

**Probing the Conformational Space of Biomolecules  
using Large Scale Molecular Dynamics Simulation**

**Dissertation**

zur Erlangung des Doktorgrades der Naturwissenschaften

(Dr. rer. nat.)

der

Naturwissenschaftlichen Fakultät II

Chemie, Physik und Mathematik

der Martin-Luther-Universität

Halle-Wittenberg

**1st Reviewer:** Prof. Dr. Daniel Sebastiani

**2nd Reviewer:** Prof. Dr. Johannes Kästner

vorgelegt von

**Herrn Thomas Kunze**

Defense Date 16.05.2025

## *Abstract*

This thesis addresses the topic of peptide structure formation and the computational generation of plausible conformations for these systems. Although the value of this concept may appear self-evident, it has the potential to be applied in a wide range of contexts. Potential avenues for further development include applications in bio-materials, medication, and disease understanding. Given that minor alterations to proteins and peptides can result in markedly disparate chemical and biochemical behaviours, it is beneficial to utilise computational techniques to gain insight into specific behaviours. The selection of an appropriate computational method is often challenging, given the wide range of available options, which differ greatly in terms of accuracy and computational cost. In certain instances, it may not be feasible to sacrifice precision in order to calculate larger molecules. In this thesis, we propose a novel approach to the unification of precision and affordable computational cost.

## List of Publications

- [1] T. Kunze, C. Dreßler, and D. Sebastiani, "Secondary Structure Formation in Hybrid Synthetic/Peptide Polymers: Insights from Molecular Dynamics Simulations," *Macromol. Theory Simulations*, Vol. 32, No. 3, 2023, pp. 1–8. doi: 10.1002/mats.202200070
- [2] M.-A. Codescu, T. Kunze, M. Weiß, M. Brehm, O. Kornilov, D. Sebastiani, and E. T. J. Nibbering, "Ultrafast Proton Transfer Pathways Mediated by Amphoteric Imidazole," *J. Phys. Chem. Lett.*, Vol. 14, No. 20, May 2023, pp. 4775–4785. doi: 10.1021/acs.jpclett.3c00595
- [3] T. Kunze, C. Dreßler, C. Lauer, W. Paul, and D. Sebastiani, "Reverse Mapping of Coarse Grained Polyglutamine Conformations from PRIME20 Sampling," *ChemPhysChem*, Vol. 25, No. 9, May 2024. doi: 10.1002/cphc.202300521
- [4] T. Kunze, C. Dressler, C. Lauer, and D. Sebastiani, "Assignment of a Physical Energy Scale for the Dimensionless Interaction Energies within the PRIME20 Peptide Model," *ChemPhysChem*, Vol. e202400592, August 2024. doi: 10.1002/cphc.202400592

## *Acknowledgments*

I'd like to start by thanking Daniel for being such a great professor to work for. He was an amazing supervisor, guiding me on equal footing with plenty of freedom and patience. While there were a few challenges along the way, I'm really grateful to have had the opportunity to work here for my PhD. Additionally, I'd like acknowledge his excellent (importing) recruitment skills. This brings me to Christian, who is an absolute delight to work with, being a genuinely great and cheerful person. Laying the foundation for most my projects, I owe him big times for the rather smooth sailing of my PhD. I also appreciate my colleagues and ex-colleagues for a pleasant work atmosphere, which sometimes lead to very extended coffee breaks. I would also like to thank my friends and family for keeping me motivated (as they would also love to have a doctor in their circle) and always having time to give some quality distraction. Lastly, I would like to thank my girl Rana for staying with me through some partially stressful times, as it is her who brightens up my (work)days the most, showing me different aspects to life, teaching me new perspectives and of course making some great CAKE (as a team). Therefore, I will take my responsibility as a Doctor very seriously.

# *Table of Contents*

Abstract	ii
List of Publications	iii
Acknowledgments	iv
Table of Contents	v
List of Figures	vi
List of Tables	viii
Nomenclature	ix
1 Introduction	1
2 Theory	3
2.1 Molecular Dynamics Simulations	3
2.2 Modelling/Treatment of Interaction Forces	9
2.3 Monte-Carlo Simulation	16
2.4 Peptides	19
2.5 Analyses	21
3 Summary of the Published Papers	23
3.1 Secondary Structure Formation in Hybrid Synthetic/Peptide Polymers: Insights from Molecular Dynamics Simulations	23
3.2 Reverse mapping of coarse-grained polyglutamine conformations from PRIME20 Sampling	26
3.3 Assignment of a physical energy scale for the dimensionless interaction energies within the PRIME20 peptide model	28
4 Paper I: Secondary Structure Formation in Hybrid Synthetic/Peptide Polymers: Insights from Molecular Dynamics Simulations	30
5 Paper II: Reverse mapping of coarse grained polyglutamine conformations from PRIME20 sampling	39
6 Paper III: Assignment of a physical energy scale for the dimensionless interaction energies within the PRIME20 peptide model	51
7 Further Research	64
8 Conclusions	76
References	78

## List of Figures

- 2.1 Schematic illustration of the MD simulation procedure. 3
- 2.2 Visualization of the influence on the sampling of a system of high (red) and low temperature (blue). 7
- 2.3 Harmonic (red) and dihedral (blue) potential with regards to the arbitrary displacement ( $x$ ) from the equilibrium values. The displacement values are  $b$ ,  $\theta$ ,  $\omega$  and  $\delta$  for bonds, angles, improper and dihedrals, respectively. 10
- 2.4 Coulomb (blue) and Lennard-Jones Potentials for arbitrary distances. Notably, the Coulomb axis has a factor of  $10^{11}$ , showing that these interactions are of a stronger nature; and the Coulomb potential is shown for particle attraction (positive and negative charge), while for two of the same charge type the Coulomb force is strongly repulsive and therefore positive. 10
- 2.5 Geometry of the PRIME20 model is visualized with the backbone being represented by 3 beads: the NH group (green bead), the  $C_\alpha$  carbon (red bead) and the CO group (blue bead), while the side chain is represented by a fourth bead (gray bead). Each position and size is specific to the individual type of amino acid. (Taken from own Publ.[3]) 12
- 2.6 Visualization of the covalent (white) and pseudo (black and yellow) bonds used in the P20 model. The pseudo bonds ensure a stable and configurational real structure. (Taken from own Publ.[3]) 13
- 2.7 Example of the improved sampling efficiency for an arbitrary potential (circles) of the metropolis algorithm (red) compared to the random walk (black). 17
- 2.8 Overview of the reaction that forms peptides from amino acids. The amino- and acid-groups of the amino acid react to form a peptide bond. 19
- 2.9 Exemplary RDFs for a binary system with mixed (black) or separated (red) phases, depending on the temperature. 21
- 2.10 Exemplary combined distribution function (CDF). (Taken from own Publ.[1]) 21
- 2.11 Exemplary Ramachandran plot of an Asp-peptide with circles indicating different secondary structure regions. 22

- 3.1 Structure of hybrid polymers composed of a homopeptide of amino acids (AA=Asp, Lys, Glu), which is connected to a short polyethylene chain. The double bond which connects two such hybrid monomers is maintained during synthesis<sup>104–106</sup>. 23
- 3.2 Exemplary combined distribution function (CDF) of  $[PE - (Lys_{10})_2]$ , which keeps a peptide dominated structure. 24
- 3.3 Exemplary combined distribution function (CDF) of  $[PE - (Lys_3)_6]$ , which adopts a ball-like structure. 24
- 3.4 Scheme describing the conversion of the coarse-grained structures into all atom geometries. Red: atoms obtained from the coarse-grained PRIME20 model, green: atoms added by simple geometric considerations. 26
- 3.5 Visualization of the peptides at the start (a-c) and end (d-f) of the simulations. The dimer region, which is crucial for hydrogen bonding, is highlighted in green, while the less significant regions for dimer structure are marked in yellow. 26
- 3.6 The hydrogen bond data set is shown with the P20 energies compared to the resulting energies from a 10 ns MD simulation. 28
- 3.7 Correlation between the coarse-grained ( $E_{P20}$ ) and atomistic ( $E_{aa}$ ) energies for a series of conformations with identical backbone hydrogen bonding states (at the coarse-grained level, here corresponding to 22 and 17 hydrogen bonds, respectively, for the left and right plots). In turn, the number of sidechain interactions varies and corresponds to -1/12 units of  $E_{P20}$  per sidechain interaction. 29

## *List of Tables*

- 2.1 Bond and pseudo-bond lengths between beads of PolyQ in the PRIME20 model with index  $i$  representing beads of the  $(i)$ th residue and index  $i+1$  representing beads of the  $(i+1)$ th residue with Sizes in Å.<sup>73</sup> 12
- 2.2 Bead diameters and square-well parameters of PolyQ in PRIME20 with sizes in Å.<sup>73</sup> 13
- 2.3 Squeeze parameters and the resulting reduced bead diameters for backbone bead interactions and interactions involving a polyglutamine side chain are given. Sizes in Å.<sup>73</sup> 13
- 3.1 Secondary structure analysis of all hybrid polymers. Shown is the integral of characteristic regions in the Ramachandran plot, which correspond to typical secondary structure motifs. 24



## Nomenclature

<i>AA</i>	amino acid
<i>AIMD</i>	Ab Initio Molecular Dynamics
<i>arb.u.</i>	arbitrary units
<i>Asp</i>	aspartic acid
<i>CDF</i>	combined distribution function
<i>cg</i>	coarse-grained
<i>COM</i>	centre of mass
<i>Coul</i>	Coulomb
<i>ff</i>	force field
<i>Glu</i>	glutamic acid
<i>HREX</i>	Hamiltonian replica exchange
<i>hs</i>	hard-sphere
<i>Lys</i>	lysine
<i>MC</i>	Monte-Carlo
<i>MD</i>	Molecular Dynamics
<i>P20</i>	PRIME20
<i>PE</i>	poly-ethylene
<i>PolyQ</i>	poly-glutamine
<i>R<sub>G</sub></i>	radius of gyration
<i>RDF</i>	radial distribution function
<i>RESP</i>	Restrained Electrostatic Potential
<i>REST</i>	replica exchange solute tempering
<i>RMSD</i>	root-mean-square deviation
<i>SAMC</i>	stochastic approach Monte-Carlo
<i>SW</i>	square-well
<i>VDW</i>	Van-der-Waals
<i>VDW</i>	van-der-Waals

## Introduction

The European Union (EU) has acknowledged the necessity to combat climate change in its constitutional framework, as outlined in Article 191: "Union policy on the environment shall contribute to pursuit of the following objectives: reserving, protecting and improving the quality of the environment, protecting human health, prudent and rational utilisation of natural resources, promoting measures at international level to deal with regional or worldwide environmental problems, and in particular combating climate change."<sup>5</sup> This led to the formulation of more concrete goals (Summary of procedural developments<sup>6</sup> is given) including a reduction in greenhouse gas emissions by 55% until 2030 and the achievement of climate neutrality by 2050. This problem has various aspects, but the task is simple: to reduce energy consumption. This involves reducing energy consumption, increasing energy efficiency (e.g. through better insulation in houses), and in some cases, replacing a product entirely.

Two sectors with significant potential for environmental improvement are construction and packaging. Concrete is responsible for approximately 8.6% of all anthropogenic emissions<sup>7</sup>, making it imperative to identify alternative production methods. Bio-mineralisation and living building materials have the potential to significantly reduce the energy requirements of construction materials by utilising the cell's proteins, sugars, and lipids, among other substances.<sup>8</sup>

Plastics on basis of non-renewable resources are broadly used in packaging<sup>9</sup>, which causes littering and environmental problems globally.<sup>10</sup> The utilisation of eco-friendly packaging as an alternative to conventional materials is a promising avenue for addressing the aforementioned issues. However, further research is required to enhance the efficiency and cost-effectiveness of bio-plastics, particularly in the context of food packaging.<sup>10,11</sup>

Although the relationship between environmentally friendly materials and medical or diagnostic applications may not be immediately apparent, all peptide-containing materials undergo significant chemical alterations in response to their surrounding environment. Consequently, these materials rely on structural changes in the peptides, which are induced by various methods including a change in temperature, pH-value, solvents or additives.<sup>12-14</sup> The resulting properties can be employed in the engineering of biomaterials to align with a particular application; how-

ever, it is similarly possible to tailor drug delivery systems, particularly copolymers, to specific targets or locations.<sup>15–18</sup> In contrast, numerous neurodegenerative diseases are characterised by the occurrence of unwanted changes in proteins, which are caused by the surrounding environment and result in the loss of normal body function. In particular, Alzheimer's, Huntington's and Parkinson's diseases highlight the necessity for further research into the unwanted aggregation of peptides and proteins thought to be the primary cause of these diseases.<sup>19–21</sup> These aggregation processes, as well as their accompanying biochemical effects, are under intense research.<sup>22,23</sup> Computational methods offer valuable insights into specific aggregation processes; however, due to their limited scope, most theoretical methods provide limited understanding beyond a single step or isolated question.<sup>24</sup> The reasons for this are that the majority of methods have specific length and time scales that they can cover, which are specifically tied to the methods' accuracy. It is therefore advantageous to combine two or more methods in order to bridge the length and time scales involved. In order to implement this approach, it is necessary to convert the model descriptions to align with the specifications of other methods. Two commonly used methods to study biomolecules are molecular dynamics (MD) simulations and Monte Carlo (MC) simulations, which can be conducted with either force fields or coarse-grained models.<sup>25–30</sup> Coarse-grained MC simulations are capable of sampling a vast portion of the phase space, whereas force field MD simulations cover a more limited region with greater resolution and also provide insights into dynamics. The combination of these two methods addresses the representability and transferability issues inherent to the coarse-grained approach, while also addressing the local phase-space sampling limitations of classical MD simulations.<sup>31–37</sup> The combination of these techniques, which are highly complementary, has already enabled investigations to be conducted at larger system and time scales.<sup>38–44</sup> Furthermore, there are applications of these techniques, such as the Iterative Boltzmann Inversion<sup>34</sup> or the Inverse Monte Carlo approach<sup>35</sup>, which produce coarse-grained parameters fitted to MD simulation properties. These techniques have been successfully applied and improved over the years.<sup>45–52</sup>

## 2.1 Molecular Dynamics Simulations

Molecular dynamics (MD) simulations represent a computational tool employed to investigate the behaviour of molecular systems at the atomic and molecular level. The technique employs the principles of classical mechanics to numerically integrate the equations of motion of a system of interacting particles over time, as shown schematically in Fig. 2.1. Furthermore, this technique enables the interactions between molecules to be modeled in different ways for example, by simple force fields or even at the quantum electron level. This is one of the reasons why MD simulations are widely used in many fields of research.

### 2.1.1 Numerical Integration of the Equations of Motion

The Starting point to calculate the evolution of a system is Newton's second law, which describes the propagation of a particle  $I$  by its mass  $m_I$  and spatial coordinates  $\mathbf{R}_I(t)$  at time  $t$  in the following way

$$m_I \mathbf{R}_I(t) = \mathbf{F}_I(t) = -\nabla V(\mathbf{R}_1(t), \mathbf{R}_2(t), \dots, \mathbf{R}_N(t)). \quad (2.1)$$

The forces  $\mathbf{F}_I(t)$  acting on the particle are defined as the gradient of the potential  $V$ , which describes the potential energy of the system with respect to all spatial coordinates of  $N$  particles.

Given that MD simulations are inherently a many-body problem, the solution to Eq. 2.1 must be obtained through numerical means.<sup>53</sup> The integration is discretised in time steps of  $\Delta t$  using the finite difference method, during which the acting forces on the particle are assumed to be constant.

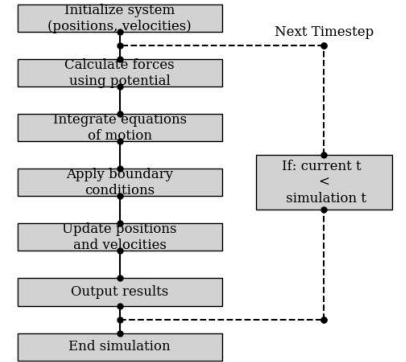
#### The Verlet algorithm

The Verlet algorithm<sup>54</sup> employs a Taylor expansion of the position  $\mathbf{R}_I$  at time  $t + \Delta t$  to derive the following result:

$$\mathbf{R}_I(t + \Delta t) = \mathbf{R}_I(t) + \Delta t \mathbf{R}_I(t) + \frac{1}{2} \Delta t^2 \mathbf{R}_I(t) + \frac{1}{6} \Delta t^3 \mathbf{R}_I(t) + O(t^{\geq 4}). \quad (2.2)$$

Doing the same for  $(t - \Delta t)$  results in:

$$\mathbf{R}_I(t - \Delta t) = \mathbf{R}_I(t) - \Delta t \mathbf{R}_I(t) + \frac{1}{2} \Delta t^2 \mathbf{R}_I(t) - \frac{1}{6} \Delta t^3 \mathbf{R}_I(t) + O(t^{\geq 4}), \quad (2.3)$$



**Figure 2.1:** Schematic illustration of the MD simulation procedure.

which can then be added to previous Eq. 2.2 to obtain after rearrangement for  $\mathbf{R}_I(t + \Delta t)$  following equation:

$$\mathbf{R}_I(t + \Delta t) = 2\mathbf{R}_I(t) - \mathbf{R}_I(t - \Delta t) + \Delta t^2 \mathbf{R}_I(t) + O(t^{\geq 4}). \quad (2.4)$$

This approach results in an equation, where the term for  $\ddot{\mathbf{R}}_I(t)$  vanishes. Accordingly, the use of terms up to the second derivative has the same accuracy as the use of all terms up to the third derivative, while neglecting everything of forth and higher order ( $O(t^{\geq 4})$ ).

The velocities  $\mathbf{v}_I$  are not explicitly given by the Eq. 2.4 but can be calculated in the following way:

$$\mathbf{v}(t) = \frac{\mathbf{R}(t + \Delta t) - \mathbf{R}(t - \Delta t)}{2\Delta t}. \quad (2.5)$$

The application of Eq. 2.5 to the initial conditions at time  $t = 0$  results in the inability to calculate the velocities. This presents a challenge when attempting to analyse the system for the conservation of energy. Furthermore, following the propagation of errors, the difference of two substantial 1st order terms employed in Eq. 2.4 results in a considerable relative error, which impairs the precision of the simulation.<sup>53,55</sup>

### The Velocity-Verlet algorithm

The Velocity-Verlet algorithm<sup>56</sup> improves its previous version by explicitly calculating position and velocities with a simple approach. Starting point is Eq. 2.3 where we substitute  $t$  by  $t + \Delta t$ , which as a result gives:

$$\mathbf{R}_I(t) = \mathbf{R}_I(t + \Delta t) - \Delta t \mathbf{R}_I(t + \Delta t) + \frac{1}{2} \Delta t^2 \mathbf{R}_I(t + \Delta t) - \frac{1}{6} \Delta t^3 \mathbf{R}_I(t) + O(t^{\geq 4}). \quad (2.6)$$

This equation is now added to Eq. 2.2, ignoring the fourth and higher order terms ( $O(t^{\geq 4})$ ), to obtain the velocity of particle  $I$  with the following term

$$\mathbf{v}_I(t + \Delta t) = \mathbf{v}_I(t) + \frac{\Delta t}{2} [\mathbf{a}_I(t) + \mathbf{a}_I(t + \Delta t)]. \quad (2.7)$$

The positions can be calculated afterwards with the obtained velocities by

$$\mathbf{R}_I(t + \Delta t) = \mathbf{R}_I(t) + \Delta t \mathbf{v}_I(t) + \frac{1}{2} \Delta t^2 \mathbf{a}_I(t). \quad (2.8)$$

This method addresses the shortcomings of the original Verlet algorithm, rendering it a superior alternative for MD simulations.

### Leap-Frog Algorithm

A different approach to improve the Verlet algorithm calculated the positions and velocities at different points in time. The leap-frog algorithm<sup>57</sup> updates the positions by:

$$\mathbf{R}_I(t + \Delta t) = \mathbf{R}_I(t) + \Delta t \mathbf{v}_I(t + \frac{1}{2} \Delta t), \quad (2.9)$$

and the velocities by:

$$\mathbf{v}_I(t + \frac{1}{2}\Delta t) = \mathbf{v}_I(t - \frac{1}{2}\Delta t) + \Delta t \mathbf{a}(t). \quad (2.10)$$

In this approach, the velocities  $\mathbf{v}_I(t + \frac{1}{2}\Delta t)$  are first calculated using Eq. 2.10 from the velocities at an earlier point in time  $t - \frac{1}{2}\Delta t$  and the acceleration of time  $t$ . Then the positions  $\mathbf{R}(t + \Delta t)$  can be updated with the just calculated velocities  $\mathbf{v}_I(t + \frac{1}{2}\Delta t)$  and the positions  $\mathbf{R}(t)$  from time  $t$  using Eq. 2.9. In order to also calculate the velocities at a synchronized time with the positions to obtain the kinetic energy contributions of the total energy, the velocities at time  $t$  can simply be calculated by:

$$\mathbf{v}_I(t) = \frac{[\mathbf{v}(t + \frac{\Delta t}{2}) + \mathbf{v}(t - \frac{\Delta t}{2})]}{2}. \quad (2.11)$$

### 2.1.2 Thermostats

Using the previously shown algorithms in Sec. 2.1.1 for a simulation of fixed number of particles  $N$  within a fixed volume  $V$ , result in the energy being a constant of motion. The ensemble is commonly referred to as the NVE ensemble. However, in most cases, it is advantageous to have a constant temperature (NVT ensemble), which is closer to experimental conditions. In simulations, this is achieved by coupling the system to a thermostat.

#### Simple Velocity Rescaling

The velocity rescaling is simplest method for a thermostat. The idea is to scale the current velocities of atom until the target temperature is achieved. The average kinetic energy  $\langle K \rangle$  is directly proportional to the temperature  $T$ :

$$\langle K \rangle = \frac{3}{2} N_f k_b T, \quad (2.12)$$

with  $N_f$  being the degrees of freedom.

The mean kinetic energy  $\langle K \rangle$  can be expressed by the mean mass and velocity  $\langle m\mathbf{v}^2 \rangle$  of all particles:

$$\langle K \rangle = \frac{1}{2} \langle m\mathbf{v}^2 \rangle. \quad (2.13)$$

The temperature of the system can be expressed by combining both Eq. 2.12 and 2.13 to obtain an expression for the temperature dependent on the mean velocity:

$$T = \frac{1}{3k_b} \langle m\mathbf{v}^2 \rangle. \quad (2.14)$$

The temperature difference  $\Delta T$  that the thermostat must to compensate for in a time step  $\delta t$  is given by:

$$\Delta T = T(t + \delta t) - T(t). \quad (2.15)$$

The velocity at time  $t + \delta t$  can be expressed by the original velocity  $\mathbf{v}(t)$  with a scaling factor  $\lambda$

$$\mathbf{v}(t + \delta t) = \lambda \mathbf{v}(t). \quad (2.16)$$

Using this expression for Eq. 2.15 and 2.14 we obtain:

$$\Delta T = \frac{m}{3k_b} [\lambda^2 \langle \mathbf{v}^2(t) \rangle - \langle \mathbf{v}^2(t) \rangle] \quad (2.17)$$

$$\Delta T = \frac{m \langle \mathbf{v}^2(t) \rangle}{3k_b} [\lambda^2 - 1]. \quad (2.18)$$

Here we can shorten the equation by using Eq. 2.14, which results into:

$$\Delta T = T(t) [\lambda^2 - 1]. \quad (2.19)$$

Using the starting temperature  $T_0$  for  $T(t + \delta t)$  results in

$$\lambda = \sqrt{\frac{T_0 - T(t)}{T(t)} + 1} \quad (2.20)$$

and shortened to

$$\lambda = \sqrt{\frac{T_0}{T(t)}}. \quad (2.21)$$

This temperature constraining method is a straightforward and expedient approach to implementing a thermostat. However, the velocity scaling within a given time step does not permit temperature fluctuations. This behaviour is far from optimal for simulating experimental conditions.

### Berendsen Thermostat

The Berendsen thermostat<sup>58</sup> uses weak coupling to scale the velocities during each time step with a specific rate of temperature ( $\frac{dT}{dt}$ ) between the instantaneous temperature  $T(t)$  and a bath temperature  $T_0$ :

$$\frac{dT}{dt} = \frac{1}{\tau_c} (T_0 - T(t)), \quad (2.22)$$

with coupling parameter  $\tau_c$  adjusting the coupling strength.

The temperature difference in a discrete form between two time steps is given by:

$$\Delta T = \frac{\delta t}{\tau_c} (T_0 - T(t)). \quad (2.23)$$

Using this term of  $\Delta T$  instead of  $[T_0 - T(t)]$  in Eq. 2.20 gives the scaling factor in the form of:

$$\lambda = \sqrt{\frac{\delta t}{\tau_c} \left( \frac{T_0}{T(t)} - 1 \right) + 1} \quad (2.24)$$

It is of utmost importance to select the coupling parameter with great care. If the coupling parameter is set too high, the thermostat has a

negligible effect on the simulation. Conversely, if the coupling parameter is set too low ( $\tau_c = \delta t$ ), the thermostat effectively becomes the velocity rescaling thermostat, which has well-documented disadvantages. The Berendsen thermostat is characterised by stability and simplicity of implementation, although it lacks a conserved quantity. The advanced velocity rescaling algorithm operates in a manner analogous to the Berendsen thermostat, employing a suitably selected random factor to ensure the correct distribution of kinetic energy, conserved energy, and ergodicity.<sup>59</sup>

### 2.1.3 Enhanced Sampling

Classical MD simulations struggle to sample the entire phase space due to the confinement to low energy regions in conformational space.<sup>60</sup> To overcome this problem, improved sampling techniques have been developed and applied to study large proteins in particular.<sup>61</sup> The most common techniques are replica exchange<sup>62</sup>, metadynamics<sup>63</sup> and simulated annealing<sup>64</sup>, which have been developed further over time for specific purposes.

#### Replica Exchange

The fundamental principle underlying the replica exchange method is the sampling of the system of interest at multiple temperatures.<sup>65</sup> Due to the elevated temperature, a greater proportion of the phase space is sampled. Subsequently, conformations that are energetically favourable are transferred to lower temperatures until they reach the desired simulation temperature. The probability of exchanging two neighbouring replicas with configurations  $x_1$  and  $x_2$  with their respective potential energies  $U$  is given by:

$$p(x_1 \leftrightarrow x_2) = \min\{1, \exp[(\beta_1 - \beta_2)(U(x_1) - U(x_2))]\}. \quad (2.25)$$

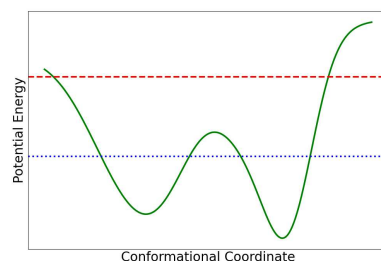
Here,  $x_2$  is the configuration in the higher temperature replica and the temperature influence is represented by  $\beta$  in form of:

$$\beta = 1/RT. \quad (2.26)$$

One disadvantage of this approach is that temperature is an extensive property, which affects both the protein and the large amounts of water that act as a solvent in the system.

#### Hamiltonian Replica Exchange

The Hamiltonian replica exchange (HREX) circumvents this problem by scaling the force field parameters instead of the temperature. The idea is that at elevated temperatures, the protein moves more freely because it has sufficient kinetic energy to overcome barriers. A similar effect is achieved by scaling down the interactions of the protein, which subsequently also allows greater mobility. The exchange probability can be calculated by the difference in potential energies of the two



**Figure 2.2:** Visualization of the influence on the sampling of a system of high (red) and low temperature (blue).



configurations  $x_1$  and  $x_2$  for both force field parameters:

$$p(x_1 \leftrightarrow x_2) = \min\left\{1, \exp\left(-\frac{[U_1(x_2) - U_1(x_1)] - [U_2(x_2) - U_2(x_1)]}{RT}\right)\right\}. \quad (2.27)$$

This generalised approach provides an avenue for further enhancements. It is possible to scale the force field parameters simultaneously with the temperature, in a manner analogous to the original replica exchange approach. It is also possible to focus on specific structural features. For instance, when a protein is solved in water, it is possible to scale only the parameters involving the protein part, thereby ignoring the vast quantities of water-water interactions that are not of interest. This approach is called replica exchange solute tempering (REST). The advanced REST2 technique, employs specific scaling for the protein-protein potential energy  $U_{PP}$  and the protein-water energy  $U_{PW}$ . This results in an acceptance ratio between neighbouring replicas of:

$$p(x_1 \leftrightarrow x_2) = \min\left\{1, (\beta_1 - \beta_2) \left[ (U_{PP}(x_2) - U_{PP}(x_1)) + \frac{\sqrt{\beta_0}}{\sqrt{\beta_1} + \sqrt{\beta_2}} (U_{PW}(x_2) - U_{PW}(x_1)) \right] \right\}. \quad (2.28)$$

The application of these techniques results in a significant reduction in the sampling time and an increase in the convergence rate, without an increase in the computational cost. The principal disadvantage of this method is the loss of dynamic properties due to the rapid change of configurations.

## 2.2 Modelling/Treatment of Interaction Forces

In computational chemistry there are different techniques, that can be used depending on the important characteristics of the system. These methods can vary drastically in computational cost and accuracy, making the correct choice crucial. Among the most commonly used methods, the quantum mechanical treatment of interaction forces is the most accurate, and is usually implemented using density functional theory. The quantum mechanical simulation is called Ab Initio Molecular Dynamics (AIMD) and includes the treatment of the electronic degrees of freedom, which comes at an immense computational cost. This treatment is for example necessary for the calculation of UV spectra, chemical reactions, including proton transfers, or spin related properties, among other applications.<sup>2,66–68</sup> Although being very accurate, due to the required computational effort there are severe limitations on the size of the system, making it unsuitable for the simulation of large biomolecules and polymers.

### 2.2.1 Classical Force Fields

The classical force fields refer to all-atom treatment of the system, this however means losing the direct effects originating from electron interactions. The potential energy is described by simple categorized potentials for intra-molecular interactions like bonds and angles, as well as inter-molecular interactions mainly van-der-Waals (VDW) and electrostatic Coulomb (Coul) interactions. In this work the CHARMM27<sup>69</sup> force field is shown as an example for this approach and also highlight some difficulties resulting from the this simpler treatment of interactions.

The total potential energy  $U_{CHARMM}$  in this force field is calculated by:

$$U_{CHARMM} = U_{bond} + U_{angle} + U_{dihedral} + U_{improp} + U_{UB} + U_{CMAP} + U_{Coul} + U_{VDW}. \quad (2.29)$$

Under the harmonic approximation, the potential energy terms for bond  $U_{bond}$ , bond angle  $U_{angle}$ , and improper torsion angle  $U_{improp}$  can be calculated as a sum of individual contributions. These calculations involve a force constant ( $k_x$ ) and consider the displacement from the equilibrium values of bond length ( $b_0$ ), bond angle ( $\theta_0$ ), or improper angle ( $\omega_0$ ).

The bond energy term, which accounts for the stretching or compression of individual bonds is given by:

$$U_{bond} = \sum_{bonds} k_b(b - b_0)^2, \quad (2.30)$$

the bond angle energy term, which captures the bending or deviation from the equilibrium angles between connected atoms is given by:

$$U_{angle} = \sum_{angles} k_\theta(\theta - \theta_0)^2, \quad (2.31)$$

and the improper energy term, which is responsible for maintaining e.g. chirality and planarity for specific atoms such as a carbonyl carbon, is given by:

$$U_{\text{improp}} = \sum_{\text{improp}} k_{\omega}(\omega - \omega_0)^2. \quad (2.32)$$

The dihedral angle potential  $U_{\text{dihedral}}$  is described for several periodicities  $n$  and with the absolute value of the force constant  $k_{\phi}$  due to the possibility of being negative for instance at *cis* conformations and can be calculated by:

$$U_{\text{dihedral}} = \sum_{\text{dihedrals}} k_{\phi}(1 + k_{\phi}\cos(n\phi - \delta)). \quad (2.33)$$

The Urey-Bradley  $U_{\text{UB}}$ , which is similar to  $U_{\text{angle}}$ , is a more refined potential, that helps describing in plane deformations and several symmetric and asymmetric bond stretching<sup>70</sup> by adding a harmonic 1-3 potential, referring to atom 1 and 3 of some angles, and is given by:

$$U_{\text{UB}} = \sum_{\text{Urey-Bradley}} k_{\text{UB}}(b_{1-3} - b_{1-3;0})^2. \quad (2.34)$$

The  $U_{\text{CMAP}}$  potential provides parameters, that adjust protein backbones with adapted cross term dihedral angle values in the following way:<sup>71</sup>

$$U_{\text{CMAP}} = \sum_{\text{residues}} u_{\text{CMAP}}(\Phi, \Psi). \quad (2.35)$$

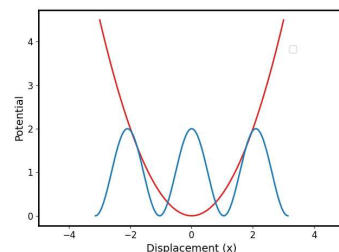
All enumerated potential were part of the bonded interactions  $U_{\text{bonded}}$ :

$$U_{\text{bonded}} = U_{\text{bond}} + U_{\text{angle}} + U_{\text{UB}} + U_{\text{dihedral}} + U_{\text{improper}} + U_{\text{CMAP}}. \quad (2.36)$$

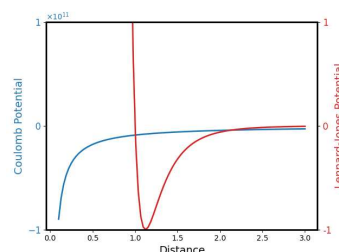
The non-bonded interactions ( $U_{\text{non-bonded}}$ ) in the system are described by two potentials: the electrostatic potential ( $U_{\text{elec}}$ ) and the Lennard-Jones potential ( $U_{\text{LJ}}$ ), which incorporates the standard 12-6 potential. The total non-bonded interaction energy is therefore just given by:

$$U_{\text{non-bonded}} = U_{\text{elec}} + U_{\text{LJ}}. \quad (2.37)$$

The Lennard-Jones potential, represented by Eq. 2.38, accounts for the attractive and repulsive forces between non-bonded pairs. It is calculated by summing over all non-bonded pairs in the system. The terms in the



**Figure 2.3:** Harmonic (red) and dihedral (blue) potential with regards to the arbitrary displacement ( $x$ ) from the equilibrium values. The displacement values are  $b$ ,  $\theta$ ,  $\omega$  and  $\delta$  for bonds, angles, improper and dihedrals, respectively.



**Figure 2.4:** Coulomb (blue) and Lennard-Jones Potentials for arbitrary distances. Notably, the Coulomb axis has a factor of  $10^{11}$ , showing that these interactions are of a stronger nature; and the Coulomb potential is shown for particle attraction (positive and negative charge), while for two of the same charge type the Coulomb force is strongly repulsive and therefore positive.

square brackets capture the interplay between the energetically optimal distance ( $r_{ij}^{\min}$ ) and the distance ( $r_{ij}$ ) between the pairs. The first term represents the repulsive interaction, and the second term represents the attractive interaction.

The electrostatic potential, described by Eq. 2.39, accounts for the long-range electrostatic interactions between charged particles. It is calculated by summing over all non-bonded pairs in the system, which involves the charges ( $q_i$  and  $q_j$ ) of the particles and the distance ( $r_{ij}$ ) between them, taking into account the permittivity of vacuum ( $\epsilon_0$ ).

$$U_{LJ} = \sum_{\text{nonb.pairs}} \epsilon_{ij} \left[ \left( \frac{r_{ij}^{\min}}{r_{ij}} \right)^{12} - 2 \left( \frac{r_{ij}^{\min}}{r_{ij}} \right)^6 \right] \quad (2.38)$$

$$U_{\text{elec}} = \sum_{\text{nonb.pairs}} \frac{q_i q_j}{\epsilon_0 r_{ij}} \quad (2.39)$$

It is important to note, that there is a distinction between the symbol  $\epsilon$ . One represents the permittivity of vacuum, denoted as  $\epsilon_0$ , while the other refers to the Lennard-Jones parameter, denoted as  $\epsilon_{ij}$ .

In this particular force field, which was optimized using a mixed arithmetic combination rule, the Lennard-Jones parameters  $\epsilon_{ij}$  and  $r_{ij}^{\min}$  are calculated in the following manner:

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j} \quad (2.40)$$

and

$$r_{ij}^{\min} = \frac{1}{2}(r_i^{\min} + r_j^{\min}). \quad (2.41)$$

In summary, the use of these simple descriptions for interaction forces makes this method potent tool to investigate large systems with an all-atom level of accuracy, however, due to the simplicity of bond interactions, it can typically not be used for simulating bond breaking or electronic excitation.

## 2.2.2 Coarse-Grained Models

Coarse-Grained models increase the speed of calculation by reducing the degrees of freedom. This includes combining multiple atoms into a single one, which removes several rotational and vibrational modes. The united-atom approach makes use of this by e.g. combining carbon atoms with the bonded hydrogen atoms. In comparison, conventional coarse-grained models use this method to a greater extent by combining e.g. whole polymer chain into a single entity. These entities are called beads instead, due to the sheer size they can no longer be recognized as an atom. In addition, coarse-grained models can change the interactions

they describe by reducing the complexity of some interactions or by combining several potentials into a single or simpler one. ...

### PRIME20 Model

The PRIME20<sup>72</sup> model was developed to study the aggregation behaviour of simple homoproteins. By employing square-well as well as hard-sphere potentials and combining multiple amino acid atoms into a single bead, the computational effort was drastically reduced.

In this model an amino acid is represented by 4 beads including 3 backbone and 1 sidechain bead. The sidechain bead R accounts for specific sidechain interactions and is positioned at the center of mass of the sidechain. The backbone comprises NH, C<sub>α</sub> and a CO bead, corresponding to the N, C and C<sub>α</sub> atoms, respectively. Given the focus on the aggregation of poly-glutamine peptides in this work, we will discuss only the relevant parameters.

The covalent bonds are represented by infinite well potentials centered around an ideal bond length  $d_{ideal}$ , while the well width accommodates for bond length fluctuations  $\Delta$  of around 2.375%. The bond potential is then given by:

$$V_{bond}(d) = \begin{cases} 0 & \text{if } d \in [d_{ideal} - \Delta, d_{ideal} + \Delta] \\ \infty & \text{otherwise} \end{cases} \quad (2.42)$$

Additionally, pseudo-bonds are implemented to maintain a trans configuration ensuring stable bonds and angles between neighbouring C<sub>α</sub> beads. These pseudo-bonds are treated similarly to covalent bonds. The ideal bond lengths for both types are summarized in Tab. 2.1.

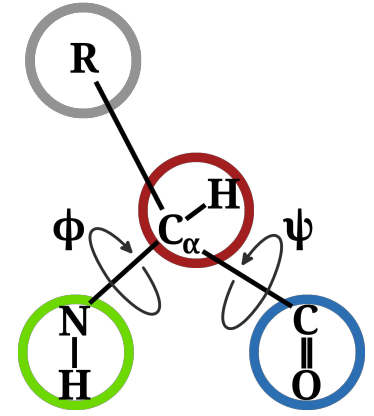
**Table 2.1:** Bond and pseudo-bond lengths between beads of PolyQ in the PRIME20 model with index  $i$  representing beads of the  $(i)$ th residue and index  $i+1$  representing beads of the  $(i+1)$ th residue with Sizes in Å.<sup>73</sup>

Bonds	NH <sub><i>i</i></sub> -C <sub><math>\alpha</math>,<i>i</i></sub> 1.46	C <sub><math>\alpha</math>,<i>i</i></sub> -CO <sub><i>i</i></sub> 1.51	CO <sub><i>i</i></sub> -NH <sub><i>i</i>+1</sub> 1.33	R <sub><i>i</i></sub> -C <sub><math>\alpha</math>,<i>i</i></sub> 1.60		
Pseudo-bonds	NH <sub><i>i</i></sub> -CO <sub><i>i</i></sub> 2.45	C <sub><math>\alpha</math>,<i>i</i></sub> -NH <sub><i>i</i>+1</sub> 2.41	CO <sub><i>i</i></sub> -C <sub><math>\alpha</math>,<i>i</i>+1</sub> 2.45	NH <sub><i>i</i></sub> -R <sub><i>i</i></sub> 2.50	C <sub><math>\alpha</math>,<i>i</i></sub> -C <sub><math>\alpha</math>,<i>i</i>+1</sub> 3.80	CO <sub><i>i</i></sub> -R <sub><i>i</i></sub> 2.56

The non-bonded interactions of peptides primarily involve of hydrogen bonding and sidechain interactions.

The steric effects of sidechains are implemented using a repulsive hard-sphere (HS) model to simulate sidechain-sidechain, sidechain-backbone and excluded volume interactions. Hydrogen bonding in the backbone and hydrophobic interactions between sidechain beads are treated as semi-finite square-well (SW) potentials in the form of:

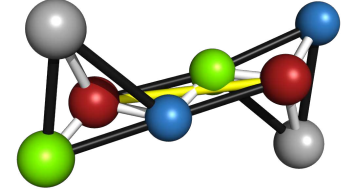
$$V_{HS}(d_{ij}) = \begin{cases} 0 & \text{if } d_{ij} > d_{ij}^{HS} \\ \infty & \text{otherwise} \end{cases} \quad (2.43)$$



**Figure 2.5:** Geometry of the PRIME20 model is visualized with the backbone being represented by 3 beads: the NH group (green bead), the C<sub>α</sub> carbon (red bead) and the CO group (blue bead), while the side chain is represented by a fourth bead (gray bead). Each position and size is specific to the individual type of amino acid. (Taken from own Publ.[3] )

$$V_{\text{SW}}(d_{ij}) = \begin{cases} 0 & \text{if } d_{ij} > d_{ij}^{\text{SW}} \\ \epsilon_{ij} & \text{if } d_{ij}^{\text{HS}} < d_{ij} < d_{ij}^{\text{SW}} \\ \infty & \text{otherwise} \end{cases} \quad (2.44)$$

Here,  $d_{ij}$  represents the distance between two beads  $i$  and  $j$ ,  $d_{ij}^{\text{HS}}$  being the hard-sphere diameter,  $\epsilon_{ij}$  the square-well depth and  $d_{ij}^{\text{SW}}$  the square-well interaction distance. Specific parameters for square-well potentials vary for different pairs of interacting side chains. The Lorentz-Berthelot combining rule was used to calculate  $d_{ij}^{\text{HS}}$  from the bead  $d^{\text{HS}}$ . The self-interaction diameter of the sidechains is also used for the sidechain-backbone interactions, which interaction values are summarized in Tab. 2.2. Additionally, hydrogen bonding is defined by a distance parameter of  $d_{ij}^{\text{SW}} = 4.5 \text{ \AA}$ , along with angle constrain between the N-H and C-O vectors, and each bead is restricted to forming one hydrogen bond.



