT.

### 2.1.5. Classical Molecular Dynamics Simulations

For many systems *ab initio* molecular dynamics simulations cannot be used because the number of atoms or the relevant time scales exceed the computational feasibility of a quantum mechanical treatment. In these cases, classical molecular dynamics simulations can be a powerful alternative. For comparison, standard DFT-based *ab initio* molecular dynamics simulations can tackle system sizes of only a few thousands of atoms and access times scales of usually no more than a few hundred picoseconds, while classical simulations employing empirical force field potentials can nowadays be carried out on systems that comprise more than a million of atoms for physical time scales that are typically in the microsecond regime. One of the main application areas of classical molecular dynamics simulations are biomolecular simulations. Here, many force field potentials have been developed that describe distinct types of systems with high accuracy, such as solvated peptides and proteins, or nucleic acids. In the following, a brief introduction to the topic will be given, where the focus will be put on empirical potentials that have been used in this thesis.

The basis of all empirical force field models is the representation of the complex potential energy hypersurface on which the nuclei move by a simple set of analytical functions which describe various parts of interatomic interactions. In the most general case, the total potential energy as a function of the nuclear positions $\mathbf{r} = \{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N\}$ is given by

$$V(\mathbf{r}) = V_{\text{bonded}}(\mathbf{r}) + V_{\text{non-bonded}}(\mathbf{r}) \ . \tag{2.49}$$

The first term incorporates functions for the treatment of covalent interactions, such as bond stretches, bond angle bends, and dihedral angle torsions. The second term describes non-bonded interactions, like Coulomb, dispersion, and exchange repulsion interactions. Sometimes force fields additionally incorporate cross-terms (class II force fields), *e.g.* terms that depend on both bonds and angles, or terms to describe polarization effects (class III force fields). However, the most commonly used force fields employ a relatively simple potential energy function:[99]

$$V(\mathbf{r}) = \sum_{\text{bonds}} k_b(b-b_0)^2 + \sum_{\text{angles}} k_0(\theta-\theta_0)^2 + \sum_{\text{torsions}} \sum_{m=1}^{M} k_{\phi,m}[\cos(m\phi-\phi_0)+1] + \sum_{\text{nonbond pairs}} \left[ \frac{q_I q_J}{r_{IJ}} + \left(\frac{\sigma_{IJ}}{r_{IJ}}\right)^{12} - \left(\frac{\sigma_{IJ}}{r_{IJ}}\right)^{6} \right] , \tag{2.50}$$

with equilibrium distances $b_0$ and angles $\theta_0$ and $\phi_0$, respectively, and force constants $k_b$, $k_0$ and $k_{\phi,m}$. The first three summations are over bonds, angles, and dihedral angles. The latter also includes 'improper' torsions, *i.e.* hindered rotations, *e.g.* to enforce the planarity in amide bonds, where the number of terms $M$ in the cosine expansion depends on the type of dihedral angle. For instance in the Amber ff99SB-ILDN force field,[100] $M$ is up to six for some angles. The last term in Equation (2.50) describes the electrostatics between the point charges $q_I$ and $q_J$ by Coulomb's law and interactions due to dispersion and exchange repulsion by a Lennard-Jones potential with parameters $\{\sigma_{IJ}\}$. It should be noted that the sum over non-bonded atom pairs excludes pairs of atoms that are separated by one (1-2 interactions) or two (1-3 interactions) bonds. Moreover, many force fields use distinct parameters to describe 1-4 interactions.

The form of Equation (2.50), the set of parameters $\{b_0, k_b,\ \theta_0,\ k_0,\ k_{\phi,m},\ \phi_0, q_I,\ \sigma_{IJ}\}$, and the definition of atom types is what defines a force field. To find optimal parameters, fits to experimental and quantum-chemistry data have to be carried out. However, not all parameters can be obtained independently from each other because the division into different interaction types is only an approximation. Nevertheless, considering only bonds and angles, this approximation works reasonably well and the force constants and equilibrium values for these degrees of freedom are often transferred between different force fields without modifications.[101] For example, note that in the course of development of some commonly used Amber force fields, bonding and angle parameters were essentially left unchanged: The parameter set for 1-2 and 1-3 interactions by Weiner and coworkers published in 1986[102,103] was transfered with only little modifications to Amber ff94,[13] Amber ff99SB,[104] and Amber ff99SB-ILDN. Note that the latter was proposed in 2010. Clearly, from today's standpoint some aspects of the original fitting protocol by Weiner et al.[102] do not represent the current state of the art. For instance, the authors used rather low level quantum-chemistry methods and a variety of different experimental techniques, such as microwave spectroscopy, NMR, and neutron diffraction, precluding a consistent definition of equilibrium values and force constants. The reason why comparable little effort was put into a more accurate determination is that the performance of a protein force field is largely dominated by torsional and non-bonding interactions. Naturally, recent developments aimed to more accurately represent these strongly correlated degrees of freedom. Moreover, a large number of benchmark simulations revealed force field specific imbalances in the description of secondary structure elements, which mostly originated from flaws in the torsional potential fitting. Unfortunately, the quest for optimal parameters is a largely underdetermined problem, and thus, there exist different 'philosophies' on how to obtain these values. For instance for re-parametrizations of Amber protein force fields, charges and van der Waals parameters have either been re-determined, or taken from the parent force field with only little modifications, followed by an elaborate derivation of torsion parameters, with most effort invested on accurately modelling the quantum-chemical torsional potential hypersurfaces.

A well-known example for which only torsional parameters have been revised, is Amber ff99SB.[104] Based on its predecessor, Amber ff94, the authors defined four distinct types of dihedral angles. As shown in Figure 2.2, $\Psi/\Phi$ denote the usual peptide backbone torsion angles, and $\Psi'$ and $\Phi'$ additionally incorporate the $C^\beta$ atom of the side chain. The optimization of dihedral angles was done in two steps. Firstly, the $\Psi/\Phi$ angle pairs were scanned for a $\text{Gly}_3$ tetrapeptide. Because the resulting 12-dimensional space is too large for an exhaustive sampling, a stochastic sampling was carried out instead to identify local minima on the potential energy surface. For each conformer, the quantum-mechanical energies were compared to the energies of the empirical potential. The final parameter minimizes the maximum absolute error (MAE)

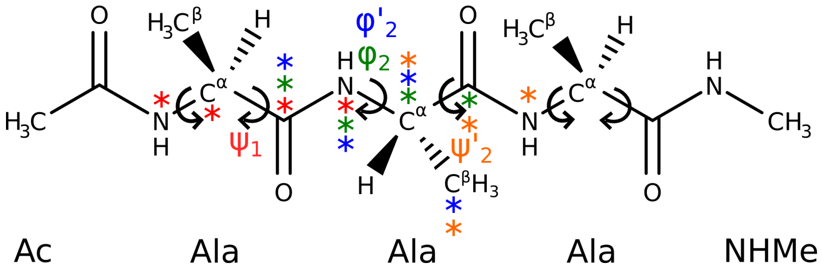


**Figure 2.2.** – Scheme of an alanine tetrapeptide (four peptide bonds) as used in the fitting of dihedral angle parameters for the Amber ff99SB force field.[104] Four different types of dihedral angles, $\Psi$, $\Phi$, $\Psi'$, and $\Phi'$ were defined for each rotatable bond. The Figure was adapted from Ref.[104]

$$\text{MAE} = \max_{i,j>i} |E^i_{\text{QM}}(j) - E^i_{\text{MM}}(j)| \; , \tag{2.51}$$

where $E^i_{\text{QM}}(j)$ and $E^i_{\text{MM}}(j)$ is the quantum-mechanical and force field energy of conformer $j$ with conformer $i$ as reference, respectively. In a second step, the space spanned by angles $\Psi'$, and $\Phi'$ was analogously scanned for the $\text{Ala}_3$ tetrapeptide, while keeping the parameters for the $\Psi/\Phi$ angles fixed. The final parameter set was again found by minimizing Equation (2.51).

### 2.1.6. Hybrid Quantum/Classical Molecular Dynamics Simulations

After having introduced *ab initio* and force field molecular dynamics simulations, the natural question is whether these two methods can be combined into a scheme that incorporates both, the high accuracy of a quantum mechanical and the numerical efficiency of a classical treatment. This approach is termed mixed quantum mechanical/molecular mechanics (QM/MM). In QM/MM a relevant subsystem, such as an active site of a protein or a solute, is treated quantum mechanically, whereas the surrounding matrix, *i.e.* the protein or the solvent molecules, are approximated by empirical force field potentials. This idea has been firstly introduced to the field of biomolecular simulations by Warshel and Levitt in 1976[105] and since been applied to a large number of research problems.

However, one of the key challenges of the QM/MM ansatz is that the coupling between the QM and MM subsystems cannot be uniquely defined. As a result, various schemes have been put forward in the literature. Most relevant for the current work are so-called additive schemes, in which the total energy of the QM/MM system is split into three parts[4]

$$E^{\text{QM/MM}} = E^{\text{QM}}(\mathbf{R}_\alpha) + E^{\text{MM}}(\mathbf{R}_I) + E^{\text{QM-MM}}(\mathbf{R}_\alpha, \mathbf{R}_I) \ . \tag{2.52}$$

The first and second term are the total energy of the quantum mechanical and classical subsystem, respectively. Their evaluation is covered in Section 2.2 and Section 2.1.5, respectively. The last term denotes the coupling energy between both parts. Here, nuclear coordinates that belong to atoms in the QM part are indicated by $\mathbf{R}_\alpha$, whereas the MM coordinates are denoted with $\mathbf{R}_I$.

Assuming that no bonds are shared between the subparts of the system, the coupling can be further divided into an electrostatic (ES) and van der Waals term (vdW):

$$E^{\text{QM-MM}} = E^{\text{QM-MM}}_{\text{ES}} + E^{\text{QM-MM}}_{\text{vdW}} \tag{2.53}$$

The short-ranged van der Waals interactions are usually described by Lennard-Jones potentials with parameters most commonly adapted from the force field that is used to evaluate $E^{\text{MM}}$. By contrast, the treatment of $E^{\text{QM-MM}}_{\text{ES}}$ is technically more elaborate. Firstly, the straightforward evaluation of the integral

$$E^{\text{QM-MM}}_{\text{ES}} = \sum_{I \in \text{MM}} q_I \int \frac{\rho(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_I|} \, d\mathbf{r} \tag{2.54}$$

can be very demanding for reasons that will be outlined below, and secondly, the interaction of electrons in the QM part with point charges in the MM part leads to an 'electron spill-out', *i.e.* an increase of electron density in the classical region and a concomitant depletion of density in the QM part. This artefact becomes increasingly dominant, if more flexible and locally less confined basis sets are used to expand the electron density, such as plane waves or Gaussian basis sets with diffuse functions.

To remedy the 'spill-out' problem, Laino et al.[106] suggested to 'smear' the classical point charges according to a Gaussian distributions (see Figure 2.3). This avoids a 'trapping' of electronic charge density in the steep potential in proximity to the point charges. Note that a similar idea is used to evaluate the electrostatic potential under periodic boundary conditions, which has been described in Section 2.1.3.

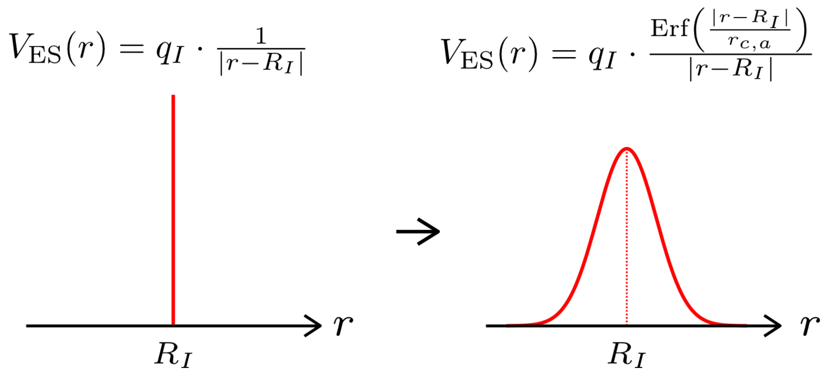

$V_{\mathrm{ES}}(r) = q_I \cdot \frac{1}{|r-R_I|}$
$V_{\mathrm{ES}}(r) = q_I \cdot \frac{\mathrm{Erf}\left(\frac{|r-R_I|}{r_{c,a}}\right)}{|r-R_I|}$
$r$
$R_I$
$r$
$R_I$

**Figure 2.3.** – Graphical illustration of the Gaussian 'smearing' of a point charge $q_I$, located at $R_I$, as employed in the method by Laino *et al.*[106] The electrostatic potential $V_{\mathrm{ES}}$ is modified according to the equations below the graphs. Note, Erf stands for the error function as defined in Equation (2.34).

Apart from charge leakage, the evaluation of the coupling energy in Equation (2.54) is non-trivial. Since in plane wave codes the QM potential is represented on a numerical grid, a brute force integration would scale with the number of grid point times the number of point charges in the system. This easily takes up as much time as needed for the solution of the electronic structure problem alone, and is therefore prohibitive.[4,106] Laino *et al.* therefore expanded the modified electrostatic potential in terms of Gaussian functions in real space

$$V_{\mathrm{ES}}(\mathbf{r},\mathbf{R}_I) = \frac{\mathrm{Erf}\left(\frac{|\mathbf{r}-\mathbf{R}_I|}{r_{c,a}}\right)}{|\mathbf{r}-\mathbf{R}_I|} = \sum_{N_g} A_g exp\left(-\left(\frac{|\mathbf{r}-\mathbf{R}_I|}{G_g}\right)^2\right) + R_{\mathrm{low}}(|\mathbf{r}-\mathbf{R}_I|)\ , \qquad (2.55)$$

where the parameters $A_g$ were obtained by a fit to the analytical potential. The advantage of using a number of Gaussian functions to represent the electrostatic potential is that a threshold (collocation region) can be defined after which the contribution to the potential vanishes. This is illustrated in Figure 2.4. It can be seen that only those MM atoms contribute to the electrostatic coupling energy that are within the collocation radius of the Gaussian at the coarsest grid level. Note that all MM atoms still contribute via the residual, long-ranged term $R_{\mathrm{low}}$. This procedure significantly reduces the number of operations compared to an evaluation of the integral in Equation (2.54) at the finest grid level. For example, if four Gaussian functions are used in the expansion, the speed-up factor is about 512, i.e. two orders of magnitude.[106] Finally, it should be mentioned that the situation is less complicated for codes that expand the electron density in Gaussian type orbitals. In these cases, the coupling is added directly to the Hamiltonian according to

$$\mathbf{H}^{\mu\nu}_{\mathrm{QM/MM}} = -\int \phi_\mu(\mathbf{r},\mathbf{R}_\alpha) \sum_{I\in\mathrm{MM}} \frac{q_I}{|\mathbf{r}-\mathbf{R}_I|} \phi_\nu(\mathbf{r},\mathbf{R}_\alpha)\, d\mathbf{r}\ . \qquad (2.56)$$

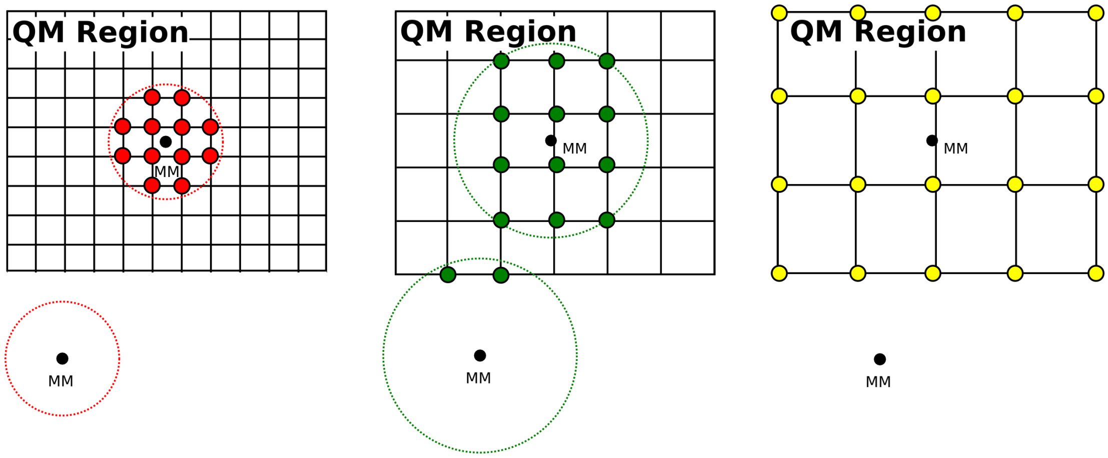


**Figure 2.4.** – Graphical illustration of the real space expansion of the MM potential according to Equation (2.55). Depicted are three grid levels, where the first two correspond to distinct Gaussian terms in the expansion. The residual, long-ranged term $R_{\text{low}}$ is represented at the coarsest grid level. The situation is illustrated for two MM atoms located inside and outside the QM region, respectively. Circles around the MM atoms correspond to the collocation radius, *i.e.* the radius beyond which the Gaussian function can be neglected. Thus, only MM atoms that are within this radius to the QM region contribute to $E_{\text{ES}}^{\text{QM-MM}}$, whereas all MM atoms contribute to the residual $R_{\text{low}}$. The Figure has been adapted from Ref.[106]

This allows for the utilization of standard prescreening routines to reduce the number of integrals over Gaussian basis functions $\phi_\mu$ and $\phi_\nu$.[106]

## 2.2. Density Functional Theory

As opposed to wave function based approaches that approximately solve the stationary, electronic Schrödinger equation

$$\hat{H}_{\text{elec}}(\mathbf{r},\mathbf{R})\Psi_n(\mathbf{r};\mathbf{R}) = E_n(\mathbf{R})\Psi_n(\mathbf{r};\mathbf{R}) \tag{2.57}$$

to yield approximated eigenstates and energies, DFT employs the electron density

$$\rho(\mathbf{r}) = N \cdot \int |\Psi_n(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)|^2 d\sigma_1 d\mathbf{x}_2 d\mathbf{x}_3 \ldots d\mathbf{x}_N \tag{2.58}$$

as the central quantity, where $\Psi_n(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ denotes an electronic wave function of state $n$ with spin coordinate $\mathbf{x}_i$ of the i-th electron, *i.e.* the direct product of spatial coordinate and spin state, $\mathbf{r}_i \cdot \sigma_i$. Because of its dependence on just three spatial coordinates, instead of $3^n$ as the electronic wave function (with $n$ being the total number of electrons), the electron density is a mathematically much less complicated quantity.

### 2.2.1. The Hohenberg-Kohn Theorems

In 1964 Hohenberg and Kohn laid the theoretical foundations of DFT by introducing their two famous Hohenberg-Kohn theorems (HKT).[107] The first HKT proofs that there cannot be two different external potentials, $V_{\text{ext}}$ and $V'_{\text{ext}}$, that differ by more than an additive constant and give rise to the same ground state electron density $\rho_0$. In other words, $\rho_0$ uniquely defines the electronic Hamiltonian (up to an additive constant), which in turn defines the ground state wave function (up to a phase factor) and all other information that can be obtained from the system. Furthermore, it follows the existence of a unique functional of the ground state energy,

$$E_0[\rho_0] = E_{\text{Ne}}[\rho_0] + F_{\text{HK}}[\rho_0] \ , \tag{2.59}$$

where $E_{\text{Ne}}[\rho_0]$ is the functional of the electrostatic interaction energy of the electrons with the nuclei and $F_{\text{HK}[\rho_0]}$ the universal functional of the electron density. Since the nuclei are at fixed positions, $E_{\text{Ne}}[\rho_0]$ is system dependent, whereas $F_{\text{HK}}[\rho_0]$, which incorporates the electron-electron and kinetic energy contribution, is system independent.

The second HKT states the existence of a variational principle for Equation (2.59), such that

$$E_0[\rho_0] \leq E_0[\widetilde{\rho}] \ , \tag{2.60}$$

with an arbitrary test density $\widetilde{\rho}$. Unfortunately, this theorem does not give any clue of how this functional can be obtained, and even worse for practical applications, Equation (2.60) does not hold if an approximation to $E_0[\rho]$ is used instead.

A slightly different formulation of the second HKT was introduced by Levy in 1979.[108] He divided the variational principle

$$E_0 = \min_{\Psi \to n} \left\langle \Psi \middle| \hat{T} + \hat{V}_{\text{ext}} + \hat{V}_{ee} \middle| \Psi \right\rangle \ , \tag{2.61}$$

into two steps. Firstly, the search is carried out over all wave functions that give rise to a particular density $\rho$. Secondly, the ground state density $\rho_0$ is identified among all the previously obtained densities as the one associated with the electronic state that has yielded the lowest energy in the first step. This can be written as

$$\begin{aligned} E_0 &= \min_{\rho\rightarrow n}\left(\min_{\Psi\rightarrow\rho}\left\langle \Psi \middle| \hat{T}+\hat{V}_{\text{ext}}+\hat{V}_{ee} \middle| \Psi \right\rangle\right) \\ &= \min_{\rho\rightarrow n}\left(F[\rho]+\int \rho(\mathbf{r})V_{\text{ext}}d\mathbf{r}\right) \ , \end{aligned} \tag{2.62}$$

with

$$F[\rho] = \min_{\Psi\rightarrow\rho}\left\langle \Psi \middle| \hat{T}+\hat{V}_{ee} \middle| \Psi \right\rangle \ . \tag{2.63}$$

In the last step of Equation (2.62) it was used that $\hat{V}_{\text{ext}}$ is just a multiplicative operator. Thus, Equation (2.62) gives a formal route of how the ground state electron density $\rho_0$ relates to the 'real' electronic ground state, defined by $\Psi_0$. This makes DFT an in-principle exact theory, even though it is impossible to derive an exact functional of the ground state energy for any realistic system.

## 2.2.2. The Kohn-Sham Approach

Only one year after the HKT have been published, W. Kohn and L. J. Sham presented an ansatz of how this functional can be actually approximated.[109] This can be acknowledged as the breakthrough of DFT, since it paved the way for its use as the nowadays probably most commonly employed electronic structure method. Their ingenious idea is based on the pragmatic strategy to calculate as much as possible of the unknown functional $E_0[\rho]$ for a reference system of $n$ non-interacting electrons and approximate the unknown remainder, which only accounts for a small contribution to the total energy. This can be expressed as

$$E_0[\rho] = T_{\text{s}}[\rho] + E_{\text{Ne}}[\rho] + J[\rho] + E_{\text{xc}}[\rho] \ , \tag{2.64}$$

where the terms are the kinetic energy of the non-interacting system $T_{\text{s}}[\rho]$, the electrostatic electron-nuclei interaction $E_{\text{Ne}}[\rho]$, the classical electron-electron interaction $J[\rho]$, and the exchange-correlation energy $E_{\text{xc}}[\rho]$ that incorporates everything unknown, *i.e.* $T_{\text{c}}[\rho]$, the self-interaction correction, and the exchange and correlation energy.

Since the solution of the Schrödinger equation of a non-interacting system is a single Slater Determinant (SD), composed of $N$ (spin) orbitals $\{\phi_i\}$, all terms in Equation (2.64) except $E_{\text{xc}}[\rho]$ can explicitly be written as

$$T_s[\rho] = -\frac{1}{2}\sum_{i=1}^{N}\int \phi_i^*(\mathbf{r})\nabla^2\phi_i(\mathbf{r})d\mathbf{r} \ , \tag{2.65}$$

$$\begin{aligned} E_{Ne}[\rho] &= -\sum_{I=1}^{M} Z_I \int \frac{\rho(\mathbf{r})}{|\mathbf{r}-\mathbf{R}_I|} d\mathbf{r} \\ &= \int v_{\text{ext}}(r)\rho(\mathbf{r}) d\mathbf{r} \ , \end{aligned} \tag{2.66}$$

$$J[\rho] = \frac{1}{2}\int\int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r} d\mathbf{r}' \ . \tag{2.67}$$

In the Kohn-Sham scheme, the density of the non-interacting reference state is then set equal to the one of the true, interacting system,

$$\rho(\mathbf{r}) = \sum_{i=1}^{N} |\phi_i(\mathbf{r})|^2 \ . \tag{2.68}$$

Minimization of Equation (2.64) under the constraint of orthonormal orbitals leads to the name Kohn-Sham (KS) equations

$$\begin{aligned} \left(-\frac{1}{2}\nabla^2 + v_{\text{ext}}(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' + v_{xc}(\mathbf{r})\right) \phi_i &= \varepsilon_i \phi_i \text{ and c.c.} \\ \left(-\frac{1}{2}\nabla^2 + v_{\text{eff}}(\mathbf{r})\right) \phi_i &= \varepsilon_i \phi_i \text{ and c.c.,} \\ \hat{\mathrm{f}}^{\text{KS}} \phi_i &= \varepsilon_i \phi_i \text{ and c.c.,} \end{aligned} \tag{2.69}$$

where the exchange-correlation potential is defined as the functional derivative of the exchange-correlation energy

$$v_{xc} = \frac{\delta E_{xc}}{\delta \rho} \ . \tag{2.70}$$

As the effective potential $v_{\text{eff}}$ itself depends on the density, the KS equations have to be solved in a self-consistent manner. It should be noted that, unlike the Hartree-Fock equations, the KS equations are in-principle exact, provided the exact form of $E_{xc}$ and with it $v_{xc}$ are available. Clearly, the performance of the model increases, if better approximations for $E_{xc}$ are used.

### 2.2.3. Approximations to the Exchange-Correlation Functional

The success of density functional theory is not only due to the fact that it is a formally exact theory, but that approximations to the exchange-correlation functional $E_{xc}[\rho]$ have been suggested that enable an efficient, yet accurate calculation of ground state properties, such as the total energy. Perhaps the model on which most of the approximate exchange-correlation functionals are based is the uniform electron gas. This system consists of a infinite number of electrons, $n \to \infty$, within an infinite volume, $V \to \infty$, with a uniform and finite electron density $N/V = \rho$.[110]

### Local Density Approximation

One of the reasons for the popularity of the uniform electron gas is that $E_{xc}[\rho]$ is known for this system to a very high degree of accuracy. The class of exchange-correlation functionals that makes use of this approximation is called local density approximation (LDA). This implies that the exchange-correlation functional is of the form[110]

$$E_{\mathrm{xc}}^{\mathrm{LDA}}[\rho] = \int \rho(\mathbf{r})\varepsilon_{\mathrm{xc}}(\rho(\mathbf{r}))\, d\mathbf{r} \ , \tag{2.71}$$

where $\varepsilon_{\mathrm{xc}}(\rho(\mathbf{r}))$ is the exchange-correlation energy per electron of a uniform electron gas with density $\rho(\mathbf{r})$. The exchange-correlation energy can be further divided in an exchange and correlation term according to

$$\varepsilon_{\mathrm{xc}}(\rho(\mathbf{r})) = \varepsilon_{\mathrm{x}}(\rho(\mathbf{r})) + \varepsilon_{\mathrm{c}}(\rho(\mathbf{r})) \ . \tag{2.72}$$

The exchange part $\varepsilon_{\mathrm{x}}(\rho(\mathbf{r}))$ can exactly be expressed as[110–112]

$$\varepsilon_{\mathrm{x}} = -\frac{3}{4}\left(\frac{3\rho(\mathbf{r})}{\pi}\right)^{\frac{3}{2}} \ , \tag{2.73}$$

However, no exact analytical expression is known for the correlation part $\varepsilon_{\mathrm{c}}$. Nonetheless, this term has been approximated to a very high degree of accuracy by Monte-Carlo simulations.[113–115]

It is rather surprising that, although the electronic charge density varies significantly in molecules, and thus is far away from an uniform distribution, LDA functionals show an acceptable performance for some important molecular properties, like equilibrium structures, harmonic frequencies and charge moments.[110] However, quite disappointingly for chemical applications, LDA performs much worse on energetics. For instance, bonding energies are usually overestimated, and are therefore rendering the method impracticable for many chemical applications.

### Generalized Gradient Approximation

A logical extension of the LDA approximation is to take into account information on the gradient of the charge distribution. This is leads to the generalized gradient approximation (GGA), whose general form can be expressed as

$$E_{\mathrm{xc}}^{\mathrm{GGA}}[\rho] = \int f(\rho, \nabla\rho)\, d\mathbf{r} \ . \tag{2.74}$$

There exist several suggestions for the explicit form of the integrand $f$, resulting in a number of different GGA functionals, where the explicit analytical form of $f$ is usually quite complicated. It should be pointed out that the integrand is often solely chosen based on an optimal performance, and sometimes not even based on a physical model.

Similarly to Equation (2.72), $E_{\mathrm{xc}}^{\mathrm{GGA}}[\rho]$ can be split into an exchange and correlation part

$$E_{\mathrm{xc}}^{\mathrm{GGA}}[\rho] = E_{\mathrm{x}}^{\mathrm{GGA}}[\rho] + E_{\mathrm{c}}^{\mathrm{GGA}}[\rho] \ . \tag{2.75}$$

In practice these two parts are often parametrized separately and in principle any correlation functional could be combined with any exchange functional. Nevertheless, only a few combinations are well established. For instance, the exchange functional by Becke[116] is usually combined with the correlation functional by Lee, Yang, and Parr[117] to the BLYP exchange-correlation functional, or with the correlation functional by Perdew[118] to form BP86. Another famous GGA functional was published by Perdew, Burke, and Ernzerhof in 1996,[119] commonly abbreviated as PBE.

The performance of these GGA functionals is much better than LDA functionals, while still retaining a local expression for the density and its gradient. If the energies of G2 data set[120,121] are computed, the absolute average errors will usually be around 5-7 kcal/mol for GGA, whereas standard LDA functionals produce an unsigned mean error of about 36 kcal/mol.[110]

#### Hybrid Functionals

The exchange and correlation contributions to LDA and GGA exchange-correlation functionals are purely local. However, it can be shown that an exact functional also incorporates a fraction of non-local exchange.[110,122,123] Since the exchange energy of a Slater determinant, i.e. an anti-symmetrized product of Kohn-Sham orbitals, can be computed exactly within the Hartree-Fock scheme, a logical extension is to incorporate this into the expression of approximate exchange-correlation functionals. This results in so-called hybrid functionals. Again, there are different ways how the exact Hartree-Fock exchange can be combined with the local corrections of LDA and GGA, and often mixing parameters are found by fits to reference sets, such as the G2 set. A notable exception, however, is the PBE0 functional introduced by Burke, Ernzerhof, and Perdew in 1997.[124] Based on perturbation theory arguments they proposed an admixture of 25 % of Hartree-Fock exchange,

$$E_{\mathrm{xc}}^{\mathrm{PBE0}} = E_{\mathrm{xc}}^{\mathrm{GGA}} + 0.25 \cdot \left(E_{\mathrm{x}}^{\mathrm{HF}} - E_{\mathrm{x}}^{\mathrm{GGA}}\right) , \tag{2.76}$$

where the PBE exchange-correlation functional was used for $E_{\mathrm{xc}}^{\mathrm{GGA}}$. Nevertheless, this should not deceive from the fact that there is no ‘true’, optimal parameter, but it is rather system dependent, and in pathological cases, like the $O_3$ molecule, the hybrid functionals give disastrous results for harmonic frequencies.[110] However, experience showed that these functionals perform in many cases superior over GGA functionals at the cost of an increased computational demand due to the computation of the exact Hartree-Fock exchange.

### 2.2.4. The LCAO Ansatz

In the Kohn-Sham approach to DFT, the electron density is expanded in terms of one-electron functions, *i.e.* Kohn-Sham orbitals. Thus, solving the electronic structure problem amounts to solving Equation (2.69) in a self-consistent manner. In principle, this coupled set of integro-differential equations can be solved purely numerical. However, this would be too demanding for most of the practical applications. Therefore, similarly to wave-function based methodologies, modern DFT codes make use of the linear combination of atomic orbitals (LCAO) approach introduced by Roothan in 1951,[125] which linearly expands the Kohn-Sham orbitals as

$$\phi_i = \sum_{\mu=1}^{L} c_{\mu i} \eta_\mu \quad . \tag{2.77}$$

The functions $\{\eta_\mu\}$ are a set of basis functions, *e.g.* atomic orbitals. As a consequence, Equation (2.69) transforms into the linear problem of finding the set of coefficients $\{c_{\mu i}\}$ that leads to a self-consistent solution. This can be compactly expressed as

$$\mathbf{F}^{\mathrm{KS}} \ \mathbf{C} = \mathbf{S} \ \mathbf{C} \ \varepsilon \ , \tag{2.78}$$

where

$$F_{\mu\nu}^{\mathrm{KS}} = \int \eta_\mu(\mathbf{r}_1) \ \hat{\mathrm{f}}^{\mathrm{KS}}(\mathbf{r}_1) \ \eta_\nu(\mathbf{r}_1) \ d\mathbf{r}_1 \tag{2.79}$$

is the Kohn-Sham matrix and

$$S_{\mu\nu} = \int \eta_\mu(\mathbf{r}_1)\eta_\nu(\mathbf{r}_1) \ d\mathbf{r}_1 \tag{2.80}$$

is called the overlap matrix, whereas the matrix $\mathbf{C}$ and the diagonal matrix $\varepsilon$ contain the expansion coefficients and orbital energies, respectively.

## 2.2.5. Basis Sets

In principle, any set of functions $\{\eta_\mu\}$ that forms a (nearly) complete basis could be used in Equation (2.77). However, only two types of basis functions are commonly used in modern electronic structure codes: (I.) Gaussian basis sets are mainly used when the electron density is localized, e.g. in molecules. (II.) Plane waves are used to describe periodic systems, such as metals that are characterized by diffuse electron clouds.

### Gaussian-Type Basis Sets

The use of Gaussian-type orbitals (GTOs) in electronic structure codes is motivated by an efficient evaluation of integrals. Since the product of two GTOs is again a GTO with modified prefactor and center, analytical expressions are available for integral handling. GTOs have the general form

$$g(\alpha, \mathbf{r}) = N x^i y^j z^k e^{-\alpha \mathbf{r}^2} \ , \tag{2.81}$$

where $N$ is a normalization constant and $x$, $y$ and $z$ are the components of $\mathbf{r}$, which ensure the proper symmetry of s, p, d, … orbitals. It is typical that a number of primitive Gaussian functions $g$ are combined to a so-called contracted Gaussian function (CGF)[126]

$$\phi_\mu^{\mathrm{CGF}}(\mathbf{r} - \mathbf{R}_A) = \sum_{p=1}^{L} d_{p\mu} g_p(\alpha_{p\mu}, \mathbf{r} - \mathbf{R}_p) \quad , \tag{2.82}$$

where $\alpha_{p\mu}$ and $d_{p,\mu}$ are the contraction exponents and coefficients of the primitive Gaussians, respectively. These parameters are commonly determined *a priori* from atomic calculations and then held fix during actual electronic structure calculations. This procedure is illustrated for the split-valence basis set 4-31G:[127] Because most chemical properties of interest are determined by the valence electrons, inner- and outer-shell orbitals are differently represented. The number 4 indicates that each core orbital is represented by one CGF which consists of 4 primitive Gaussians. The digits 31 mean that the valence orbitals are constructed from 2 CGTs, where the first (inner) function consists of 3 primitive Gaussian functions, and the outer CFG contains one Gaussian functions. Thus the 4-31G basis set consists of 2 functions for H and He, of 9 functions for Li to Ne, 13 functions for Na to Ar, etc. In the case of hydrogen, the 2 functions are

$$\phi'_{1\mathrm{s}}(\mathbf{r}) = \sum_{i=1}^{3} d'_{i,1\mathrm{s}} g_{1\mathrm{s}}(\alpha_{i,1\mathrm{s}}, \mathbf{r}) \tag{2.83}$$

$$\phi''_{1\mathrm{s}} = g_{1\mathrm{s}}(\alpha''_{1\mathrm{s}}, \mathbf{r}) \tag{2.84}$$

It turns out that just extending the number of primitive Gaussian functions in the expansion of the orbitals does not lead to balanced basis sets, if at the same time no functions of higher angular momentum are included. For instance, extending the number of primitive Gaussians beyond 6 in the expansion of the valence orbitals to form 6-31G does not lead to a significant gain in accuracy. Instead, the inclusion of higher angular momentum functions, such as adding 2p-functions to the 1s orbital of hydrogen, allows for an anisotropic electronic environment. Therefore in the nomenclature of Pople *et al.*, these functions are termed polarization functions, and indicated by a star (‘*’). One star means that polarization functions are only added to heavy atoms, *e.g.* 6-31G*, whereas two stars indicate the addition to hydrogen atoms as well, *e.g.* 6-31G**.

### Plane-Wave Basis Sets

It has already been mentioned that GTOs are particularly well-suited for isolated molecules. However, in the case of periodic systems, such as metals, the wave function is periodic as well and it is therefore necessary to account for this fact in the basis functions as well.[4,128] A natural choice is to expand the orbitals in a plane wave basis set with basis functions

$$\phi^{\mathrm{PW}}_{\mathbf{G}}(\mathbf{r}) = \frac{1}{\sqrt{\Omega}}\, e^{i\mathbf{G}\mathbf{r}} \ . \tag{2.85}$$

Here, $\Omega$ is the volume of the periodic cell and $\mathbf{G}$ is a vector in reciprocal space satisfying the periodic boundary conditions. As opposed to GTOs, the extension of a plane wave basis set can be straightforwardly controlled by the number of plane waves that are used in the expansion of the orbitals. This is controlled by the energy cutoff $E_{\mathrm{cut}}$ which limits the expansion up to a maximum $|\mathbf{G}|$-vector

$$\frac{1}{2}|\mathbf{G}|^2 \le E_{\mathrm{cut}} \ . \tag{2.86}$$

Because the considered system is periodic, the Kohn-Sham orbitals can be written as[4]

$$\phi_j(\mathbf{r},\mathbf{k}) = \exp[i\mathbf{k}\cdot\mathbf{r}]u_j(\mathbf{r},\mathbf{k}) \ . \tag{2.87}$$

The vector $\mathbf{k}$ is a vector in the first Brillouin zone and the functions $u_j$ have the periodicity of the lattice. Thus, expanding the Kohn-Sham orbitals in plane waves yields

$$\phi_j(\mathbf{r},\mathbf{k}) = \frac{1}{\sqrt{\Omega}}\sum_{\mathbf{G}} c_j(\mathbf{G},\mathbf{k}) \exp[i(\mathbf{G}+\mathbf{k})\cdot\mathbf{r}] \ . \tag{2.88}$$

Apart for their periodicity, plane waves have the advantage that the basis set size scales linearly with the system size, if used to describe the density, as opposed to quadratically in the case of GTOs. Moreover, numerically efficient algorithms for Fourier transformation, like the FFTW algorithm,[129] can be readily employed. In contrast to GTOs, plane waves are also not centered at the nuclear coordinates and therefore do not suffer from artefacts like “Pulay” forces,[130] making the gradient calculation straightforward. There is, however, also some disadvantage associated with them. Since the electronic density varies significantly in close proximity to the nuclei, a large number of plane waves would be required in this case. This would make the calculation prohibitively expensive, and therefore the use of plane waves is intimately connected to the use of pseudopotentials.

## 2.2.6. Pseudopotentials

Pseudopotentials replace the chemically inactive core electrons. Hence, the screened Coulomb potential of the nuclei, which is acting on valence electrons, and all interactions of the valence electrons with the core electrons, such as Pauli repulsion, exchange and correlation contributions, are incorporated in the pseudopotentials.[4] If properly parametrized, the use of pseudopotentials drastically simplifies the calculation of the density often with a negligible loss of accuracy. This is because only the valence electrons need to be treated explicitly, and a smaller number of plane waves is needed in order to represent the smoother pseudo-orbitals. Moreover, relativistic effects can be readily included in the potential because only the core electrons contribute significantly in this case.

A common criterion for a pseudopotential is that the density is equivalent to the ‘real’ density outside a certain cutoff radius $r_c$, and that the corresponding pseudoorbitals do not exhibit the nodal structure of the true orbitals, as depicted in Figure 2.5. Further, pseudopotentials should be transferable, *i.e.* the same potential should remain valid for all relevant atomic environments.

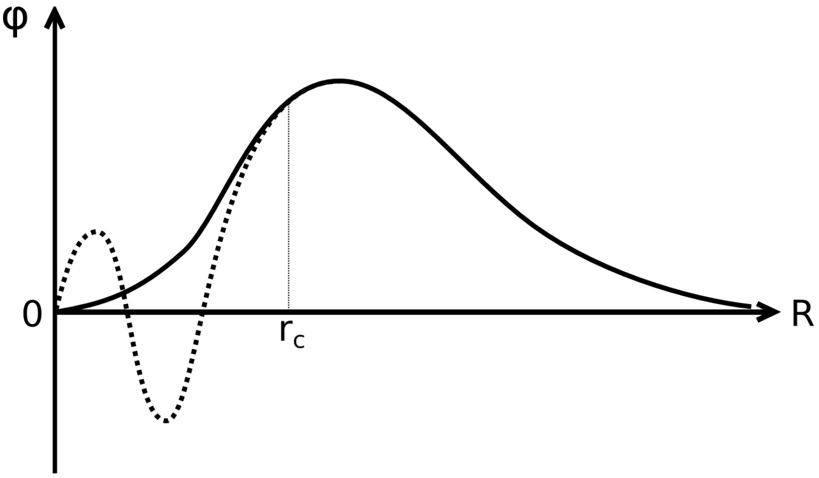


**Figure 2.5.** – Schematic depiction of a valence orbital (dashed line) and its corresponding pseudo orbital (solid line). The pseudo wave function does not exhibit the nodal structure of the real orbital in proximity to the nuclei, and both functions are identical identical beyond the cutoff radius $r_c$.

Hamann, Schlüter and Chiang (HSC) proposed a number of additional requirements that the pseudopotential and the corresponding pseudo wave functions need to satisfy:[4,131]

1. The orbital energies of the valence orbitals have to be identical for the pseudo and real Kohn-Sham orbitals
2. The pseudo and real wave functions are identical beyond a cutoff radius $r_c$
3. The pseudopotential has to be norm-conserving, i.e. the electron density within the cutoff radius $r_c$ has to be equal for pseudo and real wave functions
4. The logarithmic derivatives and first energy derivatives for pseudo and real wave functions for radii equal or greater than $r_c$ have to agree

In plane wave calculations, it is common to employ semi-local pseudopotentials $V^{\text{PP}}$ of the form

$$\begin{aligned} V^{\text{PP}}(\mathbf{r}) &= V^{\text{loc}}(r) + V^{\text{non-loc}}(r) \\ &= V^{\text{loc}}(r) + \sum_{l=0}^{l_{\max}} \sum_{m=-l}^{l} V_l(r) P_{lm}(\omega) \ . \end{aligned} \tag{2.89}$$

These pseudopotentials consist of a local and non-local term, $V^{\text{loc}}(r)$ and $V^{\text{non-loc}}(r)$, respectively, where the latter is represented by a sum over contributions from different angular momentum channels, usually s, p, d and f. The projector $P_{lm}$ projects on angular momentum functions with particular quantum numbers $l$ and $m$, and depends on the angular variables $\omega$. The local part of the pseudopotential $V^{\text{loc}}(r)$ is often chosen as

$$V^{\text{loc}}(r) = V_{l=l_{\max}}(r) \ , \tag{2.90}$$

where $l_{\mathrm{max}}$ is the highest occupied angular momentum channel in the atomic reference state that is used for parametrization. An example for a pseudopotential that falls in this class, was proposed by Goedecker, Teter, and Hutter (GTH).[132] The local and non-local part are defined as

$$V^{\mathrm{loc}}(r) = \frac{-Z_{\mathrm{ion}}}{r}\mathrm{erf}\left(\frac{\bar{r}}{\sqrt{2}}\right) + \exp\left(-\frac{1}{2}\bar{r}^2\right)\left(C_1 + C_2\bar{r}^2 + C_3\bar{r}^4 + C_4\bar{r}^6\right) \quad (2.91)$$

$$V_l^{\text{non-loc}}(\mathbf{r},\mathbf{r}') = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{m=-l}^{l} Y_{lm}(\omega)p_i^l(r)h_{ij}^l p_j^l(r)Y_{lm}^*(\omega') \quad , \quad (2.92)$$

with the Gaussian-type projectors

$$p_i^l(r) = \frac{\sqrt{2}r^{l+2(i-1)}\exp\left(-\frac{r^2}{2r_l^2}\right)}{r_l^{l+(4i-1)/2}\sqrt{\Gamma\left(l+\frac{4i-1}{2}\right)}} \quad . \quad (2.93)$$

In the expression for $V^{\mathrm{loc}}(r)$, the ionic charge is denoted with $Z_{\mathrm{ion}}$, and $\bar{r} = r/r_{\mathrm{loc}}$ is the ratio between the radial coordinate and the localization radius of the potential. $\{C_{i=1-4}\}$ are adjustable parameters. In the expression for the non-local part, $Y_{lm}$ are the spherical harmonics and $h_{ij}^l$ angular momentum dependent parameters. $\Gamma$ in Equation (2.93) is the gamma function.

