Conformational equilibria and spectroscopy of gas-phase homologous peptides from first principles

im Fachbereich Physik der Freien Universität Berlin eingereichte
Dissertation zur Erlangung des akademischen Grades
DOCTOR RERUM NATURALIUM

vorgelegt von
Dipl. Phys.
Franziska Schubert

Berlin, 2014

Erstgutachter (Betreuer): Prof. Dr. Matthias Scheffler
Zweitgutachter: Prof. Dr. Roland Netz
Tag der Disputation: 21.10.2014

ABSTRACT

Peptides and proteins fulfil crucial tasks enabling and maintaining life. Their function is directly correlated with their three-dimensional structure, which is in turn determined by their chemical composition, the amino-acid sequence. Predicting the structure of a peptide based only on its sequence information is of fundamental interest. A fully first-principles treatment free of empirical parameters would be ideal. However, this presents an ongoing challenge, due to the large system size and conformational space of most peptides.

In the present work, we address this challenge concentrating on the example of polyalanine-based peptides in the gas phase. Such studies under isolated conditions follow a bottom-up approach that allows one to investigate the intramolecular interactions important for secondary structure separate from environmental effects. Furthermore, direct benchmarks of theoretical structure predictions against experiment are facilitated.

The peptide series Ac-Ala$_n$-Lys(H$^+$), ($n \gtrsim 6$), forms α-helices in the gas phase due to a favorable interaction of the helix dipole with the positive charge at the C-terminal lysine residue. Using this design principle as a template, we explore the impact of increased structural flexibility on the conformational space due to (i) sequence length [Ac-Ala$_n$-Lys(H$^+$), $n = 19$], (ii) charge placement [Ac-Ala$_{19}$-Lys(H$^+$) versus Ac-Lys(H$^+$)-Ala$_{19}$], and (iii) backbone elongation of the monomer units as represented by β-amino acids [Ac-β^2hAla$_6$-Lys(H$^+$)]. To address the large conformational space, we develop a three-step structure-search strategy employing an unprecedented first-principles screening effort. After pre-sampling of the conformational space using a force field, thousands of structures are optimized employing density-functional theory (DFT). For this, the PBE functional is used, coupled with a pairwise correction for van der Waals interactions. For the best few structure candidates, *ab initio* replica-exchange molecular-dynamics simulations are performed in order to refine the local structural environment. It is shown that these can yield lower-energy conformations and lead to rearrangements of the hydrogen-bonding network. In order to connect to experiment, collision cross sections are calculated that link to ion mobility-mass spectrometry. Furthermore, infrared spectra are derived from *ab initio* Born-Oppenheimer molecular-dynamics simulations accounting for anharmonicities within the classical-nuclei approximation.

As expected, the 20-residue peptide Ac-Ala$_{19}$-Lys(H$^+$) forms helical structures. In contrast, placing the charge at the N-terminus [Ac-Lys(H$^+$)-Ala$_{19}$], leads to several different compact structures, which are close in energy. Such small energy differences present a challenge to the theoretical approach. Incorporating exact exchange and many-body van der Waals effects predicts the presence of only one dominant conformer, which is compatible with both experimental datasets.

In comparison to Ac-Ala$_6$-Lys(H$^+$), the β-peptide Ac-β^2hAla$_6$-Lys(H$^+$) exhibits increased conformational flexibility due to an extended monomer backbone. Out of the almost 15,000 structures optimized with DFT, no helical conformers are found in the low-energy regime. This is changed when considering vibrational free energy (300 K, harmonic approximation), which strongly favors helical conformations due to softer vibrational modes. One possible structure candidate is the H16-helix, which is compatible with both experiments. It is a unique structure as it exhibits a hydrogen-bonding pattern equivalent to the α-helix of natural peptides.

The systems considered here highlight the advances of current DFT functionals to address the large conformational space of peptides, but also the need for further development.

ZUSAMMENFASSUNG

Proteine und Peptide erfüllen wichtige Aufgaben im Stoffwechsel lebender Zellen. Ihre Funktion ist direkt an ihre dreidimensionale Struktur gekoppelt, die wiederum von der chemischen Zusammensetzung (der Aminosäuresequenz) bestimmt wird. Die Vorhersage der Struktur eines Peptides mittels allein dieser Information ist von fundamentalem Interesse. Ideal wäre eine Beschreibung nur basierend auf "ersten Prinzipien" und damit frei von empirischen Parametern. Allerdings stellt dies aufgrund der großen Konformationsräume und Systemgrößen der meisten Peptide eine schwierige Aufgabe dar.

In der vorliegenden Arbeit wird dieses Problem am Beispiel Polyalanin-basierter Peptide in der Gasphase angegangen. Derartige Studien unter isolierten Bedingungen ermöglichen es, die intramolekularen Wechselwirkungen, die kritisch für die Sekundärstruktur sind, getrennt von Einflüssen der Umgebung zu untersuchen und damit von Grund auf zu verstehen. Weiterhin werden direkte Vergleiche von theoretischen Strukturvorhersagen mit experimentellen Ergebnissen ermöglicht.

Aufgrund einer günstigen elektrostatischen Wechselwirkung des Helixdipols mit der positiven Ladung am C-terminalen Lysinrest bildet die Peptidserie Ac-Ala$_n$-Lys(H$^+$), ($n \gtrsim 6$), α-Helizes in der Gasphase. Ausgehend von diesem Designprinzip wird der Einfluss erhöhter struktureller Flexibilität auf den Konformationsraum aufgrund von (i) Sequenzlänge [Ac-Ala$_n$-Lys(H$^+$), $n = 19$], (ii) Ladungsposition [Ac-Ala$_{19}$-Lys(H$^+$) versus Ac-Lys(H$^+$)-Ala$_{19}$] und (iii) Rückgratverlängerung der Monomereinheiten [β-Aminosäuren, Ac-β^2hAla$_6$-Lys(H$^+$)] untersucht. Für die Konformationssuche wird eine dreistufige Strategie entwickelt, die auf ersten Prinzipien beruht und von ernormem Umfang ist. Anschließend an eine Struktursuche mit einem Kraftfeld werden Tausende von Strukturen auf Basis von Dichtefunktionaltheorie (DFT) optimiert. Hierfür wird das durch eine paarweise van der Waals-Korrektur erweiterte PBE Funktional benutzt. Um die Strukturvorhersage zu verfeinern, werden für die niedrigst-energetischen Strukturen anschließend *ab initio replica-exchange* Molekulardynamik-Simulationen durchgeführt. Es wird gezeigt, dass hierdurch das Wasserstoffbrückennetzwerk verändert werden kann und Strukturen mit niedrigerer Energie gefunden werden können. Die gewonnenen Ergebnisse können über Kollisions-Wirkungsquerschnitte direkt mit experimentellen Ionenmobilitätsdaten verglichen werden. Weiterhin werden Infrarot-Spektren aus *ab initio* Born-Oppenheimer Molekulardynamik-Simulationen berechnet. Hierdurch werden anharmonische Effekte in der Näherung klassischer Atomkerne berücksichtigt.

Wie erwartet bildet das Peptid Ac-Ala$_{19}$-Lys(H$^+$) mit 20 Aminosäureresten α-Helizes. Im Gegensatz dazu führt die Positionierung der Ladung am N-Terminus [Ac-Lys(H$^+$)-Ala$_{19}$] zu einer Vielzahl verschiedener kompakter Strukturtypen, die alle in einem sehr engen Energiebereich liegen. Derartig kleine Energiedifferenzen stellen eine Herausforderung für die theoretische Methode dar. Nur unter Berücksichtigung von *exact exchange* und einem Vielteilchenansatz für die van der Waals-Korrektur wird die Existenz einer einzigen Struktur vorhergesagt, die mit beiden experimentellen Datensätzen kompatibel ist.

Die Verlängerung des Monomerrückgrats vergrößert den Konformationsraum des β-Peptids Ac-β^2hAla$_6$-Lys(H$^+$) im Vergleich zu Ac-Ala$_6$-Lys(H$^+$). Unter den 15.000 DFT-optimierten Strukturen werden im niederenergetischen Bereich keine helikalen Konformere gefunden. Dies ändert sich bei Berücksichtigung freier vibronischer Energie (300 K, harmonische Näherung), da hierdurch Helizes aufgrund von weicheren Vibrationsmoden stark stabilisiert werden. Ein möglicher Strukturkandidat ist die H16-Helix, die mit den Ergebnissen beider Experimente kompatibel ist. Es ist hervorzuheben, dass diese Struktur dasselbe Wasserstoffbrückennetzwerk aufweist wie die α-Helix in natürlichen Peptiden.

Die hier betrachteten Systeme stellen zum einen die Fortschritte aktueller DFT Funktionale für die Beschreibung des Konformationsraums von Peptiden heraus, heben zum anderen aber auch die Notwendigkeit weiterer Entwicklungsarbeit hervor.

Contents

List of Abbreviations v

1 Introduction 1

I Polypeptides and (free) energy surfaces 5

2 Peptides and proteins 7
- 2.1 Interactions shaping the structure of polypeptides . . . 11
- 2.2 Structure hierarchy . . . 12
- 2.3 Secondary structure . . . 13
- 2.4 Peptidic foldamers . . . 16
- 2.5 Energy landscapes . . . 20

3 Theoretical methods to describe the energy landscape 23
- 3.1 Force fields . . . 23
- 3.2 The quantum-mechanical many-body problem . . . 27
  - 3.2.1 The Born-Oppenheimer approximation . . . 27
- 3.3 Hartree-Fock method . . . 28
- 3.4 Beyond Hartree-Fock theory: electron correlation . . . 31
  - 3.4.1 Møller-Plesset perturbation theory . . . 32
  - 3.4.2 Coupled-cluster theory . . . 33
- 3.5 Density-functional theory . . . 34
  - 3.5.1 Kohn-Sham equations . . . 35
  - 3.5.2 Approximations to the exchange-correlation functional . . . 36
    - 3.5.2.1 Local-density approximation . . . 37
    - 3.5.2.2 Generalized gradient approximation . . . 38
    - 3.5.2.3 Hybrid functionals . . . 38
  - 3.5.3 Dispersion corrections to the XC-functional approximations . . . 39
    - 3.5.3.1 TS scheme . . . 41
    - 3.5.3.2 Many-body van der Waals interactions . . . 42
- 3.6 Numeric atom-centered orbitals: `FHI-aims` . . . 45

4 Exploring energy landscapes 49
- 4.1 Molecular-dynamics simulations . . . 49
  - 4.1.1 Molecular-dynamics simulations in the canonical ensemble . . . 51

4.2 Sampling techniques . . . 53
4.2.1 Basin-hopping algorithm . . . 54
4.2.2 Replica-exchange molecuar dynamics (REMD) . . . 55
4.2.3 REMD for $(NO_3)^{-1}(HNO_3)$ + H_2O . . . 58

5 Molecular vibrations **67**
5.1 Harmonic-oscillator approximation . . . 67
5.1.1 Free energy in the harmonic oscillator-rigid rotor approximation . . . 70
5.2 Infrared (IR) spectroscopy of proteins and peptides . . . 72
5.3 Infrared (IR) spectra from the dipole time autocorrelation function . . . 74

II Large polyalanine-based peptides: structure and spectroscopy **79**

6 IR spectra from *ab initio* MD **81**
6.1 Wave-function extrapolation . . . 81
6.1.1 SCF accuracy settings 0 . . . 84
6.1.2 SCF accuracy settings 1 . . . 85
6.1.3 SCF accuracy settings 2 . . . 86
6.2 Pendry reliability factor . . . 88
6.2.1 Rigid shifts along the wavenumber axis . . . 90
6.2.2 Calculation of IR spectra . . . 90
6.2.3 Influence of the convolution . . . 91
6.3 Convergence of spectra and sensitivity for different conformers . . . 93
6.4 Comparison of different time steps . . . 95
6.5 Summary . . . 97

7 Proteins and peptides in the gas phase **99**
7.1 Experimental techniques . . . 99
7.1.1 Ion-mobility mass-spectrometry (IM-MS) . . . 100
7.1.1.1 Calculation of collision cross sections . . . 101
7.1.2 Gas-phase spectroscopy . . . 102
7.1.2.1 Infrared multiphoton dissociation (IRMPD) . . . 102
7.1.2.2 Messenger technique . . . 103
7.1.2.3 IR-UV double resonance . . . 103
7.2 Alanine-based peptides and helix formation . . . 104
7.3 Ac-Ala_{19}-Lys + H^+ vs. Ac-Lys-Ala_{19} + H^+ . . . 108

8 First-principles structure predictions for Ac-Ala_{19}-Lys + H^+ vs. Ac-Lys-Ala_{19} + H^+ **113**
8.1 Assessment of conformational search strategy . . . 114
8.1.1 Global sampling of the conformational space . . . 114
8.1.2 Local refinement . . . 117
8.1.2.1 First-principles REMD simulations . . . 121
8.2 Structure classification . . . 124
8.3 Energy landscapes: Ac-Lys-Ala_{19} + H^+ vs. Ac-Ala_{19}-Lys + H^+ . . . 125
8.4 Helical models: Ac-Lys-Ala_{19} + H^+ . . . 126

8.4.1 Helical dimers: Ac-Lys-Ala$_{19}$ + H$^+$ . . . 126
8.4.2 Helical monomers: Ac-Lys-Ala$_{19}$ + H$^+$ . . . 127
8.5 Role of a higher-level force field . . . 129
8.6 Impact of different functionals and dispersion corrections . . . 131
8.7 Summary . . . 133

9 Connecting to experiment 135
9.1 Ion-mobility mass-spectrometry . . . 135
9.2 IR spectroscopy . . . 137
9.3 Summary . . . 142

III Dealing with conformational flexibility: homologous peptides 145

10 First-principles structure predictions for Ac-β^2hAla$_6$-Lys(H$^+$) 147
10.1 Equivalent H-bonding patterns in helices of α- and β-peptides . . . 148
10.2 Assessing the conformational space of Ac-β^2hAla$_6$-Lys(H$^+$) . . . 148
10.2.1 Details of conformational analysis . . . 150
10.2.2 Assessment of the OPLSAA force field . . . 150
10.2.3 Search strategy 1: basin hopping . . . 152
10.2.3.1 Unconstrained basin-hopping search . . . 156
10.2.3.2 Constrained basin hopping: H12 . . . 163
10.2.3.3 Constrained basin hopping: H16 . . . 164
10.2.4 Summary . . . 164
10.2.5 Search strategy 2: Replica-exchange MD . . . 166
10.2.5.1 *Ab initio* REMD . . . 170
10.3 Summary . . . 171