**Figure 2.6:** Visualization of the covalent (white) and pseudo (black and yellow) bonds used in the P20 model. The pseudo bonds ensure a stable and configurational real structure. (Taken from own Publ.[3] )

**Table 2.2:** Bead diameters and square-well parameters of PolyQ in PRIME20 with sizes in  $\text{\AA}$ .<sup>73</sup>

	NH	$C_\alpha$	CO	R
$d^{\text{HS}}$	3.3	3.7	4.0	3.6
$d^{\text{SW}}$	4.5	-	4.5	6.6
$\epsilon$	-1.000	-	-1.000	-0.080

To replicate the formation of specific natural protein structures, squeeze parameters are incorporated to decrease the effective bead diameters in close proximity. The sidechain parameters are unique for each amino acid, and all parameters for PolyQ are summarized in Tab. 2.3.

**Table 2.3:** Squeeze parameters and the resulting reduced bead diameters for backbone bead interactions and interactions involving a polyglutamine side chain are given. Sizes in  $\text{\AA}$ .<sup>73</sup>

Interactions	$C_{\alpha,i}-\text{CO}_{i+1}$	$C_{\alpha,i}-\text{NH}_{i-1}$	$\text{CO}_i-\text{NH}_{i+2}$	$\text{NH}_i-\text{NH}_{i+1}$	$\text{CO}_i-\text{CO}_{i+1}$
original $d$	3.85	3.50	3.65	3.30	4.00
squeeze factor	1.1436	0.88	0.87829	0.8	0.7713
squeezed $d$	4.40286	3.08	3.2057585	2.64	3.0852
Interactions	$C_{\alpha,i-1}-R_i$	$\text{CO}_{i-1}-R_i$	$\text{NH}_{i+1}-R_i$	$C_{\alpha,i+1}-R_i$	$\text{CO}_{i-2}-R_i$
original $d$	3.65	3.8	3.45	3.65	3.8
squeeze factor	1.407	1.089	1.158	1.387	1.316
squeezed $d$	5.134	4.139	3.996	5.062	5.000

### 2.2.3 Obtaining Force Field Parameters

As force fields are less accurate interaction potentials to be able to simulate large systems, the force fields are usually optimized to target properties of specific systems. This can range from simulating solid, liquids or

gases but also optimizing parameters to fit density, heat capacity or ion mobility, among others. For the special cases of proteins there are also specific parameters added to obtain equilibrated structures, which are close to the experimentally obtained Protein Databank entries.

### Bonded Interactions

The bonded interaction potentials are typically derived from DFT or even more sophisticated calculations. To illustrate, a quantum mechanical calculation is performed for a peptide backbone dihedral potential, whereby only one specific dihedral angle is varied at a time and the resulting energy difference upon change is analysed. Subsequently, the obtained energy plot is fitted to the force field's specific dihedral potential.<sup>71</sup> It is also common to use already established force fields and adjust the parameters to fit specific properties usually obtained experimentally.<sup>74</sup>

### Atomic Partial Charges

In some cases, it is sufficient to fit the atomic partial charges to specific system properties, such as heats of vaporization or sublimation, liquid densities, or gas-phase dipole moments.<sup>75,76</sup> Otherwise, there are alternative methods for obtaining atomic charges from DFT calculations or experiments. A comprehensive overview of a broad spectrum of methods can be accessed<sup>77</sup>.

The Mulliken<sup>78</sup> or Löwdin<sup>79</sup> population analysis partitions the molecular wave function with an arbitrary orbital scheme. The Mulliken charge  $q_A^{Mul}$  is calculated using the following formula:

$$q_A^{Mul} = Z_A - \sum_{i,j \in A} P_{ij} S_{ji} \quad (2.45)$$

where  $Z_A$  is the nuclear charge,  $P_{ij}$  the one-electron matrix elements and  $S_{ji}$  the overlap matrix with basis functions  $i$  and  $j$ , where  $j$  is belonging to atom A.

The Hirshfeld<sup>80</sup> charge analyses allocates the total electron density distribution of molecules or crystals into overlapping and non-overlapping contributions of each atom. The overlap density  $\rho_{pro}$  is the sum of all densities contributing to atom A, shown in Eq. 2.46.

$$\rho_{pro} = \sum_A \rho_A \quad (2.46)$$

A weight function  $w_A$  assigns the total electron density  $\rho$  to the atoms in the following way:

$$w_A = \frac{\rho_A}{\rho_{pro}} \quad (2.47)$$

The Hirshfeld atomic partial charges are then obtained by:

$$q_A^{Hir} = Z_A - \int dV w_A \rho = Z_A - \int dV \frac{\rho_A}{\sum_A \rho_A} \rho \quad (2.48)$$

The (Restrained) Electrostatic Potential<sup>81</sup> (ESP or RESP) is a method that tries to reproduce the electrostatic Potential  $V^{ESP}(\mathbf{r})$ . The minimum of the difference obtained by a least squares fitting for grid points  $\mathbf{r}_k$  for a selected number of fit points  $N$ , provides the value for the ESP charge  $q^{ESP}$ , demonstrated in Eq. 2.49

$$\min\{q_A^{ESP}\} = \frac{1}{N} \sum_k^N [V_{QM}(\mathbf{r}_k) - V_{ESP}(\mathbf{r}_k)]^2 \quad (2.49)$$

For non periodic conditions the potential is calculated in the following way:

$$V_{ESP}(\mathbf{r}_k) = \sum_A \frac{q_A^{ESP}}{|\mathbf{R}_A - \mathbf{r}_k|}. \quad (2.50)$$

It is possible to introduce restraints  $R_{rest}$  in order to prevent the generation of unphysical values in the form of harmonic penalty functions. An example of such a function is:

$$R_{rest} = \beta \sum_m (q_m - t_m)^2, \quad (2.51)$$

where  $t_m$  represents the target value for charge  $q_m$  and  $\beta$  is the strength of the restraint.



## 2.3 Monte-Carlo Simulation

Monte-Carlo (MC) simulations or MC method is a broadly applicable technique to study problems in a statistical manner. It is applied in industrial engineering, structure prediction, computational statistics, risk analysis in economics or finance and more.<sup>82</sup>

In essence, Monte Carlo simulation is a statistical technique used to accurately study various quantities in complex systems by employing random sampling from probability distributions. Here, we focus on the field of statistical physics and chemistry.

### 2.3.1 Statistical Background

The central objective of Monte Carlo simulation is to compute the average value of a physical quantity, denoted as  $\langle A \rangle$ , within a given system. This average is determined by sampling a set of configurations  $x$  from a statistical ensemble, given by:

$$\langle A \rangle = \sum_x A(x)p(x). \quad (2.52)$$

To achieve this, a statistical weight function  $p(x)$  is employed, which quantifies the likelihood of observing a specific configuration within the ensemble. An easily deployable example of this function is the Boltzmann weight, which incorporates parameters like temperature  $T$  and the Boltzmann constant  $k_b$  to describe the probability distribution of configurations, which can be calculated by:

$$p(x, T) = \frac{e^{-H(x)/k_b T}}{Z(T)}. \quad (2.53)$$

It is worth noting, that in theory all sums theoretically are integrals, but for simplicity and applicability (numerical integration techniques) sums are used in this notation. In addition, the configurations  $x$  in theory are the vector with all degrees of freedom, which in theoretical chemistry is usually considered to be the configurations.

The Hamiltonian function  $H(x)$  plays a pivotal role in this context, calculating the total energy of a system for a given configuration. By considering the Hamiltonian, we can accurately estimate the probability of observing a particular configuration.

The partition function  $Z(T)$ , which acts as a normalization factor, represents the summation of the Boltzmann factors over all possible configurations of the system:

$$Z(T) = \sum_x e^{-H(x)/k_b T}. \quad (2.54)$$

The sampled quantity in our subset  $\langle A \rangle'$  tends towards the actual ensemble average  $\langle A \rangle$  as the subset approaches infinity, according to the law of large numbers.

$$\langle A \rangle' = \sum_{t=1}^n A(x_t) p(x_t) \xrightarrow{n \rightarrow \infty} \langle A \rangle. \quad (2.55)$$

The generation of the subset could also be done by MD simulation, where acting forces are calculated (see Sec. 2.1.1) to achieve a physically realistic behaviour of the system. However, this realistic motion of the system is computationally expensive because the whole space is sampled inefficiently. In comparison, MC simulations displace the particles randomly, resulting in a more efficient sampling, but at the expense of obtaining general dynamic properties.

### 2.3.2 Metropolis Sampling

The simplest sampling method uses uniformly-created random values for  $x$  and calculates the average quantity according to (2.55). The randomness is generated by a random number generator algorithm, which is critical for this method to have a good phase space sampling. However, due to the high dimensionality of  $x$  [each atom or coarse-grained bead each in  $x$ ,  $y$  and  $z$  direction], a lot of sampling is done with configurations with probability  $p \approx 0$ , which contribute very little to the desired average.

Importance sampling is a more refined method, that focuses more on the relevant points in space, while keeping a probability distribution similar to  $p(x)$ . In order to sample high importance regions in the space, the  $Z(T)$  has to be known, see Eq. 2.53. However this property cannot be calculated *a priori*, therefore a new technique is introduced called Markov Chains or Metropolis algorithm.<sup>83</sup> The underlying principle is based on comparing the probability of the coming step to the probability of the previous step, resulting in a canceling of the unknown function  $Z(T)$ :

$$\frac{p(x_1, T)}{p(x_2, T)} = \frac{e^{-H(x_1)/k_b T}}{Z(T)} \frac{Z(T)}{e^{-H(x_2)/k_b T}} = \frac{e^{-H(x_1)/k_b T}}{e^{-H(x_2)/k_b T}}. \quad (2.56)$$

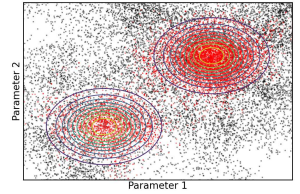
The process originates from the Master equation, which sums up the probability flows of states  $(x_i, x_j)$  for one simulation step  $(t - 1 \rightarrow t)$ , while assuming conservation of energy. This results in the probability of the state  $x_i$  being calculated by the sum of the probability of the previous state being  $x_i$  and not changing, and the change of  $x_j$  in the previous step changing to  $x_i$ :

$$p_t(x_i) = p_{t-1}x_i - \sum_{j \neq i} p_{t-1}(x_i) \omega(x_j|x_i) + \sum_{j \neq i} p_{t-1}(x_j) \omega(x_i|x_j). \quad (2.57)$$

Here,  $\omega(x_i|x_j)$  denotes the conditional transition probability to change states from  $x_j$  to  $x_i$ .

### 2.3.3 Stochastic Approximation Monte Carlo

The disadvantage of the previous Metropolis algorithm is the possibility of getting trapped in local energy minima when the landscape has multiple distinct maxima.<sup>84</sup> In addition, the algorithm relies on the



**Figure 2.7:** Example of the improved sampling efficiency for an arbitrary potential (circles) of the metropolis algorithm (red) compared to the random walk (black).

cancellation of the partition function, but knowledge of it is valuable as thermodynamic properties can be directly derived from it. The flat histogram approach, as the name suggests, attempts to obtain a uniform flat histogram of visited states. To achieve this, an approximation scheme is used where the acceptance probability depends on the current state of the probability distribution. This is necessary because the probability weights are not known *a priori*. The implementation of this idea for various situations lead to multiple successful algorithms: Umbrella sampling<sup>85</sup>, Multicanonical MC<sup>86</sup>, Well-tempered ensemble metadynamics<sup>87</sup>, Wang-Landau<sup>88</sup>, or SAMC<sup>84</sup>.

The SAMC method approximates the microcanonical configurational density of states  $g(U)$ , which is used to calculate the energy-dependent Boltzman weights:

$$p(U, T) = g(U)e^{-U/k_b T}. \quad (2.58)$$

The partition function can consequently be calculated by

$$Z(T) = \sum_U g(U)e^{-U/k_b T} \quad (2.59)$$

which can be used to obtain the canonical ensemble average of an observable  $A$  by

$$\langle A \rangle(T) = \frac{1}{Z(T)} \sum_U A(U)g(U)e^{-U/k_b T}. \quad (2.60)$$

The algorithm starts with an estimate of  $g(U)$ , which is set to  $\tilde{g}(U) = 1$  at the start. The acceptance probability from configuration  $x_1$  to  $x_2$  is calculated by the estimated density of states of their respective energies:

$$\text{acc}(x_1 \rightarrow x_2) = \min \left( 1, \frac{\tilde{g}(U(x_1))}{\tilde{g}(U(x_2))} \right). \quad (2.61)$$

The  $\tilde{g}(U)$  is updated after each MC step by

$$\ln[\tilde{g}_{t+1}(U)] = \ln[\tilde{g}_t(U)] + \gamma_t \delta_{U, U_{\text{new}}} + \pi(U), \quad (2.62)$$

with a bias  $\pi(U)$  and the convergence factor  $\gamma_t$ , which is updated by:

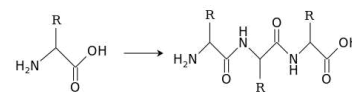
$$\gamma_t = \min \left( \gamma_0, \frac{t_0}{t} \right). \quad (2.63)$$

The convergence behaviour can be optimized by the parameters  $\gamma_0$  and  $t_0$ , and the convergence is proven if additional conditions are met.<sup>84,89,90</sup> Eq. 2.61 shows that states with low  $\tilde{g}(U)$  will be visited more often and if  $\tilde{g}(U) \approx g(U)$  the total number of visits will be equal for all  $U$ . Conversely, a flat histogram indicates uniform visit probabilities and therefore a good quality of  $\tilde{g}(U)$ .

## 2.4 Peptides

Peptides are biological (oligo-)polymers, that consists of single amino acids bound by peptide bonds, see Fig. 2.8. These peptides have an important set of backbone angles:  $\varphi$  [ $C(i) - C_{\alpha}(i) - N(i) - C(i + 1)$ ] and  $\psi$  [ $N(i - 1) - C(i) - C_{\alpha}(i) - N(i)$ ]. Important information about the secondary structure can be obtained purely from backbone dihedral angles. This information is used in Ramachandran plots<sup>91</sup>, where the two backbone dihedrals are plotted, and defined areas (sets of  $\varphi$  and  $\psi$ ) correspond to specific secondary structures, such as  $\alpha$ -helices or  $\beta$ -turns. Most commonly known are the 20 (+2)  $\alpha$ -amino acids, which are the only ones to appear in the genetic code of life.<sup>92</sup> The two additional amino acids selenocystein and pyrrolysine can be incorporated by special translation mechanisms. These amino acids can also be categorized into 6 groups depending on the chemical properties of the sidechain R. At neutral pH-value, there are positively and negatively charged sidechains, as well as uncharged sidechains with polar properties, and hydrophobic sidechains with alkane or aromatic residues. The last category consists of special cases like the smallest amino acid Glycine or sidechains with Sulphur- or Seleno- terminal groups. All naturally occurring proteinogenic amino acids are (L)-chiral and only a few naturally occurring non-proteinogenic amino acids are (D)-chiral.

Consisting of one or more long peptide chains, proteins are large biomolecules, which have various functions in living organisms. The sequence of the amino acids determines the protein's secondary and tertiary structure, which also varies in different chemical environments, influenced e.g. by additives, solvent, pH-value or temperature.<sup>12-14</sup> This change in structure is needed for example to store a hormone in a resting state.<sup>93</sup> The structure of these proteins in different environments is a reason for their functionality and obtaining knowledge about their states and transitions is highly important for medical, pharmaceutical and bio-engineering purposes. **Advances\_Poly\_Hybrid**,<sup>15-18,94,95</sup>



**Figure 2.8:** Overview of the reaction that forms peptides from amino acids. The amino- and acid-groups of the amino acid react to form a peptide bond.

### 2.4.1 Hybrid Peptide-Synthetic-Polymers

The study of hybrid peptide-containing polymers represents a distinct research area, driven by significant interest in the development of medical and environmentally friendly materials. The property design of these materials is already a significant area of research, while the synthesis of these materials is of a similar level of importance. Given the limited scope of this thesis, this chapter will concentrate on the particular characteristics of these materials. For an overview of the entire field, please refer to the sources provided. **Book\_HybridPoly\_Med**, **Book\_Peptide\_Mat**,<sup>96</sup> The incorporation of peptide segments into synthetic polymers provides them with the capacity to assemble into hierarchically organised nanoscale structures. The peptide structure formation occurs on multiple length scales, resulting from simple hydrogen bonds, secondary struc-

ture formation, and larger or tertiary structure formation. This process confers a significant advantage in the specialisation of the polymer. This is specifically useful for medical applications, for example in cancer treatment, where the ability to deliver drugs with high activity, high specificity, low toxicity and long circulation is of great benefit.<sup>98</sup>

Another noteworthy phenomenon is the self-assembly process and the potential for influencing this behavior, for instance, through amphiphilicity. In solution hybrid peptide-polymers can assemble into biomolecular nanostructures, such as micelles or vesicles.<sup>99</sup> As the peptide segment is often temperature sensitive, it is possible to have a thermoresponsive transition into, for instance, fibril structures.<sup>100</sup> Different structures are also formed in various concentrations or environments, leading to structure changes in bulk, at polar/non-polar interfaces or in thin layers.<sup>101</sup>

## 2.5 Analyses

Although molecular dynamics simulations provide precise information about the movement of atoms over time, the data obtained from these simulations is not always straightforward to interpret. There are a number of software packages that offer a range of analyses, focusing on equilibrium, dynamic or spectral properties, among others.<sup>102,103</sup>

### 2.5.1 Pair Distribution Functions

One of the most rudimentary analyses is the radial distribution function (RDF)  $g(r)$ , which shows the distance distribution of two particles. For particles  $N_i$  with their position vectors  $\vec{r}_i(t)$  the RDF is calculated using a bin width function  $\delta$ , as demonstrated in Eq. 2.64. It is also possible to calculate pair distribution functions for other properties, such as angles in angle distribution functions ADF (Eq. 2.65).

$$g_{ab}(r) = \frac{V}{N_a N_b r^2} \sum_{i=1}^{N_a} \sum_{j=i+1}^{N_b} \langle \delta(r - |\vec{r}_i(t) - \vec{r}_j(t)|) \rangle_t \quad (2.64)$$

$$ADF_{abc}(\alpha) = \frac{1}{\sin(\alpha)} \frac{1}{N_a N_b N_c} \sum_{i=1}^{N_a} \sum_{j=i+1}^{N_b} \sum_{k=j+1}^{N_c} \langle \delta[\alpha - \angle(\vec{r}_i(t) \vec{r}_j(t) \vec{r}_k(t))] \rangle \quad (2.65)$$

Two pair distribution functions can be correlated into a combined distribution function, which is visualized as a 2D contour plot, demonstrated in Fig. 2.10.

### 2.5.2 Hydrogen Bonds

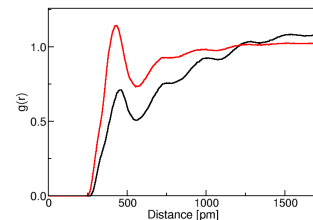
Hydrogen bonds are a driving force in peptide structure formation in classical MD simulations. As such there are several valuable properties to analyze. The lifetime of a hydrogen bond can be calculated from the averages (C) of all autocorrelation functions ACF(t), as shown in Eq. 2.66. The function  $s_i(t) = \{0, 1\}$  corresponds to a hydrogen bond criterion, which usually consists on an angle and distance parameter.

$$C(t) = \langle s_i(t) s_i(t + \tau) \rangle \quad (2.66)$$

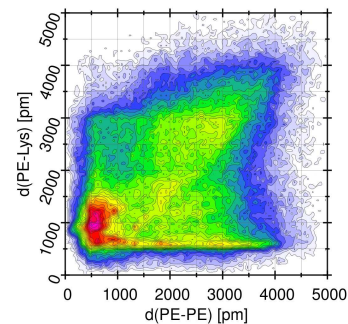
Commonly used values are a maximum angle for a hydrogen bond of  $\alpha_{HB}^{max} = 30^\circ$  and a maximum hydrogen bond distance between donor and acceptor of  $r_{HB}^{max} = 0.35 \text{ nm}$ . It is possible to obtain system specific parameters by using combined distribution functions of hydrogen bond distance and angle to show the most common conformation of the hydrogen bond.<sup>102</sup>

### 2.5.3 Protein- and Peptide-specific Analyses

Proteins and peptides exhibit a strong tendency to form secondary structures, which is why analytical techniques have been developed to identify these specific properties.



**Figure 2.9:** Exemplary RDFs for a binary system with mixed (black) or separated (red) phases, depending on the temperature.



**Figure 2.10:** Exemplary combined distribution function (CDF). (Taken from own Publ.[1])

### Radius of Gyration

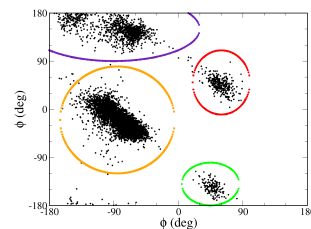
The radius of gyration ( $G_R$ ) is a structure parameter characterizing the size of a structure, which has different values for specific secondary structures like  $\alpha$ -helices and  $\beta$ -sheets. It is calculated by the distance  $R_i$  of atoms  $i$  with masses  $m_i$  from the center of mass  $R_{mean}$ :

$$R_g = \left( \frac{\sum_i |R_i - R_{mean}|^2 \cdot m_i}{\sum_i m_i} \right)^{0.5}. \quad (2.67)$$

Although the  $R_G$  is designed for an arbitrary system and is also employed in the analysis of polymers or other biomolecules, it is frequently utilized in the study of proteins and peptides.

### Ramachandran Plot

The Ramachandran plot<sup>91</sup> is a statistical evaluation of the combination of both backbone dihedral angles of a peptide. Different sections of the plot correspond to specific secondary structures. This is because, for example, in a helix, the sidechains prohibit specific backbone dihedral angles as they would result in steric hindrance. A rough example can be seen in Fig. 2.11.



**Figure 2.11:** Exemplary Ramachandran plot of an Asp-peptide with circles indicating different secondary structure regions.

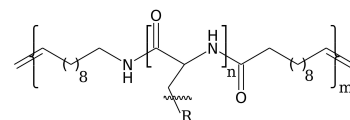
## Summary of the Published Papers

(The following chapter employs figures, which are taken from the respective publication they summarize.)

### 3.1 Secondary Structure Formation in Hybrid Synthetic/Peptide Polymers: Insights from Molecular Dynamics Simulations

In this work, the influence of the insertion of a polyethylene (PE) chain into a peptide segment in aqueous solution has been investigated by molecular dynamics simulations. Similar molecules have already been synthesised and studied by experimental groups<sup>104–106</sup>, but with protected side chains and in a different solvent. They found that the amount of  $\alpha$ -helical structures increased with increasing peptide length using Fourier-transformed infrared (FTIR) and circular dichroism (CD) spectroscopy. Given the limitations of these spectroscopic methods, we simulated the systems to understand these molecules at the atomic level. The focus was on the competition of structural driving forces between peptide hydrogen bonding and aggregation of hydrophobic PE chains in water, rather than on the secondary structures themselves. Understanding these hybrid synthetic/biopolymers could lead to new applications in medicine<sup>15–18</sup>, nanomaterials<sup>107</sup> and eco-friendly polymers<sup>94</sup>, such as those used in food packaging<sup>108</sup>.

Simulating large systems with sufficient convergence requires enhanced sampling methods to acquire a satisfactory quality of the results. Subsequently, the hybrid polymers were investigated using MD simulations with HREX<sup>60</sup> and REST2<sup>109,110</sup>. The general structure of the polymer is visualized in 3.1 and in Tab. 3.1 a list of simulated systems is given.



**Figure 3.1:** Structure of hybrid polymers composed of a homopeptide of amino acids (AA=Asp, Lys, Glu), which is connected to a short polyethylene chain. The double bond which connects two such hybrid monomers is maintained during synthesis<sup>104–106</sup>.



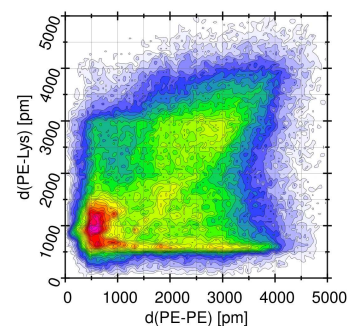
**Table 3.1:** Secondary structure analysis of all hybrid polymers. Shown is the integral of characteristic regions in the Ramachandran plot, which correspond to typical secondary structure motifs.

Hybrid Polymer	$\alpha_r$ -Helix	$\beta$ -Sheet	Other
[PE – (Lys) <sub>3</sub> ] <sub>2</sub>	0.83	0.12	0.05
[PE – (Lys) <sub>3</sub> ] <sub>6</sub>	0.73	0.15	0.12
[PE – (Lys) <sub>10</sub> ] <sub>2</sub>	0.82	0.15	0.03
[PE – (Glu) <sub>3</sub> ] <sub>2</sub>	0.79	0.18	0.03
[PE – (Glu) <sub>3</sub> ] <sub>6</sub>	0.63	0.23	0.14
[PE – (Glu) <sub>10</sub> ] <sub>2</sub>	0.88	0.10	0.02
[PE – (Asp) <sub>3</sub> ] <sub>2</sub>	0.73	0.23	0.04
[PE – (Asp) <sub>3</sub> ] <sub>6</sub>	0.62	0.17	0.21
[PE – (Asp) <sub>10</sub> ] <sub>2</sub>	0.88	0.10	0.02

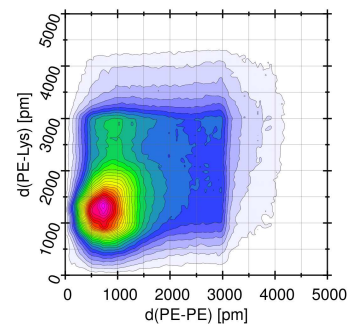
As the polymer has two vastly different components with opposing inter-molecular forces (Hydrogen bonding vs hydrophobic PE-chain aggregation), the bio-segment and the PE-segment were analyzed separately. The peptide segment was subjected to a series of commonly employed techniques, including Radius of Gyration ( $R_G$ ) and Ramachandran plots, with the objective of analysing its secondary structure. The polyethylene component was, however, analysed using standard radial distribution functions.

The results were not satisfactory, as the Ramachandran plot failed to reveal significant differences, while the RDFs highlighted some differences, but were hardly interpretable. Assessing the visual strongly varying structures, was only possible by including both structural forces and parts: the hydrogen bonding of the peptide and the hydrophobic PE-chain aggregation. Furthermore, the analysis should be readily applicable to a range of similar structures in the future, without necessitating a specific structural approach. With the PE-PE RDF showing different behaviour, we extended this analysis to a two-dimensional correlation RDF (=CDF: combined distribution function, shown in Fig. 3.2 and 3.3) with PE-PE chain interaction on the x-axis and the peptide-peptide distance also as a RDF on the y-axis. A preliminary analysis of the PE-PE RDF revealed significant differences. To gain further insight, we extended this analysis to a two-dimensional correlated RDF (CDF), which represents the PE-PE chain distances on the x-axis and the peptide-peptide distances as a RDF on the y-axis. The resulting plot, shown in Fig. 3.2 and 3.3), was presented with a colour scale, which demonstrated the distribution of all peptide-peptide distances for each PE-PE distance.

It was found that peptide chains of 10 amino acids maintain a secondary structure, albeit with a reduction in stability to approximately 50% following the analysis of intramolecular hydrogen bonding. The introduction of additional PE-chains into the peptide sequence results



**Figure 3.2:** Exemplary combined distribution function (CDF) of [PE – (Lys)<sub>10</sub>]<sub>2</sub>, which keeps a peptide dominated structure.



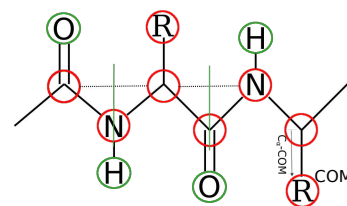
**Figure 3.3:** Exemplary combined distribution function (CDF) of [PE – (Lys)<sub>3</sub>]<sub>6</sub>, which adopts a ball-like structure.

in a complete shift in the force balance, with the previously dominant peptide interactions being superseded by a PE-chain dominated force. This results in the aggregation of PE-chains, which reduces the surface area to water. Consequently, the hybrid polymer assumes a ball-like structure. This phenomenon is also observed in peptides with the same monomer,  $[\text{PE} - (\text{AA}_3)_m]$ , but a smaller polymer length ( $m=2$  instead of  $m=6$ ).

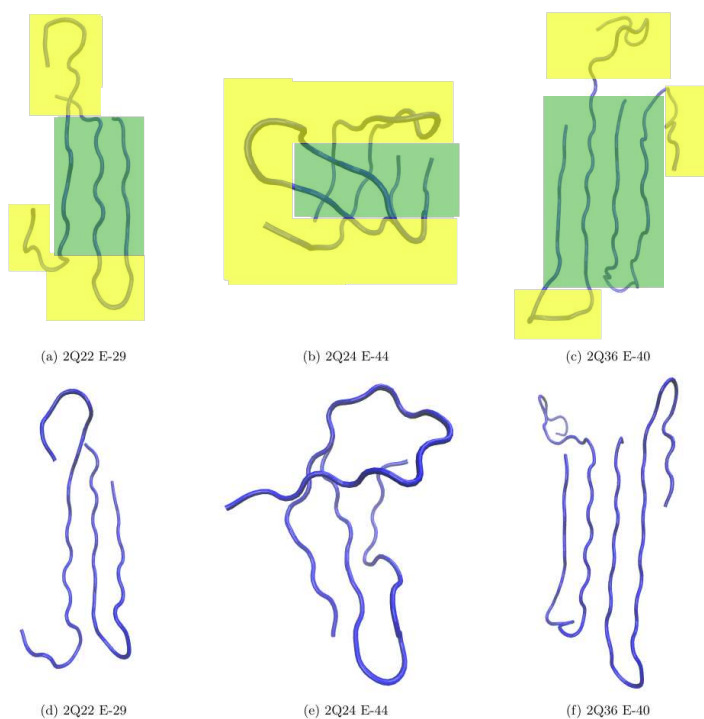
### 3.2 Reverse mapping of coarse-grained polyglutamine conformations from PRIME20 Sampling

In this work, we provide an algorithm for converting coarse-grained P20 structures into their all-atom equivalents. The method utilises the coordinates of the given P20 beads and, through geometric considerations, as illustrated in Fig. 3.4, allows for the addition of the carbonyl-oxygen and the hydrogen atoms bonded to the amino group. With regard to the sidechain, the coarse-grained bead provides only the centre of mass (COM). The all-atom sidechain group R is aligned with the vector( $C_\alpha - COM$ ) provided by the P20 beads.

The conversion algorithm was employed on a range of low-energy poly-glutamine dimers, differing in peptide length. The P20 structures generated by an SAMC calculation were found to be straightforward and reliable to convert. Following the conversion, an MD simulation was conducted to evaluate the stability of the low-energy P20 SAMC structures. It proved challenging to identify an appropriate criterion for stability, as the root-mean-square-deviation (RMSD) or radius of gyration ( $R_G$ ) were not sufficiently reliable for the range of structures under consideration. The visual stability criterion was primarily focused on the maintenance of the intermolecular shape of the dimer structure, as illustrated in Fig. 3.5.



**Figure 3.4:** Scheme describing the conversion of the coarse-grained structures into all atom geometries. Red: atoms obtained from the coarse-grained PRIME20 model, green: atoms added by simple geometric considerations.



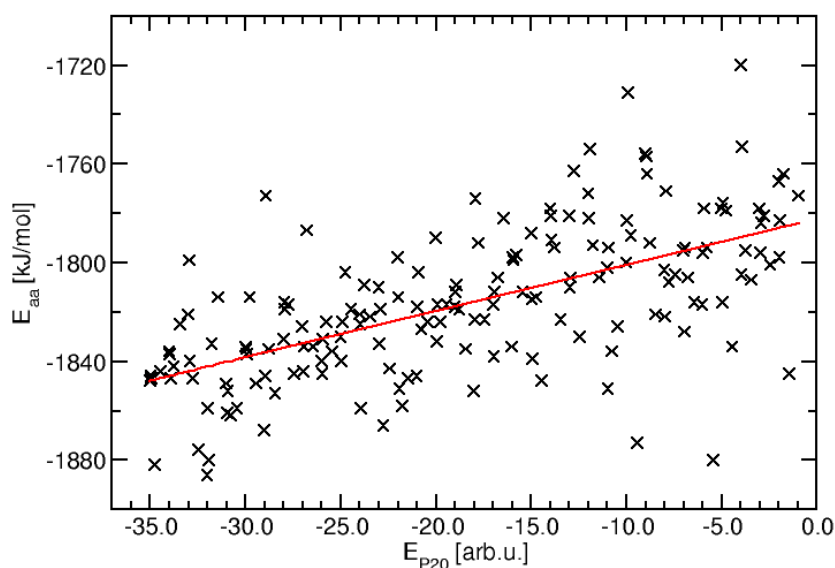
**Figure 3.5:** Visualization of the peptides at the start (a-c) and end (d-f) of the simulations. The dimer region, which is crucial for hydrogen bonding, is highlighted in green, while the less significant regions for dimer structure are marked in yellow.

This visual inspection on a chemical level refers to the inter-molecular

hydrogen bonds between the two peptides. As a result we analyzed the percentage of hydrogen bonds that remain stable during the MD simulation (=Hydrogen Bond auto-correlation function). This analysis was already more reliable than the previously mentioned RMSD or  $R_G$ . To achieve an even more satisfactory result, a differentiation between intra- and inter-molecular hydrogen bonding was made, as the focus on dimer stability gives reason to emphasize on inter-molecular interactions. Using this readily accessible analysis, revealed that around 70% of the given low-energy P20 dimer conformations were stable during the short MD simulations. This indicates the potential for a combined MC-MD approach to identify relevant peptide conformations in a more efficient manner, with reduced expenditure of resources and time.

### 3.3 Assignment of a physical energy scale for the dimensionless interaction energies within the PRIME20 peptide model

In this work we designed a calibration scheme to obtain conversion factors from the coarse-grained PRIME20 structures used for SAMC simulations to atomistic force field interactions used for simulation with full explicit solvation. The same conversion algorithm as in the previous paper was used to obtain a set of Glu<sub>26</sub>-dimer all-atom structures, which were then used for a 10 ns MD simulation to obtain locally relaxed energy estimates for each structure. The PRIME20 energy consists of two interactions: the number of backbone hydrogen bonds and the number of sidechain interactions. Two data sets were generated to analyse whether each of the two interactions could be transferred to atomistic MD energies.

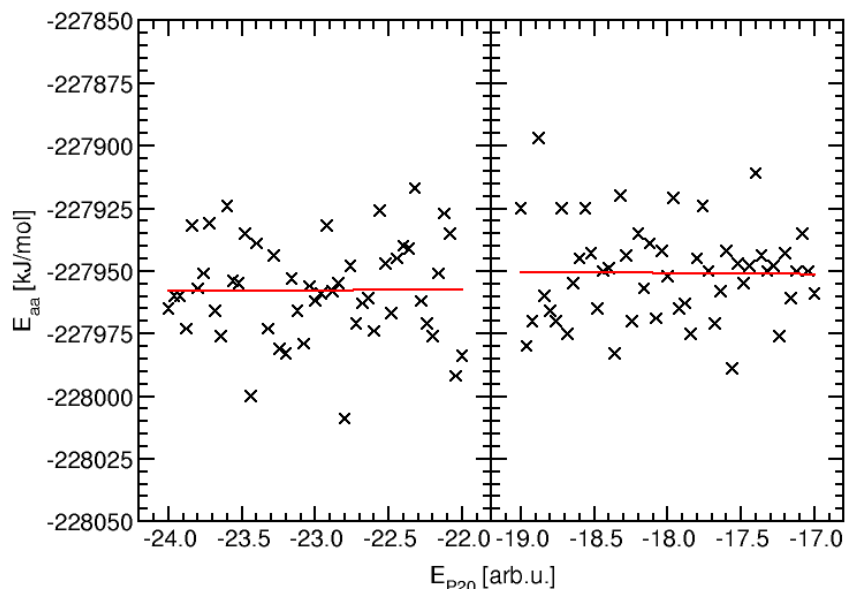


**Figure 3.6:** The hydrogen bond data set is shown with the P20 energies compared to the resulting energies from a 10 ns MD simulation.

Fig. 3.6 shows the P20 energies and their energies after conversion to an atomistic model and subsequent 10 ns MD simulation. The P20 structures have been chosen to differ by 1 arb.u., which corresponds to the energy of one hydrogen bond. The conversion factor obtained by linear regression is about 1.9 kJ/mol for a hydrogen bond in the P20 model, which is surprisingly low considering that an average hydrogen bond is typically about 20 kJ/mol. However in the simulation a hydrogen bond between NH and CO is not simply broken but changed to a peptide-solvent hydrogen bond. The energy difference corresponds to Eq. 3.1.

$$\Delta E_{P20} = E_{MD}(Pep \cdots Pep) + E_{MD}(H_2O \cdots H_2O) - 2E_{MD}(Pep \cdots H_2O). \quad (3.1)$$

The energy of the sidechain interaction was analyzed using a fixed number of hydrogen bonds and sampling the energetically close region. The results are summarized in Fig. 3.7, where the linear regression gives a slope of almost zero. Therefore there is no detectable correlation between coarse-grained and atomistic sidechain energies. On the one hand, the values of the sidechain interaction, which are 1/12th of a hydrogen bond in the coarse-grained model, may be too small to be detectable in the temperature noise in the energy. On the other hand, it is possible that during the 10 ns MD simulation the small conformational changes have an energetically larger effect than a side chain interaction.



**Figure 3.7:** Correlation between the coarse-grained ( $E_{P20}$ ) and atomistic ( $E_{aa}$ ) energies for a series of conformations with identical backbone hydrogen bonding states (at the coarse-grained level, here corresponding to 22 and 17 hydrogen bonds, respectively, for the left and right plots). In turn, the number of sidechain interactions varies and corresponds to -1/12 units of  $E_{P20}$  per sidechain interaction.

In order to verify the results, statistical analyses were carried out to ensure that the energy conversion factor was not in fact a statistical anomaly. It turned out that the MD simulation is not able to give an energy value with an error lower than the obtained 2 kJ/mol. However, when the hydrogen bonds are analysed over 35 hydrogen bonds, the required detectable energy rises to 70 kJ/mol. This value is in the range of our error bar, which we believe to be in the range of 10-20 kJ/mol.

*Paper I: Secondary Structure Formation in  
Hybrid Synthetic/Peptide Polymers: Insights  
from Molecular Dynamics Simulations*

Thomas Kunze, Christian Dreßler and Daniel Sebastiani

**Secondary Structure Formation in Hybrid Synthetic/Peptide  
Polymers: Insights from Molecular Dynamics Simulations.**

*Macromol. Theory Simul.* 2023, Vol. 32, No. 3, pp. 1-8, 2200070.

For this paper, all calculations and analyses were performed by me with guidance from (now Prof.) Christian Dreßler and under general supervision of Prof. Daniel Sebastiani. The text was written by me with support from Prof. Sebastiani and with editorial remarks from Christian Dreßler. The introductory text summarizing the previous SFB-projects (Sec. 1.1) was written solely by Prof. Daniel Sebastiani.

I hereby confirm that the use of this article is compliant with all publishing agreements, as this article is under an open access licence: <https://creativecommons.org/licenses/by/4.0/>

# Secondary Structure Formation in Hybrid Synthetic/Peptide Polymers: Insights from Molecular Dynamics Simulations

Thomas Kunze, Christian Dreßler, and Daniel Sebastiani\*

Proteins and peptides exhibit an immense variety of structures, which are generally classified according to simple structural motifs (mainly  $\alpha$  helices and  $\beta$  sheets). Considerable efforts have been invested in understanding the relationship between chemical structure (primary structure) of peptides and their spatial motifs (secondary structure). However, little is known about the possibility to interfere intentionally in these structural driving forces, for example, by inserting (short) artificial polymer chains in the peptide backbone. Structure formation on such hybrid synthetic/biochemical polymers is still an emerging field of research. Here, molecular dynamics simulations are used to illustrate the influence of inserted polyethylene segments on the secondary structure of several peptide homopolymers. A loss of structure of  $\approx 50\%$  when the peptide chain length drops to ten amino acids and a practically complete absence for even shorter peptide segments.

## 1. Introduction

### 1.1. Introduction to Project A09 of the SFB/TRR 102 “Polymers under Multiple Constraints”

This work was part of the third funding phase (2019–2023) of project A9 of the Collaborative Research Centre SFB/TRR 102, funded by the German Science Foundation (DFG). In the first two periods (2012–2015 and 2015–2019, respectively), the research focus was initially put on ions and small functional segments of biomolecules, such as chromophores, and their interaction with solvent molecules. Specifically, the first funding period (2011–2015) resulted in a series of investigations of aqueous solutions

of simple and complex ions,<sup>[1–3]</sup> in combination with their spectroscopic fingerprints.<sup>[1,4]</sup> Nevertheless, specific features of individual protein systems were already simulated as well.