## 2.3. Spectroscopic Parameters from First Principles Calculations

### 2.3.1. Calculation of NMR Chemical Shifts

If (closed-shell) molecular systems are subject to a constant magnetic field $\mathbf{B}_0$, their electron clouds start to exhibit a non-zero current density $\mathbf{j}$ producing a magnetic field $\mathbf{B}_{\mathrm{ind}}$ that is oriented anti-parallel to $\mathbf{B}_0$ at the position of the nuclei. The relation between the induced current density $\mathbf{j}(\mathbf{r})$ and the resulting additional magnetic field at position $\mathbf{s}$ is given by the Biot-Savart law, which reads in atomic units

$$\mathbf{B}_{\mathrm{ind}}(\mathbf{s}) = -\frac{1}{c^2}\int \frac{\mathbf{j}(\mathbf{r})\times(\mathbf{r}-\mathbf{s})}{|\mathbf{r}-\mathbf{s}|^3}\, d\mathbf{r} \ , \quad (2.94)$$

with $c$ being the speed of light. For nuclei with non-vanishing total spin, this effect can be indirectly measured by nuclear magnetic resonance (NMR) spectroscopy. The change of amplitude and direction of $\mathbf{B}_0$ at a nuclear position $\mathbf{R_k}$ due to the additional field $\mathbf{B}_{\mathrm{ind}}$ is expressed through the nuclear shielding tensor

$$\sigma_{ij}(\mathbf{R_k}) = -\frac{\partial B_{\mathrm{ind},i}(\mathbf{R_k})}{\partial B_{0,j}} \ . \quad (2.95)$$

In a NMR experiment, $\sigma$ can only be measured relative to an internal standard, such as tetramethylsilane (TMS) or 2,2-dimethyl-2-silapentane-5-sulfonate (DSS), whereas in quantum-chemical calculations it is obtained directly. Since the chemical shielding or shift value around a nucleus very strongly depends on its particular electronic environment, such as hybridization and hydrogen bonding situation, NMR spectroscopy is a powerful tool to study molecular systems.

For field strengths commonly used in NMR spectroscopy, the relation between $\mathbf{B}_{\text{ind}}$ and $\mathbf{B}_0$ is proportional and Equation (2.95) becomes

$$\mathbf{B}_{\text{ind}}(\mathbf{s}) = -\sigma(\mathbf{s})\mathbf{B}_0 \ . \tag{2.96}$$

Nowadays, the calculation of the chemical shielding tensor is a fairly standard procedure that is covered in many quantum-chemical text books. The following discussion on calculation of chemical shielding values is mainly based on Ref.[133]

To evaluate the current density induced by an external magnetic field in a quantum-chemical calculation, one in principle has to perform the calculation in the presence of $\mathbf{B}_0$. To this end, the momentum operator in the standard electronic Hamiltonian in the absence of external fields needs to be modified according to

$$\hat{\mathbf{p}} \rightarrow \hat{\mathbf{p}} + \mathbf{A}_0 \ . \tag{2.97}$$

$\mathbf{A}_0$ is the classical vector potential that describes the external magnetic field. If $\nabla \mathbf{A}_0 = 0$, which is always possible, the resulting Hamiltonian takes the form

$$\hat{H} = \hat{H}^{(0)} + \mathbf{A}_0\hat{\mathbf{p}} + \frac{1}{2}\mathbf{A}_0^2 \tag{2.98}$$

Unfortunately, the choice of $\mathbf{A}_0$ is still not unique and there exist many $\mathbf{A}_0$ that give rise to the same $\mathbf{B}_0$. A further possible restriction is to set the vector potential equal to

$$\mathbf{A}_0 = \frac{1}{2}\mathbf{B}_0 \times (\mathbf{r} - \mathbf{R}) \ , \tag{2.99}$$

where $\mathbf{R}$ is called the gauge origin. If one inserts Equation (2.99) into Equation (2.98) one obtains the field-dependent Hamilton operator

$$\hat{H}(\mathbf{B}_0, \mathbf{R}) = \hat{H}^{(0)} + \frac{1}{2}((\mathbf{r} - \mathbf{R}) \times \mathbf{p})\mathbf{B}_0 + \frac{1}{8}(\mathbf{B}_0 \times (\mathbf{r} - \mathbf{R}))^2 \ , \tag{2.100}$$

which also depends on the gauge origin. This dependence causes further complications as will be shown below.

As stated by Equation (2.96), the nuclear shielding tensor is independent of the magnetic field strength. For this reason, $\sigma$ can be evaluated as a small perturbation of $\hat{H}^{(0)}$, *i.e.* $\mathbf{B}_0 \rightarrow 0$. Within the framework of DFT, the effective, one-particle Kohn-Sham Hamiltonian $\hat{h}$ and the respective Kohn-Sham orbitals $\{\phi_k\}$ can then be expressed as

$$\begin{aligned} \hat{h} &= \hat{h}^{(0)} + i\hat{h}^{(1)} + \mathcal{O}(\mathbf{B}_0^2) \\ \phi_k &= \phi_k^{(0)} + i\phi_k^{(1)} + \mathcal{O}(\mathbf{B}_0^2) \ , \end{aligned} \tag{2.101}$$

where the superscripts (0) and (1) refer to the unperturbed system and to the first-order perturbation correction, respectively. Note that for real-valued Kohn-Sham orbitals, the first-order correction is purely imaginary. Since the current density vanishes in the absence of an external magnetic field, the induced current density is only expanded to first-order

$$\begin{aligned}\mathbf{j} &= \mathbf{j}^{(1)} + \mathcal{O}(\mathbf{B}_0^2)\\ &= 2\sum_{k=1}^{n}\left\{\phi_k^{(1)}\nabla\phi_k^{(0)} - \phi_k^{(0)}\nabla\phi_k^{(1)}\right\} - 2\mathbf{A}_0\sum_{k=1}^{n}\phi_k^{(0)}\phi_k^{(0)} + \mathcal{O}(\mathbf{B}_0^2)\ .\end{aligned} \tag{2.102}$$

The first term of the first-order current density is called the paramagnetic part and the second is called the diamagnetic part, because it does not require the knowledge of the $\{\phi_k^{(1)}\}$. However, note that in contrast to the total current density, this division is not unique under gauge transformation.

For the evaluation of Equation (2.102), the orbitals need to be expanded in a finite basis. The choice of the basis is obvious for $\{\phi_k^{(0)}\}$ because they are just the unperturbed Kohn-Sham orbitals. By contrast, this cannot be straight forwardly done for the $\{\phi_k^{(1}\}$. Firstly, they depend themselves on the gauge origin *via*

$$\phi_k^{(1)} \rightarrow \phi_k^{(1)} + \frac{1}{2}(\mathbf{B}_0 \times \mathbf{R})\mathbf{r}\phi_k^{(0)}\ , \tag{2.103}$$

and, secondly, the $\{\phi_k^{(1)}\}$ usually cannot be expanded in standard basis sets for any choice of the gauge origin. In practice, the gauge origin is therefore distributed over the molecular system and there are a number of schemes which are employed by standard quantum-chemical software:

- Individual Gauge for Localized Orbitals (IGLO):[134] The first-order orbitals are approximated by localized molecular orbitals with gauge centers equal to their centers of charge.

- Gauge Including Atomic Orbitals (GIAO):[135] The gauge dependence is introduced at the level of atomic basis functions. For each basis function the gauge origin is the position of the atom carrying that specific basis function.

- Continuous Set of Gauge Transformations (CSGT):[136] The first-order current density in the integral of Equation (2.94) is separately evaluated for each value of the position $\mathbf{r}$ with the gauge origin equal to the position of the atom which is closest to the particular coordinate $\mathbf{r}$.

For the results presented in this work, the GIAO method has been used whenever atom-centered basis functions were employed. Mixed atom-centered/plane wave calculations are based on the implementation by Weber *et al.*,[137] which use the IGAIM method proposed by Bader *et al.*[138] This approach is similar to the CSGT method, where in IGAIM Bader's atoms in molecules approach is used to determine the closest atomic center that is used as the gauge origin.

### 2.3.2. Calculation of Electronically Excited States

Transitions between two electronic states, *e.g.* between the ground and the first excited, are caused by electromagnetic radiation, if the photon energy matches the energy difference between these states. For conjugated organic molecules, this energy difference typically lies within the UV/VIS regime of $\approx 200-800$ nm. The relevance of electronically excited state for the current thesis lies in the altered acid-base properties of organic chromophores, such as 7-hydroxyquinoline, as outlined in Refs.[139,140] Several methods have been suggested to describe electronic excitations based on first principles, where the following discussion is based on linear-response time-dependent DFT (TDDFT). TDDFT has been proven to yield accurate absorption energies for a broad range of systems at an affordable computational cost making it amenable for routine applications, although prominent failures have been reported. The following introduction is mainly based on Refs.[141–143]

A natural starting point for the quantum-mechanical description of the interaction of a (molecular) system with a laser field is the electronic Hamiltonoperator

$$\hat{H}_{\text{elec}}(\mathbf{r},\mathbf{R}) = \hat{T}(\mathbf{r}) + \hat{V}_{\text{ee}}(\mathbf{r}) + \hat{V}_{\text{ext}}(\mathbf{r},\mathbf{R},t) \ , \tag{2.104}$$

where $\hat{T}$, $\hat{V}_{\text{ee}}$ are the standard kinetic energy and electron-electron repulsion potential operators. The time-dependent external potential operator now includes, apart from the electronic-nuclear Coulomb term, the interaction with the laser field:

$$\hat{V}_{\text{ext}}(\mathbf{r},\mathbf{R},t) = -\sum_{I=1}^{N}\sum_{i=1}^{n} \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I(t)|} + \hat{V}_{\text{laser}}(\mathbf{r},t) \tag{2.105}$$

An often made approximation for $\hat{V}_{\text{laser}}(\mathbf{r},t)$ is to treat the time-dependent laser field classically and only account for its electric component. This leads to

$$\hat{V}_{\text{laser}}(\mathbf{r},t) = Ef(t)\sin(\omega t)\sum_{i=1}^{n} \mathbf{r}_i \cdot \alpha \ . \tag{2.106}$$

Here, the vector $\alpha$ and the scalars $\omega$ and $E$ are the polarization direction, frequency and amplitude of the laser field, respectively, and the function $f(t)$ describes the pulse shape. A classical treatment of the field is justified for a high density of photons. The disregard of the magnetic component is possible, if the wavelength of the laser is much larger than the molecular system and the average distance that the electrons travels within a pulse period is small compared to the wavelength, *i.e.* $v_{\text{elec}} \ll c$, with $v_{\text{elec}}$ and $c$ being the average electron velocity and the speed of light, respectively. Moreover, the pulse period has to be small in comparison to the path the molecule travels, *i.e.* the molecule must not leave the focus during the pulse. With knowledge of the explicit form of the electronic Hamilton operator it is principally possible to propagate an initial state by solving the time-dependent Schrödinger equation. However, it is clear that this procedure can be applied to all but the smallest systems.

A slightly different view on the problem can be obtained within the framework of density functional theory. In analogy to ground-state DFT, the Runge-Gross theorem states that there is a direct mapping of the time dependent external potential

$\hat{V}_{\text{ext}}(\mathbf{r},t)$ and the time-dependent electron density $\rho(\mathbf{r},t)$.[144] Further, $\hat{V}_{\text{ext}}(\mathbf{r},t)$ uniquely determines $\hat{V}_{\text{KS}}(\mathbf{r},t)$, the time-dependent Kohn-Sham potential of a non-interacting reference system, whose density equals that of the interacting system. In cases where the potential due to the laser is small, which can be assumed for most of the standard UV/VIS experiments, it is justified to treat the laser potential as a linear perturbation, *i.e.*

$$\hat{V}_{\text{ext}}(\mathbf{r},t) = \hat{V}^{(0)}_{\text{ext}}(\mathbf{r}) + \hat{V}^{(1)}_{\text{ext}}(\mathbf{r},t) \tag{2.107}$$

The first order change in the density can then be calculated as

$$\rho^{(1)}(\mathbf{r},t) = \int\int \chi(\mathbf{r},t,\mathbf{r}',t')\hat{V}^{(1)}_{\text{ext}}(\mathbf{r}',t')\ d^3r'dt'\ . \tag{2.108}$$

An important property of the linear density response function $\chi$ is that its Fourier transform with respect to time has discrete poles at the excitation energies $\Delta E_m = E_m - E_0$ of the unperturbed system. To calculate these poles within TDDFT, one first obtains the excitation energies of the corresponding non-interacting Kohn-Sham reference system and subsequently calculates their shift towards the true excitation energies $E_m$. To this end, the linear response density is expressed analogously to Equation (2.108) in terms of the non-interacting reference system

$$\rho^{(1)}(\mathbf{r},\omega) = \int \chi_{\text{KS}}(\mathbf{r},\mathbf{r}',\omega)\hat{V}^{(1)}_{\text{KS}}(\mathbf{r}',\omega)\ d^3r'\ , \tag{2.109}$$

where the Kohn-Sham linear response function $\chi_{\text{KS}}$ takes the form

$$\chi_{\text{KS}}(\mathbf{r},\mathbf{r}',\omega) = \lim_{\eta\to 0^+}\sum_{jk}^{\infty}(f_k - f_j)\frac{\phi_j(\mathbf{r})\phi_j^*(\mathbf{r}')\phi_k(\mathbf{r}')\phi_k^*(\mathbf{r})}{\omega - (\epsilon_j - \epsilon_k) + i\eta}\ . \tag{2.110}$$

Here, $f_m$ is the occupation number of the $m$th orbital, $\phi_m$, in the Kohn-Sham ground state with orbital energy $\epsilon_m$. The time-dependent first-order Kohn-Sham potential can be decomposed into

$$V^{(1)}_{\text{KS}} = V^{(1)}_{\text{ext}}(\mathbf{r},t) + V^{(1)}_{\text{H}}(\mathbf{r},t) + V^{(1)}_{\text{xc}}(\mathbf{r},t) \tag{2.111}$$

The first order change of $V_{\text{ext}}$ is $V^{(1)}_{\text{ext}}$, where the changes in the Hartree and exchange-correlation are given as

$$V^{(1)}_{\text{H}}(\mathbf{r},t) = \int \frac{\rho^{(1)}(\mathbf{r}',t)}{|\mathbf{r}-\mathbf{r}'|}\ d^3r' \tag{2.112}$$

and

$$\begin{aligned} V^{(1)}_{\text{xc}}(\mathbf{r},t) &= \int\int \frac{\delta V_{\text{xc}}(\mathbf{r},t)}{\delta\rho(\mathbf{r}',t')}\rho^{(1)}(\mathbf{r}',t')\ d^3r'\ dt' \\ &= \int\int f_{\text{xc}}(\mathbf{r},t,\mathbf{r}',t')\rho^{(1)}(\mathbf{r}',t')\ d^3r'\ dt' \end{aligned} \tag{2.113}$$

Fourier transforming the above first-order potentials with respect to time and plugging them into the expression of the response density, Equation (2.109), yields

$$\begin{aligned}\rho^{(1)}(\mathbf{r},\omega) &= \int \chi_{\text{KS}}(\mathbf{r},\mathbf{r}',\omega) V^{(1)}_{\text{ext}}(\mathbf{r}',\omega)\; d^3r' \\ &\quad + \int\int \chi_{\text{KS}}(\mathbf{r},\mathbf{x},\omega) \left[\frac{1}{|\mathbf{x}-\mathbf{r}'|} + f_{\text{xc}}(\mathbf{x},\mathbf{r}',\omega)\right] \rho^{(1)}(\mathbf{r}',\omega)\; d^3r' d^3x\end{aligned} \tag{2.114}$$

This equation can be recast into the form

$$\int [\delta(\mathbf{r}-\mathbf{r}') - \Xi(\mathbf{r},\mathbf{r}',\omega)] \rho^{(1)}(\mathbf{r}',\omega)\; d^3r' = \int \chi_{\text{KS}}(\mathbf{r},\mathbf{r}',\omega) V^{(1)}_{\text{ext}}(\mathbf{r}',\omega) d^3r' \;, \tag{2.115}$$

where the function $\Xi$ is given by

$$\Xi(\mathbf{r},\mathbf{r}',\omega) = \int \chi_{\text{KS}}(\mathbf{r},\mathbf{r}'',\omega) \left[\frac{1}{|\mathbf{r}''-\mathbf{r}'|} + f_{\text{xc}}(\mathbf{r}'',\mathbf{r}',\omega)\right]\; d^3r'' \tag{2.116}$$

From the pole structure of the linear response function $\chi$ at the excitation frequencies $\Omega_m$ it follows that the linear density response $\rho^{(1)}(\mathbf{r},\omega)$ exhibits poles at $\omega = \Omega_m$ as well. Note that it can be shown that the poles of $\chi$ and $\chi_{\text{KS}}$ are located at different frequencies. In order to Equation (2.115) to hold, the integral operator acting on the response density on the right hand side needs to have zero eigenvalues at frequencies $\Omega_m$, or equivalently $\Xi$ in Equation (2.116) needs to have eigenvalues equal to one at the excitation frequencies. Therefore, if $\chi_{\text{KS}}$ and $f_{\text{xc}}$ would be known, it is in principle possible to obtain the excitation frequencies by solving the eigenvalue problem Equation (2.116). Most practical applications, however, do not use this equation directly, and instead (approximately) solve a set of coupled equations known as Casida equations, which can be derived directly from Equation (2.116).[145]

To conclude this part, some remarks are made on the frequency dependent exchange-correlation kernel $f_{\text{xc}}(\mathbf{r},\mathbf{r}',\omega)$. This quantity is a key ingredient for TDDFT and therefore the performance of the method will crucially dependent on the quality of approximations that are used in practice. One of the most common approximations is the adiabatic approximation

$$f^{\text{A}}_{\text{xc}}(\mathbf{r},\mathbf{r}') = \frac{\delta V^{(0)}_{\text{xc}}[\rho^{(0)}](\mathbf{r})}{\rho^{(0)}(\mathbf{r}')} \tag{2.117}$$

$$= \frac{\delta^2 E_{\text{xc}}[\rho^{(0)}]}{\delta\rho^{(0)}(\mathbf{r})\delta\rho^{(0)}(\mathbf{r}')} \tag{2.118}$$

The adiabatic kernel is frequency independent and real-valued and corresponds to the replacement of the exact kernel by its static limit at each frequency $\omega$. As a matter of fact, Equation (2.118) allows for the use of standard ground exchange-correlation functionals in TDDFT. This approximation leads to reasonably accurate excitation frequencies for a broad range of molecular systems. One of the reasons is that the adiabatic approximation fulfills the essential requirement that the frequency dependent kernel has to match the static exchange-correlation potential at the initial time $t = t_0$,[143] which is trivially satisfied by Equation (2.118).

### 2.3.3. Calculation of Infrared Spectra

Excitation of molecular systems with electromagnetic radiation in the infrared (IR) regime, i.e. between 700 and 1000 nm, drives transitions between vibrational states. By probing these resonances with IR spectroscopy it is possible to extract information on functional groups, molecular composition, protein secondary structure content, etc. from molecular systems. Further, its comparatively simple technical implementation makes it amenable to a broad range of systems and therefore this spectroscopy class is a widely applied tool in chemistry.

The calculation of IR resonances is nowadays a fairly standard task in computational chemistry. Considering molecular systems, there exists in principle two different flavors: On the one hand, IR resonances can be obtained from normal mode analysis augmented with higher order perturbative corrections. On the other hand, the IR spectrum can be calculated from the dipole autocorrelation function that can be *e.g.* obtained from *ab initio* MD simulations. Here, one approximates the quantum-mechanical dipole autocorrelation function whose Fourier transform is the spectral density

$$I(\omega) = \frac{1}{2\pi} \int \exp(-i\omega t)\, \langle \hat{\mathbf{M}}(0)\hat{\mathbf{M}}(t) \rangle_{\mathrm{NVT}} \; dt \; . \tag{2.119}$$

The dipole autocorrelation function $\langle \hat{\mathbf{M}}(0)\hat{\mathbf{M}}(t) \rangle_{\mathrm{NVT}}$ involves the time-dependent dipole operator $\hat{\mathbf{M}}(t)$, whose autocorrelation function is obtained as an average over the canonical ensemble. Equation (2.119) is readily derived in the framework of first-order time-dependent perturbation theory, where the field due to the impinging laser beam is treated within the electric dipole approximation, cf. Section 2.3.1. It is clear that Equation (2.119) has to be approximated for practical applications. The most obvious choice for *ab initio* MD simulations is

$$C_{\mathrm{QM}}(t) = \langle \hat{\mathbf{M}}(0)\hat{\mathbf{M}}(t) \rangle \tag{2.120}$$

$$\approx \langle \mathbf{M}(0)\mathbf{M}(t) \rangle = C_{\mathrm{Cl}}(t) \; , \tag{2.121}$$

where the quantum-mechanical autocorrelation function has been replaced by its classical counterpart. However, it is well known that the classical and quantum-mechanical correlation functions and their associated spectral densities feature different symmetries:

$$I_{\mathrm{QM}}(\omega) = \exp(\beta\hbar\omega) I_{\mathrm{QM}}(-\omega) \tag{2.122}$$

$$I_{\mathrm{Cl}}(\omega) = I_{\mathrm{Cl}}(-\omega) \tag{2.123}$$

Notably, the exponential prefactor reduces to 1 in the high temperature limit $\beta = 1/k_B T \rightarrow 0$. This symmetry property is known as "detailed balance" condition. A straightforward way to improve the approximation made in Equation (2.121) is therefore to restore detailed balance by introducing a suitable correction factor to the classical autocorrelation function. Ramírez *et al.* noted that the harmonic correction factor

$$Q_{\mathrm{HA}}(\omega) = \frac{\beta\hbar\omega}{1-\exp(-\beta\hbar\omega)} \tag{2.124}$$

can be motivated from the analogy between the quantum-mechanical and classical fluctuation-dissipation theorems of the dipole moment in the presence of a small time-dependent electric field $E(t)$ parallel to $\mathbf{M}$ at times $t' \leq t$:[146]

$$\langle \mathbf{M}(t) \rangle_{\mathrm{E}} = \langle \mathbf{M}(0) \rangle + \int_{-\infty}^{t} \Phi(t-t')E(t')\ dt' + \mathcal{O}(E^2)\ , \tag{2.125}$$

where $\mathbf{M}$ is either the dipole operator or the classical dipole moment. In the quantum-mechanical case $\Phi$ takes the form

$$\begin{aligned} \Phi_{\mathrm{QM}} &= -\beta \frac{d}{dt} \left\langle \frac{1}{\beta} \int_0^{\beta} \hat{\mathbf{M}}(0)\hat{\mathbf{M}}(t+i\hbar\lambda)\ d\lambda \right\rangle & (2.126) \\ &= -\beta \frac{d}{dt} C_k(t)\ , & (2.127) \end{aligned}$$

with the Kubo-transformed dipole autocorrelation function $C_k(t)$. Note that $C_k(t)$ and $C_{\mathrm{QM}}(t)$ are strictly identical, but possess different symmetry properties with respect to the time axis.[146] In the classical case, $\Phi$ is simply obtained by replacement of $C_k$ with $C_{\mathrm{Cl}}$:

$$\Phi_{\mathrm{Cl}} = -\beta \frac{d}{dt} C_{\mathrm{Cl}}(t) \tag{2.128}$$

Therefore the classical approximation to Equation (2.119) reads

$$I(\omega) \approx \frac{1}{2\pi} Q_{\mathrm{HA}}(\omega) \int \exp\left(-i\omega t\right) C_{\mathrm{Cl}}(t)\ dt\ . \tag{2.129}$$

For evaluation of $I(\omega)$ in *ab initio* MD simulations, *i.e.* under periodic boundary conditions, the system's total dipole moment is decomposed into a nuclear and electronic part. The nuclear part, $\mathbf{M}_{\mathrm{nuc}}$, is readily calculated from the nuclear charge and the position of the ion. By contrast, calculation of $\mathbf{M}_{\mathrm{ele}}$ is less straightforward because it needs in principle to be evaluated as the expectation value of the dipole operator in the given electronic state. Due to periodic boundary conditions it is not possible to define the dipole moment operator in terms of the product of charge and position operator, $e\hat{\mathbf{x}}$, since the state that results from action on a periodic wave function will not have the symmetry of the lattice. Instead, a well-established approach is to use the centers of maximally localized Wannier functions which are obtained from a unitary transformation of the Kohn-Sham orbitals.[147–149] As a result, the total dipole moment is obtained as

$$\begin{aligned} \mathbf{M} &= \mathbf{M}_{\mathrm{nuc}} + \mathbf{M}_{\mathrm{ele}} & (2.130) \\ &= e \sum_{j \in I} n_{j,val} \cdot \mathbf{r}_j^{\mathrm{nuc}} - 2e \sum_{i \in I} \mathbf{r}_i^{\mathrm{Wannier}}\ , & (2.131) \end{aligned}$$

with $\mathbf{M}_{\text{nuc}}$ and $\mathbf{M}_{\text{ele}}$ denoting the nuclear and electronic contribution to the total dipole moment. $\mathbf{r}_j^{\text{nuc}}$ and $\mathbf{r}_j^{\text{Wannier}}$ are the positions of the ions and Wannier centers, respectively, while $e$ denotes the elementary charge. The number of valence electrons of atom $j$ is $n_{j,val}$.

# 3. Overview over the Published Papers

## 3.1. Protonation Dynamics Triggered by Electronic Excitation of Photoacids and -Bases in Aqueous and Methanol Solution

Publication P1 deals with the microscopic mechanism of proton release from *N*-methyl-6-hydroxyquinolinium (6MQ) into bulk aqueous solution. Electronic excitation of 6MQ to the $S_1$ state leads to an immediate increase in acidity at its hydroxyl moiety, making 6MQ one of the strongest known photoacids.[150–153] The main goal of this project was to computationally investigate the proton pathways at the molecular level and to determine characteristic infrared absorptions of 6MQ during the process. To this end, we carried out *ab initio* MD simulations of 6MQ in explicit water solution, where, after an initial equilibration phase in the ground state, 6MQ was transferred to the excited state. In the interest of numerical efficiency, the dynamics on the true $S_1$ potential was approximated by the first triplet state $T_1$. Note that although the symmetry of the $T_1$ state differs from the true situation, quantities like the dipole moment of 6MQ and IR absorptions are only marginally affected by this approximation.[154] Moreover, the use of *ab initio* MD simulations guarantees an accurate representation of hydrogen bonding that is especially crucial for the investigation of proton transfer reactions in aqueous solution. In addition, accurate IR absorptions are readily available *via* the dipole moment autocorrelation function, obtained from maximally localized Wannier centers.[4] To elucidate PT, we firstly focussed on the situation directly after photoexcitation. It was found that in the first picoseconds after excitation the hydrogen bond between 6MQ and the proton accepting water molecule strengthens. Structurally, this could be observed from an additional peak in the radial distribution function of the 6MQ hydrogen and the surrounding water oxygen atoms. Spectroscopically, this hydrogen bond strengthening gave rise to a broad continuous IR absorption band between 1500 and 2000 $\text{cm}^{-1}$, shown in Figure 3.1. Already 1-2 ps after excitation, the proton was fully dissociated from the photoacid, resulting in the depletion of the continuous IR band, and in an altered absorption pattern in the fingerprint region from 1200 to 1500 wavenumbers (*cf.* inset in Figure 3.1). Moreover, the simulations show that the rate-limiting step for proton dissociation crucially depends upon the hydrogen bonding configuration around the proton accepting water molecule. An concerted proton transfer reaction was observed in cases where the accepting water molecule was fourfold coordinated, whereas a threefold coordination lead to a sequential proton transfer.

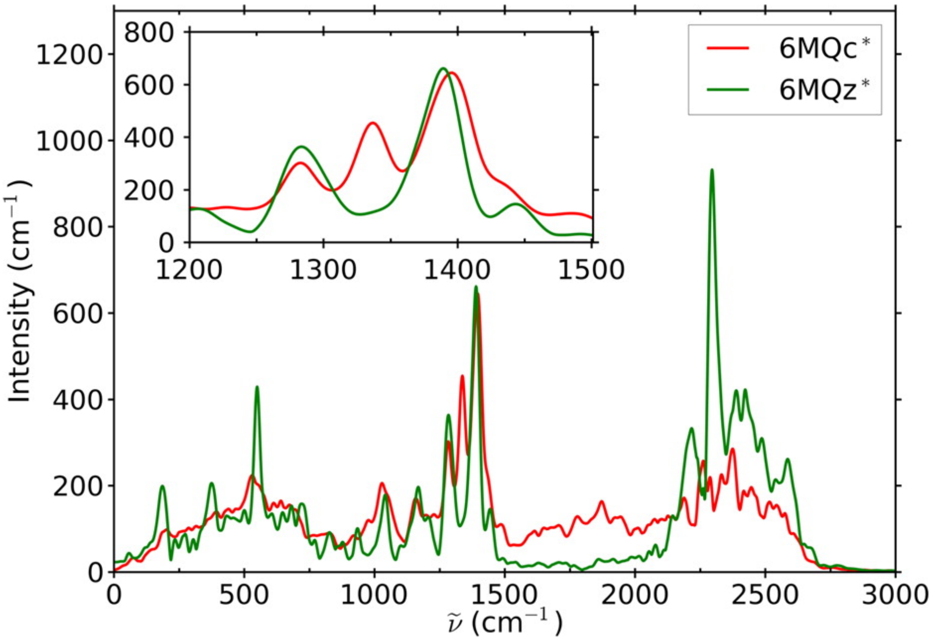


**Figure 3.1.** – Infrared spectrum of 6MQ directly after excitation (6MQc*) and after proton dissociation (6MQz*). Adapted with permission from Ref.[139] Copyright 2017 American Chemical Society.

In publication P2, the influence of distinct hydrogen bonded configurations on the microscopic charge transport was further investigated. For this purpose, we focussed on the bifunctional chromophore 7-hydroxyquinoline (7HQ).[37–39] 7HQ exhibits both photoacidic and -basic functionalities at a well-defined intramolecular distance. This allows for the formation of hydrogen bonded solvent chains that are significantly more stable than corresponding configurations in the bulk (Figure 3.2).[40] Due to its bifunctional character, 7HQ can adopt four different protonation states: The neutral (N), zwitterionic (Z), cationic (C), and anionic (A) state. Electronic excitation of 7HQ leads to the shift of the equilibrium from N to Z, where the reaction can either proceed *via* A or C. As a consequence, the tautomerization reaction can either involve the transfer of a proton (*via* A) or a negative charge, *i.e.* a hydroxyl or alcoholate ion (*via* C). In principle, the reaction can be experimentally followed by recording transient IR-active transitions, so-called marker modes, of the various protonation forms in a pump-probe experiment.[18] One difficulty, however, is that the concentration of the intermediate A or C states has to be high enough to give rise to an appreciable contribution to the overall IR signal. To achieve this, the pH value of the methanol solution was chosen such that the C or A state was already predominantly formed in the ground state. Next, the conversion into Z was probed by transient IR spectroscopy. From the reaction times summarized in Table 3.1 it becomes apparent that the conversion from A to Z is much longer than the overall N to Z reaction in neutral solution. Moreover, conversion of C is twice a fast as from N to Z suggesting that the overall mechanism preferentially involves C as an intermediate. To corroborate the observed marker bands, we further carried out a normal mode analysis of resonances in the fingerprint region. It was found that the measured absorptions are in good

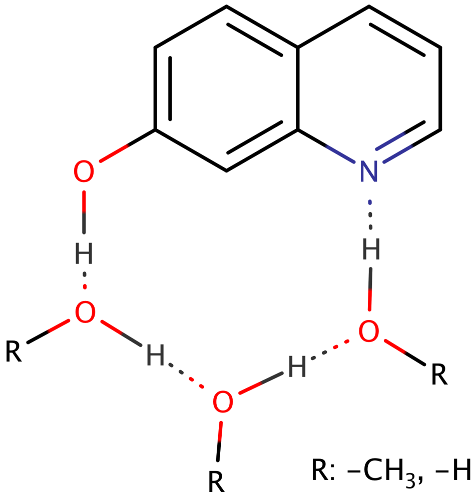


**Figure 3.2.** – 7-Hydroxyquinoline stabilizing a hydrogen bonded chain of three solvent molecules, consisting of either water and/or methanol.

agreement with those obtained from the calculations. Analysis also revealed that the vibrations in this region are well separated from solvent degrees of freedom, where small frequency shifts and changes in IR intensity of the absorption peaks result in characteristic spectroscopic patterns that can be used to discriminate the various protonation states.

**Table 3.1.** – Reaction Times Probed by Transient IR Spectroscopy

| Reaction | Reaction Time (ps) |
|---|---|
| N → Z | 330 |
| C → Z | 170 |
| A → Z | 600 |

## 3.2. Prediction of NMR Chemical Shifts of Metabolite Molecules in Aqueous Solution

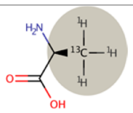
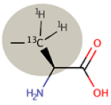
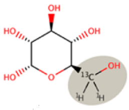
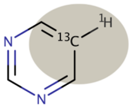
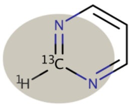
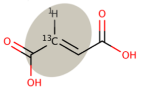
Over the last years, NMR spectroscopy has become increasingly popular for the analysis of complex metabolic mixtures resulting in a concomitant demand for more accurate NMR chemical shift predictors to identify yet unknown metabolites.[58–61] For this reason, we have suggested a new DFT based approach which significantly improves the prediction accuracy over common quantum-chemical approaches. One of the key challenges in this context is that measurements of metabolites are conducted in aqueous solution where strong solute-solvent interactions, such as hydrogen bonds, are known to have a large impact on chemical shift values. Although the explicit consideration of water molecules in quantum-chemical predictions is technically feasible, it is often computationally too demanding to be applied to the large number of candidate compounds typical for NMR metabolomics studies. Therefore in the present case, an implicit solvation model was adopted instead and systematic errors, *e.g.* those resulting from an improper description of hydrogen bonding, were accounted for *a posteriori* by applying motif-specific corrections.

The present prediction approach consists of several steps. Firstly, starting from the 2D structure of a chemical compound, a 3D conformational ensemble is generated by carrying out a Monte Carlo Multiple Minimum conformational search[155,156] using classical force field potentials. Secondly, each conformer is refined at the DFT level and its isotropic shielding value is calculated. Next, an ensemble averaged shielding value of the compound is obtained as a Boltzmann average over the conformers, where the respective weights are calculated from a thermochemical analysis (*i.a.* including zero point vibrational effects). Finally, the isotropic shielding values, $\sigma_{\mathrm{iso}}$, are converted into chemical shifts, $\delta$, according to

$$\delta = \sigma_{\mathrm{iso}} a_i + b_i \ , \tag{3.1}$$

where $a_i$ and $b_i$ are model parameters for motif $i$. For the current approach, six distinct motifs have been used depending on the direct environment of the carbon atom whose chemical shift is predicted. Note that since theoretical values are compared to $^{1}$H-$^{13}$C HSQC measurements, only carbon atoms directly bonded to at least one hydrogen are considered. Table 3.2 presents the definitions, names, and examples of the molecular motifs. The motivation behind these particular motifs was in parts based on the observation that carbon atoms directly bonded to a heteroatom, such as nitrogen, oxygen, or sulfur, exhibit larger systematic errors. This is partially a direct consequence of the lack of solute-solvent hydrogen bonds in the model. Moreover, it is has previously been found and was observed in our case as well that shielding values of aromatic carbon atoms exhibit a slightly different slope as a function of the experimental shift values than aliphatic carbons, suggesting different systematic errors in this case. In addition, methyl groups not bonded to any hetero atom show much lower prediction errors than other functional groups.

**Table 3.2.** – Definition, Names, and Examples of the Molecular Motifs Used in the Linear Regression. Adapted with permission from Ref.[157] Copyright 2017 American Chemical Society.

| Motif Definition: | *Methyl groups* not bonded to a heteroatom | *sp$^3$ carbon centers* not bonded to a heteroatom, excluding methyl groups | *sp$^3$ carbon centers* bonded to a heteroatom | *Aromatic carbon centers* not bonded to a heteroatom | *Aromatic carbon centers* bonded to a heteroatom | *sp$^2$ carbon centers* |
|---|---|---|---|---|---|---|
| Motif Name: | $\mathrm{sp}^3_{nonhet,\ met}$ | $\mathrm{sp}^3_{nonhet}$ | $\mathrm{sp}^3_{het}$ | $\mathrm{ar}_{nonhet}$ | $\mathrm{ar}_{\mathrm{het}}$ | $\mathrm{sp}^2$ |
| Examples: |  |  |  |  |  |  |

In total, 176 molecules have been used in the linear regression. The total prediction error in terms of root mean square deviations was 1.93 *vs.* 2.36 ppm for $^{13}$C and 0.154 *vs.* 0.170 ppm for $^{1}$H motif-specific and global corrections, respectively. It should be noted that although the overall improvement over the global correction might not seem substantial, the motif-specific approach allows for a more detailed estimation of prediction accuracy by attaching more or less weight to motifs with lower or higher training set RMSD values (*cf.* Table 3.3). For instance, when used for pruning the number of candidate metabolite molecules, the margin for deviations of the $\mathrm{sp}^3_{\mathrm{nonhet,met}}$ should be chosen smaller than for the $\mathrm{sp}^2$ motif because the RMSD value of the latter is three times as high as the former.

**Table 3.3.** – Root Mean Square Deviations (in ppm) of Predicted Chemical Shifts w.r.t. Experiment

| Nucleus | $\mathrm{sp}^3_{\mathrm{nonhet,met}}$ | $\mathrm{sp}^3_{\mathrm{nonhet}}$ | $\mathrm{sp}^3_{\mathrm{het}}$ | $\mathrm{ar}_{\mathrm{nonhet}}$ | $\mathrm{ar}_{\mathrm{het}}$ | $\mathrm{sp}^2$ |
|---|---|---|---|---|---|---|
| $^{13}$C | 1.15 | 1.31 | 1.76 | 2.05 | 2.32 | 3.03 |
| $^{1}$H | 0.079 | 0.118 | 0.239 | 0.146 | 0.177 | 0.214 |

It should be finally mentioned that for the current training set the performance of the DFT-based approach is superior over the empirical NMRPredict software[66] ($\mathrm{RMSD}_{\mathrm{NMRPredict}}$ 3.15/0.255 ppm). Therefore in metabolite identification, the latter should be used first to generate candidate compounds, whereas the DFT-based prediction should be used for a further pruning.

## 3.3. Impact of the $\mathrm{Phe}_{19}$-to-$\mathrm{Lys}_{19}$ Mutation in Aβ(1-40) on the Fibril Structure

The goal of this project was to understand structural changes of amyloid β (1-40) (Aβ(1-40)) fibrils as a result of the $\mathrm{Phe}_{19}$-to-$\mathrm{Lys}_{19}$ mutation. Misfolding of Aβ into amyloid fibrils is one of the hallmarks of Alzheimer's disease (AD), where Aβ fibrils constitute the main component of so-called senile plaques, proteinaceous, extracellular deposits in the cerebral cortex of AD patients' brains. Investigations of Aβ fibrils at the atomic level are inherently difficult as they do not exhibit long-range order, precluding the structural elucidation with X-ray crystallography. Moreover, they tend to be polymorphic, *i.e.* their actual conformational structure highly depends on particular growing conditions, like pH value, salt concentration, and temperature.[45] Sometimes this even leads to coexistence of several polymorphs in the same sample. Despite these difficulties, a number of atomic models of Aβ have been proposed based on solid-state NMR measurements or, more recently, on cryo-EM experiments.[46–53] However, many questions remain unanswered. For instance, the relation between local structural motifs, such as salt bridges or hydrophobic contacts, and cytotoxicity is still elusive. While nowadays accepted that smaller, transient aggregates of Aβ are the most toxic species, it is not clear how this property structurally manifests. In this context, Das *et al.* recently investigated a series of non-physiological mutations altering the $Phe_{19}$-$Leu_{34}$ contact in Aβ(1-40) fibrils.[57] Surprisingly, all peptide variants turned out being completely non-toxic, although all but one still formed fibrils with similar secondary structure content.[44,56] This leads to the conclusion that already subtle changes in the peptide sequence cause tremendous differences in physiological properties, but comparably small changes of aggregation behavior. The aim of the present work was to further understand the structural effects of these mutations. To this end, we studied a particular variant of the database, namely $Phe_{19}$-to-$Lys_{19}$ Aβ(1-40) in its fibril state.

By combining structural alignment modeling with MD simulations, two fibril models, M1 and M2, have been created, depicted in Figure 3.3. Note that in both models, the positively charged lysine is pointing towards the solvent, whereas the original phenylalanine in the WT is directed inside the hydrophobic fibril interior. Starting point for the modeling was in both cases the wild type (WT) Aβ(1-40) fibril structure proposed by Tycko *et al.*.[158] The mutant models differ with respect to each other in the side chain orientation in proximity to the mutated site. In the case of model M1, the entire lower β-sheet, involving residues 9 to 23, is flipped by 180° about the peptide strand axis compared to the WT structure. On the other hand in model M2, only peptide backbone dihedral angles in vicinity of position 19 have been changed, resulting in a locally disordered backbone.

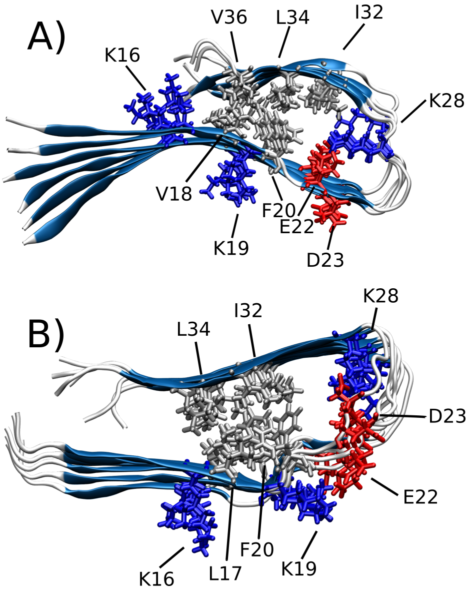


**Figure 3.3.** – Structural models M1 (A) and M2 (B) of mutated $Phe_{19}$-to-$Lys_{19}$ Aβ(1-40) fibrils after 600 ns of MD simulation. Key residues are drawn in stick representation. Hydrophobic, positively, and negatively charged side chains are highlighted in silver, blue, and red, respectively. Adapted with permission from Ref.[159] Copyright 2017 American Chemical Society.