11 Conformational preferences: Ac-β^2hAla$_6$-Lys(H$^+$) vs. Ac-Ala$_6$-Lys(H$^+$) 175
11.1 Results for Ac-β^2hAla$_6$-Lys(H$^+$) . . . 175
11.2 Comparison of Ac-β^2hAla$_6$-Lys(H$^+$) and Ac-Ala$_6$-Lys(H$^+$) . . . 179
11.3 Impact of different functionals and dispersion corrections . . . 183
11.4 Summary . . . 184

12 Connecting to experiment 185
12.1 Ion mobility-mass spectrometry . . . 185
12.2 Infrared multiphoton dissociation (IRMPD) spectra . . . 186
12.3 α-Peptide Ac-Ala$_6$-Lys(H$^+$) . . . 187
12.4 β-Peptide Ac-β^2hAla$_6$-Lys(H$^+$) . . . 190
12.4.1 Discussion . . . 190

13 Conclusions and outlook 195

Appendices 199

A Extra details for part II **201**
A.1 IR spectra derived from force-field MD simulations 201
A.2 RMSD of structures relaxed with different methods 201
A.3 Ion-mobility mass-spectrometry measurements for Ac-Lys-Ala_{19} + H^+ 203
A.4 Second experimental IR spectrum for Ac-Lys-Ala_{19} + H^+ 204
A.5 Comparison of IR spectra 204

B Extra details for part III **207**
B.1 Convergence of vibrational normal modes 207
B.2 Analysis of energetic contributions 207
B.3 IRMPD spectra: raw data vs. smoothed data 207
B.4 Comparison of IR spectra based on the Pendry R-factor 209

Eidesstattliche Versicherung **211**

Curriculum vitæ **213**

Acknowledgements **215**

Bibliography **217**

List of Abbreviations

ACFD Adiabatic-connection fluctuation-dissipation 42, 43

ATD Arrival time distribution 100, 110, 136, 143, 185, 186, 188, 203

BO Born Oppenheimer 49, 67, 81

BSSE Basis set superposition error 46

CCS Collision cross section 100, 101, 105, 107, 110, 135–137, 142, 143, 148, 186–188, 190, 193, 194, 196, 197

CD Circular dichroism 18

CFDM Coupled fluctuating-dipole model 42

CI Configuration interaction 31, 33

CP Car Parrinello 82

DFT Density-functional theory 3, 4, 10, 18, 26, 34, 36, 38–41, 43–46, 53, 81, 90, 106–108, 114, 117, 119, 121, 126, 127, 129, 131, 133, 134, 148, 150, 152, 155, 157, 158, 161, 164–166, 168, 173, 175, 194, 195, 197, 201

EAF Exchange attempt frequency 66

EHSS Exact hard-sphere scattering 101, 135

ESI Electrospray ionization 100

FEL Free-electron laser 102

FES Free-energy surface 22, 64, 70

`FHI-aims` Fritz Haber Institute *ab initio* molecular simulation 26, 45–47, 117

FWHM Full width at half maximum 91, 103

GEA Gradient-expansion approximation 38

GGA Generalized gradient approximation 38–40, 47, 90

HEG Homogeneous electron gas 37

IM-MS Ion mobility-mass spectrometry 3, 100, 105–108, 110–114, 126, 129–131, 133, 135–137, 142, 143, 147, 184, 185, 187, 188, 190, 193, 195, 196, 203

IR Infrared 4, 64, 66, 67, 69, 70, 72–77, 81, 82, 86, 88, 90, 91, 93, 95, 97, 102, 103, 107, 108, 111, 127, 137, 140, 142, 143, 148, 185–188, 190, 193, 194, 196, 197, 201, 204

IRMPD Infrared multiphoton dissociation 3, 81, 88, 89, 102, 103, 108, 111–114, 126, 135, 137, 142, 143, 147, 184–186, 188, 190, 195, 204

IVR Intramolecular vibrational redistribution 103

LDA Local-density approximation 37–40, 47

LEED Low-energy electron diffraction 88

MAE Mean absolute error 42, 129

MD Molecular dynamics 2, 4, 47, 49–51, 53–55, 58, 59, 67, 75, 77, 81, 82, 86, 90, 93, 95, 97, 101, 105, 107, 108, 110, 114, 140, 143, 171, 185, 190, 193, 196, 201, 204

NAO Numeric atom-centered orbital 45, 46

NMR Nuclear magnetic resonance 18

PA Projection approximation 101, 102, 135, 188, 190

PES Potential-energy surface 2, 4, 22, 23, 26, 28, 49, 54, 64, 66, 67, 113, 114, 130, 131, 163, 176, 179, 183, 198

QHO Quantum harmonic oscillator 42, 43

REMD Replica-exchange molecular dynamics 3, 53–55, 58, 59, 61, 63, 64, 66, 114, 115, 121, 125–127, 133, 134, 148, 165, 166, 168, 170, 171, 173, 175, 193, 195, 196

RMSD Root mean square deviation 22, 115, 121, 125, 130, 131, 133, 157, 166, 173, 183

RPA Random-phase approximation 42, 43

SCF Self-consistent field 82–86, 97, 117

SGE Schrödinger equation 23, 26–29, 31, 35, 69

SPPS Solid-phase peptide synthesis 8

TJM Trajectory method 101, 135, 188, 190

vdW van der Waals 39, 40, 42, 103, 106, 129, 131, 132, 134, 183, 195, 196

XC Exchange correlation 37, 40, 42, 137, 209

ZPE Zero-point energy 59, 71

1 Introduction

Proteins are biomolecules that play a key role in virtually all biochemical processes in the cells of living organisms. They are the molecular machines that carry out the versatile and essential tasks encoded in genes: As enzymes they catalyze biochemical reactions and as ion pumps they govern the transport of ions through membranes. They are involved in signaling processes, e.g., as receptor proteins on the outside of cell membranes or as antibodies of the immune system recognizing and tagging foreign targets for destruction. Apart from that, they transport molecules within a cell or through the body to the places where they are actually needed and also have structural functions, e.g., in the cytoskeleton, the scaffold of the cell. The reason why proteins can carry out these vast and versatile amounts of tasks stems from their ability to adopt well-defined, specific shapes, where the functional groups are arranged in such a way that they can selectively interact with other molecules. Understanding the mechanisms behind this and the physical code that links the chemical formula of a protein to its actual function is an active field of research[1–3], equally challenging to biologists, chemists, and physicists.

The molecular building blocks of proteins are the amino acids, each containing an amino and a carboxy group (see Fig. 1.1a). They enclose a carbon atom that is linked to a side chain specific to each of the 20 natural amino acids. When the amino and the carboxy group of two amino acids react with each other, a peptide bond is formed. The linear polymers arising by the linkage of amino-acid residues via such peptide bonds are called (poly)peptides or proteins (see Fig. 1.1b). The continuous sequence of covalently bound atoms is referred to as the backbone of the protein, where the sequence of amino acids is known as the primary structure.

In his landmark experiments in the early 1960s, Anfinsen[4, 5] found that folding, the process that takes the protein from the denatured to its three-dimensional, functional shape, is reversible. On these grounds, he formulated his thermodynamic hypothesis[5, 6] that the three-dimensional native structure of a protein is the state where the system's free energy has its global minimum

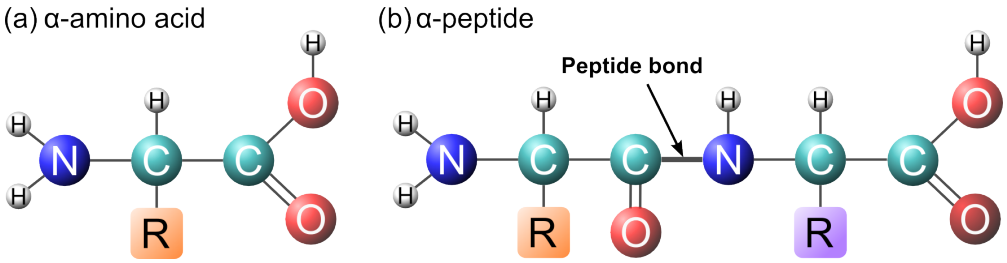


Figure 1.1: A schematic representation of (a) a natural α-amino acid with side chain R and (b) a peptide with two amino-acid residues linked by a peptide bond. The color coding (nitrogen: blue, hydrogen: white, carbon: cyan, oxygen: red) will be used throughout this thesis.

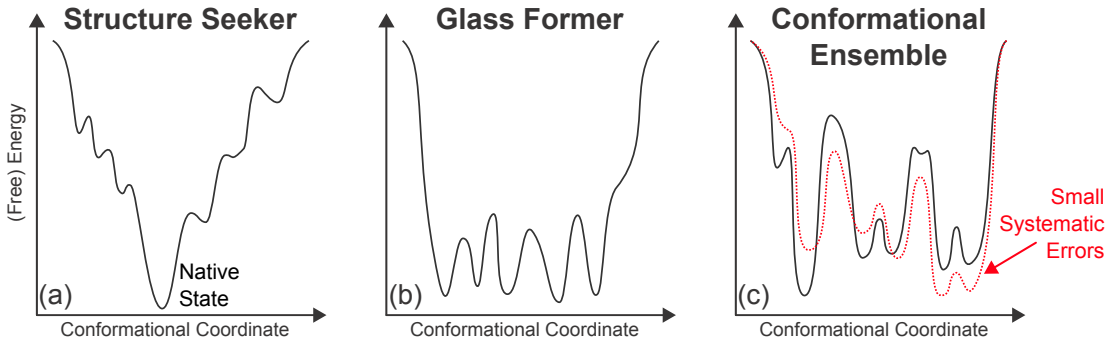


Figure 1.2: Schematic examples of the (free) energy landscape of (a) a structure seeker, (b) a glass former, and (c) a conformational ensemble.

and thus, is determined by its amino-acid sequence in a given environment. This structure should be unique, stable (against small changes of the environment), and accessible within biological time scales. Today it is accepted that folding is guided by a funnel-shaped (free) energy landscape, with the process of protein structure formation governed by the free-energy gradient[7]. The case of a structure seeker, with only one steep folding funnel pointing to one ordered native state, is but one limiting case[8] of possible free-energy landscapes (see Fig. 1.2 a). The other extreme would be a sawtooth-shaped landscape with many local minima that are similar in energy, yielding an unstructured glass former[8] (Fig. 1.2 b), depending on the temperature and the barrier heights separating the local minima. Another scenario is the conformational ensemble (Fig. 1.2 c) with a relatively flat free-energy surface exhibiting several low-energy conformers separated by barriers that are distinct but not insurmountable. Conformers that are higher in energy than the lowest-energy state, but that are still thermally accessible, may be of importance in the context of molecular recognition[9].

Experiments observing protein folding suffer from the problem that high structural resolution is very difficult to obtain together with sufficient temporal resolution. However, folding simulations based on molecular dynamics (MD) provide high-resolution temporal and structural data of the evolved trajectories[1]. On the other hand, reliable folding simulations have to deal with three basic challenges: sufficient sampling of the conformational space, a high-accuracy description of the potential-energy surface (PES) and robust data analysis[3]. While average folding times lie in the range of milliseconds,[1] to obtain accurate trajectories, the equations of motions have to be integrated with time steps of the order of femtoseconds leading to $\approx 10^{12}$ time steps in total. Furthermore, to obtain good statistics many folding events have to be sampled. This results in an excessive computational demand paired with the problem of large system sizes (approximately 10^5 atoms using explicit-solvent simulations)[3]. Recently, due to initiatives such as Folding@Home[10], a distributed-computing project where (private) people share idle computer time of their (private) resources, or ANTON, a computer specifically designed for molecular-dynamics simulations, millisecond simulations have become possible[3, 11, 12]. These simulations, and in general most of the simulations for protein-related problems, are performed using force fields. Force fields are empirical functions with fitted parameters that describe the PES of a system based on the knowledge of the nuclear positions. However, the fitting

[1]There are also proteins that fold much slower, while there are also fast folders, which obtain their native state on the order of microseconds.

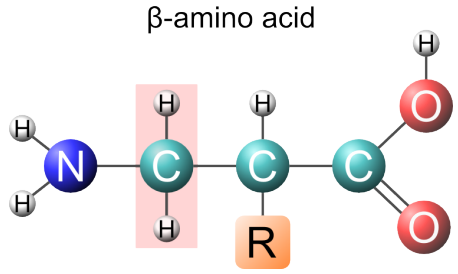


Figure 1.3: A schematic representation of a β-amino acid. The additional methylene CH_2 group in the backbone is highlighted with a light pink background.

process and the use of standard functional forms can lead to problems of limited accuracy and transferability[13, 14]. To resolve subtle energy differences such as present, e.g., in conformational ensembles (see dotted red line in Fig. 1.2c), a first-principles method based on the solution of the many-body Schrödinger equation would be desirable, treating all conformations on an equal footing. However, such approaches are computationally much more demanding.

In this work, we assess the challenging problem of predicting the structure or structural ensemble of a peptide on a *quantitative* level, employing an unprecedented first-principles screening effort. For this, we use density-functional theory (DFT) with the PBE[15] exchange-correlation functional explicitly corrected for long-range dispersion interactions[16] (PBE+vdW). Given the huge conformational space of peptides, an efficient and reliable search technique has to be developed. We obtain *global* sampling of the structure space by performing replica-exchange molecular dynamics (REMD) simulations with a force field. Then we follow up with a *local* refinement, relaxing thousands of structures with the PBE+vdW functional. In order to find the lowest-energy structures of the respective basins, we perform PBE+vdW REMD simulations. For validation purposes, we compare our structure predictions to experimental ion mobility-mass spectrometry (IM-MS) and infrared multiphoton dissociation (IRMPD) data.

Specifically, we focus on two challenging application cases:

(i) We examine a 20-residue peptide system that is big enough to partially show tertiary structure. In contrast to previous studies in our group[17], this peptide is not only much larger, but also presents a very complex landscape with no *a priori* knowledge about its structural preferences existing.

(ii) We further investigate the impact of backbone elongation of the amino-acid building blocks as represented by β-peptides.

Just as the natural α-peptides, β-peptides belong to the group of homologous peptides, which are composed of the homologous amino acids. A homologous series, in organic chemistry, consists of compounds that differ in length by one methylene (CH_2) group. As illustrated in Fig. 1.3, a β-amino acid has one additional methylene group between the amino and the carboxyl group compared to a natural (α)-amino acid [cf. Fig. 1.1(a)]. This backbone extension makes a β-peptide more flexible as it yields one additional torsional degree of freedom per residue, resulting in an even more complex conformational space. Another effect of this modification is that, compared to α-peptides, β-peptides are more stable against proteases[18–20], which are enzymes that cleave peptide bonds. This is interesting with respect to the possible use of β-peptides for pharmaceutical purposes. In fact, it has already been shown that β-peptides are

able to modulate native protein-protein interactions[18, 19, 21–24]. In this thesis, we investigate how the conformational space of a β-peptide is influenced by its increased flexibility. For this, we focus on a comparison between a natural α-peptide and its equivalent β-peptide obtained by exchanging the α-amino acids with the corresponding β-amino acids.