In the second funding period (2015–2019), the focus shifted to the investigation of conformational variations of small amyloidogenic proteins due to specific mutations which modify the equilibrium of intra-protein interactions and protein–solvent interactions.<sup>[5]</sup> In parallel, the interactions of salt ions with solvating water molecules and the interaction equilibrium between a typical protein salt bridge and the solvating water molecules complemented this picture.<sup>[6–8]</sup> Special attention was again paid to the importance of spectroscopic signatures, both regarding infrared<sup>[6,7]</sup> and NMR

spectroscopy.<sup>[5,9]</sup> Finally, during the third funding period (2019–2023), the project has evolved toward two distinct directions: general polymer structure formation and functional hybrid organic/inorganic copolymers in the context of energy storage/conversion.

The first direction is the continuation of the collaborations on polymer structure formation, which comprises the analysis of secondary structure perturbation of peptide chains via flexible polyethylene segments (research part of this article), the understanding of induced helicity in an achiral polymer via a single chiral monomer within the polymer (collaboration with project A03), and the development of reverse coarse-graining scheme for benchmarking the sampling quality of a coarse-grained simulation run of systems such as polyglutamine (collaboration with project A07), as well as the interplay between polymer structure (here: cellulose) and the solubility in different kinds of solvents.<sup>[10]</sup>

On the other hand, a second direction has emerged from the consideration of novel types of polymers, specifically hybrid organic/inorganic copolymers. On the example of crosslinked organic/sulfur compounds, we have started an investigation about local packing effects and lithium diffusion/lithiation reactions in the presence of lithium ions.<sup>[11]</sup> Beforehand, a preliminary study dealt with the understanding of lithium diffusion in amorphous thiophene.<sup>[12]</sup> This research line is presently intensified and will give rise to novel projects in the context of renewable energies.

Concluding this survey of topic of this project A09, we want to spotlight a methodological project line that has emerged from the successful collaboration with project A07 (W. Paul). In our joint study of polyglutamine and its conformational distribution in aqueous solution,<sup>[13]</sup> we have established an inverse coarse

T. Kunze, D. Sebastiani  
 Martin-Luther University Halle-Wittenberg  
 Institute for Physical Chemistry  
 Von-Danckelmann-Platz 4, 06120 Halle/Saale, Germany  
 E-mail: daniel.sebastiani@chemie.uni-halle.de

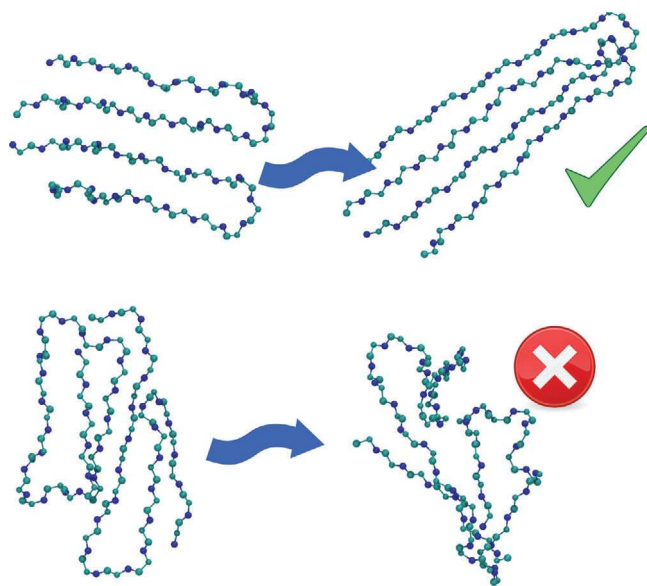
C. Dreßler  
 Institute of Physics  
 Ilmenau University of Technology  
 Weimarer Straße 32, 98684 Ilmenau, Germany

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/mats.202200070>

© 2023 The Authors. Macromolecular Theory and Simulations published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/mats.202200070





**Figure 1.** Illustration of the result of the inverse coarse-graining procedure from the PRIME20 peptide model to atomistic resolution, on the example of a (Gln)<sub>22</sub> dimer. Two example conformations are shown: on the left, the atomistic structure directly reconstructed from the PRIME20 conformation; on the right, the corresponding structure after 10 ns of equilibration via atomistic molecular dynamics simulations (in aqueous solution). The upper conformation remains stable (apart from a global rotation), while the lower structure changes significantly, marking an enthalpically unfavorable conformation of the PRIME20 sampling.

graining protocol for the PRIME20 coarse grained protein model.<sup>[14]</sup> We have designed a computational scheme to re-establish a tentative atomistic structure of the glutamine oligopeptide after conformational sampling with at the coarse-grained level using the PRIME20 model. The tentative structures are then equilibrated for a short period of 10 ns using all-atom molecular dynamics simulations, and the degree of structural deviation from the tentative conformation is measured. Using this scheme, we are able to validate the phase space sampling quality of the PRIME20 interaction scheme.

An illustration of the application potential of our scheme is shown in **Figure 1** on the example of a glutamine-22 dimer. Two structures (upper left and lower left) have been obtained from the PRIME20 conformational sampling scheme, using the inverse coarse-graining algorithm. Subsequently, molecular dynamics simulation have been performed for 10 ns in aqueous solution, yielding the atomistic conformations shown in the upper right and lower right parts of Figure 1. While the upper conformation has only rotated in space but otherwise remained unchanged, the lower structure has changed significantly during the equilibration run. Such a situation indicates that the particular conformation as obtained from the PRIME20 conformational sampling does not represent a local minimum of the potential energy landscape and should thus be discarded. Preliminary results indicate that the majority of the conformations generated by the PRIME20 sampling approach are “good” structures which represent stable conformations also within the atomistic equilibration run. Particular focus will be put on the correlation between the

energies computed at the coarse grained level and those obtained from the atomistic molecular dynamics simulation.

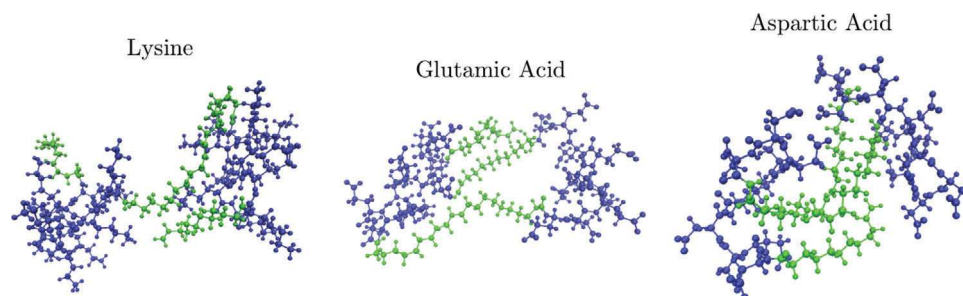
## 1.2. Introduction to this Research

Proteins have a significant role in our body, ranging from metabolism, defense mechanisms to structural functions in skin and bones amongst many others. With a wide range of functions there is also the possibility of multiple malfunctions, noticeable by the huge amount of diseases caused by misfolded proteins. Notoriously Alzheimer's,<sup>[15]</sup> Huntington's,<sup>[16]</sup> and Parkinson's<sup>[17]</sup> are some of the most commonly known neurodegenerative diseases attributed to the toxicity after a structural change and following aggregation.<sup>[18]</sup> This transition can occur for multiple reasons, including temperature, pH-value, additives, and solvent.<sup>[19–21]</sup>

A therapeutic approach curing these diseases consists of molecules similar to the aggregating proteins but with the ability to destabilize the aggregating conformation.<sup>[22]</sup> This was already tried with so called  $\beta$ -sheet breaker inhibiting and dissolving amyloid- $\beta$  structures, which are the cause for Alzheimer's disease.<sup>[18,23]</sup> Since there has not been any major success for this complex problem yet, new models and approaches are discussed.<sup>[24,25]</sup> Such conformation inducing compounds can often be found as peptides, especially as copolymers to obtain a specific drug delivery system.<sup>[26–29]</sup> The use of peptide containing hybrid copolymers,<sup>[30]</sup> not only enables a possibly nontoxic drug delivery system, it can also influence the secondary structure to well-defined physical characteristics.<sup>[31–35]</sup>

The synthetic preparation of hybrid polymers can be achieved by acyclic diene metathesis (ADMET) polymerization in hexafluoroisopropanol (HFIP) or trifluoroethanol (TFE), which takes already prepared oligopeptides, adding alkenes on both sides to polymerize multiple of these molecules by metathesis.<sup>[30,36,37]</sup> The base oligopeptides can be obtained by ring-opening polymerization (ROP), however full natural proteins were previously only available by the Merrifield's method, which can only produce small peptides at high costs.<sup>[38–40]</sup> By advancement in chemical protein synthesis powerful alternative synthesis methods are now available to produce large previously unobtainable proteins, however this approach deals with different difficulties such as solubility and purification among others.<sup>[41]</sup>

For industrial purposes, the research of hybrid polymers is focused on effectively enhancing or substituting widely used bulk polymers with biopolymers to increase biodegradability and environmentally friendly production.<sup>[42]</sup> This includes materials for packaging even in the food industry,<sup>[43]</sup> polymers which are able to form nanomaterials<sup>[44]</sup> used in water purification<sup>[45]</sup> or medical supply like implants.<sup>[46]</sup> Combining synthetic and biopolymers gives a whole new class of materials, that has the prospect to utilize the best properties of both. One of the main advantages of incorporating biological segments into hybrid polymers is the structural control, for example, given by the secondary structure of peptide segments.<sup>[47]</sup> This control includes regulated structure transitions caused by pH- or temperature-change among other things.<sup>[48]</sup> However, the application of hybrid peptide-copolymers is mostly limited to drug delivery applications as of now,<sup>[49]</sup> but



**Figure 2.** Snapshots of molecular structures of  $[\text{PE}-(\text{AA})_{10}]_2$  molecules (left to right: Lys, Glu, Asp) during the molecular dynamics simulation. The explicit solvent molecules are not shown.

upon further research could also have an impact on functional surfaces or biomineralization.<sup>[50]</sup>

For structural analysis a set of different techniques is feasible, including circular dichroism (CD) as well as Fourier transformed infrared (FTIR) spectroscopy in solution and solid-state.<sup>[51–53]</sup>

In this work, the interplay of synthetic and peptide segments in hybrid polymers is investigated. We focus on the insertion of a specific type of synthetic polymer segment into a series of model peptides. We determine characteristic conformational motifs and their dependence on the chemistry of the hybrid molecule, specifically amino acid (AA) type and length. A particular focus lies on the attempt to identify conformational patterns induced by the different geometric chain flexibility, but also the type of philicity of the polymer segments: while the PE chains are the textbook example of hydrophobic molecules, the charged peptide segments have a distinctly hydrophilic character.

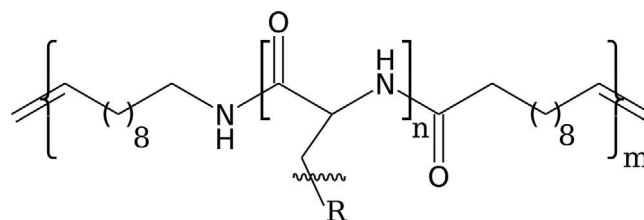
An illustration of three hybrid polymers investigated is given in **Figure 2**.

Studies on the structure of homopeptides have been done for a long time, often by using circular dichroism spectroscopy. For lysine (Lys) oligomers the structure is random for short peptides and gets more helical content with increasing length and even more so at higher pH values.<sup>[54]</sup> Glutamic acid (Glu) peptides show similar behavior, transitioning from random to helical structure starting at  $n = 10$  and having a full helical structure after  $n = 50$ .<sup>[55]</sup>

With the insertion of a PE chain into the continuous homopeptide, stronger hydrophobic forces get introduced into the molecule. These new integrated forces are opposed to the peptide intra-molecular and solvent hydrogen bonds and will therefore result in conformations with reduced solvent accessible surface area compared to peptides with similar overall length. In **Figure 2** common established structures are visualized. The green PE chains show multiple possible positions in regard to peptides and other PE chains. While this qualitative trend is obvious, we presently have little quantitative understanding of how exactly the peptide conformation is influenced by the geometric flexibility and the hydrophobic forces due to the artificial PE chains. This is the primary aspect that is addressed in the following, using atomistic molecular dynamics simulations.

## 2. Experimental Section

The chemical structure of the small model hybrid systems are shown in **Figure 3**, with peptides connected by a PE chain similar



**Figure 3.** Structure of hybrid polymers composed of a homopeptide with amino acid (AA=Asp, Lys, Glu), which is connected to a short polyethylene chain. The double bond which connects two such hybrid monomers is maintained during synthesis.<sup>[30,36,37]</sup>

**Table 1.** Summary of computational parameters for all sets of hybrid polymers. While the temperature  $T_{\text{max}}$  of the highest of  $N_{\text{repl}}$  replica differs for the systems, the temperature  $T_{\text{system}}$  for the calculation of the lowest replica, which is the one considered for analysis, remains the same. (AA=Asp, Lys, Glu).

Polymer type	$T_{\text{max}}$ [K]	$N_{\text{repl}}$	$T_{\text{system}}$ [K]
$[\text{PE}-(\text{AA})_{10}]_2$	500	16	300
$[\text{PE}-(\text{AA})_3]_2$	700	16	300
$[\text{PE}-(\text{AA})_3]_6$	600	16	300

to experimentally investigated hybrid polymers.<sup>[30,36,37]</sup> The peptides were varied in length and amino acid type (AA=Asp, Lys, Glu).

The hybrid polymers were simulated with classical molecular dynamics simulations using the enhanced sampling technique Hamiltonian replica exchange<sup>[56]</sup> with solute tempering 2 (REST2)<sup>[57,58]</sup> implemented by PLUMED<sup>[59]</sup> into GROMACS version 2019.1.<sup>[60,61]</sup> The CHARMM27<sup>[62,63]</sup> force field was used to calculate the hybrid polymer interactions and solvation effects of water were calculated with the TIP3P water model. Each type of hybrid molecule was put elongated into cubic box filled with water, ranging from 6 to 10 nm depending on the hybrid molecule's size. All amino acid sidechains were charged, therefore the whole system was neutralized afterward with chloride or sodium ions. Overall 16 replica per system were used starting at 300 K and ranging up to 500 to 700 K, again dependent on the system size to achieve average exchange probabilities ranging from 10% to 40%, summarized in **Table 1**. With a Lincs<sup>[64]</sup> fourth order linear constraint for covalent hydrogen bonds the time step was kept at 0.5 fs for technical reasons at early equilibration with replica

**Table 2.** Secondary structure analysis of all hybrid polymers and references. Shown is the integral of characteristic regions in the Ramachandran plot, which correspond to typical secondary structure motifs.

Hybrid polymer	$\alpha$ -helix	$\beta$ -sheet	Other
[PE-(Asp) <sub>3</sub> ] <sub>2</sub>	0.74	0.22	0.04
[PE-(Asp) <sub>3</sub> ] <sub>6</sub>	0.61	0.18	0.21
[PE-(Asp) <sub>10</sub> ] <sub>2</sub>	0.83	0.15	0.02
6× Asp <sub>3</sub>	0.55	0.32	0.13
2× Asp <sub>10</sub>	0.73	0.26	0.01
[PE-(Glu) <sub>3</sub> ] <sub>2</sub>	0.75	0.20	0.05
[PE-(Glu) <sub>3</sub> ] <sub>6</sub>	0.65	0.22	0.13
[PE-(Glu) <sub>10</sub> ] <sub>2</sub>	0.87	0.11	0.02
6× Glu <sub>3</sub>	0.49	0.38	0.13
2× Glu <sub>10</sub>	0.82	0.17	0.01
[PE-(Lys) <sub>3</sub> ] <sub>2</sub>	0.79	0.14	0.07
[PE-(Lys) <sub>3</sub> ] <sub>6</sub>	0.70	0.19	0.11
[PE-(Lys) <sub>10</sub> ] <sub>2</sub>	0.81	0.15	0.04
6× Lys <sub>3</sub>	0.50	0.35	0.15
2× Lys <sub>10</sub>	0.92	0.02	0.06

exchange attempt every 200 steps. The simulations were run for 60 ns and analyzed with built-in GROMACS tools for hydrogen bonding, radius of gyration ( $R_g$ ) and Ramachandran plots. All other analysis was performed by TRAVIS<sup>[66,67]</sup> and visualization was carried out by VMD.<sup>[65]</sup> Analysis started after 20 ns of initial calculation. The [PE-(Glu)<sub>3</sub>]<sub>6</sub> system was run for 90 ns, however no significant difference in any of the relevant conformational distributions was found. The reference calculation for the homopeptides used the same parameters, however without the replica exchange method. The termini of the peptide were capped by *N*-methyl (NME) and acetyl (ACE) to reduce the effect of the termini and as a result have a better reference.

## 3. Results

### 3.1. Hybrid Polymer with Single Embedded PE Chain

Our focus lies on the formation of structural motifs at a more general level, specifically the characterization of hybrid polymers in terms of the spatial arrangement of the peptide and PE segments. With this goal in mind we have determined structural parameters, which we have found to be suitable to describe (and discriminate) the overall conformations of our hybrid polymers. For reference we analyzed the peptides without the PE-chains, shown as “6× AA<sub>3</sub>” and “2× AA<sub>10</sub>” in Table 2.

One of the main characteristics of peptide secondary structures is the formation of helical or hairpin structure (helices, sheet-like structures). Our starting point was therefore to analyze the degree of helicity, which the hybrid polymers are able to establish. To derive an approximate conformational statistics, we calculated the Ramachandran plots and quantified the secondary structure regions approximately. The Ramachandran plot itself is a representation of the backbone angles ( $\psi$ ,  $\phi$ ), thereby each secondary structure correlates to a specific region in the plot. Accumulating all points in these specific regions leads to a roughly

estimated distribution of secondary structures. The results in Table 2 show similar behavior, observing roughly 75%  $\alpha$ - and 15%  $\beta$ -structures for all types of molecules. Some hybrids show more equal behavior, for example, [PE-(AA)<sub>10</sub>]<sub>2</sub> and [PE-(AA)<sub>3</sub>]<sub>2</sub>. Visual inspection however shows differences between the amino acids comparing the snapshots in Figures 4 and 2, which are not represented by the Ramachandran plot. Hence, it turns out this broadly used analysis does not give satisfactory structural classification for very short peptide segments. The underlying reason is that short peptides exhibit characteristic backbone angles without adopting the corresponding secondary structure motifs.

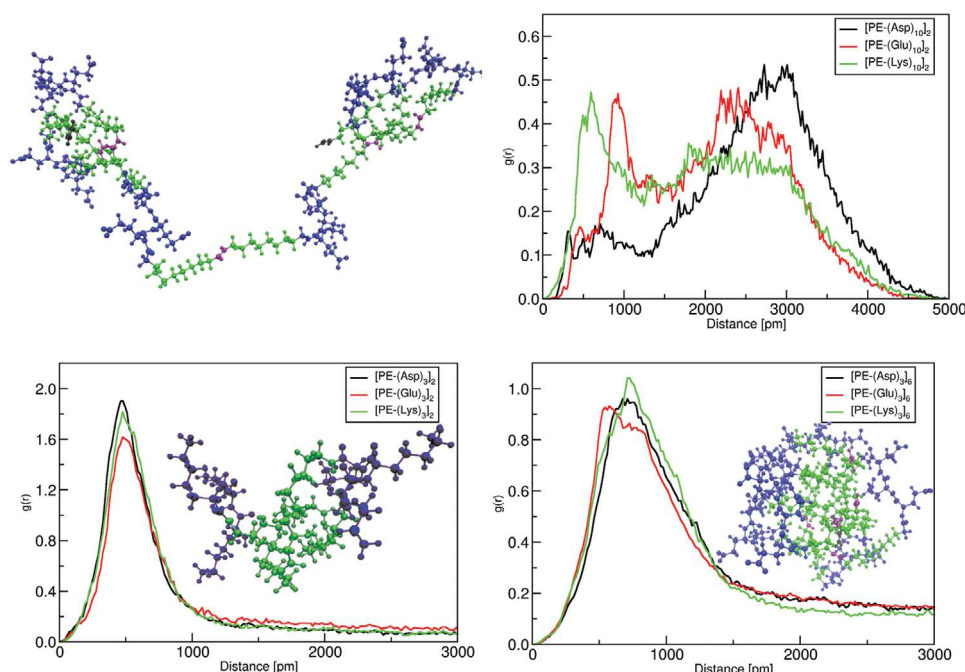
Another aspect of the structure formation for our hybrid systems is the behavior of the PE chains. Consequently, we calculated radial distribution functions (RDF) of the PE chains  $d(\text{PE-PE})$ . The polyethylene–polyethylene distance  $d(\text{PE-PE})$  is calculated as the distance between the central carbon atoms of the polyethylene chains. These central carbon atoms are marked in purple in the snapshot in Figure 4 (top left). The number as well as the positions of the maxima of the RDFs differ for each amino acid AA = Asp, Glu, Lys in the hybrid polymer series [PE-(AA)<sub>10</sub>]<sub>2</sub>. In contrast, the RDFs of the small peptide–polyethylene hybrid molecules of the type [PE-(AA)<sub>3</sub>]<sub>2</sub> have similar shapes and only one distinct maximum. The RDFs of the hybrid-polymers of the type [PE-(AA)<sub>3</sub>]<sub>6</sub> show a rather similar behavior compared to the RDFs of the hybrid-polymer [PE-(AA)<sub>3</sub>]<sub>2</sub>. Based on the analysis of the RDFs, we can conclude that bigger peptides have greater PE–PE distances and are stronger affected by the type of the amino acid. In order to explain the origin of these observation, we have therefore switched to more complex geometric parameters in order to capture more detail of the genuine peptide structure via the intercalated PE chains. Specifically, we have found two particular distances that are able to describe and discriminate our hybrid polymers. The first parameter is the distance between centers of masses of a PE chain and a peptide segment  $d(\text{PE-AA})$  with AA=Asp, Glu, Lys. Note that each pair of PE chain/peptide segment is considered (not only the adjacent chain). The other parameter  $d(\text{PE-PE})$  is the distance between the center of mass of a given PE chain to the center of mass of another, not necessarily adjacent, PE chain.

With the two given parameters  $d(\text{PE-PE})$  and  $d(\text{PE-AA})$  the structure was analyzed by a combined distribution function<sup>[66,67]</sup>  $g(d(\text{PE-AA}), d(\text{PE-PE}))$ . This function represents the 2D probability distribution for the simultaneous occurrence of a given PE–AA distance and a given PE–PE distance.

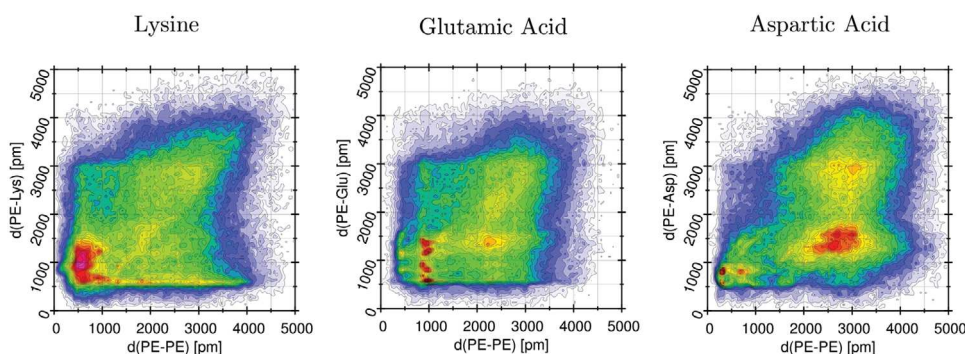
A particular feature of this set of parameters is the explicit incorporation of molecular philicity: The combined distribution function of  $d(\text{PE-PE})$  and  $d(\text{PE-AA})$  allow the analysis of aggregation effects due to segments of same philicity (PE–PE) and opposite philicity (PE–AA). Notably, the distances between PE chains also give information about the secondary structure, since the PE segments are the linking parts between peptides. This way, a linear elongated peptide will yield larger PE–PE distances, while secondary structure features like helices or turns will result in nearby PE–PE distances.

The results of the radial distribution functions  $g(d(\text{PE-AA}), d(\text{PE-PE}))$  for the [PE-(AA)<sub>10</sub>]<sub>2</sub> systems are shown in Figure 5. The hybrid polymers exhibit distinctly different patterns for the three amino acids. Each type has an individual pattern of highly





**Figure 4.** Radial distribution functions of the distances  $d(\text{PE-PE})$  between two polyethylene chains for the different hybrid polymers. Herein, the polyethylene–polyethylene distance  $d(\text{PE-PE})$  is calculated as the center of mass of the PE-chain. Note that the positions of the terminating polyethylen chains are also included into the calculation of the RDFs.



**Figure 5.** Combined radial distribution functions of  $[\text{PE}-(\text{AA})_{10}]_2$  for the centers of masses  $d(\text{PE-PE})$  and  $d(\text{PE-AA})$ .

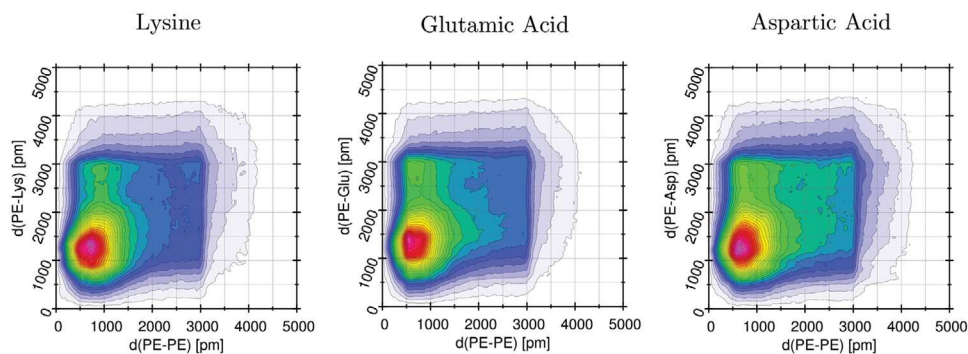
populated conformations in the combined distribution functions. For example, PE-Glu shows multiple local maxima at  $d(\text{PE-PE})=1$  nm with  $d(\text{PE-AA})$  ranging from 1 to 3 nm. For aspartic acid, on the other hand, no such pattern of local maxima is visible; instead, the system has a broad peak at (3 nm, 2.5 nm). It should be noted that these combined distribution functions are averaged over about 50 ns, and show considerable evolution during this simulation period. A series of distribution functions for shorter time windows is given in Supporting Information for illustration.

This analysis cannot capture the full conformational space of the hybrid polymers due to incomplete statistical sampling convergence of the molecular dynamics simulation. However, the visible population of spatially extended conformations for early and late stages of the simulation (see Supporting Information) indicates that the structural sampling is most probably sufficient for a qualitative analysis.

The very distinct pattern for larger peptide segments observed in our simulations show that the actual chemistry of the amino acid type is decisive for the conformational distribution. Therefore, these hybrid polymers with a chain length of ten amino acids can be considered as peptide dominated structures, and the insertion of the synthetic polymer segments does not eliminate the individual structures from the peptide segment. The structure formation is clearly influenced by the polyethylene chains, but the peptide segment is long enough so that the individual chemistry of the amino acid side chains is able to contribute significantly to the secondary structure formation process.

### 3.2. Hybrid Polymer with Multiple Embedded PE Chain

A natural follow-up question is how much the structure of the peptide strand is changed if the peptide is “interrupted” multiple



**Figure 6.** Combined radial distribution functions of  $[\text{PE}-(\text{AA})_3]_6$  for the centers of masses  $d(\text{PE}-\text{PE})$  and  $d(\text{PE}-\text{AA})$ .

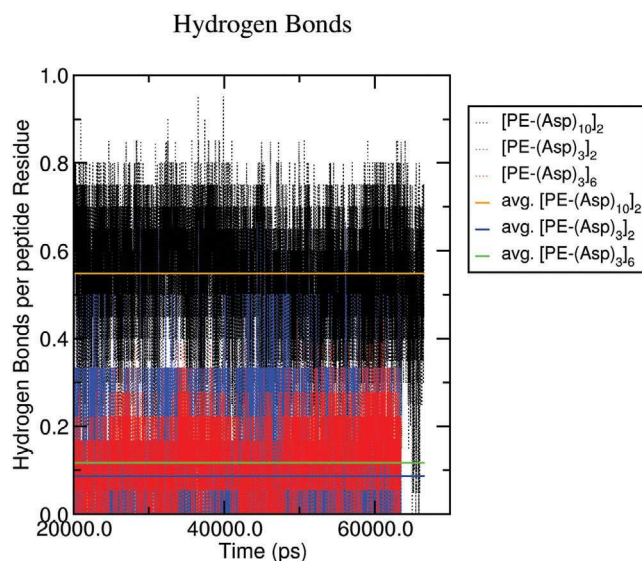
times. We therefore switched from  $[\text{PE}-(\text{AA})_{10}]_2$  to  $[\text{PE}-(\text{AA})_3]_6$  molecules having a similar number of peptide residues but severely different peptide lengths. A preliminary hypothesis is that the hydrophobic properties of the PE chain overcome the peptide interaction and lead to a PE dominated structure. While this hypothesis is not supported by the Ramachandran plots in Table 2, we observe a visually quite strong difference in conformational pattern between  $[\text{PE}-(\text{AA})_{10}]_2$  and  $[\text{PE}-(\text{AA})_3]_6$ . Again, we utilize the combined distribution function  $g(d(\text{PE}-\text{AA}), d(\text{PE}-\text{PE}))$  shown in **Figure 6**. At first view, the patterns for  $[\text{PE}-(\text{AA})_3]_6$  look clearly distinct from the corresponding combined distribution function from the  $[\text{PE}-(\text{AA})_{10}]_2$  systems.

The combined distribution functions are independent of the amino acid type, which is at variance with the observations for the  $[\text{PE}-(\text{AA})_{10}]_2$  polymers. The favorable regions are also less scattered and show only a single maximum. Furthermore, the PE-PE distances are also more confined to lower distances compared to the PE-AA distance distribution, which suggests the peptide part to be more flexible for  $[\text{PE}-(\text{AA})_3]_6$  compared to  $[\text{PE}-(\text{AA})_{10}]_2$ . Besides the reduced length of the intermediate AA unit, this is caused by a lower amount of formed hydrogen bonds, shown by the average hydrogen bonds per residue over time in **Figure 7**. Due to the charged sidechains of the peptide segments the hydrogen bonds are considered to be only formed by the backbone. Overall, this indicates a structure dominated by the hydrophobic PE chains.

### 3.3. Influence of Total Chain Length

As a complementary question, we have investigated whether the conformational pattern of our hybrid polymers depends significantly on the number of repetitions of the elementary building block, that is, the number of monomer units  $m$  in  $[\text{PE}-(\text{AA})_n]_m$ . To this purpose, we have taken the  $[\text{PE}-(\text{AA})_3]_6$  systems (AA=Lys, Glu, Asp) and reduced the polymer index from  $m = 6$  to  $m = 2$ , which corresponds to the minimal meaningful version of this hybrid polymers.

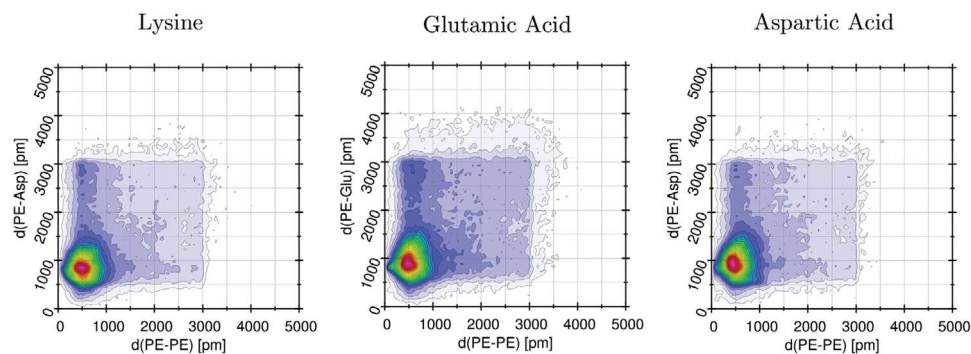
The combined spatial distribution functions of the two characteristic effective distances  $d(\text{PE}-\text{AA})$  and  $d(\text{PE}-\text{PE})$  in the three systems is shown in **Figure 8**. The pattern shows an almost vanishing probability for PE-PE distances larger than 1 nm, and similarly low probabilities for PE-AA distances beyond 1.5 nm. No significant variations are observed upon exchange of the amino



**Figure 7.** Time curve of the average number of backbone hydrogen bonds of one peptide residue to any other peptide for all  $[\text{PE}-(\text{Asp})_n]_m$ . This includes hydrogen bonds to peptides in the same unit, as well as hydrogen bonds to other peptide units, which are interrupted by a PE-chain. The number of hydrogen bonds is divided by the number of peptide residues to compare the systems, which is indicated by “per peptide residue.”

acid Lys/Glu/Asp. The considerable reduction of the distance distribution to the central peak near (0.5 nm, 1.0 nm) is even more apparent when comparing to the hybrid polymers of the same type but triple length (i.e.,  $[\text{PE}-(\text{AA})_3]_6$ ) in **Figure 6**. In the latter, the intermediate distances up to about 3 nm are visibly more frequent in the distribution. It should be noted that while the overall shape of the distribution function is still quite similar between  $m = 2$  and  $m = 6$ , the change in polymer length has a considerably larger effect than the variation of amino acid type.

Geometrically, the (PE-PE, PE-AA) peak distances around (0.5 nm, 1.0 nm) correspond to directly adjacent polymer strands. We have checked whether this spatial proximity is rather accidental or a direct consequence of hydrogen bonding, which is the most prominent structural driving force in peptides and proteins. The temporal evolution of the number of hydrogen bonds per amino acid during our trajectories is shown in **Figure 7**. The pattern for the hybrid polymers with short amino acid sequences



**Figure 8.** Combined radial distribution functions of  $[\text{PE}-(\text{AA})_3]_2$  for the centers of masses  $d(\text{PE}-\text{PE})$  and  $d(\text{PE}-\text{AA})$ .

(i.e.,  $n = 3$ , shown in red and blue) are very similar. This confirms the picture observed in the combined radial distribution functions (Figures 6 and 8). However, the polymers with longer amino acid chains ( $n = 10$ ) show a hydrogen bonding probability that is about five times higher. With an absolute value around 0.6, this probability is already close to the expected limit for a regular peptide of 0.8–0.9.

These findings illustrate that while the  $[\text{PE}-(\text{AA})_3]_6$  hybrid polymer has about the same total number of amino acids as  $[\text{PE}-(\text{AA})_{10}]_2$ , it behaves structurally much more like the minimalist  $[\text{PE}-(\text{AA})_3]_2$  polymer. On the other hand, the  $[\text{PE}-(\text{AA})_{10}]_2$  system is already similar to a regular peptide, despite the considerable perturbation from the polyethylene segment.

## 4. Conclusion

We have modeled the conformational space of a series of short peptides and their persistence upon perturbation of their peptide sequence by insertion of short, highly flexible polyethylene segments. Specifically, we have performed molecular dynamics simulations of  $[\text{PE}-(\text{AA})_n]_{\text{lm}}$  (with  $\text{AA}=\text{Asp, Lys, Glu}$ ;  $n=\{3, 10\}$ ;  $m=\{2, 6\}$ ) in aqueous solution. The analysis of one- and 2D radial distribution functions and hydrogen bonds of these peptide/polymer segments shows that the secondary structure response to the inserted polyethylene chain is quite different for the three amino acid types. Upon frequent insertion (corresponding to a very short length  $n = 3$  of the peptide chain), we are unable to observe any meaningful secondary structure, independently of amino acid type and total length of the hybrid polymer.

Our results show that it is possible to locally suppress secondary structure motifs in peptides by means of inserting short synthetic polymer segments into the chain, and that this perturbation is restricted to about five to ten amino acids into the peptide.

## Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

## Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project-ID 189853844—SFB/TRR 102.

Open access funding enabled and organized by Projekt DEAL.

## Conflict of Interest

The authors declare no conflict of interest.

## Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## Keywords

hybrid bio/synthetic polymers, molecular dynamics simulations, peptide secondary structure

Received: November 10, 2022

Revised: February 6, 2023

Published online:

- [1] H. Elgabarty, M. Wolff, A. Glaubitz, D. Hinderberger, D. Sebastiani, *Phys. Chem. Chem. Phys.* **2013**, 15, 16082.
- [2] C. Allolio, N. Salas-Illanes, Y. S. Desmukh, M. R. Hansen, D. Sebastiani, *J. Phys. Chem. B* **2013**, 117, 9939.
- [3] T. Watermann, H. Elgabarty, D. Sebastiani, *Phys. Chem. Chem. Phys.* **2014**, 16, 6146.
- [4] C. Schiffmann, D. Sebastiani, *Phys. Status Solidi B* **2012**, 249, 368.
- [5] F. Hoffmann, J. Adler, B. Chandra, K. R. Mote, G. Bekçioğlu-Neff, D. Sebastiani, D. Huster, *J. Phys. Chem. Lett.* **2017**, 8, 4740.
- [6] M. V. Vener, A. V. Odinokov, C. Wehmeyer, D. Sebastiani, *J. Chem. Phys.* **2015**, 142, 215106.
- [7] G. Bekçioğlu-Neff, C. Allolio, Y. S. Desmukh, M. R. Hansen, D. Sebastiani, *ChemPhysChem* **2016**, 17, 1166.
- [8] S. Pylaeva, M. Brehm, D. Sebastiani, *Sci. Rep.* **2018**, 8, 13626.
- [9] F. Hoffmann, D.-W. Li, D. Sebastiani, R. Brüscheiler, *J. Phys. Chem. A* **2017**, 121, 3071.
- [10] M. Brehm, M. Pulst, J. Kressler, D. Sebastiani, *J. Phys. Chem. B* **2019**, 123, 3994.
- [11] R. Kiani, D. Sebastiani, P. Partovi-Azar, *ChemPhysChem* **2021**, 23, 202100519.
- [12] P. Partovi-Azar, D. Sebastiani, *Batteries Supercaps* **2019**, 2, 695.
- [13] S. Pylaeva, A. Böker, H. Elgabarty, W. Paul, D. Sebastiani, *ChemPhysChem* **2018**, 19, 2931.
- [14] M. Cheon, I. Chang, C. K. Hall, *Proteins: Struct., Funct., Bioinf.* **2010**, 78, 2950.
- [15] R. H. N. Kalaria, S. I. Harik, *J. Neurochem.* **1989**, 53, 1083.