Experimentally, the fibril structure of $Phe_{19}$-to-$Lys_{19}$ Aβ(1-40) has been analyzed by solid-state NMR. Due to the low natural abundance of $^{13}C$, the sample has been isotopically labeled, giving only access to a limited number of residues at the same time. Side chain contacts appear as cross peaks in $^{13}C$-$^{13}C$ DARR NMR spectra resulting from strong dipolar coupling interactions. As indicated in Figure 3.4A), a contact between the isotopically labeled residues $Phe_{20}$ and $Leu_{34}$ has been observed. This could be computationally confirmed by evaluating a distance criterion over the course of the MD simulations. From the populations of side chain contacts presented in Figure 3.4B) it becomes apparent that the $Phe_{20}$ and $Leu_{34}$ contact in model M1 is formed during the entire simulation time of 400 ns, whereas in model M2 it is only populated about 20 %. This is a strong indication in favor of the local side chain orientation in model M1 providing stark evidence that the $Lys_{19}$ side chain is indeed directed outside the hydrophobic fibril interior. It should be emphasized that the observed $Phe_{20}$/$Leu_{34}$ contact has not been reported so far in any physiological Aβ(1-40) variant. Instead, the $Phe_{19}$/$Leu_{34}$ contact has been detected in fibrils as well as smaller oligomers[160,161] and protofibrils.[55] Thus, it seems indicative that the change of cytotoxicity results from the perturbation at position 19, offering alternative strategies for the treatment of Alzheimer's disease by specifically targeting the $Phe_{19}$/$Leu_{34}$ contact.

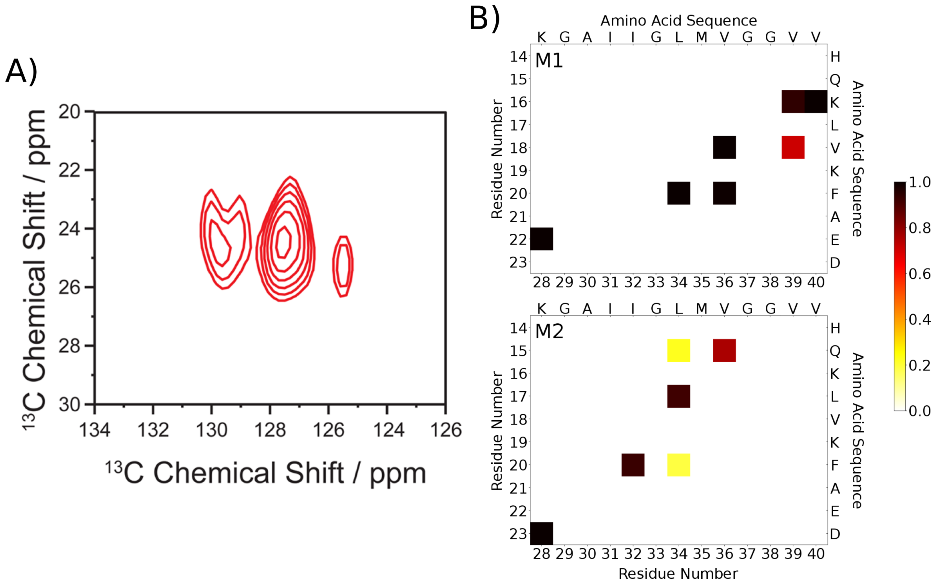


**Figure 3.4.** – A) Section of a $^{13}$C—$^{13}$C DARR NMR spectrum of $Phe_{19}$-to-$Lys_{19}$ Aβ(1-40) fibrils indicating the side chain contact between $Phe_{20}$ and $Leu_{34}$. B) Contact maps showing the time averaged population of side chain contacts between residues belonging to two opposing β-strands in models M1 and M2. A contact was defined by a distance of $\leq$ 3.2 Å between any two atoms of the residues. Adapted with permission from Ref.[159] Copyright 2017 American Chemical Society.

# 4. Conclusion

This thesis deals with the complex competition between hydrophilic and hydrophobic interactions in molecular systems and the resulting structural, dynamical as well as spectroscopic signatures. In small systems, comprised of a few hundreds of atoms, the atomistic details of protonation dynamics in aqueous solution could be successfully determined as a function of picosecond-fluctuations of the hydrogen bond network. In much larger biochemical systems, *i.e.* amyloid β peptide fibrils, the conformational statistics could be related to the interaction between specific peptide side chains, in particular in view of their polarity. For both small and large systems, the computational prediction of spectroscopic fingerprints is crucial for the validation and an in-depth understanding of structure and dynamics. For this purpose, this thesis addresses very complementary spectroscopic techniques, namely time-resolved infrared and NMR spectroscopy.

The computational investigations covered in this thesis were conducted by using molecular dynamics simulations as the primary methodology. While *ab initio* molecular dynamics simulations permitted insights into the complex hydrogen bonding dynamics and charge transfer phenomena in bulk solutions on the picosecond time scale, more approximate force field molecular dynamics simulations allowed for the treatment of peptide aggregates in explicit solvent on time scales in the nanosecond regime. Inherent to both simulation techniques is that thermodynamic and spectroscopic quantities could be readily computed in well-defined ensembles and compared to measurements. Hence, direct relationships between the underlying atomic or molecular structure and dynamics of model systems and experimental observables have been established.

The first part of the thesis addresses laser-triggered proton transfer processes in aqueous and methanol solutions. For the investigated systems, impinging UV laser light leads to the electronic excitation of co-solvated photoacids and -bases which in turn alter the pH value of the solution. In particular, the excitation of $N$-methyl-6-hydroxyquinolinium (6MQ) results in a very drastic increase in acidity, making it one of the strongest known photoacids. In the present work, it was found that the dissociation mechanism of 6MQ crucially depends on the specific hydrogen bonding configuration of the proton accepting water molecule. More precisely, a fourfold coordination leads to a nearly concerted transfer to the neighboring solvent molecule, whereas in the case of a threefold coordination the proton remains at the initial water molecule for typically 1-5 ps. Moreover, the deprotonation of 6MQ is accompanied by characteristic changes of infrared resonances that have been calculated from *ab initio* molecular dynamics trajectories. Remarkably, the computed IR spectra reveal a broad continuous absorption band emerging immediately after excitation of 6MQ in the range from 1500 to 2000 $\mathrm{cm}^{-1}$. This spectral feature was shown to result from the anharmonic vibration of the loosely attached proton at the chromophore. After completed photoacid dissociation, additional distinct absorption changes of 6MQ in the fingerprint region between 1300 and 1400 $\text{cm}^{-1}$ have been identified that might be used as markers in pump-probe type infrared experiments. In a next step, to investigate spacial constraints on proton transfer, the chromophore was changed to 7-hydroxyquinoline (7HQ). 7HQ is a bifunctional chromphore, *i.e.* it carries both photoacidic and -basic moieties at a well-defined intramolecular distance. Here, the transfer reaction in methanol solution was computationally investigated. By analyzing normal mode frequencies of 7HQ in its different charge states, the experimental assignment of marker vibrations could be rationalized in terms of frequency shifts of distinct vibrational modes as a result of (de-)protonation. Due to the bifunctional character of 7HQ, the reaction can either proceed *via* the transfer of a proton or a methanolate ion. While experimental results suggest a sequential transfer of a methanolate molecule in bulk solution, theoretical activation energies for a concerted reaction along a hydrogen bonded chain of three methanol molecules indicate a proton transfer, demonstrating the intricate dependence of charge transfer processes on the particular solvent configurations.

In the second part of the thesis, a numerically efficient approach for the prediction of nuclear magnetic resonance (NMR) chemical shifts has been devised. The aim of this work was to enhance the prediction accuracy for small to medium sized metabolite molecules in aqueous solution over common quantum-chemical prediction schemes. For this purpose, molecular motif-specific corrections were applied to calculated isotropic shielding values in order to improve the accuracy of the resulting chemical shifts. Depending on the direct environment of the query nucleus, the chemical shift was computed by scaling and shifting the shielding value according to motif-specific parameters. These parameters were obtained from linear regression analysis of experimental shifts with respect to calculated shielding values for a database comprising more than 170 compounds. In the interest of numerical efficiency, only a limited number of low energy conformers was considered for each compound, as opposed to the full ensemble, which is *e.g.* accessible from molecular dynamics simulation trajectories. To further reduce the numerical effort, solvent-solute interactions were only implicitly taken into account. The introduced approach outperforms molecular-motif independent corrections as well as empirical predictors. Most importantly, it allows for a more detailed estimation of prediction accuracy. For instance, the excellent agreement obtained for methyl and aromatic $^{13}$C and $^{1}$H that are not directly bonded to a heteroatom, such as O, N, S, or P, can be exploited in the calculation of matching scores for metabolite identification. In such a scoring scheme, where calculated chemical shifts of candidate structures are compared to experimental values of the unknown compound, chemical shift deviations for these motifs should be given a higher weight than deviations for other motifs for which larger errors have been observed in the linear regression analysis.

The third project focusses on the structural consequences of a single residue mutation in amyloid β (Aβ) (1-40) fibrils. It was investigated how the replacement of the hydrophobic amino acid phenylalanine (Phe) with the positively charged lysine (Lys) at position 19 impacts the structure of the peptide in its aggregated state. This is of particular biochemical relevance, since misfolding and aggregation of Aβ is one of the hallmarks of Alzheimer's disease (AD). Remarkably, it was previously found that the herein investigated, non-physiological $\mathrm{Phe}_{19}$-to-$\mathrm{Lys}_{19}$ mutation completely eliminates cytotoxicity, suggesting that distinct structural motifs are responsible for the devastating physiological effects caused by aggregated wild-type Aβ peptides. To elucidate these structural alterations in more detail, two atomic fibril models of Aβ $\mathrm{Phe}_{19}$-to-$\mathrm{Lys}_{19}$ have been created and compared to solid-state NMR measurements. The main finding was a previously unreported contact between side chains of $\mathrm{Phe}_{20}$ and $\mathrm{Leu}_{34}$, being a strong indication of a fully solvated Lys side chain oriented outside the hydrophobic fibril interior. From this result it becomes clearer that the suppression of the $\mathrm{Phe}_{19}$ and $\mathrm{Leu}_{34}$ contact, which is formed in most of the known wild type Aβ oligomer, protofibril and fibril structures, is causative for the abolishment of toxicity. This may also suggest a novel strategy for the design of new drugs against AD to specifically target the $\mathrm{Phe}_{19}$-$\mathrm{Leu}_{34}$ contact.

# 5. References

[1] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press, New York, USA, 1991.

[2] H. J. Berendsen, *Simulating the Physical World: Hierarchical Modeling from Quantum Mechanics to Fluid Dynamics*, Cambridge University Press, Cambridge, UK, 2007.

[3] M. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation*, Oxford University Press, New York, USA, 2010.

[4] D. Marx and J. Hutter, *Ab initio Molecular Dynamics: Basic Theory and Advanced Methods*, Cambridge University Press, New York, 2009.

[5] B. Kirchner and J. Vrabec, *Multiscale Molecular Methods in Applied Chemistry*, Springer, Berlin, Germany, 2012, vol. 307.

[6] J. VandeVondele, U. Borstnik and J. Hutter, Linear Scaling Self-Consistent Field Calculations with Millions of Atoms in the Condensed Phase, *J. Chem. Theory Comput.*, 2012, **8**, 3565–3573.

[7] R. O. Dror, R. M. Dirks, J. Grossman, H. Xu and D. E. Shaw, Biomolecular Simulation: A Computational Microscope for Molecular Biology, *Annu. Rev. Biophys.*, 2012, **41**, 429–452.

[8] J. Nasica-Labouze, P. H. Nguyen, F. Sterpone, O. Berthoumieu, N.-V. Buchete, S. Coté, A. De Simone, A. J. Doig, P. Faller, A. Garcia *et al.*, Amyloid $\beta$-protein and Alzheimer's Disease: When Computer Simulations Complement Experimental Studies, *Chem. Rev.*, 2015, **115**, 3518.

[9] D. Marx, M. E. Tuckerman, J. Hutter and M. Parrinello, The Nature of the Hydrated Excess Proton in Water, *Nature*, 1999, **397**, 601–604.

[10] R. Vuilleumier and D. Borgis, Proton Conduction: Hopping along Hydrogen Bonds, *Nat. Chem.*, 2012, **4**, 432–433.

[11] P. L. Geissler, C. Dellago, D. Chandler, J. Hutter and M. Parrinello, Autoionization in Liquid Water, *Science*, 2001, **291**, 2121–2124.

[12] M. Tuckerman, K. Laasonen, M. Sprik and M. Parrinello, Ab Initio Molecular Dynamics Simulation of the Solvation and Transport of Hydronium and Hydroxyl Ions in Water, *J. Chem. Phys.*, 1995, **103**, 150–161.

[13] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.

[14] R. B. Best, X. Zhu, J. Shim, P. E. Lopes, J. Mittal, M. Feig and A. D. MacKerell Jr, Optimization of the Additive CHARMM All-Atom Protein Force

Field Targeting Improved Sampling of the Backbone $\varphi$, $\psi$ and Side-Chain $\chi 1$ and $\chi 2$ Dihedral Angles, *J. Chem. Theory Comput.*, 2012, **8**, 3257.

[15] C. Oostenbrink, A. Villa, A. E. Mark and W. F. Van Gunsteren, A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6, *J. Comput. Chem.*, 2004, **25**, 1656–1676.

[16] Y.-L. S. Tse, C. Knight and G. A. Voth, An Analysis of Hydrated Proton Diffusion in Ab Initio Molecular Dynamics, *J. Chem. Phys.*, 2015, **142**, 014104.

[17] K. Adamczyk, M. Prémont-Schwarz, D. Pines, E. Pines and E. T. J. Nibbering, Real-Time Observation of Carbonic Acid Formation in Aqueous Solution, *Science*, 2009, **326**, 1690–1694.

[18] O. F. Mohammed, D. Pines, J. Dreyer, E. Pines and E. T. Nibbering, Sequential Proton Transfer Through Water Bridges in Acid-Base Reactions, *Science*, 2005, **310**, 83–86.

[19] O. F. Mohammed, D. Pines, E. T. J. Nibbering and E. Pines, Base-Induced Solvent Switches in Acid-Base Reactions, *Angew. Chem. Int. Ed.*, 2007, **119**, 1480–1483.

[20] O. F. Mohammed, D. Pines, E. Pines and E. T. J. Nibbering, Aqueous Bimolecular Proton Transfer in Acid-Base Neutralization, *Chem. Phys.*, 2007, **341**, 240–257.

[21] K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror and D. E. Shaw, Systematic Validation of Protein Force Fields Against Experimental Data, *PloS One*, 2012, **7**, e32131.

[22] R. Vuilleumier and D. Borgis, Transport and Spectroscopy of the Hydrated Proton: A Molecular Dynamics Study, *J. Chem. Phys.*, 1999, **111**, 4251–4266.

[23] D. Marx, Proton Transfer 200 Years after von Grotthuss: Insights from Ab Initio Simulations, *ChemPhysChem*, 2006, **7**, 1848–1870.

[24] O. Markovitch, H. Chen, S. Izvekov, F. Paesani, G. A. Voth and N. Agmon, Special Pair Dance and Partner Selection: Elementary Steps in Proton Transport in Liquid Water, *J. Phys. Chem. B*, 2008, **112**, 9456–9466.

[25] J. Xu, Y. Zhang and G. A. Voth, Infrared Spectrum of the Hydrated Proton in Water, *J. Phys. Chem. Lett.*, 2010, **2**, 81–86.

[26] W. Kulig and N. Agmon, A ‘Clusters-in-Liquid’ Method for Calculating Infrared Spectra Identifies the Proton-Transfer Mode in Acidic Aqueous Solutions, *Nat. Chem.*, 2013, **5**, 29–35.

[27] T. E. Decoursey, Voltage-Gated Proton Channels and Other Proton Transfer Pathways, *Physiol. Rev.*, 2003, **83**, 475–579.

[28] L. Vilčiauskas, M. E. Tuckerman, G. Bester, S. J. Paddison and K.-D. Kreuer, The Mechanism of Proton Conduction in Phosphoric Acid, *Nat. Chem.*, 2012, **4**, 461–466.

[29] K.-D. Kreuer, S. J. Paddison, E. Spohr and M. Schuster, Transport in Proton Conductors for Fuel-Cell Applications: Simulations, Elementary Reactions, and Phenomenology, *Chem. Rev.*, 2004, **104**, 4637–4678.

[30] R. Jorn, J. Savage and G. A. Voth, Proton Conduction in Exchange Membranes Across Multiple Length Scales, *Acc. Chem. Res.*, 2012, **45**, 2002–2010.

[31] N. Agmon, The Grotthuss Mechanism, *Chem. Phys. Lett.*, 1995, **244**, 456–462.

[32] D. Marx, A. Chandra and M. E. Tuckerman, Aqueous Basic Solutions: Hydroxide Solvation, Structural Diffusion, and Comparison to the Hydrated Proton, *Chem. Rev.*, 2010, **110**, 2174–2216.

[33] A. Hassanali, F. Giberti, J. Cuny, T. D. Kühne and M. Parrinello, Proton Transfer Through the Water Gossamer, *Proc. Natl. Acad. Sci.*, 2013, **110**, 13723–13728.

[34] F. Messina, M. Prémont-Schwarz, O. Braem, D. Xiao, V. S. Batista, E. T. J. Nibbering and M. Chergui, Ultrafast Solvent-Assisted Electronic Level Crossing in 1-Naphthol, *Angew. Chem. Int. Ed.*, 2013, **52**, 6871–6875.

[35] M. Rini, B.-Z. Magnes, E. Pines and E. T. J. Nibbering, Real-Time Observation of Bimodal Proton Transfer in Acid-Base Pairs in Water, *Science*, 2003, **301**, 249–352.

[36] M. Rini, D. Pines, B.-Z. Magnes, E. Pines and E. T. J. Nibbering, Bimodal Proton Transfer in Acid-Base reactions in Water, *J. Chem. Phys. A*, 2004, **122**, 9593–9610.

[37] P. Thistlethwaite and P. Corkill, Direct Observation of Photoautomerism Kinetics in 7-Quinolinol by Picosecond Spectroscopy, *Chem. Phys. Lett.*, 1982, **85**, 317–321.

[38] M. Itoh, T. Adachi and K. Tokumura, Time-Resolved Fluorescence and Absorption Spectra and Two-Step Laser Excitation Fluorescence of the Excited-State Proton Transfer in the Methanol Solution of 7-Hydroxyquinoline, *J. Am. Soc. Chem.*, 1984, **106**, 850–855.

[39] E. Bardez, Excited-State Proton Transfer in Bifunctional Compounds, *Isr. J. Chem.*, 1999, **39**, 319–332.

[40] G. Bekçioğlu, C. Allolio and D. Sebastiani, Water Wires in Aqueous Solutions from First-Principles Calculations, *J. Phys. Chem. B*, 2015, **119**, 4053–4060.

[41] F. Chiti and C. M. Dobson, Protein Misfolding, Functional Amyloid, and Human Disease, *Annu. Rev. Biochem.*, 2006, **75**, 333–366.

[42] F. Chiti and C. M. Dobson, Amyloid Formation by Globular Proteins under Native Conditions, *Nat. Chem. Biol.*, 2009, **5**, 15–22.

[43] R. Wetzel, S. Shivaprasad and A. D. Williams, Plasticity of Amyloid Fibrils, *Biochemistry*, 2007, **46**, 1.

[44] J. Adler, H. A. Scheidt, M. Krüger, L. Thomas and D. Huster, Local Interactions Influence the Fibrillation Kinetics, Structure and Dynamics of A$\beta$ (1–40) but Leave the General Fibril Structure Unchanged, *Phys. Chem. Chem. Phys.*, 2014, **16**, 7461–7471.

[45] A. T. Petkova, R. D. Leapman, Z. Guo, W.-M. Yau, M. P. Mattson and R. Tycko, Self-Propagating, Molecular-Level Polymorphism in Alzheimer's $\beta$-Amyloid Fibrils, *Science*, 2005, **307**, 262–265.

[46] A. K. Paravastu, R. D. Leapman, W.-M. Yau and R. Tycko, Molecular

Structural Basis for Polymorphism in Alzheimer's $\beta$-Amyloid Fibrils, *Proc. Natl. Acad. Sci.*, 2008, **105**, 18349–18354.

[47] I. Bertini, L. Gonnelli, C. Luchinat, J. Mao and A. Nesi, A New Structural Model of A$\beta$40 Fibrils, *J. Am. Chem. Soc.*, 2011, **133**, 16013–16022.

[48] A. T. Petkova, Y. Ishii, J. J. Balbach, O. N. Antzutkin, R. D. Leapman, F. Delaglio and R. Tycko, A Structural Model for Alzheimer's $\beta$-Amyloid Fibrils Based on Experimental Constraints from Solid State NMR, *Proc. Natl. Acad. Sci.*, 2002, **99**, 16742–16747.

[49] J. M. Lopez del Amo, M. Schmidt, U. Fink, M. Dasari, M. Fändrich and B. Reif, An Asymmetric Dimer as the Basic Subunit in Alzheimer's Disease Amyloid $\beta$ Fibrils, *Angew. Chem. Int. Ed.*, 2012, **124**, 6240–6243.

[50] J.-X. Lu, W. Qiang, W.-M. Yau, C. D. Schwieters, S. C. Meredith and R. Tycko, Molecular Structure of $\beta$-Amyloid Fibrils in Alzheimer's Disease Brain Tissue, *Cell*, 2013, **154**, 1257–1268.

[51] Y. Xiao, B. Ma, D. McElheny, S. Parthasarathy, F. Long, M. Hoshi, R. Nussinov and Y. Ishii, A $\beta$(1-42) Fibril Structure Illuminates Self-Recognition and Replication of Amyloid in Alzheimer's Disease, *Nat. Struct. Mol. Biol.*, 2015, **22**, 499–505.

[52] M. A. Wälti, F. Ravotti, H. Arai, C. G. Glabe, J. S. Wall, A. Böckmann, P. Güntert, B. H. Meier and R. Riek, Atomic-Resolution Structure of a Disease-Relevant A$\beta$ (1–42) Amyloid Fibril, *Proc. Natl. Acad. Sci.*, 2016, **113**, E4976–E4984.

[53] L. Gremer, D. Schölzel, C. Schenk, E. Reinartz, J. Labahn, R. B. Ravelli, M. Tusche, C. Lopez-Iglesias, W. Hoyer, H. Heise *et al.*, Fibril Structure of Amyloid-$\beta$ (1–42) by Cryo–Electron Microscopy, *Science*, 2017, **358**, 116–119.

[54] A. R. A. Ladiwala, J. C. Lin, S. S. Bale, A. M. Marcelino-Cruz, M. Bhattacharya, J. S. Dordick and P. M. Tessier, Resveratrol Selectively Remodels Soluble Oligomers and Fibrils of Amyloid A$\beta$ into Off-Pathway Conformers, *J. Biol. Chem.*, 2010, **285**, 24228–24237.

[55] H. A. Scheidt, I. Morgado, S. Rothemund, D. Huster and M. Fändrich, Solid-State NMR Spectroscopic Investigation of A$\beta$ Protofibrils: Implication of a $\beta$-Sheet Remodeling upon Maturation into Terminal Amyloid Fibrils, *Angew. Chem. Int. Ed.*, 2011, **50**, 2837–2840.

[56] J. Adler, M. Baumann, B. Voigt, H. A. Scheidt, D. Bhowmik, T. Häupl, B. Abel, P. K. Madhu, J. Balbach, S. Maiti *et al.*, A Detailed Analysis of the Morphology of Fibrils of Selectively Mutated Amyloid $\beta$ (1–40), *ChemPhysChem*, 2016, **17**, 2744–2753.

[57] A. K. Das, A. Rawat, D. Bhowmik, R. Pandit, D. Huster and S. Maiti, An Early Folding Contact Between Phe19 and Leu34 is Critical for Amyloid-$\beta$ Oligomer Toxicity, *ACS Chem. Neurosci.*, 2015, **6**, 1290–1295.

[58] G. N. Gowda and D. Raftery, Can NMR solve some significant challenges in metabolomics?, *J. Magn. Reson.*, 2015, **260**, 144–160.

[59] T. W.-M. Fan and A. N. Lane, Applications of NMR Spectroscopy to Systems

Biochemistry, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2016, **92**, 18–53.

[60] J. L. Markley, R. Brüschweiler, A. S. Edison, H. R. Eghbalnia, R. Powers, D. Raftery and D. S. Wishart, The Future of NMR-Based Metabolomics, *Curr. Opin. Biotechnol.*, 2017, **43**, 34–40.

[61] K. Bingol, L. Bruschweiler-Li, D. Li, B. Zhang, M. Xie and R. Brüschweiler, Emerging New Strategies for Successful Metabolite Identification in Metabolomics, *Bioanalysis*, 2016, **8**, 557–573.

[62] W. B. Dunn, D. I. Broadhurst, H. J. Atherton, R. Goodacre and J. L. Griffin, Systems Level Studies of Mammalian Metabolomes: The Roles of Mass Spectrometry and Nuclear Magnetic Resonance Spectroscopy, *Chem. Soc. Rev.*, 2011, **40**, 387–426.

[63] O. Beckonert, H. C. Keun, T. M. Ebbels, J. Bundy, E. Holmes, J. C. Lindon and J. K. Nicholson, Metabolic Profiling, Metabolomic and Metabonomic Procedures for NMR Spectroscopy of Urine, Plasma, Serum and Tissue Extracts, *Nat. Protoc.*, 2007, **2**, 2692–2703.

[64] H. K. Kim, Y. H. Choi and R. Verpoorte, NMR-Based Metabolomic Analysis of Plants, *Nat. Protoc.*, 2010, **5**, 536–549.

[65] M. S. Palmnas and H. J. Vogel, The Future of NMR Metabolomics in Cancer Therapy: Towards Personalizing Treatment and Developing Targeted Drugs?, *Metabolites*, 2013, **3**, 373–396.

[66] A. J. Brandolini, *NMRPredict Modgraph Consultants, Ltd, 1348 Graham Place, Escondido, CA 92129. http://www. modgraph-usa. com.*, 2006.

[67] *ACD/NMR Predictors, Advanced Chemistry Development.*

[68] G. Bekçioğlu-Neff, C. Allolio, Y. S. Desmukh, M. R. Hansen and D. Sebastiani, Dynamical Dimension to the Hofmeister Series: Insights from First-Principles Simulations, *ChemPhysChem*, 2016, **17**, 1166–1173.

[69] H. Elgabarty, P. Schmieder and D. Sebastiani, Unraveling the Existence of Dynamic Water Channels in Light-Harvesting Proteins: Alpha-C-Phycocyanobilin In Vitro, *Chem. Sci.*, 2013, **4**, 755–763.

[70] M. W. Lodewyk, M. R. Siebert and D. J. Tantillo, Computational Prediction of 1H and 13C Chemical Shifts: A Useful Tool for Natural Product, Mechanistic, and Synthetic Organic Chemistry, *Chem. Rev.*, 2011, **112**, 1839–1862.

[71] P. H. Willoughby, M. J. Jansma and T. R. Hoye, A Guide to Small-Molecule Structure Assignment Through Computation of 1H and 13C NMR Chemical Shifts, *Nat. Prot.*, 2014, **9**, 643–661.

[72] D. J. Tantillo, Walking in the Woods With Quantum Chemistry–Applications of Quantum Chemical Calculations in Natural Products Research, *Nat. Prod. Rep.*, 2013, **30**, 1079–1086.

[73] R. A. Cormanich, M. Bühl and R. Rittner, Understanding the Conformational Behaviour of Ac-Ala-NHMe in Different Media. A joint NMR and DFT Study, *Org. Biomol. Chem.*, 2015, **13**, 9206–9213.

[74] D. Flaig, M. Maurer, M. Hanni, K. Braunger, L. Kick, M. Thubauville and C. Ochsenfeld, Benchmarking Hydrogen and Carbon NMR Chemical Shifts at

HF, DFT, and MP2 Levels, *J. Chem. Theory Comput.*, 2014, **10**, 572–578.

[75] I. A. Konstantinov and L. J. Broadbelt, Regression Formulas for Density Functional Theory Calculated 1H and 13C NMR Chemical Shifts in Toluene-d 8, *J. Phys. Chem. A*, 2011, **115**, 12364–12372.

[76] G. K. Pierens, 1H and 13C NMR Scaling Factors for the Calculation of Chemical Shifts in Commonly Used Solvents Using Density Functional Theory, *J. Comp. Chem.*, 2014, **35**, 1388–1394.

[77] E. Benassi, Benchmarking of Density Functionals for a Soft but Accurate Prediction and Assignment of 1H and 13C NMR Chemical Shifts in Organic and Biological Molecules, *J. Comp. Chem.*, 2017, **38**, 87–92.

[78] T. Zhu, J. Z. Zhang and X. He, Automated Fragmentation QM/MM Calculation of Amide Proton Chemical Shifts in Proteins with Explicit Solvent Model, *J. Chem. Theory Comput.*, 2013, **9**, 2104–2114.

[79] J. C. Butcher, *Numerical Methods for Ordinary Differential Equations*, John Wiley & Sons, Ltd, Chichester, UK, 2008.

[80] R. Hockney, S. Goel and J. Eastwood, Quiet High-Resolution Computer Models of a Plasma, *J. Comp. Phys.*, 1974, **14**, 148–158.

[81] W. C. Swope, H. C. Andersen, P. H. Berens and K. R. Wilson, A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters, *J. Chem. Phys.*, 1982, **76**, 637–649.

[82] H. Berendsen and W. Van Gunsteren, *Practical Algorithms for Dynamic Simulations*, North-Holland, Amsterdam, 1986.

[83] C. W. Gear, *Numerical Initial Value Problems in Ordinary Differential Equations*, Prentice Hall PTR, USA, 1971.

[84] L. Verlet, Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules, *Phys. Rev.*, 1967, **159**, 98.

[85] H. J. Berendsen, J. v. Postma, W. F. van Gunsteren, A. DiNola and J. Haak, Molecular Dynamics with Coupling to an External Bath, *J. Chem. Phys.*, 1984, **81**, 3684–3690.

[86] H. C. Andersen, Molecular Dynamics Simulations at Constant Pressure and/or Temperature, *J. Chem. Phys.*, 1980, **72**, 2384–2393.

[87] S. Nosé, A Molecular Dynamics Method for Simulations in the Canonical Ensemble, *Mol. Phys.*, 1984, **52**, 255–268.

[88] W. G. Hoover, Canonical Dynamics: Equilibrium Phase-Space Distributions, *Phys. Rev. A*, 1985, **31**, 1695.

[89] G. J. Martyna, M. L. Klein and M. Tuckerman, Nosé–Hoover Chains: The Canonical Ensemble via Continuous Dynamics, *J. Chem. Phys.*, 1992, **97**, 2635–2643.

[90] G. Bussi, D. Donadio and M. Parrinello, Canonical Sampling through Velocity Rescaling, *J. Chem. Phys.*, 2007, **126**, 014101.

[91] G. J. Martyna, M. E. Tuckerman, D. J. Tobias and M. L. Klein, Explicit Reversible Integrators for Extended Systems Dynamics, *Mol. Phys.*, 1996, **87**,

1117–1157.

[92] S. Nosé and M. Klein, Constant Pressure Molecular Dynamics for Molecular Systems, *Mol. Phys.*, 1983, **50**, 1055–1076.

[93] M. Abraham, B. Hess, D. Spoel and E. Lindahl, *GROMACS User Manual version 5.0. 1. 2014*, 2015.

[94] M. Parrinello and A. Rahman, Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method, *J. Appl. Phys.*, 1981, **52**, 7182–7190.

[95] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers, *SoftwareX*, 2015, **1**, 19–25.

[96] P. P. Ewald, Die Berechnung Optischer und Elektrostatischer Gitterpotentiale, *Ann. Phys.*, 1921, **369**, 253–287.

[97] U. Essman, L. Perera, M. L. Berkowitz, T. A. Darden, H. Lee and L. G. Pedersen, A Smooth Particle Mesh Ewald Method, *J. Chem. Phys.*, 1995, **103**, 8577–8593.

[98] M. Del Ben, M. Schönherr, J. Hutter and J. VandeVondele, Bulk Liquid Water at Ambient Temperature and Pressure from MP2 Theory, *J. Phys. Chem. Lett.*, 2013, **4**, 3753–3759.

[99] J. W. Ponder and D. A. Case, Force Fields for Protein Simulations, *Adv. Protein Chem.*, 2003, **66**, 27–85.

[100] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field, *Proteins: Struct., Funct., Bioinf.*, 2010, **78**, 1950–1958.

[101] A. R. Leach, *Molecular Modelling: Principles and Applications*, Pearson Education Ltd., Essex, UK, 2001.

[102] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta and P. Weiner, A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins, *J. Am. Chem. Soc.*, 1984, **106**, 765–784.

[103] S. J. Weiner, P. A. Kollman, D. T. Nguyen and D. A. Case, An All Atom Force Field for Simulations of Proteins and Nucleic Acids, *J. Comput. Chem.*, 1986, **7**, 230–252.

[104] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters, *Proteins: Struct., Funct., Bioinf.*, 2006, **65**, 712–725.

[105] A. Warshel and M. Levitt, Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme, *J. Mol. Biol.*, 1976, **103**, 227–249.

[106] T. Laino, F. Mohamed, A. Laio and M. Parrinello, An Efficient Real Space Multigrid QM/MM Electrostatic Coupling, *J. Chem. Theory Comput.*, 2005, **1**, 1176–1184.

[107] P. Hohenberg and W. Kohn, Inhomogeneous Electron Gas, *Phys. Rev.*, 1964, **136**, 864.

[108] M. Levy, Universal Variational Functionals of Electron Densities, First-Order Density Matrices, and Natural Spin-Orbitals and Solution of the v-Representability Problem, *Proc. Natl. Acad. Sci.*, 1979, **76**, 6062–6065.

[109] W. Kohn and L. J. Sham, Self-Consistent Equations Including Exchange and Correlation Effects, *Phys. Rev.*, 1965, **140**, A1133.

[110] W. Koch and M. C. Holthausen, *A Chemist's Guide to Density Functional Theory, Second Edition*, Wiley-VCH, Weinheim, Germany, 2008.

[111] F. Bloch, Bemerkung zur Elektronentheorie des Ferromagnetismus und der Elektrischen Leitfähigkeit, *Z. Phys.*, 1929, **57**, 545–555.

[112] P. A. M. Dirac, Note on Exchange Phenomena in the Thomas Atom, *Math. Proc. Cambridge*, 1930, **26**, 376–385.

[113] D. M. Ceperley and B. Alder, Ground State of the Electron Gas by a Stochastic Method, *Phys. Rev. Lett.*, 1980, **45**, 566.

[114] S. H. Vosko, L. Wilk and M. Nusair, Accurate Spin-Dependent Electron Liquid Correlation Energies for Local Spin Density Calculations: A Critical Analysis, *Can. J. Phys.*, 1980, **58**, 1200–1211.

[115] J. P. Perdew and Y. Wang, Accurate and Simple Analytic Representation of the Electron-Gas Correlation Energy, *Phys. Rev. B*, 1992, **45**, 13244.

[116] A. D. Becke, Density Functional Calculations of Molecular Bond Energies, *J. Chem. Phys.*, 1986, **84**, 4524–4529.

[117] C. Lee, W. Yang and R. G. Parr, Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density, *Phys. Rev. B*, 1988, **37**, 785.

[118] J. P. Perdew, Density-Functional Approximation for the Correlation Energy of the Inhomogeneous Electron Gas, *Phys. Rev. B*, 1986, **33**, 8822.

[119] J. P. Perdew, K. Burke and M. Ernzerhof, Generalized Gradient Approximation Made Simple, *Phys. Rev. Lett.*, 1996, **77**, 3865.

[120] L. A. Curtiss, K. Raghavachari, G. W. Trucks and J. A. Pople, Gaussian-2 Theory for Molecular Energies of First- and Second-Row Compounds, *J. Chem. Phys.*, 1991, **94**, 7221–7230.

[121] L. A. Curtiss, K. Raghavachari, P. C. Redfern and J. A. Pople, Assessment of Gaussian-2 and Density Functional Theories for the Computation of Enthalpies of Formation, *J. Chem. Phys.*, 1997, **106**, 1063–1079.

[122] A. D. Becke, A New Mixing of Hartree–Fock and Local Density-Functional Theories, *J. Chem. Phys.*, 1993, **98**, 1372–1377.

[123] A. D. Becke, Density-Functional Thermochemistry. III. The Role of Exact Exchange, *J. Chem. Phys.*, 1993, **98**, 5648.

[124] K. Burke, M. Ernzerhof and J. P. Perdew, The Adiabatic Connection Method: A Non-Empirical Hybrid, *Chem. Phys. Lett.*, 1997, **265**, 115–120.

[125] C. C. J. Roothaan, New Developments in Molecular Orbital Theory, *Rev. Mod. Phys*, 1951, **23**, 69.

[126] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, General Publishing Company, Ltd., Canada, 1989.

[127] R. Ditchfield, W. J. Hehre and J. A. Pople, Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules, *J. Chem. Phys.*, 1971, **54**, 724–728.

[128] C. Kittel, *Introduction to Solid State Physics*, Wiley, 2005.

[129] M. Frigo and S. G. Johnson, Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on, 1998, pp. 1381–1384.

[130] P. Pulay, Derivative Methods in Quantum Chemistry, *Adv. Chem. Phys.*, 1987, **69**, 241–286.

[131] D. R. Hamann, M. Schlüter and C. Chiang, Norm-Conserving Pseudopotentials, *Phys. Rev. Lett.*, 1979, **43**, 1494–1497.

[132] S. Goedecker, M. Teter and J. Hutter, Separable Dual-Space Gaussian Pseudopotentials, *Phys. Rev. B*, 1996, **54**, 1703–1710.

[133] M. Kaupp, B. Michael, V. G. Malkin *et al.*, *Calculation of NMR and EPR Parameters: Theory and Applications*, John Wiley & Sons, 2006.

[134] W. Kutzelnigg, Theory of Magnetic Susceptibilities and NMR Chemical shifts in Terms of Localized Quantities, *Isr. J. Chem.*, 1980, **19**, 193–200.

[135] R. Ditchfield, Molecular Orbital Theory of Magnetic Shielding and Magnetic Susceptibility, *J. Chem. Phys.*, 1972, **56**, 5688–5691.

[136] T. A. Keith and R. F. Bader, Calculation of Magnetic Response Properties using a Continuous Set of Gauge Transformations, *Chem. Phys. Lett.*, 1993, **210**, 223–231.

[137] V. Weber, M. Iannuzzi, S. Giani, J. Hutter, R. Declerck and M. Waroquier, Magnetic Linear Response Properties Calculations with the Gaussian and Augmented-Plane-Wave Method, *J. Chem. Phys.*, 2009, **131**, 014106.

[138] T. Keith and R. Bader, Calculation of Magnetic Response Properties using Atoms in Molecules, *Chem. Phys. Lett.*, 1992, **194**, 1–8.

[139] G. Bekçioğlu, F. Hoffmann and D. Sebastiani, Solvation-Dependent Latency of Photoacid Dissociation and Transient IR Signatures of Protonation Dynamics, *J. Phys. Chem. A*, 2015, **119**, 9244–9251.

[140] F. Hoffmann, M. Ekimova, G. Bekçioğlu-Neff, E. T. Nibbering and D. Sebastiani, Combined Experimental and Theoretical Study of the Transient IR Spectroscopy of 7-Hydroxyquinoline in the First Electronically Excited Singlet State, *J. Phys. Chem. A*, 2016, **120**, 9378–9389.

[141] C. Fiolhais, F. Nogueira and M. A. Marques, *A Primer in Density Functional Theory*, Springer-Verlag, Berlin, 2003, vol. 620.

[142] M. Petersilka, U. Gossmann and E. Gross, Excitation Energies from Time-Dependent Density-Functional Theory, *Phys. Rev. Lett.*, 1996, **76**, 1212.

[143] C. A. Ullrich, *Time-Dependent Density-Functional Theory: Concepts and Applications*, Oxfort University Press, New York, 2011.

[144] E. Runge and E. K. Gross, Density-Functional Theory for Time-Dependent

Systems, *Phys. Rev. Lett.*, 1984, **52**, 997.

[145] M. E. Casida, C. Jamorski, K. C. Casida and D. R. Salahub, Molecular Excitation Energies to High-Lying Bound States from Time-Dependent Density-Functional Response Theory: Characterization and Correction of the Time-Dependent Local Density Approximation Ionization Threshold, *J. Chem. Phys.*, 1998, **108**, 4439–4449.

[146] R. Ramírez, P. P. Kumar, D. Marx *et al.*, Quantum Corrections to Classical Time-Correlation Functions: Hydrogen Bonding and Anharmonic Floppy Modes., *J. Chem. Phys.*, 2004, **121**, 3973–3983.

[147] N. Marzari and D. Vanderbilt, Maximally Localized Generalized Wannier Functions for Composite Energy Bands, *Phys. Rev. B*, 1997, **56**, 12847.

[148] P. L. Silvestrelli and M. Parrinello, Water Molecule Dipole in the Gas and in the Liquid Phase, *Phys. Rev. Lett.*, 1999, **82**, 3308.

[149] P. L. Silvestrelli and M. Parrinello, Structural, Electronic, and Bonding Properties of Liquid Water from First Principles, *J. Chem. Phys.*, 1999, **111**, 3572–3580.

[150] M. Veiga-Gutiérrez, A. Brenlla, C. Carreira Blanco, B. Fernández, S. A. Kovalenko, F. Rodríguez-Prieto, M. Mosquera and J. L. P. Lustres, Dissociation of a Strong Acid in Neat Solvents: Diffusion Is Observed after Reversible Proton Ejection Inside the Solvent Shell, *J. Phys. Chem. B*, 2013, **117**, 14065–14078.

[151] E.-A. Gould, A. V. Popov, L. M. Tolbert, I. Presiado, Y. Erez, D. Huppert and K. M. Solntsev, Excited-State Proton Transfer in N-Methyl-6-Hydroxyquinolinium Salts: Solvent and Temperature Effects, *Phys. Chem. Chem. Phys.*, 2012, **14**, 8964–8973.

[152] A. V. Popov, E.-A. Gould, M. A. Salvitti, R. Hernandez and K. M. Solntsev, Diffusional Effects on the Reversible Excited-State Proton Transfer. From Experiments to Brownian Dynamics Simulations, *Phys. Chem. Chem. Phys.*, 2011, **13**, 14914–14927.

[153] E. Bardez, A. Chatelain, B. Larrey, B. Valeur *et al.*, Photoinduced Coupled Proton and Electron Transfers. 1. 6-Hydroxyquinoline, *J. Phys. Chem.*, 1994, **98**, 2357–2366.

[154] C. Allolio, M. Sajadi, N. P. Ernsting and D. Sebastiani, An Ab Initio Microscope: Molecular Contributions to the Femtosecond Time-Dependent Fluorescence Shift of a Reichardt-Type Dye, *Angew. Chem. Intl. Ed.*, 2013, **52**, 1813–1816.

[155] G. Chang, W. Still and W. Guida, An Internal Coordinate Monte Carlo Method for Searching Conformational Space, *J. Am. Chem. Soc.*, 1989, **111**, 4379–4386.

[156] M. Saunders, K. Houk, Y. D. Wu, W. C. Still, M. Lipton, G. Chang and W. C. Guida, Conformations of Cycloheptadecane. A Comparison of Methods for Conformational Searching, *J. Am. Chem. Soc.*, 1990, **112**, 1419–1427.