For both the flexible β-peptides and the large natural peptides (20 residues), we clearly face the limit of what can be achieved by first-principles electronic-structure methods today. Compared to empirical methods such as force fields, probably the most important advantage of DFT is its wider range of validity due to its quantum-mechanical foundation. Still, the exchange-correlation functional is only approximately known. We assess the exchange-correlation functional applied (PBE+vdW) and, along these lines, point out directions in which the theory can be improved.

The peptides dealt with in this thesis are based on polyalanine, where alanine (Ala) is a relatively simple amino acid with a methyl group (CH_3) as the side chain R (cf. Fig. 1.1). Additionally, the peptides contain lysine (Lys) residues, whose side chains have a protonated amino group (CH_2-CH_2-CH_2-CH_2-NH_3^+). All simulations are performed under isolated conditions. This allows for a direct benchmark of first-principles simulations for feasible system sizes against sufficiently detailed experiments under the exact same conditions.

This thesis is divided into three parts. Part I (Polypeptides and (free) energy surfaces) details the theoretical background: After giving a general introduction about (non)natural polypeptides (Chapter 2), we discuss methods to describe the PES (Chapter 3) and methods to explore it (Chapter 4). Chapter 5 deals with the computation of infrared (IR) spectra and the calculation of free energies.

In the second part, we focus on natural polyalanine-based peptides (Large polyalanine-based peptides: structure and spectroscopy). In Chapter 6, we present benchmarks for IR spectra obtained from first-principles MD simulations and in Chapter 7, we explain experimental techniques relevant to this work and how to connect our theoretical results to the experimental data. Chapters 8 and 9 focus on the conformational search and the comparison of Ac-Ala_{19}-Lys + H^+ versus Ac-Lys-Ala_{19} + H^+, where the lysine residue is located at the C- and the N-terminus, respectively. The position of the protonated lysine residue has a critical impact on the structure. While Ac-Ala_{19}-Lys + H^+ is clearly α-helical[25–28], Ac-Lys-Ala_{19} + H^+ favors a rather compact conformation. We also show how sensitive the specific conclusion is to the details of the theoretical approach employed, including the results of more advanced methods in targeted calculations.

Part III of this thesis (Dealing with conformational flexibility: homologous peptides) describes the comparison of the structure space of natural versus non-natural β-peptides with an extended backbone. More specifically, we compare Ac-Ala_6-Lys(H^+) and Ac-β^2hAla_6-Lys(H^+). While Chapter 10 focuses on the conformational search for Ac-β^2hAla_6-Lys(H^+), in Chapter 11, we compare the results for the two peptides and connect the findings to experiment in Chapter 12.

Chapter 13 gives the conclusions and an outlook.

Part I

Polypeptides and (free) energy surfaces

2 PEPTIDES AND PROTEINS

As mentioned in the introduction, proteins are biopolymers with amino acids as their monomeric unit. As the name implies, amino acids are (carboxylic) acids containing an amino group. A schematic representation of an amino acid was given in the introduction in Fig. 1.1a. At physiological pH values, the carboxyl and the amino group of an amino acid are both ionized (zwitterionic form), whereas in the gas phase they are neutral. Figure 1.1a depicts an amino acid in its non-zwitterionic form. The Cα atom has four different substituents, which makes it chiral. As a result, the same amino acid can adopt two configurations that differ in the spatial arrangement of the atoms around the chiral center Cα. These two configurations constitute mirror images (see Fig. 2.1) called enantiomers. The two enantiomers of amino acids are denoted as L- and D-amino acids by convention, where in nature almost exclusively the L-type occurs.

The amino group of one amino acid and the carboxyl group of another amino acid can formally react with each other as schematically illustrated in Fig. 2.2.[1] The amino acids form a bond between the amide nitrogen and the carbonyl carbon atom. It is called amide bond and links two amino-acid *residues*. Formally, a water molecule is released upon the reaction. The group of atoms C(=O)–N(H) is referred to as a peptide link or the peptide group. The linear polymers obtained by the linkage of amino acids via amide bonds in a head to tail fashion are called peptides. In this case, the amide bond is also referred to as peptide bond. Peptides with up to ≈ 10 amino-acid residues are known as oligopeptides. Larger peptides are typically referred to as polypeptides and polypeptides of more than 50-100 residues are denoted as proteins.

The covalently bound atom series $\cdots$–C(=O)–N(H)–Cα–$\cdots$ constitute the peptide's backbone. As the amino acids in a peptide are all linked in the same way, a peptide has two defined termini. One terminus comprises an amino group and is referred to as the N-terminus, while the other one involves a carbonyl group and is consequently denoted as the C-terminus.

In the cells of living organisms, the synthesis of proteins takes place at the ribosomes.[2] We

[1] In practice, the formation of a peptide bond is much more complicated and needs to be catalysed. We will get back to this question later in this chapter.

[2] There also exists non-ribosomal peptide synthesis, which is catalysed by special enzymes called synthetases. The peptides produced in this way are typically very short (up to 50 residues).

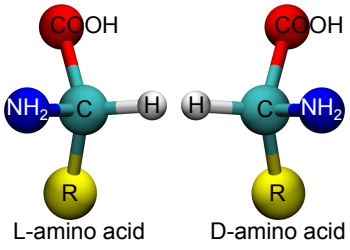


Figure 2.1: A schematic illustration of the two enantiomers for amino acids: L- and D-configuration.

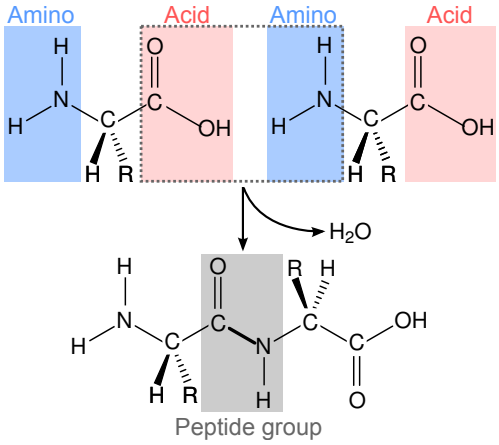


Figure 2.2: Schematic illustration of the formal reaction of two amino acids forming a peptide bond. Upon this reaction a water molecule is released.

will only give a brief account here. More details can be found in virtually any biochemistry textbook, e.g., in the book by Voet and Voet[29]. According to the *central dogma of molecular biology*[30][3] the flow of information corresponds to DNA → RNA → protein. In the first step, a process known as transcription, the nucleotide sequence of the DNA strand is copied to an mRNA molecule. The mRNA is transported to the ribosomes, which catalyze the synthesis of the proteins. The genetic code is a dictionary that relates the nucleotide sequence of the DNA/mRNA to the polypeptide amino-acid sequence. It has a triplet character with three nucleotides forming a codon that specifies one particular amino acid. Although there have been many amino acids found in living organisms (more than 700[29]), proteins are composed of only 20 amino acids referred to as standard amino acids.[4,5] They are depicted in Fig. 2.3 together with their three- and one-letter code, by which they are frequently referred to. Their side chains have different properties, which play an important role for protein function. The amino acids depicted in the first row of Fig. 2.3 have a charged side chain at physiological pH values: arginine, histidine, and lysine are positively charged, while aspartic acid and glutamic acid are negatively charged. The second row of Fig. 2.3 illustrates amino acids whose side chain is polar, but not charged at physiological pH values. The so-called special cases are shown in the third row. Glycine's side chain is a hydrogen atom, which makes it the only non-chiral amino acid. The side chain of proline is cyclic and involves the imine nitrogen atom, restricting its conformational freedom. The side chain of cysteine contains a thiol group. The thiol groups of two cysteines can form disulfide bonds, which have an important impact on protein structure: they can link two individual peptide chains or create a cross-link within the same chain[29]. The fourth row of Fig. 2.3 illustrates amino acids with a hydrophobic side chain.

The flexibility of peptides originates mostly from rotations around single bonds, involving changes of dihedral angles. A dihedral or torsional angle involves 4 atoms, A–B–C–D, and is defined as the angle between the two planes spanned by the atoms A, B, C and B, C, D, respectively, as depicted in Fig. 2.4. When looking along the rotational axis, the dihedral angle is

[3]This term was coined by Francis Crick, who was awarded the Nobel Prize in medicine in 1962 for his model of the DNA structure (together with James Watson and Maurice Wilkins). He used the word dogma due to a misconception. Later he said that he rather meant a hypothesis than a (religious) doctrine that is unquestionable.

[4]The genetic code ciphers the 20 standard amino acids. There are two additional aminoacids occuring in proteins of eukaryotes, selenocysteine and pyrrolysine, which are coded by different mechanisms. After the translation, proteins often undergo posttranslational modifications. These include attaching other molecules (sugars, lipids, ...) or modifications of the amino acids.

[5]In the laboratory, peptides can be synthesized by a technique known as solid-phase peptide synthesis (SPPS). A detailed account of this method can be found in standard textbooks, e.g., Ref. [31].

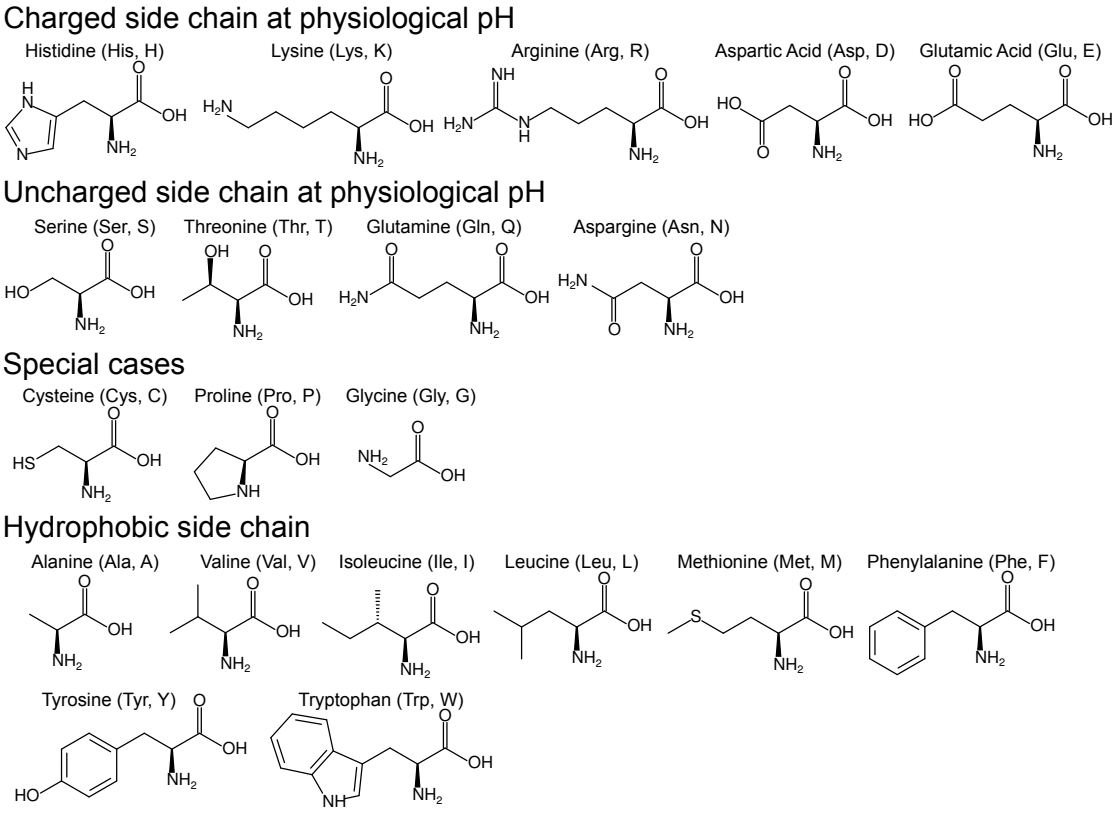


Figure 2.3: Summary of the 20 standard proteinogenic amino acids. Carbon atoms (and their attached hydrogen atoms) are not shown explicitly, but represented by kinks.

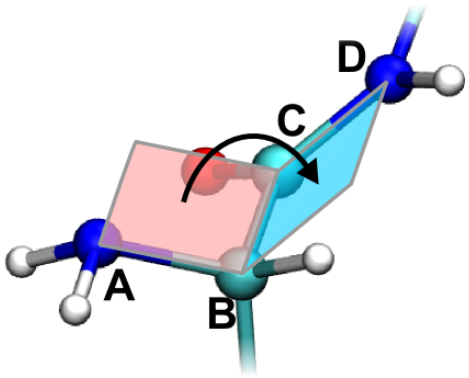


Figure 2.4: Dihedral angle between the two planes spanned by atoms A, B, C, and B, C, D.

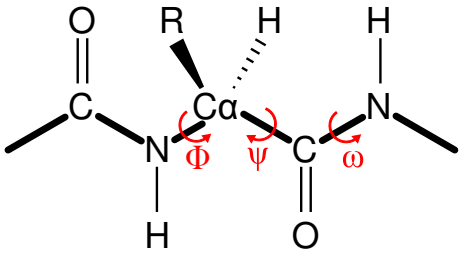


Figure 2.5: Backbone dihedral angles for natural α-peptides.

positive (negative) by standard convention if the end atom D in the back lies in a clockwise (anti-clockwise) direction from the front reference atom A. The dihedral angles of a peptide's backbone, termed ϕ, ψ, and ω, are illustrated in Fig. 2.5. The torsional angle ω denotes the rotation about the peptide bond. However, the partial double-bond character of the peptide bond prohibits free rotation about the bond axis and renders the peptide group rigid and planar. Thus, ω has only two distinct values: *cis* (0°) and *trans* (180°). Due to steric hindrance, the *cis* conformation occurs only very rarely and almost all peptide residues assume the *trans* conformation[32]. The only exception is proline, where about 5% of the residues take the *cis* conformation[32, 33]. This is owed to the cyclic nature of its side chain: in both *cis* and *trans* conformations, the Cβ atom of the preceding side chain encounters a carbon atom of the proline residue (either the Cα atom in the *cis* conformation or the Cδ atom of the ring in the *trans* conformation) yielding a smaller energy difference between the two states than for other residues[29].