- [16] H. Y. Zoghbi, H. T. Orr, *Annu. Rev. Neurosci.* **2000**, *23*, 217.
- [17] C. M. Lill, C. Klein, *Nervenarzt* **2017**, *88*, 345.
- [18] C. Soto, *FEBS Lett.* **2001**, *498*, 204.
- [19] E. M. Bradbury, B. G. Carpenter, H. Goldman, *Biopolymers* **1968**, *6*, 837.
- [20] A. I. Arunkumar, T. K. Kumar, T. Sivaraman, C. Yu, *Int. J. Biol. Macromol.* **1997**, *21*, 299.
- [21] A. I. Arunkumar, T. K. S. Kumar, C. Yu, *Int. J. Biol. Macromol.* **1997**, *21*, 223.
- [22] C. Soto, *J. Mol. Med.* **1999**, *77*, 412.
- [23] C. Soto, *CNS Drugs* **1999**, *12*, 347.
- [24] E. Karran, B. De Strooper, *Nat. Rev. Drug Discov.* **2022**, *21*, 306.
- [25] G. B. Frisoni, D. Altomare, D. R. Thal, F. Ribaldi, R. van der Kant, R. Ossenkoppele, K. Blennow, J. Cummings, C. van Duijn, P. M. Nilsson, P. Y. Dietrich, P. Scheltens, B. Dubois, *Nat. Rev. Neurosci.* **2022**, *23*, 53.
- [26] M. Fändrich, C. M. Dobson, *EMBO J.* **2002**, *21*, 5682.
- [27] A. Lavasanifar, J. Samuel, G. S. Kwon, *Adv. Drug Delivery Rev.* **2002**, *54*, 169.
- [28] A. Lalatsa, A. G. Schätzlein, M. Mazza, T. B. H. Le, I. F. Uchegbu, *J. Controlled Release* **2012**, *161*, 523.
- [29] N. Liu, B. Li, C. Gong, Y. Liu, Y. Wang, G. Wu, *Colloids Surf., B* **2015**, *136*, 562.
- [30] J. Freudenberger, W. H. Binder, *Macromolecules* **2019**, *52*, 4534.
- [31] A. Carlsen, S. Lecommandoux, *Curr. Opin. Colloid Interface Sci.* **2009**, *14*, 329.
- [32] J. P. Chen, I. M. Chu, M. Y. Shiao, B. R. S. Hsu, S. H. Fu, *J. Ferment. Bioeng.* **1998**, *86*, 185.
- [33] M. Metzke, N. O'Connor, S. Maiti, E. Nelson, Z. Guan, *Angew. Chem., Int. Ed.* **2005**, *44*, 6529.
- [34] K. T. Al-Jamal, W. T. Al-Jamal, J. T. Wang, N. Rubio, J. Buddle, D. Gathercole, M. Zloh, K. Kostarelos, *ACS Nano* **2013**, *7*, 1905.
- [35] P. A. Taylor, A. Jayaraman, *Annu. Rev. Chem. Biomol. Eng.* **2020**, *11*, 257.
- [36] M. B. Canalp, W. H. Binder, *RSC Adv.* **2020**, *10*, 1287.
- [37] M. B. Canalp, A. Meister, W. H. Binder, *RSC Adv.* **2019**, *9*, 21707.
- [38] R. B. Merrifield, *J. Am. Chem. Soc.* **1963**, *85*, 2149.
- [39] C. Ho Wong, S. C. Zimmerman, *Chem. Commun.* **2013**, *49*, 1679.
- [40] C. Bonduelle, *Polym. Chem.* **2018**, *9*, 1517.
- [41] H. Sun, A. Brik, *Acc. Chem. Res.* **2019**, *52*, 3361.
- [42] S. Mohan, O. S. Oluwafemi, N. Kalarikkal, S. Thomas, S. P. Songca, in *Recent Advances in Biopolymers* (Ed: F. K. Perveen), Intech Open, London **2015**, Ch. 3.
- [43] V. Siracusa, I. Blanco, S. Romani, U. Tylewicz, P. Rocculi, M. Dalla Rosa, *J. Appl. Polym. Sci.* **2012**, *125*, S2.
- [44] A. Panáček, L. Kvítek, R. Prucek, M. Kolář, R. Večeřová, N. Pizúrová, V. K. Sharma, T. Nevěčná, R. Zbořil, *J. Phys. Chem. B* **2006**, *110*, 16248.
- [45] M. U. Sankar, S. Aigal, S. M. Maliyekkal, A. Chaudhary, Anshup, A. A. Kumar, K. Chaudhari, T. Pradeep, *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 8459.
- [46] R. Rebelo, M. Fernandes, R. Figueiro, *Proc. Eng.* **2017**, *200*, 236.
- [47] G. Fuks, R. M. Talom, F. Gauffre, *Chem. Soc. Rev.* **2011**, *40*, 2475.
- [48] K. E. Gebhardt, S. Ahn, G. Venkatachalam, D. A. Savin, *J. Colloid Interface Sci.* **2008**, *317*, 70.
- [49] M. Morell, J. Puiggali, *Polymers* **2013**, *5*, 188.
- [50] S. Matsumura, A. R. Hlil, C. Lepiller, J. Gaudet, D. Guay, Z. Shi, S. Holdcroft, A. S. Hay, *J. Polym. Sci. Part A Polym. Chem.* **2008**, *46*, 7207.
- [51] S. M. Kelly, T. J. Jess, N. C. Price, *Biochim. Biophys. Acta - Proteins Proteomics* **2005**, *1751*, 119.
- [52] J. Kong, S. Yu, *Acta Biochim. Biophys. Sin.* **2007**, *39*, 549.
- [53] W. K. Surewicz, H. H. Mantsch, D. Chapman, *Biochemistry* **1993**, *32*, 389.
- [54] A. Yaron, E. Katchalski, A. Berger, G. D. Fasman, H. A. Sober, *Biopolymers* **1971**, *10*, 1107.
- [55] M. Rinaudo, A. Domard, *J. Am. Chem. Soc.* **1976**, *98*, 6360.
- [56] G. Bussi, *Mol. Phys.* **2014**, *112*, 379.
- [57] L. Wang, R. Friesner, B. Berne, *J. Phys. Chem. B* **2011**, *115*, 9431.
- [58] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia, M. Parrinello, *Comput. Phys. Commun.* **2009**, *180*, 1961.
- [59] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, *Comput. Phys. Commun.* **2014**, *185*, 604.
- [60] H. J. C. Berendsen, D. van der Spoel, R. van Drunen, *Comput. Phys. Commun.* **1995**, *91*, 43.
- [61] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. Berendsen, *J. Comput. Chem.* **2005**, *26*, 1701.
- [62] N. Foloppe, A. D. MacKerell, *J. Comput. Chem.* **2000**, *21*, 86.
- [63] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kucsera, F. T. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kucsera, D. Yin, M. Karplus, *J. Phys. Chem. B* **1998**, *102*, 3586.
- [64] B. Hess, *J. Chem. Theory Comput.* **2008**, *4*, 116.
- [65] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.* **1996**, *14*, 33.
- [66] M. Brehm, B. Kirchner, *J. Chem. Inf. Model.* **2011**, *51*, 2007.
- [67] M. Brehm, M. Thomas, S. Gehrke, B. Kirchner, *J. Chem. Phys.* **2020**, *152*, 164105.

*Paper II: Reverse mapping of coarse grained polyglutamine conformations from PRIME20 sampling*

Thomas Kunze, Christian Dreßler, Christian Lauer, Wolfgang Paul and Daniel Sebastiani

**Reverse mapping of coarse grained polyglutamine conformations from PRIME20 sampling**

*ChemPhysChem* 2024, Vol. 25, e202300521.

For this paper the PRIME20 structures were supplied by Christian Lauer, under supervision of Prof. Wolfgang Paul. The conversion algorithm was coded by Prof. Christian Dreßler with support of me by supplying input- and reference-files as well as providing application testing. All calculations and analysis from the given P20 structures were performed by me under the supervision of Prof. Daniel Sebastiani. The main text was written by me in collaboration with Prof. Daniel Sebastiani. Christian Lauer, Prof. Wolfgang Paul and Christian Dreßler assisted in the writing by guidance and providing paragraphs regarding their own contributions.

I hereby confirm that the use of this article is compliant with all publishing agreements, as this article is under an open access licence: <https://creativecommons.org/licenses/by-nc-nd/4.0/>



# Reverse Mapping of Coarse Grained Polyglutamine Conformations from PRIME20 Sampling

Thomas Kunze,<sup>[b]</sup> Christian Dreßler,<sup>[a]</sup> Christian Lauer,<sup>[b]</sup> Wolfgang Paul,<sup>[b]</sup> and Daniel Sebastiani<sup>\*[b]</sup>

An inverse coarse-graining protocol is presented for generating and validating atomistic structures of large (bio-) molecules from conformations obtained via a coarse-grained sampling method. Specifically, the protocol is implemented and tested based on the (coarse-grained) PRIME20 protein model (P20/SAMC), and the resulting all-atom conformations are simulated using conventional biomolecular force fields. The phase space sampling at the coarse-grained level is performed with a stochastic approximation Monte Carlo approach. The method

is applied to a series of polypeptides, specifically dimers of polyglutamine with varying chain length in aqueous solution. The majority (> 70%) of the conformations obtained from the coarse-grained peptide model can successfully be mapped back to atomistic structures that remain conformationally stable during 10 ns of molecular dynamics simulations. This work can be seen as the first step towards the overarching goal of improving our understanding of protein aggregation phenomena through simulation methods.

## 1. Introduction

Proteins are one of the key constituents of life on our planet. Composed of specific amino acid sequences,<sup>[1,2]</sup> they perform a large part of bio-relevant functionality in all living organisms. On the other hand, protein malfunction is at the origin of numerous diseases, among many others, Alzheimer's,<sup>[3]</sup> Huntington's<sup>[4]</sup> and Parkinson's<sup>[5]</sup> disease. One of the problematic processes in this context is their unwanted aggregation, e.g., into amyloid fibers.<sup>[6,7]</sup>

This aggregation process, its local biochemical prerequisites, and also kinetic and mechanical aspects are the subject of an ongoing intense research effort.<sup>[8,9]</sup> In this context, computational methods are an important clue to the qualitative and quantitative understanding of the numerous individual elements of the aggregation process.<sup>[10]</sup> However, computational methods generally address only one particular step or one isolated question of the process, as there are no theoretical approaches that capture the vast complexity of the aggregation in a comprehensive way, i.e. with atomistic resolution on the picosecond timescale, chemical accuracy, hours of simulated times and including macroscopic effects like crowding.<sup>[11]</sup> There

are continuously attempts made in the theory community to "bridge" computational scales, be it length scales, time scales, or accuracy and chemical resolution levels. These attempts normally consist of combining two or more established methods from different regions on those scales, and the theoretical challenge is to yield a consistent description of the system of interest across these methods, meaning that the two distinct methods must be enabled to "hand over" the system forth and back in a consistent manner.

In this context, we present here a protocol that enables the transfer of biomolecular systems of intermediate size between two specific simulation methods which are based on slightly different resolution levels (atomistic versus coarse-grained structures) and different interaction potentials (biomolecular force fields versus hard sphere-type potentials). Therefore, part of the representability and transferability problems of the quasi-global coarse-grained (CG) sampling gets addressed by the local spatio-temporal phase space coverage of the classical force field MD simulations.<sup>[12–18]</sup>

Specifically, we combine atomistic molecular dynamics simulations with a Monte-Carlo sampling scheme based on the coarse-grained PRIME20 protein model. The difficulty of this combination of simulation methods is the loss of atomistic resolution in the PRIME20 scheme which needs to be reverted and the partial simplification of repulsive and attractive interactions which need to reintroduce the energetic and entropic contributions of the neglected degrees of freedom into the coarse-grained potential. Especially the use of implicit solvent for biomolecules on aqueous solution may lead to a thermodynamically incorrect weighting of conformations of different nature.

Both Monte Carlo and Molecular Dynamics (MD) simulation were extensively used in the past to study bio molecules.<sup>[19–24]</sup> Several hybrid approaches already combine these two methods, because Monte Carlo and MD simulations are highly complementary techniques.<sup>[25–31]</sup> While Monte Carlo methods are a

[a] Prof. Dr. C. Dreßler  
Institut für Physik, Ilmenau University of Technology  
Weimarer Straße 32, 98693 Ilmenau, Germany

[b] T. Kunze, C. Lauer, Prof. Dr. W. Paul, Prof. Dr. D. Sebastiani  
Faculty of Natural Sciences II,  
Martin-Luther University Halle-Wittenberg  
Von-Danckelmann-Platz 4, 06120 Halle, Germany  
E-mail: daniel.sebastiani@chemie.uni-halle.de

Supporting information for this article is available on the WWW under <https://doi.org/10.1002/cphc.202300521>

© 2024 The Authors. ChemPhysChem published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

suitable tool to probe large parts of the conformational space of bio molecules, MD simulations are able to calculate the local structure fluctuations and dynamics of a given peptide configuration. In this work, we will combine the coarse-grained polymer model PRIME20 which has successfully been used in Stochastic Approximation Monte Carlo simulations (P20/SAMC) and an all atom MD simulation. The coarse-grained Monte Carlo model can be used to identify a set of low energy structures, which is not possible from a classical MD trajectory due to the limited length of the simulations. All atom MD simulations starting from the structures obtained from the Monte Carlo method will reveal the full atomistic picture including, e.g., solvation by explicit water molecules. The dynamical properties, such as the evolution of the hydrogen bond network, can be studied in that way and atomistic MD simulation will automatically incorporate entropic contributions of degrees of freedom which had been averaged over in the coarse grained description. In this way, the molecular dynamics simulations will act as validation and a posteriori correction tool for the thermodynamic weighting function for configurations delivered by the Monte Carlo simulations.

There are successful examples for the combination of MD and MC methods. The Inverse Monte Carlo approach<sup>[16]</sup> or the Iterative Boltzmann Inversion<sup>[32]</sup> can produce coarse-grained parameters fitted to MD simulation properties such as radial distribution functions. These and similar such methods were successfully improved and used to study a variety of topics.<sup>[15,17,33–40]</sup>

## 2. Coarse-Grained Model

The atomistic description of AMBER03<sup>[41]</sup> follows the general force field approach. In order to compare this already established technique, we have to introduce the characteristics of the PRIME20 model.

The PRIME20 model is a 4-bead model, where each amino acid is represented by 3 backbone beads and 1 side chain bead, as shown in Figure 1. The backbone beads refer to the NH bead, the C<sub>α</sub> bead and the CO bead. They are located at the C<sub>α</sub> position, the C position and the N position, respectively. The

side chain bead R is located at the center of mass of the side chain, while its position and size is specific for the amino acid it represents. Here, we will focus on the parameters relevant for polyglutamine (PolyQ), which are obtained from the complete list of parameters for the PRIME20 model.<sup>[42]</sup>

Covalent bonds are represented as white sticks on the right side in Figure 1. They are modeled as infinite well potentials around an ideal bond length. The width of the well allows for bond length fluctuations  $\Delta$  of 2.375% from the ideal value:

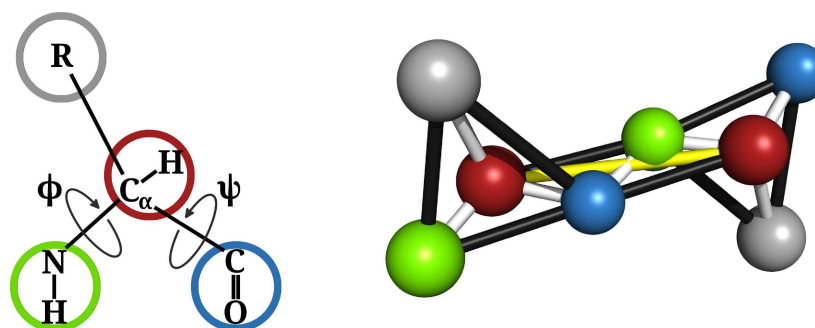
$$V_{\text{bond}}(d) = \begin{cases} 0 & \text{if } d \in [d_{\text{ideal}} - \Delta, d_{\text{ideal}} + \Delta] \\ \infty & \text{otherwise} \end{cases} \quad (1)$$

Here  $d$  represents the distance between two bonds,  $d_{\text{ideal}}$  is the ideal bond length and  $\Delta = 0.02375 d_{\text{ideal}}$ . PRIME20 utilizes pseudo-bonds between beads separated by two covalent bonds to stabilize bond angles, and between consecutive C<sub>α</sub> beads to keep the peptide in a *trans* configuration. Pseudo-bonds behave in the same way as covalent bonds and are represented by black and yellow sticks in Figure 1. Bond and pseudo-bond lengths for PolyQ are listed in Table 1.

Non-bonded bead interactions separate into two types. On the one hand, there are excluded volume interactions between multiple backbone beads and between backbone and side chain beads. They are modeled as hard-sphere (HS) repulsions. On the other hand, there are hydrophobic interactions between side chain beads as well as hydrogen bond formation between NH and the CO bead, which are modeled as semi-infinite square well potentials:

$$V_{\text{HS}}(d_{ij}) = \begin{cases} 0 & \text{if } d_{ij} > d_{ij}^{\text{HS}} \\ \infty & \text{otherwise} \end{cases} \quad (2)$$

$$V_{\text{sw}}(d_{ij}) = \begin{cases} 0 & \text{if } d_{ij} > d_{ij}^{\text{sw}} \\ \varepsilon_{ij} & \text{if } d_{ij}^{\text{HS}} < d_{ij} < d_{ij}^{\text{sw}} \\ \infty & \text{otherwise} \end{cases} \quad (3)$$



**Figure 1.** Geometry of the PRIME20 model. The backbone is represented by 3 beads: the NH group (green bead), the C<sub>α</sub> carbon (red bead) and the CO group (blue bead). The side chain is represented by the fourth bead (gray bead). Its position and size is specific for the individual type of amino acid. On the left the assignment of atoms to beads and the dihedral angles are shown. On the right the geometry of a PRIME20 dimer is shown. White sticks represent covalent bonds. Black and yellow sticks represent pseudo-bonds that stabilize the structure. The size of the beads is not true to scale.

**Table 1.** Bond and pseudo-bond lengths between beads of PolyQ in PRIME20. Here, the index  $i$  represents beads of the  $(i)$ th residue and the index  $i + 1$  represents beads of the  $(i + 1)$ th residue. Sizes in Å.

Bonds	NH <sub>i</sub> -C <sub>α,i</sub>	C <sub>α,i</sub> -CO <sub>i</sub>	CO <sub>i</sub> -NH <sub>i+1</sub>	R <sub>i</sub> -C <sub>α,i</sub>		
	1.46	1.51	1.33	1.60		
Pseudo-bonds	NH <sub>i</sub> -CO <sub>i</sub>	C <sub>α,i</sub> -NH <sub>i+1</sub>	CO <sub>i</sub> -C <sub>α,i+1</sub>	NH <sub>i</sub> -R <sub>i</sub>	C <sub>α,i</sub> -C <sub>α,i+1</sub>	CO <sub>i</sub> -R <sub>i</sub>
	2.45	2.41	2.45	2.50	3.80	2.56

where  $d_{ij}$  is the distance between beads  $i$  and  $j$ ,  $d_{ij}^{\text{HS}}$  is the hard-sphere diameter,  $d_{ij}^{\text{SW}}$  is the square-well interaction distance and  $\varepsilon_{ij}$  is the square-well depth. For interactions between side-chain beads, the 3 functional parameters ( $d_{ij}^{\text{HS}}$ ,  $d_{ij}^{\text{SW}}$  and  $\varepsilon_{ij}$ ) have specific values for each pair of interacting side-chain beads  $i$  and  $j$ . For hard-sphere repulsion interactions we use the Lorentz-Berthelot combining rule to calculate  $d_{ij}^{\text{HS}}$  from the beads  $d^{\text{HS}}$ . As side-chain diameters are only defined for side-chain-side-chain interactions, we use their self-interaction diameter for side-chain-backbone interactions. The self-interaction value of  $d_{ij}^{\text{HS}}$  and  $d_{ij}^{\text{SW}}$  are shown in Table 2.

For the formation of hydrogen bonds between NH and CO beads additional conditions, next to being within square-well interaction distance  $d_{ij}^{\text{SW}} = 4.5$  Å, have to be satisfied. Firstly, both beads considered for the hydrogen bond formation are not already involved in another hydrogen bond, and secondly there is an angle constraint between the N–H and the C–O vector.

In the model described up to this point, beads in close proximity along the chain will overlap in a way that prevents the formation of certain protein structures found in nature. To solve this shortcoming, *squeeze parameters* are introduced, which reduce the effective diameters of beads in close proximity along the chain. There are squeeze parameters for 10

different bead interactions. These parameters applied to side chain beads are specific for each amino acid and the glutamine parameters are shown in Table 3. For a detailed description of hydrogen bond formation as well as squeeze parameter implementation in the PRIME20 model we refer to the following Refs. [42,43].

The energy scale in the model is defined by the hydrogen bond strength  $\varepsilon_{\text{HB}} = -1$ . Side-chain interaction energies are given relative to  $\varepsilon_{\text{HB}}$  (see  $\varepsilon$  in Table 2). Physical energies  $E'$  and temperatures  $T'$  can be retrieved from the reduced quantities ( $E$  and  $T$ ) by assigning a value to  $\varepsilon_{\text{HB}}$ :  $E' = \varepsilon_{\text{HB}}E$  and  $T' = \varepsilon_{\text{HB}}T/k_{\text{B}}$ .

Both, the coarse-grained MC as well as the MD approach are established techniques, which can be applied separately for the investigation of the polypeptide aggregation. The combination of these two methods requires the careful design of mutual interfaces. In the first part of the manuscript, we will present a possible pathway to transfer coarse-grained structures of two polyglutamine strands into all atom geometries by a general applicable protocol. In the second part, we will start from the converted all atom structures to perform molecular dynamics simulations and discuss the stability of the P20/SAMC structures. The importance and relevance of establishing protocols for the back- and forth-conversion of structures between the coarse-grained model and all atoms MD simulation was already shown in various applications, especially for biomolecular and micellar systems.<sup>[44–58]</sup>

**Table 2.** Bead diameters and square-well parameters of PolyQ in PRIME20. Sizes in Å.

	NH	$\text{C}_{\alpha}$	CO	R
$d^{\text{HS}}$	3.3	3.7	4.0	3.6
$d^{\text{SW}}$	4.5	–	4.5	6.6
$\varepsilon$	–1.000	–	–1.000	–0.080

**Table 3.** Squeeze parameters and resulting reduced bead diameters for backbone bead interactions and interactions involving a polyglutamine side chain. Sizes in Å.

Interactions	$\text{C}_{\alpha,i}\text{-CO}_{i+1}$	$\text{C}_{\alpha,i}\text{-NH}_{i-1}$	$\text{CO}_i\text{-NH}_{i+2}$	$\text{NH}_i\text{-NH}_{i+1}$	$\text{CO}_i\text{-CO}_{i+1}$
original $d$	3.85	3.50	3.65	3.30	4.00
squeeze factor	1.1436	0.88	0.87829	0.8	0.7713
squeezed $d$	4.40286	3.08	3.2057585	2.64	3.0852
Interactions	$\text{C}_{\alpha,i-1}\text{-R}_i$	$\text{CO}_{i-1}\text{-R}_i$	$\text{NH}_{i+1}\text{-R}_i$	$\text{C}_{\alpha,i+1}\text{-R}_i$	$\text{CO}_{i-2}\text{-R}_i$
original $d$	3.65	3.8	3.45	3.65	3.8
squeeze factor	1.407	1.089	1.158	1.387	1.316
squeezed $d$	5.134	4.139	3.996	5.062	5.000

### 3. Results

#### 3.1. Conversion of coarse-grained into all atom structures

Our goal is to develop a protocol for the back-conversion of conformations obtained from the coarse-grained peptide interaction model PRIME20 into atomistic structures. The concept of our protocol is illustrated in Figure 2. The PRIME20 scheme provides simulation data which contains coordinates for the backbone carbon and nitrogen atoms, as well as the center of mass (COM) coordinates of the side chain residues of the peptide. The illustration in Figure 3 indicates these with red circles. The atoms labeled with green circles are not provided by the PRIME20 scheme, and the center of mass of a residue R of course lacks the coordinates for the individual atoms.

The concept of our back-mapping scheme is to derive the coordinates of the carbonyl oxygens and the nitrogen protons from the peptide backbone directly from the backbone carbon coordinates, by assuming equilibrium bond distances and a planar geometry with respect to the two adjacent backbone atoms. For the other atoms in the amino acid residues R, the coordinate of the initial carbon atom is computed in the same way, and the orientation of the residue is defined by the connection vector from the backbone  $C_\alpha$  atom to the center of mass from the PRIME20 simulation data (see Figure 3). For the

initial conversion step, we assume the molecular equilibrium conformation for the amino acid residue as such, so that the anchor point (via the center of mass) and the orientation (via the  $C_\alpha$ -COM vector) are sufficient to reconstruct the coordinates of the full residue.

The atomic coordinates computed in this way are tentative values, which lead to considerable misalignments in the peptide structure. The most common problem is that atoms from two adjacent amino acid residues are too close to each other. However, our protocol turns out to yield reasonable values for the start of a short geometry optimization cycle, in the sense that the standard optimization algorithms are able to respond to the close-proximity-misalignments and reorient the amino acid residues away from each other by maintaining the overall peptide structure as proposed by the coarse-grained scheme. It should be noted that while the resulting atomistic peptide geometry is technically possible, it is not for granted that this conformation is locally stable from a thermodynamical perspective. The latter aspect is addressed in a second stage within our back-mapping scheme.

To grasp the structural deviation from our back-mapping method, we have calculated the root-mean square displacement (RMSD) comparing the P20/SAMC resulted structure to the geometry optimized structures for the MD simulations. Similar to the structure conversion, we only compared the N,

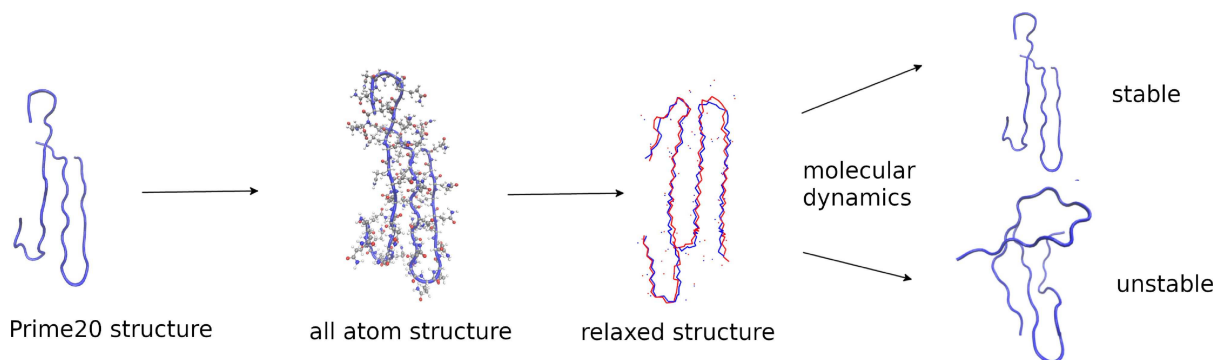


Figure 2. Visualization of the central process for the generation of data in this article.

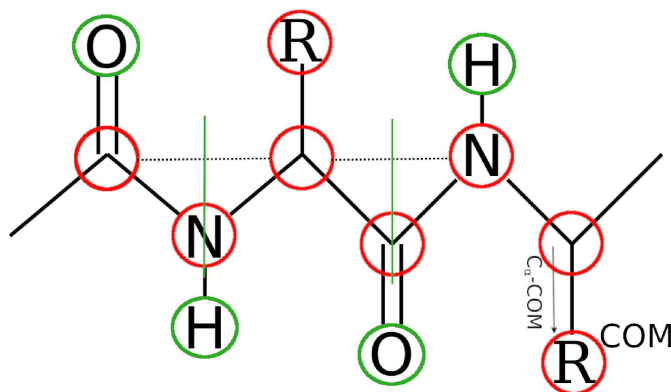


Figure 3. Scheme describing the conversion of the coarse-grained structures into all atom geometries. Red: atoms obtained from the coarse-grained PRIME20 model, green: atoms added by simple geometric considerations.

$C_{\text{Carbonyl}}$ ,  $C_{\alpha}$  atoms and the sidechain COM to the respective beads of the PRIME20 model. The RMSD data for all conformations is given in Table 4, as well as visual examples for change in structure caused by the energy optimization in Figure 4. We observe relatively similar and actually quite small displacements for all calculated structures, which on the one hand shows the back-mapping technique is reliable, on the other hand shows the PRIME20 structures are near a local energy minimum instead of being geometries that will relax considerably upon energy minimization.

### 3.2. Relaxed geometry of the all atom peptide structures

In order to cover a broad variety of systems for the back-mapping protocol, we generated a series of 21 solvated peptide dimers ( $\text{Glu}_n$ )<sub>2</sub> with varying length ( $n \in 14, 16, 18, 20, 22, 24, 26, 28, 36$ ) within the P20/SAMC simulation framework. For each of the nine dimer systems, up to four conformations were selected from the P20/SAMC scheme for inverse coarse-graining. In order to test the back-mapping protocol, we picked generally low-energy conformation along with a few extremely low-energy conformations, so that both “easy”, in the sense of typical aggregated peptide conformations, and “difficult” conforma-

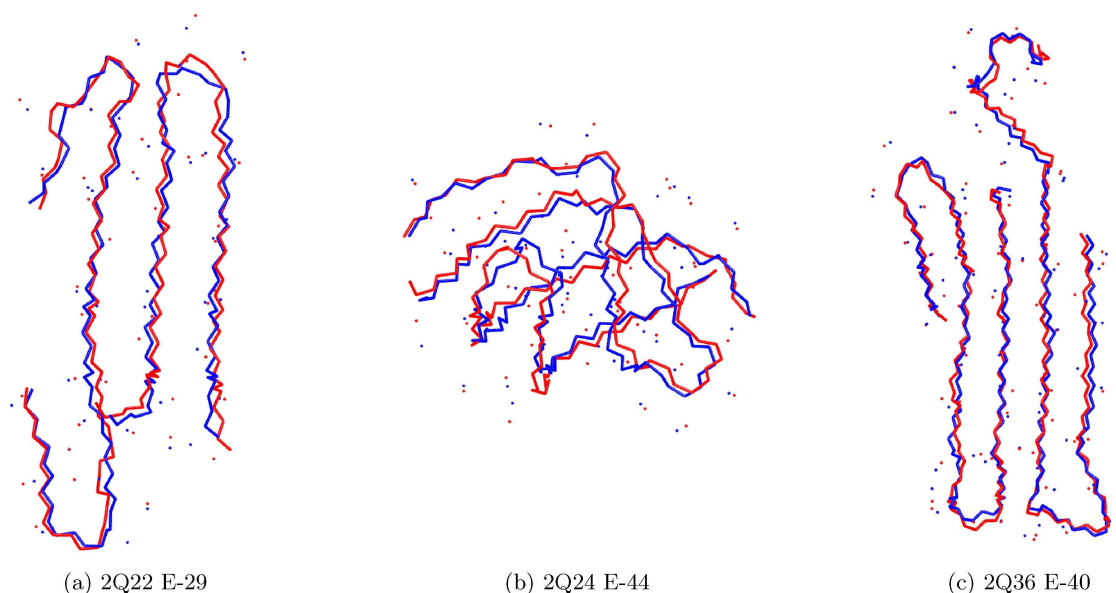
tions, in the sense of very uncommon peptide features, were processed and back-mapped to atomistic structures. The terms “easy” indicates typical aggregated conformations as was determined via analysis of the hydrogen bond contact probabilities within the PRIME20 scheme. “Difficult” conformations are of the lowest energy found in single P20/SAMC simulation runs. This makes them more likely to contain sterically demanding atomistic features such as highly rigid hydrogen bond networks. A complete list of the investigated peptide dimers including their chain lengths and energies calculated within the P20/SAMC model is given in Table 4. In the table, the canonical expectation value  $\langle U \rangle_T = 1/Z_U \sum_U U g(U) e^{-\beta U}$  of the configuration energy at room temperature is given. It is derived from the density of states  $g(U)$  of the PolyQ systems. One can see, that  $\langle U \rangle_T$  increases when going to systems of longer chain lengths. Performing MD simulations at room temperature on conformations of configuration energies far below  $\langle U \rangle_T$  has implications on the expected mechanical stability in MD. The further away from  $\langle U \rangle_T$  a configurations energy is, the more likely it will be unstable in the MD simulation. However, for the MD simulation run lengths of 10 ns (see SI), possible metastability in configurations can be found.

In the next step, we added explicit solvent molecules to the all atom structures and performed geometry optimizations.

**Table 4.** Overview of all calculated systems with the canonical expectation value of the configurational energy  $\langle U \rangle_T$  at room temperature. Furthermore, including MD energy properties, visual stability and a comparison of visual and  $\text{ACF}_{\text{hb}}$  stability. Green color shows agreement between both, red disagreement and black cases, where visual inspection was not fully distinguishable/ accessible (n.a.), for unstable (x) and stable (o) structures.

System	$\langle U \rangle_T (T=300 \text{ K})$	Visual Stability	Stability $hb_{\text{inter}}$	RMSD
–23.92	–1.68	unstable	x	0.94
–18.88	–3.01	stable	o	0.79
–21.88	–3.01	half-stable (n.a.)	x	0.74
–25.00	–3.40	stable	o	0.65
–26.36	–3.40	stable	o	0.78
–30.00	–3.40	unstable	x	0.83
–27.00	–8.05	stable	o	0.72
–30.00	–8.05	half-stable (n.a.)	o	0.92
–29.36	–20.48	stable	o	0.76
–30.00	–20.48	stable	o	0.57
–37.48	–20.48	stable	o	0.90
–30.32	–19.86	stable	o	0.72
–31.40	–19.86	stable	o	0.76
–33.92	–19.86	stable	o	0.73
–44.04	–19.86	unstable	x	0.90
–35.00	–31.48	stable	o	0.79
–44.96	–31.48	unstable	x	0.85
–36.12	–20.63	unstable	x	0.72
–38.24	–20.63	unstable	o	0.76
–40.00	–37.23	stable	o	0.75
–40.00	–37.23	stable	o	0.67





**Figure 4.** Comparison of the PRIME20 structure (red) and the backmapped, geometry optimized all atom structures (blue) used as starting point for the MD simulation. All PRIME20 beads are visualized and their respective MD atoms:  $C_{\alpha}$ , N, C and the sidechain centre of mass  $COM_{side}$ .

To this purpose the all atom peptide dimer structures were centered in a  $4\text{ nm} \times 4\text{ nm} \times 4\text{ nm}$  simulation box, and water molecules were added until a density between  $1.00\text{--}1.07\text{ g/cm}^3$  was reached. After solvating the peptide dimers, we performed geometry optimizations of the all atom structures using the program package GROMACS and the force field AMBER03.<sup>[41]</sup>

The calculation of force field energies was successful, in the sense that all calculations converged rapidly, for each of the converted all atom systems and the atomic positions of the peptide dimers were relaxed with respect to the minimization of the energy. For the comparison to the initial coarse grained P20/SAMC structures, the geometry optimized all atom peptide dimer structures were again reconverted into the coarse grained structures. The root mean square deviation of the coarse grained peptide dimers between before and after geometry optimization is given in Table 4.

In Figure 4, we show for three selected examples initial coarse-grained structures from the P20/SAMC calculations and the relaxed and back mapped all atom structures. The initially obtained coarse-grained peptide dimer structures and the geometry optimized all atom peptide dimer geometries are in good agreement.

In conclusion, both the back-conversion of the coarse-grained peptide dimer structures into atomic configurations and the subsequent local geometry optimizations with explicit aqueous solvation were successful and resulted in structurally acceptable conformations with a very good structural similarity to the original (coarse-grained) configurations.

All individual steps within our backmapping protocol are summarized in Figure 2. Our approach can be used for the automatic generation of fully solvated initial structures for all atom molecular dynamics simulations from coarse-grained P20/SAMC model geometries. In the future, we plan to extend our

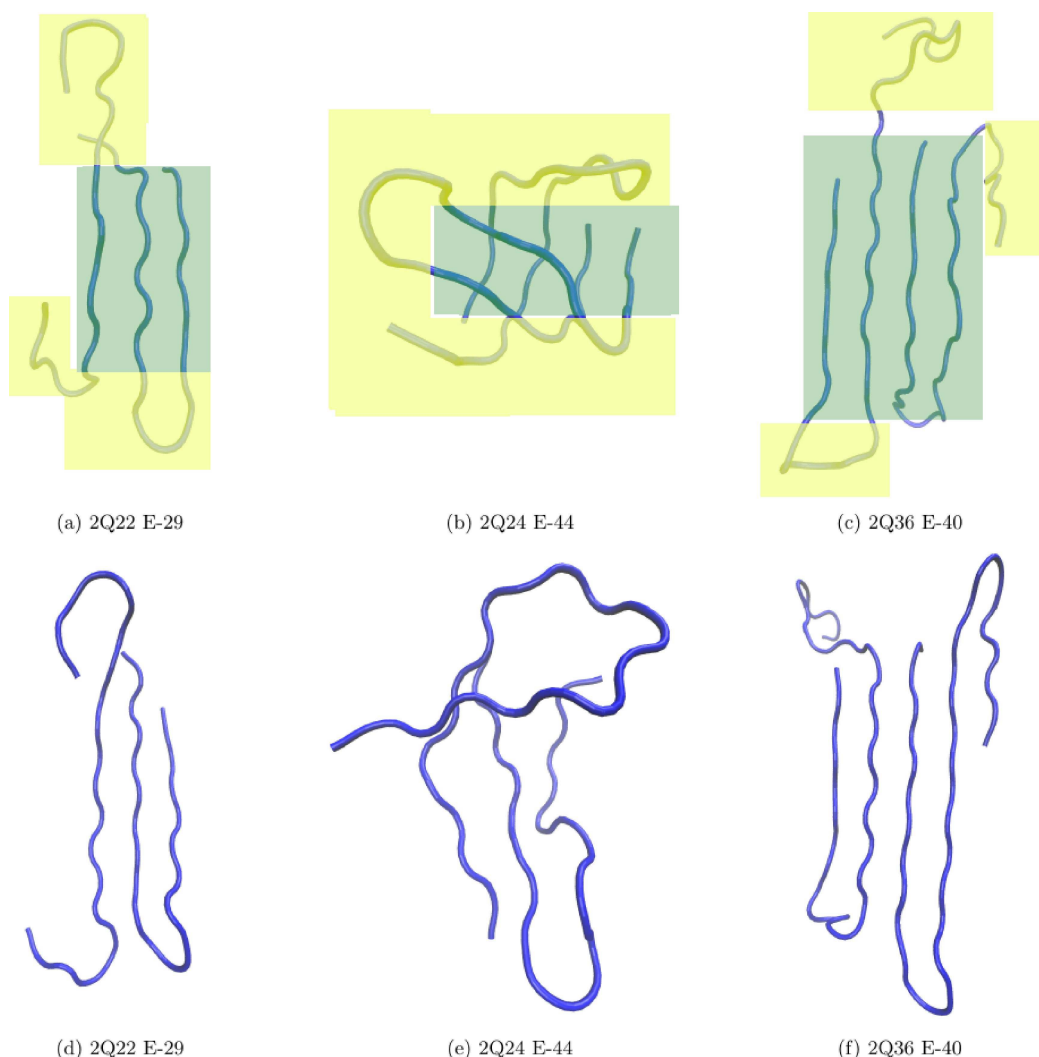
approach to peptide structures formed by other amino acids than glutamine.

### 3.3. Molecular dynamics simulation of initial dipeptide configurations obtained from the P20/SAMC calculations

We have visually inspected the peptide dimer structures provided by the P20/SAMC sampling before and after the molecular dynamics relaxations in order to characterize the structures on an empirical level as “stable” or “unstable”. We have focused on the strength of structural changes within the stronger hydrogen-bonded central regions of the peptides. The hydrogen bonding can be either at the peptide backbone level (NHO, both intramolecular and intermolecular, corresponding to beta-hairpin structures and collinear peptide strand conformations, respectively) or between amino acid sidechains (mainly intermolecular), see Figure 5.

Regarding the visual discrimination between “stable” and “unstable”, we have started by defining a “core” and a “peripheral” part of the dimer (green and yellow shaded areas in Figure 5). The core region is the part that contains direct peptide contacts, and would be the nucleation area for further aggregation of additional peptides. The peripheral regions are peptide segments that are fully solvated and/or localized outside the direct attachment region for additional peptides. The classification “stable” vs. “unstable” is now applied based on the structural integrity of the core region, i.e. its persistence after the short MD simulation.

The empirical classification of all 21 peptide dimer conformations in terms of “stable” or “unstable” is given in Table 4. The atomic coordinates of the first and last frame are also reported as raw data in the SI. A qualitative observation from



**Figure 5.** Visualization of the peptides at the start (a-c) and end (d-f) of the simulations.

this first visual classification step is that if the initial P20/SAMC structure incorporates parallel peptide strands (either intramolecular, in a hairpin conformation, or intermolecular by just parallel backbone segments) connected through hydrogen bonds, then the structure of the aggregated peptide strands remained stable during the atomistic relaxation run. Examples for such configurations are the structures Figure 5(a)/(d) and (c)/(f).

Another empirical observation from our visual inspections is that as soon as one of the coarse-grained peptides adopts a quasi-spherical shape, the dimer interaction is inhibited and the resulting configuration turns out to be unstable under molecular dynamics equilibration. A typical example for such a structure is represented in Figure 5(b)/(e), where both peptides lose their initial P20/SAMC conformation (Figure 5(b)) after relaxation 5(e). A possible explanation for this observation is the implicit treatment of solvation within the PRIME20 model. More compact (i.e. rather spherical) conformations tend to maximize

the intramolecular contacts of the peptide and to minimize the surface area towards the solvent. Within the explicit solvation used for the atomistic molecular dynamics simulations, the enthalpic benefit of peptide-solvent interactions is stronger, and thus the tendency to form compact structures is weaker. Independently of the solvation influence, the P20/SAMC calculations produces very low-energy structures, which are associated to exist at lower temperatures. However, the model P20 model is optimized for proteinogenic structures at room temperature, which could lead to unphysical structures at the low temperature range. This behaviour is reflected in comparing the  $\langle U \rangle_T$  to the actual system energy. In most cases, only the lowest energy was not stable during the MD simulations.

## 3.4. Hydrogen bond dynamics

As a complementary perspective regarding the dimer stability, we now look for a physical property that can be quantified a bit better compared to a mere visual inspection. We chose to look at the intermolecular hydrogen bonds between the peptides, in particular considering their temporal stability. Therefore, we calculated the autocorrelation function of all intermolecular hydrogen bonds and its time evolution. This function indicates how many of the initial hydrogen bonds (at  $t = 0$ ) have remained intact after a given time (e.g. during the full simulation of 10 ns). The data is shown in Figure 6 for a selection of dimer configurations. Each line corresponds to a given starting structure from the PRIME20 sampling, and those structures that have been visually characterized as “stable” are represented as full lines, while “unstable” structures are shown as dashed lines.

While there is a certain amount of numerical noise, a plateau value is reached for most of the dimers after around 3 ns. Afterwards, we observe fluctuations around those plateaus, which corresponds to hydrogen bond breaking and reformation processes. Interestingly, our initial empirical assessment in terms of stability is fully confirmed by this semi-quantitative analysis: all “stable” structures yield a highly preserved hydrogen bond network (i.e. little decay of the

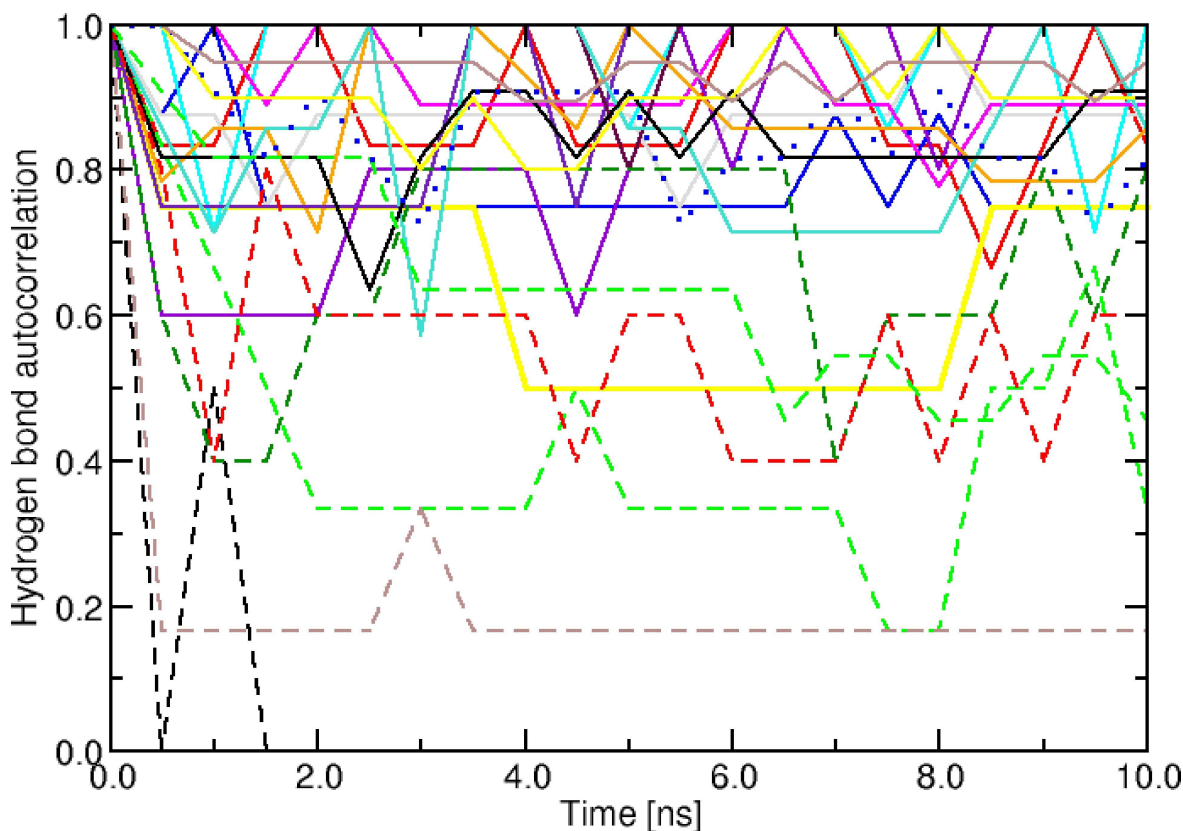
**Table 5.** Averages of all unstable or stable autocorrelation functions of the intermolecular hydrogen bonds.