[157] F. Hoffmann, D.-W. Li, D. Sebastiani and R. Brüschweiler, Improved Quantum Chemical NMR Chemical Shift Prediction of Metabolites in Aqueous Solution toward the Validation of Unknowns, *J. Phys. Chem. A*, 2017, **121**, 3071–3078.

[158] A. K. Paravastu, R. D. Leapman, W.-M. Yau and R. Tycko, Molecular

Structural Basis for Polymorphism in Alzheimer's $\beta$-Amyloid Fibrils, *Proc. Natl. Acad. Sci.*, 2008, **105**, 18349–18354.

[159] F. Hoffmann, J. Adler, B. Chandra, K. R. Mote, G. Bekçioğlu-Neff, D. Sebastiani and D. Huster, Perturbation of the F19-L34 Contact in Amyloid $\beta$ (1-40) Fibrils Induces Only Local Structural Changes but Abolishes Cytotoxicity, *J. Phys. Chem. Lett.*, 2017, **19**, 4740–4745.

[160] B. Sarkar, V. S. Mithu, B. Chandra, A. Mandal, M. Chandrakesan, D. Bhowmik, P. K. Madhu and S. Maiti, Significant Structural Differences between Transient Amyloid-$\beta$ Oligomers and Less-Toxic Fibrils in Regions Known To Harbor Familial Alzheimer$'$ s Mutations, *Angew. Chem. Int. Ed.*, 2014, **53**, 6888–6892.

[161] B. Chandra, D. Bhowmik, B. K. Maity, K. R. Mote, D. Dhara, R. Venkatramani, S. Maiti and P. K. Madhu, Major Reaction Coordinates Linking Transient Amyloid-$\beta$ Oligomers to Fibrils Measured at Atomic Level, *Biophys. J.*, 2017, **113**, 805–816.

# A. Publications

Below is a list of publications in peer-reviewed journals that I have co-authored. The individual contributions and responsibilities of the respective authors are listed in detail.

[P1] G. Bekçioğlu, F. Hoffmann, D. Sebastiani
**Solvation-Dependent Latency of Photoacid Dissociation and Transient IR Signatures of Protonation Dynamics**;
*J. Phys. Chem. A* **119**, 9244–9251 (2015)

G. Bekçioğlu prepared the setup and conducted the molecular dynamics simulations. She carried out the part of the analysis that dealt with the aqueous solvation of the chromophore and the proton dissociation mechanism. Further, she wrote those parts of the manuscript that concern the aforementioned parts of the analysis.
I contributed to the manuscript by calculating and interpreting infrared spectra of the chromophore in its various protonation states from molecular dynamics simulation data. Moreover, I wrote the part of the manuscript that deals with the interpretation of infrared spectra.
D. Sebastiani supervised the work and contributed to the redaction of the manuscript.

[P2] F. Hoffmann, M. Ekimova, G. Bekçioğlu-Neff, E. Nibbering, D. Sebastiani
**Combined Experimental and Theoretical Study of the Transient IR Spectroscopy of 7-Hydroxyquinoline in the First Electronically Excited Singlet State**;
*J. Phys. Chem. A* **120**, 9378–9389 (2016)

In this project, I conducted the quantum-chemical calculations and carried out the vibrational analysis. Furthermore, I analyzed the electronic and structural effects of excitation and wrote the main part of the manuscript.
M. Ekimova conducted the transient infrared experiments and analyzed experimental data. Moreover, she contributed to the production of the graphs.
G. Bekçioğlu-Neff contributed quantum-chemical calculations in the initial phase of the project. Further, she had a significant part in the design of the study through a number of fruitful discussions regarding the computational setup.
E. Nibbering and D. Sebastiani supervised the experimental and computational part of the project, respectively. Moreover, they contributed to the preparation of the manuscript.

[P3] F. Hoffmann, Da-Wei Li, D. Sebastiani and R. Brüschweiler
**Improved Quantum Chemical NMR Chemical Shift Prediction of Metabolites in Aqueous Solution toward the Validation of Unknowns**;
*J. Phys. Chem. A* **121**, 3071–3078 (2017)

I carried out quantum-chemical NMR shielding calculations and derived linear scaling parameters for the conversion of the shielding values to NMR chemical shifts. Moreover, I wrote the main part of the manuscript.
D.-W. Li contributed by assigning experimentally measured chemical shifts of a large number of metabolite molecules and by compiling this information into a database. Moreover, he carried out empirical chemical shift predictions, which I compared to the quantum-chemical predictions so as to assess the accuracy of the MOSS-DFT approach.
D. Sebastiani contributed through discussions regarding the technical realization in the initial phase of the project and took part in the redaction of the manuscript.
R. Brüschweiler designed and supervised the study and wrote parts of the manuscript.

[P4] Felix Hoffmann, Juliane Adler, Bappaditya Chandra, Kaustubh R. Mote, G. Bekçioğlu-Neff, Daniel Sebastiani, and Daniel Huster
**Perturbation of the F19-L34 Contact in Amyloid $\beta$ (1-40) Fibrils Induces Only Local Structural Changes but Abolishes Cytotoxicity**;
*J. Phys. Chem. Lett.* **8**, 4740–4745 (2017)

I mainly designed the computational part of the study, created the atomistic models of F19K amyloid β (1-40) fibrils, and carried out the MD simulations. Moreover, I wrote those parts of the manuscript dealing with the computational models.

Juliane Adler, Bappaditya Chandra, Kaustubh R. Mothe synthesised the peptides and carried out the solid-state NMR measurements on the fibrils.

G. Bekçioğlu-Neff contributed to the project in its initial phase through preliminary MD simulations and discussions.

Daniel Sebastiani supervised the computational part of the project.

Daniel Huster designed and supervised the study, and wrote parts of the manuscript.

# Solvation-Dependent Latency of Photoacid Dissociation and Transient IR Signatures of Protonation Dynamics

Gül Bekçioğlu, Felix Hoffmann, and Daniel Sebastiani

Reprinted with permission from

J. Phys. Chem. A **2015**, 119, 9244–9251.

Copyright 2017 American Chemical Society.

`http://dx.doi.org/10.1021/acs.jpca.5b05438`

THE JOURNAL OF
PHYSICAL CHEMISTRY A

Article

pubs.acs.org/JPCA

# Solvation-Dependent Latency of Photoacid Dissociation and Transient IR Signatures of Protonation Dynamics

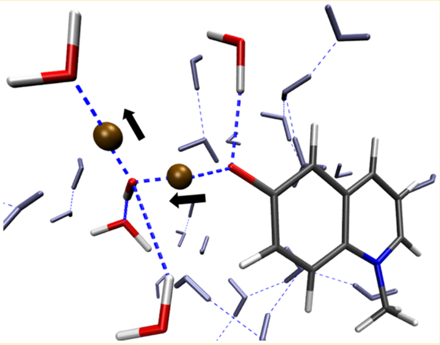
Gül Bekçioğlu,[†,‡] Felix Hoffmann,[‡] and Daniel Sebastiani*,[‡]

[†]Physics Department, Freie Universität Berlin, Arnimallee 14, 14195 Berlin, Germany

[‡]Institut für Chemie, Martin-Luther-Universität Halle-Wittenberg, Von-Danckelmann-Platz 4, 06120 Halle (Saale), Germany

*Supporting Information*

**ABSTRACT:** We elucidate the characteristic proton pathways and the transient infrared signatures of intermediate complexes during the first picoseconds of photoinduced protonation dynamics of a photoacid (*N*-methyl-6-hydroxyquinolinium) in aqueous solution from first-principles molecular dynamics simulations. Our results indicate that the typical latency time between photoexcitation and proton dissociation ranges from 1 ps to longer time time scales (~100 ps). The rate-limiting step for the actual dissociation of the proton into the solvent is the solvation structure of the first accepting water molecule. The nature of the proton pathway in water (stepwise or concerted) is not unique but determined by the coordination number of the accepting water molecules along the hydrogen bond chain. We find a characteristic uncommon infrared mode at ~1300 cm$^{-1}$ of the transient photobase-Eigen cation complex immediately after photodissociation that we predict to be observable experimentally in time-resolved IR spectroscopy. A broad continuous absorption band from 1500 to 2000 cm$^{-1}$ arises from the acidic proton imminently before dissociation.



## 1. INTRODUCTION

Photoinduced proton transfer (PT) in aqueous solutions is of fundamental interest in a large variety of chemical and biological processes such as the storage of energy and radiation-induced damage of DNA.[1−6] In aqueous solution, where the hydrogen bond interactions are very extended, PT can be very complex due to the structural rearrangement of water molecules. Water-mediated photoinduced PT may happen in a concerted or stepwise mechanism from the acid to the base over large distances.[7−14] Central in these mechanisms is the key role of the surrounding solvent shells, which facilitate the PT by fluctuations of the hydrogen bond network or breaking and formation of hydrogen bonds. These fluctuations are also considered to play important roles for the stabilization of reaction products and are involved in the reaction coordinate for proton dissociation mechanisms. However, the number of theoretical and experimental studies investigating these mechanisms have remained moderate.[15]

Ultrafast studies of photoinduced proton dissociation in aqueous solutions have utilized the class of organic molecules called photoacids.[16−23] A suitable molecule for this purpose is the "super" photoacid *N*-methyl-6-hydroxyquinolinium cation (6MQc) depicted in Figure 1. 6MQc exhibits a large p*K* drop from 7.2 in the ground state to −7 in the excited state (6MQc*).[24−29] Pérez-Lustres et al. reported proton dissociation kinetics of 6MQc* in aqueous solution with time-resolved fluorescence studies. It was shown that the excited-state proton dissociation barrier is 2 kJ mol$^{-1}$.[29] However, the fundamental question of how the intra- and intermolecular factors affect photoacidity and PT rate still remain under debate.[29−32]

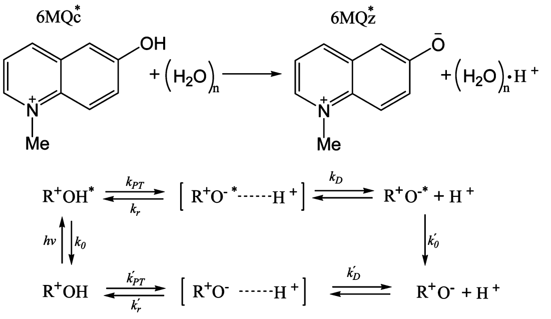


**Figure 1.** Excited-state PT and relaxation processes of excited 6MQc species.

The seminal work by Eigen and Weller[33−35] has frequently been utilized to discuss acid dissociation reactions and rates. The general kinetic approach, based on this model for acid−base reactions in aqueous solutions, consists of two step process: (a) The acidic proton is transferred from the acid to a solvent molecule after photoexcitation to form a contact ion pair, and (b) subsequently the contact ion pair is separated by diffusion (see Figure 1).[36] However, many details of the PT

**Received:** June 8, 2015
**Revised:** August 17, 2015
**Published:** August 17, 2015

ACS Publications © 2015 American Chemical Society

DOI: 10.1021/acs.jpca.5b05438
*J. Phys. Chem. A* 2015, 119, 9244−9251

from photoacid to water remain elusive, for example, the structure and stability of the contact ion pair formed by the photoacid and the hydrated proton, as well as the hydrogen bonding network around a solvated photosensitive probe molecule.

In the present paper, we are particularly interested in the role water plays in promoting photoinduced acid dissociation and in the hydrogen bonding rearrangements that accompany it. We put the chromophore in the triplet state ($T_1$) for the simulation of the excited-state dynamics; that is, the first excited state ($S_1$) was approximated with a $T_1$ surface. Our $T_1$ description is motivated by an increased numerical stability and a reduced computational effort compared to a real $S_1$ calculation. Of course, this protocol represents an approximation of the experimental situation. However, we have extensively benchmarked the validity of this approximation for 6MQc (see Supporting Information) and 6MQz[37] and verified that it correctly represents the most significant physical intramolecular interactions. The central question addressed in this work is the direct determination of elementary dissociation steps that lead to the final product: a fully solvated proton and a deprotonated photoacid. We computed IR spectra to follow the individual steps during the dissociation processes that occur in a time range of femtoseconds to tens of picoseconds.[38]

## 2. COMPUTATIONAL DETAILS

The molecular dynamics simulations were staged in a cubic, periodic box with a side length of 16.06 Å hosting one 6MQc molecule and 130 $H_2O$ molecules at a density of $d$ = 1.00 g/cm$^3$. First, we run 20 ps trajectory in NVT ensemble at 350 K. Then, we started five independent ground-state simulations ∼10 ps each in NVE ensemble from the previous simulation. From these NVE trajectories, 16 excited-state trajectories were started. We run these excited-state trajectories for ∼10 ps if no PT occurred and for ∼20 ps if PT occurred. For all calculations, we used GPW[39] scheme as implemented in the CP2K[40] software package. The BLYP[41,42] functional with a TZVP valence basis set, Goedecker[43] potentials, and a 350 Ry plane-wave cutoff was employed. With the increase in temperature we hope to balance overstructuring effects, found in water simulations at lower temperature.[44,45] For the very same reason, we used DFT-D2 dispersion corrections.[46] The simulation temperature of 350 K corresponds to a physical temperature of 300−320 K. For the ground-state simulations, we used restricted Kohn−Sham density functional theory (DFT) and a time step of 1 fs, while for the excited-state simulations, we used unrestricted Kohn−Sham DFT, a multiplicity of 3, and a time step of 1 fs. All protons in the system were given the mass of deuterium to allow us to use 1 fs time step.

All DFT-based static electronic structure and normal mode calculations were performed by using Gaussian 09 program.[47] The conventional DFT and normal mode calculations using hybrid exchange-correlation functional B3LYP[48] and BLYP[41,42] functional were perfomed within the presence of the CPCM model.[49] The basis set was TZVP for these calculations.

To study IR absorptions of the system, the electron density was subjected on the fly to a localization procedure employing maximally localized Wannier functions during the excited-state simulations.[50] Wannier centers can be regarded as centers of charge density of electron pairs in local orbitals and allow therefore for a chemical intuitive interpretation in terms of bonds and lone pairs. Moreover, it is well-known that this approach readily enables the calculation of molecular dipole moments in a condensed-phase system.[51] For instance, the dipole moment of molecule $I$ can be calculated via

$$\mu_I = \mu_I^{\text{el}} + \mu_I^{\text{nuclei}} \tag{1}$$

$$= -2e\sum_{i\in I}\mathbf{r}_i^{\text{Wannier}} + e\sum_{j\in I} n_{j,\text{val}}\cdot\mathbf{r}_j^{\text{nuclei}} \tag{2}$$

with $\mu_I^{\text{el}}$ and $\mu_I^{\text{nuclei}}$ designating the electronic and nuclei contribution to the molecular dipole moment. $\mathbf{r}_i^{\text{Wannier}}$ and $\mathbf{r}_j^{\text{nuclei}}$ are the positions of the Wannier centers and the ions, respectively; $e$ denotes the elementary charge, and $n_{j,\text{val}}$ indicates the number of valence electrons of the respective atom at position $\mathbf{r}_j^{\text{nuclei}}$.

This, in turn, allows one to compute the classical dipole autocorrelation function, whose Fourier transform is related to the product of absorption coefficient per unit length and refractive index, that is, the IR spectrum, via

$$\alpha(\omega)n(\omega) = F(\omega)\int_{-\infty}^{\infty}\mathrm{d}t\langle\mu(0)\mu(t)\rangle e^{-i\omega t} \tag{3}$$

where $F(\omega) = \beta\,\omega^2/6Vc\varepsilon_0$ and $\mu$ are the total dipole moment of the simulation box. Note that eq 3 is the classical approximation of the formerly quantum dipole autocorrelation function taking into account the harmonic quantum correction factor.[52]

IR absorptions for 6MQc, 6MQc*, and 6MQz* are calculated based on a spectral decomposition approach utilizing the cross-correlation function of the total dipole moment with the molecular dipole moment of the solute.[38]

Accordingly, eq 3 is modified by replacing the dipole autocorrelation function with the cross-correlation function yielding

$$I_I(\omega) = F(\omega)\int_{-\infty}^{\infty}\mathrm{d}t\langle\mu(0)\mu_I(t)\rangle e^{-i\omega t} \tag{4}$$

with the molecular dipole moment $\mu_I$, where the index $I$ denotes the set of atoms belonging to 6MQc, 6MQc*, and 6MQz*, respectively. For further analysis of the IR absorptions, we calculated the Fourier transform of the C−O bond autocorrelation function defined as

$$P_{\text{CO}}(\omega) = \int_{-\infty}^{\infty}\mathrm{d}t\langle\dot{d}_{\text{CO}}(0)\dot{d}_{\text{CO}}(t)\rangle e^{-i\omega t} \tag{5}$$

where $\dot{d}_{\text{CO}}$ is the time-derivative of the C−O bond distance. The resulting peaks should be approximately at positions of the normal modes where this bond vibration contributes, assuming that this vibration is not highly anharmonic. Shifts in $P_{\text{CO}}(\omega)$ can be regarded as a measure of bond strength variations, where a blue shift corresponds to a strengthening and a red shift corresponds to a weakening of the C−O bond.

## 3. RESULTS

### 3.1. Aqueous Solvation of 6MQc.

Five of 16 trajectories showed successful proton dissociation within 1−2 ps after photoexcitation. To verify the effects of solvent reorganization upon excitation, we computed the radial distribution functions (RDFs) of 6MQc-oxygen and water hydrogen atoms $g(\text{O}_{6\text{MQc}}\text{-D}_\text{w})$, and the acidic 6MQc-hydrogen and water oxygen atoms $g(\text{D}_{6\text{MQc}}\text{-O}_\text{w})$ from the ground- and excited-state trajectories. Our findings are reported in Figure 2. First, we observe a small peak arising at 1.1 Å $g(\text{D}_{6\text{MQc}}\text{-O}_\text{w})$ in the excited state due to the RO···D···$OD_2$ bond; that is, the proton is partly shared by

DOI: 10.1021/acs.jpca.5b05438
J. Phys. Chem. A 2015, 119, 9244−9251

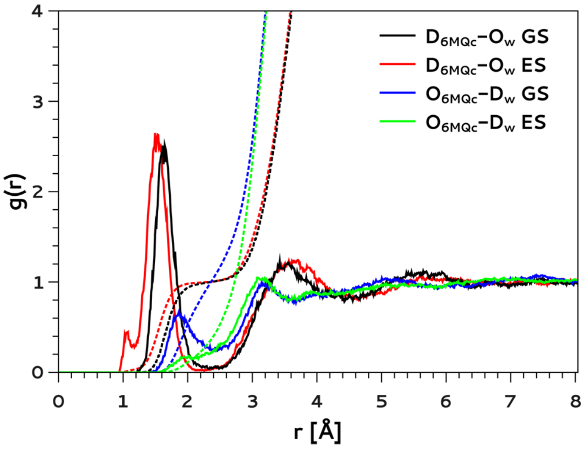


**Figure 2.** RDFs of 6MQc-hydrogen and water-oxygen atoms ($D_{6MQc}$-$O_w$) and 6MQc-oxygen and water-hydrogen atoms the water molecules ($O_{6MQc}$-$D_w$) are in the ground (black and blue) and the excited (red and green) states, respectively. The integration number of the distribution functions (dashed lines) is also reported.

the two oxygen atoms constituting the first stage of acid dissociation. The apparent shifts of the first peak and the following minimum at 2.25 Å in the excited state indicate that the ROD···$OD_2$ bond shortens by 0.2 Å due to the acidic proton becoming more positively charged and more disposed to H-bonding upon excitation. Further, strengthening of H-bonding does not seem to extend beyond the first solvation shell. Instead, one observes an increased order as represented by increase in the coordination number in the second solvation shell. The integration over the peaks at 3.6 Å in the $g(D_{6MQc}$-$O_w)$ corresponds to 8.8 and 9.4 water molecules in the ground and excited state, respectively. On the other hand, $g(O_{6MQc}$-$D_w)$ shows the effect of elongation of the hydrogen bond to the 6MQc*-oxygen due to the charge rearrangements around the lone pairs of the oxygen atom in the excited state (see Supporting Information and Figure 4a). However, this bond reforms as a subsequent solvent stabilization of the product, completing the PT reaction. Similar findings have been already discussed in experimental studies in terms of solvatochromic shifts (a blue shift) in the steady-state fluorescence spectra of undissociated photoacids.[53] Complementary to these findings, we found no obvious structural differences between the contact ion pair geometry for the successful and unsuccessful proton dissociation (see Supporting Information).

Another interesting observation in our simulations is the back-protonation reaction (geminate reaction) following the proton dissociation from 6MQc* after ~10 ps photodissociation. It was shown by time-resolved fluorescence experiments that the deviation from an exponential fluorescence decay results in the existence of the back-protonation reaction in the excited state.[54] Such a process, in turn, gives rise to a multiexponential time-resolved fluorescence decay, rather than a purely exponential form.

**3.2. Proton Dissociation Mechanism.** We now augment our analysis by elucidating the intake of protons into the H-bond network to provide a more complete picture of the long-range PT reaction in bulk aqueous solution. This fundamental interest lies at the heart of the microscopic character of the structure and mobility of liquid water. The basic PT reaction can be explained by a von Grotthuss mechanism involving a sequential hopping mechanism of protons to neighboring water molecules.[4−6] However, the actual PT reaction has been discussed to be either concerted or sequential in the bulk solution.[55,56] Although both mechanisms are plausible, concerted proton charge rearrangements occur in an assembly of several water molecules linked by hydrogen bonds for smaller PT distances, and a change into sequential transfer then is anticipated when the PT distance increases. From an atomistic point of view, the question of sequential or concerted PT can be translated into the comparison of typical lifetimes of particular protonation states in relation to the lifetimes of the involved hydrogen bonds.[55]

To verify the relationship between concerted and sequential PT mechanism, we employ, in Figure 3, the RDFs of the proton acceptor oxygen atom (denoted O*) and the hydrogen atoms of water molecules. The RDFs ($g_{O*-D}(r)$) are computed by averaging over 20 fs intervals before a successful PT occurs. We interpret $g_{O*-D}(r_{min})$ up to the relevant minimum of the first solvation peaks, characterized by a radius $r_{min}$. Note here that the peaks in $g_{O-D}(r)$ below 2.45 Å are generally considered as an indicator of a H-bond in bulk water. Figure 3 shows that in the case of a stepwise mechanism, there is a broad first peak with a minimum at 1.65 Å due to the strong O–D···O* bond. There is no further contribution from the other first solvation shell water molecules. However, in the case of the concerted mechanism, the O* atom accepts one additional H-bond, which is responsible for the small peak at 2.2 Å. This additional peak, in turn, requires an extension of the coordination radius to $r_{min}$ = 2.45 Å. In this regard, the actual character of the PT reaction depends on how the proton is taken into the H-bonded network: an increase in the coordination number changes the character of the PT event. In particular, when the initial coordination of the acceptor oxygen is ~fourfold ($n(r_{min})$ = 3.7 in Figure 3), the concerted PT is favored in this mechanism, and the subsequent PT occurs within ~0.1−0.5 ps. However, if the accepting water molecule is initially only threefold coordinated ($n(r_{min})$ = 3.1 in Figure 3), it will hold the proton for typically ~1−5 ps, which eventually yields a stepwise migration. Figure 4 illustrates the difference between these

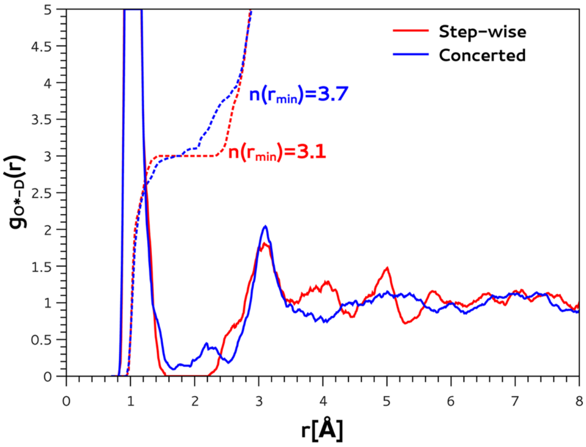


**Figure 3.** RDF and coordination numbers of a proton acceptor oxygen atom and the hydrogen atom of water molecules (O*-D) corresponding to the concerted mechanism (blue line and dashed line, respectively), and sequential hopping mechanism (red line and dashed line, respectively). The coordination numbers ($n(r_{min})$) are computed by characterizing the minimum of the first solvation shell of $g_{O*-D}(r)$.

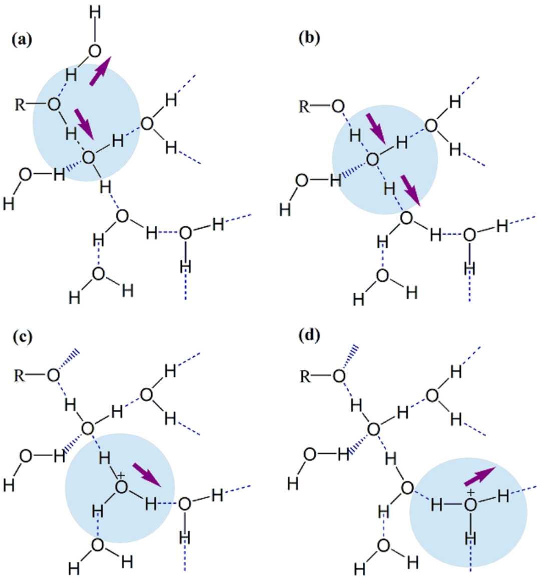


**Figure 4.** Scheme illustrating the PT mechanisms of 6MQc* in aqueous solution. (a) Formation of contact ion pair and elongation of hydrogen bond between 6MQc-oxygen atom and water molecule; (b) transient structures involving the concerted motion of two protons; (c, d) a threefold coordinated water molecule holding the excess proton.

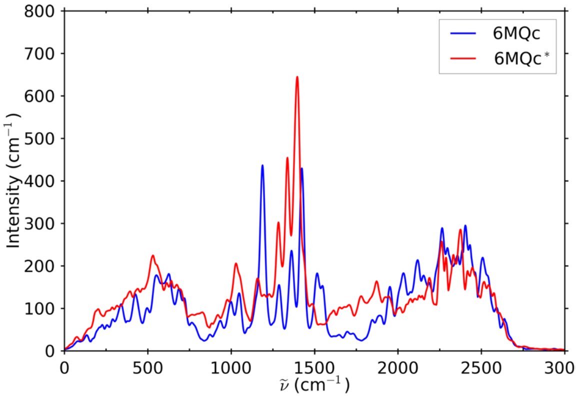


**Figure 5.** Infrared spectra for 6MQc and 6MQc* averaged over five and 11 trajectories, respectively. The simulation length was 10 ps in each case.

mechanisms regarding the coordination number. Our findings suggest that the concerted PT occurs when the proton receiving oxygen (O*) accepts a hydrogen bond from another water molecule before PT takes place. In contrast, the stepwise transfer involves a single O—H bond breaking/forming at a time. Our trajectories reveal that 60% of the proton dissociation reactions are concerted and that 40% occur in a stepwise manner. In all cases, the concerted events start from a fourfold coordinated water, and the stepwise process start from a threefold coordinated water.

A similar observation on the nature of the proton pathway in water was also reported by Tuckermann et al.[55] concerning the PT mechanism of the hydrated proton and in the study of Maurer et al.[57] by investigating the ultrafast acid dissociation and acid—base neutralization reactions. Therefore, we conclude that an ultrafast (subps, i.e. concerted) long-distance proton separation from the acid requires an initial fourfold coordination of all involved water molecules.

**3.3. Infrared Spectral Shifts.** We computed the IR spectra of 6MQc, 6MQc*, and 6MQz* to investigate effects of electronic excitation and proton dissociation on the IR absorptions. The spectra of 6MQc and 6MQc* are presented in Figure 5, corresponding to negative and early positive delay times in transient IR measurements. Figure 5 reveals slight red shifts by ~30 cm$^{-1}$ of most absorption peaks within 1000 to 1500 cm$^{-1}$ (mainly C—C and C—O stretching) upon electronic excitation (cf. Supporting Information for mode assignment). More precisely, peaks at 1000, 1180, 1355, 1420, and 1510 cm$^{-1}$ in the ground-state spectrum are shifted to 970, 1150, 1330, 1390, and 1480 cm$^{-1}$ in the excited-state spectrum. These red shifts are in line with the fact that C—C bonds in the aromatic rings get lengthened on average, as shown in Figure S7 in the Supporting Information. Conversely, a shortening of C—C bonds perpendicular to the direction of largest extension of 6MQc was observed consistent with findings of Agmon et al., who found a roughly alternating pattern of lengthening and shortening for 2-naphtol upon excitation to $S_1$.[58] A quantitative comparison of these shifts with transient IR measurements, however, is not directly possible due to the use of the $T_1$ state in our case.

Further, Figure 5 reveals intensity changes upon electronic excitation at 1180 cm$^{-1}$ and between 1280 and 1400 cm$^{-1}$, where intensity is decreased in the former and increased in the latter case. In contrast, only minor changes appear in the C—D stretching range ~2300 to 2700 cm$^{-1}$ and for frequencies below 700 cm$^{-1}$.

Most interesting, however, the O—D stretching vibration gives rise to a continuous absorption band between 1500 and 2000 cm$^{-1}$ in the excited-state spectrum caused by the flat potential of the O—D···OD$_2$ hydrogen bond. Additionally, the chromophore O—D bond is drastically weakened in the excited state, whereas the hydrogen bond to the coordinating water is strengthened, as shown in Figure 2. Note here that, in contrast to spectra obtained from molecular dynamics (MD) simulations that allow for a fully anharmonic treatment, this continuous absorption band cannot be obtained from normal mode calculations, even with perturbative corrections to the harmonic frequencies, where it always appears as a localized peak (see Supporting Information).

In Figure 5, the corresponding O—D absorption in the ground state coalesces with the peak between 1800 and 2800 cm$^{-1}$ (mainly C—D vibrations) as suggested by normal mode calculations with explicit water molecules (see Supporting Information). Therefore, the O—D absorption is located at higher frequencies than in the excited-state MD spectrum. The O—D stretching absorption in the normal mode spectrum without consideration of explicit solvent molecules is expectedly red-shifted due to the absence of hydrogen bonds to the solvent.

To attain further evidence that the continuous absorption band between 1500 and 2000 cm$^{-1}$ in the excited-state spectrum is indeed due to the O—D stretching vibration, we computed the IR spectrum for 6MQc* based on dipole moments calculated from atomic Hirshfeld charges,[59] leaving the trajectories unchanged. In Figure 6, it can be seen that the continuous absorption vanishes by setting the partial charge on the acidic hydrogen atom to zero, suggesting that the broad absorption band is dominated by the dynamics of this atom, which is comparably loosely bonded in 6MQc*. However, note here that peak patterns and intensities differ slightly compared

DOI: 10.1021/acs.jpca.5b05438
J. Phys. Chem. A 2015, 119, 9244−9251

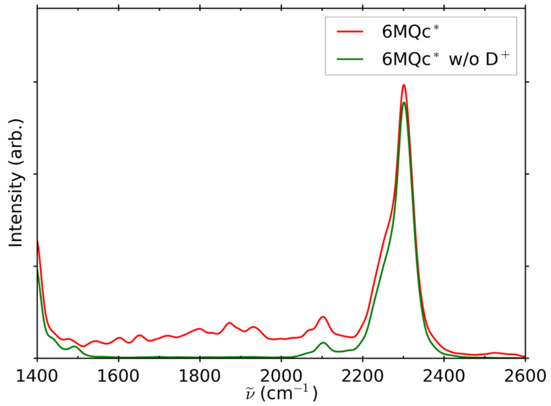


**Figure 6.** Infrared spectrum between 1400 and 2600 $cm^{-1}$ computed from atomic Hirshfeld charges based on trajectories of 6MQc*. The spectrum labeled 6MQc* refers to the total chromophore, whereas the spectrum 6MQc* without $D^+$ was obtained by setting the partial charge on the proton to zero.

to Figure 5, since the spectrum based on Hirshfeld charges misses electronic polarization effects and the underlying dipole moments differ compared to the Wannier center approach. Our findings are consistent with previous studies that ascribed absorption band to the motion of the weakly bonded proton in contact ion pairs and hydrated protons of ground-state acids as well as photoacids.[60−62] Moreover, experimental results for the photoacid 8-hydroxy-1,3,6-pyrenetrisulfonic trisodium salt (HPTS) suggest that the amplitude of the continuous absorption peak is largest at the onset of proton dissociation, while the contribution of the hydronium ion, which is not included in our spectral calculations, contributes with a smaller amplitude.[62]

Having discussed the impact of electronic excitation at the onset of proton dissociation, we now focus on the question how the IR absorptions change after the proton is released to a solvating water molecule. Figure 7 shows the spectrum of 6MQc* and 6MQz*, where the inset presents the absorptions between 1200 and 1500 $cm^{-1}$. The most significant change in IR absorption upon proton dissociation is the depletion of the continuous absorption band between 1500 and 2300 $cm^{-1}$, providing further evidence that this feature originates from the acidic proton. Moreover, the bands between 1250 and 1550 $cm^{-1}$, originating mostly from bond stretching vibrations of the ring atoms and the C−O bond, are significantly affected when proton dissociation occurs. In particular, the peak at 1330 $cm^{-1}$ is absent in the spectrum of the dissociated photoacid (cf. inset Figure 7). Notably, this feature is consistent for all single-trajectory spectra of the protonated and deprotonated chromophore, based on which we calculated the average spectra shown in Figure 7. Further, we analyzed this feature by means of normal-mode analysis (see Supporting Information) and found that the variation in peak patterns is due to changes of normal modes dominated by C−C stretching vibrations of the ring atoms, as illustrated in Figures S10 and S12 in the Supporting Information. In the region between 1250 and 1550 $cm^{-1}$, we observe an altered peak structure in the normal mode spectra due to the absence of the proton. In the MD spectra, the stretching vibrations between 1200 and 1500 $cm^{-1}$ (labeled with D-I in the Supporting Information) are most distinct, and therefore we can regard the vanishing of the absorption at 1330 $cm^{-1}$ as a consequence of these altered stretching modes. Note here that these unique assignments cannot be further resolved in our MD spectra due to the limited sampling time. This observation in turn suggests that the peak pattern around 1330 $cm^{-1}$ might serve as a marker for the protonation state in transient IR experiments. A similar feature has already been discussed in an experimental study of HPTS.[61] Further, we investigated this feature for 6MQz* and 6MQz by means of normal mode calculations (see Supporting Information). We found that the peak pattern around 1330 $cm^{-1}$ differs for 6MQz compared to 6MQz* and therefore regard this marker to be characteristic for the excited 6MQz* only.

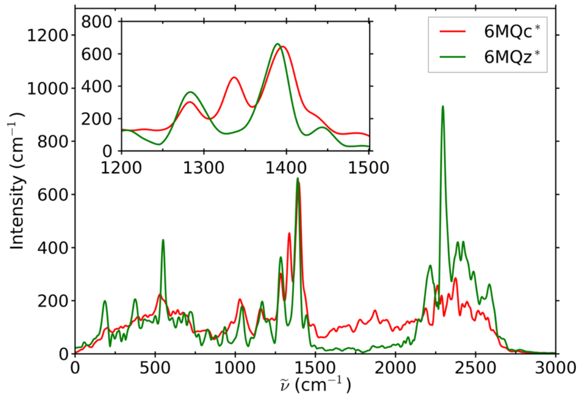


**Figure 7.** Infrared spectrum of 6MQc* and 6MQz* averaged over 11 and four trajectories, respecticvely. For 6MQz*, spectra were computed only within the time range of complete dissociation.

In addition, we investigated the effect of the proton dissociation on the C−O bond vibration. To this end, we computed the Fourier transform of the velocity auto-correlation function according to eq 5, which is depicted in Figure 8. We note that the vibrational frequency of the C−O bond cannot easily be extracted from the MD spectrum in Figure 7 due to the superposition of bands caused by other vibrations. In contrast, bands in the power spectrum of the distance auto-correlation function should approximately occur at positions of the respective normal modes, where the C−O bond vibration contributes. A comparison of the power spectra of 6MQc* and 6MQz* in Figure 8 reveals that only slight frequency shifts occur upon proton dissociation. However, the average distance

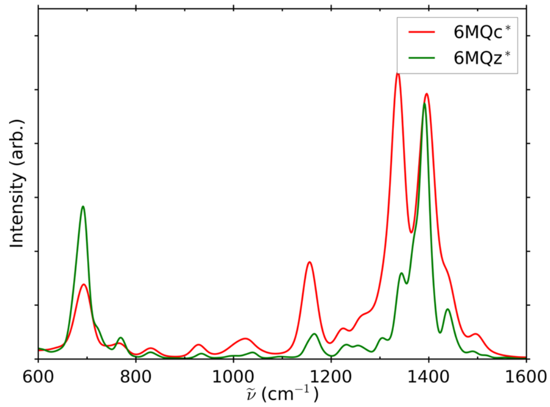


**Figure 8.** Fourier transform of the C−O bond distance velocity autocorrelation of 6MQc* and 6MQz* averaged over 11 and four trajectories, respectively.

of the C−O bond in the simulations gets shortened by ∼0.05 Å as shown in Figure S7 in the Supporting Information. The fact that this shortening does not lead to a blue shift of the vibrational frequencies as in the normal modes spectra can be explained by the strong hydrogen bonding of water molecules to the partially negatively charged oxygen, which restrict the motions of the C−O moiety (cf. normal mode H at 1413 $cm^{-1}$ for 6MQ* and mode I at 1445 $cm^{-1}$ for 6MQz* in Supporting Information).

Regarding the intensities of the individual (single-trajectory based) power spectra underlying the averaged spectrum in Figure 8, we found strong variations for the two protonation states and for each trajectory suggesting a complex dependency on the actual atomic conformation and hydrogen bonding configuration around the hydroxyl group.

As mentioned before, all presented MD spectra are computed based on simulations of fully deuterated samples. A comparison with experimental data obtained from non-deuterated samples therefore must take into account isotope effects. We investigated absorptions for nondeuterated samples by comparing the normal mode spectra for 6MQc, 6MQc*, and 6MQz* (see Supporting Information) and found blue shifts of absorptions above 2000 $cm^{-1}$ by ∼1000 $cm^{-1}$ in the nondeuterated case. This suggests that C−H stretch and O−H stretch vibrations will be observed at higher frequencies in experimental spectra of nondeuterated samples. We found no overall absorption shift below 2000 $cm^{-1}$ but instead observed a reordering of normal modes resulting in a distinct absorption pattern compared to the deuterated case (cf. Supporting Information). As a consequence, the possible marker at 1300 $cm^{-1}$ indicative of 6MQc* or 6MQz* in the deuterated case will be different for nondeuterated samples. The normal mode spectra in the latter case feature as well an altered but distinct peak pattern upon proton release in the fingerprint region below 2000 $cm^{-1}$. Therefore, this regime might be suited as a marker region in transient IR measurements of the non-deuterated chromophore similarly as in the deuterated case.

## 4. CONCLUSION

We have investigated the photoinduced proton dissociation processes from the super photoacid 6MQc in aqueous solution from first-principles MD simulations. Besides the prototypical red-shifting effects of excitation on the intramolecular modes of the solute, we find a broad IR absorption line from 1500 to 2000 $cm^{-1}$ after excitation and before deprotonation of the acid. Our analysis shows that this line is due to the loose bonding situation of the acidic proton just before dissociation. Right afterward, we observe the transient formation of a contact ion pair with a characteristic IR signature of ∼1300 $cm^{-1}$ that we predict to be observable experimentally, provided a sufficient time resolution (femtosecond) is possible.

We also investigated hydrogen bond rearrangements upon electronic excitation and structural differences between the contact ion pair geometries. In these terms, the rate-limiting step for the proton transfer to the first solvent molecule is found not to be the initial acid−water barrier but instead the solvation structure around the first water molecule. Further, we elucidate the nature of the Grotthuss-style dissociation/transport pathway in terms of stepwise mechanism versus concerted motion across multiple water molecules. Our findings indicate that the concerted mechanism is possible if the accepting water molecule is fourfold coordinated. In the case of only threefold coordination, the character of the protonation pathway is rather stepwise.

Our results will support the experimental assignment of transient IR signatures from femtosecond spectroscopy, providing an atomistic interpretation of the underlying short-lived molecular conformations of such photoacids. The findings on the coordination dependence of the nature of the proton pathway illustrate why the long-standing debate on the stepwise versus concerted mechanisms will prevail, as there both mechanisms can coexist for flexible coordination states of the involved water molecules.

## ■ ASSOCIATED CONTENT

### Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpca.5b05438.

> Computational details and validation of the ab initio molecular dynamics simulations (computed properties in the $S_0$, $S_1$, and $T_1$ states), as well as the assignment of IR resonances of the compounds. (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

*E-mail: daniel.sebastiani@chemie.uni-halle.de.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We are grateful to Dr. E. T. J. Nibbering (Max-Born-Institut, Berlin) for useful discussions and suggestions and for reading an early version of the manuscript. Computing infrastructure was provided by the KIT-Hochleistungsrechner of the Steinbuch Centre for Computing. This work was supported by Leibniz Graduate School of Molecular Biophysics Berlin, Kekulé-Stipendium, and German Research Foundation (DFG) under Grant No. SE 1008/11-1.

## ■ REFERENCES

(1) Bort, G.; Gallavardin, T.; Ogden, D.; Dalko, P. I. From One-Photon to Two-Photon Probes: "Caged" Compounds, Actuators, and Photoswitches. *Angew. Chem., Int. Ed.* **2013**, *52*, 4526−4537.

(2) Bucher, D. B.; Schlueter, A.; Carell, T.; Zinth, W. Watson-Crick Base Pairing Controls Excited-State Decay in Natural DNA. *Angew. Chem., Int. Ed.* **2014**, *53*, 11366−11369.

(3) Conrad, K. S.; Manahan, C. C.; Crane, B. R. Photochemistry of Flavoprotein Light Sensors. *Nat. Chem. Biol.* **2014**, *10*, 801−809.

(4) Morrone, J. A.; Tuckerman, M. E. Ab Initio Molecular Dynamics Study of Proton Mobility in Liquid Methanol. *J. Chem. Phys.* **2002**, *117*, 4403−4413.

(5) Izvekov, S.; Voth, G. A. Ab Initio Molecular-Dynamics Simulation of Aqueous Proton Solvation and Transport Revisited. *J. Chem. Phys.* **2005**, *123*, 044505.

(6) Geissler, P.; Dellago, C.; Chandler, D.; Hutter, J.; Parrinello, M. Ab Initio Analysis of Proton Transfer Dynamics in (H2O)3H+. *Chem. Phys. Lett.* **2000**, *321*, 225−230.

(7) Hassanali, A. A.; Cuny, J.; Ceriotti, M.; Pickard, C. J.; Parrinello, M. The Fuzzy Quantum Proton in the Hydrogen Chloride Hydrates. *J. Am. Chem. Soc.* **2012**, *134*, 8557−8569.

(8) Tuckerman, M. E.; Marx, D.; Klein, M. L.; Parrinello, M. On the Quantum Nature of the Shared Proton in Hydrogen Bonds. *Science* **1997**, *275*, 817−820.

(9) Ludueña, G. A.; Sebastiani, D. Possibility of Coherent Delocalized Nuclear Quantum States of Protons in $Li_2NH$. *J. Phys. Chem. Lett.* **2010**, *1*, 3214−3218.

(10) Ceriotti, M.; Bussi, G.; Parrinello, M. Nuclear Quantum Effects in Solids Using a Colored-Noise Thermostat. *Phys. Rev. Lett.* **2009**, *103*, 030603.

(11) Ceriotti, M.; Miceli, G.; Pietropaolo, A.; Colognesi, D.; Nale, A.; Catti, M.; Bernasconi, M.; Parrinello, M. Nuclear Quantum Effects in Ab Initio Dynamics: Theory and Experiments for Lithium Imide. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2010**, *82*, 174306.