Due to the restrictions for ω, the structure of peptides is most importantly characterized by the dihedral angles ϕ and ψ. However, due to steric clashes not all (ϕ/ψ) angle pairs are possible. This was first analytically determined in the seminal work by Ramachandran in 1963. In his original paper[34], he varied the ϕ and ψ angles of dipeptides searching for steric interferences between all atoms. For this, he considered the atoms as hard spheres and a steric clash was said to occur when two atoms that are not covalently bound come closer than the sum of their van der Waals radii. Obviously, this depends on the choice of the van der Waals radii. Ramachandran used two datasets in his original work: normally allowed and outer limit distances. From his results he could then define normally allowed and outer limit regions separated from prohibited regions in a plot of the (ϕ/ψ) space that became known as the Ramachandran plot or Ramachandran diagram. These regions coincide remarkably well with the ψ and ϕ data angle pairs for known peptides[34]. The Ramachandran plot has been revisited and refined in various efforts, e.g., in Ref. [35–37]. Figure 2.6 A) shows a Ramachandran plot based on experimental data from Ref. [36] and Fig. 2.6 B) illustrates a Ramachandran diagram for > 4000 conformations of a polyalanine-based peptide relaxed based on density-functional theory (DFT) using the PBE+vdW functional (see Chapter 3) in this work. The white regions denote forbidden regions in the (ϕ/ψ)-space. The most important backbone conformations of proteins, helices and β-sheets, are highlighted. They will be explained in more detail in Section 2.3.

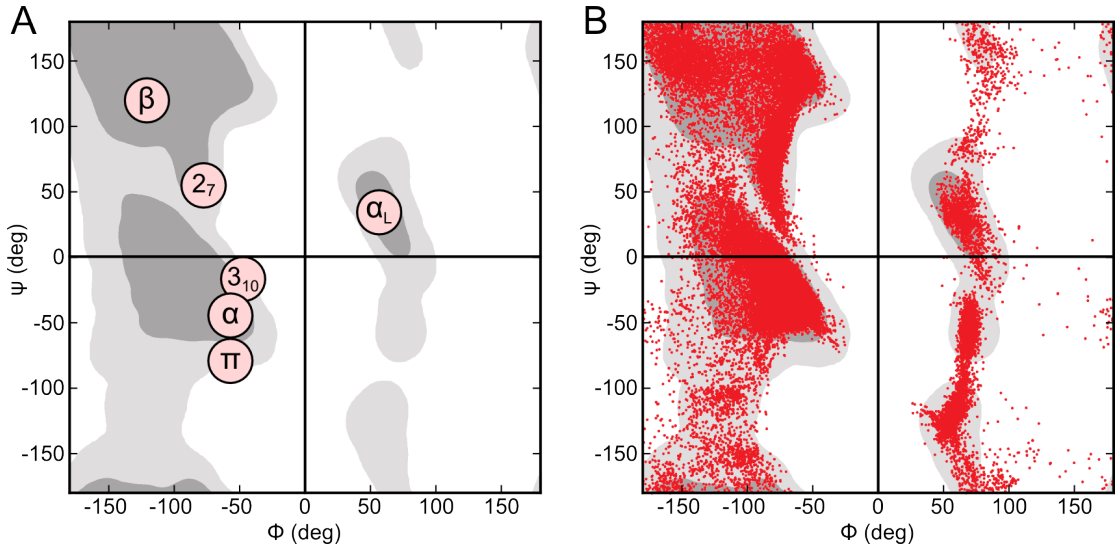


Figure 2.6: A) Ramachandran contour plot based on experimental data for 500 proteins provided with Ref. [36]. The graph was generated using the scripts by Peter Cock[38]. The inner contour defines the "favored" region (dark grey), which includes 98% of the data and the outer contour specifies the "allowed" region (light grey), which contains 99.95% of the data. The position of typical secondary-structure backbone conformations (see section 2.3) of proteins, namely helices and β-sheets are given. B) Ramachandran plot for alanine-based polypeptide conformations relaxed with density-functional theory (PBE+vdW, see Chapter 3) overlaid on the contour plot shown in A. The data comprises more than 4000 structures of the 20-residue peptide Ac-Lys-Ala_{19} + H^+, where each red dot represents one dihedral angle pair.

The Ramachandran plots for different amino-acid residues differ only slightly. The diagrams for residues with a side chain that is branched at Cβ generally exhibit smaller allowed regions than the diagrams for residues that are unbranched at Cβ. Due to its cyclic side chain, proline is the most conformationally restricted amino-acid residue, while glycine, which exhibits only a hydrogen atom as side chain, has the largest conformational freedom with the largest allowed regions in the Ramachandran plot. Furthermore, the Ramachandran plot of glycine is symmetric in (ϕ/ψ) space.

2.1 Interactions shaping the structure of polypeptides

The amino-acid sequence describes the covalent topology of a polypeptide. For the actual three-dimensional structure, however, also non-covalent interactions play a crucial role. Apart from steric hindrance and *intermolecular* interactions, such as between peptide and solvent, there are many *intramolecular* interactions that play an important role for the formation and stabilization of peptide structure. Among these are (intramolecular) electrostatic interactions, which are ubiquitous. Charged residues, e.g., can form ion pairs (or salt bridges). Dipole-dipole interactions are important as well, since many constituents of a polypeptide exhibit a permanent dipole, most importantly the peptide group (approximately 3.5 Debye[39]). In helices, e.g., the C(=O)–N(H) groups are aligned with their dipole moments summing up to a significant macro dipole moment. Furthermore, permanent dipole moments induce dipole moments in other molecules or atoms leading to an attractive interaction.

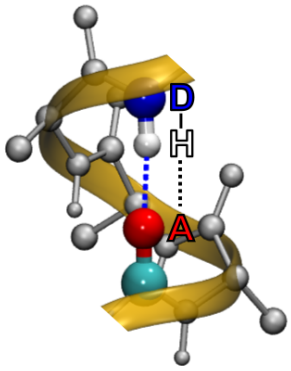


Figure 2.7: Hydrogen bond between a carbonyl group and an NH group in the backbone of an α-helical peptide.

Due to quantum effects, atoms have an instantaneous fluctuating electron density. The associated dipole (and higher) moment(s) polarizes neighboring atoms yielding an attractive interaction known as London dispersion forces[40] or van der Waals interactions.[6] The interaction is very subtle. However, due to the large number of interatomic contacts, they play a crucial role for structure formation and stability. Sometimes, especially in the chemistry community, the term van der Waals interactions is also used to refer to interactions between permanent dipoles as well as permanent and induced dipoles. However, in this thesis, we use the definition that is common in physics, in which van der Waals interactions exclusively correspond to the London dispersion forces.

Another class of important interactions are hydrogen bonds or H-bonds. Hydrogen bonds are formed between a hydrogen atom, which is covalently bound to a donor D, and an acceptor A, which has a lone-pair electron cloud. They are assigned a direction that points from the donor to the acceptor. The donor group D-H is weakly acidic, while the acceptor A is weakly basic. In polypeptides, most importantly nitrogen, oxygen and sometimes sulfur atoms act as donors and as well as acceptors in hydrogen bonds. A hydrogen bond is represented as D–H$\cdots$A. An example of a hydrogen bond between a carbonyl group and an NH group in the backbone of an α-helical peptide is illustrated in Fig. 2.7. A hydrogen bond is more directional than a purely electrostatic interaction but less than a pure covalent bond. Hydrogen bonds (D–H$\cdots$ A) are often linear, with the donor group D–H oriented in the direction of the lone-pair electron orbital of the acceptor, but deviations from this ideal are common. In this thesis, unless stated otherwise, a hydrogen bond is considered to be present if the distance between a hydrogen atom and an acceptor is less than 2.5 Å.

Hydrogen bonds show significant cooperativity phenomena[41–44]. For example, it was found that hydrogen bonds in an infinite α-helical chain are strengthened by a factor of two compared to an isolated hydrogen bond[41].

2.2 Structure hierarchy

The structure of proteins can be classified into four levels. This nomenclature was introduced by Linderstrøm-Lang in 1951[45].

- The **primary structure** is the sequence and number of amino acids of a polypeptide, i.e., the covalent scaffold of the protein.

[6] Named after the Dutch physicist Johannes Diderik van der Waals.

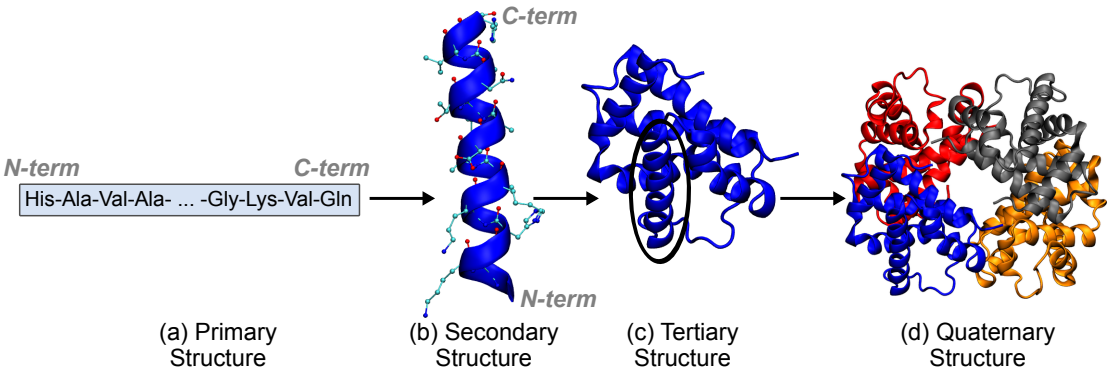


Figure 2.8: Different levels of protein structure formation: primary structure, secondary structure, tertiary structure, and quaternary structure. The example illustrated in this picture is human hemoglobin taken from the protein database (PDB ID: 1GZX). The primary and secondary-structure illustration comprise residues 54 to 72. The ribbon runs along the polypeptide backbone as a guide to the eye, with a thicker representation for helical segments. For clarity, the tertiary and quaternary structure are illustrated without atoms in a cartoon representation. The blue subunit is chain A. Each subunit/chain is colored differently to show the arrangement of the subunits in the quarternary structure.

- The **secondary structure** denotes the local three-dimensional backbone conformation of the polypeptide, not considering the conformation of the side chains. The three main secondary-structure elements are helices, pleated sheets, and turns. They are the building blocks of the tertiary structure and will be discussed in more detail in Section 2.3. Helices and pleated sheets are usually linked by turns, or by segments of the backbone that are less easy to describe (although not necessarily less structured), often referred to as loops.
- The **tertiary structure** is the overall three-dimensional structure of a single polypeptide chain. It arises through the association of the secondary-structure building blocks along with the spatial arrangement of the side chains.
- Different separate polypeptide chains can associate together via non-covalent interactions or disulfide bonds in a defined spatial arrangement. This is referred to as the **quaternary structure**, where the individual polypeptide chains are called subunits.

Figure 2.8 illustrates the four different structure levels of proteins, based on the example of human hemoglobin. The sequence of residues 54 to 72 (primary structure, Fig. 2.8a) forms a helix (secondary structure, Fig. 2.8b). This helix is one of the secondary-structure building blocks that makes up the three-dimensional structure (tertiary structure, Fig. 2.8c) of chain A, one of the four subunits of hemoglobin. The assembly of the four individual subunits is referred to as the quaternary structure (Fig. 2.8d).

2.3 Secondary structure

The term *secondary structure* refers to the local three-dimensional conformation of the backbone of a polypeptide without considering the spatial arrangement of the side chains. There are three main elements of secondary structure, namely helices, sheets, and turns. They are called secondary-structure building blocks as they constitute the building blocks for the three-dimensional overall shape of the polypeptide (tertiary structure).

Helices are periodic structures, which are stabilized by (periodic) hydrogen bonds between the carbonyl and NH groups in the backbone of the peptide. Examples of different helices are depicted in Fig. 2.9. We note that in this thesis, we follow the CPK coloring convention named after Robert Corey, Linus Pauling, and Walter Koltun[46]. Red spheres denote oxygen atoms, white spheres are used for hydrogen and blue denotes nitrogen. The color used for carbon differs in different visualization programs. In this thesis, the VMD program[47] is used for rendering in most cases, which uses cyan for carbon by default.

The most prominent helix type, the α-helix, was predicted by Pauling, Corey, and Branson in 1951[48, 49]. They discovered it by systematically searching for all possible hydrogen-bonding patterns of a single polypeptide chain based on four assumptions: (a) all residues are equivalent (without regard to the side chains), (b) the planarity of the peptide group, (c) a distance of 2.72 Å between the acceptor (oxygen) and the donor (nitrogen) of the hydrogen bond, and (d) a deviation from the linear arrangement of D-H$\cdots$A with $\sphericalangle$(N,H,O)=180° by less than 30°. The α-helix is characterized by a hydrogen-bonding pattern with hydrogen bonds periodically formed between every i and $i+4$ residue. The approximate dihedral angles are $\phi = -57°$ and $\psi = -47°$. An alternative way to denote helices is by giving the number of atoms involved in the hydrogen-bonded pseudocycle as a subscript together with the number of residues comprised in one turn. In the α-helix every hydrogen-bonded pseudocycle involves 13 atoms. One turn of the helix comprises 3.6 residues with a rise along the helix axis per turn (*pitch*) of 5.4 Å. Thus, the α-helix is also denoted as a 3.6_{13}-helix. A helix is chiral, i.e., it can be either left or right handed.[7] For L-amino-acid residues the α-helix is right-handed due to sterical hindrances of the side chains in the corresponding left-handed helix. In turn D-amino-acid residues form a left-handed helix. The chirality of the helix thus depends on the chirality of the amino-acid residues.

Apart from the α-helix, Fig. 2.9 also shows other helix types and a scheme which illustrates the corresponding hydrogen bond patterns. In the 2.2_7-helix each residue i forms a hydrogen bond with residue $i+2$. One turn is comprised of 2.2 residues and the hydrogen-bonded pseudocycles contain 7 atoms. The 3_{10}-helix comprises 3 residues per turn and 10 atoms in the pseudocycles formed by the hydrogen bonds between every i and $i+3$ residue. It has a pitch of 6.0 Å and a smaller helix diameter than the α-helix. The π-helix, on the other hand, has a larger helix diameter than the α-helix. Here, hydrogen bonds are periodically formed between the i and the $i+5$ residue. While the α-helix makes up about 31% of the secondary structure of proteins[29], the 3_{10}-helix is only occasionally found, mostly at the termini of α-helices. The π-helix occurs very rarely and the 2.2_7-helix has never been observed. As illustrated in the scheme in Fig. 2.9, in all helix types, the direction of the hydrogen bonds (defined to point from the donor to the acceptor) points from the C- to the N-terminus of the peptide. This leads to a helix dipole pointing in the same direction – due to the alignment of the hydrogen-bonded C(=O)–N(H) groups along the helical axis, the individual dipole moments sum up to an overall dipole moment of the helix, which points from the C- to the N-terminus (see Fig. 2.9). For steric reasons, helical structures with periodic hydrogen bonds pointing in the opposite direction, i.e., from the N- to the C-terminus do not occur. However, as we shall see in the next section, for peptides with an artificially extended backbone (homologous peptides) such types of helices have been observed.