Stability	Average
stable	0.84
unstable	0.51

autocorrelation function), while the “unstable” structures all exhibit a rapid decay and large fluctuations. The average values of the autocorrelation functions are listed in Table 5.

Here, we have looked at the hydrogen bonding autocorrelation functions merely with a qualitative eye, as a complementary semi-quantitative tool in addition to the classification of stable/unstable structures as discussed above. We have explicitly avoided to fit the hydrogen bond autocorrelation functions shown in Figure 6 to exponentials (yielding a numerical hydrogen bond lifetime), as we believe this would imply a quantitative relaxation time measure, which, however, is simply not reflected by the raw data (to our belief).

Additionally, we have also calculated the same autocorrelation functions but for intramolecular and intra/inter hydrogen bonds combined (all data given in the Supporting Information). However, with our focus on the peptide dimer stability, the intermolecular hydrogen bonds had most significance.



**Figure 6.** Intermolecular backbone hydrogen bonding autocorrelation function (percentage of hydrogen bonds of initial structure that are preserved) over the whole MD simulation time for all simulated peptides. Straight lines resemble stable structures, dashed lines unstable structures and dotted lines show the structure, which is visually unstable but hydrogen bonding suggests a stable structure.



### 3.5. Discussion and Outlook

The overall picture of our simulations confirms the reliability of the P20/SAMC method. From the thermodynamically most representative conformations generated from the P20/SAMC approach, physical meaningful configurations remained stable during the MD simulations, while unphysical peptide dimer structures were unstable. A particularly characteristic shape of highly unstable structures resembles a sphere, and these structures can be caused by the specific way in which the solvent is represented within the PRIME20 model. Instead of a chemically specific solvent interaction (which would depend on the actual chemical environment, i.e. whether there are actual particles in the vicinity), the PRIME20 model incorporates solvent effects by reducing the interaction strength between actual particles. As an example, the energetic strength of a hydrogen bond is chosen considerably lower than the normal chemical value of around 20 kJ/mol. Since the side chains are normally more solvent exposed than the peptide backbone, those hydrogen bonds carry an even lower energy contribution. The stable peptide aggregates formed in most cases extended hydrogen bond patterns between parallel peptide strands. While the categorization of coarse-grained structures into “stable” and “unstable” types is nontrivial from a quantitative point of view, it turned out that a more qualitative perspective is (in our opinion) sufficient to capture whether the conformation is essentially chemically reasonable.

Thus, the combination of the coarse-grained MC and MD simulations is suited to identify and investigate the local dynamics of stable aggregates of peptide strands. The P20/SAMC model allows to sample efficiently the entire phase space, while the all atom molecular dynamics simulations enable the probing of the geometric as well as the dynamic properties of the local minimum energy structures. As a side effect, molecular dynamics helps to validate the reliability of the P20/SAMC low energy structures by exclusion of unstable geometries from further analysis.

In the next step, we plan to extend our protocol for the conversion of coarse-grained into all atom structures towards peptides composed of other amino acids compared to glutamine.

The algorithm for the conversion of coarse-grained into all atom structures could be applied to all 21 PRIME20 polyglutamine structures without any changes. Subsequent relaxation of the coordinates of the fully solvated peptide dimer structures was possible using the GROMACS program package. The resulting geometry optimized structures were in good agreement with the initial P20/SAMC geometries. This is in particular remarkable, because the transferability to an all atom approach was not explicitly intended during the development of the PRIME20 model.

This work can be seen as the first step towards the overarching goal of improving the understanding of peptide aggregation using the PRIME20 model. In this development step, we have demonstrated how to convert coarse-grained P20/SAMC structures into all atom structures for MD simulations. Although the back mapping was possible, the resulting

coarse-grained structures could not be used for energy calculations within the PRIME20 model. The reason being the use of square-well potential and many cutoff values for inter- and intramolecular distances that have to be fulfilled by a peptide geometry to be a valid PRIME20 structure. Fluctuating configurations from finite temperature molecular dynamics simulations do often not fulfill these strict cut off criterions.

### 4. Conclusions

We have designed and implemented a reverse coarse-graining approach for the back-mapping of atomistic structures into conformations obtained from a united-atom scheme (PRIME20 approach) that is suitable for large-scale Monte-Carlo based conformational sampling. The reverse coarse-graining method is straightforward to implement for regular proteins/peptides and allows for a subsequent exploitation of atomistic molecular structures generated from the extensive conformational search done at the coarse-grained level.

We have validated the approach with a series of shorter peptide dimers via a conformational stability analysis using molecular dynamics simulations. It turns out that the majority – but not all – of the conformations delivered from the large-scale conformational sampling are “good” structures that remain stable for at least 10 ns of simulation. As a side result, we have found that a visual empirical assessment of the conformations yields stability estimates which are in good agreement with a more quantitative analysis in terms of the persistence of the intermolecular hydrogen bond network. All structures that were visually assessed as “unphysical” turned out to be unstable during the molecular dynamics simulations.

Our approach provides a further layer of atomistic detail to the coarse grained simulation of structurally challenging systems, combining the large-scale phase space sampling capability of the coarse-grained Monte Carlo method with the better accuracy and the atomistic resolution available at the molecular dynamics level.

### Computational Details

We used the PRIME20 model to perform coarse grained Monte-Carlo simulations for dimers of polyglutamine with chains length  $n$  between 14 and 36 amino acids. The simulation method we used is the Stochastic Approximation Monte Carlo (SAMC) method. It is an advanced flat-histogram Monte Carlo method which aims for a flat visitation histogram of energy states. In achieving this, it avoids getting stuck in local energy minima as can be the case with conventional Monte Carlo methods. SAMC achieves the even visitation of energy states by approximating the density of states (DOS)  $g(U)$  with respect to the potential energy  $U$ . The DOS describes the number of states in the system that belong to a given energy interval  $[U, U + \Delta U]$ . It then uses the DOS in its acceptance criterion: for an SAMC move from configuration  $x$  with the energy  $U(x)$  to configuration  $x'$  with the energy  $U(x')$ , the move is accepted with the probability  $\min(1, \tilde{g}(U(x'))/\tilde{g}(U(x)))$ .  $\tilde{g}(U)$  is the current estimate for the DOS. After the move is rejected or accepted,  $\tilde{g}(U)$  is updated according to  $\tilde{g}(U(x_{\text{new}})) = \tilde{g}(U(x_{\text{new}})) + \gamma_{\text{tr}}$ , where  $x_{\text{new}} = x'$  if the move was accepted and  $x_{\text{new}} = x$  if the move was

rejected. The modification factor  $\gamma_t$  goes to 0, for time  $t \rightarrow \infty$ .  $t$  is measured in MC steps. Additional conditions have to be met in order for the DOS to converge.<sup>[59–61]</sup> After a sufficiently accurate  $g(U)$  was obtained, further MC runs with a fixed DOS were performed. With the flat visitation histogram of energy states, snapshots at various energies were collected in multiple simulation runs of  $10^9$  MC steps.

Four different MC move types are used in the SAMC simulations. A local displacement move, which moves a single bead in a randomly chosen direction by a random distance, with a maximal displacement of 0.02 Å. A pivot rotation move, which randomly chooses a residue and rotates either its  $\Psi$  or  $\Phi$  angle by a random amount and direction. Furthermore, two moves are implemented to manipulate the relative position of the two chains in the system: a whole-chain rotation and a whole-chain translation move. After every move, the new configuration must be in agreement with the PRIME20s constraints on bond-lengths and excluded volumes. Similar to already successful calculations,<sup>[62]</sup> we simulated polyglutamine dimer systems with chain lengths  $N \in (14, 16, 18, 20, 22, 24, 26, 28, 36)$ .  $N$  refers to the number of residues in a single chain. For shorter chains ( $N \in (14, 16, 18, 20, 22, 24, 26)$ ) the cubic simulation box was of length  $L = 112.5$  Å and for longer chains ( $N \in (28, 36)$ ) the box was of length  $L = 150$  Å. The simulation box was periodic in all directions. This translates to a milli-molar concentration, which is close to in vitro experiments on polyglutamine aggregation.

The coarse-grained low-energy structures resulting from the PRIME20 simulations listed in Table 4 were translated into all-atom structures with both termini charged and were directly suitable for calculations. These structures were then explicitly solvated using the standard GROMACS<sup>[63,64]</sup> solvation tool; it should be noted that this solvation algorithm resulted in varying numbers of water molecules for different geometries. After an initial energy minimization (emtol=100; emstep=0.1; niter=20) for all atoms, a 10 ns NVT MD simulation with a 0.5 fs time step was performed at 300 K using velocity rescaling with 0.1 ps time constant, Lincs 4th order constraint<sup>[65]</sup> for covalent hydrogen bonds and the AMBER03<sup>[41]</sup> force field, while water interactions were represented by the TIP3P<sup>[66]</sup> water model. The Verlet cutoff-scheme and periodic-boundary conditions were used, and electrostatics were calculated with PME using potential-shift-Verlet for the coulomb modifier.

The energies and radii of gyration  $R_G$  were calculated by GROMACS tools,<sup>[63,64]</sup> and visualization was performed with VMD.<sup>[67]</sup> The first 2 ns were treated as initial equilibration and not used for GROMACS analysis. The hydrogen bond autocorrelation functions were calculated with a python script; the persistence of all hydrogen bonds determined in the initial structure was checked every 1 ns along the trajectory, by means of a combined distance/angle criterion. Note that we explicitly checked for temporary ruptures of hydrogen bonds, i.e. the autocorrelation function can increase again if a hydrogen bond is only shortly broken.

## Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 189853844 – TRR 102. Open Access funding enabled and organized by Projekt DEAL.

## Conflict of Interests

The authors declare no conflict of interest.

## Data Availability Statement

The data that support the findings of this study are openly available in Reverse mapping of coarse grained polyglutamine conformations from PRIME20 sampling at <https://github.com/thomascookies/Reverse-mapping-of-coarse-grained-polyglutamine-conformations-from-PRIME20-sampling>, reference number 0.

**Keywords:** backmapping · coarse-grained · molecular dynamics simulations · monte carlo simulation · peptide secondary structure · PRIME20

- [1] M. Huntley, G. B. Golding, *J. Mol. Evol.* **2000**, 51 (2), 131.
- [2] C. E. Pearson, R. R. Sinden, *Curr. Opin. Struct. Biol.* **1998**, 8 (3), 321.
- [3] R. H. N. Kalaria, S. I. Harik, *J. Neurochem.* **1989**, 53 (4), 1083.
- [4] H. Y. Zoghbi, H. T. Orr, *Annu. Rev. Neurosci.* **2000**, 23, 217.
- [5] C. M. Lill, C. Klein, *Nervenarzt* **2017**, 88 (4), 345.
- [6] C. Soto, *FEBS Lett.* **2001**, 498 (2–3), 204.
- [7] P. H. Nguyen, A. Ramamoorthy, B. R. Sahoo, J. Zheng, P. Faller, J. E. Straub, L. Dominguez, J. E. Shea, N. V. Dokholyan, A. de Simone, B. Ma, R. Nussinov, S. Najafi, S. T. Ngo, A. Loquet, M. Chiricotto, P. Ganguly, J. McCarty, M. S. Li, C. Hall, Y. Wang, Y. Miller, S. Melchionna, B. Habenstein, S. Timr, J. Chen, B. Hnath, B. Strodel, R. Kaye, S. Lesné, G. Wei, F. Sterpone, A. J. Doig, P. Derreumaux, *Chem. Rev.* **2021**, 121 (4), 2545.
- [8] A. M. Morris, M. A. Watzky, R. G. Finke, *Biochim. Biophys. Acta Proteins Proteomics* **2009**, 1794 (3), 375.
- [9] J. A. Housmans, G. Wu, J. Schymkowitz, F. Rousseau, *FEBS J.* **2023**, 290 (3), 554.
- [10] S. Navarro, S. Ventura, *Curr. Opin. Struct. Biol.* **2022**, 73, 102343.
- [11] R. O. Dror, R. M. Dirks, J. P. Grossman, H. Xu, D. E. Shaw, *Annu. Rev. Biophys.* **2012**, 41 (1), 429.
- [12] D. Rosenberger, M. Hanke, N. F. Van der Vegt, *Eur. Phys. J. Spec. Top.* **2016**, 225 (8–9), 1323.
- [13] H. J. Risselada, S. J. Marrink, *Phys. Chem. Chem. Phys.* **2009**, 11 (12), 2056.
- [14] E. Brini, V. Marcon, N. F. Van der Vegt, *Phys. Chem. Chem. Phys.* **2011**, 13 (22), 10468.
- [15] D. Reith, M. Pütz, F. Müller-Plathe, *J. Comput. Chem.* **2003**, 24 (13), 1624.
- [16] A. P. Lyubartsev, A. Laaksonen, *Phys. Rev. E* **1995**, 52 (4), 3730.
- [17] S. Izvekov, G. A. Voth, *J. Phys. Chem. B* **2005**, 109 (7), 2469.
- [18] J. W. Mullinax, W. G. Noid, *J. Phys. Chem. C* **2010**, 114 (12), 5661.
- [19] M. Karplus, J. Kuriyan, *Proc. Natl. Acad. Sci. USA* **2005**, 102 (19), 6679.
- [20] M. Bendahmane, K. P. Bohannon, M. M. Bradberry, T. C. Rao, M. W. Schmidtke, P. S. Abbineni, N. L. Chon, S. Tran, H. Lin, E. R. Chapman, J. D. Knight, A. Anantharam, *Mol. Biol. Cell* **2018**, 29 (7), 834.
- [21] S. Sharma, M. Lindau, *Proc. Natl. Acad. Sci. USA* **2018**, 115 (50), 12751.
- [22] R. M. Henry, C. H. Yu, T. Rodinger, R. Pomès, *J. Mol. Biol.* **2009**, 387 (5), 1165.
- [23] L. K. Scarbath-Evers, S. Jähnigen, H. Elgabarty, C. Song, R. Narikawa, J. Matsik, D. Sebastiani, *Phys. Chem. Chem. Phys.* **2017**, 19 (21), 13882.
- [24] F. Hoffmann, J. Adler, B. Chandra, K. R. Mote, G. Bekçioğlu-Neff, D. Sebastiani, D. Huster, *J. Phys. Chem. Lett.* **2017**, 8 (19), 4740.
- [25] I. Kurisaki, S. Tanaka, *Phys. Chem. Chem. Phys.* **2022**, 24 (17), 10575.
- [26] M. S. Barhaghi, B. Crawford, G. Schwing, D. J. Hardy, J. E. Stone, L. Schwiebert, J. Potoff, E. Tajkhorshid, *J. Chem. Theory Comput.* **2022**, 18 (8), 4983.
- [27] H. J. Woo, A. R. Dinner, B. Roux, *J. Chem. Phys.* **2004**, 121 (13), 6392.
- [28] I. Y. Ben-Shalom, C. Lin, T. Kurtzman, R. C. Walker, M. K. Gilson, *J. Chem. Theory Comput.* **2019**, 15 (4), 2684.
- [29] M. S. Bodnarchuk, M. J. Packer, A. Haywood, *ACS Med. Chem. Lett.* **2020**, 11 (1), 77.
- [30] G. A. Ross, E. Russell, Y. Deng, C. Lu, E. D. Harder, R. Abel, L. Wang, *J. Chem. Theory Comput.* **2020**, 16 (10), 6061.
- [31] S. Pylaeva, A. Böker, H. Elgabarty, W. Paul, D. Sebastiani, *ChemPhysChem* **2018**, 19 (21), 2931.
- [32] D. Reith, M. Pütz, F. Müller-Plathe, *J. Comput. Chem.* **2003**, 24 (13), 1624.
- [33] W. G. Noid, J.-W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, H. C. Andersen, *J. Chem. Phys.* **2008**, 128, 24.
- [34] E. Brini, N. F. Van der Vegt, *J. Chem. Phys.* **2012**, 137, 154113.

- [35] P. Ganguly, N. F. A. Van der Vegt, *J. Chem. Theory Comput.* **2013**, 9 (12), 5247.
- [36] L. C. Jacobson, R. M. Kirby, V. Molinero, *J. Phys. Chem. B* **2014**, 118 (28), 8190.
- [37] J.-w. Shen, C. Li, N. F. Van der Vegt, C. Peter, *J. Chem. Theory Comput.* **2011**, 7 (6), 1916.
- [38] M. Langeloth, T. Sugii, M. C. Böhm, F. Müller-plathe, *J. Chem. Phys.* **2015**, 143, 243158.
- [39] S. Jain, S. Garde, S. K. Kumar, *Ind. Eng. Chem. Res.* **2006**, 45 (16), 5614.
- [40] C.-C. Fu, P. Kulkarni, S. Shell, G. Leal, *J. Chem. Phys.* **2012**, 137, 164106.
- [41] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, P. Kollman, *J. Comput. Chem.* **2003**, 24 (16), 1999.
- [42] A. Böker, W. Paul, *J. Phys. Chem. B* **2022**, 126 (38), 7286.
- [43] A. Böker, Ph.D. thesis, Martin-Luther-University Halle-Wittenberg, **2019**.
- [44] J. Peng, C. Yuan, R. Ma, Z. Zhang, *J. Chem. Theory Comput.* **2019**, 15 (5), 3344.
- [45] S. D. Peroukidis, D. G. Tsalikis, M. G. Noro, I. P. Stott, V. G. Mavrantzas, *J. Chem. Theory Comput.* **2020**, 16 (5), 3363.
- [46] M. Feig, P. Rotkiewicz, A. Kolinski, J. Skolnick, C. L. Brooks III, *Proteins Struct. Funct. Genet.* **2000**, 41 (1), 86.
- [47] B. Hess, S. Leo, N. Van der Vegt, K. Kremer, *Soft Matter* **2006**, 2 (5), 409.
- [48] A. P. Heath, L. E. Kavrakli, C. Clementi, *Proteins Struct. Funct. Bioinf.* **2007**, 68 (3), 646.
- [49] C. Peter, K. Kremer, *Soft Matter* **2009**, 5 (22), 4357.
- [50] S. M. Gopal, S. Mukherjee, Y.-m. Cheng, M. Feig, *Proteins Struct. Funct. Bioinf.* **2009**, 78 (5), 1266.
- [51] A. J. Rzepiela, L. V. Schäfer, N. Goga, H. J. Risselada, A. H. De Vries, S. J. Marrink, *J. Comput. Chem.* **2010**, 31 (6), 1333.
- [52] P. J. Stansfeld, M. S. P. Sansom, *J. Chem. Theory Comput.* **2011**, 7 (4), 1157.
- [53] P. Brocos, P. Mendoza-Espinosa, R. Castillo, J. Mas-Oliva, Á. Piñeiro, *Soft Matter* **2012**, 8 (34), 9005.
- [54] T. A. Wassenaar, K. Pluhackova, R. A. Bo, S. J. Marrink, D. P. Tieleman, *J. Chem. Theory Comput.* **2014**, 10 (3), 676.
- [55] L. E. Lombardi, M. A. Marti, L. Capece, *Bioinformatics* **2016**, 32 (8), 1235.
- [56] M. Machado, S. Pantano, *Bioinformatics* **2016**, 32 (10), 1568.
- [57] S. Poblete, S. Bottaro, G. Bussi, *Biochem. Biophys. Res. Commun.* **2018**, 498 (2), 352.
- [58] M. Shimizu, S. Takada, *J. Chem. Theory Comput.* **2018**, 14 (3), 1682.
- [59] F. Liang, *J. Stat. Phys.* **2006**, 122 (3), 511.
- [60] F. Liang, C. L. Liu, R. J. Carroll, *J. Am. Stat. Assoc.* **2007**, 102 (477), 305.
- [61] T. Shakhov, S. Zablotskiy, A. Böker, V. Ivanov, W. Paul, *Eur. Phys. J. Spec. Top.* **2017**, 226 (4), 705.
- [62] C. Lauer, W. Paul, *Macromol. Theory Simul.* **2023**, 2200075, 1.
- [63] H. J. C. Berendsen, D. Van der Spoel, R. Van Drunen, *Comput. Phys. Commun.* **1995**, 91 (1–3), 43.
- [64] D. Van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. Berendsen, *J. Comput. Chem.* **2005**, 26 (16), 1701.
- [65] B. Hess, *J. Chem. Theory Comput.* **2008**, 4 (1), 116.
- [66] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, 79 (2), 926.
- [67] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **1996**, 14 (1), 33.

Manuscript received: November 10, 2023  
Revised manuscript received: February 1, 2024  
Accepted manuscript online: February 5, 2024  
Version of record online: March 28, 2024

*Paper III: Assignment of a physical energy scale for the dimensionless interaction energies within the PRIME20 peptide model*

Thomas Kunze, Christian Lauer, Christian Dreßler, and Daniel Sebastiani

**Assignment of a physical energy scale for the dimensionless interaction energies within the PRIME20 peptide model.**

*ChemPhysChem* 2024, e202400592.

The PRIME20 structures were provided by Christian Lauer. The conversion of the structures (using the previously published algorithm) and all calculations were performed by me under the supervision of Prof. Daniel Sebastiani. The main text was written by me in collaboration with Prof. Daniel Sebastiani. Christian Lauer and Prof. Christian Dreßler assisted in the writing by providing guidance and paragraphs regarding their own contributions.

I hereby confirm that the use of this article is compliant with all publishing agreements, as this article is under an open access licence: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

The shown article is the accepted version, not the final version published by the journal.

## Accepted Article

**Title:** Assignment of a Physical Energy Scale for the Dimensionless Interaction Energies within the PRIME20 Peptide Model

**Authors:** Thomas Kunze, Christian Dressler, Christian Lauer, and Daniel Sebastiani

This manuscript has been accepted after peer review and appears as an Accepted Article online prior to editing, proofing, and formal publication of the final Version of Record (VoR). The VoR will be published online in Early View as soon as possible and may be different to this Accepted Article as a result of editing. Readers should obtain the VoR from the journal website shown below when it is published to ensure accuracy of information. The authors are responsible for the content of this Accepted Article.

**To be cited as:** *ChemPhysChem* **2024**, e202400592

**Link to VoR:** <https://doi.org/10.1002/cphc.202400592>

# Assignment of a Physical Energy Scale for the Dimensionless Interaction Energies within the PRIME20 Peptide Model

Thomas Kunze Christian Lauer Christian Dreßler Daniel Sebastiani\*

T. Kunze, C. Lauer, C. Dreßler, D. Sebastiani\*,  
Address: Martin-Luther University Halle-Wittenberg  
Faculty of Natural Sciences II  
Von-Danckelmann-Platz 4 06120 Halle/Saale  
Sachsen-Anhalt Germany  
Ilmenau University of Technology  
Email Address:  
thomas.kunze@chemie.uni-halle.de  
christian.dressler@tu-ilmenau.de  
christian.lauer@physik.uni-halle.de  
daniel.sebastiani@chemie.uni-halle.de

**Keywords:** *Backmapping, Coarse-Grained, Molecular Dynamics Simulations, Monte Carlo Simulation, Peptide Interactions, PRIME20*

We present a calibration scheme to determine the conversion factors from a coarse-grained stochastic approximation Monte Carlo approach using the PRIME20 peptide interaction model to atomistic force-field interaction energies at full explicit aqueous solvation. The conversion from coarse-grained to atomistic structures was performed according to our previously established inverse coarse-graining protocol. We provide a physical energy scale for both the backbone hydrogen bonding interactions and the sidechain interactions by correlating the dimensionless energy descriptors of the PRIME20 model with the energies averaged over molecular dynamics simulations. The conversion factor for these interactions turns out to be around 2kJ/mol for the backbone interactions, and zero for the sidechain interactions. We discuss these surprisingly small values in terms of their molecular interpretation.

## 1 Introduction

Protein malfunction can lead to various diseases including Alzheimer's,<sup>1</sup> Huntington's,<sup>2</sup> and Parkinson's<sup>3</sup> disease. One problem in this context is the unwanted aggregation of proteins, where the result of that process can lead to the formation of amyloid fibers.<sup>4,5</sup>

Computational methods play a crucial role in qualitatively and quantitatively understanding the numerous individual elements of the aggregation process.<sup>6</sup>

However, the complexity of aggregation requires the combination of multiple theoretical methods to achieve accuracy while maintaining reasonable timescales.

In our previous work,<sup>7</sup> we provided a protocol that allows the transfer of bio-molecular systems of intermediate size between two specific simulation methods. This approach combines two different resolution levels (atomistic vs coarse-grained) and two different interaction potentials (bio-molecular force fields vs hard-sphere potentials). Combining these two methods addresses the representability and transferability problems of the quasi-global coarse-grained (CG) sampling by local spatio-temporal phase space coverage of the classical force field molecular dynamics (MD) simulations.<sup>8–14</sup> In detail, our approach combines an MC sampling scheme based on the P20 protein model with MD simulations to regain atomistic accuracy by reintroducing energetic and entropic contributions neglected by the CG potential. Furthermore, explicit solvent interactions may result in a more thermodynamically accurate weighting of the conformations.

Both MC and MD simulations have been extensively used in the past to study biomolecules.<sup>15–20</sup> As they are highly complementary techniques, several hybrid approaches already combine these two methods.<sup>21–27</sup> Monte Carlo methods are a suitable tool for exploring large parts of the conformational space of biomolecules.

Meanwhile, MD simulations can model the local structural fluctuations and dynamics of a given peptide configuration. By starting from structures obtained from the Monte-Carlo method, the subsequent MD simulations will provide the atomistic view, further enhanced by explicit water solvation. This allows for the examination of the dynamic characteristics of hydrogen bond networks by automatically including the entropic effects of atomistic degrees of freedom.

## 2 Computational Methods

### 2.1 Stochastic Approximation Monte-Carlo Simulation

The Stochastic Approximation Monte Carlo (SAMC)<sup>28,29</sup> method, which was developed as a mathematical formulation of the Wang-Landau<sup>30</sup> algorithm, was used for the simulation of a *Glu*<sub>26</sub>-dimer. The objective of the SAMC is to achieve a flat visitation histogram of energy states. This approach avoids the problem of getting stuck in local energy minima, that can occur with standard MC simulations. The SAMC achieves an even visitation of energy states by approximating the microcanonical density of states (DOS)  $g(U)$  with respect to the potential energy  $U$ . The DOS describes the number of states in the system that belong to a given energy interval  $[U, U + \Delta U]$ . SAMC then uses the DOS in its acceptance criterion: for an SAMC move from configuration  $x$  with energy  $U(x)$  to configuration  $x'$  with the energy  $U(x')$ , the move is accepted with a probability of:

$$acc(x'|x) = \min \left( 1, \frac{\tilde{g}(U(x))}{\tilde{g}(U(x'))} \right), \quad (1)$$

with  $\tilde{g}(U)$  being the current estimate for the DOS. After the move is rejected or accepted,  $\tilde{g}(U)$  is updated according to:

$$\tilde{g}(U(x_{\text{new}})) = \tilde{g}(U(x_{\text{new}})) + \gamma_t, \quad (2)$$

where  $x_{\text{new}} = x'$  if the move was accepted and  $x_{\text{new}} = x$  if the move was rejected. The modification factor  $\gamma_t$  goes to 0 for time  $t \rightarrow \infty$ , according to:

$$\gamma_t = \min \left( \gamma_0, \frac{t_0}{t} \right), \quad (3)$$

with  $t$  being measured in MC steps. The convergence of the SAMC algorithm was proven when additional conditions were fulfilled.<sup>28,29,31</sup> Simulations were run until a sufficiently accurate  $g(U)$  was obtained, with  $\gamma_t < 10^{-7}$ . Afterwards, multiple production MC runs with a fixed DOS and over  $10^9$  MC steps each were performed to collect configuration snapshots over the system's entire energy range. Four different MC move types were used in the SAMC simulations. Firstly, a local displacement move, which moves a single bead in a randomly chosen direction by a random distance, with a maximal displacement of 0.02 Å. Secondly, a pivot rotation move, which randomly chooses a residue and rotates either its  $\Psi$  or  $\Phi$  angle by a random amount and direction. Additionally, two moves are implemented to manipulate the relative position of the two chains in the system: a whole-chain rotation and a whole-chain translation move. After every move, the new configuration must be in agreement with the PRIME20's constraints on bond lengths and excluded volumes. Similar to already successful calculations,<sup>32</sup> we simulated polyglutamine dimer systems with a chain consisting of 26 glutamine residues. A cubic simulation box with length  $L = 150$  Å was used, which was periodic in all directions. This translates to a millimolar concentration, which is close to in vitro experiments on polyglutamine aggregation.

In the PRIME20 model, there are peptide backbone-backbone interactions of amplitude one as well as sidechain-X interactions (X=backbone or sidechain) of amplitude 0.08. In the concept of this coarse-grained interaction model, no specific microscopic nature of these interactions is specified, which means both hydrogen bonding and hydrophobic interactions are represented by this effective interaction strength. In our system, however, all three interaction types (peptide backbone-backbone, sidechain-backbone or sidechain-sidechain) are actually hydrogen bonds. The PRIME20 interaction model contains two distinct types of intra- and inter-peptide interactions: backbone hydrogen bonds and sidechain interactions.



These interaction types contribute 1.0 and 0.08 arbitrary energy units to the PRIME20 total energy expression, respectively, for each molecular group that actually interacts in the local geometry of a given glutamine structure:

$$E_{P20} = -1N_{\text{backbone}} - 0.08N_{\text{sidechains}}. \quad (4)$$

In order to adequately sample this “space of interactions” contained in the ensemble of coarse-grained structures generated by the MC simulations, we have generated subsets of conformations in such a way that each pair of values for the amplitude of the two interaction types ( $N_{\text{backbone}}$ ,  $N_{\text{sidechains}}$ ) is well represented in the ensemble of configurations used as input for our inverse coarse-graining protocol.

## 2.2 Molecular Dynamics Simulation

In previous work, a protocol for the back-conversion of conformations obtained from the coarse-grained peptide interaction model PRIME20 to atomistic structures was developed. The PRIME20 scheme provides simulation data which contains coordinates for the backbone carbon and nitrogen atoms, as well as the center of mass (COM) coordinates of the side chain residues of the peptide, which are indicated by red circles in Fig. 1. The atoms labeled with green circles are not provided, however with our previously published algorithm, we derive the coordinates of the carbonyl oxygens and the nitrogen protons in the peptide backbone directly from the backbone carbon coordinates by assuming planar NH-C-CO geometry. For the sidechain R, which is only one bead provided in the PRIME20 model, the coordinate of the initial carbon atom is computed by adjacent NH and CO groups, and the orientation of the residue is defined by the connection vector from the backbone  $C_\alpha$  atom to the center of mass from the PRIME20 simulation data. We assume molecular equilibrium conformation for the amino acid residues, so that the anchor point (via the center of mass) and the orientation (via the  $C_\alpha$ -COM vector) are sufficient to reconstruct the coordinates of the full residue.

The atomic coordinates of the *Glu*<sub>26</sub>-dimer computed this way lead to considerable misalignments in the 3D structure of the peptides. The most common problem is that atoms from two adjacent amino acid residues are too close to each other. However, the protocol turned out to yield reasonable values for the start of a short geometry optimization cycle. The standard optimization algorithms are able to respond to close-proximity misalignments and reorient the amino acid residues away from each other while maintaining the overall peptide structure proposed by the coarse-grained scheme. It should be noted that while the resulting atomistic peptide geometry is technically possible, it is not guaranteed that this conformation is locally stable from a thermodynamical perspective. The latter aspect was addressed in our previous work of the back-mapping scheme.<sup>7</sup>

For each PRIME20 energy data point, a *Glu*<sub>26</sub>-dimer structure was randomly selected from the provided MC structure set and converted into an all-atom structure, similar to previous research. More specifically, the coarse-grained structures resulting from the PRIME20 MC simulations were translated into all-atom structures with both termini charged and were directly suitable for calculations. These structures were then explicitly solvated with 6700 water molecules using the standard GROMACS<sup>33,34</sup> solvation tool. After an initial energy minimization (emtol=100; emstep=0.1; niter=20) for all atoms, a 10 ns NVT MD simulation with a 0.5 fs time step was performed at 300 K using velocity rescaling with a 0.1 ps time constant, Lincs 4th order constraint<sup>35</sup> for covalent hydrogen bonds, and the AMBER03<sup>36</sup> force field, while water interactions were represented by the TIP3P<sup>37</sup> water model. The Verlet cutoff scheme and periodic boundary conditions were used, and electrostatics were calculated with PME using potential-shift Verlet for the Coulomb modifier.

As a reference simulation, 6700 water molecules were simulated with the same MD parameters, but with a slightly smaller box to achieve a similar density. The average energy obtained was -215346 kJ/mol. A short MD simulation of a single *Glu*<sub>26</sub>-peptide resulted in an average energy of -5392 kJ/mol. Therefore, our simulation with 6700 water molecules and 2 peptides has a reference energy of -226130 kJ/mol. This reference energy was used for visual clarity in our plots.



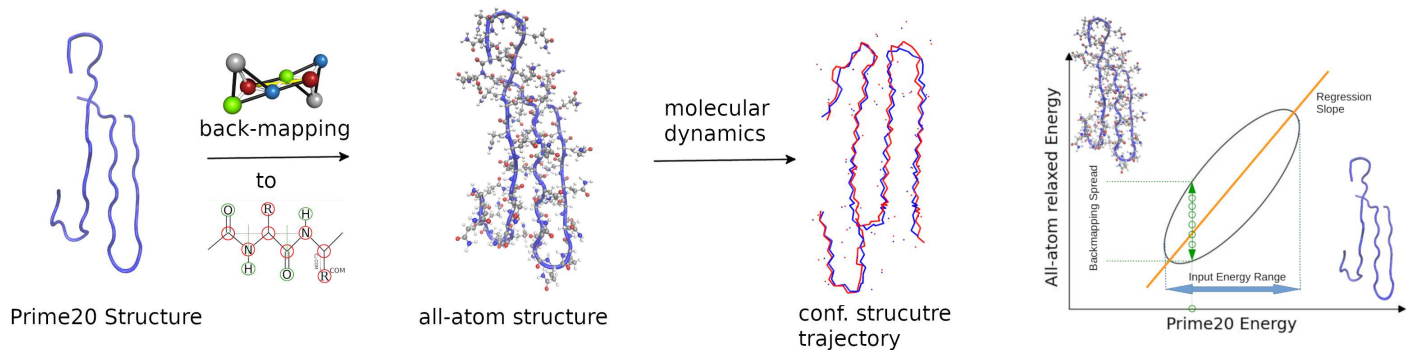


Figure 1: The central process for the generation of data in this article is visualized: starting by back-mapping<sup>7</sup> Prime20 structures to all-atom structures and running short MD simulations to compare the energies for both techniques.

### 3 Results

#### 3.1 Density of States of the coarse-grained conformational space

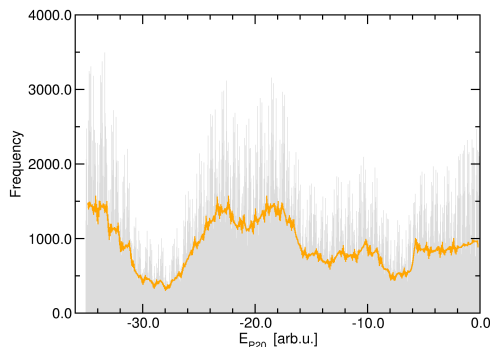


Figure 2: Distribution of the (dimensionless) energies  $E_{P20}$  of the ensemble of coarse-grained structures ( $N \approx 800000$ ) generated using the PRIME20 interaction model. The orange line is the running average (for an energy window of 0.5 arb.u.).

We have computed the energy histogram of the ensemble of initial coarse-grained structures that were generated with the flat-histogram Monte-Carlo sampling scheme at the PRIME20 level of theory (see Fig. 2). We use the dimensionless energy units provided by the PRIME20 interaction model, which combine inter-peptide backbone hydrogen bonding and side chain interaction energies with specific relative weights. Although the distribution is not strictly flat, it has no characteristic internal structure, and shows that the sampling protocol provides a sufficient number of conformations for any given energy value. In order to exclude any hidden bias in this distribution, we also analyzed its Fourier transform (see SI for details), which revealed no particular spectral features.

In the PRIME20 model, the total energy ( $E_{P20}$ ) is composed of a larger contribution due to backbone hydrogen bonding and a smaller contribution due to side chain interactions, with a non-trivial commensurability, see the Methods section. To verify the balanced distribution of the weaker side-chain interactions contributions, we have additionally calculated the density of states of the total PRIME20 energies  $E_{P20}$  *modulo* the hydrogen bonding contributions (i.e. considering only the side chain interactions, represented by the fractional part of  $E_{P20}$ ). This projected density of states is given in the Supporting Information. Again, this distribution function shows no distinct spectral peaks, indicating an adequate statistical representation of all amplitudes for this weaker interaction type.

This preliminary statistical analysis of the underlying conformational space of our peptide dimer in terms of its energy distribution shows that there are no “forbidden” energy ranges with low densities of states. In particular, also the thermodynamically unfavourable conformations (i.e. those with energies near  $E_{P20}=0$  arb.u.) are well represented in the manifold. In this sense, we are confident that our basic data is reasonably unbiased and does not need to be weighted or corrected a posteriori. Hence, we conclude that the initial Monte-Carlo sampling at the PRIME20 level can be considered converged for our purposes.

### 3.2 Energy correlation between coarse-grained and atomistic models

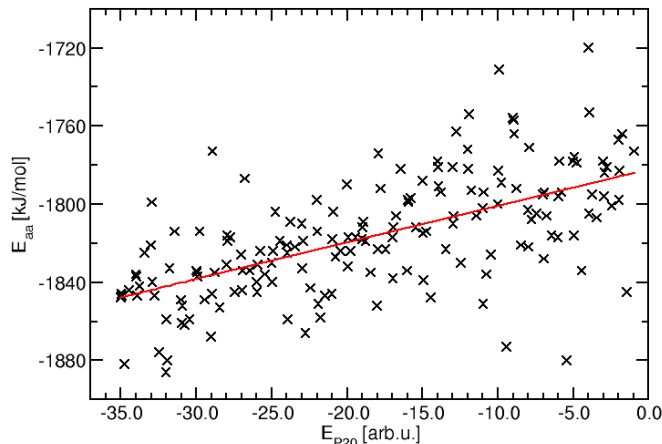


Figure 3: Correlation of  $E_{P20}$  and  $E_{aa}$  for the hydrogen bond series with linear regression analysis. The slope of the regression is  $1.9 [kJ \cdot mol^{-1} / arb_{P20}.u.]$ .

The central goal of this work is to investigate the correlation between the dimensionless energies of the coarse-grained peptide structures generated under the PRIME20 model and the (regular dimensional) energies of the locally relaxed all-atom conformations. The all-atom energies are obtained from our reverse coarse-graining protocol<sup>7</sup> by means of a preliminary geometry-optimization and a subsequent 10 ns molecular dynamics simulation (at constant ambient temperature) at the all-atom force-field level. The instantaneous total energy values during the MD simulation are then averaged, yielding the final energy value at the all-atom level. Such a correlation allows to assign an effective physical energy value to the dimensionless energy scale used by the coarse-grained interaction model.

The raw correlation as well as the linear fit are shown in Fig. 3. Clearly, a positive correlation is recognizable, i.e. structures with more positive PRIME20 energies correspond to conformations with more positive force-field energies. However, the variations of the final all-atom energies are quite large, and even exceed the systematic dependence of  $E_{aa}$  on  $E_{P20}$ . It should be noted that there is of course also a statistical error bar associated with every single data point  $E_{aa}$ ; this aspect will be addressed later on in this article.

The correlation between coarse-grained PRIME20 energies  $E_{P20}$  and locally relaxed all-atom conformations  $E_{aa}$  is obtained as  $1.9 \text{ kJ/mol}$  per PRIME20 energy unit. At first sight, this value is considerably lower than the typical energy of a hydrogen bond of  $20 \text{ kJ/mol}$  (one energy unit in the PRIME20 interaction model corresponds to one intermolecular  $\text{NH} \cdots \text{OC}$  peptide hydrogen bond). However, the reference situation is not simply a broken peptide hydrogen bond. Instead, both the NH and CO hydrogen bonding partner will establish hydrogen bonds to liquid water from the solvent, but in turn break a water-water hydrogen bond. The true situation is of course even more involved, as the coordination

numbers of the water molecules may differ between the two situation (i.e. a water can donate two hydrogen bonds to other water molecules, but is less likely to bond ).

Thus, one PRIME20 energy unit corresponds to the difference between these two competing situations:

$$\Delta E_{P20} = E_{MD}(\text{Pep} \cdots \text{Pep}) + E_{MD}(\text{H}_2\text{O} \cdots \text{H}_2\text{O}) - 2E_{MD}(\text{Pep} \cdots \text{H}_2\text{O}). \quad (5)$$

Hence, a comparably small value of 1.9 kJ/mol makes perfect sense as the effective intermolecular peptide hydrogen bond energy difference.