(12) Benoit, M.; Marx, D. The Shapes of Protons in Hydrogen Bonds Depend on the Bond Length. *ChemPhysChem* **2005**, *6*, 1738−1741.

(13) Marx, D.; Benoit, M.; Parrinello, M. Tunnelling and Zero-Point Motion in High-Pressure Ice. *Nature* **1998**, *392*, 258−261.

(14) Tuckerman, M. E.; Marx, D. Heavy-Atom Skeleton Quantization and Proton Tunneling in "Intermediate-Barrier" Hydrogen Bonds. *Phys. Rev. Lett.* **2001**, *86*, 4946−4949.

(15) Leiderman, P.; Genosar, L.; Huppert, D. Excited-State Proton Transfer: Indication of Three Steps in the Dissociation and Recombination Process. *J. Phys. Chem. A* **2005**, *109*, 5965−5977.

(16) Bekçioğlu, G.; Allolio, C.; Ekimova, M.; Nibbering, E. T. J.; Sebastiani, D. Competition between Excited State Proton and OH- Transport via a Short Water Wire: Solvent Effects Open the Gate. *Phys. Chem. Chem. Phys.* **2014**, *16*, 13047−13051.

(17) Bekçioğlu, G.; Allolio, C.; Sebastiani, D. Water Wires in Aqueous Solutions from First-Principles Calculations. *J. Phys. Chem. B* **2015**, *119*, 4053−4060.

(18) Rini, M.; Magnes, B.-Z.; Pines, E.; Nibbering, E. T. J. Real-Time Observation of Bimodal Proton Transfer in Acid-Base Pairs in Water. *Science* **2003**, *301*, 249−352.

(19) Mohammed, O. F.; Pines, D.; Dreyer, J.; Pines, E.; Nibbering, E. T. J. Sequential Proton Transfer Through Water Bridges in Acid-Base Reactions. *Science* **2005**, *310*, 83−86.

(20) Adamczyk, K.; Prémont-Schwarz, M.; Pines, D.; Pines, E.; Nibbering, E. T. J. Real-Time Observation of Carbonic Acid Formation in Aqueous Solution. *Science* **2009**, *326*, 1690−1694.

(21) Allolio, C.; Sajadi, M.; Ernsting, N.; Sebastiani, D. An Ab Initio Microscope: Molecular Contributions to the Femtosecond Time-Dependent Fluorescence Shift of a Reichardt-Type Dye. *Angew. Chem., Int. Ed.* **2013**, *52*, 1813−1816.

(22) Guglielmi, M.; Tavernelli, I.; Rothlisberger, U. On the Proton Transfer Mechanism in Ammonia-Bridged 7-Hydroxyquinoline: A TDDFT Molecular Dynamics Study. *Phys. Chem. Chem. Phys.* **2009**, *11*, 4549−4555.

(23) Xia, S.-H.; Xie, B.-B.; Fang, Q.; Cui, G.; Thiel, W. Excited-State Intramolecular Proton Transfer to Carbon Atoms: Nonadiabatic Surface-Hopping Dynamics Simulations. *Phys. Chem. Chem. Phys.* **2015**, *17*, 9687−9697.

(24) Veiga-Gutiérrez, M.; Brenlla, A.; Carreira Blanco, C.; Fernández, B.; Kovalenko, S. A.; Rodríguez-Prieto, F.; Mosquera, M.; Lustres, J. L. P. Dissociation of a Strong Acid in Neat Solvents: Diffusion is Observed after Reversible Proton Ejection inside the Solvent Shell. *J. Phys. Chem. B* **2013**, *117*, 14065−14078.

(25) Gould, E.-A.; Popov, A. V.; Tolbert, L. M.; Presiado, I.; Erez, Y.; Huppert, D.; Solntsev, K. M. Excited-State Proton Transfer in N-Methyl-6-Hydroxyquinolinium Salts: Solvent and Temperature Effects. *Phys. Chem. Chem. Phys.* **2012**, *14*, 8964−8973.

(26) Popov, A. V.; Gould, E.-A.; Salvitti, M. A.; Hernandez, R.; Solntsev, K. M. Diffusional Effects on the Reversible Excited-State Proton Transfer. From Experiments to Brownian Dynamics Simulations. *Phys. Chem. Chem. Phys.* **2011**, *13*, 14914−14927.

(27) Bardez, E.; Boutin, P.; Valeur, B. Photoinduced Biprotonic Transfer in 4-Methylumbelliferone. *Chem. Phys. Lett.* **1992**, *191*, 142−148.

(28) Bardez, E.; Chatelain, A.; Larrey, B.; Valeur, B. Photoinduced Coupled Proton and Electron Transfers. 1. 6-Hydroxyquinoline. *J. Phys. Chem.* **1994**, *98*, 2357−2366.

(29) Pérez-Lustres, J. L.; Rodriguez-Prieto, F.; Mosquera, M.; Senyushkina, T. A.; Ernsting, N. P.; Kovalenko, S. A. Ultrafast Proton Transfer to Solvent: Molecularity and Intermediates from Solvation- and Diffusion-Controlled Regimes. *J. Am. Chem. Soc.* **2007**, *129*, 5408−5418.

(30) Poizat, O.; Bardez, E.; Buntinx, G.; Alain, V. Picosecond Dynamics of the Photoexcited 6-Methoxyquinoline and 6-Hydroxyquinoline Molecules in Solution. *J. Phys. Chem. A* **2004**, *108*, 1873−1880.

(31) Bardez, E.; Fedorov, A.; Berberan-Santos, M. N.; Martinho, J. M. G. Photoinduced Coupled Proton and Electron Transfers. 2. 7-Hydroxyquinolinium Ion. *J. Phys. Chem. A* **1999**, *103*, 4131−4136.

(32) Kim, T. G.; Topp, M. R. Ultrafast Excited-State Deprotonation and Electron Transfer in Hydroxyquinoline Derivatives. *J. Phys. Chem. A* **2004**, *108*, 10060−10065.

(33) Eigen, M. Proton Transfer, Acid-Base Catalysis and Enzymatic Hydrolysis. *Angew. Chem., Int. Ed. Engl.* **1964**, *3*, 1−19.

(34) Weller, A. Fast Reactions of Excited Molecules. *Prog. React. Kinet.* **1961**, *1*, 187.

(35) Eigen, M.; Kruse, W.; Maeyer, L. D. Über den Zustand des Protons (Hydroniumions) in Wässriger Lösung. *Prog. React. Kinet.* **1964**, *2*, 285.

(36) Thomas, V.; Maurer, P.; Iftimie, R. On the Formation of Proton-Shared and Contact Ion Pair Forms during the Dissociation of Moderately Strong Acids: An Ab Initio Molecular Dynamics Investigation. *J. Phys. Chem. B* **2010**, *114*, 8147−8155.

(37) Allolio, C.; Sebastiani, D. Approaches to the Solvation of the Molecular Probe N-Methyl-6-Quinolone in its Excited State. *Phys. Chem. Chem. Phys.* **2011**, *13*, 16395−16403.

(38) Iftimie, R.; Tuckerman, M. E. Decomposing Total IR Spectra of Aqueous Systems into Solute and Solvent Contributions: A Computational Approach Using Maximally Localized Wannier Orbitals. *J. Chem. Phys.* **2005**, *122*, 214508.

(39) Lippert, G.; Hutter, J.; Parrinello, M. A Hybrid Gaussian and Plane Wave Density Functional Scheme. *Mol. Phys.* **1997**, *92*, 477−487.

(40) Hutter, J.; Iannuzzi, M.; Schiffmann, F.; VandeVondele, J. CP2K: Atomistic Simulations of Condensed Matter Systems. *WIREs. Comput. Mol. Sci.* **2014**, *4*, 15−25.

(41) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098.

(42) 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: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789.

(43) Goedecker, S.; Teter, M.; Hutter, J. Separable Dual-Space Gaussian Pseudopotentials. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 1703−1710.

(44) VandeVondele, J.; Mohamed, F.; Krack, M.; Hutter, J.; Sprik, M.; Parrinello, M. The Influence of Temperature and Density Functional Models in Ab Initio Molecular Dynamics Simulation of Liquid Water. *J. Chem. Phys.* **2005**, *122*, 014515.

(45) Sorenson, J. M.; Hura, G.; Glaeser, R. M.; Head-Gordon, T. What Can X-ray Scattering Tell Us about the Radial Distribution Functions of Water? *J. Chem. Phys.* **2000**, *113*, 9149−9161.

(46) Grimme, S. Semiempirical GGA-Type Density Functional Constructed with a Long-Range Dispersion Correction. *J. Comput. Chem.* **2006**, *27*, 1787−1799.

(47) Frisch, M. J. et al. *Computer code Gaussian 09;* Gaussian, Inc: Wallingford, CT, 2011.

(48) Becke, A. D. Density-Functional Thermochemistry. 3. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(49) Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **1998**, *102*, 1995−2001.

(50) Silvestrelli, P. L.; Parrinello, M. Water Molecule Dipole in the Gas and in the Liquid Phase. *Phys. Rev. Lett.* **1999**, *82*, 3308.

(51) Thomas, M.; Brehm, M.; Fligg, R.; Vöhringer, P.; Kirchner, B. Computing Vibrational Spectra from Ab Initio Molecular Dynamics. *Phys. Chem. Chem. Phys.* **2013**, *15*, 6608−22.

(52) Ramirez, R.; Lopez-Ciudad, T.; Kumar P, P.; Marx, D. Quantum Corrections to Classical Time-Correlation Functions: Hydrogen Bonding and Anharmonic Floppy Modes. *J. Chem. Phys.* **2004**, *121*, 3973.

(53) Agmon, N. Elementary Steps in Excited-State Proton Transfer. *J. Phys. Chem. A* **2005**, *109*, 13−35.

(54) Pines, E.; Huppert, D. Geminate Recombination Proton-Transfer Reactions. *Chem. Phys. Lett.* **1986**, *126*, 88−91.

(55) Berkelbach, T. C.; Lee, H.-S.; Tuckerman, M. E. Concerted Hydrogen-Bond Dynamics in the Transport Mechanism of the Hydrated Proton: A First-Principles Molecular Dynamics Study. *Phys. Rev. Lett.* **2009**, *103*, 238302.

(56) Tuckerman, M. E.; Laasonen, K.; Sprik, M.; Parrinello, M. Ab Initio Molecular Dynamics Simulation of the Solvation and Transport of H3O+ and OH- Ions in Water. *J. Phys. Chem.* **1995**, *99*, 5749−5752.

(57) Maurer, P.; Thomas, V.; Iftimie, R. A Computational Study of Ultrafast Acid Dissociation and Acid-Base Neutralization Reactions. II. The Relationship between the Coordination State of Solvent Molecules and Concerted versus Sequential Acid Dissociation. *J. Chem. Phys.* **2011**, *134*, 094505.

(58) Agmon, N.; Rettig, W.; Groth, C. Electronic Determinants of Photoacidity in Cyanonaphthols. *J. Am. Chem. Soc.* **2002**, *124*, 1089.

(59) Hirshfeld, F. L. Bonded-Atom Fragments for Describing Molecular Charge Densities. *Theor. Chem. Acc.* **1977**, *44*, 129−138.

(60) Iftimie, R.; Tuckerman, M. E. The Molecular Origin of the ”Continuous” Infrared Absorption in Aqueous Solutions of Acids: A Computational Approach. *Angew. Chem., Int. Ed.* **2006**, *45*, 1144−7.

(61) Cox, M. J.; Timmer, R. L. A.; Bakker, H. J.; Park, S.; Agmon, N. Distance-Dependent Proton Transfer along Water Wires Connecting Acid-Base Pairs. *J. Phys. Chem. A* **2009**, *113*, 6599−6606.

(62) Siwick, B. J.; Bakker, H. J. On the Role of Water in Intermolecular Proton Transfer Reactions. *J. Am. Chem. Soc.* **2007**, *129*, 13412−13420.

DOI: 10.1021/acs.jpca.5b05438
*J. Phys. Chem. A* 2015, 119, 9244−9251

# Combined Experimental and Theoretical Study of the Transient IR Spectroscopy of 7-Hydroxyquinoline in the First Electronically Excited Singlet State

Felix Hoffmann, Maria Ekimova, Gül Bekçioğlu-Neff, Erik T. J. Nibbering, and Daniel Sebastiani

Reprinted with permission from

J. Phys. Chem. A **2016**, 120, 9378–9389.

Copyright 2017 American Chemical Society.

`http://dx.doi.org/10.1021/acs.jpca.6b07843`

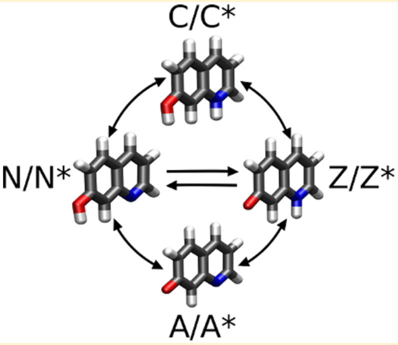
THE JOURNAL OF PHYSICAL CHEMISTRY A
Article
pubs.acs.org/JPCA

# Combined Experimental and Theoretical Study of the Transient IR Spectroscopy of 7-Hydroxyquinoline in the First Electronically Excited Singlet State

Felix Hoffmann,[†] Maria Ekimova,[‡] Gül Bekçioğlu-Neff,[¶,†] Erik T. J. Nibbering,[*,‡] and Daniel Sebastiani[*,†]

[†]Institut für Chemie, Martin-Luther-Universität Halle-Wittenberg, Von-Danckelmann-Platz 4, 06120 Halle (Saale), Germany
[‡]Max Born Institut für Nichtlineare Optik und Kurzzeitspektroskopie, Max Born Strasse 2A, 12489 Berlin, Germany
[¶]Physics Department, Freie Universität Berlin, Arnimallee 14, 14195 Berlin, Germany

*Supporting Information*



**ABSTRACT:** The photophysics of 7-hydroxyquinoline (7HQ) in protic media results from an interplay of acid−base chemistry, prompted by the effects of photoacidity of the hydroxyl group and photobasicity of the nitrogen atom in the quinoline aromatic system. With ultrafast IR spectroscopic measurements, we follow the proton transfer dynamics of 7HQ in its four possible charged forms in methanol solution. Using deuterated methanol as solvent, we determine deuteron transfer rates from the neutral to the zwitterionic form to be 330 ps, those from the cationic form to the zwitterionic form to be 170 ps, and those from the anionic form to the zwitterionic form to be 600 ps. We compare the observed IR-active fingerprint marker patterns in the electronic ground state and the first electronically excited $^1L_b$-state with those calculated using density functional theory and time-dependent density functional theory, respectively, and find good correspondence between experimental and calculated transitions. The calculations provide insight into the nature of electronic excitation of these four different charged forms of 7HQ, suggesting the key role of electronic charge distribution changes upon electronic excitation of 7HQ and hydrogen bond changes at the donor hydroxyl and acceptor nitrogen moieties.

## 1. INTRODUCTION

Proton transfer (PT) reactions play a key role in numerous chemical and biological processes as diverse as the autoionization of water,[1] the von Grotthuss mechanism,[2−7] proton conductivity in hydrogen fuel cells,[8−10] hydroxide ion transport,[11] acid−base neutralization reactions,[12−15] and PT through membranes.[16] The underlying microscopic mechanisms and associated dynamics of PT can be rather involved for such condensed-phase systems, showing an interplay of elementary PT steps, solvent shell rearrangements, and diffusional motions. Structurally well-defined systems appear to be a promising means for a better access to the intrinsic PT steps.

In this regard, bifunctional chromophores with proton-donating and -accepting moieties at fixed distances have been proven valuable tools in time-resolved PT studies as they allow for controlled triggering of the reaction with a UV pump pulse.[17−36] These special classes of photoacids (or photobases), showing excited-state proton transfer (ESPT) from the donating to the accepting groups, have been studied in protic solvents, such as water, ice, and alcohols, and even in alcohol/ether mixtures. PT has often been understood to occur via a prearranged number of water or alcohol molecules, akin to the concept of “wires” or “bridges”.[13−15,37−41] The actual ESPT mechanism might be either stepwise or concerted, depending on solvent polarity and solvent shell configurations surrounding the photoacid or photobase, with a key role played by the dynamical aspects of hydrogen bond rearrangements.[3,6,15,42]

A prime example for this class of bifunctional photoacids is 7-hydroxyquinoline (7HQ) (Figure 1).[17−35] In the electronic ground state, 7HQ can adopt four equilibrium species, a neutral enol form (N), which is the predominant form in the ground state ($S_0$), a deprotonated enol-anion (A), an imine protonated cation (C), and a keto or zwitterionic form (Z).[43,44] In aqueous solution, the C form has been reported to be the dominant species at low pH conditions (pH < 2), and the A form is generated for basic solutions (pH > 10). Whereas the N form dominates in most other solvents, both the N and Z forms are quantitatively formed under aqueous solution conditions at neutral pH. Because of having four species in equilibrium through acid/base chemistry, one can consider different Förster cycles that 7HQ can undergo by photoexcitation (see Figure 1b): (I) N → N* → A* → A → N; (II) C → C* → Z* → Z → C; (III) A → A* → Z* → Z → A; (IV) N → N* → C* → C → N. Förster cycles (I) and (II) follow the photoacid pathway, that is, electronic excitation is followed by ESPT from the hydroxyl moiety to the solvent, and conversion back to the

**Received:** August 3, 2016
**Revised:** September 28, 2016
**Published:** November 7, 2016

ACS Publications © 2016 American Chemical Society

DOI: 10.1021/acs.jpca.6b07843
*J. Phys. Chem. A* 2016, 120, 9378−9389

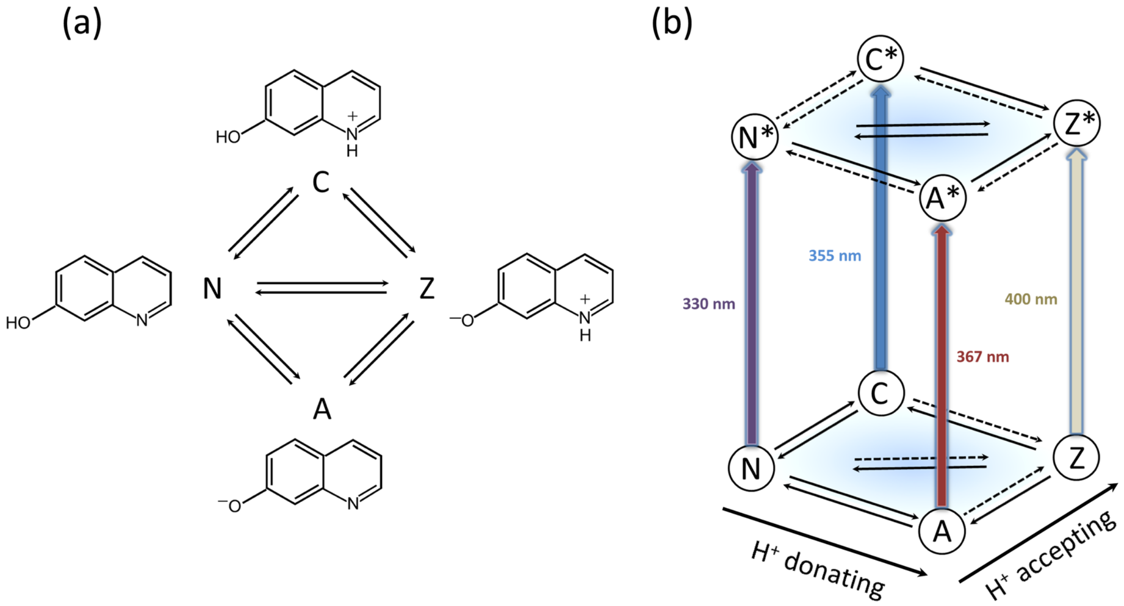


**Figure 1.** (a) Different charged forms of 7HQ: neutral form (N), zwitterionic form (Z), cationic form (C), and anionic form (A). (b) PT pathways connecting the respective N, C, A, and Z species in the $S_0$ and $S_1$ states as well as locations of the electronic transitions.

electronic ground state is accompanied by the concomitant proton back transfer. In contrast, cycles (III) and (IV) are prompted by the effect of photobasicity, that is, the increased basicity of the quinoline nitrogen upon electronic excitation, leading to proton abstraction from the solvent. In addition to that, the combined effect of photoacidity of the hydroxyl moiety and the photobasicity of the quinoline nitrogen can result in a Förster cycle N → N* → Z* → Z → N. Here it remains an open question whether the route of N → Z* can occur in a direct concerted fashion or stepwise via either the Z* or A* charged states.

Reaction rates are clearly high in aqueous solution; Bardez and co-workers have derived an N* to Z* conversion at neutral pH ≈ 7 to be $>10^{11}$ $s^{-1}$, that is, with a time constant of <10 ps.[23] Instead, at low pH < 1, the cationic C* form seems to show only an electronic excited-state decay with a time constant of 9.7 ns. The reason for this is that at such low pH conditions the effective acid/base equilibrium remains on the acid side. An estimate for the real proton transfer to solvent (PTTS) time constant of 18.1 ps was made from a kinetic modeling of measurements performed at higher pH conditions. Even though a similar effect of predominant electronically excited-state decay was observed for the A* form under high pH ≈ 13 conditions, an estimate of the time constant of proton abstraction by the 7HQ nitrogen atom of the photoexcited A* form of 181 ps was derived for neutral pH conditions. Kwon and Mohammed concluded from time-resolved fluorescence measurements on 7HQ in water that a fast picosecond component can be ascribed to a water wire arrangement through which the proton is transferred.[33] These short picosecond time scales for the different PT steps of 7HQ in the $S_1$ state in aqueous solution are in strong contrast to the reported ESPT rates from the N* to the Z* tautomer, resulting from photoexcitation of the N form to the $S_1$ state with a pump pulse at 330 nm, observed to occur with a time constant of 170 ps in methanol.[17−20,22] Even slower rates have been reported for 7HQ dissolved in alcohol/ether or alcohol/alkane mixtures.[25,27,29,32,33,45]

For 7HQ-$(NH_3)_3$ in the gas phase, a combined electron transfer and PT, that is, a net hydrogen transfer, was postulated to occur upon electronic excitation to the $S_1$ state along “ammonia wires”.[46] This net hydrogen transfer is a result of a nonadiabatic transition from the $^1\pi\pi^*$ to a $\pi\sigma^*$ Rydberg state.[46,47] Instead, in polar protic media, in particular, water, a net PT takes place for 7HQ remaining in the $^1\pi\pi^*$ $S_1$ state. Our previous studies showed that a hydrogen-bonded water chain connecting hydroxyl and nitrogen moieties is stable on the picosecond time scale for 7HQ in aqueous solution.[48] This “water wire” preformed in the electronic ground state acts as a relay for ESPT directly upon initiation of the reaction through photoexcitation of 7HQ. Others have suggested that similar solvent configurations are responsible for facilitating efficient PT in methanol solution and in methanol clusters.[17−19,49,50] In some of these works, the importance of proton tunneling has been emphasized as the underlying mechanism for ESPT.[18]

Until now, time-resolved studies of 7HQ and other hydroxyquinolines have been performed utilizing probing techniques of electronic transitions by time-resolved emission or transient UV/vis absorption. Here we report results of time-resolved probing of IR-active transitions of the different forms of 7HQ in methanol solution. We show results obtained by photoexcitation of the C, N, and A forms of 7HQ in methanol and identify fingerprint marker bands by a comparison of experimental and theoretical results.

## 2. EXPERIMENTAL AND COMPUTATIONAL DETAILS

**2.1. Steady-State and Transient Spectroscopic Measurements.** 7HQ (99%) was purchased from ACROS Organics, the deuterated solvents $CD_3OD$ (99.8%), DMSO-$d_6$ (99.8%), $D_2O$ (99.95%), and $D_2SO_4$ (96−98 wt % in $D_2O$, 99.5% D atom) were from Deutero, and NaOH (99%) was from Merck. All chemicals were used without further purification.

Linear absorption and fluorescence spectra were obtained using a PerkinElmer spectrometer and a JOBIN YVON Horiba fluorolog, respectively. The concentration of 7HQ was

DOI: 10.1021/acs.jpca.6b07843
J. Phys. Chem. A 2016, 120, 9378−9389

approximately $10^{-5}$ mol/L, providing an optical density between 0.5 and 0.6 in a 10 mm thick quartz cuvette. Steady-state FT-IR spectra were recorded with a Varian 640 FT-IR spectrometer. The solutions were placed between 1 mm $CaF_2$ windows with a sample thickness of 100 $\mu$m fixed by Teflon spacers. The typical concentration of 7HQ for the pump−probe and FT-IR measurements was 40−50 mM. Deuteration of the hydroxy group of 7-HQ in DMSO-$d_6$ was obtained by adding a 100-fold excess amount of $D_2O$. Sulfuric acid and sodium hydroxide in concentrations of 100 and 500 mM, respectively, were used to prepare the cationic (C) and anionic (A) forms of 7-HQ in $CD_3OD$, respectively. Full deprotonation/protonation was confirmed by comparing to steady-state electronic absorption spectra measured under the same sample conditions.

Ultrafast UV pump−IR probe measurements were performed as described previously.[51] In short, the excitation beam was generated from an optical parametric amplifier (TOPAS-C, Light Conversion), pumped by an 800 nm beam (Tsunami oscillator with Spitfire Pro regenerative and booster amplifier stages, Spectra Physics), producing 50 fs pulses at 330 nm and a 1 kHz repetition rate. The pump pulses were stretched by propagating the UV beam through a cuvette with water, resulting in a pump−probe cross-correlation width of about 350 fs. The temporal resolution was dominated by the stretched pump pulse as the probe pulse duration was typically 100 fs and group velocity mismatch between pump and probe pulses typically only leads to a temporal resolution of 150−200 fs with this experimental setup. The pulse delay between pump and probe beams was achieved with a delay stage, before the beams were focused onto the sample. The UV pulses were focused to 200 $\mu$m spot sizes with pulse energies of ~1 $\mu$J at the sample. Tunable IR pulses were generated by a home-built double-pass optical parametric amplifier, followed by difference frequency mixing of the signal and idler. The output beam was sent to a ZnSe wedge, and the reflected pulses from its front and back surfaces were used as probe and references beams, respectively. Both probe and reference pulses were focused onto the sample by means of an off-axis parabolic mirror (focal diameter 200 $\mu$m), and after the samples were dispersed in a polychromator (2 $cm^{-1}$ resolution in the 1200−1700 $cm^{-1}$ spectral range) and spectrally resolved, absorbance changes were recorded using a liquid-nitrogen-cooled HgCdTe double-array detector (2 × 64 pixels). A peristaltic pump was used to circulate the sample through a flow cell (1 mm thick $CaF_2$ windows separated by a Teflon spacer of 50 $\mu$m). To confirm that no sample degradation had occurred during the experiments, steady-state UV/vis and FT-IR absorption spectra of the samples were recorded before and after each pump−probe measurement.

**2.2. Computational Details.** All calculations were carried out with Gaussian09[52] on the DFT/TDDFT[53] level of theory using the B3LYP[54] exchange−correlation functional together with D3 dispersion correction[55] and the TZVP basis set. In order to capture solvation effects, the conductor polarized continuum model was adopted using methanol as the solvent.[56] In addition, we explicitly included three methanol molecules solvating the polar groups of 7HQ, as depicted for N/N* in Figure 2. The methanol molecules were assembled in a "wired" configuration to allow for PT from the proton-donating to the -accepting chromophore moiety. This choice was motivated by earlier studies by us in aqueous solution where we found three membered solvent configurations to be the shortest and most stable solvent wires connecting proton-donating and -accepting groups in 7HQ.[48] In accordance with experiments, all methanol molecules as well as the hydroxy and nitrogen moieties of the chromophore have been deuterated. Geometry and transition state optimizations in the $S_0$ and $S_1$ states were performed using $10^{-6}$ au as the convergence criterion for the RMS force without imposing symmetry restrictions. Vertical excitations were calculated from the respective geometries optimized in the electronic ground state.

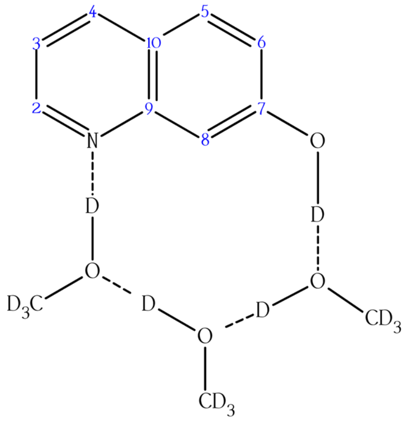


**Figure 2.** Schematic of N/N* showing the explicit solvent configuration. Deuterons are labeled with D, and deuteron bonds are indicated with dashed lines.

In order to account for anharmonicities in the experimental spectra, we adjusted the harmonic frequencies in two ways. First, normal-mode frequencies were corrected based on a perturbation theory approach as implemented in Gaussian 09. For TDDFT calculations, we additionally obtained a global scaling factor of 0.9763 derived from a linear fit to the perturbation corrected frequencies in the ground state. Hereby we assume that systematic errors in TDDFT are similar to those in ground-state DFT, which has been confirmed recently by Liu and co-workers[57] on a test set of organic and inorganic molecules. We also compared the effect of other scaling factors on the results. For instance, a general scaling factor suggested by Merrick and co-workers[58] of 0.9688 leads to a red shift of about 10 $cm^{-1}$ compared to our approach and therefore does not turn out superior in this case.

The estimated reaction time constants presented in Table 2 were obtained from uncorrected harmonic frequencies based on excited-state and excited-state transition-state-optimized geometries, respectively. For more details, see eq 1 in the Supporting Information.

## 3. RESULTS

**3.1. Electronic Absorption and Emission Spectra.** As shown in Figure 1, acid−base chemistry of 7HQ connects the different charged forms and tautomers. Depending on the solvent medium and pH value, one can predominantly form the neutral tautomer N in either methanol or DMSO, or by adding small equimolar amounts of $H_2SO_4$ in methanol, 7HQ converts into the cationic species C, whereas by adding NaOH, the anion A is formed. This behavior is clear from the measured electronic absorption spectra (Figure 3), where for methanol and DMSO one observes only the $^1L_b$ transition band of the N tautomer peaking at 330 nm. The first electronic transition band of the cationic form is located at 355 nm, whereas the

DOI: 10.1021/acs.jpca.6b07843
J. Phys. Chem. A 2016, 120, 9378−9389

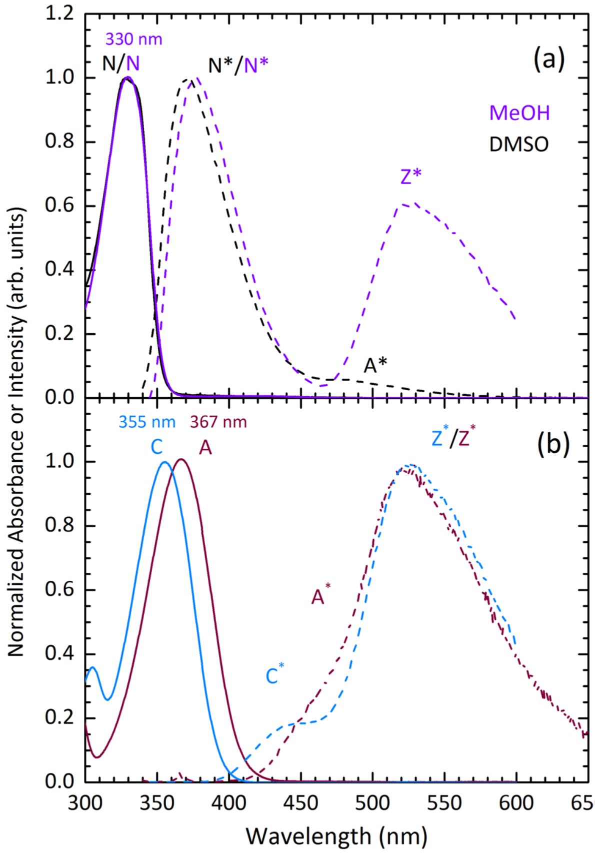


**Figure 3.** Electronic absorption (solid curves) and emission (dashed curves) spectra of 7HQ. (a) Spectra of 7HQ dissolved in methanol (purple curves) and DMSO (black curves); (b) spectra of 7HQ in methanol, after addition of sulfuric acid (cyan curves) or sodium hydroxide (brown curves).

anionic form has its first electronic transition band peaking at 367 nm. Similar observations were reported on by Bardez for aqueous solutions of 7HQ.[23] An interesting detail is the efficient formation of the cationic form C by adding about equimolar amounts of $H_2SO_4$ compared to 7HQ in methanol, whereas in water, much higher proton concentrations are necessary to quantitatively shift the equilibrium to the C form.

Fluorescence emission spectra of 7HQ provide first insight into the ESPT dynamics. As reported before,[19,20,22] 7HQ excited to the N* state exhibits dual fluorescence emission properties in methanol, as indicated by the fluorescence bands of the N* and the Z* forms (Figure 3a). On the basis of the reported value for PT dynamics of 7HQ in methanol (170 ps time constant, i.e. at least 1 order of magnitude longer than that in water), it appears that electronic excitation of the N tautomer results in acid−base reaction dynamics, installing a new equilibrium between N* and Z* in the $S_1$ state of 7HQ. Instead, 7HQ in DMSO does predominantly emit in its N* tautomer state, although a tail in the emission may hint at formation of the anionic A* form, suggesting a minute fraction of electronically excited 7HQ releasing a proton to the solvent. Dual fluorescence is also observed for the C* and A* forms; however, here the PT product Z* tautomer dominates the emission spectrum (Figure 3b).

We additionally corroborated the assignments of absorbance and emission bands by calculating vertical excitation and emission energies. As seen in Table 1, we find good agreement between experimental and calculated values, where deviations are within the known error of the TDDFT/B3LYP.[59,60] From the percentage of HOMO−LUMO character, it is clear that excitations are of $\pi\pi^*$ character. Further, analysis of the electronic transition dipole moments for N, C, A, and Z confirms that the electronic transition is of $^1L_b$ nature for all 7HQ species (see the Supporting Information for details).

**Table 1. Calculated Vertical Absorption Energies and Experimental UV Absorption Maxima (upper part) as Well as Vertical Emission Energies and Experimental Fluorescence Energies (lower part) for N, C, and A along with Percentages of HOMO−LUMO Character**[a]

| excitation | calcd. (nm) | HOMO−LUMO (%) | expt. (nm) |
|---|---|---|---|
| N → N* | 320.0 | 94.1 | 330 |
| Z → Z* | 406.5 | 95.6 | |
| C → C* | 343.6 | 94.0 | 355 |
| A → A* | 381.2 | 96.5 | 367 |
| N* → N | 374.0 | 97.7 | 375 |
| Z* → Z | 500.3 | 98.1 | 521 |
| C* → C | 416.3 | 97.7 | 440 |
| A* → A | 452.2 | 98.0 | 470 |

[a] For experimental values, compare Figure 3.

**3.2. Transient IR Measurements.** Transient IR spectra have been recorded for 7HQ in deuterated solvents $CD_3OD$ or DMSO-$d_6$ (with small amounts of $D_2O$ added to warrant H/D exchange at the hydroxyl group of 7HQ in the $S_0$ state), monitoring the fingerprint region from 1300 to 1650 $cm^{-1}$. Figure 4 shows the transient IR spectra recorded at pulse delay times from 0 ps to 1 ns, where the absorbance change $\Delta A = -\log_{10}(T/T_0)$ is plotted as a function of probe frequency for different pump−probe delays ($T$;$T_0$: sample transmission with and without excitation). Besides typical bleach components due to decreased absorption of vibrational transitions of 7HQ in the $S_0$ state (shown as the black dashed line in Figure 5), one can clearly observe a broad-band spectral feature at early delay times decaying with subpicosecond and picosecond components. This feature can (considering its dynamical characteristics) be related to multiphoton excitation effects of the solvent, with a minor possible contribution of the O−D stretching mode of the N* or C* forms or the N−H stretching mode of C* of 7HQ in the $S_1$ state or even an O−D stretching mode of a solvent molecule hydrogen-bonded to the nitrogen lone pair of 7HQ in the case of the N* and A* species. We note that a broad-band absorption is also present for the N* species in DMSO-$d_6$, where the hydrogen bond between the nitrogen lone pair and a solvent molecule is absent. The multiphoton excitation effects have been further characterized by doing power-dependent studies and by doing experiments with different pump pulse durations while keeping the pump pulse energies the same. These additional test experiments were done on both solutions and pure solvents. It appeared that the broad featureless transient absorption was predominantly caused by multiphoton excitation effects in the solvent, in line with experience gathered with many femtosecond UV/IR pump−probe experiments previously performed. To identify marker modes for the different electronic excited-state species, we now compare the results obtained for the different solutions.

Figure 4a shows the transient IR spectra obtained by optical excitation of N → N* to the $S_1$ state of N* in $CD_3OD$. At early

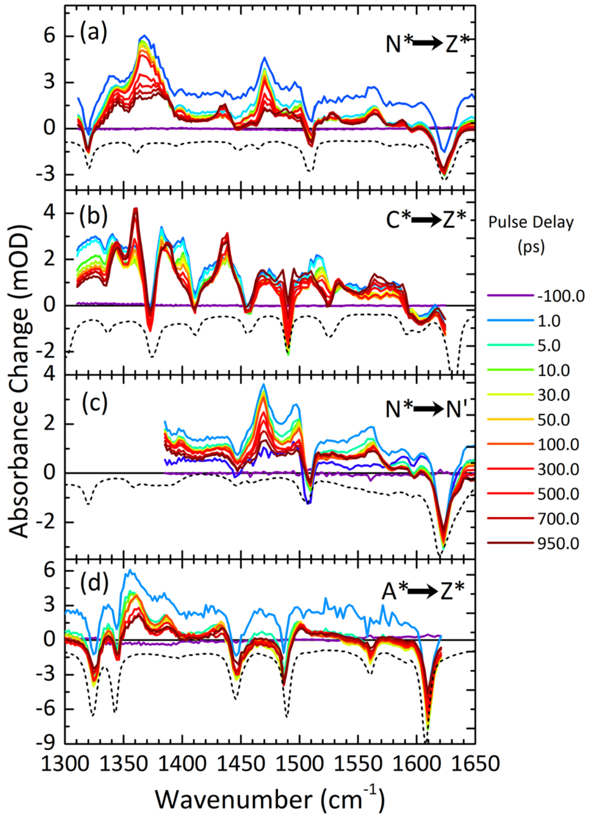


**Figure 4.** Transient IR spectra of 7HQ recorded after electronic excitation of (a) N in methanol-$d_4$, (b) C in methanol-$d_4$, (c) N in DMSO-$d_6$, and (d) A in methanol-$d_4$ as a function of pulse delay between the UV pump and IR probe pulses. Inverted steady-state FT-IR spectra, indicating the positions of ground-state bleach signals, are shown as black dashed lines.

picosecond delay times, we observe IR-active fingerprint transitions of N* at 1341, 1366, 1398, 1470, 1501, and 1564 cm$^{-1}$. These transient marker mode transitions decay on a time scale of hundreds of picoseconds but do not diminish to zero magnitude. For comparison, in Figure 4c, we show the transient IR spectra of 7HQ in DMSO-$d_6$ solution (with a small amount of $D_2O$ added to ascertain full H/D exchange at the hydroxyl group of 7HQ). These small amounts of $D_2O$ did not affect the overall photophysics of 7HQ in DMSO-$d_6$, as evidenced by a comparison of the fluorescence emission spectra of 7HQ recorded in neat DMSO and with small amounts of $D_2O$ added. The measurements show IR-active transitions of the N* species at N* 1397, 1469, 1497, and 1561 cm$^{-1}$ (we have only monitored the spectral region at >1385 cm$^{-1}$). Realizing that the maximum scanning range of our delay stage amounts to 1 ns and the fact that N* cannot act as a photoacid in DMSO-$d_6$, the observed decay (with a time constant of 1.5 ± 0.3 ns; see Figure 5c) should reflect a different relaxation channel. Because the decay of fingerprint bands marking the N* tautomer is not accompanied by a recovery of bleach signals indicative of the electronic ground state of N, we tentatively ascribe the decay of N* to be due to electronically excited-state level crossing to a (presumably) dark state by internal conversion or intersystem crossing to a triplet state.

In addition to the decay dynamics of marker transitions of the N* tautomer of 7HQ in $CD_3OD$, an increase of IR absorption occurs at particular frequency positions anticipated to be a signature for the Z* species at (1345), 1362, 1384, 1440, (1470) 1489, 1529, (1564), and 1588 cm$^{-1}$. To determine whether this assignment is correct, we determine the temporal characteristics of these marker mode transitions. Figure 5a shows kinetic traces measured at 1474 and 1502 cm$^{-1}$, showing the decay kinetics of the reactant N*, as well as the rise of a product species measured at 1440 cm$^{-1}$. Both the decay of N* and the rise of the product species can be fitted with a single-exponential function with a time constant of 320 ± 30 ps. Comparing this result with previously reported PT rates for N* → Z* in methanol to be 170 ps,[17−20,22] we deduce a kinetic isotope effect of 1.9 ± 0.2 for H/D exchange. This finding strongly supports the notion that the observed fingerprint bands of the reaction product can be assigned to the Z* tautomer.

Noting that the N* → Z* deuteron transfer reaction dynamics does not imply a full 100% quantum yield but a value rather close to 30−40%, as the fluorescence emission spectra (Figure 3a) imply, one should take into account that both the N* and Z* species can have IR-active transitions of their respective marker modes at similar detection frequencies, as will become apparent from the vibrational analysis below. To have alternative access to the identification of marker mode transitions of the Z* tautomer, we have performed transient IR spectroscopic measurements upon electronic excitation of the cationic C species (prepared by adding equimolar amounts of $D_2SO_4$ to a solution of 7HQ in $CD_3OD$). Figure 4b shows the measured transient IR spectra, where we observe, besides bleach signals at frequency positions indicative of the IR-active transitions of C, electronically excited state bands of C* at 1326, 1341, 1359, 1382, 1401, 1436 (broad), 1474, 1515, 1546, and 1580 cm$^{-1}$. These IR-active marker transitions decay with a time constant of 160 ± 5 ps (using a single-exponential decay function; see Figure 5b). Concomittant with the decay of marker transitions of the C* species, we detect a single-exponential rise (with the same time constant) of marker band transitions at 1344, 1361, 1388, 1439, 1467, 1486, 1510, 1533, 1564, (1579), and 1588 cm$^{-1}$. These values are consistent with the picture that electronic excitation of C induces a deuteron transfer reaction C* → Z*.

To complete the picture, we have also studied the transient response upon electronic excitation of A in $CD_3OD$ (prepared by adding small amounts of NaOH to a solution of 7HQ in $CD_3OD$). Figure 4d shows the transient IR spectra, where we observe, besides pronounced bleach signals of A in the electronic ground state, IR-active transitions marking A* at 1357 and 1386 cm$^{-1}$. These bands exhibit a frequency upshift of 4−6 cm$^{-1}$ within the first 10 ps that we ascribe to vibrational cooling effects in the $S_1$ state of A*. We find additional marker bands of A* at 1435 and 1501 cm$^{-1}$. Figure 5d shows that the decay of A* is clearly much slower; single-exponential fitting results in a time constant of 600 ± 50 ps. Knowing that the fluorescence spectrum of A* indicates that A* converts to a large extent into the Z* tautomer on the time scale of the fluorescence lifetime (Figure 3b), we have looked for transient IR signals indicative of the formation of the Z* tautomer. Only at 1530 cm$^{-1}$ is a rise of transient IR absorbance apparent (Figure 5d), which exhibits a rise kinetics with a similar time constant as those of the decaying band of A* at 1504 cm$^{-1}$. A more conclusive assignment of the transient IR signals of A* can only be made when the A species of 7HQ is investigated for pulse delay times clearly longer than 1 ns.