[7] The helix is called right (left) handed, if the helix spiral follows the direction of the remaining fingers when the right (left) thumb is pointing along the helical axis.

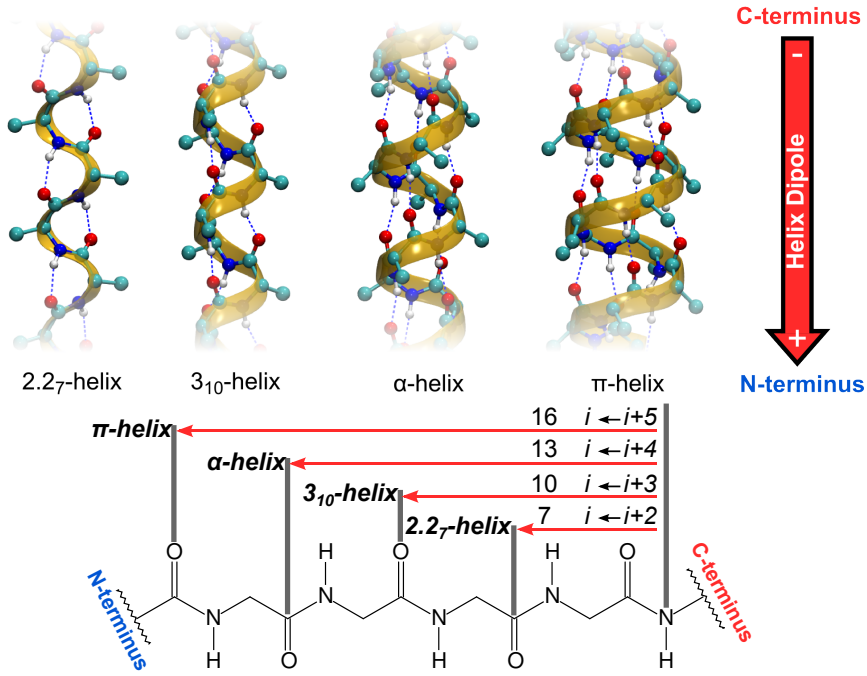


Figure 2.9: A schematic example of different helix types: 2.2_7-, 3_{10}-, α- and π-helix. The backbone of the polypeptide chain is highlighted by a yellow ribbon. Hydrogen bonds are indicated by dashed blue lines and hydrogen atoms that are attached to carbon atoms are omitted for clarity. The lower panel shows a scheme of the corresponding hydrogen-bonding patterns. Each kink represents a carbon atom (with its attached hydrogen atoms).

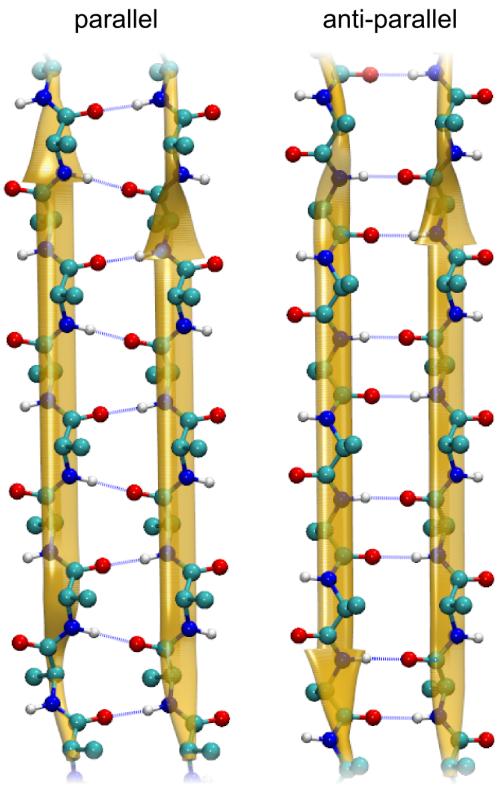


Figure 2.10: Schematic example of a parallel and an anti-parallel β pleated sheet. Hydrogen bonds are indicated by dashed lines and hydrogen atoms that are attached to carbon atoms are omitted for clarity. The arrows depict the direction in which the strand runs, i.e., from the N- to the C-terminus.

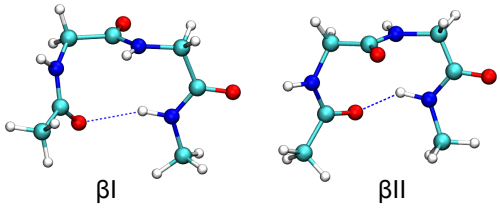


Figure 2.11: Illustration of two β turns: βI and βII.

The other basic secondary-structure type, the β pleated sheet, was likewise discovered by Pauling and Corey in 1951[49, 50]. In contrast to helices, it involves hydrogen bonds between separate individual polypeptide chains. There are two types of β pleated sheets: (a) the parallel β sheet, where the two strands that are hydrogen bonded extend in the same direction, and (b) the antiparallel β sheet, where the two strands that are hydrogen bonded extend in opposite directions. Schematic examples of both types of β sheets are illustrated in Fig. 2.10. In the antiparallel sheet, the hydrogen bonds are perpendicular to the chain direction, while for the parallel β sheet they are diagonal. For both types, the side chains of the residues on adjacent chains extend in the same direction, pointing alternately along opposite sides of the strand. The dihedral angles of the individual residues of the strands are not 180°, but (in an ideal sheet) rather $\phi = -119°$ and $\psi = 113°$ for the parallel sheet and $\phi = -139°$ and $\psi = 135°$ for the antiparallel sheet[29]. Hence, when viewed from the side, the sheets are not flat, but look rather pleated, which is where the name "pleated sheet" originates from. The number of strands found in sheets ranges between 2 to 22 with an average of 6[29].

The third basic class of secondary-structure elements are turns. They are non-repetitive and reverse the direction of a polypeptide chain. Turns come in different flavors. The most important type of turns is the β-turn, which involves 4 consecutive residues i, $i+1$, $i+2$, and $i+3$. The original classification of β-turns goes back to Venkatachalam[51]. He categorized them according to the dihedral angles of residues $i+1$ and $i+2$. Based on solely theoretical considerations, he determined three general classes. They split into six categories βI, $\beta I'$, βII, $\beta II'$, βIII, and $\beta III'$[51], with the categories denoted with a prime being the corresponding mirror images of the backbone conformation. In β-turns, the Cα atoms of residue i and $i+3$ are in close contact (< 7 Å[52–54]), often coming along with a hydrogen bond formed between the C(=O) of the i and the N(H) of the $i+3$ residue. Other turn types exhibit analogous features: α turns usually exhibit a hydrogen bond between the i and $i+4$ residue and π turns between the i and $i+5$ residue. However, the H-bond in turns is often disrupted and its existence is not necessary for the segment to be characterized as a turn. The most widely occurring motifs are the βI- and βII-turn, which differ by a flip of 180° of the central peptide group. They are schematically illustrated in Fig. 2.11. In proteins β-turns are often found in antiparallel β pleated sheets reversing the direction of the strand. This constellation is called β-hairpin.

2.4 Peptidic foldamers

Despite the versatility of sequence, structure, and function that proteins feature, biology spans only a small part of the chemical space. It is a longstanding idea to increase the biological toolbox by synthetic polymers with unique functions. Since the mid-1990s, non-natural polymers that obtain a compact fold (foldamers[55, 56]) have more and more entered the scientific spotlight. This was initiated from materials science, especially the interest in nylon and nylon derivatives [57–64]. Nylon (derivatives) and peptides are chemically related as they both consist of monomers that are linked by amide bonds. Nylon-2, e.g., corresponds to a polyglycine chain. A schematic representation of the chemical formula of nylon-(m+1) is given in Fig. 2.12. If one (or more) of the carbon atoms is substituted with a side chain, one speaks of a nylon derivative.

The pivotal trigger for the promotion of foldamers into an active field of research, however, were the findings from Seebach's[65, 66] and Gellman's groups [55, 67], who showed that

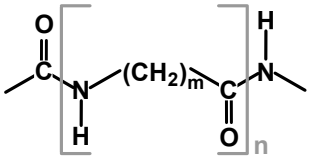


Figure 2.12: Schematic representation of the chemical formula of nylon-(m+1). Nylon is a polymer, where the monomer units are linked by amide bonds. It is named according to the number of carbon atoms in the monomer unit, namely m+1. Nylon-2, e.g., thus corresponds to a polyglycine chain.

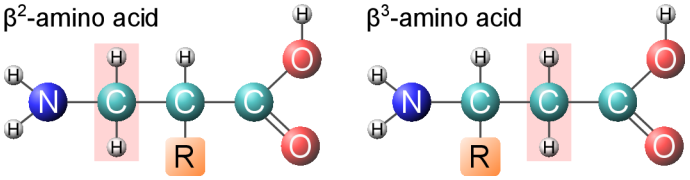


Figure 2.13: A schematic representation of a β^2- and β^3-amino acid with side chain R in its non-zwitterionic form. The additional methylene group is highlighted in pink.

oligomers composed of non-natural amino acids can adopt defined helical structures. The term foldamer was coined by Gellman[55, 56] and describes *"any polymer with a strong tendency to adopt a specific compact conformation"*[56]. With respect to proteins, the term "compact" rather refers to the three-dimensional (tertiary) structure. However, the building blocks of the tertiary structure are the secondary-structure elements. Thus, the first step in constructing a foldamer has to be to identify synthetic oligomers that have a conformational preference similar to a regular secondary-structure element (helices, turns, or sheets)[56]. The field of foldamer research rapidly branched out (see Ref. [18, 68–78] and many references therein). This work, however, specifically deals with peptidic foldamers, i.e., foldamers with monomeric units that can formally be derived from natural amino acids. There are various possibilities to derive new monomers from a natural α-amino acid such as exchanging atom types and altering or shifting the side chain. Alternatively, one could imagine a building block with an extended backbone. Insertion of methylene (CH_2) groups between the amino group and the carboxylic acid results in the class of *homologous* amino acids since, in organic chemistry, a series whose members differ in length by one CH_2 group is known as a homologous series. Natural amino acids occurring in native proteins are the first members of such a homologous series. They have an amino group attached to the α-carbon (Cα) and are thus denoted as α-amino acids. Amino acids exhibiting one additional CH_2 group have the amino group linked to the Cβ atom and, hence, are referred to as β-amino acids. Analogously, γ- and δ-amino acids have 2 and 3 additional methylene groups, and so on. They are commonly referred to as ω-amino acids[74]. Correspondingly, oligomers composed of these amino acids are referred to as ω-peptides. In this thesis, the focus is primarily on β-peptides. As depicted in Fig. 2.13, there are two different substitution patterns for homologous β-amino acids. The side chain can be either substituted at the second or at the third carbon position, leading to a β^2- and β^3-amino acid, respectively. For example, a β^2-amino acid derived from alanine (Ala) by backbone homologation is denoted as β^2hAla following the nomenclature in the literature[69, 79], where "h" stands for "homo".

Of the homologous series, β-amino acids are the closest relatives to the natural α-amino acids and have become the figurehead of foldamer research[77]. The additional methylene group in the backbone yields one additional torsional degree of freedom per residue compared to the natural α-amino acids making the β-peptide's backbone more flexible. The backbone dihedral angles of β-peptides are illustrated in Fig. 2.14 with the additional torsional angle denoted as θ (see Fig. 2.5 for a comparsion with α-peptides). Various secondary structure motifs have been found in β-peptides and also hybrid α/β-peptides, covered in many reviews[18, 19, 69, 70, 72, 73, 77, 79–

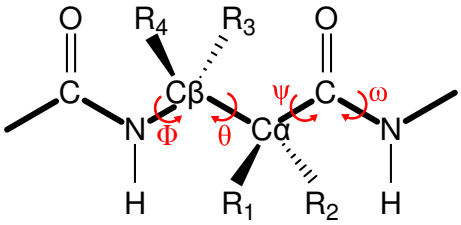


Figure 2.14: Backbone dihedral angles for β-peptides.

81] (see also many references in Ref. [74]). Besides turn and sheet motifs[18, 79], different helical patterns have been experimentally observed[69, 77, 79] in solution (mostly in MeOH, but also in H_2O[82, 83]) and in the solid state based on X-ray crystallography, nuclear magnetic resonance (NMR), and circular dichroism (CD) spectroscopy. For small isolated (gas-phase) α/β-peptides and β-peptides, combined ultra violet-infrared (UV-IR) techniques have also been used to elucidate the structure and identify single conformers[84–88]. This method will be further described in Section 7.1.2.

Along with experimental studies many computational studies on β-peptides have been performed, as reviewed in Ref. [74]. The group of van Gunsteren has conducted numerous force-field based studies (see, e.g., Ref. [78, 89–100]) often in close collaboration with the Seebach group[79, 89, 95–97, 99]. One of the first first-principles based studies was published in 1996 by Alemán and co-workers using Hartree-Fock calculations (see Chapter 3) to investigate β-aspartate[63], which is a nylon-3 derivative. After this, many conformational studies were conducted based on Hartree-Fock, MP2, and DFT calculations for various types of β-peptides[68, 101–104], also specifically investigating helices with H-bonds alternatingly pointing in opposite directions (mixed helices)[68, 103, 105, 106]. In the following, we shall describe in more detail which kinds of helices in β-peptides have been experimentally found or theoretically predicted.

Figure 2.15 illustrates schematic examples of different β-peptidic helix types and their corresponding hydrogen-bonding patterns. The helices are denoted by the number of atoms in the hydrogen-bonded pseudocycles. As mentioned earlier, by convention, a hydrogen bond D-H$\cdots$A has a direction, which points from the donor to the acceptor (D$\rightarrow$A). The upper part of the scheme in Fig. 2.15 shows helical patterns where the H-bonds point in opposite sequence direction. In the corresponding helices this leads to a helix dipole pointing in opposite sequence direction as well. This is analogous to the helices found in natural α-peptides (described in Section 2.3). Hydrogen bonds between the i and the $i+2$ residue result in 8-membered pseudocycles (H8-helix). A H12-helix has periodic H-bonds between the i and the $i+3$ residue, a H16-helix between the i and the $i+4$ residue, and a H20-helix between the i and the $i+5$ residue.

In β-polypeptides, helices with hydrogen bonds pointing in the other direction, namely along the strand direction, have also been found[69, 77, 79]. This yields a helix dipole, which points from the N- to the C-terminus. Such helices are illustrated in the lower part of the scheme in Fig. 2.15. Hydrogen bonds between the i and the $i+1$ residue lead to 10-membered pseudocycles (H10-helix). The H14-helix has H-bonds between the i and $i+2$ residue, and the H18-helix has H-bonds between the i and $i+3$ residue.