However, the problem remains that different coarse-grained structures with virtually no energy difference (e.g. one PRIME20 energy unit) typically yield all-atom conformations that exhibit considerable energetic deviations (of ten times the corresponding all-atom energy difference, i.e.  $10 \times 2 \text{ kJ/mol} = 20 \text{ kJ/mol}$ ). This variability represents a challenge for the physical interpretation of the energy landscape generated and sampled by the PRIME20 interaction model; most likely, the coarse-graining approach suffers from not recognizing many of the more subtle energetic effects of structural deformations of the peptides. Examples of such effects include torsional and angular potentials along the peptide backbone chain, but also steric effects related to the actual size of solvent molecules (e.g. an area with space for 1.9 water molecules can only be filled with one water molecule, which in an all-atom description will result in a force that tending to reduce the volume of that area).

### 3.3 Sidechain Interaction

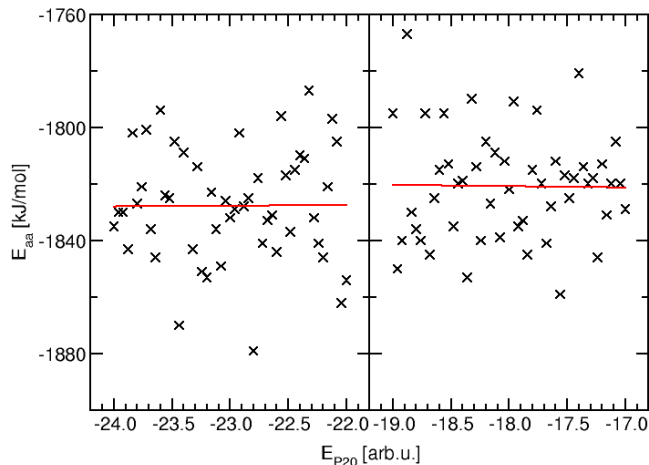


Figure 4: Correlation between the coarse-grained ( $E_{P20}$ ) and atomistic ( $E_{aa}$ ) energies for a series of conformations with identical backbone hydrogen bonding states (at the coarse-grained level, here corresponding to 22 and 17 hydrogen bonds, respectively, for the left and right plots). In turn, the number of sidechain interactions varies and corresponds to  $-1/12$  units of  $E_{P20}$  per sidechain interaction.

In Fig. 4, we plot the energies at the coarse-grained and at the atomistic level for a series of conformations that have an identical number of backbone hydrogen bond interactions, characterized by “large” energy steps (one arbitrary unit) at the coarse-grained level, but different numbers of sidechain interactions, characterized by “small” energy steps ( $1/12$  of an arbitrary unit). Each of these conformations was processed through our inverse coarse graining protocol, so that each atomistic energy represents an average value obtained during a 10 ns molecular dynamics simulation. Here, we have arbitrarily chosen two specific values for the number of backbone hydrogen bond interactions (22 and 17, respectively, for the two plots in Fig. 4).

We observe a correlation between the (coarse-grained) sidechain interactions and the atomistic energies with practically zero slope. While the atomistic energies are statistically quite scattered with a distribution width of around  $\pm 20 \text{ kJ/mol}$ , the correlation slope is below  $1 \text{ kJ/mol}$  per  $E_{P20}$  energy unit in both

cases. The reason for this weak correlation is that it is statistically challenging to detect a correlation of the order of 1/12 of a hydrogen bond (identified in the previous section as corresponding to an atomistic energy of 2 kJ/mol, resulting in 0.2 kJ/mol for the expected sidechain interaction) in the presence of numerical noise of the order of 20 kJ/mol. From a chemical perspective, even the short MD simulations within our equilibration protocol (10 ns) result in conformational changes that are energetically more important than a single sidechain interaction energy. Hence, we consider the actual energetic conversion factor of the PRIME20 sidechain interactions to be zero. Notably, this does not mean that the sidechain interactions have zero interaction strength, but rather that the correlation of the PRIME20 interaction scheme with the true (atomistic) interaction energy is small.

### 3.4 Analysis of statistical errors / numerical uncertainties of the atomistic MD simulations

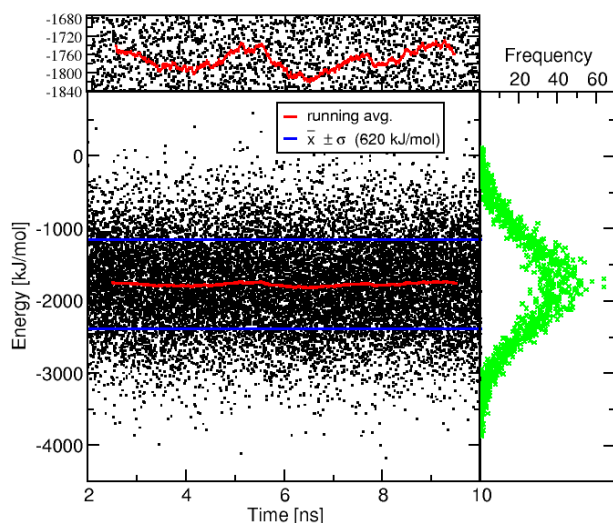


Figure 5: Energy fluctuation during the MD simulation was analyzed using histograms, running averages and the one-sigma interval ( $\bar{x} \pm \sigma$ ).

Fig. 5 shows the time evolution of the total energy during a typical MD simulation. The energy fluctuates in a range of around 4000 kJ/mol, while the one-sigma interval is about 1200 kJ/mol. Since our goal is to evaluate the conversion relationship between the P20 energies and the MD energies, we first want to investigate the accuracy of the determination of the average energy based on a 10 ns MD simulation. In other words, we want to check how effective is the averaging of the considerable instantaneous total energy fluctuations during the MD runs, compared to the energy variations between the different P20 structures. As a simple estimate of the numerical error due to the averaging of the discrete energy values, we calculated the energy averages for a randomly selected subset of the MD snapshots with about half of the data set size.

We also calculated for the same data set the standard error of the mean (SEM  $\sigma_{\bar{x}}$ ), which is given by:

$$\sigma_{\bar{x}} \approx \frac{\sigma_x}{\sqrt{N}} \quad (6)$$

Since our energy data points are highly correlated at short times, it is not appropriate to use the number of MD steps for  $N$ . Instead, we propose to use the number of typical hydrogen bond lifetimes (10 ps for relaxation of the hydrogen bond network of liquid water) for this quantity; for a simulation time of 10 ns, this results in  $N=10\text{ns}/10\text{ps}=1000$ . The use of the longer relaxation times corresponding to the peptide groups would lead to a "more-than-local equilibration", however our idea behind this entire backmapping approach is to leave the overall structure (as delivered by the coarse-grained model) unchanged as much as possible (i.e. doing only a local equilibration to avoid steric incompatibilities).

Using our standard deviation  $\sigma = 620 \text{ kJ/mol}$  and the resulting  $N=1000$  gives us an estimated energy error:

$$\sigma_{\bar{x}} \approx \frac{620 \text{ kJ/mol}}{\sqrt{1000}} \approx 19.6 \text{ kJ/mol} \quad (7)$$

Thus, the formal statistical uncertainty for the calculation of average all-atom energy for a given P20 starting structure during the MD simulation is obtained as  $\pm 20 \text{ kJ/mol}$ . For comparison, using instead a time interval of 1 ps for the assumed lifetime of a given MD simulation would give an estimated energy error of only  $6.2 \text{ kJ/mol}$ . It is interesting to note, that another estimate can be obtained visually from the running average (red line) in Fig. 5. A closer inspection reveals fluctuations of about  $\pm 40 \text{ kJ/mol}$ , which is in a similar range to the estimate from Eq. 7.

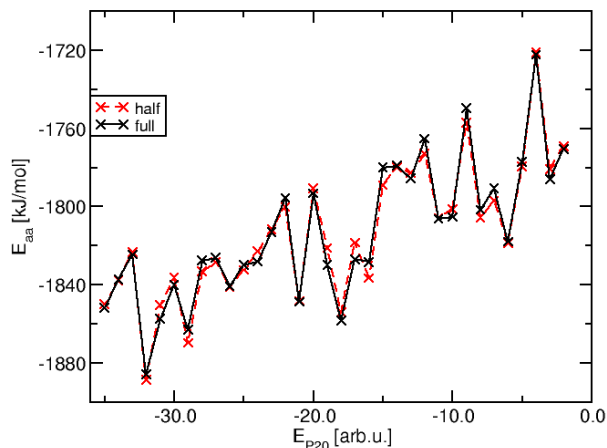


Figure 6: Validation of the statistical averaging accuracy from our MD simulations.

In Fig. 6 the average atomistic total energy for a series of PRIME20 converted structures is shown for two averaging protocols: first using all MD snapshots (black) or only half of the available number of snapshots (red), selected randomly from the entire MD trajectory. This comparison is intended to illustrate the accuracy of the statistical averaging from a different perspective.

The averaging error this way turns out to be considerably smaller than the statistical error obtained previously (see Fig. 5) derived from the explicit energy distribution. Therefore, we believe that our energy averaging protocol based on the 10 ns MD simulations is sufficient to yield converged average energy values with an accuracy around  $5 \text{ kJ/mol}$ . We want to stress, that this is not an accurate statistical error but rather a consistency check that not obvious bias is generated by our approach.

## 4 Conclusion

We have determined the energy scale conversion factors from the coarse-grained protein interaction model PRIME20 to all-atom energies at the common force-field level using explicit solvation and local conformational equilibration. Using a previously established protocol for the structure conversion,<sup>7</sup> we have generated an ensemble of conformations using stochastic approximation Monte Carlo sampling. We subsequently computed atomistic energies for each value of the coarse-grained interaction descriptor (peptide backbone hydrogen bonding and sidechain interaction) by averaging over a set of about ten different coarse-grained conformations, equilibrating each conformation for about 10 ns via molecular dynamics simulations.

Our central result is that the atomistic physical energy scale for the backbone hydrogen bonding interaction of the PRIME20 model (which uses dimensionless energy units) is obtained as 2 kJ/mol per backbone interaction and virtually zero per sidechain interaction. This energy scale appears comparably small at first sight but is explained in terms of its interpretation as relative energies with respect to competing interactions (peptide to solvent). Our results confirm previous findings about salt bridges in peptides.<sup>38</sup> We validate our findings by carefully estimating our statistical errors in the determination of the average atomistic energy values using several statistical techniques. Eventually, our results will allow for an insightful interpretation of structures generated using the coarse-grained PRIME20 interaction model.

## 5 Acknowledgement

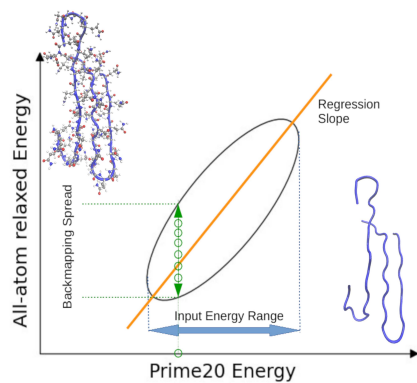
We would like to thank Prof. Wolfgang Paul for his help and guidance regarding this project.

## References

- <sup>1</sup> R. H. N. Kalaria, S. I. Harik, *J. Neurochem.* **1989**, *53*, 4 1083.
- <sup>2</sup> H. Y. Zoghbi, H. T. Orr, *Annu. Rev. Neurosci.* **2000**, *23* 217.
- <sup>3</sup> C. M. Lill, C. Klein, *Nervenarzt* **2017**, *88*, 4 345.
- <sup>4</sup> C. Soto, *FEBS Lett.* **2001**, *498*, 2-3 204.
- <sup>5</sup> P. H. Nguyen, A. Ramamoorthy, B. R. Sahoo, J. Zheng, P. Faller, J. E. Straub, L. Dominguez, J. E. Shea, N. V. Dokholyan, A. de Simone, B. Ma, R. Nussinov, S. Najafi, S. T. Ngo, A. Loquet, M. Chiricotto, P. Ganguly, J. McCarty, M. S. Li, C. Hall, Y. Wang, Y. Miller, S. Melchionna, B. Habenstein, S. Timr, J. Chen, B. Hnath, B. Strodel, R. Kayed, S. Lesné, G. Wei, F. Sterpone, A. J. Doig, P. Derreumaux, *Chem. Rev.* **2021**, *121*, 4 2545.
- <sup>6</sup> S. Navarro, S. Ventura, *Curr. Opin. Struct. Biol.* **2022**, *73* 102343.
- <sup>7</sup> T. Kunze, C. Dreßler, C. Lauer, W. Paul, D. Sebastiani, *ChemPhysChem* **2024**, *25*, e202300521.
- <sup>8</sup> D. Rosenberger, M. Hanke, N. F. van der Vegt, *Eur. Phys. J. Spec. Top.* **2016**, *225*, 8-9 1323.
- <sup>9</sup> H. J. Risselada, S. J. Marrink, *Phys. Chem. Chem. Phys.* **2009**, *11*, 12 2056.
- <sup>10</sup> E. Brini, V. Marcon, N. F. Van Der Vegt, *Phys. Chem. Chem. Phys.* **2011**, *13*, 22 10468.
- <sup>11</sup> D. Reith, M. Pütz, F. Müller-Plathe, *J. Comput. Chem.* **2003**, *24*, 13 1624.
- <sup>12</sup> A. P. Lyubartsev, A. Laaksonen, *Phys. Rev. E* **1995**, *52*, 4 3730.
- <sup>13</sup> S. Izvekov, G. A. Voth, *J. Phys. Chem. B* **2005**, *109*, 7 2469.
- <sup>14</sup> J. W. Mullinax, W. G. Noid, *J. Phys. Chem. C* **2010**, *114*, 12 5661.

- <sup>15</sup> M. Karplus, J. Kuriyan, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 19 6679.
- <sup>16</sup> M. Bendahmane, K. P. Bohannon, M. M. Bradberry, T. C. Rao, M. W. Schmidtke, P. S. Abbineni, N. L. Chon, S. Tran, H. Lin, E. R. Chapman, J. D. Knight, A. Anantharam, *Mol. Biol. Cell* **2018**, *29*, 7 834.
- <sup>17</sup> S. Sharma, M. Lindau, *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 50 12751.
- <sup>18</sup> R. M. Henry, C. H. Yu, T. Rodinger, R. Pomès, *J. Mol. Biol.* **2009**, *387*, 5 1165.
- <sup>19</sup> L. K. Scarbath-Evers, S. Jähnigen, H. Elgabarty, C. Song, R. Narikawa, J. Matysik, D. Sebastiani, *Phys. Chem. Chem. Phys.* **2017**, *19*, 21 13882.
- <sup>20</sup> F. Hoffmann, J. Adler, B. Chandra, K. R. Mote, G. Bekçioğlu-Neff, D. Sebastiani, D. Huster, *J. Phys. Chem. Lett.* **2017**, *8*, 19 4740.
- <sup>21</sup> I. Kurisaki, S. Tanaka, *Phys. Chem. Chem. Phys.* **2022**, *24*, 17 10575.
- <sup>22</sup> M. S. Barhaghi, B. Crawford, G. Schwing, D. J. Hardy, J. E. Stone, L. Schwiebert, J. Potoff, E. Tajkhorshid, *J. Chem. Theory Comput.* **2022**, *18*, 8 4983.
- <sup>23</sup> H. J. Woo, A. R. Dinner, B. Roux, *J. Chem. Phys.* **2004**, *121*, 13 6392.
- <sup>24</sup> I. Y. Ben-Shalom, C. Lin, T. Kurtzman, R. C. Walker, M. K. Gilson, *J. Chem. Theory Comput.* **2019**, *15*, 4 2684.
- <sup>25</sup> M. S. Bodnarchuk, M. J. Packer, A. Haywood, *ACS Med. Chem. Lett.* **2020**, *11*, 1 77.
- <sup>26</sup> G. A. Ross, E. Russell, Y. Deng, C. Lu, E. D. Harder, R. Abel, L. Wang, *J. Chem. Theory Comput.* **2020**, *16*, 10 6061.
- <sup>27</sup> S. Pylaeva, A. Böker, H. Elgabarty, W. Paul, D. Sebastiani, *ChemPhysChem* **2018**, *19*, 21 2931.
- <sup>28</sup> F. Liang, *J. Stat. Phys.* **2006**, *122*, 3 511.
- <sup>29</sup> F. Liang, C. L. Liu, R. J. Carroll, *J. Am. Stat. Assoc.* **2007**, *102*, 477 305.
- <sup>30</sup> F. Wang, D. P. Landau, *Phys. Rev. Lett.* **2001**, *86*, 10 2050.
- <sup>31</sup> T. Shakirov, S. Zablotskiy, A. Böker, V. Ivanov, W. Paul, *Eur. Phys. J. Spec. Top.* **2017**, *226*, 4 705.
- <sup>32</sup> C. Lauer, W. Paul, *Macromol. Theory Simulations* **2023**, *2200075* 1.
- <sup>33</sup> H. J. C. Berendsen, D. van der Spoel, R. van Drunen, *Comput. Phys. Commun.* **1995**, *91*, 1-3 43.
- <sup>34</sup> D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. Berendsen, *J. Comput. Chem.* **2005**, *26*, 16 1701.
- <sup>35</sup> B. Hess, *J. Chem. Theory Comput.* **2008**, *4*, 1 116.
- <sup>36</sup> Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, P. Kollman, *J. Comput. Chem.* **2003**, *24*, 16 1999.
- <sup>37</sup> W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 2 926.
- <sup>38</sup> S. Pylaeva, M. Brehm, D. Sebastiani, *Sci. Rep.* **2018**, *8*, 1.

## Table of Contents



A physical energy scale is presented for the dimensionless interaction energies within the PRIME20 coarse-grained peptide model. Specifically, the conversion of backbone hydrogen bond energies and sidechain interaction energies into all-atom structures is investigated.



## Further Research

Marius-Andrei Codesc, Thomas Kunze, Moritz Weiß, Martin Brehm, Oleg Kornilov, Daniel Sebastiani\*, and Erik T. J. Nibbering\*

### **Ultrafast Proton Transfer Pathways Mediated by Amphoteric Imidazole**

*J. Phys. Chem. Lett.* 2023, Vol. 13, No. 20, pp. 4775-4785.

I was also involved in a collaborative project investigating the proton transfer of a molecule that exhibits amphoteric properties upon electronic excitation. My contribution showed different proton transfer energies of the excited molecules for multiple possible pathways as well as from the solvent.

I hereby confirm that the use of this article is compliant with all publishing agreements, as this article is under an open access licence: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

# Ultrafast Proton Transfer Pathways Mediated by Amphoteric Imidazole

Marius-Andrei Codescu, Thomas Kunze, Moritz Weiß, Martin Brehm, Oleg Kornilov, Daniel Sebastiani,\* and Erik T. J. Nibbering\*



Cite This: *J. Phys. Chem. Lett.* 2023, 14, 4775–4785



Read Online

ACCESS |



Metrics & More

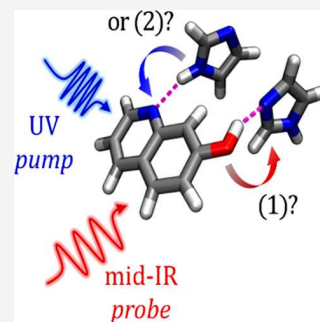


Article Recommendations



Supporting Information

**ABSTRACT:** Imidazole, being an amphoteric molecule, can act both as an acid and as a base. This property enables imidazole, as an essential building block, to effectively facilitate proton transport in high-temperature proton exchange membrane fuel cells and in proton channel transmembrane proteins, enabling those systems to exhibit high energy conversion yields and optimal biological function. We explore the amphoteric properties of imidazole by following the proton transfer exchange reaction dynamics with the bifunctional photoacid 7-hydroxyquinoline (7HQ). We show with ultrafast ultraviolet-mid-infrared pump–probe spectroscopy how for imidazole, in contrast to expectations based on textbook knowledge of acid–base reactivity, the preferential reaction pathway is that of an initial proton transfer from 7HQ to imidazole, and only at a later stage a transfer from imidazole to 7HQ, completing the 7HQ tautomerization reaction. An assessment of the molecular distribution functions and first-principles calculations of proton transfer reaction barriers reveal the underlying reasons for our observations.



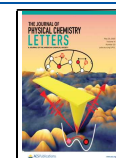
Amphoterism is the ability of molecules to act both as an acid and as a base. An example is water ( $\text{H}_2\text{O}$ ) that can both donate a proton, becoming a hydroxide anion ( $\text{OH}^-$ ), and accept a proton, forming the hydronium ion ( $\text{H}_3\text{O}^+$ ). In a similar way, amphoterism governs the acid and base properties of other protic solvents such as alcohols ( $\text{ROH}$ ) and amines ( $\text{RNH}_2$ ) and heterocyclic aromatic molecular systems like imidazole ( $\text{C}_3\text{N}_2\text{H}_4$ ). Another example is offered by ampholytes, molecules that have both acidic and basic groups, such as the amino acid  $\text{H}_2\text{N-RCH-CO}_2\text{H}$ , where tautomerism between a neutral form and a zwitterionic form is augmented with additional anionic and cationic forms. Amphoterism efficiently facilitates proton transport pathways by consecutive proton exchange steps, making it a key factor in the underlying microscopic mechanism of the von Grotthuss mechanism in water<sup>1–9</sup> and in other protic solvents.<sup>10,11</sup> In the von Grotthuss picture, the excess proton “jumps” sequentially along the solvent molecules at a pace much faster than what the Stokes–Einstein hydrodynamic diffusion model predicts for an individual protonated solvent molecule. Proton exchange is also understood to occur in a sequential von Grotthuss-like fashion in acid dissociation<sup>12–16</sup> and in acid–base neutralization reactions in protic solvents,<sup>17–22</sup> as well as for proton transport in phosphoric acid,<sup>23</sup> imidazole,<sup>24</sup> and imidazole derivatives.<sup>25,26</sup> Imidazole derivatives can be major constituents in hydrogen fuel cells,<sup>27,28</sup> for instance, as excellent proton carriers in high-temperature proton exchange membrane fuel cells (HT-PEMFC).<sup>29</sup> Finally, imidazole, being the functional group of the amino acid histidine, is a crucial building block in numerous biological systems, efficiently enabling energy transport,<sup>30</sup> signal transduction,<sup>31</sup> or pH regulation.<sup>32–34</sup>

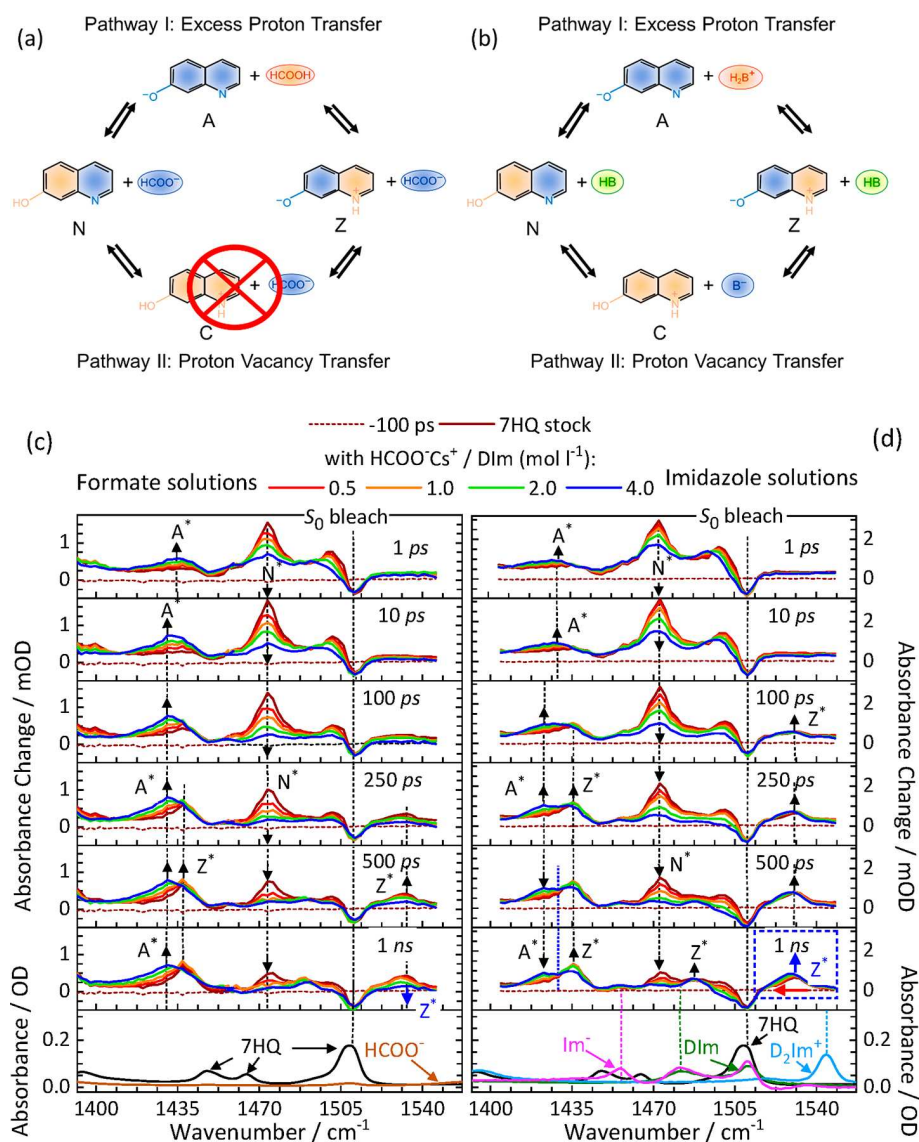
In this Letter, we report our findings for imidazole as a means for ultrafast proton transport in a methanol solution. The reasons behind the particular choice of the two molecular compounds (methanol and imidazole) for the proton-conducting material are possibly not immediately obvious. In the context of (industrial) proton exchange membrane fuel cells, liquid water is a very common choice as the proton conductor, and in practice, it is realized in the form of water channels in an otherwise hydrophobic polymer matrix with sulfonic acid end groups (the famous NAFION material concept). These materials reach their functional limits at the boiling point of liquid water, which is why considerable effort is being dedicated to finding water-free proton-conducting materials. In contrast to water as such, mid-sized organic molecules (such as imidazole and hydroxylated alkanes) could be attached directly as side chains to the polymer backbone, which would reduce or eliminate the functional degradation due to evaporation at increased temperatures. In our experimental and computational setup, a direct consideration of polymers is not possible, which is why we resort to the corresponding molecular systems. The vision is that learning the local mechanisms of protonation dynamics in these

**Received:** March 3, 2023

**Accepted:** April 26, 2023

**Published:** May 15, 2023





**Figure 1.** Comparison of the acid–base reaction pathways between the N and Z tautomers and the ionic A and C species of 7HQ, reacting with (a) formate ion/formic acid or (b) the amphoteric  $H_2B^+/HB/B^-$ , where HB can be imidazole, or the solvent  $H_2O$  or  $CH_3OH$ . Excess proton transfer pathway I can occur with the bases formate anion and imidazole but also with the solvent reacting with 7HQ. Note that for formate solutions proton vacancy pathway II can occur only with 7HQ exclusively reacting with the solvent, not with formate as the active reaction partner, whereas for imidazole solutions, both pathways are possible for imidazole and the solvent reacting with 7HQ. The transient UV-pump-IR-probe spectra are shown as a function of the base added to the deuterated methanol solution at particular pulse delay times for 7HQ reacting with (c) the formate anion or (d) imidazole. The dashed lines in the plots indicate the transient response recorded at  $-100$  ps, showing the baseline in these measurements.

materials will enable or improve the rational design of real (polymeric) water-free proton-conducting materials.

For our purposes, we use 7-hydroxyquinoline (7HQ) to initiate proton transfer reactions and follow the dynamics using femtosecond UV-pump-IR-probe spectroscopy. The objective is to measure the reaction dynamics of “tight” contact and “loose” solvent-separated reaction pairs that can be prepared under well-defined conditions. The underlying mechanisms of proton transfer dynamics of such “tight” and “loose” photoacid–carboxylate complexes have been found to occur on time scales of hundreds of femtoseconds and several picoseconds, respectively. This difference in time scales is understood to be due to a single-step event with possible solvent shell rearrangements in the case of “tight” reaction pairs,<sup>35</sup> whereas a sequential mechanism from a proton-

donating photoacid via the water solvent bridge to the accepting base necessitates hydrogen bond rearrangements facilitating the proton hops along the water bridge.<sup>17</sup> As the reaction dynamics time scales of these “tight” and “loose” complexes are clearly distinct from those of the photoacid molecules reacting with base molecules after more extensive configurational and diffusional motions that will involve a large number of hydrogen bond and solvent shell rearrangements taking place on time scales of hundreds of picoseconds (or longer), the latter fraction has time-dependent characteristics that rather can be regarded as being due to reaction kinetics without providing detailed insight into the elementary steps that underlie the proton exchange.

Amphoterism is also at play in the acid–base equilibrium of 7-hydroxyquinoline (7HQ) in the electronic  $S_0$  ground state

for both aqueous and methanol solutions, where for an aqueous solution under neutral pH  $\sim 7$  conditions both the neutral (N) and zwitterionic (Z) tautomer occur quantitatively, at a low pH of  $< 2$  the cationic form C dominates, and at a high pH of  $> 10$  the anionic form A is formed (see Figure 1).<sup>36,37</sup> 7HQ is a so-called bifunctional photoacid, for which electronic excitation of the  $S_0 \rightarrow S_1$  transition of the neutral 7HQ tautomer N makes the molecule both a photoacid and a photobase. The  $pK_a$  value of the OH group decreases by 8 units, making 7HQ a strong acid, whereas the  $pK_b$  value of the quinoline nitrogen site changes by 5 units, making 7HQ also a stronger base. These properties strongly dictate the acid–base behavior in the first electronic excited state and photoinduced proton transfer dynamics of 7HQ in protic solvents.<sup>11,36–42</sup>

Recent results obtained in a combined ultrafast infrared spectroscopic and ab initio quantum molecular dynamics study of 7HQ in water/methanol mixtures have shown that on a microscopic level proton transport takes place from the proton-accepting quinoline group to the proton-donating OH group via a methoxide/hydroxide transport mechanism on a time scale of tens to hundreds of picoseconds.<sup>11</sup> Via the addition of another acid or base, it is possible to change the preference of the proton transfer pathways from a hydrolysis/methanolysis (solvolysis) “proton vacancy” mechanism (with 7HQ following the  $N^* \rightarrow C^* \rightarrow Z^*$  route) to a protolysis “excess proton” mechanism (with 7HQ transforming from  $N^* \rightarrow A^* \rightarrow Z^*$ ). Our first results were obtained with the formate anion, which promptly accepts the proton from the OH group of 7HQ upon electronic excitation of the bifunctional photoacid, when in the proximity of 7HQ under either “tight” contact reaction pair or “loose” solvent-separated reaction pair conditions.<sup>43</sup> Our results obtained with the formate anion as an additive did not support a full quantitative transformation of 7HQ following the  $N^* \rightarrow A^* \rightarrow Z^*$  pathway. In this work, we report on the possible role of imidazole as a mediator in the different proton transport pathways that 7HQ can follow, as always with a close interplay with and or even direct involvement of the nearest solvent molecules. Here we will show that amphoterism is at play with imidazole, acting both as a proton acceptor and as a proton donor, as opposed to *N*-methylimidazole that has been used as base in proton transfer studies with 7-hydroxy-4-(trifluoromethyl)-1-coumarin.<sup>44</sup> In the case of amphoteric imidazole, one cannot a priori assume that the dominant reaction pathway of 7HQ changes from the solvolysis (methanolysis) to the protolysis pathway when imidazole is added to a solution of 7HQ in a methanol solution, as imidazole can be the active reaction partner of 7HQ in both possible acid–base reaction routes. Whereas empirical free energy–reactivity relationships will provide clear hints about this matter for “loose” 7HQ–imidazole reaction pairs, there is no straightforward assessment available for the “tight” 7HQ–imidazole reaction pairs that may be envisaged to be present in solutions with high imidazole concentrations, as proton transfer reactions for “tight” acid–base reaction pairs may well have a low reaction barrier or may even be barrierless.

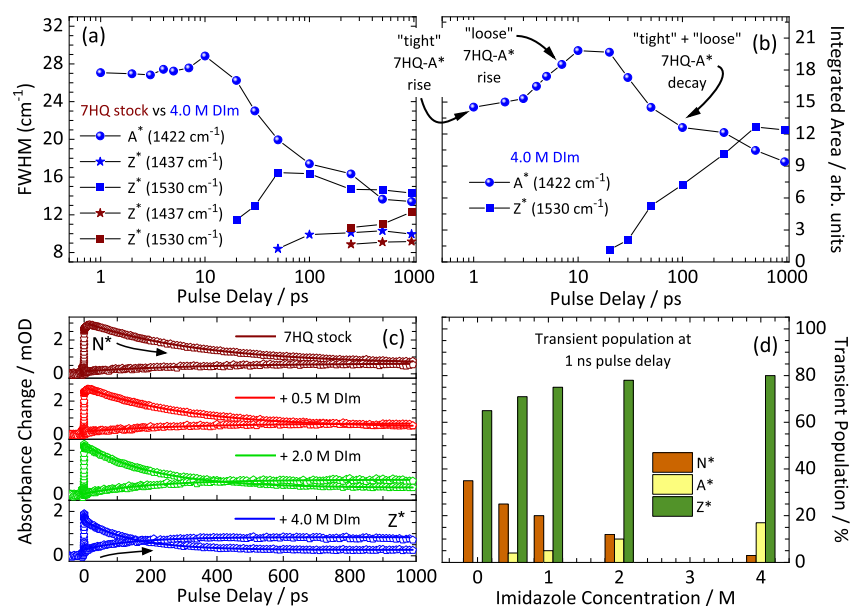
We follow the ultrafast proton transfer dynamics between 7HQ and imidazole as a function of time upon electronic excitation of the neutral 7HQ tautomer N at 330 nm. By using ultrafast infrared spectroscopy, we can follow the dynamical behavior of the different forms of ampholyte 7HQ in the first electronic excited state, namely,  $N^*$  and  $Z^*$  tautomers and charged  $C^*$  and  $A^*$  species.<sup>11,42,43</sup> The IR-active normal

modes specific to each species of 7HQ in the  $S_1$  state have been identified and characterized,<sup>42</sup> allowing the acid–base reaction dynamics along the protolysis or solvolysis pathways in deuterated methanol ( $CD_3OD$ ) to be distinctly followed<sup>11</sup> (see Figure 1) and steered.<sup>43</sup> Figure 1 also provides an overview of the transient UV/IR pump–probe spectra recorded for the 7HQ–imidazole pair at specific pulse delay times when particular steps along the possible proton transfer pathways are anticipated to occur. We show here the spectral region of 1390–1560  $cm^{-1}$  where the marker bands for different charged and tautomer species of 7HQ can be most easily discerned (a broader spectral range is presented in the Supporting Information).  $N^*$  has a strong IR-active transition at 1475  $cm^{-1}$ ;  $Z^*$  displays two IR-active bands at 1440 and 1530  $cm^{-1}$ , while  $A^*$  appears with a broad band at 1430  $cm^{-1}$  but is narrower and frequency downshifted to 1422  $cm^{-1}$  at longer pulse delays. The pulse delay-dependent magnitudes of these marker bands are directly proportional to the transient population of the 7HQ species in the  $S_1$  state. For comparison, the transient UV/IR pump–probe spectra are depicted with those of a previously reported experiment on the 7HQ–formate photoacid–base system<sup>43</sup> and the stock solution of 7HQ.<sup>11,42</sup>

The following observations on proton transfer dynamics of 7HQ, as grasped from the transient response of the IR-active marker bands of 7HQ obtained with imidazole as a base, are much like those realized with formate: (1) an initial decrease in the level of  $N^*$  and the appearance of  $A^*$  within the time resolution, (2) a further decrease in the level of  $N^*$  and the appearance of more  $A^*$  on the time scale of a few picoseconds, (3) the magnitudes of the early time components of the decrease in the level of  $N^*$  and the increase in the level of  $A^*$  increase with imidazole concentration, and (4) the magnitude of the  $N^*$  marker band decreases further at longer time scales of hundreds of picoseconds, scaling with the base concentration and approaching zero for the highest base concentration used in these experiments (4.0 M).

Distinctly different behavior can also be deduced from the comparison between the 7HQ–imidazole and 7HQ–formate results at the high base concentration depicted in Figure 1. Whereas in the 7HQ–formate case the transient  $A^*$  marker band indicates that the proton transfer kinetics predominantly halts at the  $A^*$  anion, for the 7HQ–imidazole case the reaction proceeds further. The appearance of the  $Z^*$  marker bands occurs on a time scale of hundreds of picoseconds, whereas the  $A^*$  marker band has by then frequency downshifted from 1430 to 1422  $cm^{-1}$  and diminished in spectral breadth, suggesting a decrease in the transient population of  $A^*$  on this long time scale. The initial large spectral breadth and subsequent narrowing of IR-active marker bands on picosecond time scales have often been observed for ultrafast photoinduced reactions<sup>45</sup> and ascribed to initial excess vibrational excitation (i.e., the molecules with increased internal vibrational population numbers are “hot”), followed by vibrational cooling by energy dissipation to the solvent shell molecules. Such phenomena are typically observed for IR-active vibrational modes of chromophores undergoing photoinduced chemical reactions. As anharmonic coupling constants have typically a negative sign, the IR-active fingerprint of initial “hot” molecules appears to be frequency downshifted, and with the vibrational cooling process, the spectral narrowing of the IR-active fingerprint modes is accompanied by a frequency upshift. However, in the case of 7HQ reacting with imidazole,





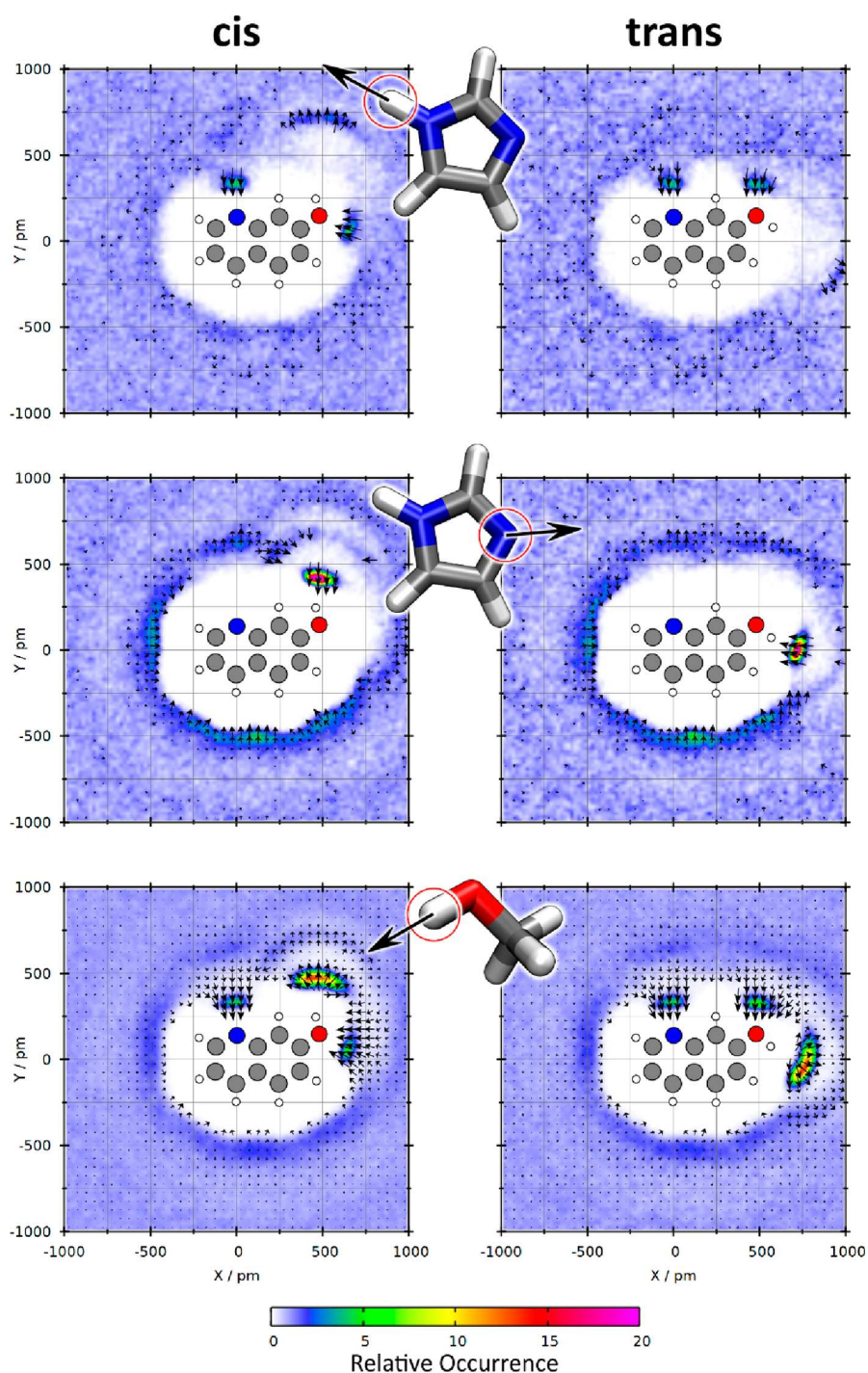
**Figure 2.** (a) Full widths at half-maximum (fwhm) of 7HQ marker bands for the  $A^*$  and  $Z^*$  species as a function of pulse delay time for the 7HQ stock and 4 M imidazole solutions. (b) Integrated areas of the  $A^*$  and  $Z^*$  marker bands (at 1422 and 1530  $\text{cm}^{-1}$ , respectively), as a function of pulse delay time. (c) Transient kinetics of the 7HQ  $N^*$  and  $Z^*$  tautomers as a function of the DIm imidazole concentration. (d) Absolute population fractions of the  $N^*$  and  $Z^*$  tautomers and the  $A^*$  anion at a 1 ns pulse delay time, derived from the transient UV/IR pump–probe spectra, as a function of imidazole concentration. Note that the curves depicted in panels a and b are shown with logarithmic scaling of the x-axis, whereas in panel c, a normal scaling has been used, to highlight the early time components of “tight” and “loose” complexes in panels a and b and the long time components in panel c.

the frequency shift occurs in the opposite direction. An alternative explanation for our observations on the  $A^*$  marker band around 1422–1430  $\text{cm}^{-1}$  on a time scale from picoseconds to several tens of picoseconds may be found in a possible distribution of hydrogen bond configurations directly upon proton/deuteron transfer from 7HQ to imidazole, converting  $N^*$  into  $A^*$ , followed by hydrogen bond and solvent shell rearrangements. Instead, the two marker bands of  $Z^*$  at 1440 and 1530  $\text{cm}^{-1}$  exhibit a small increase of spectral width with a large pulse delay for a high imidazole concentration compared to what has been observed for 7HQ without an added base. The 1470  $\text{cm}^{-1}$  marker band of the  $N^*$  species of 7HQ shows a small increase in spectral width but negligible changes in frequency position with an increase in imidazole concentration. Even though a normal mode analysis of the fingerprint modes of the four different 7HQ species,  $N^*$ ,  $C^*$ ,  $A^*$ , and  $Z^*$ , has been performed<sup>42</sup> and these normal modes have been found to be predominantly governed by C–C, C=C, C–N, and C=N stretching displacements of the C and N atoms in the aromatic quinoline ring parts and C–H bending motions of 7HQ, a proper analysis of the role of hydrogen bonding and solvent shell rearrangements can be performed only when the first solvent shell molecules are also included in these normal mode characterizations. Only further in-depth mode analysis of the IR-active marker vibrations of  $N^*$ ,  $A^*$ , and  $Z^*$  as well as an assessment of the distribution of the configurations of the 7HQ–imidazole reaction pairs and their temporal characteristics, which can be grasped by ab initio molecular dynamics simulations, may shed light on the underlying reasons for this interesting observation.