DOI: 10.1021/acs.jpca.6b07843
J. Phys. Chem. A 2016, 120, 9378−9389

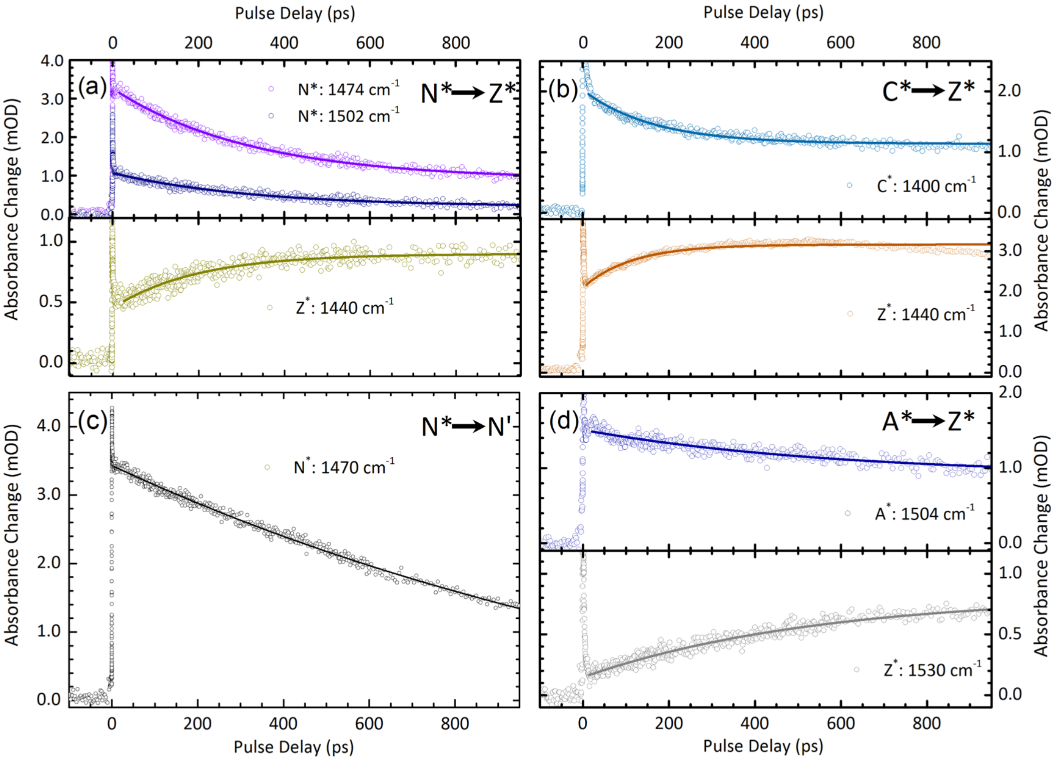


**Figure 5.** (a) Transient kinetics of the N* → Z* conversion following marker bands for N* and Z*. (b) Transient kinetics of the C* → Z* conversion following marker bands for C* and Z*. (c) Transient population kinetics of the N* tautomer (after excitation of N in DMSO-$d_6$). (d) Transient kinetics of the A* → Z* conversion following marker bands for A* and Z*.

**3.3. ESPT Pathway.** In order to get some insights about the PT pathway, we estimated reaction rate constants with harmonic transition state theory.[61] To this end, we calculated activation energies ($\Delta E_a$) and normal-mode frequencies for the tautomerization reaction of N* to Z* along a solvent wire configuration (see Figure 2). This was inspired by earlier studies on ultrafast PT in aqueous solutions suggesting that ultrafast PT can be facilitated by these configurations.[13,41,48,62]

Two plausible pathways exist for the tautomerization reaction of N* to Z*. The first starts with deprotonation of the hydroxy moiety and can formally be characterized as a PT mechanism, whereas the second pathway is initiated by protonation at the nitrogen site of 7HQ, resulting in a methoxide ion transport (MT). We note that the coexistence of both mechanisms in neutral solutions was proposed earlier theoretically[62,63] and experimentally.[21,28] Due to the predefined solvent "wire" configuration, both mechanism proceed concerted and therefore provide an ultrafast reaction channel for the tautomerization of N*.

Table 2 presents estimated reaction time constants τ for PT and MT along with respective reaction and zero-point energy-corrected activation energies. Because the lower activation energy for PT results in an about 7.7 times lower estimated reaction time, the predominant mechanism is likely of PT nature. Moreover, because the calculated reaction times underestimate the experimentally observed rising times significantly, an accurate determination of time constants should consider larger solvent clusters or even bulk conditions as well as solvent fluctuations at finite temperatures. This is ongoing work in our group and lies beyond the scope of the current study. However, likely both concerted and stepwise mechanisms coexist in bulk solution, as suggested for aqueous solutions.[33]

**Table 2. Zero-Point Energy-Corrected Activation and Reaction Energies (kcal mol$^{-1}$) of the Reaction N* → Z* for PT and MT[a]**

| $\Delta E_r$(kcal/mol) | $\Delta E_a$ (kcal/mol) | | τ (ps) | |
|---|---|---|---|---|
| | PT | MT | PT | MT |
| −15.8 | 1.4 | 2.6 | 9.2 | 70.5 |

[a] Reaction time constants τ (ps) were estimated based on harmonic transition state theory.[61]

**3.4. Vibrational Analysis.** In order to further corroborate the assignments of absorption maxima to marker modes for the various 7HQ forms, we conducted normal-mode analysis of the steady-state and transient IR spectra.

A comparison of the normal-mode frequencies with the experimental steady-state spectrum for N within 1200−1700 cm$^{-1}$ is presented in Figure 6. Overall good agreement between experimental absorptions and calculated normal-mode frequencies is achieved with root-mean-square errors (RMSEs) of 8.3 and 7.2 cm$^{-1}$ for perturbation theory corrected and scaled frequencies, respectively. In the observed frequency window, IR-active transitions mainly originate from in-plane bending vibrations, which are delocalized over the entire chromophore molecule. For lower frequencies, elongation amplitudes are mainly on hydrogen atoms, whereas for increasing frequencies, contributions from the carbon backbone become more pronounced. For resonances at around 1600 cm$^{-1}$, vibrations in addition start to become more localized in either of the aromatic rings. Moreover, experimental absorptions at 1216,

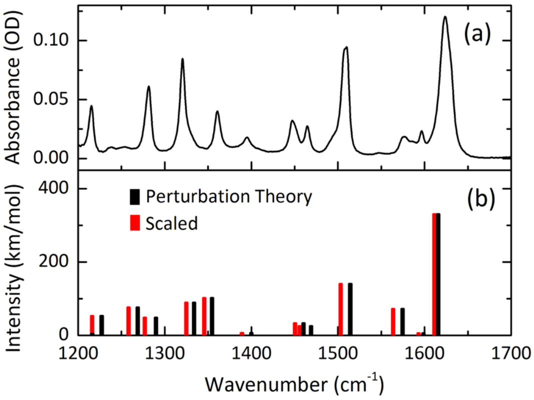


**Figure 6.** Experimental steady-state FT-IR spectrum (a) and calculated normal modes (b) of N in the fingerprint region between 1200 and 1700 cm$^{-1}$. To account for anharmonicities, normal modes have been corrected either by perturbation theory calculations or by a global scaling factor of 0.9763.

1282, 1360, 1465, 1508, and 1576 cm$^{-1}$ involve bending vibrations between $C_8$–O–D atoms. Interestingly, however, the corresponding normal-mode frequencies do not exhibit larger deviations from the experimental absorptions compared to other modes. Therefore, major anharmonicities due to strong hydrogen bonding to the solvent are of minor magnitude.

The vibrational analysis additionally confirms that the recorded fingerprint region 1200−1700 cm$^{-1}$ is well separated from solvent contributions as displacements on methanol atoms are negligibly small. However, already below 1140 cm$^{-1}$, modes start to exhibit significant contributions from solvent atoms. On the other hand, modes at larger wavenumbers than 1700 cm$^{-1}$ are characterized by C–H stretch vibrations of methanol. At wavenumbers even above 2000 cm$^{-1}$, normal-mode vibrations exhibit strong coupling of solvent and chromophore degrees of freedom. This is especially true for O–D stretch vibrations of 7HQ, which are found between 2000 and 2200 cm$^{-1}$. Hence, it is likely that anharmonicities due to solvent hydrogen bonding play a more important role in this frequency range.

Having demonstrated the good agreement of normal modes with the experimental absorptions in the electronic ground state, we continue with the vibrational analysis of the transient IR spectra. Figure 7 presents the normal-mode assignments in the first electronically excited state to experimental absorbance changes for the various 7HQ forms at 30 and 500 ps time delay. We find that experimental absorption maxima correlate well with calculated frequencies with a RMSE for N* of 14.1 cm$^{-1}$. We note that the somewhat larger RMSE compared to the ground state can mostly account for the lower resolution of the transient spectra and additional spectral overlap effects, such as, for example, ground-state bleach and multiphoton excitation components. Especially in the case of the anion, the assignment

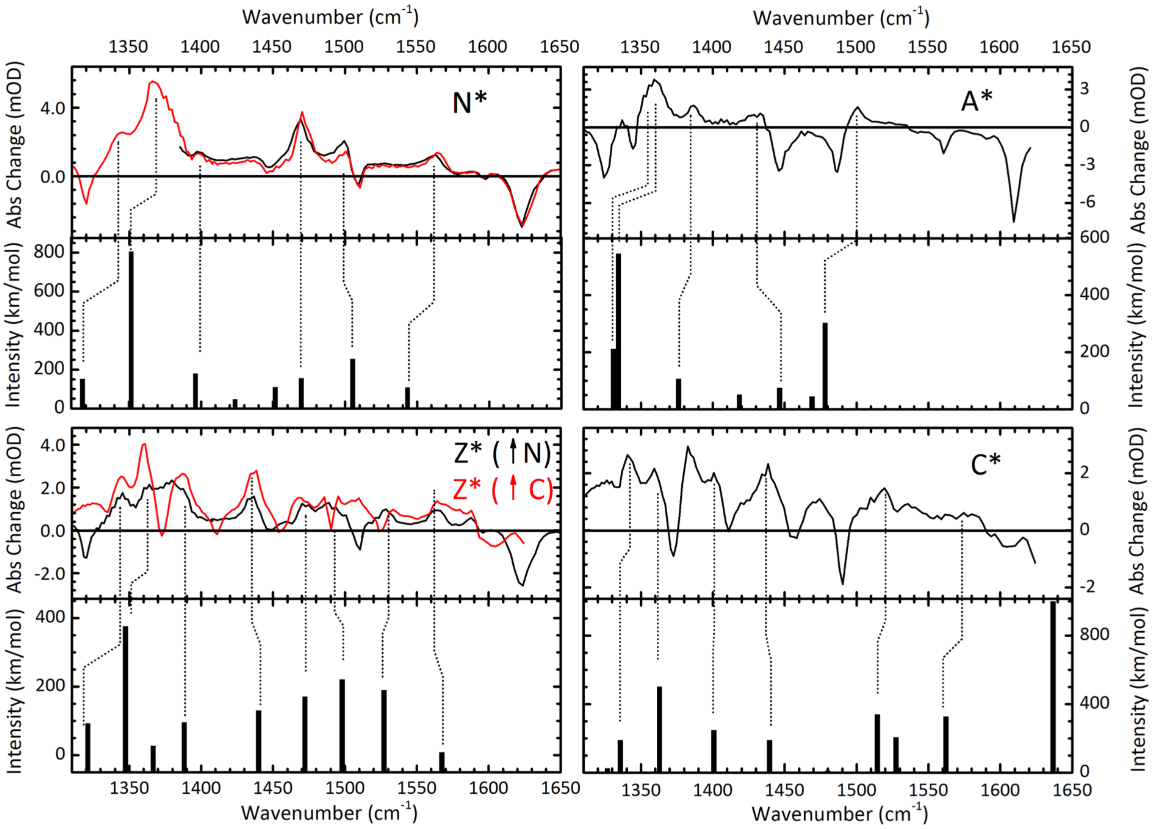


**Figure 7.** Normal-mode assignments to marker modes of N*, Z*, C*, and A* in the transient IR spectra. Normal-mode absorptions are represented as sticks in the lower half of the spectra and scaled by 0.9763 to account for anharmonic effects. Experimental spectra for N*, C*, and A* were recorded at 30 ps delay in neutral, acidic, and basic methanol solution, respectively. The spectrum of Z* was recorded at a 500 ps time delay in neutral (black line) and acidic solution (red line). In addition, the spectrum of N* has been recorded in DMSO-$d_6$ (red line).

to normal modes is hampered by the substantial ground-state bleach signals. Furthermore, it is known that TDDFT is less accurate in calculating vibrational frequencies for electronically excited states, as DFT is for the electronic ground state, where RMSEs are about 50% larger with respect to the ground state.[57]

Within the recorded range of 1310−1650 $cm^{-1}$ are seven (A*) to nine normal modes (Z* and C*), like in the ground state, all of which are associated with in-plane ring deformations, where vibrations are typically delocalized over the entire chromophore molecule (see the Supporting Information for details). Moreover, elongations on hydrogen atoms are dominant in the lower part of the observed range, whereas displacements of the carbon backbone become increasingly dominant in the higher-frequency part close to 1600 $cm^{-1}$. In the case of N*, bending vibrations of $C_8$−O−D only have contributions for normal modes in the lower region of the observed frequency window, namely, at 1317 and 1351 $cm^{-1}$.

Most modes of the various 7HQ forms exhibit large similarities among each other regarding their displacement vectors. For instance, the normal mode of N* at 1424 $cm^{-1}$ is related to modes close to 1440 $cm^{-1}$ in the spectra of Z*, C*, and A*. This mode is therefore assigned to the experimental absorption maxima at 1440, 1436, and 1435 $cm^{-1}$ in Z*, C*, and A*, respectively. Likewise, the normal modes assigned to the experimental absorption band at 1470 $cm^{-1}$ in N* and Z* feature similar displacement vectors.

## 4. DISCUSSION

**4.1. Deuteron Transfer Equilibria and Kinetics of 7HQ.** Figure 1 shows the general reaction scheme of 7HQ. Only because of different $pK_a$ values for the N, C, A, and Z forms of 7HQ in the $S_0$ and $S_1$ states, PT kinetics occurs, as is expected for photoacid/photobase systems. Now when starting with an initially prepared transient population of N* (using a UV pump pulse exciting N), one can envisage three different routes for PT: (I) a concerted N* → Z* pathway directly transporting a deuteron from the hydroxyl group to the aromatic ring nitrogen group via a methanol solvent bridge; (II) a sequential N* → C* → Z* deuteron transfer pathway consisting of deuteron pick-up by the aromatic ring nitrogen group from the solvent followed by a later deuteron transfer from the hydroxyl group to the solvent; and (III) a sequential N* → A* → Z* deuteron transfer pathway consisting of deuteron transfer from the hydroxyl group to the solvent followed by a later deuteron pick-up by the aromatic ring nitrogen group from the solvent. Kinetically, these three different pathways may be hard to distinguish when having only access to the population of the reactant N* and product Z* species. This remains to be the case even when marker modes for either the C* or A* species are spectrally overlapping with those of the N* or Z* tautomers or when respective time constants for the different steps prevent a significant transient population of the C* or A* charged forms.

The observed time constant for the C* → Z* deuteron dissociation time is 170 ps in deuterated methanol, whereas the A* → Z* deuteron pick-up time from a nearby solvent molecule is 600 ps. The time constant for the overall N* → Z* deuteron dissociation/abstraction lies in between: 330 ps (Figure 5). From the observed kinetic isotope effect for the latter reaction of 1.9, one can conclude a sequential process to be the dominating mechanism as larger values have been reported for tunneling PT pathways[74,75] (as one would expect to be the case for concerted pathway I). Noting that pathway III implies a rather slow second deuteron solvent abstraction step, incompatible with the observed overall time constant for the N* → Z* tautomerization in deuterated methanol, we argue that this route can be part of only the N* → Z* reaction pathway when dynamics of the second A* → Z* step becomes significantly faster when a deuteron is already present along the methanol bridge. That this can indeed be the case is suggested by the calculated reaction time constants presented in Table 2. Here, the data indicate that a concerted reaction proceeds mainly according to the PT mechanism, that is, A* is formed in a first step followed by deuteron pick-up. However, due to the cluster model, it remains elusive which mechanism is present for a stepwise reaction path. It follows therefore that a separate measurement of the N* → A* deuteron dissociation time and a more realistic modeling of solvent effects is necessary to unequivocally determine whether the sequential N* → A* → Z* reaction pathway is a key pathway for the N* → Z* tautomerization. On the other hand, the observed time constants for C* → Z* deuteron dissociation to the solvent and N* → Z* tautomerization suggest that for pathway II only a minor transient population for the C* of at most 26% of the initially excited population of N* is generated. This estimate comes from a sequential population kinetics scheme with 100% quantum yields for the individual reaction steps. In reality, however, these numbers will be even lower, taking into account that the N* → Z* tautomerization is in fact an equilibration reaction leading to transient populations of similar magnitudes for N* and Z*. Our transient spectra shown in Figure 4 do not reveal a significantly large signal for the C* species at frequency positions of its respective marker mode transitions. However, the fact that for the first step in pathway II, that is, N* → C*, the first deuteron abstraction from the solvent is slower than the deuteron dissociation to the solvent, that is, the second step C* → Z*, is consistent with reported values for PT by photoacids to the solvent, vs PT by photobases from the solvent, in aqueous solutions.[23,76] These trends appear to hold also for methanol solutions. Realizing that $pK_a$ values typically change 4−5 units when going from methanol to water, it appears to be a worthwhile endeavor to explore the reaction dynamics of 7HQ in aqueous solution using transient IR spectroscopy as the observed kinetics will speed up by at least 1 order of magnitude.

**4.2. Normal-Mode Analysis.** On the basis of the cluster model, we assigned experimental marker modes of N, C, A, and Z in the electronic ground and excited states to normal modes. By this, we relate the contributions of distinct molecular vibrations to the marker modes so as to facilitate a more detailed understanding of the intricate photophysics of 7HQ. Further, the comparison of calculated frequencies with experimental absorptions provides an additional assessment for the adequacy of the cluster model to describe certain aspects of the processes following electronic excitation and to delineate in part the effect of hydrogen-bonding on the solvent.

Overall good agreement between steady-state experimental and calculated normal-mode spectra in the electronic ground state (see Figure 6) is suggested by RMSEs of about 7−8 $cm^{-1}$, which is comparable with other studies on photoacids[64−66] and organic molecules.[67] This finding in turn suggests that pronounced anharmonic effects are of minor magnitude within the recorded frequency window of 1200−1700 $cm^{-1}$. Notably, mode-mixing effects with the $\delta_{OD}$ or $\delta_{ND}$ bending and solvent degrees of freedom as well as associated anharmonic effects due to hydrogen bonding remain of minor magnitudes. Therefore, the absence of solvent coupling as indicated by the normal-mode coordinates suggests that the chosen cluster model is indeed adequate to describe the fundamental absorptions and to delineate distinct vibrational contributions within this window. Outside of the experimentally recorded frequency range, however, normal modes feature a significant intermixing of solvent and chromophore vibrations. Especially modes that involve OD stretching vibrations show a strong coupling suggesting that normal-mode analysis in regions outside of 1200−1700 $cm^{-1}$ might be less adequate and anharmonic and solvent dynamics effects have to be taken into account explicitly, for example, by means of molecular dynamics simulations.[35,68]

In the electronically excited state, assignments exhibit larger RMSEs (14.1 $cm^{-1}$ for N*) as compared to the ground state (7.2 $cm^{-1}$ for N) due to the lower resolution of the measured spectra and additional effects like ground-state bleaching. Moreover, TDDFT is slightly less accurate compared to corresponding ground-state calculations. Liu and co-workers[57] reported based on a set of 24 small organic and inorganic molecules that the RMSE between measured and scaled frequencies increased by about 10−20 $cm^{-1}$ when scaling factors derived from ground-state calculations were applied. Only slight improvements over ground-state scaling factors were reported for those derived from excited-state calculations. Nevertheless, a reasonable correlation of calculated and experimental absorptions is observed in our case. Moreover, an additional guide for resonance assignment is provided by the fact that the frequency range of 1310−1650 $cm^{-1}$ is well separated from higher-energy vibrations like C−D stretch vibrations of methanol or O−D stretch vibrations of 7HQ that occur above 2000 $cm^{-1}$. We note, however, some degree of ambiguity for the assignment of close-lying marker modes to normal-mode vibrations that are separated by less than 20 $cm^{-1}$.

From the assignment of experimental marker bands in the transient spectra to normal-mode vibrations in Figure 7 and by comparing their respective elongation vectors presented in Figure S3, it becomes clear that a significant number of vibrations are structurally related. For instance, the normal mode at 1423.72 $cm^{-1}$ in N* features similar elongation vectors compared to the mode at 1439.48 $cm^{-1}$ in C*. Thus, the occurrence of marker modes can be understood in terms of small, but often observable, frequency shifts with associated changes in IR transition cross sections caused by the distinct deuteration states. It should be noted that modes with no elongation vector on the reactive sites are shifted as well, indicating a rather complicated dependence of frequency shifts and normal-mode positions in the fingerprint region of 7HQ.

### 4.3. Electronic and Structural Effects of Excitation.

Comparison of experimental and calculated electronic absorption and emission energies shown in Table 1 reveals excellent agreement and confirms the assignment of bands to the various deuteration states of 7HQ. From transition dipole moments (see Figure S1), it is clear that transitions can be characterized as $^1L_b$, in accordance with earlier findings in aqueous solution.[23] The role of charge redistributions for the various processes can be investigated by analyzing the electron density and partial charge differences before and after electronic excitation (see the Supporting Information). Our results show that upon excitation electron density is withdrawn from the hydroxy moiety toward the nitrogen-containing ring, where, more precisely, it is decreased at the bond shared by O and $C_8$ and at the O atom itself. On the other hand, density is increased at N, $C_2$, $C_4$, and $C_6$. A similar observation is made from the analysis of partial charge differences ($\Delta\delta$) before and after excitation, computed from Hirshfeld partitioning[69] of the electron density (see Table S1). The largest change in $\Delta\delta$ is observed on atoms $C_2$ and $C_4$, which become more negative, as well as atoms $C_8$ and O, which become more positive after excitation. It is noteworthy that the sum of $|\Delta\delta|$ is largest for the anionic and zwitterionic forms of 7HQ, which is in line with earlier reports that photoacidity is mainly due to redistribution effects of the photobase.[68,70−73]

Table 3 presents changes of the $C_7$−O bond length and heavy atom distances of the 7HQ oxygen and nitrogen atom to the closest methanol oxygen atom, respectively, upon excitation. It can be seen that excitation leads in general to a shortening of the C−O bond, where the largest change is observed for N*. Moreover, for N and C, hydrogen bonds between the solvent and the OD group of 7HQ become shorter, whereas they increase in the case of Z and A. Analogously for N and A, the N−O distance decreases, whereas it lengthens for C and Z, although the change is less pronounced than that at the OD site. This effect is expected as it is a direct consequence of the increased acidity and decreased basisity in N, C and Z, A, correspondingly, upon photoexcitation and a concomitant strengthening/weakening of 7HQ−methanol hydrogen bonds. Psciuk and co-workers[68] obtained similar results for 2-naphthol, where they observed a shortening of the hydrogen bond between the hydroxy moiety and an acetonitrile molecule in a number of nonpolar solvents. Their results show in addition a clear correlation between the shortening of the hydrogen bond in the excited state (around 0.08 Å in their study) and a red shift of the OH stretch frequency of about 100 $cm^{-1}$, providing further evidence for a strengthening of the hydrogen bond.

**Table 3. Distances in Å for the $C_7$−O Bond and the 7HQ Oxygen and Nitrogen to the Closest Methanol Oxygen Atom**[a]

| | $C_7$−O | $O_c \cdots O_m$ | $N_c \cdots O_m$ |
|---|---|---|---|
| N | 1.35 | 2.66 | 2.74 |
| N* | −0.04 | −0.12 | −0.05 |
| C | 1.33 | 2.62 | 2.84 |
| C* | −0.02 | −0.12 | 0.06 |
| Z | 1.28 | 2.62 | 2.77 |
| Z* | −0.01 | 0.08 | 0.02 |
| A | 1.29 | 2.58 | 2.78 |
| A* | −0.01 | 0.11 | −0.04 |

[a] Dots indicate hydrogen bonds between 7HQ and methanol, where the subscript indicates either the atom number or whether the atom belongs to methanol (M) or 7HQ (C). Bond lengths of excited 7HQ forms are with respect to ground-state geometries.

## 5. CONCLUSIONS

We investigated transient IR-active transitions of the various deuteration forms of 7HQ in methanol solution in the first electronically excited state. Depending on the *p*D value, we monitored the formation of the distinct 7HQ forms (N*, C*, A*, and Z*) by following their characteristic fingerprint marker bands. Vibrational analysis reveals good agreement between experimentally measured IR-active transitions and calculated normal-mode frequencies, indicating a minor role of anharmonicities for this spectral region. Further, analysis shows that vibrations in the recorded range are well separated from solvent degrees of freedom, where differences in deuteration states lead to small frequency shifts and, concomitantly, to changes of IR transition cross sections. Estimates of reaction time constants for N* →Z* suggest preference for a deuteron transport mechanism for a concerted reaction, that is, the deuteron is released first, contrasting a situation where the reaction starts with a pick-up of a deuteron from a methanol molecule. Conversely, the experimental time constants suggest that the reaction proceeds largely in a stepwise manner, and experimental rates are in better agreement with hydroxide ion transport. In order to resolve this contradiction, additional measurements of the N* → Z* pathway and a more realistic modeling of solvent effects are necessary to unequivocally determine whether A* is indeed the intermediate in the tautomerization reaction. So far, our measurements revealed higher PT rates for C* → Z* than those for A* → Z*, as evidenced by the time evolution of marker modes in the transient spectra. Our results on partial charge differences for the various forms of 7HQ support the notion that conjugate photoacidity/photobasicity is dominated by charge redistribution effects in the conjugated photobase/photoacid, respectively. In contrast, changes in hydrogen bond lengths are more pronounced in N, C, and A than those in Z, suggesting that the structural relaxation of 7HQ following electronic excitation leads to stronger solvent reorganizations in the case of N, C, and A.

## ■ ASSOCIATED CONTENT

### Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpca.6b07843.

> Figures showing transition dipole moments, partial charges and difference electron densities; details of harmonic transition state theory; and plots of normal-mode coordinates of N*, C*, A*, and Z* (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

*E-mail:nibberin@mbi-berlin.de. Phone: (+49)-(0)30-6392-1477 (E.T.J.N.).

*E-mail: daniel.sebastiani@chemie.uni-halle.de. Phone: (+49)-(0)345-55-25836 (D.S.).

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work has benefitted from financial support by the German Science Foundation (Project Number NI 492/13-1/SE 1008/11-1). F.H. thanks the Fonds der Chemischen Industrie for a Kekulé fellowship.

## ■ REFERENCES

(1) Geissler, P. L.; Dellago, C.; Chandler, D.; Hutter, J.; Parrinello, M. Autoionization in Liquid Water. *Science* **2001**, *291*, 2121−2124.

(2) Marx, D.; Tuckerman, M.; Hutter, J.; Parrinello, M. The Nature of the Hydrated Excess Proton in Water. *Nature* **1999**, *397*, 601−604.

(3) Vuilleumier, R.; Borgis, D. Transport and Spectroscopy of the Hydrated Proton: A Molecular Dynamics Study. *J. Chem. Phys.* **1999**, *111*, 4251−4266.

(4) Marx, D. Proton Transfer 200 Years After Von Grotthuss: Insights From Ab Initio Simulations. *ChemPhysChem* **2007**, *8*, 209−210.

(5) Markovitch, O.; Chen, H.; Izvekov, S.; Paesani, F.; Voth, G. A.; Agmon, N. Special Pair Dance and Partner Selection: Elementary Steps in Proton Transport in Liquid Water. *J. Phys. Chem. B* **2008**, *112*, 9456−9466.

(6) Xu, J.; Zhang, Y.; Voth, G. A. Infrared Spectrum of the Hydrated Proton in Water. *J. Phys. Chem. Lett.* **2011**, *2*, 81−86.

(7) Kulig, W.; Agmon, N. A 'Clusters-In-Liquid' Method for Calculating Infrared Spectra Identifies the Proton-Transfer Mode in Acidic Aqueous Solutions. *Nat. Chem.* **2012**, *5*, 29−35.

(8) Vilciauskas, L.; Tuckerman, M. E.; Bester, G.; Paddison, S. J.; Kreuer, K.-D. The Mechanism of Proton Conduction in Phosphoric Acid. *Nat. Chem.* **2012**, *4*, 461−466.

(9) Kreuer, K. D.; Paddison, S. J.; Spohr, E.; Schuster, M. Transport in Proton Conductors for Fuel-Cell Applications: Simulations, Elementary Reactions, and Phenomenology. *Chem. Rev.* **2004**, *104*, 4637−4678.

(10) Jorn, R.; Savage, J.; Voth, G. A. Proton Conduction in Exchange Membranes Across Multiple Length Scales. *Acc. Chem. Res.* **2012**, *45*, 2002−2010.

(11) Tuckerman, M. E.; Marx, D.; Parrinello, M. The Nature and Transport Mechanism of Hydrated Hydroxide in Aqueous Solution. *Nature* **2002**, *417*, 925−929.

(12) Genosar, L.; Cohen, B.; Huppert, D. Ultrafast Direct Photoacid-Base Reaction. *J. Phys. Chem. A* **2000**, *104*, 6689−6698.

(13) Mohammed, O. F.; Pines, D.; Dreyer, J.; Pines, J.; Nibbering, E. T. J. Sequential Proton Transfer Through Water Bridges in Acid-Base Reactions. *Science* **2005**, *310*, 83−86.

(14) Mohammed, O. F.; Pines, D.; Nibbering, E. T. J.; Pines, E. Base-Induced Solvent Switches in Acid-Base Reactions. *Angew. Chem.* **2007**, *119*, 1480−1483.

(15) Pines, D.; Nibbering, E. T. J.; Pines, E. Monitoring the Microscopic Molecular Mechanisms of Proton Transfer in Acid-Base Reactions in Aqueous Solutions. *Isr. J. Chem.* **2015**, *55*, 1240−1251.

(16) Decoursey, T. E. Voltage-Gated Proton Channels and Other Proton Transfer Pathways. *Physiol. Rev.* **2003**, *83*, 475−579.

(17) Thistlethwaite, P.; Corkill, P. Direct Observation of Photo-automerism Kinetics in 7-Quinolinol by Picosecond Spectroscopy. *Chem. Phys. Lett.* **1982**, *85*, 317−321.

(18) Thistlethwaite, P. J. Solvation Effects in the Phototautomerization of 7-Quinolinol. *Chem. Phys. Lett.* **1983**, *96*, 509−512.

(19) Itoh, M.; Adachi, T.; Tokumura, K. Time-Resolved Fluorescence and Absorption Spectra and Two-Step Laser Excitation Fluorescence of the Excited-State Proton Transfer in the Methanol Solution of 7-Hydroxyquinoline. *J. Am. Chem. Soc.* **1984**, *106*, 850−855.

(20) Konijnenberg, J.; Ekelmans, G. B.; Huizer, A. H.; Varma, C. A. G. Mechanism and Solvent Dependence of the Solvent-Catalysed Pseudo-intraMolecular Proton Transfer of 7-Hydroxyquinoline in the First Electronically Excited Singlet State and in the Ground State of Its Tautomer. *J. Chem. Soc., Faraday Trans. 2* **1989**, *85*, 39−51.

(21) Lee, S.-I.; Jang, D.-J. Proton Transfer of Aqueous 7-Hydroxyquinoline in the First Excited Singlet, Lowest Triplet, and Ground States. *J. Phys. Chem.* **1995**, *99*, 7537−7541.

(22) Nakagawa, T.; Kohtani, S.; Itoh, M. Picosecond Fluorescence and Two-Step LIF Studies of the Excited-State Proton Transfer in Methanol Solutions of 7-Hydroxyquinoline and Methyl-Substituted 7-Hydroxyquinolines. *J. Am. Chem. Soc.* **1995**, *117*, 7952−7957.

(23) Bardez, E. Excited-State Proton Transfer in Bifunctional Compounds. *Isr. J. Chem.* **1999**, *39*, 319−332.

(24) Kwon, O. H.; Kim, T. G.; Lee, Y. S.; Jang, D. J. Biphasic Tautomerization Dynamics of Excited 7-Hydroxyquinoline in Reverse Micelles. *J. Phys. Chem. B* **2006**, *110*, 11997−12004.

(25) Kwon, O.-H.; Lee, Y.-S.; Yoo, B. K.; Jang, D.-J. Excited-State Triple Proton Transfer of 7-Hydroxyquinoline Along a Hydrogen-Bonded Alcohol Chain: Vibrationally Assisted Proton Tunneling. *Angew. Chem., Int. Ed.* **2006**, *45*, 415−419.

(26) Bhattacharya, B.; Samanta, A. Excited-State Proton-Transfer Dynamics of 7-Hydroxyquinoline in Room Temperature Ionic Liquids. *J. Phys. Chem. B* **2008**, *112*, 10101−10106.

(27) Park, Y. S.; Kim, B.; Lee, Y. S.; Kwon, O. H.; Jang, D.-J. Triple Proton Transfer of Excited 7-Hydroxyquinoline Along a Hydrogen-Bonded Water Chain in Ethers: Secondary Solvent Effect on the Reaction Rate. *Photochem. Photobiol. Sci.* **2009**, *8*, 1611−1617.

(28) Presiado, I.; Erez, Y.; Gepshtein, R.; Huppert, D. Excited-State Proton Transfer and Proton Reactions of 6-Hydroxyquinoline and 7-Hydroxyquinoline in Water and Ice. *J. Phys. Chem. C* **2009**, *113*, 20066−20075.

(29) Park, S.-Y.; Lee, Y.-S.; Kwon, O.-H.; Jang, D.-J. Proton Transport of Water in Acid-Base Reactions of 7-Hydroxyquinoline. *Chem. Commun.* **2009**, 926−928.

(30) Park, S.-Y.; Jang, D.-J. Accumulated Proton-Donating Ability of Solvent Molecules in Proton Transfer. *J. Am. Chem. Soc.* **2010**, *132*, 297−302.

(31) Al-Lawatia, N.; Husband, J.; Steinbrecher, T.; Abou-Zied, O. K. Tautomerism in 7-Hydroxyquinoline: A Combined Experimental and Theoretical Study in Water. *J. Phys. Chem. A* **2011**, *115*, 4195−4201.

(32) Kang, B.; Ko, K. C.; Park, S.-Y.; Jang, D.-J.; Lee, J. Y. Solvent Effect on the Excited-State Proton Transfer of 7-Hydroxyquinoline Along a Hydrogen-Bonded Ethanol Dimer. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6332−6339.

(33) Kwon, O.-H.; Mohammed, O. F. Water-Wire Catalysis in Photoinduced Acid-Base Reactions. *Phys. Chem. Chem. Phys.* **2012**, *14*, 8974−8980.

(34) Lim, H.; Jeong, H.; Park, S.-Y.; Lee, J. Y.; Jang, D.-J. Excited-State Proton-Relay Dynamics of 7-Hydroxyquinoline Controlled by Solvent Reorganization in Room Temperature Ionic Liquids. *Phys. Chem. Chem. Phys.* **2012**, *14*, 218−224.

(35) Bekçioğlu, G.; Hoffmann, F.; Sebastiani, D. Solvation-Dependent Latency of Photoacid Dissociation and Transient IR Signatures of Protonation Dynamics. *J. Phys. Chem. A* **2015**, *119*, 9244−9251.

(36) Ditkovich, J.; Mukra, T.; Pines, D.; Huppert, D.; Pines, E. Bifunctional Photoacids: Remote Protonation Affecting Chemical Reactivity. *J. Phys. Chem. B* **2015**, *119*, 2690−2701.

(37) Mohammed, O. F.; Pines, D.; Nibbering, E. T. J.; Pines, E. Base-Induced Solvent Switches in Acid-Base Reactions. *Angew. Chem.* **2007**, *119*, 1480−1483.

(38) Mohammed, O. F.; Pines, D.; Pines, E.; Nibbering, E. T. J. Aqueous Bimolecular Proton Transfer in Acid-Base Neutralization. *Chem. Phys.* **2007**, *341*, 240−257.

(39) Hynes, J. T. Physical Chemistry: The Peripatetic Proton. *Nature* **2007**, *446*, 270−273.

(40) Rini, M.; Magnes, B.-Z.; Pines, E.; Nibbering, E. T. J. Real-Time Observation of Bimodal Proton Transfer in Acid-Base Pairs in Water. *Science* **2003**, *301*, 349−352.

(41) Cox, M. J.; Timmer, R. L. A.; Bakker, H. J.; Park, S.; Agmon, N. Distance-Dependent Proton Transfer Along Water Wires Connecting Acid-Base Pairs. *J. Phys. Chem. A* **2009**, *113*, 6599−6606.

(42) Kiefer, P. M.; Hynes, J. T. Theoretical Aspects of Tunneling Proton Transfer Reactions in a Polar Environment. *J. Phys. Org. Chem.* **2010**, *23*, 632−646.

(43) Mason, S. F. The Tautomerism of N-Heteroaromatic Hydroxy-Compounds. Part III. Ionisation Constants. *J. Chem. Soc.* **1958**, 674−685.

(44) Mason, S. F.; Philp, J.; Smith, B. E. Prototropic Equilibria of Electronically Excited Molecules. Part II. 3-, 6-, and 7-Hydroxyquinoline. *J. Chem. Soc. A* **1968**, 3051−3056.

(45) Park, S.-Y.; Jang, D.-J. Accumulated Proton-Donating Ability of Solvent Molecules in Proton Transfer. *J. Am. Chem. Soc.* **2010**, *132*, 297−302.

(46) Tanner, C.; Manca, C.; Leutwyler, S. Probing the Threshold to H Atom Transfer Along a Hydrogen-Bonded Ammonia Wire. *Science* **2003**, *302*, 1736−1739.

(47) Fernandez-Ramos, A.; Martinez-Nunez, E.; Vazquez, S. A.; Rios, M. A.; Estevez, C. M.; Merchan, M.; Serrano-Andres, L. Hydrogen Transfer vs Proton Transfer in 7-Hydroxy-Quinoline·$(NH_3)_3$: A CASSCF/CASPT2 Study. *J. Phys. Chem. A* **2007**, *111*, 5907−5912.

(48) Bekçioğlu, G. B.; Allolio, C.; Sebastiani, D. Water Wires in Aqueous Solutions From First-Principles Calculations. *J. Phys. Chem. B* **2015**, *119*, 4053−4060.

(49) Matsumoto, Y.; Ebata, T.; Mikami, N. Structure and Photoinduced Excited State Keto-Enol Tautomerization of 7-Hydroxyquinoline-(CH3OH) N Clusters. *J. Phys. Chem. A* **2002**, *106*, 5591−5599.

(50) Thut, M.; Tanner, C.; Steinlin, A.; Leutwyler, S. Time-Dependent Density Functional Theory as a Tool for Isomer Assignments of Hydrogen-Bonded Solute· Solvent Clusters. *J. Phys. Chem. A* **2008**, *112*, 5566−5572.

(51) Adamczyk, K.; Prémont-Schwarz, M.; Pines, D.; Pines, E.; Nibbering, E. T. J. Real-Time Observation of Carbonic Acid Formation in Aqueous Solution. *Science* **2009**, *326*, 1690−1694.

(52) Frisch, M.; Trucks, G.; Schlegel, H. B.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G.; et al. *Gaussian 09*, revision A.02; Gaussian Inc.: Wallingford, CT, 2009.

(53) Runge, E.; Gross, E. K. U. Density-Functional Theory for Time-Dependent Systems. *Phys. Rev. Lett.* **1984**, *52*, 997−1000.

(54) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(55) Grimme, S. E.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parameterization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

(56) Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **1998**, *102*, 1995−2001.

(57) Liu, J.; Liang, W. Analytical Approach For the Excited-State Hessian in Time-Dependent Density Functional Theory: Formalism, Implementation, and Performance. *J. Chem. Phys.* **2011**, *135*, 184111.

(58) Merrick, J. P.; Moran, D.; Radom, L. An Evaluation of Harmonic Vibrational Frequency Scale Factors. *J. Phys. Chem. A* **2007**, *111*, 11683−11700.

(59) Ullrich, C. A. *Time-Dependent Density-Functional Theory: Concepts and Applications*; Oxford University Press: New York, 2011.

(60) Jacquemin, D.; Wathelet, V.; Perpéte, E. A.; Adamo, C. Extensive TD-DFT Benchmark: Singlet-Excited States of Organic Molecules. *J. Chem. Theory Comput.* **2009**, *5*, 2420−2435.

(61) Vineyard, G. H. Frequency Factors and Isotope Effects in Solid State Rate Processes. *J. Phys. Chem. Solids* **1957**, *3*, 121−127.

(62) Bekçioğlu, G.; Allolio, C.; Ekimova, M.; Nibbering, E. T. J.; Sebastiani, D. Competition Between Excited State Proton and OH-Transport via a Short Water Wire: Solvent Effects Open the Gate. *Phys. Chem. Chem. Phys.* **2014**, *16*, 13047−13051.

(63) Cui, Y.; Zhao, H.; Zhao, J.; Li, P.; Song, P.; Xia, L. The Excited-State Multiple Proton Transfer Mechanism of the 7-Hydroxyquinoline-$(CH_3OH)_3$ Cluster. *New J. Chem.* **2015**, *39*, 9910−9917.

(64) de Vivie-Riedle, R.; De Waele, V.; Kurtz, L.; Riedle, E. Ultrafast Excited-State Proton Transfer of 2-(2′-Hydroxyphenyl)-Benzothiazole: Theoretical Analysis of the Skeletal Deformations and the Active Vibrational Modes. *J. Phys. Chem. A* **2003**, *107*, 10591−10599.

(65) Mohammed, O. F.; Dreyer, J.; Magnes, B.-Z.; Pines, E.; Nibbering, E. T. Solvent-Dependent Photoacidity State of Pyranine Monitored by Transient Mid-Infrared Spectroscopy. *ChemPhysChem* **2005**, *6*, 625−636.

(66) Luber, S.; Adamczyk, K.; Nibbering, E. T. J.; Batista, V. S. Photoinduced Proton Coupled Electron Transfer in 2-(2′-Hydroxyphenyl)-Benzothiazole. *J. Phys. Chem. A* **2013**, *117*, 5269−5279.

(67) Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*; Wiley-VCH: Weinheim, Germany, 2015.

(68) Psciuk, B. T.; Prémont-Schwarz, M.; Koeppe, B.; Keinan, S.; Xiao, D.; Nibbering, E. T.; Batista, V. S. Correlating Photoacidity to Hydrogen-Bond Structure by Using the Local O-H Stretching Probe in Hydrogen-Bonded Complexes of Aromatic Alcohols. *J. Phys. Chem. A* **2015**, *119*, 4800−4812.

(69) Hirshfeld, F. L. Bonded-Atom Fragments for Describing Molecular Charge Densities. *Theor. Chem. Acc.* **1977**, *44*, 129−138.

(70) Pines, D.; Pines, E. *Solvent Assisted Photoacidity*; Wiley-VCH: Weinheim, Germany, 2007; Vol. *1*; p 377.