The most studied helical structure found in β-peptides is the H14-helix[69, 77, 79]. It has been observed in the solid state by X-ray crystallography, but also in solution by NMR and CD spectroscopy, mostly in methanol (MeOH), but also in water[82, 83]. Seebach and co-workers

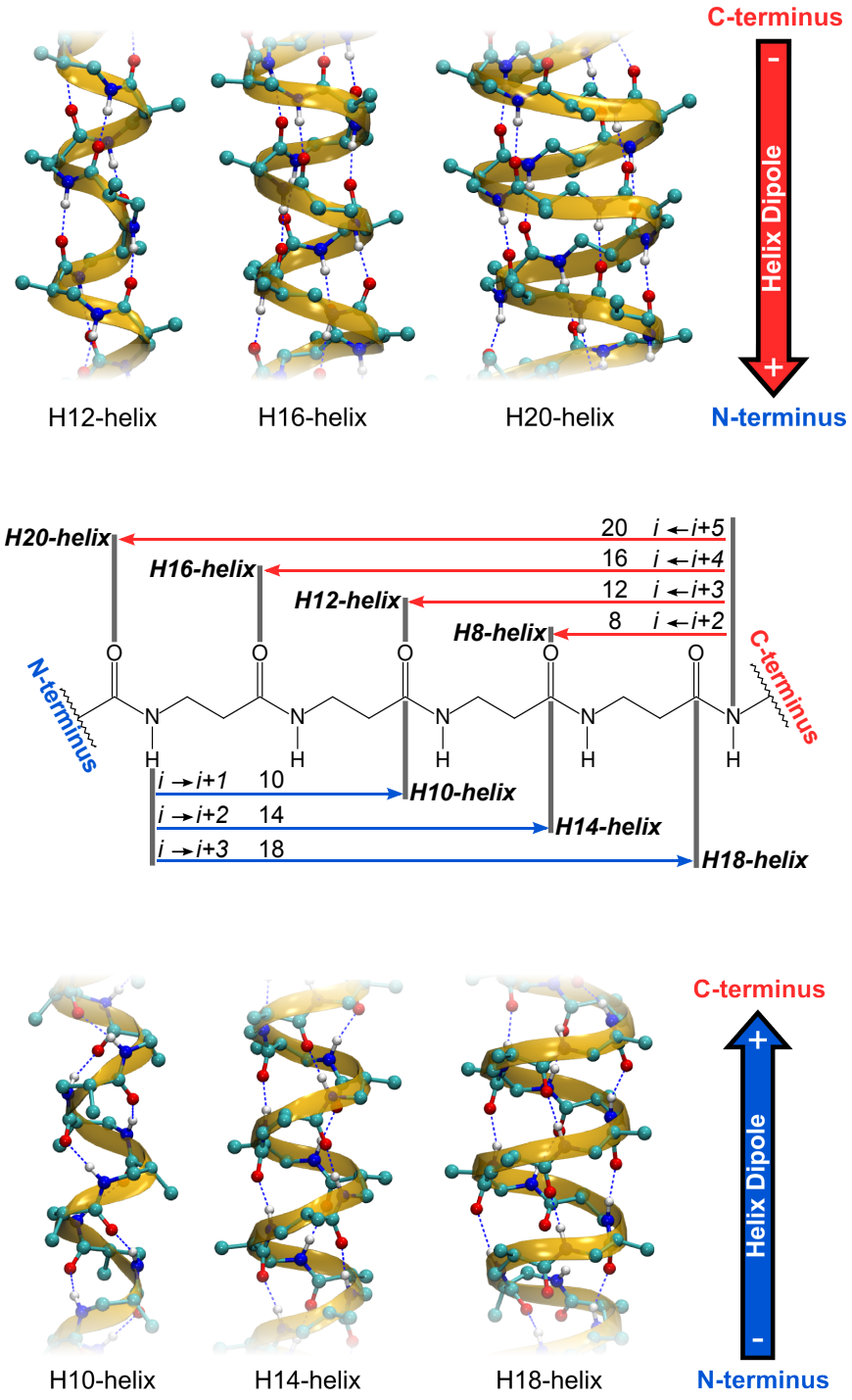


Figure 2.15: Examples of different helix types for β-peptides. Hydrogen atoms that are attached to carbon atoms are omitted for clarity. The middle panel shows a scheme of the corresponding hydrogen-bonding patterns. Each kink represents a carbon atom (with its attached hydrogen atoms).

even found a H14-helix for an icosapeptide comprising all 20 different proteinogenic amino-acid side chains[107]. A helix with 12-membered hydrogen-bonded pseudocycles (see Fig. 2.15), the H12-helix, was discovered by Gellman and co-workers[67] for a conformationally restricted β-peptide, where both the Cα and the Cβ are involved in a five-membered ring. Similar conformationally restricted β-peptides have been shown to form H10-helices[108] and H8-helices[109]. Recently, Fülöp and coworkers experimentally observed a H18-helix and confirmed it by Hartree-Fock calculations[110]. The H20-helix has not been found, yet. For the H16-helix there are hints from fibre diffraction studies of a nylon derivative[62, 111] and a Hartree-Fock study[74]. In sequences with alternating β^2- and β^3-units, Seebach and co-workers[112, 113] found a mixed helix, comprised of alternating 10- and 12-membered hydrogen-bonded rings with the hydrogen bonds pointing in alternating directions. Such mixed helices are also referred to as β-helices due to the similarities of their hydrogen-bonding pattern with a β-sheet. A well-known example of a β-helix occuring in α-peptides is Gramicidin A[114–116] with alternating 20- and 22-atom membered hydrogen-bonded pseudocycles.[8]

As illustrated in the introduction, proteins, and also peptides, play a key role in virtually all biochemical processes in the body. They are able to adopt very specific and diverse tasks, making them extremely interesting for medical application. However, the use of natural peptides as drugs suffers from several problems[117]:

1. their instability against proteases, which are enzymes that cleave peptide bonds,
2. their poor oral bioavailability, i.e., only a small fraction of the actual dose is eventually absorbed after oral ingestion (e.g., due to too large molecular masses or/and the lack of appropriate transport mechanisms),
3. their short excretion life times through kidney and liver, and
4. their interaction with multiple receptors due to their conformational flexibility (and not only the ones that would be intended).

On the contrary to natural peptides, non-natural peptides seem to be promising candidates for drug design. It could be shown that the stability against proteases is increased for homologous peptides (see Refs. [18–20] and references therein). There are hints that small β-peptides are orally bioavailable and have excretion times that are larger than for natural peptides[18]. Furthermore, several studies showed that β-peptides and heterogeneous α/β-peptides can be used to modulate native protein-protein interactions[18, 19, 21–24, 118]. Based on a combined experimental and theoretical study (force fields), Michel *et al.*[23] have identified a β-peptide that has the potential to prevent the inhibitation of a cancer suppressor protein (p53) by oncoproteins.

2.5 Energy landscapes

As mentioned in the introduction, in their pioneering experiments in the early 1960s, Christian Anfinsen and co-workers found that the folding of proteins is reversible[4]. Proteins can fold into their native structure, denature and then re-fold again to the same state[4, 6]. From these findings, Anfinsen inferred that the native state of a protein is coded in its amino-acid sequence

[8] This peptide consists of alternating L- and D-residues. It is not synthesized at the ribosomes, but by non-ribosomal peptide synthesis.

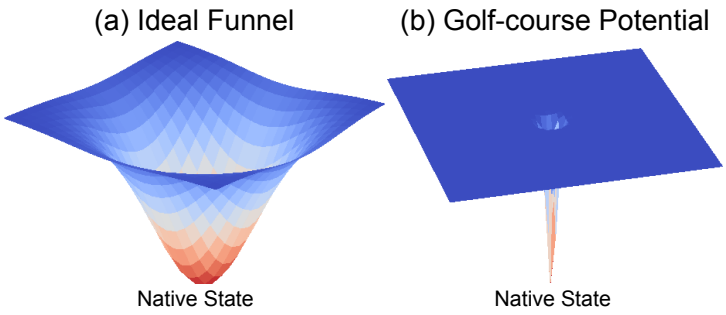


Figure 2.16: Schematic representation of different energy landscapes: (a) an idealized funnel and (b) Levinthal's golf-course potential.

in a given environment (pH value, temperature, solvent, ...)[4–6]. He furthermore deduced that the native state has to be the global minimum of the system's free energy. It is unique, stable (against small changes of the environment), and accessible on biological time scales[5]. This statement became known as the thermodynamic hypothesis.

Cyrus Levinthal[119, 120] argued that if a protein randomly searches its possible conformations for the native state, it will never find it, just like finding a needle in the haystack[7]. This can be put into a mathematical framework (see, e.g., Ref.[121]): If there is a protein sequence comprised of 101 amino acids with 3 microstates per peptide linkage, this would lead to 3^{100} possible conformations. Even if the protein was able to sample the conformations very fast, say at a speed of 10^{13} per second, it would still take the protein 10^{27} years to sample all of them, i.e., longer than the age of the universe. However, proteins do fold on biological time scales – otherwise life would be impossible. This became known as Levinthal's paradox[121, 122]. As a solution to this paradox, Levinthal suggested the existence of folding pathways, which are defined as a unique sequence of steps that take the protein from its denatured state to its folded state[120]. Various efforts have been undertaken to identify folding pathways. These efforts are reviewed, e.g., in Ref.[7]. However, Levinthal's paradox is inherently flawed as it assumes all conformations to be searched with the same probability, i.e., randomly. If an energy bias for the conformations is introduced, it can be shown that folding times reduce to biological time scales[121]. In the mid-nineties a new view[123, 124] of protein folding emerged, moving away from folding pathways to an energy-landscape perspective[7, 125–127]. This was decisively triggered by the Letter to Nature by Karplus[9] and co-workers[128] who were the first to give a detailed account on the free-energy surface of a protein. They used a 27-bead polymer model and performed Monte Carlo simulations to fold the protein. Their results showed that not all conformations were sampled upon folding. This means that folding times (measured in Monte Carlo steps) are smaller than the times estimated by Levinthal. Furthermore, there are various folding pathways that take the denatured protein to the native state instead of one unique pathway[123, 128]. This resolves Levinthal's paradox – Dill and Chan describe it with the picture of skiers on a mountainside: although all of them start from different points on the mountain, they will all eventually reach the same valley[7](see Fig. 2.16a). Folding is not guided by one single unique pathway, but there exists a folding funnel, exhibiting multiple parallel pathways. In contrast, Levinthal's picture can be compared to a golf-course potential[7](see Fig. 2.16b), a flat surface (all conformations having the same energy) with only one deep dip (the native state). On such a landscape, the golf ball would idle around endlessly before finding the hole. However, the landscape is not flat and thus, the search of the protein is not random, but directed to the

[9]Martin Karplus was awarded the Nobel Prize for chemistry in 2013.

native state by the change in energy upon changing the atomic coordinates. Energy-landscape theory has nowadays developed into a general theory focusing not only on proteins, but rather unifying different research fields such as proteins, glasses, and molecular clusters[129]. Many details that go beyond the scope of this summary here are covered in the textbook by David Wales[130].

The potential-energy surface (PES) denotes the potential energy of a system, e.g., a protein, as a function of all its atomic coordinates: $V = V(\{\boldsymbol{R}_I\})$ with $\boldsymbol{R}_I$ denoting the positions of the atomic nuclei and $I = 1 \dots N_{\mathrm{at}}$, where N_{at} is the number of atoms. Considering a $(3N_{\mathrm{at}}+1)$-dimensional space, $V(\{\boldsymbol{R}_I\})$ is a surface in this space, depending on $3N_{\mathrm{at}}$ coordinates with the $(3N_{\mathrm{at}}+1)$th dimension being the value of the potential energy. Local minima of the PES are points where the gradient of the potential energy vanishes and where every infinitesimally small variation of the coordinates will lead to an increase of $V(\{\boldsymbol{R}_I\})$. The lowest-energy minimum is denoted as the global minimum of the PES. Methods to calculate $V(\{\boldsymbol{R}_I\})$ will be addressed in Chapter 3.

At physiological conditions, the quantity that the protein or peptide aims to minimize is the free energy. In order to understand and describe the thermodynamic properties of proteins and peptides the PES is thus often projected onto the free-energy surface (FES) using a set of (reaction) coordinates or order parameters $\{X_i\}$. The FES is the actual surface that is explored during the folding or structure formation process. In contrast to the PES, which is a high-dimensional function depending on $3N_{\mathrm{at}}$ coordinates, the FES is typically described by only one or two order parameters $\{X_i\}$ (at least less than $3N_{\mathrm{at}}$). All other degrees of freedom enter the FES as averages for fixed values of the order parameters. Frequently chosen order parameters $\{X_i\}$ include the number of hydrogen bonds, the radius of gyration, the electric dipole moment, or the root mean square deviation (RMSD) of different conformations.

The free energy $F(\{X_i\})$ of a specific state $\{X_i\}$ is related to the probability $P(\{X_i\})$ that this state is occupied via[131–133]:

$$F(\{X_i\}) - F(\{X_i'\}) = -k_{\mathrm{B}}T[\ln P(\{X_i\}) - \ln P(\{X_i'\})] \quad , \tag{2.1}$$

with $\{X_i'\}$ being a reference state. We here denote the free energy with F as we refer to the Helmholtz free energy[131]

$$F(T,V) = U - TS \quad , \tag{2.2}$$

where U is the internal energy averaged over all states of the ensemble, T is the temperature, and S is the entropy. If the pressure p instead of the volume V is kept constant, the relevant thermodynamical potential is the Gibbs free energy

$$G(T,p) = U - TS + pV \quad . \tag{2.3}$$

In this thesis, we use the Helmholtz free energy as the experiments of polypeptides in the gas phase (as discussed in Chapter 7) are performed at $p \simeq 0$. Moreover, we are concerned with energy differences, where for different conformers the term $p\Delta V$ can be neglected due to the extreme dilution of the peptides, so that $\Delta G = \Delta F$.

3 THEORETICAL METHODS TO DESCRIBE THE ENERGY LANDSCAPE

This chapter intends to give a tutorial overview of the different theoretical methods used in this work to describe the potential-energy surface (PES) of peptides and proteins. Within the scope of the present thesis, this overview neither claims, nor aims for completeness. The interested reader is referred to one of the many textbooks giving in-depth descriptions of the field, e.g., Refs. [134–137].

The present chapter starts with a description of empirical models of the potential-energy function, called force fields. Force fields are classical models that do not take into account the electronic structure explicitly with the advantage of being computationally cheap compared to first-principles methods, which are described in the subsequent sections. First principles (or *ab initio*) means that the quantum-mechanical many-body problem is solved based only on the fundamental physical laws [Dirac or Schrödinger equation (SGE)], possibly including physically motivated approximations, but without using model Hamiltonians or relying on empirical parameters. First-principles methods enable us to obtain a much more reliable PES than force fields. However, at the same time, they are much more computationally expensive.