Whereas in our earlier studies<sup>11,42</sup> we have been able to successfully analyze the observed transient population kinetics by the time-dependent magnitude of the  $N^*$ ,  $A^*$ , and  $Z^*$

species, we found here that this approach has led to inconsistent results. The reason is that in these previous studies the population kinetics predominantly occurred on longer time scales of hundreds of picoseconds, whereas possible changes in the band shape and shifts in frequency position are known to typically take place on subpicosecond time scales or time scales of a few picoseconds, albeit not necessarily for all IR-active marker bands.<sup>43</sup> To discern the correct transient population dynamics of 7HQ upon electronic excitation of the  $N$  tautomer at early pulse delay times, the transient mid-infrared absorption bands have been analyzed using a Gaussian line shape fitting procedure (for details, see the Supporting Information). It turns out that this procedure is necessary to correctly determine the population dynamics of the  $A^*$  anion that is generated in the “tight” and “loose” complexes at the early pulse delay times.

Figure 2 shows the results of our analysis of the transient population of the  $N^*$  and  $Z^*$  tautomer and the  $A^*$  anion species in the 7HQ–imidazole reaction. The outcome of the Gaussian line shape fitting of the 1422–1430  $\text{cm}^{-1}$  marker band of  $A^*$  and of the two marker bands of  $Z^*$  (1437 and 1530  $\text{cm}^{-1}$ ) is shown in panels a and b of Figure 2 for the 4.0 M imidazole case, which we have plotted on a logarithmic scale for the pulse delay x-axis, to accentuate the early time dynamics. Indeed, the full width at half-maximum (fwhm) of this band decreases by a factor of 2.5 within 50–100 ps (Figure 2a), which we tentatively ascribe to hydrogen bond reorganization and solvent shell rearrangement dynamics of the imidazole/imidazolium units in the 7HQ–imidazole product pairs on this time scale. Moreover, the integrated intensity of the  $A^*$  marker band has a major component appearing within the time resolution, and an additional increase on the picosecond time scale (see Figure 2b). Similar early time components can also be observed in the decay of

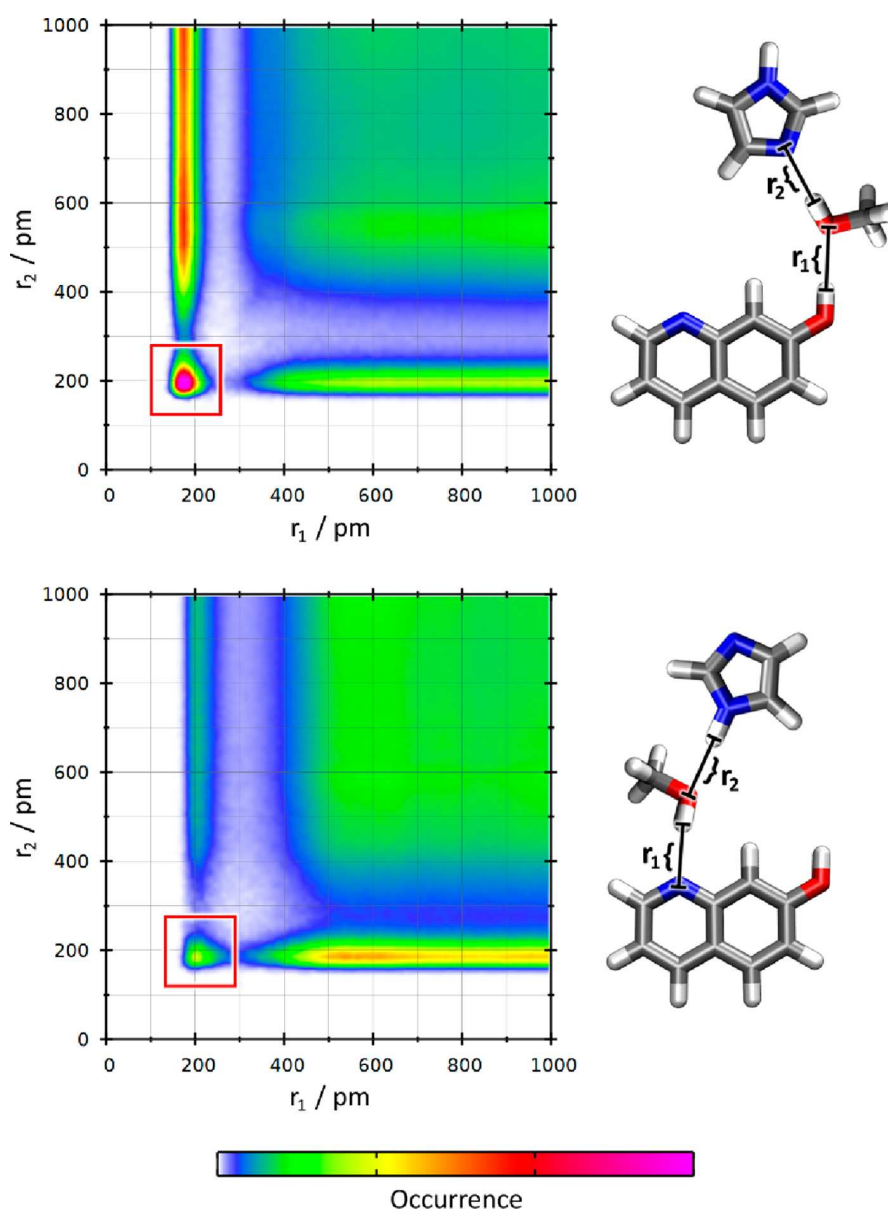


**Figure 3.** Average particle density of imidazole (HIm) and methanol (CH<sub>3</sub>OH) molecules around a *cis*- or *trans*-7HQ molecule (see the color scale) together with the average orientation of the displayed vectors of HIm and CH<sub>3</sub>OH (see the arrows).

N\*, and together with the observations of the increase in the level of A\*, a consistent picture of “tight” contact and “loose” solvent-separated 7HQ–imidazole reaction pairs emerges, the relative fraction for these increasing with imidazole concentration. Interestingly, at longer pulse delay times, the integrated intensity of the A\* marker band decays with a time constant that appears within the error margin of the experimental data to be identical to the increase in the integrated intensity of the Z\* marker band at 1530 cm<sup>−1</sup> (Figure 2b). Additional components in the decay of N\* occurring on a time scale of hundreds of picoseconds appear to be similar to components

in the rise of Z\* (Figure 2c). Results from exponential fits of these long time components are summarized in Table S2.

The experimental results clearly display a change in the reaction dynamics of 7HQ when going from 0.0 M (stock solution) to 4.0 M imidazole. The N\* to Z\* tautomerization accelerates in the presence of imidazole, and a switch from the solvolysis (methanolysis) “proton hole” pathway toward the protolysis “excess proton” pathway is apparent upon addition of more imidazole to the solution. To determine the relative reaction yields along these two different routes, we have used the results from the marker band fitting to Gaussian profiles



**Figure 4.** Combined distribution function depicting the probability of finding a certain distance (see the sketch) between 7HQ and CH<sub>3</sub>OH and between CH<sub>3</sub>OH and HIm on the horizontal and vertical axes, respectively.

(see also Figure S3). For the stock solution, we observe an equilibration of the fraction ( $f_i$ ) of photoexcited 7HQ–N\* molecules from  $f_{N^*}(\tau = 0 \text{ ps}) = 1$  to  $f_{N^*}(\tau = 1 \text{ ns}) = 0.35$ . Realizing that for the “proton hole” pathway, with 7HQ following the  $N^* \rightarrow C^* \rightarrow Z^*$  tautomerization route, only a significant transient population of the N\* and Z\* tautomers has been observed at any pulse delay for 7HQ in deuterated methanol (CD<sub>3</sub>OD),<sup>11,42,43</sup> we can then correlate the value of the observed transient absorbance of the Z\* 1540 cm<sup>−1</sup> band with a transient population value:  $f_{Z^*}(\tau = 1 \text{ ns}) = 0.65$  [with  $f_{Z^*}(\tau = 0 \text{ ps}) = 0$ ]. Assuming that the IR cross sections of the integrated 7HQ marker bands are not affected when going from the (0.0 M imidazole) stock solution to 4.0 M imidazole in CD<sub>3</sub>OD, we learn then that with a 1 ns pulse delay the fraction of Z\* increases to  $f_{Z^*}(\tau = 1 \text{ ns}) = 0.80 \pm 0.03$  while that of N\* is much smaller [ $f_{N^*}(\tau = 1 \text{ ns}) = 0.03 \pm 0.03$ ]. This means that the value observed for the integrated band intensity of the A\* anion with a 1 ns pulse delay is significant:  $f_{A^*}(\tau = 1$

ns) =  $0.17 \pm 0.03$ . Now comparing the long delay value of the integrated band value of A\* ( $9.4 \pm 0.2$ ) with that of its maximum value ( $19.8 \pm 0.2$ ) reached at 10 ps, we learn that under 4.0 M imidazole conditions a transient population build-up fraction of A\* reaches a value of  $f_{A^*}(\tau = 10 \text{ ps}) = 0.36 \pm 0.07$ . This value is consistent with the decrease in the transient population of N\* from  $f_{N^*}(\tau = 0 \text{ ps}) = 1$  to  $f_{N^*}(\tau = 10 \text{ ps}) = 0.45 \pm 0.05$ , due to the fast deuteron transfer reaction for “tight” and “loose” complexes, as derived from the early time decay components of the N\* marker band at 1470 cm<sup>−1</sup>.

The decay of the integrated area of the A\* marker band at time scales of hundreds of picoseconds correlates well with the rise of that of the Z\* marker band points strongly to the dominant occurrence of the  $N^* \rightarrow A^* \rightarrow Z^*$  pathway at 4.0 M imidazole, yet we cannot unequivocally exclude the possibility that a significant fraction still follows the  $N^* \rightarrow C^* \rightarrow Z^*$  pathway. One could interpret the long time dynamical behavior of the N\*, A\*, and Z\* marker bands in experimental



observations as being indicative of further fractions of the 7HQ  $N^*$  tautomer following the  $N^* \rightarrow A^* \rightarrow Z^*$  pathway, for which the first proton/deuteron transfer step takes more time due to rearrangements of the imidazole molecules toward the proton/deuteron-donating 7HQ–OH site with possible interstitial solvent molecules, including partial rotational and diffusional motions. After these rearrangements, a proton/deuteron transfer rate may be inferred from free energy–reactivity relationships (see Table S3), albeit this under the assumption of spherical shapes for acids and bases without explicit molecular structure, to model the long time components in diffusion-assisted photoacid–base reactions in protic solvents.<sup>11,42,43,46–52</sup> On the contrary, the alternative option of imidazole acting as a proton/deuteron donor toward the 7HQ–quinoline nitrogen site also must be considered. As the acidity of imidazole as a proton donor is quite similar to that of the solvent methanol, one can also argue that a major fraction of 7HQ still follows the  $N^* \rightarrow C^* \rightarrow Z^*$  pathway, but now with the first step initiated by imidazole acting as proton donor forming the imidazolate anion. On the basis of our experimental results, we can exclude an ultrafast proton/deuteron transfer on subpicosecond time scales or time scales of a few tens of picoseconds for the  $N^* \rightarrow C^* \rightarrow Z^*$  pathway, because  $C^*$  is not observed at these short pulse delay times and  $Z^*$  is only formed with a time constant of 150 ps at 4.0 M imidazole. This implies, even for 4.0 M imidazole conditions, an absence of “tight” complexes when the preexisting hydrogen bond between the 7HQ quinoline nitrogen site and N–H group of imidazole comprises a barrierless proton transfer reaction coordinate or when these “tight” complexes have a major abundance it has a major reaction barrier with a 150 ps reaction time constant as a result. To discern which of these two options prevails, we now present our results on classical molecular dynamics simulations and electronic excited state quantum chemical calculations.

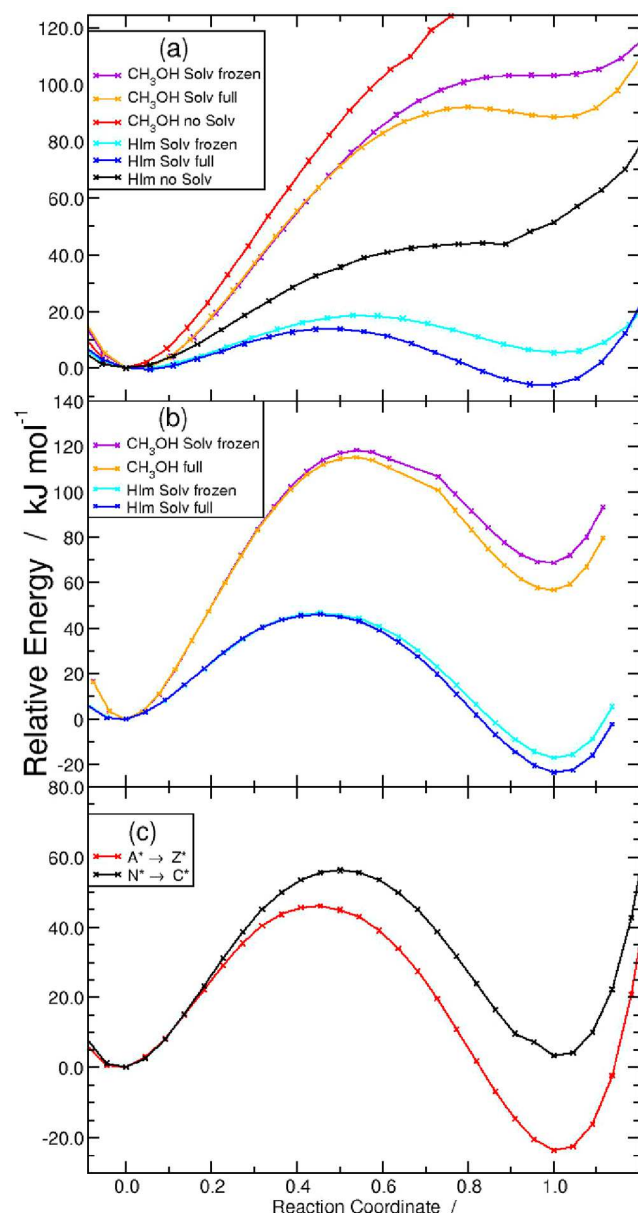
More insight into the geometric aspects of specific 7HQ–imidazole reaction pair configurations can be obtained from the atomic and molecular spatial distribution functions of solvent molecules around 7HQ, which can be derived from molecular dynamics simulations. Figure 3 shows the average particle densities of methanol and imidazole around the hydrogen bond-donating and -accepting sites of 7HQ. These results show a pronounced occurrence of having the 7HQ–OH group acting as a hydrogen bond donor, for the *cis* and *trans* configurations of the hydroxyl group, to a specific methanol or imidazole molecule acting with their lone pairs as a hydrogen bond acceptor. Here it is good to note that the distribution of dihedral angle functions of 7HQ is ~60% *cis*-7HQ and ~40% *trans*-7HQ, irrespective of the concentration of imidazole used in our experiments. The dihedral angles have rather narrow distributions of the *cis*-7HQ and *trans*-7HQ configurations with their maxima in the plane of the 7HQ–aromatic backbone (see the Supporting Information). As a result, the spatial distribution functions of methanol and imidazole are confined to a limited range around the hydrogen bond-donating and -accepting sites of 7HQ. This fact also validates the assumption that a planar projection of the solvent distribution functions, as shown in Figure 3, provides all essential information in a transparent fashion.

The fact that the magnitudes of the calculated spatial distribution functions for methanol or imidazole being donors of hydrogen bonds to the 7HQ hydroxyl and quinoline nitrogen lone pairs are approximately 3–4 times lower than

those where methanol or imidazole acts as a hydrogen bond acceptor strongly suggests an accordingly smaller hydrogen bond interaction strength for these two different types of hydrogen bonds that 7HQ can have. This also means that in the context of preformed “tight” contact reaction pairs imidazole is more likely to act with its nitrogen lone pair as a base to the 7HQ hydroxyl acidic group than imidazole is to act with its N–H group as an acid to donate the proton to the 7HQ quinoline nitrogen site. To determine whether a similar situation occurs for the “loose” solvent-separated reaction pairs, we computed the two-dimensional distribution functions from our molecular dynamics simulations (Figure 4). Here, again, a larger relative occurrence is apparent for those “loose” complexes where the solvent and imidazole are interacting as hydrogen bond acceptors to the 7HQ–OH group. Instead, in the other “loose” complexes, the solvent and imidazole act as hydrogen bond donors to the 7HQ–quinoline nitrogen. The combined distribution functions also show that the hydrogen bond distances are indicative of weak hydrogen bond strengths,<sup>53–56</sup> as expected for 7HQ in the electronic ground state.

The observed population kinetics obtained from the experimental ultrafast UV/IR pump–probe results shows that for high imidazole concentrations the  $N^* \rightarrow A^* \rightarrow Z^*$  pathway of 7HQ is the dominant pathway for the “tight” and “loose” reaction pair fractions. The particle density and distribution functions derived strongly point to a significant fraction of imidazole being hydrogen bonded at both the 7HQ–OH and 7HQ–quinoline nitrogen sites for these “tight” and “loose” reaction pairs. These findings yield a series of new questions about the relative reaction rates for individual proton transfer steps. (1) Why is the deprotonation of  $N^*$  into the solvent faster toward imidazole than toward methanol? (2) Why is the protonation of  $A^*$  at the nitrogen site faster when taken from imidazole than from methanol as the proton donor? (3) Why is the protonation at the nitrogen site from imidazole faster for  $A^*$  than for  $N^*$ , whereas the opposite is the case for methanol?<sup>42</sup> To provide insight into these factors that control the reaction rates of individual steps and the overall relative importance of the  $N^* \rightarrow A^* \rightarrow Z^*$  and  $N^* \rightarrow C^* \rightarrow Z^*$  pathways, we have determined the proton transfer energy profiles for the distinct reaction steps for the “tight” hydrogen-bonded 7HQ–HIm complexes (see Figure 5). For the reaction path, we used a linear interpolation between the optimized reactant and product geometries. This path was sampled with vertical excitation energy calculations at the time-dependent density functional theory (TD-DFT) level. The solvent screening effect was incorporated via effective polarizable continuum methods using both the dielectric constants corresponding to electronic polarization only ( $\epsilon_\infty = 4.5$ ,<sup>57</sup> inspired from the observation that the protonation reaction is considerably faster than the geometric solvent reorientation dynamics) and, for comparison, the dielectric constant corresponding to full solvent relaxation ( $\epsilon_0 = 32.63$ <sup>58</sup>). Special care was taken to correctly follow the proper electronic excited state energy profile (i.e., the one that corresponds to the  $S_1$  state of 7HQ) that does not always represent the lowest vertical excitation along the reaction path. Our calculations show that from an enthalpy perspective, the imidazole molecules are considerably more likely to accept or donate a proton than are methanol molecules. The reaction barriers with methanol as the proton acceptor or donor are so much higher (on the order of 80 kJ/mol) that whenever imidazole is





**Figure 5.** Energy paths for (a) oxygen site deprotonation onto HB =  $\text{CH}_3\text{OH}$ , HIm ( $\text{N}^* + \text{HB} \rightarrow \text{A}^* + \text{H}_2\text{B}^+$ ) and (b) proton abstraction from HB onto the nitrogen site ( $\text{A}^* + \text{HB} \rightarrow \text{Z}^* + \text{B}^-$ ) in both cases with different solvation influences. (c) Energy path for proton abstraction of HIm onto  $\text{A}^*$  and  $\text{N}^*$ .

present in the solvent, the reaction will preferentially follow those pathways where imidazole is actively involved as the proton acceptor [ $\text{N}^* + \text{HB} \rightarrow \text{A}^* + \text{H}_2\text{B}^+$  (Figure 5a)] or the proton donor [ $\text{A}^* + \text{HB} \rightarrow \text{Z}^* + \text{B}^-$  (Figure 5b)]. In terms of relative reaction rates, we estimate that the reaction via imidazole will occur  $10^8$  times faster than via methanol, assuming that the vibrational relaxation after vertical excitation results in a locally increased kinetic energy of the proton corresponding to a proton temperature of 600 K. Finally, in the comparison of proton abstraction by either  $\text{N}^*$  or  $\text{A}^*$  from imidazole, it follows that the barrier of the  $\text{A}^* + \text{HB} \rightarrow \text{Z}^* + \text{B}^-$  pathway is lower than that of the  $\text{N}^* + \text{HB} \rightarrow \text{C}^* + \text{B}^-$  pathway (with HB being imidazole). In this case, the relative reaction rates can be estimated to differ by a factor of 50 (considering a relative barrier difference of 10  $\text{kJ/mol}$  at a

proton temperature of 300 K, as the reacting proton does not originate from 7HQ and therefore does not have an increased temperature).

A close connection between acid–base reaction rates has been defined by Marcus using the BEBO free energy–reactivity relation,<sup>46</sup> and ultrafast proton transfer studies have shown that this relationship appears to confirm predicted proton transfer rates with those derived from experiments.<sup>11,49–51</sup> The current understanding is that conclusions can be drawn about only solvent-mediated proton transfer, i.e., for “loose” solvent-separated acid–base reaction pairs. Interestingly, despite this caveat, an assessment of the acid–base reactivity of 7HQ with imidazole based on a comparison of the differences in  $\text{pK}_a$  values,  $\Delta\text{pK}_a$  (see Table S3), also provides for the case of “tight” acid–base reaction pairs a clear indication of why  $\text{N}^*$  reacts faster with imidazole than with methanol as the proton acceptor or why  $\text{A}^*$  reacts faster with imidazole than with methanol as the proton donor and why abstraction of a proton from imidazole by  $\text{A}^*$  is faster than that by  $\text{N}^*$ . We argue that with the current energy barrier calculations we can provide a proper estimate in quantitative terms of the extent to which a close similarity exists between the quantum chemical calculations of acid–base reaction pairs and the semiempirical BEBO free energy relationship.

We note that the energy barriers, the  $\Delta\text{pK}_a$  values, and the observed reaction rates all point to imidazole being a faster proton transporter than methanol, even though the underlying mechanisms for proton transport are similar.<sup>10,24</sup> We suggest that the individual steps in charge separation are more efficiently mediated by the larger aromatic imidazole both as a proton acceptor and as a proton donor (due to a greater delocalization of the charge on the imidazole cation and anion, respectively; this feature also applies for the aromatic 7HQ molecule) than by methanol where the charge is much more localized. Ultimately, our findings for the energy barriers of the individual proton transfer steps with imidazole as the proton donor and proton acceptor strongly suggest that the  $\text{N}^* \rightarrow \text{A}^* \rightarrow \text{Z}^*$  pathway is followed by those 7HQ molecules that are already in the “tight” and “loose” reaction pair configurations and is the more likely route followed by those 7HQ molecules where larger orientational rearrangements are necessary before a reaction with imidazole can proceed.

We have attempted to identify the underlying reasons for the shape of the energy profiles in terms of the degree of charge delocalization in the solvent molecules (methanol and imidazole), in view of providing an intuitive qualitative answer to the three questions about the proton transfer reaction rates. For this purpose, we have computed the changes in the partial charges for the  $\text{A}^* + \text{HB} \rightarrow \text{Z}^* + \text{B}^-$  reaction for the oxygen/nitrogen atoms of B. While for methanol the oxygen partial charge changes from  $-0.54$  to  $-0.98$ , the imidazole nitrogen charge changes from  $-0.13$  to  $-0.39$ . Interestingly, the other nitrogen atom in imidazole remains practically unchanged, and its partial charge changes from  $-0.35$  to  $-0.39$ . The negative charge is thus equally shared between the two nitrogen atoms, while in  $\text{CH}_3\text{O}^-$ , the entire charge is carried by the single oxygen atom. This illustrates the considerably better intramolecular delocalization of the anionic excess charge for imidazole compared to methanol, which in turn explains why the proton donation capability is so much better for imidazole. A similar argument holds for the 7HQ deprotonation reactions. Along this line of argument, we anticipate similar effects will play an important role in proton transport in mixed

water/imidazole and mixed methanol/imidazole solutions without the photoacid chromophore, but much less in neat imidazole or imidazole derivatives.<sup>24,26</sup>

In conclusion, our joint experimental–theoretical study illustrates that it is possible to manipulate the pathway of acid–base proton exchange reactions in a controlled way, exploiting the different microsolvation properties and proton affinities of suitably chosen solvent molecules. Here, we have used the bifunctional photoacid 7HQ in combination with methanol and imidazole as solvent components, as opposed to the more common aqueous solvation. In comparison to the water environment, we observe specific changes in the reaction pathways and reaction dynamics when 7HQ is transformed from the N\* state to the final Z\* state. Previous studies of proton transfer pathways of 7HQ in methanol have shown that the acquisition of a proton from the solvent methanol by the 7HQ quinoline nitrogen site is the first and rate-determining step followed by a fast proton donation by the 7HQ hydroxyl OH group to the solvent (with the reaction rate of the first N\* → C\* step being 2 orders of magnitude smaller than that of the second C\* → Z\* step). Adding imidazole as a reaction partner quantitatively changes the time ordering of the proton transfer steps, with ultrafast proton donation of the 7HQ hydroxyl OH group to imidazole occurring on a subpicosecond time scale or a time scale of a few picoseconds, and only at clearly longer time scales does the acquisition of a proton by the 7HQ quinoline nitrogen site from a nearby imidazole complete the 7HQ N\* → Z\* tautomerization reaction. The sequential N\* → A\* → Z\* pathway has, irrespective of the 7HQ–imidazole reaction pair configurations being “tight” contact or “loose” solvent-separated, a 2–3 order of magnitude difference in reaction rates between the first N\* → A\* step and the second A\* → Z\* step. In light of our findings of this large difference in reaction rates for imidazole as the proton acceptor compared to imidazole as the proton donor, it would be interesting to further explore this aspect in those cases in which imidazole derivatives mediate proton transfer, such as in HT-PEMFC or transmembrane proton channel proteins.

## EXPERIMENTAL AND COMPUTATIONAL METHODS

The Supporting Information includes details about the experimental setup used;<sup>11,42,43</sup> free energy–reactivity assessments (using reported<sup>42,59,60</sup> or derived<sup>61–63</sup> pK<sub>s</sub> values); molecular dynamics simulations with LAMMPS<sup>64</sup> and PACKMOL<sup>65</sup> packages, and using optimized potentials for liquid simulations all-atom (OPLS-aa) force fields,<sup>66</sup> restrained electrostatic potentials (RESP),<sup>67</sup> and a Nosé–Hoover chain thermostat;<sup>68–70</sup> trajectory analysis using the TRAVIS,<sup>71,72</sup> VMD,<sup>73</sup> and Tachyon<sup>74</sup> program packages; and proton transfer energy profiles calculated with ORCA<sup>58</sup> on the TD-DFT level with a solvent polarizable continuum model (CPCM).<sup>75</sup>

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpclett.3c00595>.

Experimental details and computational details, steady state and transient IR spectra, examples of Gaussian line shape fits of A\* and Z\* marker bands, exponential fits of long time behavior kinetics, free energy–reactivity relationships, acidities and reactivities as estimated for

7HQ and imidazole in H<sub>2</sub>O, CH<sub>3</sub>OH, and CD<sub>3</sub>OD solutions, and dihedral distribution functions (PDF)

## AUTHOR INFORMATION

### Corresponding Authors

Erik T. J. Nibbering – Max Born Institut für Nichtlineare Optik und Kurzzeitspektroskopie, 12489 Berlin, Germany; [orcid.org/0000-0001-5874-8052](https://orcid.org/0000-0001-5874-8052); Email: [nibberin@mbi-berlin.de](mailto:nibberin@mbi-berlin.de)

Daniel Sebastiani – Institut für Chemie, Martin-Luther-Universität Halle-Wittenberg, 06120 Halle (Saale), Germany; [orcid.org/0000-0003-2240-3938](https://orcid.org/0000-0003-2240-3938); Email: [daniel.sebastiani@chemie.uni-halle.de](mailto:daniel.sebastiani@chemie.uni-halle.de)

### Authors

Marius-Andrei Codescu – Max Born Institut für Nichtlineare Optik und Kurzzeitspektroskopie, 12489 Berlin, Germany

Thomas Kunze – Institut für Chemie, Martin-Luther-Universität Halle-Wittenberg, 06120 Halle (Saale), Germany

Moritz Weiß – Institut für Chemie, Martin-Luther-Universität Halle-Wittenberg, 06120 Halle (Saale), Germany

Martin Brehm – Institut für Chemie, Martin-Luther-Universität Halle-Wittenberg, 06120 Halle (Saale), Germany; [orcid.org/0000-0002-6861-459X](https://orcid.org/0000-0002-6861-459X)

Oleg Kornilov – Max Born Institut für Nichtlineare Optik und Kurzzeitspektroskopie, 12489 Berlin, Germany; [orcid.org/0000-0002-3343-2614](https://orcid.org/0000-0002-3343-2614)

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jpclett.3c00595>

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

The authors cordially acknowledge the financial support from the Deutsche Forschungsgemeinschaft (DFG) under Project 263266015 (Project codes SE 1008/11-2 and NI 492/13-2) and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (ERC Grant Agreement 788704). M.B. acknowledges financial support by the Deutsche Forschungsgemeinschaft (DFG) through Project Br 5494/1-3.

## REFERENCES

- (1) de Grotthuss, C. J. T. Sur la Décomposition de l'Eau et des Corps Qu'elle Tient en Dissolution à l'Aide de l'Electricité Galvanique. *Ann. Chim.* **1806**, LVIII, 54–74.
- (2) von Grotthuss, T. Über die Chemische Wirksamkeit des Lichts und der Elektrizität; Besonders über einen Merkwürdigen Neuen Gegensatz dieser Wirksamkeit, den das Licht auf Gewisse Substanzen Äußert, je Nachdem es entweder aus Nicht-Oxidirenden Körpern oder aus der Atmosphärischen Luft Unmittelbar in Dieselben und aus Letzteren in Jene Eindringt. *Jahresverhandlungen der Kurländischen Gesellschaft für Literatur und Kunst* **1819**, 1, 119–194.
- (3) Agmon, N. The Grotthuss Mechanism. *Chem. Phys. Lett.* **1995**, 244, 456–462.
- (4) Tuckerman, M.; Laasonen, K.; Sprik, M.; Parrinello, M. Ab-Initio Molecular-Dynamics Simulation of the Solvation and Transport of H<sub>3</sub>O<sup>+</sup> and OH<sup>−</sup> Ions in Water. *J. Phys. Chem.* **1995**, 99, 5749–5752.
- (5) Tuckerman, M.; Laasonen, K.; Sprik, M.; Parrinello, M. Ab-Initio Molecular-Dynamics Simulation of the Solvation and Transport of Hydronium and Hydroxyl Ions in Water. *J. Chem. Phys.* **1995**, 103, 150–161.

- (6) Marx, D.; Tuckerman, M. E.; Hutter, J.; Parrinello, M. The Nature of the Hydrated Excess Proton in Water. *Nature* **1999**, *397*, 601–604.
- (7) Vuilleumier, R.; Borgis, D. Transport and Spectroscopy of the Hydrated Proton: A Molecular Dynamics Study. *J. Chem. Phys.* **1999**, *111*, 4251–4266.
- (8) Schmitt, U. W.; Voth, G. A. The Computer Simulation of Proton Transport in Water. *J. Chem. Phys.* **1999**, *111*, 9361–9381.
- (9) Marx, D. Proton Transfer 200 Years after von Grothuss: Insights from Ab Initio Simulations. *ChemPhysChem* **2006**, *7*, 1848–1870.
- (10) Morrone, J. A.; Tuckerman, M. E. Ab Initio Molecular Dynamics Study of Proton Mobility in Liquid Methanol. *J. Chem. Phys.* **2002**, *117*, 4403–4413.
- (11) Ekimova, M.; Hoffmann, F.; Bekçioğlu-Neff, G.; Rafferty, A.; Kornilov, O.; Nibbering, E. T. J.; Sebastiani, D. Ultrafast Proton Transport between a Hydroxy Acid and a Nitrogen Base Along Solvent Bridges Governed by the Hydroxide/Methoxide Transfer Mechanism. *J. Am. Chem. Soc.* **2019**, *141*, 14581–14592.
- (12) Andot, K.; Hynes, J. T. HCl Acid Ionization in Water - a Theoretical Molecular Modeling. *J. Mol. Liq.* **1995**, *64*, 25–37.
- (13) Geissler, P. L.; Dellago, C.; Chandler, D.; Hutter, J.; Parrinello, M. Autoionization in Liquid Water. *Science* **2001**, *291*, 2121–2124.
- (14) Park, J. M.; Laio, A.; Iannuzzi, M.; Parrinello, M. Dissociation Mechanism of Acetic Acid in Water. *J. Am. Chem. Soc.* **2006**, *128*, 11318–11319.
- (15) Hassanali, A.; Giberti, F.; Cuny, J.; Kühne, T. D.; Parrinello, M. Proton Transfer through the Water Gossamer. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 13723–13728.
- (16) Pines, D.; Ditikovich, J.; Mukra, T.; Miller, Y.; Kiefer, P. M.; Daschakraborty, S.; Hynes, J. T.; Pines, E. How Acidic Is Carbonic Acid? *J. Phys. Chem. B* **2016**, *120*, 2440–2451.
- (17) 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.
- (18) Mohammed, O. F.; Pines, D.; Nibbering, E. T. J.; Pines, E. Base-Induced Solvent Switches in Acid-Base Reactions. *Angew. Chem., Int. Ed.* **2007**, *46*, 1458–1469.
- (19) 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.
- (20) 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.
- (21) Thomas, V.; Rivard, U.; Maurer, P.; Bruhacs, A.; Siwick, B. J.; Iftimie, R. Concerted and Sequential Proton Transfer Mechanisms in Water-Separated Acid-Base Encounter Pairs. *J. Phys. Chem. Lett.* **2012**, *3*, 2633–2637.
- (22) Rivard, U.; Thomas, V.; Bruhacs, A.; Siwick, B.; Iftimie, R. Donor-Bridge-Acceptor Proton Transfer in Aqueous Solution. *J. Phys. Chem. Lett.* **2014**, *5*, 3200–3205.
- (23) Vilčiauskas, 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.
- (24) Long, Z. R.; Atsango, A. O.; Napoli, J. A.; Markland, T. E.; Tuckerman, M. E. Elucidating the Proton Transport Pathways in Liquid Imidazole with First-Principles Molecular Dynamics. *J. Phys. Chem. Lett.* **2020**, *11*, 6156–6163.
- (25) Luduena, G. A.; Kühne, T. D.; Sebastiani, D. Mixed Grothuss and Vehicle Transport Mechanism in Proton Conducting Polymers from Ab Initio Molecular Dynamics Simulations. *Chem. Mater.* **2011**, *23*, 1424–1429.
- (26) Cosby, T.; Vicars, Z.; Heres, M.; Sangoro, J. Associating Imidazoles: Elucidating the Correlation between the Static Dielectric Permittivity and Proton Conductivity. *Phys. Rev. Lett.* **2018**, *120*, 136001.
- (27) Kreuer, K. D.; Fuchs, A.; Ise, M.; Spaeth, M.; Maier, J. Imidazole and Pyrazole-Based Proton Conducting Polymers and Liquids. *Electrochim. Acta* **1998**, *43*, 1281–1288.
- (28) 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.
- (29) Melchior, J. P.; Majer, G.; Kreuer, K. D. Why Do Proton Conducting Polybenzimidazole Phosphoric Acid Membranes Perform Well in High-Temperature PEM Fuel Cells? *Phys. Chem. Chem. Phys.* **2017**, *19*, 601–612.
- (30) Grigorieff, N.; Ceska, T. A.; Downing, K. H.; Baldwin, J. M.; Henderson, R. Electron-Crystallographic Refinement of the Structure of Bacteriorhodopsin. *J. Mol. Biol.* **1996**, *259*, 393–421.
- (31) Decoursey, T. E. Voltage-Gated Proton Channels and Other Proton Transfer Pathways. *Physiol. Rev.* **2003**, *83*, 475–579.
- (32) Hu, F. H.; Luo, W. B.; Hong, M. Mechanisms of Proton Conduction and Gating in Influenza M2 Proton Channels from Solid-State NMR. *Science* **2010**, *330*, 505–508.
- (33) Okada, A.; Miura, T.; Takeuchi, H. Protonation of Histidine and Histidine-Tryptophan Interaction in the Activation of the M2 Ion Channel from Influenza A Virus. *Biochemistry* **2001**, *40*, 6053–6060.
- (34) Acharya, R.; Carnevale, V.; Fiorin, G.; Levine, B. G.; Polishchuk, A. L.; Balannik, V.; Samish, I.; Lamb, R. A.; Pinto, L. H.; DeGrado, W. F.; Klein, M. L. Structure and Mechanism of Proton Transport through the Transmembrane Tetrameric M2 Protein Bundle of the Influenza A Virus. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 15075–15080.
- (35) 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.
- (36) Mason, S. F. The Tautomerism of N-Heteroaromatic Hydroxy-Compounds. 3. Ionisation Constants. *J. Chem. Soc.* **1958**, 674–685.
- (37) Mason, S. F.; Philp, J.; Smith, B. E. Prototropic Equilibria of Electronically Excited Molecules. 2. 3-, 6-, and 7-Hydroxyquinoline. *J. Chem. Soc. A* **1968**, 3051–3056.
- (38) Lee, S.-I.; Jang, D.-J. Proton Transfers of Aqueous 7-Hydroxyquinoline in the First Excited Singlet, Lowest Triplet, and Ground-States. *J. Phys. Chem.* **1995**, *99*, 7537–7541.
- (39) Tokumura, K.; Natsume, M.; Nakagawa, T.; Hashimoto, M.; Yuzawa, T.; Hamaguchi, H.; Itoh, M. Time-Resolved Infrared Study of Ground-State Phototautomer Formed in the Excited-State Proton Transfer of 7-Hydroxyquinoline in Methanol. *Chem. Phys. Lett.* **1997**, *271*, 320–326.
- (40) Bardez, E. Excited-State Proton Transfer in Bifunctional Compounds. *Isr. J. Chem.* **1999**, *39*, 319–332.
- (41) Kwon, O.-H.; Mohammed, O. F. Water-Wire Catalysis in Photoinduced Acid-Base Reactions. *Phys. Chem. Chem. Phys.* **2012**, *14*, 8974–8980.
- (42) Hoffmann, F.; Ekimova, M.; Bekçioğlu-Neff, G.; Nibbering, E. T. J.; Sebastiani, D. 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.
- (43) Codescu, M.-A.; Weiss, M.; Brehm, M.; Kornilov, O.; Sebastiani, D.; Nibbering, E. T. J. Switching between Proton Vacancy and Excess Proton Transfer Pathways in the Reaction between 7-Hydroxyquinoline and Formate. *J. Phys. Chem. A* **2021**, *125*, 1845–1859.
- (44) Amoroso, G.; Taylor, V. C. A.; Duchi, M.; Goodband, E.; Oliver, T. A. A. Following Bimolecular Excited-State Proton Transfer between Hydroxycoumarin and Imidazole Derivatives. *J. Phys. Chem. B* **2019**, *123*, 4745–4756.
- (45) Nibbering, E. T. J.; Fidler, H.; Pines, E. Ultrafast Chemistry: Using Time-Resolved Vibrational Spectroscopy for Interrogation of Structural Dynamics. *Annu. Rev. Phys. Chem.* **2005**, *56*, 337–367.
- (46) Marcus, R. A. Theoretical Relations among Rate Constants Barriers and Bronsted Slopes of Chemical Reactions. *J. Phys. Chem.* **1968**, *72*, 891–899.
- (47) Pines, E.; Fleming, G. R. Proton-Transfer in Mixed Water Organic-Solvent Solutions - Correlation between Rate, Equilibrium-



- Constant, and the Proton Free-Energy of Transfer. *J. Phys. Chem.* **1991**, *95*, 10448–10457.
- (48) Pines, E. The Kinetic Isotope Effect in the Photo-Dissociation Reaction of Excited-State Acids in Aqueous Solutions. In *Isotope Effects in Chemistry and Biology*; Kohen, A., Limbach, H.-H., Eds.; CRC Taylor & Francis: Boca Raton, FL, 2006; pp 451–464.
- (49) 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.
- (50) Munitz, N.; Avital, Y.; Pines, D.; Nibbering, E. T. J.; Pines, E. Cation-Enhanced Deprotonation of Water by a Strong Photobase. *Isr. J. Chem.* **2009**, *49*, 261–272.
- (51) Prémont-Schwarz, M.; Barak, T.; Pines, D.; Nibbering, E. T. J.; Pines, E. Ultrafast Excited-State Proton-Transfer Reaction of 1-Naphthol-3,6-Disulfonate and Several 5-Substituted 1-Naphthol Derivatives. *J. Phys. Chem. B* **2013**, *117*, 4594–4603.
- (52) Spies, C.; Shomer, S.; Finkler, D.; Pines, D.; Pines, E.; Jung, G.; Huppert, D. Solvent Dependence of Excited-State Proton Transfer from Pyranine-Derived Photoacids. *Phys. Chem. Chem. Phys.* **2014**, *16*, 9104–9114.
- (53) Novak, A. Hydrogen Bonding in Solids. Correlation of Spectroscopic and Crystallographic Data. *Struct. Bonding (Berlin)* **1974**, *18*, 177–216.
- (54) Hadži, D.; Bratos, S. Vibrational Spectroscopy of the Hydrogen Bond. In *The Hydrogen Bond: Recent Developments in Theory and Experiments*; Schuster, P., Zundel, G., Sandorfy, C., Eds.; Structure and Spectroscopy; North Holland: Amsterdam, 1976; Vol. II, pp 565–611.
- (55) Mikenda, W.; Steinböck, S. Stretching Frequency vs. Bond Distance Correlation of Hydrogen Bonds in Solid Hydrates: A Generalized Correlation Function. *J. Mol. Struct.* **1996**, *384*, 159–163.
- (56) Libowitzky, E. Correlation of O-H Stretching Frequencies and O-HO Hydrogen Bond Lengths in Minerals. *Monatsh. Chem.* **1999**, *130*, 1047–1059.
- (57) Liu, X. M.; Yu, J. S. Characterization of the Dielectric Properties of Water and Methanol in the D-Band Using a Quasi-Optical Spectroscopy. *Sci. Rep.* **2019**, *9*, 18562.
- (58) Neese, F. Software Update: The ORCA Program System-Version 5.0. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2022**, *12*, e1606.
- (59) Konijnenberg, J.; Ekkelmans, G. B.; Huizer, A. H.; Varma, C. Mechanism and Solvent Dependence of the Solvent-Catalyzed Pseudo-Intramolecular Proton-Transfer of 7-Hydroxyquinoline in the 1st Electronically Excited Singlet State and in the Ground-State of Its Tautomer. *J. Chem. Soc. Faraday Trans. 2* **1989**, *85*, 39–51.
- (60) Walba, H.; Isensee, R. W. Acidity Constants of Some Arylimidazoles and Their Cations. *J. Org. Chem.* **1961**, *26*, 2789–2791.
- (61) Rived, F.; Roses, M.; Bosch, E. Dissociation Constants of Neutral and Charged Acids in Methyl Alcohol. The Acid Strength Resolution. *Anal. Chim. Acta* **1998**, *374*, 309–324.
- (62) Rived, F.; Canals, I.; Bosch, E.; Rosés, M. Acidity in Methanol-Water. *Anal. Chim. Acta* **2001**, *439*, 315–333.
- (63) Cox, B. G. *Acids and Bases: Solvent Effects on Acid-Base Strengths*; Oxford University Press: Oxford, U.K., 2013.
- (64) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular-Dynamics. *J. Comput. Phys.* **1995**, *117*, 1–19.
- (65) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A Package for Building Initial Configurations for Molecular Dynamics Simulations. *J. Comput. Chem.* **2009**, *30*, 2157–2164.
- (66) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- (67) Golze, D.; Hutter, J.; Iannuzzi, M. Wetting of Water on Hexagonal Boron Nitride@Rh(111): A QM/MM Model Based on Atomic Charges Derived from Nano-Structured Substrates. *Phys. Chem. Chem. Phys.* **2015**, *17*, 14307–14316.
- (68) Nosé, S. A Unified Formulation of the Constant Temperature Molecular-Dynamics Methods. *J. Chem. Phys.* **1984**, *81*, 511–519.
- (69) Nosé, S. A Molecular-Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984**, *52*, 255–268.
- (70) Martyna, G. J.; Klein, M. L.; Tuckerman, M. Nosé-Hoover Chains - the Canonical Ensemble Via Continuous Dynamics. *J. Chem. Phys.* **1992**, *97*, 2635–2643.
- (71) Brehm, M.; Kirchner, B. TRAVIS - a Free Analyzer and Visualizer for Monte Carlo and Molecular Dynamics Trajectories. *J. Chem. Inf. Model.* **2011**, *51*, 2007–2023.
- (72) Brehm, M.; Thomas, M.; Gehrke, S.; Kirchner, B. TRAVIS-a Free Analyzer for Trajectories from Molecular Simulation. *J. Chem. Phys.* **2020**, *152*, 164105.
- (73) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics Model.* **1996**, *14*, 33–38.
- (74) Stone, J. E. An Efficient Library for Parallel Ray Tracing and Animation. Master's Thesis, Missouri University of Science and Technology, Rolla, MO, 1998.
- (75) 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.