(71) Agmon, N.; Rettig, W.; Groth, C. Electronic Determinants of Photoacidity in Cyanonaphthols. *J. Am. Chem. Soc.* **2002**, *124*, 1089−1096.

(72) Hynes, J. T.; Tran-Thi, T.-H.; Granucci, G. Intermolecular Photochemical Proton Transfer in Solution: New Insights and Perspectives. *J. Photochem. Photobiol., A* **2002**, *154*, 3−11.

(73) Granucci, G.; Hynes, J. T.; Millie, P.; Tran-Thi, T.-H. A Theoretical Investigation of Excited-State Acidity of Phenol and Cyanophenols. *J. Am. Chem. Soc.* **2000**, *122*, 12243−12253.

(74) Pines, E. In *Isotope Effects in Chemistry and Biology*; Kohen, A., Limbach, H.-H., Eds.; CRC Press: Boca Raton, FL, 2006.

(75) Salna, B.; Benabbas, A.; Sage, J. T.; van Thor, J.; Champion, P. M. Wide-Dynamic-Range Kinetic Investigations of Deep Proton Tunnelling in Proteins. *Nat. Chem.* **2016**, *8*, 874−880.

(76) Munitz, N.; Avital, Y.; Pines, D.; Nibbering, E. T.; Pines, E. Cation-Enhanced Deprotonation of Water by a Strong Photobase. *Isr. J. Chem.* **2009**, *49*, 261−272.

# Improved Quantum Chemical NMR Chemical Shift Prediction of Metabolites in Aqueous Solution toward the Validation of Unknowns

Felix Hoffmann, Da-Wei Li, Daniel Sebastiani, and Rafael Brüschweiler

Reprinted with permission from

J. Phys. Chem. A **2017**, 121, 3071–3078.

Copyright 2017 American Chemical Society.

`http://dx.doi.org/10.1021/acs.jpca.7b01954`

THE JOURNAL OF
PHYSICAL CHEMISTRY A

Article

pubs.acs.org/JPCA

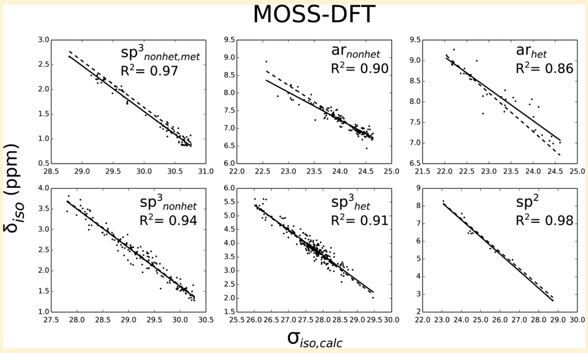
# Improved Quantum Chemical NMR Chemical Shift Prediction of Metabolites in Aqueous Solution toward the Validation of Unknowns

Felix Hoffmann,[†] Da-Wei Li,[‡] Daniel Sebastiani,[†] and Rafael Brüschweiler*[,‡,§,∥]

[†]Institute of Chemistry, Martin-Luther-University Halle-Wittenberg, von-Danckelmann-Platz 4, 06120 Halle, Germany
[‡]Campus Chemical Instrument Center, The Ohio State University, Columbus, Ohio 43210, United States
[§]Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210, United States
[∥]Department of Biological Chemistry and Pharmacology, The Ohio State University, Columbus, Ohio 43210, United States

*Supporting Information*

**ABSTRACT:** A quantum-chemistry based protocol, termed MOSS-DFT, is presented for the prediction of $^{13}$C and $^{1}$H NMR chemical shifts of a wide range of organic molecules in aqueous solution, including metabolites. Molecular motif-specific linear scaling parameters are reported for five different density functional theory (DFT) methods (B97-2/pcS-1, B97-2/pcS-2, B97-2/pcS-3, B3LYP/pcS-2, and BLYP/pcS-2), which were applied to a large set of 176 metabolite molecules. The chemical shift root-mean-square deviations (RMSD) for the best method, B97-2/pcS-3, are 1.93 and 0.154 ppm for $^{13}$C and $^{1}$H chemical shifts, respectively. Excellent results have been obtained for chemical shifts of methyl and aromatic $^{13}$C and $^{1}$H that are not directly bonded to a heteroatom (O, N, S, or P) with RMSD values of 1.15/0.079 and 1.31/0.118 ppm, respectively. This study not only demonstrates how NMR chemical shift in aqueous environment can be improved over the commonly used global linear scaling approach, but also allows for motif-specific error estimates, which are useful for an improved chemical shift-based verification of metabolite candidates of metabolomics samples containing unknown components.



## 1. INTRODUCTION

The analysis of complex metabolic mixtures of a wide range of biological systems using NMR spectroscopy has become increasingly popular over the past decade in the context of the rapidly growing field of metabolomics.[1−4] This is in part due to the ability to simultaneously provide detailed spectroscopic information on many different metabolites in the same sample. In that way, new insights into the state of biological systems as well as into metabolic pathways are possible.[5−8]

The concomitant development of databases and web-based query tools, such as the Complex Mixture Analysis by NMR (COLMAR) database,[9] The Human Metabolome Database (HMDB)[10] and the Biological Magnetic Resonance Data Bank (BMRB),[11] has further increased the usefulness of NMR spectroscopy and enabled the automated assignment of metabolites. However, the identification of unknown metabolites, which are metabolites that give rise to signals in the NMR spectra but that have not been identified previously or are not part of commonly accessible NMR databases, remains a key challenge. In this regard, empirical chemical shift predictors, such as NMRPredict[12] as used by the MNova software,[13] or the ACD/NMR predictor[14] are useful to efficiently compare experimental chemical shift information with predicted chemical shifts of a large number of candidate structures, but their accuracy is determined by the nature and size of the underlying database.

As a consequence, one often encounters excellent prediction results for structures that are part of the underlying database, but structures (or substructures) that have not been considered during the fitting procedure are usually predicted with significantly reduced accuracy. Therefore, predictors that are less dependent on or even entirely independent of databases, potentially allow for the more balanced and more accurate chemical shift prediction for both known and unknown structures. In this context, quantum-chemical (QC) methods offer a promising alternative. However, their computational cost is in most cases several orders of magnitude larger and they require the use of high performance computing resources when a large number of predictions is needed. To keep the computational costs manageable, several approximations are usually employed, both regarding the system under study and the QC method itself. Although other procedures have been put forward in the literature,[15,16] a widely used approach is to

**Received:** February 28, 2017
**Revised:** April 7, 2017
**Published:** April 7, 2017

© 2017 American Chemical Society

calculate shielding constants only for a limited number of relevant conformers and to account for solvent effects implicitly. Among the various QC methods, density functional theory (DFT) has been shown to yield both the accuracy and numerical efficiency to allow for NMR chemical shift calculations on a routine basis.[17−24] Nevertheless, recent developments in wave function-based methods, especially MP2, offer potentially useful alternatives for small to medium-sized molecules.[25−27]

The calculated shielding constants can be converted to chemical shifts in three ways: (i) by subtracting them from the shielding constant of an internal reference, such as tetramethylsilane (TMS) or 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS), (ii) by intermediate references or multi-standards, and (iii) by means of (linear) regression. Whereas the first two approaches usually rely on (often fortuitous) error cancellation between shielding constants calculated for the query and reference atom, linear regression allows for the correction of systematic errors as a function of the shielding constant. Several studies have reported excellent agreements with experimental shifts by using DFT together with linear regression and provided fitting parameters for several exchange-correlation/basis set combinations,[20,28−30] where some of them used multiple motif-dependent standards.[31] However, most of them were conducted in organic solvents, such as chloroform, for a relatively small number of molecules. While these results are helpful in the field of general organic chemistry, they are somewhat less practical for metabolomics studies, because NMR chemical shifts are usually measured in aqueous solution. Naturally, the presence of water poses additional challenges for the prediction of NMR chemical shifts of metabolites, especially if combined with implicit solvent models. For example, chemical shifts of atoms in close proximity to hydrogen bond donors or acceptors will have different systematic errors than shifts of those atoms that are further away.[32] Moreover, molecules that allow for intramolecular hydrogen bonding often exhibit large chemical shift deviations when implicit solvent models are employed. This is because conformational search and geometry optimization in implicit solvent favor geometries that are stabilized by intramolecular hydrogen bonds, which often leads to a biased conformational ensemble compared to the one in explicit water. As a result, atoms in different microscopic solvation environments will have different systematic errors.

In the present work, we report a DFT-based protocol, termed molecular motif-specific scaling of density-functional-theory-based chemical shifts (MOSS-DFT), for NMR chemical shift predictions, which is based on a set of 176 molecules that are relevant for metabolomics studies. Chemical shielding constants are converted to chemical shifts by a motif-specific linear regression approach of calculated shielding constants to experimental chemical shift measured with 2D $^1$H−$^{13}$C HSQC spectroscopy. In the first part, details about the MOSS-DFT database construction and the computational methods are given. Thereafter, we provide the definition of motifs and compare our prediction results with those from earlier studies that employ global linear regressions. In the last part, we discuss the influence of basis set size and the exchange-correlation functional by comparing the results to various functional-basis set combinations.

## 2. METHODS

For the MOSS-DFT database construction, a number of molecules were randomly picked from the COLMAR database. We used Open Babel[33] to generate 3D coordinates from 2D structures and to adjust the protonation state of ionizable groups to pH 7, that is, all carboxylic, phosphonic and sulfonic acid groups were deprotonated and all amino groups were protonated. For the calculation of the chemical shifts we largely follow the approach described previously.[21] All molecules were subjected to a conformational search, which was conducted with the MacroModel program[34] that is part of the Schrödinger suite. The OPLS 2005 force field[35] was used together with an implicit solvent model for water.[34] For sampling, the Monte Carlo Multiple Minimum (MCMM) algorithm was chosen[36,37] with the maximum step number set to 5000. Other program options have been set to their default values or set automatically by MacroModel, which is part of the automatic setup. To avoid unphysical conformational ensembles due to the implicit solvation model, the resulting conformers were checked for intramolecular hydrogen bonds, defined by a donor (D)−acceptor (A) distance of less than 3.5 Å and a D−H−A angle of 180° ± 30°. In the case of intramolecular hydrogen bonding in at least one of the conformers, the compound was removed from the database. In addition, molecules with long carbon hydride chains were excluded, if the conformational search yielded a large number (>100) of similarly low energy structures. The conformers obtained from the conformational search were further optimized at the DFT level using the Gaussian 09 program.[38] The B3LYP[39] exchange correlation functional was used together with the D3 dispersion correction[40] and the def2-TZVP basis set.[41] In order to capture solvation effects, the conductor polarized continuum model (CPCM) was adopted using water as solvent.[42,43] To ensure well-converged local minimum geometries, extremely tight convergence criteria have been used along with an ultrafine integration grid. Convergence to local minimum structures has been additionally monitored by normal-mode analysis at the same level of theory. In cases where two initially different conformers had essentially identical total energies ($\Delta E \leq 10^{-9}$ hartree), it was assumed that the DFT optimization led to the same minimum structure and calculations were continued only for one of the structures. The conformer population $p_i$ at 298.15 K was estimated from a Boltzmann analysis

$$p_i = \frac{e^{-E_i/RT}}{\sum_{j=1}^{N} e^{-E_j/RT}} \tag{1}$$

where $E_i$ denotes the relative free energy (electronic + thermal free energy) of conformer $i$ as estimated from the thermochemical analysis in Gaussian 09 with respect to the most stable conformer. $RT$ is the product of the ideal gas constant and the absolute temperature, respectively, and $N$ is the total number of distinct conformers as obtained from geometry optimization. NMR shielding constants were calculated based on the gauge-independent atomic orbitals (GIAO) approach as implemented in Gaussian 09.[44−48] We calculated NMR shielding constants at the DFT level employing five different functional-basis set combinations (see Table 3). The computed NMR shielding values have been averaged according to the weights of the Boltzmann analysis. Because calculated shielding constants have been referenced to chemical shifts obtained from $^1$H−$^{13}$C HSQC experiments, the MOSS-DFT database only contains $^{13}$C−$^1$H

**Table 1. Definitions, Names, and Examples of the Molecular Motifs Used in the Linear Regression Approach**

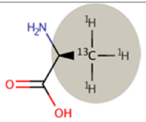
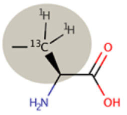
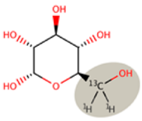
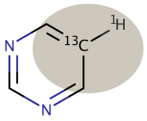
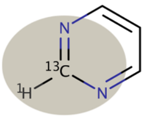
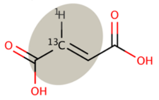
| | | | | | | |
|---|---|---|---|---|---|---|
| Motif Definition: | *Methyl groups* not bonded to a heteroatom | *$sp^3$ carbon centers* not bonded to a heteroatom, excluding methyl groups | *$sp^3$ carbon centers* bonded to a heteroatom | *Aromatic carbon centers* not bonded to a heteroatom | *Aromatic carbon centers* bonded to a heteroatom | *$sp^2$ carbon centers* |
| Motif Name: | $sp^3_{nonhet, met}$ | $sp^3_{nonhet}$ | $sp^3_{het}$ | $ar_{nonhet}$ | $ar_{het}$ | $sp^2$ |
| Examples: |  |  |  |  |  |  |

**Table 2. RMSDs with Respect to Experimental NMR Chemical Shifts along with Linear Regression Parameters for the B97-2/pcS-3 Method**

| nucleus | | $sp^3_{nonhet,met}$ | $ar_{nonhet}$ | $ar_{het}$ | $sp^3_{nonhet}$ | $sp^3_{het}$ | $sp^2$ | total |
|---|---|---|---|---|---|---|---|---|
| $^{13}C$ | a | −0.9532 | −0.8433 | −0.8634 | −0.9277 | −0.9261 | −0.9936 | |
| | b | 176.5288 | 172.4185 | 173.6563 | 171.0732 | 172.1137 | 177.9458 | |
| | RMSD | 1.15 | 1.31 | 1.76 | 2.05 | 2.32 | 3.03 | 1.93 |
| $^{1}H$ | a | −0.9244 | −0.7854 | −0.7725 | −0.9399 | −0.9181 | −0.9677 | |
| | b | 29.2860 | 26.0863 | 26.0812 | 29.8153 | 29.2769 | 30.4447 | |
| | RMSD | 0.079 | 0.118 | 0.239 | 0.146 | 0.177 | 0.214 | 0.154 |
| | no. of $^{13}C-^{1}H$ pairs | 87 | 192 | 35 | 141 | 285 | 16 | 756 |

pairs of covalently bonded atoms. The calculated shielding constants of methyl and methylene protons have been averaged to reduce misassignments in the case of magnetically inequivalent protons. In our calculations, about 4% of the atoms had chemical shift deviations larger than 7 and 0.6 ppm for $^{13}C$ and $^{1}H$, respectively. A closer inspection of the problematic compounds suggests in some cases misassignments in the experimental spectra or obvious problems in conformational sampling, such as errors in dihedral angles, especially for some tertiary carbon atoms. However, for the majority of outliers it was not easily possible to identify the specific cause for the deviation, such as less common functional groups. To prevent biased linear regression parameters, the molecules containing those atoms have been removed from the database, and the linear fit was repeated for the remaining 176 molecules. Finally, to convert the shielding constants to chemical shifts, a motif-specific linear regression approach was used according to

$$\delta = \sigma_{iso} a_i + b_i \qquad (2)$$

where $\sigma_{iso}$ denotes the shielding constant; $a_i$ and $b_i$ are model parameters for motif $i$ (details on motifs are given below), and $\delta$ is the chemical shift value.

## 3. RESULTS AND DISCUSSION

### 3.1. Motif-Specific Linear Regressions for the More Accurate Chemical Shift Prediction.

Table 1 shows the definition and names of motifs that are used to convert the calculated shielding constants to chemical shifts. We assigned each $^{13}C-^{1}H$ pair to one of the motifs based on the chemical environment of its carbon atom. That is, we determined in a first step whether the carbon atom is $sp^3$- or $sp^2$-hybridized, and subsequently, whether it is bonded to any element other than carbon or hydrogen. For $sp^2$ carbons, we further discriminated between aromatic and aliphatic carbons, where we assigned aromatic atoms to $ar_{het}$ and $ar_{nonhet}$ depending on whether they are bonded to a heteroatom or not. Aliphatic $sp^2$ atoms was assigned to the $sp^2$ motif. Because of the relatively small number of atoms within this motif, it was not further divided into hetero- or nonhetero-bonded carbons. Note, these definitions are mutually exclusive; that is, every $^{13}C-^{1}H$ pair is assigned to only one of the motifs. Our motivation for the present classification is 3-fold: we aim (I) to correct for errors due to the implicit modeling of solute-water interactions that will affect carbons bonded to heteroatoms differently than carbon atoms bonded to other carbon atoms, (II) to identify and obtain groups of atoms with superior or inferior prediction accuracy, and optimize their scaling parameters to further reduce errors within these groups, and (III) to account for slightly different slopes of aliphatic and aromatic carbons and protons. As part of the fitting process, we carried out a quadratic fit for each of the motifs and found no significant improvements.

The RMSD values and slopes of the linear regressions for the best performing method, B97-2/pcS-3, are reported in Table 2, where the considered RMSD intervals were limited to 0−7 ppm ($^{13}C$) and 0−0.6 ppm ($^{1}H$), respectively. Our findings show a good overall performance of the MOSS-DFT approach as indicated by RMSD values of 1.93 and 0.154 ppm for $^{13}C$ and $^{1}H$, respectively. A particularly good prediction is achieved for motif $sp^3_{nonhet,met}$ with RMSD values of 1.15 and 0.079 for $^{13}C$ and $^{1}H$, respectively. Therefore, the chemical shift prediction of methyls that are not bonded to any heteroatom should be given a higher weight when such information is used for the identification and validation of unknown metabolites.

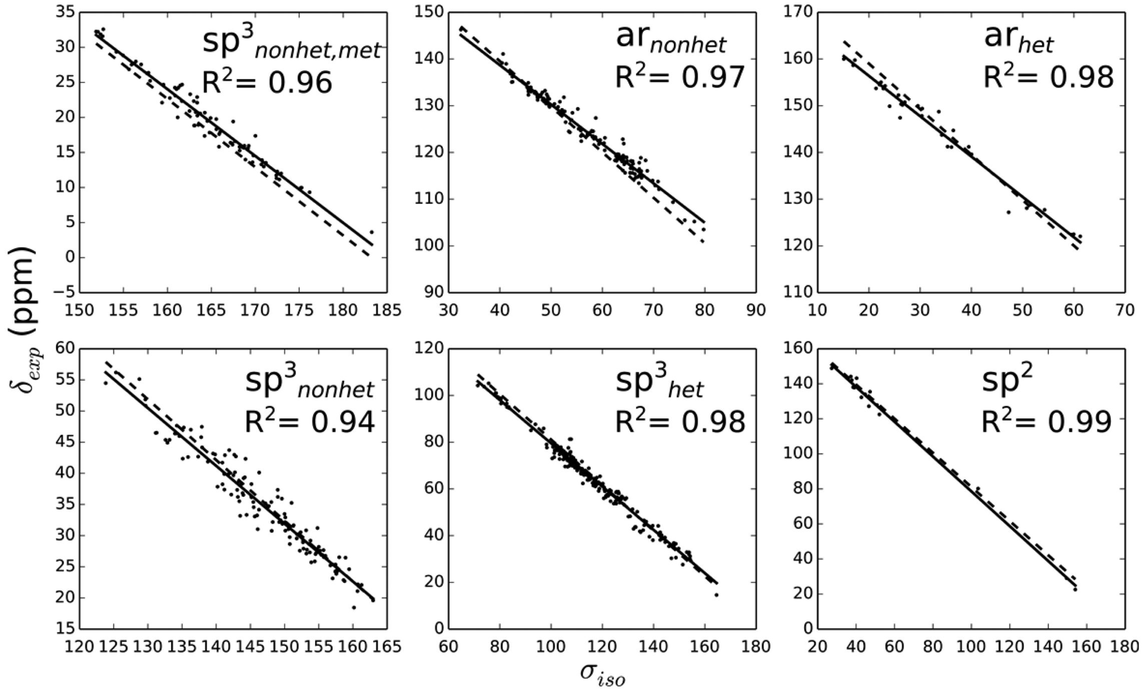


**Figure 1.** Experimental chemical shifts versus computed isotropic shielding constants of $^{13}$C using the B97-2/pcS-3 method. The solid lines belong to the best fit for each motif, while the dashed lines correspond to the global fit.

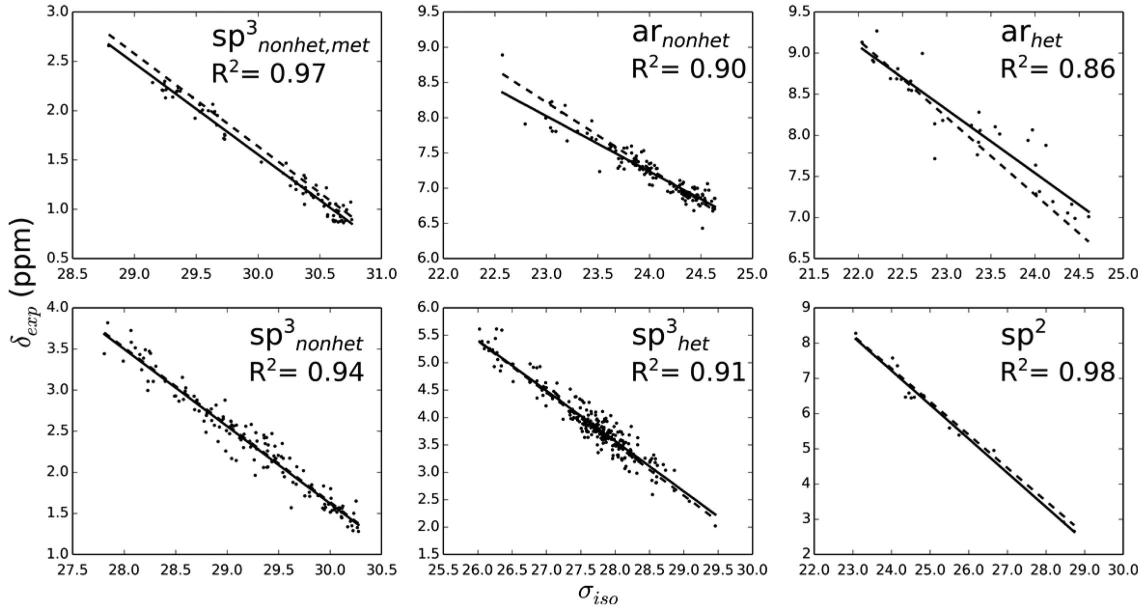


**Figure 2.** Experimental chemical shifts versus computed isotropic shielding constants of $^1$H using the B97-2/pcS-3 method. The solid lines belong to the best fit for each type, while the dashed lines correspond to the global fit.

As a general trend, a higher prediction accuracy is achieved for $^{13}$C−$^1$H pairs that are not directly bonded to a heteroatom, with carbons in the group $ar_{het}$ being a notable exception. The worst prediction results are found for $sp^2$ carbons and protons, as well as for protons within the $ar_{het}$ group. Nevertheless, with the MOSS-DFT prediction approach we obtained an accuracy that is comparable to earlier studies of organic solvents using global scaling factors and intercepts.[20,24,28−30] We also note that in our case the RMSD values for global scaling are 2.36 and 0.170 ppm for $^{13}$C and $^1$H, respectively, and thus significantly larger than those for the motif-specific scaling. To cross-validate our approach, a new training set was created by randomly excluding 20% of the compounds from the original MOSS-DFT training set and assigning them to a test set. The obtained linear regression parameters for the new training set and the RMSD values of the test set are reported in Table S3. We find overall RMSD values of 1.94 and 0.133 ppm for $^{13}$C and $^1$H, respectively, which are well comparable to the ones reported above. In addition, a comparison of the linear regression parameters in Table S3 with the original MOSS-DFT parameters shows only marginal differences, which confirms that the MOSS-DFT parameters are robust, i.e. they are insensitive to the exact choice of the database. As a result, a good prediction accuracy can be expected when applying the MOSS-DFT parameters to targets outside of the training set, provided that the conformational ensemble of the query compound has been obtained in a similar way.

Figures 1 and 2 show the correlation between shielding constants and experimental chemical shifts within the considered RMSD interval. The high $R^2$ coefficients indicate very good correlations considering the narrow chemical shift ranges of most of the motifs. Usually, better correlations are found for $^{13}$C than for $^{1}$H. This is not unexpected, since the chemical shift ranges are much smaller in the latter case. Furthermore, there are significant differences between global and motif-specific fits. The largest discrepancies between global and motif-specific slopes are found for $\mathrm{ar}_{nonhet}$ and $\mathrm{ar}_{het}$, indicating larger systematic errors in these cases. It has been previously proposed for Hartree–Fock that the larger degree of electron correlation in aromatic compounds is responsible for larger systematic errors,[49] but this is still under debate for DFT. Nevertheless, Table 2 shows that the RMSD values of the aromatic groups are mostly lower than those for $\mathrm{sp}^3_{nonhet}$, which illustrates the importance of the motif-specific corrections used here. Another interesting observation is found for motif $\mathrm{sp}^3_{nonhet,met}$: the correlation plots for $^{13}$C and $^{1}$H show almost parallel "best fit" lines for global and motif-specific regressions. At this point, it is unclear whether this result is intrinsic to our model (i.e., GIAO–DFT with implicit water model and the use of optimized geometries) or whether it is of purely statistical nature. The use of a larger sample size with a concomitant reduction of the spread of the error distribution might be able to resolve this issue.

### 3.2. Choice of Basis Set and Exchange-Correlation Functional.

The computational cost of chemical shift predictions is especially crucial in the field of metabolomics as often large sets of candidate compounds exist for which predictions are needed. Clearly, the computational time required for any QC method to calculate shielding constants will depend on the size of the basis set. To find an acceptable trade-off between computational cost and prediction accuracy, the dependence of RMSD values for $^{13}$C and $^{1}$H chemical shifts on the basis set size has been examined. The results for the sequence pcS-1, pcS-2 and pcS-3 are shown in Table 3. Note that the average computational time for a single NMR shielding calculation increases from 1 min for pcS-1 to around 4 h for pcS-3 on 12 cores. It is found that the RMSD value for $^{13}$C chemical shifts decreases by 0.09 ppm when the pcS-2 basis set is used instead of pcS-1. The use of pcS-3 only yields a marginal improvement of 0.03 ppm. Although, on an absolute scale, the differences between RMSD values of $^{1}$H are comparable to those of $^{13}$C, it is clear that changes in the second digit after the decimal point are already a major accuracy gain for $^{1}$H. As for $^{13}$C, our results indicate a consistent improvement for $^{1}$H chemical shifts when going to larger basis sets, where the difference between pcS-1 and pcS-2 is as large as 0.058 ppm. However, unlike for $^{13}$C, the use of B97-2/pcS-3 still leads to a small improvement by 0.01 ppm compared to B97-2/pcS-2. It is also interesting to note that there is no obvious relation between a scaling factor closer to one and a lower RMSD value (see Tables S1 and S2). Nevertheless, one observes convergence of the scaling factors in the case of $^{13}$C for pcS-2 to pcS-3. Thus, we can conclude that the accuracy gain results from an improved correlation between calculated shielding constants and experimental chemical shifts, but not from the reduction of the method's systematic error as reflected by the scaling factor.

Table 3. RMS Deviations of Chemical Shifts for Different Exchange-Correlation Functionals and Basis Sets

| XC functional/basis set | RMSD $^{13}$C (ppm) | RMSD $^{1}$H (ppm) |
|---|---|---|
| B97-2[50]/pcS-1[51] | 2.05 | 0.222 |
| B97-2/pcS-2[51] | 1.96 | 0.164 |
| B97-2/pcS-3[51] | 1.93 | 0.154 |
| B3LYP[39]/pcS-2 | 2.03 | 0.165 |
| BLYP[52−54]/pcS-2 | 2.62 | 0.166 |

In addition to the basis set convergence, we compared the performances of two other exchange-correlation functionals. The results are reported in Table 3. The first functional, B3LYP, is one of the most widely used hybrid DFT functionals for organic molecules, whereas the second, BLYP, is a GGA functional often used in explicit solvent *ab initio* molecular dynamics simulations and NMR shielding calculations due to its numerical efficiency. Compared to B97-2/pcS-2, both functionals perform slightly worse for $^{13}$C with RMSDs of 0.07 and 0.66 ppm for B3LYP and BLYP, respectively. Interestingly, the prediction performance of both functionals for $^{1}$H is consistent with that of B97-2. Unlike the basis set dependence, our findings show a correlation between a lower RMSD and a slope closer to one (see Tables S1 and S2), which indicates a smaller systematic error.

### 3.3. Comparison with Empirical Predictions.

Since the time needed for QC predictions is typically orders of magnitude larger than for empirical predictions, the former need to deliver a substantial accuracy increase to justify their use in metabolomics studies. To verify this, we compared the RMSD values of our MOSS-DFT method to empirical predictions in Table 4. Likewise, for the QC results above, all molecules that contained $^{13}$C/$^{1}$H atoms, whose empirically predicted shifts deviated by more than 7 ppm/0.6 ppm, have been omitted in the RMSD calculation. We note, however, that a completely unbiased comparison between both methods is not possible in our case, because it would require an independent validation set and therefore the knowledge of the training set of the NMRPredict program, which is not available. Our findings show an improvement of more than 1.2 ppm for $^{13}$C and 0.1 ppm for $^{1}$H of the QC method in terms of RMSD over the empirical NMRPredict program. Even by using only a global correction, our QC predictor still outperforms the empirical predictor. However, predictions that are closely related to those in the training set are usually predicted by the empirical method with a very high accuracy, whereas structures that are dissimilar to the training set tend to be predicted

Table 4. RMSD Values for MOSS-DFT (B97-2/pcS-3) and NMRPredict Chemical Shift Predictions with respect to experiment

| | | |
|---|---|---|
| $^{13}$C | $\mathrm{RMSD}_{\mathrm{MOSS\text{-}DFT}}$ | 1.93 ppm |
| | $\mathrm{RMSD}_{\mathrm{NMRPredict}}$ | 3.15 ppm[a] |
| $^{1}$H | $\mathrm{RMSD}_{\mathrm{MOSS\text{-}DFT}}$ | 0.154 ppm |
| | $\mathrm{RMSD}_{\mathrm{NMRPredict}}$ | 0.255 ppm[a] |

[a] Outliers above 7 ppm ($^{13}$C)/0.6 ppm ($^{1}$H) have been excluded

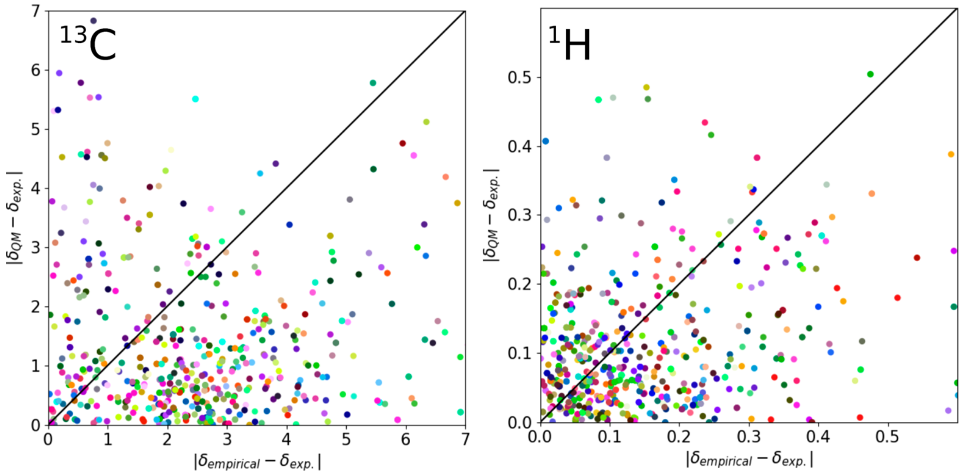


**Figure 3.** Absolute deviations of predictions by MOSS-DFT versus NMRPredict for $^{13}C$ and $^{1}H$. The diagonal is shown to guide the eye. Points marked with the same color belong to the same molecule. Note that the chemical shift error ranges for $^{13}C$ and $^{1}H$ have been limited to the intervals indicated in the figure.

significantly less accurately. We also investigated the correlation between the predictions by MOSS-DFT and NMRPredict (Figure 3) and find that there is no obvious correlation between the two. Nevertheless, both methods seem to be complementary in the sense that there are only few predictions that have large errors in both cases. The low probability of large errors for either method could be used as additional scoring information as part of the identification protocol. For example, if both predictors show a relatively large error for a candidate structure, it can be assigned a lower matching score.

## 4. CONCLUSION

In this work, we have presented a new DFT-based chemical shift prediction approach in aqueous solution, MOSS-DFT, that is specifically suited for the analysis of metabolites. To correct for systematic errors, a motif-specific model was introduced, which goes beyond a common global scaling correction. Our approach is based on a set of 176 molecules, where $^{13}C$—$^{1}H$ atom pairs have been classified into 6 motifs based on hybridization, aromaticity, and heteroatom bonding of the carbon atom. Subsequently, linear regression parameters have been derived separately for each of the motifs. We found a total RMSD value with respect to experimental data of 1.93/0.154 ppm for $^{13}C$/$^{1}H$. This value is 0.43/0.016 ppm lower than that of the global linear correction. As a general trend, predictions of atoms that are not bonded to a heteroatom exhibit a lower error. The best motif-specific RMSD values were obtained for $sp^3_{nonhet,met}$ and $ar_{nonhet}$ with 1.15/0.079 ppm and 1.31/0.118 ppm for $^{13}C$/$^{1}H$ atoms. The nonaromatic $sp^2$ atoms have the highest RMSD (3.03 ppm) for $^{13}C$, whereas for $^{1}H$ the highest RMSD of 0.239 ppm was found for $ar_{het}$. A comparison of different functional/basis set combinations suggests that B97-2/pcS-2 is most economical, although B97-2/pcS-3 is slightly more accurate for $^{1}H$. In the case of $^{1}H$, no significant improvement is found for B97-2 and B3LYP functionals over BLYP. A limitation of the MOSS-DFT approach is for molecules that form intramolecular hydrogen bonds. Therefore, future developments and applications of numerically efficient protocols should focus on the improved sampling and selection of relevant conformations for further improvements toward accurate calculations of chemical shifts of a growing spectrum of synthetic and naturally occurring molecules.

## ■ ASSOCIATED CONTENT

### Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpca.7b01954.

Histograms of absolute deviations, linear regression parameters for employed DFT methods, linear regression parameters and RMSD values of cross-validation analysis, list of molecule IDs used in the test set, and table with molecule IDs and structures of the full MOSS-DFT set (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

*(R.B.) E-mail: bruschweiler.1@osu.edu.

### ORCID

Rafael Brüschweiler: 0000-0003-3649-4543

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

F.H. thanks the Fonds der Chemischen Industrie for a Kekulé fellowship. This work was supported by the National Institutes of Health (Grant R01GM066041 to R.B.). We thank the Ohio Supercomputer Center and the Martin-Luther University Halle-Wittenberg for providing us high performance computing resources.

## ■ REFERENCES

(1) Nagana Gowda, G. A.; Raftery, D. Can NMR solve some significant challenges in metabolomics? *J. Magn. Reson.* **2015**, *260*, 144−60.

(2) Fan, T. W.; Lane, A. N. Applications of NMR spectroscopy to systems biochemistry. *Prog. Nucl. Magn. Reson. Spectrosc.* **2016**, *92−93*, 18−53.

(3) Markley, J. L.; Brüschweiler, R.; Edison, A. S.; Eghbalnia, H. R.; Powers, R.; Raftery, D.; Wishart, D. S. The future of NMR-based metabolomics. *Curr. Opin. Biotechnol.* **2017**, *43*, 34−40.

(4) Bingol, K.; Bruschweiler-Li, L.; Li, D.; Zhang, B.; Xie, M.; Brüschweiler, R. Emerging new strategies for successful metabolite identification in metabolomics. *Bioanalysis* **2016**, *8*, 557−73.

(5) Dunn, W. B.; Broadhurst, D. I.; Atherton, H. J.; Goodacre, R.; Griffin, J. L. Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chem. Soc. Rev.* **2011**, *40*, 387−426.

(6) Beckonert, O.; Keun, H. C.; Ebbels, T. M.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat. Protoc.* **2007**, *2*, 2692−703.

(7) Kim, H. K.; Choi, Y. H.; Verpoorte, R. NMR-based metabolomic analysis of plants. *Nat. Protoc.* **2010**, *5*, 536−49.

(8) Palmnas, M. S.; Vogel, H. J. The future of NMR metabolomics in cancer therapy: towards personalizing treatment and developing targeted drugs? *Metabolites* **2013**, *3*, 373−96.

(9) Bingol, K.; Li, D. W.; Bruschweiler-Li, L.; Cabrera, O. A.; Megraw, T.; Zhang, F.; Brüschweiler, R. Unified and isomer-specific NMR metabolomics database for the accurate analysis of (13)C-(1)H HSQC spectra. *ACS Chem. Biol.* **2015**, *10*, 452−9.

(10) Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; et al. HMDB 3.0–The Human Metabolome Database in 2013. *Nucleic Acids Res.* **2013**, *41*, D801−7.

(11) Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; et al. BioMagResBank. *Nucleic Acids Res.* **2008**, *36*, D402−8.

(12) *NMRPredict*; Modgraph Consultants Ltd..

(13) *Mnova*; Mestrelab Research.

(14) *ACD/NMR Predictors*; Advanced Chemistry Development.

(15) Kwan, E. E.; Liu, R. Y. Enhancing NMR prediction for organic compounds using molecular dynamics. *J. Chem. Theory Comput.* **2015**, *11*, 5083−5089.

(16) Dracinsky, M.; Möller, H. M.; Exner, T. E. Conformational sampling by ab initio molecular dynamics simulations improves NMR chemical shift predictions. *J. Chem. Theory Comput.* **2013**, *9*, 3806−3815.

(17) Bekcioglu-Neff, G.; Allolio, C.; Desmukh, Y. S.; Hansen, M. R.; Sebastiani, D. Dynamical dimension to the Hofmeister series: insights from first-principles simulations. *ChemPhysChem* **2016**, *17*, 1166−73.

(18) Elgabarty, H.; Schmieder, P.; Sebastiani, D. Unraveling the existence of dynamic water channels in light-harvesting proteins: alpha-C-phycocyanobilin in vitro. *Chem. Sci.* **2013**, *4*, 755−763.

(19) Banyai, D. R.; Murakhtina, T.; Sebastiani, D. NMR chemical shifts as a tool to analyze first principles molecular dynamics simulations in condensed phases: the case of liquid water. *Magn. Reson. Chem.* **2010**, *48* (Suppl. 1), S56−S60.

(20) Lodewyk, M. W.; Siebert, M. R.; Tantillo, D. J. Computational prediction of 1H and 13C chemical shifts: a useful tool for natural product, mechanistic, and synthetic organic chemistry. *Chem. Rev.* **2012**, *112*, 1839−62.

(21) Willoughby, P. H.; Jansma, M. J.; Hoye, T. R. A guide to small-molecule structure assignment through computation of 1H and 13C NMR chemical shifts. *Nat. Protoc.* **2014**, *9*, 643−60.

(22) Tantillo, D. J. Walking in the woods with quantum chemistry-applications of quantum chemical calculations in natural products research. *Nat. Prod. Rep.* **2013**, *30*, 1079−86.

(23) Cormanich, R. A.; Buhl, M.; Rittner, R. Understanding the conformational behaviour of Ac-Ala-NHMe in different media. A joint NMR and DFT study. *Org. Biomol. Chem.* **2015**, *13*, 9206−13.

(24) Flaig, D.; Maurer, M.; Hanni, M.; Braunger, K.; Kick, L.; Thubauville, M.; Ochsenfeld, C. Benchmarking hydrogen and carbon NMR chemical shifts at HF, DFT, and MP2 Levels. *J. Chem. Theory Comput.* **2014**, *10*, 572−8.

(25) Gauss, J.; Werner, H. J. NMR chemical shift calculations within local correlation methods: the GIAO-LMP2 approach. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2083−2090.

(26) Loibl, S.; Schutz, M. NMR shielding tensors for density fitted local second-order Møller-Plesset perturbation theory using gauge including atomic orbitals. *J. Chem. Phys.* **2012**, *137*, 084107.

(27) Maurer, M.; Ochsenfeld, C. A linear- and sublinear-scaling method for calculating NMR shieldings in atomic orbital-based second-order Møller-Plesset perturbation theory. *J. Chem. Phys.* **2013**, *138*, 174104.

(28) Konstantinov, I. A.; Broadbelt, L. J. Regression formulas for density functional theory calculated 1H and 13C NMR chemical shifts in toluene-d8. *J. Phys. Chem. A* **2011**, *115*, 12364−72.

(29) Pierens, G. K. 1H and 13C NMR scaling factors for the calculation of chemical shifts in commonly used solvents using density functional theory. *J. Comput. Chem.* **2014**, *35*, 1388−94.

(30) Benassi, E. Benchmarking of density functionals for a soft but accurate prediction and assignment of 1H and 13C NMR chemical shifts in organic and biological molecules. *J. Comput. Chem.* **2017**, *38*, 87−92.

(31) Sarotti, A. M.; Pellegrinet, S. C. A multi-standard approach for GIAO 13C NMR calculations. *J. Org. Chem.* **2009**, *74*, 7254−7260.

(32) Zhu, T.; Zhang, J. Z.; He, X. Automated fragmentation QM/MM calculation of amide proton chemical shifts in proteins with explicit solvent model. *J. Chem. Theory Comput.* **2013**, *9*, 2104−14.

(33) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminf.* **2011**, *3*, 33.

(34) *MacroModel*, Schrödinger Release 2014; Schrödinger: New York, 2014.

(35) Banks, J. L.; Beard, H. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Felts, A. K.; Halgren, T. A.; Mainz, D. T.; Maple, J. R.; et al. Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J. Comput. Chem.* **2005**, *26*, 1752−80.

(36) Chang, G.; Guida, W. C.; Still, W. C. An internal coordinate Monte-Carlo method for searching conformational space. *J. Am. Chem. Soc.* **1989**, *111*, 4379−4386.

(37) Saunders, M.; Houk, K. N.; Wu, Y. D.; Still, W. C.; Lipton, M.; Chang, G.; Guida, W. C. Conformations of cycloheptadecane - a comparison of methods for conformational searching. *J. Am. Chem. Soc.* **1990**, *112*, 1419−1427.

(38) 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., et al., *Gaussian 09*; Gaussian, Inc.: Wallingford, CT, 2009.

(39) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(40) 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. *J. Chem. Phys.* **2010**, *132*, 154104.

(41) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297−305.

(42) Barone, V.; Cossi, M. Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. *J. Phys. Chem. A* **1998**, *102*, 1995−2001.

(43) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.* **2003**, *24*, 669−81.

(44) London, F. Quantum theory of interatomic currents in aromatic compounds. Théorie quantique des courants interatomiques dans les combinaisons aromatiques. *J. Phys. Radium* **1937**, *8*, 397−409.

(45) McWeeny, R. Perturbation theory for the Fock-Dirac density matrix. *Phys. Rev.* **1962**, *126*, 1028.

(46) Ditchfield, R. Self-consistent perturbation theory of diamagnetism: I. A gauge-invariant LCAO method for NMR chemical shifts. *Mol. Phys.* **1974**, *27*, 789−807.

(47) Wolinski, K.; Hinton, J. F.; Pulay, P. Efficient implementation of the gauge-independent atomic orbital method for NMR chemical-shift calculations. *J. Am. Chem. Soc.* **1990**, *112*, 8251−8260.