3.1 FORCE FIELDS

Empirical potential-energy functions, called "force fields", are widely used in computer simulations of peptides and proteins[1, 2, 138]. A force field is constituted by a functional expression used to describe the potential energy and the corresponding parameters that enter it. The latter are determined by fitting to a set of experimental and/or theoretical data from quantum-mechanical calculations. In fact, there are many different force fields. However, most of the widely used force fields employ a similar form for the energy expression and similar techniques for determining the parameters. For a more detailed description, we thus concentrate on one representative of the standard force fields, namely the OPLS-AA force field[139] as it is used in the present thesis. The given formulae follow Ref. [139].

The molecular interactions are divided into non-bonded interactions and bonded interactions, where the latter comprise contributions from bonds, angles, and torsions. The overall potential-energy function is given as:

$$E_{\mathrm{tot}} = E_{\mathrm{bond}} + E_{\mathrm{angle}} + E_{\mathrm{torsion}} + E_{\mathrm{non-bonded}} \quad . \tag{3.1}$$

The potential-energy term for the non-bonded interactions contains a Coulomb term accounting

for electrostatic interactions and a Lennard-Jones potential term accounting for dispersion interactions and Pauli repulsion:

$$E_{\mathrm{non-bonded}}(\mathrm{AB}) = \sum_{I}^{\text{on A}} \sum_{J}^{\text{on B}} f_{IJ} \left[q_I q_J \frac{e^2}{R_{IJ}} + 4\epsilon_{IJ} \left(\frac{\sigma_{IJ}^{12}}{R_{IJ}^{12}} - \frac{\sigma_{IJ}^{6}}{r_{IJ}^{6}} \right) \right] \quad . \tag{3.2}$$

The sum runs over all atoms I on molecule A and all atoms J on molecule B. For intramolecular non-bonded interactions, the same expression is employed with $I < J$ to avoid double counting. The scaling factor f_{IJ} is 0 if I and J are separated by less than three bonds, 0.5 if they are separated by three bonds and 1 otherwise. The charges of the atomic units are denoted by q_I and R_{IJ} is the distance between atom I and J. The parameters ϵ_{IJ} and σ_{IJ} describe the shape of the Lennard-Jones potential. As a first step in the development of the OPLS-AA force field, most of the parameters were adopted from the OPLS-UA force field[139–141], the predecessor of OPLS-AA. UA stands for "united atom" and means that not all atoms are treated explicitly [as it is done in the OPLS-AA (all-atom) approach]. In fact, the hydrogen atoms attached to aliphatic carbon atoms are treated implicitly by adjusting the parameters for the carbon atoms accordingly. For the OPLS-AA force field, the parameters that were adopted from the OPLS-UA force field were refitted to the properties of organic liquids[139].

The bonded interactions include contributions associated with bond stretching, angle bending and dihedral-angle rotations. Bond and angle deformations are described by harmonic springs connecting the atoms, with K_r and K_θ denoting the corresponding spring constants:

$$E_{\mathrm{bond}} = \sum_{\mathrm{bonds}} K_r (R - R_{\mathrm{eq}})^2 \quad , \tag{3.3}$$

$$E_{\mathrm{angle}} = \sum_{\mathrm{angles}} K_\theta (\theta - \theta_{\mathrm{eq}})^2 \quad , \tag{3.4}$$

R is the distance between the bonded atoms and R_{eq} denotes the equilibrium distance. Analogously, θ is the angle between the atoms with θ_{eq} being the equilibrium angle. Most of the force constants K_r and K_θ were adopted from the (pre-existing) Amber force field[139, 142]. The energy as a function of the dihedral angles is represented by a Fourier series:

$$E_{\mathrm{torsion}} = \sum_i \frac{V_1^i}{2}[1 + \cos(\phi_i)] + \frac{V_2^i}{2}[1 - \cos(2\phi_i)] + \frac{V_3^i}{2}[1 + \cos(3\phi_i)] \quad , \tag{3.5}$$

where the sum runs over all torsional angles ϕ_i. The parameters V_1^i, V_2^i, and V_3^i were obtained from fitting to MP2 data for the alanine dipeptide[143] (see Section 3.3 for a description of the MP2 method).

As well as the OPLS force field, other popular and widely used standard force fields include Amber[142, 144] and CHARMM[145–147]. The Amoeba force field, which is developed in the group of Jay W. Ponder[148], is a "next generation" force field moving away from the fixed-charge model to a description that takes polarization effects into account.

Due to the high computational cost of first-principles methods, force fields are currently the only feasible approach to sample the conformational space of large peptides or whole proteins. However, their reliability is generally restricted by two limitations[138]: additivity and transferability. Additivity is related to the potential-energy function, where it is assumed

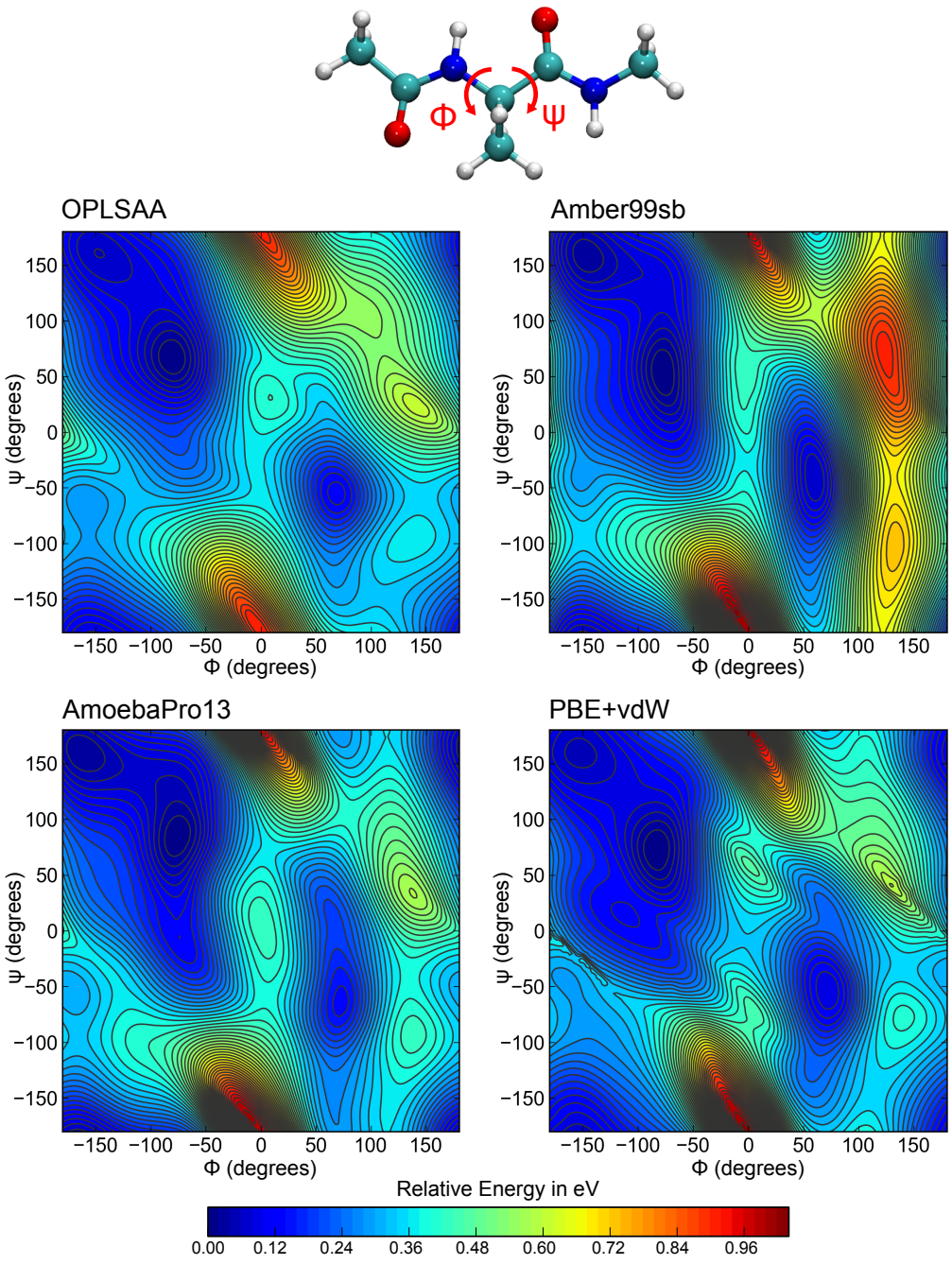


Figure 3.1: Top: Structural representation of the alanine dipeptide and its dihedral angles ϕ and ψ. Bottom: Contour maps of the alanine-dipeptide Ramachandran (ϕ,ψ)-surface computed with the OPLSAA, Amber99sb, AmoebaPro13 force fields and at the density-functional theory (DFT) level of theory using the PBE+vdW functional. A grid spacing of 10° was used. Contour lines are drawn every 20 meV. The color code gives the energy relative to the respective global minimum. For the PBE+vdW plot, we increased the grid spacing to 5° for ϕ between -180° and -120° and for ψ between -80° and 0° to obtain a better resolution. However, as slight changes in the angles involve large changes in energy, there are still some small artifacts visible in this region.

that, for all systems, the potential energy can be expressed as a sum over different contributions with a rather simple physical interpretation. Transferability refers to the parameters. They are determined based on a (necessarily limited) set of systems and structures and it is questionable to what extent they can be used to describe a much wider range of systems. In fact, it is well known that the detailed PES of different force fields differ (manifested, e.g., in Ramachandran plots[13, 14]), and that larger-scale conformational properties can also deviate[149–152]. We illustrate this here based on the example of the Ramachandran surface of the alanine dipeptide that we computed with the OPLSAA[143], Amber99sb[153] and the higher-level polarizable AmoebaPro13[154] force fields (see Fig. 3.1). For a comparison, we also calculated the same at a first-principles level of theory (density-functional theory (DFT) with the PBE+vdW functional, explained in more detail in Section 3.5). We explored the (ϕ, ψ) dihedral-angle space using a two-dimensional grid with a spacing of 10° for both dihedrals. For each of the 1296 grid points, we performed a geometry relaxation with constrained ϕ and ψ values using the respective method. For the force-field part, version 6.2 of the Tinker program[155] was used and for the PBE+vdW calculations, we employed the "Fritz Haber Institute *ab initio* molecular simulation" (`FHI-aims`) code with the PLUMED[156] interface (see Section 3.6 for more details on `FHI-aims`).

The contour plots of the Ramachandran surface obtained with the different methods are depicted in Fig. 3.1. While they look qualitatively similar, they indeed show differences, e.g., in the position of the local minima and maxima. The local minimum at around ($\phi = -110^\circ$,$\psi = 10^\circ$) of the PBE+vdW surface is not found by any of the force fields. Still, both the OPLSAA and the AmoebaPro13 data resemble more closely the PBE+vdW results than Amber99sb, with PBE+vdW being the highest-level theoretical method tested here. In this context, it is interesting to note that both the torsional parameters of the OPLSAA and the AmoebaPro13 force field have been determined by fitting to the Ramachandran surface of the alanine dipeptide obtained using (single-point) MP2 calculations (the MP2 method will be explained in Section 3.4). The torsional parameters are traditionally determined as the last step in the parametrization process so that the total force-field energy is effectively fitted to reflect the training data, in this case the Ramachandran surface of the alanine dipeptide. On the other hand, for the parametrization of the torsional parameters of the Amber99sb force field, no alanine dipeptide data was used, but MP2 energy differences of alanine and glycine tetrapeptide conformers. Together with the fact that the OPLSAA and the AmoebaPro13 data are more similar to the higher-level PBE+vdW results this highlights once again the transferability problem of force fields. They perform well for the set of structures they were actually parametrized for, but transferability to different sets of structures can not be guaranteed. In the following, we will discuss first-principles methods for describing the PES of molecules based on the solution of the SGE. Such approaches have a wider range of validity due to their more rigorous quantum-mechanical footing.

3.2 THE QUANTUM-MECHANICAL MANY-BODY PROBLEM

The properties of a piece of matter, composed of nuclei and electrons, are determined by the fundamental laws of quantum mechanics. The SGE[157] in its time-independent form reads:[1]

$$\hat{\mathcal{H}}\Psi_n(\{\boldsymbol{R}_I\},\{\boldsymbol{x}_i\}) = \mathcal{E}_n\Psi_n(\{\boldsymbol{R}_I\},\{\boldsymbol{x}_i\}) \quad , \tag{3.6}$$

where Ψ_n are the eigenvectors and $\mathcal{E}_n$ the corresponding eigenvalues, with $n = 0$ corresponding to the ground state. Ψ_n is a many-body wave function depending on the set of coordinates of the nuclei ($\{\boldsymbol{R}_I\}, I = 1 \dots N_{\text{at}}$) and the set of spatial coordinates $\boldsymbol{r}_i$ and spin coordinates σ_i of the electrons with $\{\boldsymbol{x}_i\} = \{(\boldsymbol{r}_i, \sigma_i)\}$ with $i = 1 \dots N_{\text{el}}$. N_{el} is the number of electrons and N_{at} the number of atoms in the system. $\hat{\mathcal{H}}$ denotes the Hamiltonian operator, here written in atomic units[2]:

$$\hat{\mathcal{H}} = \underbrace{-\sum_I^{N_{\text{at}}} \frac{\nabla_I^2}{2M_I}}_{\hat{T}_{\text{n}}} \underbrace{-\sum_i^{N_{\text{el}}} \frac{\nabla_i^2}{2}}_{\hat{T}_{\text{e}}} + \underbrace{\frac{1}{2}\sum_I^{N_{\text{at}}}\sum_{J\neq I}^{N_{\text{at}}} \frac{Z_I Z_J}{|\boldsymbol{R}_I - \boldsymbol{R}_J|}}_{\hat{V}_{\text{nn}}} + \underbrace{\frac{1}{2}\sum_i^{N_{\text{el}}}\sum_{j\neq i}^{N_{\text{el}}} \frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|}}_{\hat{V}_{\text{ee}}} \underbrace{-\sum_i^{N_{\text{el}}}\sum_I^{N_{\text{at}}} \frac{Z_I}{|\boldsymbol{r}_i - \boldsymbol{R}_I|}}_{\hat{V}_{\text{en}}} \quad . \tag{3.7}$$

M_I is the mass and Z_I the atomic number of the corresponding nucleus. $\hat{T}_{\text{n}}$ is the kinetic energy operator for the nuclei and $\hat{T}_{\text{e}}$ the electronic kinetic energy operator. $\hat{V}_{\text{ee}}$ denotes the electron-electron interaction, $\hat{V}_{\text{nn}}$ the nuclear-nuclear interaction, and $\hat{V}_{\text{en}}$ the electron-nuclear interaction. Although we can write down a mathematical correct framework that determines the properties of the system under study, the critical point is to find the many-body wave function, which is a complex object depending on all nuclear and electronic coordinates. In order to solve the SGE for realistic systems, it is thus crucial to find suitable approximations. The first approximation usually applied is the Born-Oppenheimer approximation, which is discussed in the next section.