## Conclusions

This thesis deals with the atomistic computational modeling of structural and dynamic properties of peptides and peptide-derived polymers. The central focus of the work is on the formation of secondary structure elements based on elementary interactions at the level of the primary structure, with the help of enhanced atomistic molecular dynamics simulations. At this level, one of the challenges for theoretical/computational approaches is that systems of increasingly complex intrinsic structure exhibit increasingly large fluctuations in their microstructure and microdynamics. As a consequence, experimentally accessible (physico-chemical) properties are ensemble averages of increasingly complex phase space structures, where atomistic parameters (e.g. typical hydrogen bonding distances) can no longer be derived from experimental data. In turn, the quantitative calculation of such experimental data on the basis of atomistic approaches becomes increasingly difficult, since their phase space convergence becomes more resource-intensive to achieve. In this thesis, this challenge is addressed by a step-wise adaptation of the complexity of the simulated structures, starting from small molecules in a multi-component solvent, via local structures of hybrid synthetic/biochemical copolymers, up to large-scale conformational sampling of solvated polypeptides at the transition to coarse-grained simulation techniques.

Specifically, the properties that were addressed in these areas in the present thesis work are (a) microsolvation phenomena, including local solvent micro-phase-separation; (b) the impact of structural flexibility (controlled by chain length variations) on the ability to form secondary structures; and (c) the accuracy of enhanced phase space sampling techniques for a highly flexible solvated polypeptide using coarse-grained simulations using effective interaction potentials, and their back-transformation to realistic atomistic conformations.

The first system investigates the insertion of flexible PE segments into peptides. This insertion of hydrophobic segments disrupts the hydrogen bonds of the peptide backbone and reduces or ultimately eliminates secondary structure formation of the peptide. Notably, as the complexity and size of the system increases, a single hydrogen bond makes a smaller contribution to the overall chemical behaviour. However, when studying a specific hydrogen bond, it is possible to model the system at a more sophisticated level to increase accuracy, as it was done in Chap. 7.

The second system investigates the efficient simulation of larger rather than smaller systems by combining two complementary techniques of coarse-grained Monte Carlo simulations and classical all-atom Molecular Dynamics simulations. This approach exploits the advantages of both methods, a good and efficient phase-space sampling from the coarse-grained MC simulations, while the MD simulation provides more degrees of freedom and explicit solvation to obtain more accurate results. This makes it possible to study large biomolecules at the all-atom level, which is usually limited by computational cost. An algorithm has been developed to convert structures between the coarse-grained and atomistic models. In practice, the majority of coarse-grained low energy structures obtained by MC simulations retained their secondary structure after the short MD simulation. In addition, an energy conversion parameter was obtained to give a physical energy scale to the coarse-grained structures.

This work lays the groundwork for further research on large biomolecular systems, especially for medical and material science purposes, as a simple but robust conversion algorithm makes it possible to investigate large systems in a more time-efficient manner.

## References

- 1 T. Kunze, C. Dreßler, and D. Sebastiani, "Secondary Structure Formation in Hybrid Synthetic/Peptide Polymers: Insights from Molecular Dynamics Simulations," *Macromol. Theory Simulations*, Vol. 32, No. 3, 2023, pp. 1–8. doi: 10.1002/mats.202200070 cited on p. iii
- 2 M.-A. Codescu, T. Kunze, M. Weiß, M. Brehm, O. Kornilov, D. Sebastiani, and E. T. J. Nibbering, "Ultrafast Proton Transfer Pathways Mediated by Amphoteric Imidazole," *J. Phys. Chem. Lett.*, Vol. 14, No. 20, May 2023, pp. 4775–4785. doi: 10.1021/acs.jpcllett.3c00595 cited on pp. iii, 9
- 3 T. Kunze, C. Dreßler, C. Lauer, W. Paul, and D. Sebastiani, "Reverse Mapping of Coarse Grained Polyglutamine Conformations from PRIME20 Sampling," *ChemPhysChem*, Vol. 25, No. 9, May 2024. doi: 10.1002/cphc.202300521 cited on p. iii
- 4 T. Kunze, C. Dressler, C. Lauer, and D. Sebastiani, "Assignment of a Physical Energy Scale for the Dimensionless Interaction Energies within the PRIME20 Peptide Model," *ChemPhysChem*, Vol. e202400592, August 2024. doi: 10.1002/cphc.202400592 cited on p. iii
- 5 Union European, *THE TREATY ON THE FUNCTIONING OF THE EUROPEAN UNION*, 2012. cited on p. 1
- 6 Union European, "SUBMISSION BY SPAIN AND THE EUROPEAN COMMISSION ON BEHALF OF THE EUROPEAN UNION AND ITS MEMBER STATES," March 2020, 2023, pp. 1–9. cited on p. 1
- 7 S. A. Miller, A. Horvath, and P. J. Monteiro, "Readily implementable techniques can cut annual CO2 emissions from the production of concrete by over 20%," *Environ. Res. Lett.*, Vol. 11, No. 7, 2016. doi: 10.1088/1748-9326/11/7/074029 cited on p. 1
- 8 C. M. Heveran, S. L. Williams, J. Qiu, J. Artier, M. H. Hubler, S. M. Cook, J. C. Cameron, and W. V. Srubar, "Biom mineralization and Successive Regeneration of Engineered Living Building Materials," *Matter*, Vol. 2, No. 2, 2020, pp. 481–494. doi: 10.1016/j.matt.2019.11.016 cited on p. 1
- 9 N. I. Ibrahim, F. S. Shahar, M. T. H. Sultan, A. U. M. Shah, S. N. A. Safri, and M. H. M. Yazik, "Overview of Bioplastic Introduction and Its Applications in Product Packaging," *Coatings*, Vol. 11, No. 1423, November 2021. doi: 10.2307/j.ctv11cw45p.12 cited on p. 1

cited on p. 1



- 10 J. Nilsen-Nygaard, E. N. Fernández, T. Radusin, B. T. Rotabakk, J. Sarfraz, N. Sharmin, M. Sivertsvik, I. Sone, and M. K. Pettersen, "Current status of biobased and biodegradable food packaging materials: Impact on food quality and effect of innovative processing technologies," *Compr. Rev. Food Sci. Food Saf.*, Vol. 20, No. 2, March 2021, pp. 1333–1380. doi: 10.1111/1541-4337.12715
- 11 H. Moustafa, A. M. Youssef, N. A. Darwish, and A. I. Abou-Kandil, "Eco-friendly polymer composites for green packaging: Future vision and challenges," *Compos. Part B Eng.*, Vol. 172, No. May, 2019, pp. 16–25. doi: 10.1016/j.compositesb.2019.05.048 cited on p. 1
- 12 E. M. Bradbury, B. G. Carpenter, and H. Goldman, "Conformational studies of polymers and copolymers of L-aspartate esters. I. Preparation and solution studies," *Biopolymers*, Vol. 6, No. 6, 1968, pp. 837–850. doi: 10.1002/bip.1968.360060607 cited on pp. 1, 19
- 13 A. I. Arunkumar, T. K. Kumar, T. Sivaraman, and C. Yu, "Acetonitrile-induced conformational transitions in poly-L-lysine," *Int. J. Biol. Macromol.*, Vol. 21, No. 4, 1997, pp. 299–305. doi: 10.1016/S0141-8130(97)00065-2 cited on pp. 1, 19
- 14 A. I. Arunkumar, T. K. S. Kumar, and C. Yu, "Specificity of helix-induction by 2,2,2-trifluoroethanol in polypeptides," *Int. J. Biol. Macromol.*, Vol. 21, No. 3, 1997, pp. 223–230. doi: 10.1016/S0141-8130(97)00064-0 cited on pp. 1, 19
- 15 M. Fändrich and C. M. Dobson, "The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation," *EMBO J.*, Vol. 21, No. 21, 2002, pp. 5682–5690. doi: 10.1093/emboj/cdf573 cited on pp. 2, 19, 23
- 16 A. Lavasanifar, J. Samuel, and G. S. Kwon, "Poly(ethylene oxide)-block-poly(L-amino acid) micelles for drug delivery," *Adv. Drug Deliv. Rev.*, Vol. 54, No. 2, 2002, pp. 169–190. doi: 10.1016/S0169-409X(02)00015-7 cited on pp. 2, 19, 23
- 17 A. Lalatsa, A. G. Schätzlein, M. Mazza, T. B. H. Le, and I. F. Uchegbu, "Amphiphilic poly(l-amino acids) - New materials for drug delivery," *J. Control. Release*, Vol. 161, No. 2, 2012, pp. 523–536. doi: 10.1016/j.jconrel.2012.04.046 cited on pp. 2, 19, 23
- 18 N. Liu, B. Li, C. Gong, Y. Liu, Y. Wang, and G. Wu, "Colloids and Surfaces B: Biointerfaces A pH- and thermo-responsive poly ( amino acid )-based drug delivery system," *Colloids Surf., B*, Vol. 136, 2015, pp. 562–569. cited on pp. 2, 19, 23
- 19 R. H. N. Kalaria and S. I. Harik, "Reduced Glucose Transporter at the Blood-Brain Barrier and in Cerebral Cortex in Alzheimer Disease," *J. Neurochem.*, Vol. 53, No. 4, 1989, pp. 1083–1088. doi: 10.1111/j.1471-4159.1989.tb07399.x cited on p. 2
- 20 C. M. Lill and C. Klein, "Epidemiologie und Ursachen der Parkinson-Erkrankung," *Nervenarzt*, Vol. 88, No. 4, 2017, pp. 345–355. doi: 10.1007/s00115-017-0288-0 cited on p. 2
- 21 H. Y. Zoghbi and H. T. Orr, "Glutamine Repeats and Neurodegeneration," *Annu. Rev. Neurosci.*, Vol. 23, No. 1, March 2000, pp. 217–247. doi: 10.1146/annurev.neuro.23.1.217 cited on p. 2

- 22 A. M. Morris, M. A. Watzky, and R. G. Finke, "Protein aggregation kinetics, mechanism, and curve-fitting: A review of the literature," *Biochim. Biophys. Acta - Proteins Proteomics*, Vol. 1794, No. 3, 2009, pp. 375–397. doi: 10.1016/j.bbapap.2008.10.016 cited on p. 2
- 23 J. A. Housmans, G. Wu, J. Schymkowitz, and F. Rousseau, "A guide to studying protein aggregation," *FEBS J.*, Vol. 290, No. 3, 2023, pp. 554–583. doi: 10.1111/febs.16312 cited on p. 2
- 24 A. Rahman, B. Saikia, C. R. Gogoi, and A. Baruah, "Advances in the understanding of protein misfolding and aggregation through molecular dynamics simulation," *Prog. Biophys. Mol. Biol.*, Vol. 175, No. August, November 2022, pp. 31–48. doi: 10.1016/j.pbiomolbio.2022.08.007 cited on p. 2
- 25 M. Karplus and J. Kuriyan, "Molecular dynamics and protein function," *Proc. Natl. Acad. Sci. U. S. A.*, Vol. 102, No. 19, 2005, pp. 6679–6685. doi: 10.1073/pnas.0408930102 cited on p. 2
- 26 M. Bendahmane, K. P. Bohannon, M. M. Bradberry, T. C. Rao, M. W. Schmidtke, P. S. Abbineni, N. L. Chon, S. Tran, H. Lin, E. R. Chapman, J. D. Knight, and A. Anantharam, "The synaptotagmin C2B domain calcium-binding loops modulate the rate of fusion pore expansion," *Mol. Biol. Cell*, Vol. 29, No. 7, 2018, pp. 834–845. doi: 10.1091/mbc.E17-11-0623 cited on p. 2
- 27 S. Sharma and M. Lindau, "Molecular mechanism of fusion pore formation driven by the neuronal SNARE complex," *Proc. Natl. Acad. Sci. U. S. A.*, Vol. 115, No. 50, 2018, pp. 12751–12756. doi: 10.1073/pnas.1816495115 cited on p. 2
- 28 R. M. Henry, C. H. Yu, T. Rödinger, and R. Pomès, "Functional Hydration and Conformational Gating of Proton Uptake in Cytochrome c Oxidase," *J. Mol. Biol.*, Vol. 387, No. 5, 2009, pp. 1165–1185. doi: 10.1016/j.jmb.2009.02.042 cited on p. 2
- 29 L. K. Scarbath-Evers, S. Jähnigen, H. Elgabarty, C. Song, R. Narikawa, J. Matysik, and D. Sebastiani, "Structural heterogeneity in a parent ground-state structure of AnPixJg2 revealed by theory and spectroscopy," *Phys. Chem. Chem. Phys.*, Vol. 19, No. 21, 2017, pp. 13882–13894. doi: 10.1039/c7cp01218g cited on p. 2
- 30 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.*, Vol. 8, No. 19, September 2017, pp. 4740–4745. doi: 10.1021/acs.jpclett.7b02317 cited on p. 2
- 31 D. Rosenberger, M. Hanke, and N. F. van der Vegt, "Comparison of iterative inverse coarse-graining methods," *Eur. Phys. J. Spec. Top.*, Vol. 225, No. 8-9, 2016, pp. 1323–1345. doi: 10.1140/epjst/e2016-60120-1 cited on p. 2
- 32 H. J. Risselada and S. J. Marrink, "Curvature effects on lipid packing and dynamics in liposomes revealed by coarse grained molecular dynamics simulations," *Phys. Chem. Chem. Phys.*, Vol. 11, No. 12, 2009, pp. 2056–2067. doi: 10.1039/b818782g cited on p. 2

cited on p. 2

- 33 E. Brini, V. Marcon, and N. F. Van Der Vegt, "Conditional reversible work method for molecular coarse graining applications," *Phys. Chem. Chem. Phys.*, Vol. 13, No. 22, 2011, pp. 10 468–10 474. doi: 10.1039/c0cp02888f
- 34 D. Reith, M. Pütz, and F. Müller-Plathe, "Deriving effective mesoscale potentials from atomistic simulations," *J. Comput. Chem.*, Vol. 24, No. 13, October 2003, pp. 1624–1636. doi: 10.1002/jcc.10307 cited on p. 2
- 35 A. P. Lyubartsev and A. Laaksonen, "Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach," *Phys. Rev. E*, Vol. 52, No. 4, October 1995, pp. 3730–3737. doi: 10.1103/PhysRevE.52.3730 cited on p. 2
- 36 S. Izvekov and G. A. Voth, "A multiscale coarse-graining method for biomolecular systems," *J. Phys. Chem. B*, Vol. 109, No. 7, 2005, pp. 2469–2473. doi: 10.1021/jp044629q cited on p. 2
- 37 J. W. Mullinax and W. G. Noid, "A Generalized-Yvon - Born - Green Theory for Determining Coarse-Grained Interaction," *J. Phys. Chem. C*, Vol. 114, No. 12, 2010, pp. 5661–5674. cited on p. 2
- 38 I. Kurisaki and S. Tanaka, "Computational prediction of heteromeric protein complex disassembly order using hybrid Monte Carlo/molecular dynamics simulation," *Phys. Chem. Chem. Phys.*, Vol. 24, No. 17, 2022, pp. 10 575–10 587. doi: 10.1039/d2cp00267a cited on p. 2
- 39 M. S. Barhaghi, B. Crawford, G. Schwing, D. J. Hardy, J. E. Stone, L. Schwiebert, J. Potoff, and E. Tajkhorshid, "Py-MCMD: Python Software for Performing Hybrid Monte Carlo/Molecular Dynamics Simulations with GOMC and NAMD," *J. Chem. Theory Comput.*, Vol. 18, No. 8, 2022, pp. 4983–4994. doi: 10.1021/acs.jctc.1c00911 cited on p. 2
- 40 H. J. Woo, A. R. Dinner, and B. Roux, "Grand canonical Monte Carlo simulations of water in protein environments," *J. Chem. Phys.*, Vol. 121, No. 13, 2004, pp. 6392–6400. doi: 10.1063/1.1784436 cited on p. 2
- 41 I. Y. Ben-Shalom, C. Lin, T. Kurtzman, R. C. Walker, and M. K. Gilson, "Simulating Water Exchange to Buried Binding Sites," *J. Chem. Theory Comput.*, Vol. 15, No. 4, 2019, pp. 2684–2691. doi: 10.1021/acs.jctc.8b01284 cited on p. 2
- 42 M. S. Bodnarchuk, M. J. Packer, and A. Haywood, "Utilizing Grand Canonical Monte Carlo Methods in Drug Discovery," *ACS Med. Chem. Lett.*, Vol. 11, No. 1, 2020, pp. 77–82. doi: 10.1021/acsmchemlett.9b00499 cited on p. 2
- 43 G. A. Ross, E. Russell, Y. Deng, C. Lu, E. D. Harder, R. Abel, and L. Wang, "Enhancing Water Sampling in Free Energy Calculations with Grand Canonical Monte Carlo," *J. Chem. Theory Comput.*, Vol. 16, No. 10, 2020, pp. 6061–6076. doi: 10.1021/acs.jctc.0c00660 cited on p. 2
- 44 S. Pylaeva, A. Böker, H. Elgabarty, W. Paul, and D. Sebastiani, "The Conformational Ensemble of Polyglutamine-14 Chains: Specific Influences of Solubility Tail and Chromophores," *ChemPhysChem*, Vol. 19, No. 21, September 2018, pp. 2931–2937. doi: 10.1002/cphc.201800551 cited on p. 2

- 45 W. G. Noid, J.-W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen, "The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models," *J. Chem. Phys.*, Vol. 128, No. 24, June 2008. doi: 10.1063/1.2938860 cited on p. 2
- 46 E. Brini and N. F. van der Vegt, "Chemically transferable coarse-grained potentials from conditional reversible work calculations," *J. Chem. Phys.*, Vol. 137, No. 154113, 2012. cited on p. 2
- 47 P. Ganguly and N. F. A. van der Vegt, "Representability and Transferability of Kirkwood Buffer Iterative Boltzmann Inversion Models for Multicomponent Aqueous Systems," *J. Chem. Theory Comput.*, Vol. 9, No. 12, 2013, pp. 5247–5256. cited on p. 2
- 48 L. C. Jacobson, R. M. Kirby, and V. Molinero, "How Short Is Too Short for the Interactions of a Water Potential? Exploring the Parameter Space of a Coarse-Grained Water Model Using Uncertainty Quantification," *J. Phys. Chem. B*, Vol. 118, No. 28, 2014, pp. 8190–8202. cited on p. 2
- 49 J.-w. Shen, C. Li, N. F. van der Vegt, and C. Peter, "Transferability of Coarse Grained Potentials: Implicit Solvent Models for Hydrated Ions," *J. Chem. Theory Comput.*, Vol. 7, No. 6, 2011, pp. 1916–1927. cited on p. 2
- 50 M. Langeloth, T. Sugii, M. C. Böhm, and F. Müller-plathe, "The glass transition in cured epoxy thermosets: A comparative molecular dynamics study in coarse-grained and atomistic resolution," *J. Chem. Phys.*, Vol. 143, No. 243158, 2015. doi: 10.1063/1.4937627 cited on p. 2
- 51 S. Jain, S. Garde, and S. K. Kumar, "Do Inverse Monte Carlo Algorithms Yield Thermodynamically Consistent Interaction Potentials?" *Ind. Eng. Chem. Res.*, Vol. 45, No. 16, August 2006, pp. 5614–5618. doi: 10.1021/ie060042h cited on p. 2
- 52 C.-C. Fu, P. Kulkarni, S. Shell, and G. Leal, "A test of systematic coarse-graining of molecular dynamics simulations: Thermodynamic properties," *J. Chem. Phys.*, Vol. 137, No. 164106, 2012. cited on p. 2
- 53 A. Fulst and C. Schwermann, *Molekulardynamiksimulation*. Westfälische Wilhelms-Universität Münster, 2014, pp. 1–28. cited on pp. 3, 4
- 54 L. Verlet, "Computer "experiments" on classical fluids. II. Equilibrium correlation functions," *Phys. Rev.*, Vol. 165, No. 1, 1968, pp. 201–214. doi: 10.1103/PhysRev.165.201 cited on p. 3
- 55 J. Leszczynski, *Handbook of computational chemistry*. 2012, pp. 1–1430. doi: 10.1007/978-94-007-0711-5 cited on p. 4
- 56 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.*, Vol. 76, No. 1, January 1982, pp. 637–649. doi: 10.1063/1.442716 cited on p. 4
- 57 R. W. Hockney, "Plasma Physics: Methods in Computational Physics," Vol. 9, 1970, pp. 135–211. cited on p. 4  
cited on p. 6

- 58 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *J. Chem. Phys.*, Vol. 81, No. 8, October 1984, pp. 3684–3690. doi: 10.1063/1.448118
- 59 G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," *J. Chem. Phys.*, Vol. 126, No. 1, 2007. doi: 10.1063/1.2408420 cited on p. 7
- 60 G. Bussi, "Hamiltonian replica exchange in GROMACS : a flexible implementation," *Mol. Phys.*, Vol. 112, No. 3–4, 2014, pp. 379–384. doi: 10.1080/00268976.2013.824126 cited on pp. 7, 23
- 61 E. Hruska, J. R. Abella, F. Nüske, L. E. Kavraki, and C. Clementi, "Quantitative comparison of adaptive sampling methods for protein dynamics," *J. Chem. Phys.*, Vol. 149, No. 24, December 2018. doi: 10.1063/1.5053582 cited on p. 7
- 62 Y. Sugita and Y. Okamoto, "Replica-exchange molecular dynamics method for protein folding," *Chem. Phys. Lett.*, Vol. 314, No. 1-2, November 1999, pp. 141–151. doi: 10.1016/S0009-2614(99)01123-9 cited on p. 7
- 63 A. Laio and M. Parrinello, "Escaping free-energy minima," *Proc. Natl. Acad. Sci.*, Vol. 99, No. 20, October 2002, pp. 12 562–12 566. doi: 10.1073/pnas.202427399 cited on p. 7
- 64 S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by Simulated Annealing," *Science (80-. )*, Vol. 220, No. 4598, May 1983, pp. 671–680. doi: 10.1126/science.220.4598.671 cited on p. 7
- 65 D. J. Earl and M. W. Deem, *Parallel tempering: Theory, applications, and new perspectives*, August 2005 URL: <http://arxiv.org/abs/physics/0508111> 20 <http://dx.doi.org/10.1039/B509983H> cited on p. 7
- 66 C. Dreßler, J. Hänseroth, and D. Sebastiani, "Coexistence of Cationic and Anionic Phosphate Moieties in Solids: Unusual but Not Impossible," *J. Phys. Chem. Lett.*, Vol. 14, No. 32, August 2023, pp. 7249–7255. doi: 10.1021/acs.jpcllett.3c01521 cited on p. 9
- 67 C. Dreßler and D. Sebastiani, "Effect of anion reorientation on proton mobility in the solid acids family CsH<sub>y</sub>XO<sub>4</sub> (X = S, P, Se, y = 1, 2) from ab initio molecular dynamics simulations," *Phys. Chem. Chem. Phys.*, Vol. 22, No. 19, 2020, pp. 10 738–10 752. doi: 10.1039/C9CP06473G cited on p. 9
- 68 R. Siewertsen, H. Neumann, B. Buchheim-Stehn, R. Herges, C. Näther, F. Renth, and F. Temps, "Highly Efficient Reversible ZE Photoisomerization of a Bridged Azobenzene with Visible Light through Resolved S<sub>1</sub> (nπ\*) Absorption Bands," *J. Am. Chem. Soc.*, Vol. 131, No. 43, November 2009, pp. 15 594–15 595. doi: 10.1021/ja906547d cited on p. 9
- 69 N. Foloppe and A. D. MacKerell, "All-Atom Empirical Force Field for Nucleic Acids: I. Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data," *J. Comput. Chem.*, Vol. 21, No. 2, 2000, pp. 86–104. doi: 10.1002/(SICI)1096-987X(20000130)21:2<86::AID-JCC2>3.0.CO;2-G cited on p. 9

cited on p. 10



- 70 A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus, "All-atom empirical potential for molecular modeling and dynamics studies of proteins," *J. Phys. Chem. B*, Vol. 102, No. 18, 1998, pp. 3586–3616. doi: 10.1021/jp973084f
- 71 A. D. Mackerell, M. Feig, and C. L. Brooks, "Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulation," *J. Comput. Chem.*, Vol. 25, No. 11, 2004, pp. 1400–1415. doi: 10.1002/jcc.20065 cited on pp. 10, 14
- 72 M. Cheon, I. Chang, and C. K. Hall, "Extending the {PRIME} model for protein aggregation to all 20 amino acids," *Proteins: Struct., Funct., Bioinf.*, Vol. 78, No. 14, July 2010, pp. 2950–2960. doi: 10.1002/prot.22817 cited on p. 12
- 73 A. Böker and W. Paul, "Thermodynamics and Conformations of Single Polyalanine, Polyserine, and Polyglutamine Chains within the PRIME20 Model," *J. Phys. Chem. B*, Vol. 126, No. 38, September 2022, pp. 7286–7297. doi: 10.1021/acs.jpcc.2c04360 cited on pp. 12, 13
- 74 M. S. Miller, W. K. Lay, S. Li, W. C. Hacker, J. An, J. Ren, and A. H. Elcock, "Reparametrization of Protein Force Field Nonbonded Interactions Guided by Osmotic Coefficient Measurements from Molecular Dynamics Simulations," *J. Chem. Theory Comput.*, Vol. 13, No. 4, April 2017, pp. 1812–1826. doi: 10.1021/acs.jctc.6b01059 cited on p. 14
- 75 W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids," *J. Am. Chem. Soc.*, Vol. 118, No. 45, November 1996, pp. 11 225–11 236. doi: 10.1021/ja9621760 cited on p. 14
- 76 J. R. Maple, M.-J. Hwang, T. P. Stockfisch, U. Dinur, M. Waldman, C. S. Ewig, and A. T. Hagler, "Derivation of class II force fields. I. Methodology and quantum force field for the alkyl functional group and alkane molecules," *J. Comput. Chem.*, Vol. 15, No. 2, February 1994, pp. 162–182. doi: 10.1002/jcc.540150207 cited on p. 14
- 77 J. Meister and W. H. E. Schwarz, "Principal Components of Ionicity," *J. Phys. Chem.*, Vol. 98, No. 33, August 1994, pp. 8245–8252. doi: 10.1021/j100084a048 cited on p. 14
- 78 R. S. Mulliken, "Electronic Population Analysis on LCAO–MO Molecular Wave Functions. I," *J. Chem. Phys.*, Vol. 23, No. 10, October 1955, pp. 1833–1840. doi: 10.1063/1.1740588 cited on p. 14
- 79 P.-O. Löwdin, "On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals," *J. Chem. Phys.*, Vol. 18, No. 3, March 1950, pp. 365–375. doi: 10.1063/1.1747632 cited on p. 14

- 80 F. L. Hirshfeld, "Bonded-atom fragments for describing molecular charge densities," *Theor. Chim. Acta*, Vol. 44, No. 2, 1977, pp. 129–138. doi: 10.1007/BF00549096
- 81 C. I. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman, "A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model," *J. Phys. Chem.*, Vol. 97, No. 40, October 1993, pp. 10 269–10 280. doi: 10.1021/j100142a004 cited on p. 15
- 82 D. P. Kroese, T. Brereton, T. Taimre, and Z. I. Botev, "Why the Monte Carlo method is so important today," *WIREs Comput. Stat.*, Vol. 6, No. 6, November 2014, pp. 386–392. doi: 10.1002/wics.1314 cited on p. 16
- 83 N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of State Calculations by Fast Computing Machines," *J. Chem. Phys.*, Vol. 21, No. 6, June 1953, pp. 1087–1092. doi: 10.1063/1.1699114 cited on p. 17
- 84 F. Liang, C. L. Liu, and R. J. Carroll, "Stochastic approximation in Monte Carlo computation," *J. Am. Stat. Assoc.*, Vol. 102, No. 477, 2007, pp. 305–320. doi: 10.1198/016214506000001202 cited on pp. 17, 18
- 85 G. Torrie and J. Valleau, "Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling," *J. Comput. Phys.*, Vol. 23, No. 2, February 1977, pp. 187–199. doi: 10.1016/0021-9991(77)90121-8 cited on p. 18
- 86 U. H. E. Hansmann and Y. Okamoto, "Finite-size scaling of helix–coil transitions in poly-alanine studied by multicanonical simulations," *J. Chem. Phys.*, Vol. 110, No. 2, January 1999, pp. 1267–1276. doi: 10.1063/1.478169 cited on p. 18
- 87 M. Bonomi and M. Parrinello, "Enhanced Sampling in the Well-Tempered Ensemble," *Phys. Rev. Lett.*, Vol. 104, No. 19, May 2010, p. 190 601. doi: 10.1103/PhysRevLett.104.190601 cited on p. 18
- 88 F. Wang and D. P. Landau, "Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States," *Phys. Rev. Lett.*, Vol. 86, No. 10, March 2001, pp. 2050–2053. doi: 10.1103/PhysRevLett.86.2050 cited on p. 18
- 89 F. Liang, "A Theory on Flat Histogram Monte Carlo Algorithms," *J. Stat. Phys.*, Vol. 122, No. 3, February 2006, pp. 511–529. doi: 10.1007/s10955-005-8016-8 cited on p. 18
- 90 T. Shakirov, S. Zablotskiy, A. Böker, V. Ivanov, and W. Paul, "Comparison of Boltzmann and Gibbs entropies for the analysis of single-chain phase transitions," *Eur. Phys. J. Spec. Top.*, Vol. 226, No. 4, 2017, pp. 705–723. doi: 10.1140/epjst/e2016-60326-1 cited on p. 18
- 91 G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, "Stereochemistry of polypeptide chain configurations," *J. Mol. Biol.*, Vol. 7, No. 1, 1963, pp. 95–99. doi: 10.1016/S0022-2836(63)80023-6 cited on pp. 19, 22
- 92 A. Ambrogelly, S. Palioura, and D. Söll, *Natural expansion of the genetic code*, January 2007 URL: <https://www.nature.com/articles/nchembio847> cited on p. 19  
cited on p. 19



- 93 M. Gopalswamy, A. Kumar, J. Adler, M. Baumann, M. Henze, S. T. Kumar, M. Fändrich, H. A. Scheidt, D. Huster, and J. Balbach, "Structural characterization of amyloid fibrils from the human parathyroid hormone," *Biochim. Biophys. Acta - Proteins Proteomics*, Vol. 1854, No. 4, April 2015, pp. 249–257. doi: 10.1016/j.bbapap.2014.12.020
- 94 S. Mohan, O. S. Oluwafemi, N. Kalarikkal, S. Thomas, and S. P. Songca, "Biopolymers – Application in Nanoscience and Nanotechnology," *Recent Adv. Biopolym.*, 2015. cited on pp. 19, 23
- 95 R. Rebelo, M. Fernandes, and R. Fangueiro, "Biopolymers in Medical Implants: A Brief Review," *Procedia Eng.*, Vol. 200, 2017, pp. 236–243. doi: 10.1016/j.proeng.2017.07.034 cited on p. 19
- 96 S. Maude, L. Tai, R. Davies, B. Liu, S. Harris, P. Kocienski, and A. Aggeli, *Peptide-Based Materials*, T. Deming, Ed., ser. Topics in Current Chemistry. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, Vol. 310, p. 27. doi: 10.1007/978-3-642-27139-7 cited on p. 19
- 97 C. Alemán, A. Bianco, and M. Venanzi, *Peptide Materials*, C. Alemán, A. Bianco, and M. Venanzi, Eds. Wiley, May 2013, pp. 1–433. doi: 10.1002/9781118592403
- 98 Q. Sun, Z. Yang, and X. Qi, "Design and Application of Hybrid Polymer-Protein Systems in Cancer Therapy," *Polymers*, Vol. 15, No. 9, May 2023, p. 2219. doi: 10.3390/polym15092219 cited on p. 20
- 99 H. G. Börner and H. Schlaad, "Bioinspired functional block copolymers," *Soft Matter*, Vol. 3, No. 4, 2007, pp. 394–408. doi: 10.1039/B615985K cited on p. 20
- 100 O. D. Krishna and K. L. Kiick, "Protein- and peptide-modified synthetic polymeric biomaterials," *Pept. Sci.*, Vol. 94, No. 1, January 2010, pp. 32–48. doi: 10.1002/bip.21333 cited on p. 20
- 101 J. Y. Shu, B. Panganiban, and T. Xu, "Peptide-Polymer Conjugates: From Fundamental Science to Application," *Annu. Rev. Phys. Chem.*, Vol. 64, No. 1, April 2013, pp. 631–657. doi: 10.1146/annurev-physchem-040412-110108 cited on p. 20
- 102 M. Brehm, M. Thomas, S. Gehrke, and B. Kirchner, "TRAVIS—A free analyzer for trajectories from molecular simulation," *J. Chem. Phys.*, Vol. 152, No. 16, 2020, p. 164105. doi: 10.1063/5.0005078 cited on p. 21
- 103 D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. Berendsen, "GROMACS: Fast, flexible, and free," *J. Comput. Chem.*, Vol. 26, No. 16, 2005, pp. 1701–1718. doi: 10.1002/jcc.20291 cited on p. 21
- 104 J. Freudenberger and W. H. Binder, "Multisegmented Hybrid Polymer Based on Oligo-Amino Acids: Synthesis and Secondary Structure in Solution and in the Solid State," *Macromolecules*, Vol. 52, No. 12, 2019, pp. 4534–4544. doi: 10.1021/acs.macromol.9b00684 cited on p. 23
- 105 M. B. Canalp and W. H. Binder, "Lysine ( carboxybenzyl ) s : Synthesis and investigations of secondary structure," *RSC Adv.*, Vol. 10, No. 3, 2020, pp. 1287–1295. doi: 10.1039/c9ra09189k cited on p. 23  
cited on p. 23

- 106 M. B. Canalp, A. Meister, and W. H. Binder, "Secondary structure of end group functionalized oligomeric-L-lysines: Investigations of solvent and structure dependent helicity," *RSC Adv.*, Vol. 9, No. 38, 2019, pp. 21 707–21 714. DOI: 10.1039/c9ra03099a
- 107 A. Panáček, L. Kvítek, R. Prucek, M. Kolář, R. Večeřová, N. Pizúrová, V. K. Sharma, T. Nevěčná, and R. Zbořil, "Silver colloid nanoparticles: Synthesis, characterization, and their antibacterial activity," *J. Phys. Chem. B*, Vol. 110, No. 33, 2006, pp. 16 248–16 253. DOI: 10.1021/jp063826h cited on p. 23
- 108 V. Siracusa, I. Blanco, S. Romani, U. Tylewicz, P. Rocculi, and M. Dalla Rosa, "Poly(lactic acid)-modified films for food packaging application: Physical, mechanical, and barrier behavior," *J. Appl. Polym. Sci.*, Vol. 125, No. S2, 2012. DOI: 10.1002/app cited on p. 23
- 109 L. Wang, R. Friesner, and B. Berne, "Replica Exchange with Solute Scaling: A more efficient version of Replica Exchange with Solute Tempering (REST2)," *J Phys Chem B.*, Vol. 115, No. 30, 2011, pp. 9431–9438. DOI: 10.1038/jid.2014.371 cited on p. 23
- 110 M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia, and M. Parrinello, "PLUMED: A portable plugin for free-energy calculations with molecular dynamics," *Comput. Phys. Commun.*, Vol. 180, No. 10, 2009, pp. 1961–1972. DOI: 10.1016/j.cpc.2009.05.011 cited on p. 23

# Curriculum Vitae

## Personal Data

NAME: Thomas Kunze  
PLACE OF BIRTH:  
DATE OF BIRTH:

## Academic education

2020 – 2025	PhD student at Martin-Luther-University Halle Theoretical Chemistry Group of Prof. Sebastiani
2020	MASTER OF SCIENCE IN CHEMISTRY University of Halle Grade:
2019	BACHELOR OF SCIENCE IN CHEMISTRY University of Halle Grade:
2015	ABITUR (A LEVELS) Georg-Cantor-Gymnasium Halle Grade: Additional academic course in chemistry at Fachhochschule Merseburg

.....  
Ort, Datum / Place, Date

.....  
Signature/Unterschrift

### **Eidesstattliche Erklärung / *Declaration under Oath***

Ich erkläre an Eides statt, dass ich die Arbeit selbstständig und ohne fremde Hilfe verfasst, keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

*I declare under penalty of perjury that this thesis is my own work entirely and has been written without any help from other people. I used only the sources mentioned and included all the citations correctly both in word or content.*

---

Datum / Date

---

Unterschrift des Antragstellers / *Signature of the applicant*