DOI: 10.1021/acs.jpca.7b01954

(48) Cheeseman, J. R.; Trucks, G. W.; Keith, T. A.; Frisch, M. J. A comparison of models for calculating nuclear magnetic resonance shielding tensors. *J. Chem. Phys.* **1996**, *104*, 5497−5509.

(49) Facelli, J. *Encyclopedia of Nuclear Magnetic Resonance*; Grant, D. M., Harris, R. K., Eds.; John Wiley & Sons: London, 2002; Vol. 9, pp 323−333.

(50) Wilson, P. J.; Bradley, T. J.; Tozer, D. J. Hybrid exchange-correlation functional determined from thermochemical data and ab initio potentials. *J. Chem. Phys.* **2001**, *115*, 9233−9242.

(51) Jensen, F. Basis set convergence of nuclear magnetic shielding constants calculated by density functional methods. *J. Chem. Theory Comput.* **2008**, *4*, 719−27.

(52) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098−3100.

(53) Lee, C. T.; Yang, W. T.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron-density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789.

(54) Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. Results obtained with the correlation-energy density functionals of Becke and Lee, Yang and Parr. *Chem. Phys. Lett.* **1989**, *157*, 200−206.

# Perturbation of the F19-L34 Contact in Amyloid β (1-40) Fibrils Induces Only Local Structural Changes but Abolishes Cytotoxicity

Felix Hoffmann, Juliane Adler, Bappaditya Chandra, Kaustubh R. Mote, Gül Bekçioğlu-Neff, Daniel Sebastiani, and Daniel Huster

Reprinted with permission from

J. Phys. Chem. Lett. **2017**, 8, 4740–4745.

Copyright 2017 American Chemical Society.

`http://dx.doi.org/10.1021/acs.jpclett.7b02317`

The Journal of Physical Chemistry Letters
Letter
pubs.acs.org/JPCL

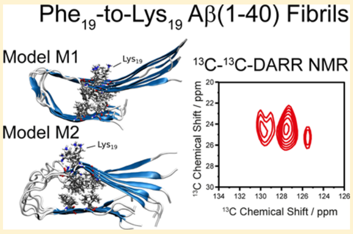
# Perturbation of the F19-L34 Contact in Amyloid $\beta$ (1-40) Fibrils Induces Only Local Structural Changes but Abolishes Cytotoxicity

Felix Hoffmann,*[†] Juliane Adler,[‡] Bappaditya Chandra,[§] Kaustubh R. Mote,[∥] Gül Bekçioğlu-Neff,[†] Daniel Sebastiani,[†] and Daniel Huster*[‡]

[†]Department of Chemistry, Martin-Luther Universität Halle-Wittenberg, von-Danckelmann-Platz 4, 06120 Halle/Saale, Germany
[‡]Institute for Medical Physics and Biophysics, Leipzig University, Härtelstr. 16-18, 04107 Leipzig, Germany
[§]Tata Institute of Fundamental Research, Homi Bhabha Road, Colaba, Mumbai 400 005, India
[∥]TIFR Centre for Interdisciplinary Sciences, Tata Institute of Fundamental Research, 21 Brundavan Colony, Narsingi, Hyderabad 500 075, India

*Supporting Information*

**ABSTRACT:** We explored structural details of fibrils formed by a mutated amyloid $\beta$ (A$\beta$(1-40)) peptide carrying a $Phe_{19}$ to $Lys_{19}$ mutation, which was shown to completely abolish the toxicity of the molecule. Computer models suggest that the positively charged $Lys_{19}$ side chain is expelled from the hydrophobic fibril interior upon fibrillation. This can be accommodated by either a 180° flip of the entire lower $\beta$-strand (model M1) or local perturbations of the secondary structure in the direct vicinity of the mutated site (model M2). This is accompanied by the formation of a new salt bridge between $Glu_{22}$ and $Lys_{28}$ in model M1. Experimentally, a novel contact between $Phe_{20}$ and $Leu_{34}$ as well as the significant structural perturbation of residues 20−23 could be confirmed. However, the mutated fibrils do not show the formation of any salt bridges. This demonstrates that although morphologically very robust, local perturbations of the A$\beta$(1-40) sequence lead to moderate structural alterations with tremendous impact on the physiological importance of these aggregates, which may suggest alternative strategies for the development of a remedy against Alzheimer's disease.



Alzheimer's disease (AD) is the most common form of dementia, affecting an estimated 5.5 million people in the U.S. alone and causing costs of $259 billion alone for the U.S. healthcare system in 2017.[1] The hallmark of AD is the misfolding and aggregation of the amyloid $\beta$ (A$\beta$) peptide to insoluble fibrils that proceeds via the formation of soluble oligomers which represent the most common toxic forms of A$\beta$ aggregates.[2−6] A multitude of research approaches has been dedicated to the characterization of A$\beta$ in its various aggregation stages over the last two decades,[7−11] but our molecular-level understanding of the disease and the aggregation process is still incomplete. This is especially unsatisfactory as residue-specific contacts might exist that are crucial for the toxicity of A$\beta$ aggregates, which could be targeted by drugs that selectively suppress these interactions.[12] In this regard, Das et al. showed that several A$\beta$(1-40) variants that alter the hydrophobic contact between phenylalanine 19 (F19) and leucine 34 (L34) completely abolished the cytotoxicity of A$\beta$(1-40).[13] This clearly suggests that the formation of toxic intermediate structures crucially depends on the F19-L34 contact, found in almost all structural studies of A$\beta$(1-40) fibrils.[14−24] It is particularly intriguing that the aforementioned peptides mutated in position F19 all formed fibrils. While fibril formation occurred at quite different rates but with relatively similar free energies of formation,[25] X-ray diffraction and transmission electron microscopy data indicated a high degree of similarity in the morphology between the fibrils of mutated and wildtype (WT) A$\beta$.[25,26] Thus, to understand the impact of mutations of the crucial F19-L34 contact, specifically with regard to its importance on the toxic intermediate structures, more detailed investigations are required. Here we studied in detail the mutatated F19 to lysine 19 (K19) variant, termed F19K A$\beta$(1-40), on a molecular level and provide a basis for the molecular understanding of the structural alterations as compared with the WT, allowing us to deduce to the role of this contact in the formation of toxic intermediate A$\beta$ aggregates.

In our approach, we investigated the A$\beta$(1-40) F19K mutation by combining molecular modeling and molecular dynamics (MD) simulations with solid-state NMR experiments. To this end, we started from an existing structural model of WT A$\beta$(1-40) fibrils,[17] based on which we created two distinct models for the F19K mutation, as shown in Figure 1. Note that residues 1 to 8 are not part of the models because they are missing in the underlying structural data. These residues did not show any stable secondary structure and are considered to be mobile and exhibit a random coil conformation.[27] We note that more recently published WT A$\beta$(1-40) models by Tycko

**Received:** August 31, 2017
**Accepted:** September 14, 2017
**Published:** September 14, 2017

ACS Publications © XXXX American Chemical Society

DOI: 10.1021/acs.jpclett.7b02317
J. Phys. Chem. Lett. 2017, 8, 4740−4745

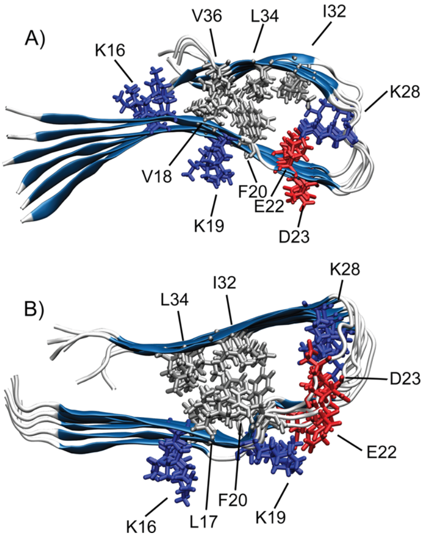


**Figure 1.** Structural models M1 (A) and M2 (B) of mutated F19K A$\beta$(1-40) fibrils after 600 ns of MD simulation. Key residues are drawn in stick representation. Hydrophobic, positively, and negatively charged side chains are highlighted in silver, blue, and red, respectively. Note that residues 1 to 8 are not considered in the models because they were not part of the underlying WT structural model.[17]

et al.[20] and Bertini et al.[16] were either grown from brain-seeded fibrils and thus adopt a different structure than fibrils grown under in vitro conditions[20] or are lacking the D23-K28 salt bridge, contradicting our experimental results for WT A$\beta$(1-40) fibrils. Moreover, although structural models with intertwined protofilaments or antiparallel $\beta$-sheets, similar to the Osaka[28] or the Iowa mutant model,[29] respectively, cannot be ruled out.[30] These structures have not been considered here because the experimental data obtained so far suggest a very high degree of similarity between F19K and WT A$\beta$(1-40) morphologies.

In addition to the postulation of the underlying WT model, three main assumptions for the modeling of the F19K structure were made: (I) The newly introduced K19 side chain is solvated and thus not directed inside the hydrophobic interior of the fibril. (II) The overall structure of the protofilament units is similar to those in the WT fibril; that is, two facing parallel $\beta$-sheets are connected by a turn region that comprise similar sequence regions as in the WT, in agreement with the common structural models.[16,20] (III) Those residues in the upper $\beta$-sheet interact with residues in the lower $\beta$-sheet of the second next peptide strand. Therefore, side chain contacts within the hydrophobic interior are formed intermolecularly. Fibrils are modeled as two opposing copies of the cross-$\beta$ monomers (Figure S1).

The scheme in Figure 2A illustrates the side chain orientations, $\beta$-sheet content, and the molecular contacts between residues in the hydrophobic interior of models M1 and M2. Although the models show distinct side chain orientations, both structures exhibit a hydrophobic core region of the same size involving two residues of the lower and two residues of the upper $\beta$-sheet, respectively. In case of model M1, hydrophobic contacts of <3.2 Å are formed between V18 and V36 as well as between F20 and L34/V36, whereas in model M2 the contacts are formed between side chains of L17 and L34 as well as between F20 and I32/L34. In addition, Figure 2B shows the populations of these inter-residual contacts as calculated from the MD simulations. It becomes apparent that the hydrophobic packing is slightly more compact

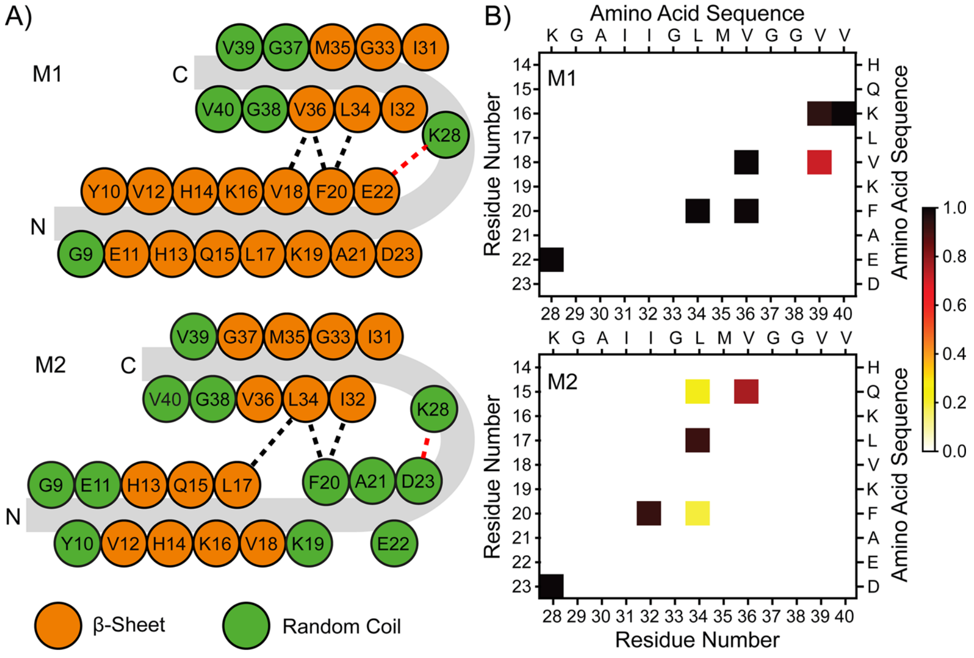


**Figure 2.** (A) Scheme showing the side chain orientations of selected residues in models M1 and M2. The secondary structure information was calculated from the MD simulations trajectories using the DSSP algorithm.[31] Black dashed lines indicate hydrophobic contacts and red dashed lines indicate the salt bridges E22-K28 and D23-K28, respectively. (B) Contact maps showing the time averaged population of side chain contacts between residues belonging to the two opposing $\beta$-strands in models M1 and M2. A contact is defined by a distance of ≤3.2 Å between any two atoms of the residues.

DOI: 10.1021/acs.jpclett.7b02317
J. Phys. Chem. Lett. 2017, 8, 4740−4745

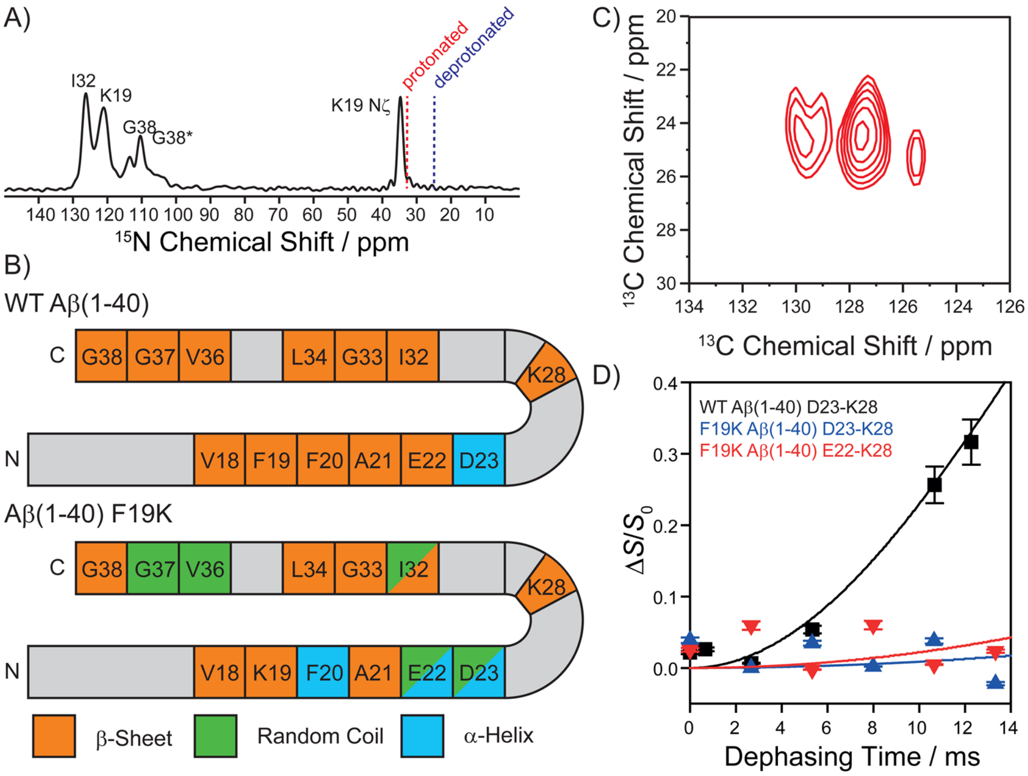


**Figure 3.** Summary of the experimental findings of the main structural features of fibrils formed by the A*β*(1-40) F19K mutant in comparison to WT A*β*(1-40) fibrils. (A) $^{15}$N CPMAS NMR spectrum of fibrils formed by the A*β*(1-40) F19K mutant (sample V2, labeled residues: K19, I32, G38) indicating the protonation state of the $^{15}$N*ζ* side chain. (B) Summary of the assignment of the secondary chemicals shifts to secondary structure elements. (C) Section of a $^{13}$C−$^{13}$C DARR NMR spectrum (mixing time 600 ms) of fibrils grown from F19K mutated A*β*(1-40) peptides indicating the molecular contact between F20 and L34. (D) REDOR dephasing curves for WT A*β*(1-40) labeled in D23 and K28 (black), F19K A*β*(1-40) labeled in D23 and K28 (blue), and F19K A*β*(1-40) labeled in E22 and K28 (red). Lines represent best fit simulations carried out for 4.0 (black), 7.0 (blue), and 6.0 Å (red); the latter $^{13}$C−$^{15}$N distances are not detectable in REDOR experiments.

in model M1 than in model M2. In the first case, all contacts are ∼100% populated, whereas the distance criterion between F20 and L34 in the latter is only fulfilled during 30% of the simulation time. The contact maps also reveal that each model features a salt bridge in the turn region connecting the two *β*-sheets. In model M1, the salt bridge is formed between E22 and K28 and in model M2 between D23 and K28 with 100% population. Regarding the presence of charged or polar side chains in the hydrophobic core, an additional feature is observed in model M2: Here Q15 is engaged in a contact with V36, suggesting that even polar side chains can be tolerated to a certain degree in the hydrophobic core region.

Finally, we investigated the relative stability of the models by means of a generalized Born surface area analysis. The results are presented in Table S1. We found that model M1 is ∼27.6 kJ/mol per strand more stable than model M2. If the two structures were in a direct equilibrium with each other, this free-energy difference would result in a population ratio of approximately 9:1 for models M1 and M2, respectively, as calculated from Monte Carlo simulations.[32]

Next, we carried out solid-state NMR experiments to investigate specific structural features of fibrils formed by the A*β*(1-40) F19K mutant. First, the protonation state of the K19 N*ζ* side chain was investigated by $^{15}$N CPMAS NMR spectroscopy. Figure 3A shows the 1D $^{15}$N NMR spectrum of fibrils of the A*β*(1-40) F19K mutant, where I32, K19, and G38 were U-$^{13}$C/$^{15}$N labeled. The side chain signal of K19 N*ζ* side is detected at 34.7 ppm, which clearly indicates the fully protonated (and thus positively charged) state of the amine. Reference values for the protonated and deprotonated states, which are separated by an approximately −7.5 ppm upfield shift, are indicated.[33]

Next, we investigated the secondary structure of selected amino acids of fibrils formed by the A*β*(1-40) F19K mutant in comparison with WT A*β*(1-40) fibrils. Several peptides with alternating $^{13}$C/$^{15}$N labeling schemes were converted into fibrils and subjected to solid-state NMR measurements, and signal assignment was achieved in 2D $^{13}$C-$^{13}$C DARR and $^{13}$C-$^{15}$N HetCor NMR experiments. A complete list of the chemical shifts for the respective amino acids is given in Table S2. The structural interpretation of the secondary chemical shift values is illustrated in Figure 3B and Figure S4. In WT A*β*(1-40) fibrils, all investigated amino acids showed pronounced *β*-sheet character in accordance with the structural model. Only D23, which is involved in the salt bridge with K28, showed *α*-helical conformation. K28 was not studied for WT A*β*(1-40) here, but according to the literature,[16,17] it shows *β*-sheet-like chemical shifts. In fibrils of the A*β*(1-40) F19K mutant, significant chemical shift changes are observed. In the upper *β*-strand, residues G37 and V36 are in random-coil structure, while for I32 two polymorphs are observed with *β*-sheet or random-coil chemical shifts, respectively. More drastic are the chemical shift alterations in the lower *β*-strand. F20 shows a backbone conformation in agreement with *α*-helix, and E22 and D23 provide chemical shifts at the boundary between random coil and $\alpha$-helix. The mutated K19 is found in $\beta$-strand conformation. This indicates that, in particular, the secondary structure of residues 20 through 23 is significantly perturbed by the F19K mutation.

Next, we investigated if a hydrophobic contact between the side chains of F20 and L34 was formed, as suggested in the MD models. This would be a possible structural consequence of the K19 side chain rotating out of the hydrophobic fibrillary interior and pointing toward the aqueous phase. Figure 3C displays a section of a $^{13}$C-$^{13}$C DARR NMR spectrum of fibrils grown from the A$\beta$(1-40) F19K mutant with $^{13}$C labels at F20, E22, and L34. Cross peaks between the F20 ring carbons and the L34 C$\gamma$, C$\delta$, and C$\varepsilon$ signals are clearly identified, suggesting a molecular contact between the F20 and L34 side chains.

Finally, $^{13}$C—$^{15}$N frequency selective REDOR NMR experiments were conducted to confirm the presence of the D23-K28 salt bridge in WT A$\beta$(1-40) fibrils and to explore the salt bridge preference between K28 and either E22 or D23 in the A$\beta$(1-40) F19K variant. For WT A$\beta$(1-40) fibrils, the increase in the $\Delta S/S_0$ as a function of REDOR dephasing time (Figure 3D) can be fitted with a $^{15}$N—$^{13}$C distance of 4.0 Å, indicative of the formation of a stable salt bridge in agreement with the literature.[34] However, no salt bridge could be detected between either E22-K28 or D23-K28 in fibrils formed by the A$\beta$(1-40) F19K mutant.

We also determined $^{1}$H-$^{13}$C$\alpha$ order parameters of WT and mutated F19K A$\beta$(1-40) fibrils, which provide information about the dynamics in the peptide backbone (Figure S5, Tables S2 and S3).[27] Overall, relatively high order parameters are measured, which are quite similar for WT and mutated fibrils and also agree reasonably well between experiment and simulation. Interesting differences are found for F19 in the WT and K19 in the mutant. Although the K19 side chain points out of the hydrophobic interior in contrast with F19, its order parameters are significantly higher than for F19, indicating that the structure of the F19K must be very rigid in this region. In contrast, the flipped architecture of F20, pointing outward in WT and inward in fibrils of the F19K mutant, has no influence on the fluctuations of the C$\alpha$-H bond vector. D23, which is engaged in the salt bridge with K28 in WT A$\beta$(1-40) fibrils, has a lower order in the F19K mutant, where no such salt bridge is found. Similarly, E22 is more rigid in the F19K mutant than in the WT. Finally, somewhat higher order is detected for G33 and V36 in the F19K mutant. Order parameters determined from the MD for G37 and G38 are significantly lower than determined in the experiment. This discrepancy might be a result of different spatial arrangements of the cross-$\beta$ units with respect to each other or due to the lower water content in the experiment.

Cellular studies have shown that mutations of residue F19 forming the important F19-L34 hydrophobic contact of A$\beta$(1-40) completely abolish toxicity and hamper membrane binding.[13] The structural basis of this intriguing biological effect remains enigmatic as fibrils of very similar morphology are formed by all investigated mutated peptides.[25,26] Here we focus on the F19K mutant of A$\beta$(1-40) and compare structural features of MD models with experimental studies. Replacing a hydrophobic phenylalanine residue by a lysine is accompanied by a high free-energy penalty of 18.9 kJ/mol considering an experimentally determined octanol—water hydrophobicity scale[35] and assuming that the N$\zeta$ amide remains fully protonated, which was confirmed experimentally. This suggests that the K19 side chain is rotated out of the hydrophobic core of the oligomer/fibril, which could be confirmed experimentally by detecting a previously unreported hydrophobic contact between F20 and L34. Such a scenario would require significant secondary or tertiary structure changes in the vicinity of the mutated side. Our two MD models considered either local secondary structure alterations associated with the inside-out flip of K19 and the outside-in exchange F20 or a 180° rotation of the entire lower $\beta$-strand of the A$\beta$(1-40) structure (Figure 1). The latter would be energetically more favorable by about −9.4 kJ/mol considering side chain hydrophobicity.[35] To explore if such a substantial structural alteration was occurring in the F19K mutants, we investigated the secondary structure of the fibrils formed by A$\beta$(1-40) F19K peptides. Indeed, significant local secondary structure changes were detected, especially involving residues 20—23 more in agreement with the local K19 inside-out and F20 outside-in exchange. Furthermore, a salt bridge between E22 and K28, as predicted in the flipped $\beta$-strand model, was not detected; however, also the salt bridge between D23 and K28 known for mature WT A$\beta$ fibrils was absent. It is obvious that the models, which start from the mature WT fibril structure, are not expected to fully reproduce all experimentally found features, as the assembly process of the experimental preparation is not reproduced in the simulation due to the extended time scales of the aggregation process.

It is structurally understandable that F20 can replace the mutated position 19 and maintain the important hydrophobic contacts to I32, L34, or V36. The positively charged K19 side chain must be expelled from the hydrophobic interior of the A$\beta$ fibrils, avoiding a high-energy penalty of such an arrangement. However, this exchange causes significant alterations in the lower $\beta$-strand and abolishes the D23-K28 salt bridge, which is discussed as a structure stabilizing element of mature A$\beta$ fibrils.[36] However, as shown previously, fibrils of the A$\beta$(1-40) F19K variant are only slightly destabilized by ~0.5 kJ/mol compared with the WT.[25] Nevertheless, all structural alterations especially relevant for the oligomeric level are crucial enough to abolish toxicity and prevent membrane binding of the mutated A$\beta$ peptides.

The abolished toxicity of the F19K A$\beta$ mutant must be the result of structural changes on the oligomer/protofibrillar level. Here our understanding is still rather incomplete. A partially folded structure of A$\beta$(1-40) in solution revealed a molecular contact between F19 and G38 stabilizing a C-terminal hairpin,[8] indicating the importance of the F19 side chain for the formation of intermediate structures. In this structure, residues 13 to 23 of A$\beta$(1-40) form a $3_{10}$ helix, which is possibly echoed in the significant perturbation of the $\beta$-strand structure that we observed for residues 20—23 (Figure 3B). Another study also reported the absence of the F19-L34 contact in large oligomers of A$\beta$(1-40).[37] Controversially, the F19-L34 contact was clearly shown in other preparation of smaller oligomers[36,38] and also in protofibrils of A$\beta$(1-40).[22] In A$\beta$(1-42) oligomers, contacts of F19 with L34[10] or Ile31/Ala30[39,40] have been observed, again underlining the importance of the F19 side chain in the structure formation in intermediate A$\beta$ species. From these and our findings, it becomes clearer that a perturbation of residue F19 is causative for the major changes in the observed A$\beta$ toxicity.

It should be emphasized that the mutations created in position 19 of A$\beta$ peptides were not motivated to study biological effects per se, although the high relevance of this position for the toxicity was demonstrated. In the human genome, there is indeed a single nucleotide polymorphism (SNP) known involving position F690 in the amyloid precursor protein (APP), which corresponds to position 19 in Aβ(1-40). The dbSNP database (https://www.ncbi.nlm.nih.gov/projects/SNP/) lists a known exchange of phenylalanine to leucine (rs201724975). It is, however, not known whether individuals who carry this mutation are susceptible to developing AD. The results of our study may also suggest an alternative strategy in the development of a remedy against the disease. Instead of finding antibodies that target a specific epitope of oligomeric Aβ structures, small molecules that specifically affect the F19-L34 hydrophobic contact may represent an attractive alternative strategy.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpclett.7b02317.

> Computational and experimental details, figure illustrating the NMR chemical shifts and their structural meaning, tables containing chemical shifts, $^{1}\text{H}-^{13}\text{C}\alpha$ order parameters, and calculated order parameters of models M1 and M2. (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

*F.H.: felix.hoffmann@chemie.uni-halle.de.
*D.H.: daniel.huster@medizin.uni-leipzig.de.

### ORCID

Daniel Sebastiani: 0000-0003-2240-3938
Daniel Huster: 0000-0002-3273-0943

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

D.H. thanks Profs. P. Hildebrand, P. K. Madhu, and T. Schöneberg for helpful discussions. The study was supported by the Deutsche Forschungsgemeinschaft (SFB TRR 102, A6 and A9). F.H. thanks the Fonds der Chemischen Industrie for a Kekulé fellowship. We are indebted to the Center for Information Services and High Performance Computing (ZIH) at the Dresden University of Technology for computational resources.

## ■ REFERENCES

(1) Alzheimer'ss Association. Alzheimer's disease facts and figures. *Alzheimer's Dementia* **2017**, *13*, 325−373.

(2) Haass, C.; Selkoe, D. J. Soluble protein oligomers in neurodegeneration: lessons from the alzheimer's amyloid β-peptide. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 101−112.

(3) Näslund, J. Correlation between elevated levels of amyloid β-peptide in the brain and cognitive decline. *JAMA* **2000**, *283*, 1571.

(4) Lue, L.-F.; Kuo, Y.-M.; Roher, A. E.; Brachova, L.; Shen, Y.; Sue, L.; Beach, T.; Kurth, J. H.; Rydel, R. E.; Rogers, J. Soluble amyloid β peptide concentration as a predictor of synaptic Change in Alzheimer's disease. *Am. J. Pathol.* **1999**, *155*, 853−862.

(5) McLean, C. A.; Cherny, R. A.; Fraser, F. W.; Fuller, S. J.; Smith, M. J.; Vbeyreuther, K.; Bush, A. I.; Masters, C. L. Soluble pool of Aβ amyloid as a determinant of severity of neurodegeneration in Alzheimer's disease. *Ann. Neurol.* **1999**, *46*, 860−866.

(6) Wang, J.; Dickson, D. W.; Trojanowski, J. Q.; Lee, V. M. Y. The levels of soluble versus insoluble brain Aβ distinguish Alzheimer's disease from normal and pathologic aging. *Exp. Neurol.* **1999**, *158*, 328−337.

(7) DeToma, A. S.; Salamekh, S.; Ramamoorthy, A.; Lim, M. H. Misfolded proteins in Alzheimer's disease and type II diabetes. *Chem. Soc. Rev.* **2012**, *41*, 608−621.

(8) Vivekanandan, S.; Brender, J. R.; Lee, S. Y.; Ramamoorthy, A. A Partially folded structure of amyloid-β(1−40) in an aqueous environment. *Biochem. Biophys. Res. Commun.* **2011**, *411*, 312−316.

(9) Sato, T.; Kienlen-Campard, P.; Ahmed, M.; Liu, W.; Li, H.; Elliott, J. I.; Aimoto, S.; Constantinescu, S. N.; Octave, J.-N.; Smith, S. O. Inhibitors of amyloid toxicity based on β-sheet packing of Aβ40 and Aβ42. *Biochemistry* **2006**, *45*, 5503−5516.

(10) Ahmed, M.; Davis, J.; Aucoin, D.; Sato, T.; Ahuja, S.; Aimoto, S.; Elliott, J. I.; Van Nostrand, W. E.; Smith, S. O. Structural conversion of neurotoxic amyloid-β1-42 oligomers to fibrils. *Nat. Struct. Mol. Biol.* **2010**, *17*, 561−567.

(11) Wälti, M. A.; Ravotti, F.; Arai, H.; Glabe, C. G.; Wall, J. S.; Böckmann, A.; Güntert, P.; Meier, B. H.; Riek, R. Atomic-resolution structure of a disease-relevant Aβ(1-42) amyloid fibril. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, E4976−E4984.

(12) Kachlishvili, K.; Dave, K.; Gruebele, M.; Scheraga, H. A.; Maisuradze, G. G. Eliminating a protein folding intermediate by tuning a local hydrophobic contact. *J. Phys. Chem. B* **2017**, *121*, 3276−3284.

(13) Das, A. K.; Rawat, A.; Bhowmik, D.; Pandit, R.; Huster, D.; Maiti, S. An early folding contact between Phe19 and Leu34 is critical for amyloid-β oligomer toxicity. *ACS Chem. Neurosci.* **2015**, *6*, 1290−1295.

(14) Tycko, R.; Wickner, R. B. Molecular structures of amyloid and prion fibrils: consensus versus controversy. *Acc. Chem. Res.* **2013**, *46*, 1487−1496.

(15) Petkova, A. T.; Ishii, Y.; Balbach, J. J.; Antzutkin, O. N.; Leapman, R. D.; Delaglio, F.; Tycko, R. A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 16742−16747.

(16) Bertini, I.; Gonnelli, L.; Luchinat, C.; Mao, J.; Nesi, A. A new structural model of Aβ40 fibrils. *J. Am. Chem. Soc.* **2011**, *133*, 16013−16022.

(17) Paravastu, A. K.; Leapman, R. D.; Yau, W. M.; Tycko, R. Molecular structural sasis for polymorphism in Alzheimer's β-amyloid fibrils. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 18349−18354.

(18) Lopez del Amo, J. M.; Schmidt, M.; Fink, U.; Dasari, M.; Fändrich, M.; Reif, B. An asymmetric dimer as the dasic subunit in Alzheimer's disease amyloid β fibrils. *Angew. Chem., Int. Ed.* **2012**, *51*, 6136−6139.

(19) Petkova, A. T.; Yau, W.-M.; Tycko, R. Experimental constraints on quaternary structure in Alzheimer's β-amyloid fibrils. *Biochemistry* **2006**, *45*, 498−512.

(20) Lu, J.-X.; Qiang, W.; Yau, W.-M.; Schwieters, C. D.; Meredith, S. C.; Tycko, R. Molecular structure of β-amyloid fibrils in Alzheimer's disease brain tissue. *Cell* **2013**, *154*, 1257−1268.

(21) Ladiwala, A. R. A.; Lin, J. C.; Bale, S. S.; Marcelino-Cruz, A. M.; Bhattacharya, M.; Dordick, J. S.; Tessier, P. M. Resveratrol selectively remodels soluble oligomers and fibrils of amyloid Aβ into off-pathway conformers. *J. Biol. Chem.* **2010**, *285*, 24228−24237.

(22) Scheidt, H. A.; Morgado, I.; Rothemund, S.; Huster, D.; Fändrich, M. Solid-state NMR spectroscopic investigation of Aβ protofibrils: implication of a β-sheet remodeling upon maturation into terminal amyloid fibrils. *Angew. Chem., Int. Ed.* **2011**, *50*, 2837−2840.

(23) Carballo-Pacheco, M.; Strodel, B. Advances in the simulation of protein aggregation at the atomistic scale. *J. Phys. Chem. B* **2016**, *120*, 2991−2999.

(24) Rossetti, G.; Magistrato, A.; Pastore, A.; Persichetti, F.; Carloni, P. Structural properties of polyglutamine aggregates investigated via molecular dynamics simulations. *J. Phys. Chem. B* **2008**, *112*, 16843−16850.

(25) Adler, J.; Baumann, M.; Voigt, B.; Scheidt, H. A.; Bhowmik, D.; Häupl, T.; Abel, B.; Madhu, P. K.; Balbach, J.; Maiti, S.; Huster, D. A detailed analysis of the morphology of fibrils of selectively mutated amyloid β (1−40). *ChemPhysChem* **2016**, *17*, 2744−2753.

(26) Adler, J.; Scheidt, H. A.; Krüger, M.; Thomas, L.; Huster, D. Local interactions influence the fibrillation kinetics, structure and dynamics of Aβ(1−40) but leave the general fibril structure unchanged. *Phys. Chem. Chem. Phys.* **2014**, *16*, 7461.

(27) Scheidt, H. A.; Morgado, I.; Rothemund, S.; Huster, D. Dynamics of amyloid β fibrils revealed by solid-state NMR. *J. Biol. Chem.* **2012**, *287*, 2017−2021.

(28) Schütz, A. K.; Vagt, T.; Huber, M.; Ovchinnikova, O. Y.; Cadalbert, R.; Wall, J.; Güntert, P.; Böckmann, A.; Glockshuber, R.; Meier, B. H. Atomic-resolution three-Dimensional structure of amyloid β fibrils bearing the Osaka mutation. *Angew. Chem., Int. Ed.* **2015**, *54*, 331−335.

(29) Qiang, W.; Yau, W. M.; Luo, Y.; Mattson, M. P.; Tycko, R. Antiparallel β-sheet architecture in Iowa-mutant β-amyloid fibrils. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 4443−4448.

(30) Schledorn, M.; Meier, B. H.; Böckmann, A. Alternative salt bridge formation in Aβ - a hallmark of early-onset Alzheimer's disease? *Front. Mol. Biosci.* **2015**, *2*, 14.

(31) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577−2637.

(32) Raz, Y.; Adler, J.; Vogel, A.; Scheidt, H. A.; Häupl, T.; Abel, B.; Huster, D.; Miller, Y. The influence of the ΔK280 mutation and N- or C-terminal extensions on the structure, dynamics, and fibril morphology of the tau R2 repeat. *Phys. Chem. Chem. Phys.* **2014**, *16*, 7710.

(33) Platzer, G.; Okon, M.; McIntosh, L. P. pH-dependent random Coil $^1$H, $^{13}$C, and $^{15}$N chemical shifts of the ionizable amino acids: a guide for protein pK a measurements. *J. Biomol. NMR* **2014**, *60*, 109−129.

(34) Mithu, V. S.; Sarkar, B.; Bhowmik, D.; Chandrakesan, M.; Maiti, S.; Madhu, P. K. $Zn^{2+}$ Binding disrupts the Asp23-Lys28 salt bridge without altering the hairpin-shaped cross-β structure of Aβ42 amyloid aggregates. *Biophys. J.* **2011**, *101*, 2825−2832.

(35) Wimley, W. C.; White, S. H. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat. Struct. Mol. Biol.* **1996**, *3*, 842−848.

(36) Sarkar, B.; Mithu, V. S.; Chandra, B.; Mandal, A.; Chandrakesan, M.; Bhowmik, D.; Madhu, P. K.; Maiti, S. Significant structural differences between transient amyloid-β oligomers and less-toxic fibrils in regions known to harbor familial Alzheimer's mutations. *Angew. Chem., Int. Ed.* **2014**, *53*, 6888−6892.

(37) Prade, E.; Barucker, C.; Sarkar, R.; Althoff-Ospelt, G.; Lopez del Amo, J. M.; Hossain, S.; Zhong, Y.; Multhaup, G.; Reif, B. Sulindac sulfide induces the formation of large oligomeric aggregates of the Alzheimer's disease amyloid-β peptide which exhibit reduced neurotoxicity. *Biochemistry* **2016**, *55*, 1839−1849.

(38) Chandra, B.; Bhowmik, D.; Maity, B. K.; Mote, K. R.; Dhara, D.; Venkatramani, R.; Maiti, S.; Madhu, P. K. Major reaction coordinates linking transient amyloid-β oligomers to fibrils measured at atomic level. *Biophys. J.* **2017**, *113*, 805−816.

(39) Tay, W. M.; Huang, D.; Rosenberry, T. L.; Paravastu, A. K. The Alzheimer's amyloid-β(1-42) peptide forms off-pathway oligomers and fibrils that are distinguished structurally by intermolecular organization. *J. Mol. Biol.* **2013**, *425*, 2494−2508.

(40) Lendel, C.; Bjerring, M.; Dubnovitsky, A.; Kelly, R. T.; Filippov, A.; Antzutkin, O. N.; Nielsen, N. C.; Härd, T. A hexameric peptide barrel as building block of amyloid-β protofibrils. *Angew. Chem., Int. Ed.* **2014**, *53*, 12756−12760.

# B. Academic curriculum vitae

# Felix Hoffmann – Curriculum Vitae

| | | | |
|---|---|---|---|
| **Address** | Südstr. 23, | **Phone** | +49 178 6313054 |
| | 06110 Halle/Saale, | **Email** | felix.hoffmann@chemie.uni-halle.de |
| | Germany | **Google Scholar** | https://goo.gl/6FuEdp |
| **Date of Birth** | November 19, 1987 | | |
| **Nationality** | German | | |

## Education

**Doctoral Researcher (PhD)** **January 2014 - Present**

*Martin-Luther University Halle-Wittenberg, Germany – Institute of Chemistry*
*Advisor: Prof. Dr. Daniel Sebastiani*

Thesis Title: Structural Driving Forces and Their Spectroscopic Signatures: From Protonation Dynamics to Amyloid Formation

**Visiting Scholar** **June 2016 - November 2016**

*The Ohio State University, US – Department of Chemistry and Biochemistry*

**M. Sc. in Chemistry** **October 2011 - September 2013**

*Ruhr-University Bochum, Germany – Faculty of Chemistry and Biochemistry*
*Advisors: Dr. Jochen Blumberger, Dr. Jörg Behler, Prof. Dr. Dominik Marx*

Thesis Title: Assessment of the Accuracy of Constrained Density Functional Theory for the Calculation of Electronic Couplings

**Visiting Scholar** **April 2013 - September 2013**

*University College London, UK – Department of Physics and Astronomy*

**B. Sc. in Chemistry** **October 2008 - September 2011**

*Ruhr-University Bochum, Germany – Faculty of Chemistry and Biochemistry*
*Advisors: Dr. Jörg Behler, Prof. Dr. Dominik Marx*

Thesis Title: DFT-Untersuchungen zur Adsorption von Weinsiäure auf Kupferoberflächen

## Scholarships

**Kekulé PhD Scholarship of the Fonds der Chemischen Industrie** **June 2014 - May 2016**

**Cusanuswerk Scholarship (Grundförderung)** **October 2009 - September 2013**

## Professional Memberships

**Integrated Research Training Group of the SFB-TRR 102 (interdisciplinary graduate school)** **January 2015 - Present**

**Gesellschaft Deutscher Chemiker** **June 2007 - Present**

**Deutsche Physikalische Gesellschaft** **June 2007 - Present**

## Language Skills

**English:** Advanced proficiency

**German:** Native speaker

# Publications

- **F. Hoffmann**, J. Adler, G. Bekçioğlu-Neff, D. Sebastiani, and D. Huster
  *Abolishment of the F19-L34 Contact in Amyloid* β *(1-40) Fibrils Induces Major Structural Changes with Possible Implications for Cytotoxicity*, J. Phys. Chem. Lett., 2017, 8 (19), 4740–4745
  URL: http://dx.doi.org/10.1021/acs.jpclett.7b02317
- **F. Hoffmann**, Da-Wei Li, D. Sebastiani, and R. Brüschweiler
  *Improved Quantum Chemical NMR Chemical Shift Prediction of Metabolites in Aqueous Solution toward the Validation of Unknowns* J. Phys. Chem. A, 2017, 121 (16), 3071–3078
  URL: http://dx.doi.org/10.1021/acs.jpca.7b01954
- **F. Hoffmann**, M. Ekimova, G. Bekçioğlu-Neff, E. Nibbering, and D. Sebastiani
  *Combined Experimental and Theoretical Study of the Transient IR Spectroscopy of 7-Hydroxyquinoline in the First Electronically Excited Singlet State* J. Phys. Chem. A, 2016, 120 (47), 9378–9389 URL: http://dx.doi.org/10.1021/acs.jpca.6b07843
- G. Bekçioğlu-Neff, **F. Hoffmann**, and D. Sebastiani
  *Solvation-Dependent Latency of Photoacid Dissociation and Transient IR Signatures of Protonation Dynamics* J. Phys. Chem. A, 2015, 119 (35), 9244–9251
  URL: http://dx.doi.org/10.1021/acs.jpca.5b05438
- A. Kubas, **F. Hoffmann**, A. Heck, H. Oberhofer, M. Elstner, and J. Blumberger
  *Electronic Couplings for Molecular Charge Transfer: Benchmarking CDFT, FODFT and FODFTB against High-Level Ab Initio Calculations* J. Chem. Phys., 2014, 140, 104105
  URL: http://dx.doi.org/10.1063/1.4867077
- F. Gajdos, S. Valner, **F. Hoffmann**, J. Spencer, M. Breuer, A. Kubas, M. Dupuis, and J. Blumberger
  *Ultrafast Estimation of Electronic Couplings for Electron Transfer between Pi-Conjugated Organic Molecules* J. Chem. Theory Comput., 2014, 10, 4653
  URL: http://dx.doi.org/10.1021/ct500527v

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Hilfsmittel und Quellen angefertigt habe.

Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen sind, habe ich als solche kenntlich gemacht. Des Weiteren versichere ich, dass ich diese Arbeit an keiner anderen Institution eingereicht habe.

Mir ist bekannt, dass bei Angabe falscher Aussagen die Prüfung als nicht bestanden gilt. Mit meiner Unterschrift versichere ich die Richtigkeit dieser Angaben und erkenne die rechtlichen Grundlagen an.

Felix Hoffmann

Datum