3.2.1 THE BORN-OPPENHEIMER APPROXIMATION

The physical reasoning behind the Born-Oppenheimer approximation[160] is based on the huge mass difference between electrons and nuclei. The lightest nucleus, a single proton, has a mass m_{P} that is about 1800 times larger than the mass of an electron m_{e}. This suggests that electronic and nuclear motion can be (approximately) decoupled. Upon a possible motion of the nuclei, the electrons – or strictly speaking the electronic wave function – will instantaneously adjust to the new nuclear positions. Meanwhile, they will always stay in the same electronic state; the movement of the nuclei does not induce electronic transitions. This is the "adiabatic" or Born-Oppenheimer approximation. Mathematically, this decoupling of electronic and nuclear motions can be realized by a product ansatz of the total wave function:

$$\Psi_{\text{BO}}(\{\boldsymbol{R}_I\},\{\boldsymbol{x}_i\}) = \Theta(\{\boldsymbol{R}_I\})\Phi_n(\{\boldsymbol{R}_I\},\{\boldsymbol{x}_i\}) \quad , \tag{3.8}$$

[1] One should bare in mind that the universal equation naturally accounting for the spin of the electrons and relativistic effects is the Dirac equation[158, 159]. However, solving the SGE is suitable for the problems treated in this thesis.
[2] We will use Hartree atomic units throughout the rest of this thesis, unless explicitly stated differently.

where $\Phi_n(\{\boldsymbol{R}_I\},\{\boldsymbol{x}_i\})$ denotes the electronic part of the wave function and $\Theta(\{\boldsymbol{R}_I\})$ indicates the nuclear wave function. $\Phi_n(\{\boldsymbol{R}_I\},\{\boldsymbol{x}_i\})$ is a solution to the electronic SGE

$$\hat{\mathcal{H}}^{\mathrm{e}}\Phi_n(\{\boldsymbol{R}_I\},\{\boldsymbol{x}_i\}) = E_n^{\mathrm{e}}(\{\boldsymbol{R}_I\})\Phi_n(\{\boldsymbol{R}_I\},\{\boldsymbol{x}_i\}) \quad , \tag{3.9}$$

with the electronic Hamiltonian

$$\hat{\mathcal{H}}^{\mathrm{e}} = \hat{T}_{\mathrm{e}} + \hat{V}_{\mathrm{ee}} + \hat{V}_{\mathrm{en}} \quad . \tag{3.10}$$

$\Phi_n(\{\boldsymbol{R}_I\},\{\boldsymbol{x}_i\})$ defines the electronic wave function for a fixed configuration of the nuclei, where the eigenvalues $E_n^{\mathrm{e}}(\{\boldsymbol{R}_I\})$ depend on the nuclear positions. As $\hat{\mathcal{H}}$ does not direcly act on the nuclear coordinates $\{\boldsymbol{R}_I\}$, they are solely parameters in $\Phi_n(\{\boldsymbol{R}_I\},\{\boldsymbol{x}_i\})$.

If the ansatz $\Psi_{\mathrm{BO}}(\{\boldsymbol{R}_I\},\{\boldsymbol{x}_i\})$ (Eq. 3.8) is put into the SGE, one finds:

$$\begin{aligned}
\hat{\mathcal{H}}\Theta(\{\boldsymbol{R}_I\})\Phi_n(\{\boldsymbol{R}_I\},\{\boldsymbol{x}_i\}) &= (\hat{T}_{\mathrm{n}} + \hat{V}_{\mathrm{nn}} + \hat{\mathcal{H}}^{\mathrm{e}})\Theta(\{\boldsymbol{R}_I\})\Phi_n(\{\boldsymbol{R}_I\},\{\boldsymbol{x}_i\}) \\
&= \Phi_n(\{\boldsymbol{R}_I\},\{\boldsymbol{x}_i\})\cdot\hat{V}_{\mathrm{nn}}\Theta(\{\boldsymbol{R}_I\}) \\
&\quad +\Phi_n(\{\boldsymbol{R}_I\},\{\boldsymbol{x}_i\})\cdot E_n^{\mathrm{e}}(\{\boldsymbol{R}_i\})\Theta(\{\boldsymbol{R}_I\}) \\
&\quad +\Phi_n(\{\boldsymbol{R}_I\},\{\boldsymbol{x}_i\})\cdot\hat{T}_{\mathrm{n}}\Theta(\{\boldsymbol{R}_I\}) \\
&\quad -\sum_I^{N_{\mathrm{at}}}\frac{1}{2M_I}\Theta(\{\boldsymbol{R}_I\})\nabla_I^2\Phi_n(\{\boldsymbol{R}_I\},\{\boldsymbol{x}_i\}) \\
&\quad -\sum_I^{N_{\mathrm{at}}}\frac{1}{M_I}[\nabla_I\Theta(\{\boldsymbol{R}_I\})][\nabla_I\Phi_n(\{\boldsymbol{R}_I\},\{\boldsymbol{x}_i\})] \quad .
\end{aligned} \tag{3.11}$$

The assumption of the Born-Oppenheimer approximation is to neglect the last two terms. They result from the nuclear kinetic-energy operator acting on the nuclear coordinates in $\Phi_n(\{\boldsymbol{R}_I\},\{\boldsymbol{x}_i\})$ and decrease with $\frac{1}{M_I}$[136]. Thus, in the limit of infinite proton mass the approximation becomes exact. In this case, the nuclei move in an effective potential

$$V(\{\boldsymbol{R}_I\}) = V_{\mathrm{nn}}(\{\boldsymbol{R}_I\}) + E_n^{\mathrm{e}}(\{\boldsymbol{R}_I\}) \quad . \tag{3.12}$$

For $n=0$, i.e., for the electronic ground state, $E_0^{\mathrm{e}}(\{\boldsymbol{R}_I\})$, this effective potential is called the Born-Oppenheimer surface or potential-energy surface (PES) V_{BO}. A more detailed account of the Born-Oppenheimer approximation can be found, e.g., in the book by J. Kohanoff[136].

The Born-Oppenheimer approximation reduces the quantum many-body problem in Eq. 3.6 to an electronic-structure problem that involves solving the electronic time-independent SGE, Eq. 3.9. The following sections focus first on wave function-based approaches, which aim at finding accurate approximations to the electronic wave function. Subsequently, density-based approaches are discussed, where not the wave function is the focus of interest, but the electronic density.

3.3 Hartree-Fock method

The electron-electron interaction $\hat{V}_{\mathrm{ee}} = \frac{1}{2}\sum_i^{N_{\mathrm{el}}}\sum_{j\neq i}^{N_{\mathrm{el}}}\frac{1}{|\boldsymbol{r}_i-\boldsymbol{r}_j|}$ couples the spatial coordinates of the electrons. Thus, a simple product ansatz of one-electron wave functions generally does

not yield a solution to the SGE. However, one can pursue such an approach to find an upper boundary for the ground-state energy E_0 of the system (Hartree method[161, 162]). According to the variational principle the expectation value of the energy for any trial wave function (not equal to the ground-state wave function) is always larger than the ground-state energy E_0:

$$E_0 < E_{\mathrm{Hartree}} = \left\langle \Phi_{\mathrm{Hartree}} \left| \hat{\mathcal{H}}^{\mathrm{e}} \right| \Phi_{\mathrm{Hartree}} \right\rangle \quad , \tag{3.13}$$

where Φ_{Hartree} is the product function of one-particle wave functions yielding the lowest energy under the contraint of $\langle \Phi_{\mathrm{Hartree}} | \Phi_{\mathrm{Hartree}} \rangle = 1$. However, electrons are fermionic particles; according to the Pauli principle their wave function has to be antisymmetric, i.e., it has to change sign upon the exchange of two particle coordinates:

$$\Phi(\dots, \boldsymbol{x}_i, \dots, \boldsymbol{x}_j, \dots) = -\Phi(\dots, \boldsymbol{x}_j, \dots, \boldsymbol{x}_i, \dots) \quad . \tag{3.14}$$

An extension to the Hartree-method is the Hartree-Fock method, where the wave function is written as a Slater determinant of single-electron orbitals[163, 164]. The mathematical form of a determinant naturally yields the antisymmetry of the wave function with

$$\Phi^{\mathrm{HF}}(\{\boldsymbol{x}_i\}) = \frac{1}{\sqrt{N_{\mathrm{el}}!}} \begin{vmatrix} \varphi_1(\boldsymbol{x}_1) & \varphi_2(\boldsymbol{x}_1) & \cdots & \varphi_{N_{\mathrm{el}}}(\boldsymbol{x}_1) \\ \varphi_1(\boldsymbol{x}_2) & \varphi_2(\boldsymbol{x}_2) & \cdots & \varphi_{N_{\mathrm{el}}}(\boldsymbol{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_1(\boldsymbol{x}_{N_{\mathrm{el}}}) & \varphi_2(\boldsymbol{x}_{N_{\mathrm{el}}}) & \cdots & \varphi_{N_{\mathrm{el}}}(\boldsymbol{x}_{N_{\mathrm{el}}}) \end{vmatrix} \quad , \tag{3.15}$$

where $\{\boldsymbol{x}_i\} = \{(\boldsymbol{r}_i, \sigma_i)\}$ again captures the spatial and spin coordinates for electron i, while $\{\varphi_i\}$ indicate a set of orthonormalized one-electron spin orbitals. The Hartree-Fock energy can then be written as

$$\begin{aligned} E^{\mathrm{HF}} = \left\langle \Phi^{\mathrm{HF}} \left| \hat{\mathcal{H}}^{\mathrm{e}} \right| \Phi^{\mathrm{HF}} \right\rangle \quad = \quad & -\sum_i^{N_{\mathrm{el}}} \int \varphi_i^*(\boldsymbol{x}_i) \frac{\nabla_i^2}{2} \varphi_i(\boldsymbol{x}_i) \mathrm{d}\boldsymbol{x}_i \\ & -\sum_i^{N_{\mathrm{el}}} \sum_I^{N_{\mathrm{at}}} \int \frac{Z_I}{|\boldsymbol{r}_i - \boldsymbol{R}_I|} |\varphi_i(\boldsymbol{x}_i)|^2 \, \mathrm{d}\boldsymbol{x}_i \\ & + \frac{1}{2} \sum_i^{N_{\mathrm{el}}} \sum_j^{N_{\mathrm{el}}} \underbrace{\int\int \frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \overbrace{\varphi_i^*(\boldsymbol{x}_i)\varphi_j^*(\boldsymbol{x}_j)\varphi_i(\boldsymbol{x}_i)\varphi_j(\boldsymbol{x}_j)}^{|\varphi_i(\boldsymbol{x}_i)|^2 |\varphi_j(\boldsymbol{x}_j)|^2} \mathrm{d}\boldsymbol{x}_i \mathrm{d}\boldsymbol{x}_j}_{\text{Coulomb integral } J_{ij}} \\ & - \frac{1}{2} \sum_i^{N_{\mathrm{el}}} \sum_j^{N_{\mathrm{el}}} \underbrace{\int\int \frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \varphi_i^*(\boldsymbol{x}_i)\varphi_j^*(\boldsymbol{x}_j)\varphi_i(\boldsymbol{x}_j)\varphi_j(\boldsymbol{x}_i) \mathrm{d}\boldsymbol{x}_i \mathrm{d}\boldsymbol{x}_j}_{\text{Exchange integral } K_{ij}} \quad . \end{aligned} \tag{3.16}$$

The first term can be associated with the kinetic energy of the electrons. The second term involves the electron-nuclei interaction, i.e., the Coulomb energy of the electronic charge density in the electric field generated by the nuclei. The third term is denoted as the *Hartree energy* and is composed of a sum over the so-called "Coulomb integrals" as highlighted in the equation. Its physical interpretation is a classical Coulomb energy between two charge distributions. The only difference to the fifth term, the *exchange energy*, is that the coordinates of the spin orbitals in the

integrals are exchanged. Hence, the name "exchange integrals". The *exchange energy* associated with the exchange integrals is a quantum phenomenon, which does not have a classical physical explanation. It arises through the ansatz of the wave function as an antisymmetrized product of one-electron spin orbitals and couples only electrons in the same spin state.[3] Thus, the energy increases if electrons with the same spin come closer to each other. This fulfils Pauli's principle, which in a single-particle picture states that two electrons with the same spin cannot occupy the same state. The self-interaction, namely the case $j = i$, exactly cancels in the Coulomb and exchange integrals. Thus, $i = j$ can be included in both sums, where the Hartree-Fock method itself remains self-interaction free by construction. With the expression for the energy at hand, the variational principle can be applied to find the Slater determinant yielding the lowest energy. The variational expression reads:

$$\delta \left[E^{\mathrm{HF}} - \sum_{i}^{N_{\mathrm{el}}} \sum_{j}^{N_{\mathrm{el}}} \lambda_{ij} \left(\int \varphi_i^*(\boldsymbol{x}_i) \varphi_j(\boldsymbol{x}_i) \mathrm{d}\boldsymbol{x}_i - \delta_{ij} \right) \right] = 0 \qquad . \tag{3.17}$$

The variation of the energy with respect to infinitesimal small changes of the spin-orbitals $\delta\varphi_i^*(\boldsymbol{x}_i)$ needs to be zero under the additional constraint of ortho-normalized spin orbitals. The latter is accounted for by employing Lagrange multipliers. When this is explicitly carried out, one arrives at a set of single-particle equations for the spin orbitals, termed the "Hartree-Fock equations". In this way, the many-body problem is reduced to a set of coupled effective one-particle equations[136][4]:

$$\begin{aligned} \hat{\mathcal{F}}_i \varphi_i(\boldsymbol{x}_i) &= \left(-\frac{\nabla_i^2}{2} - \sum_{I}^{N_{\mathrm{at}}} \frac{Z_I}{|\boldsymbol{r}_i - \boldsymbol{R}_I|} \right) \varphi_i(\boldsymbol{x}_i) \\ &\quad + \sum_{j}^{N_{\mathrm{el}}} \underbrace{\left(\int \varphi_j^*(\boldsymbol{x}_j) \frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \varphi_j(\boldsymbol{x}_j) \mathrm{d}\boldsymbol{x}_j \right) \varphi_i(\boldsymbol{x}_i)}_{\hat{J}_{ij}\varphi_i(\boldsymbol{x}_i)} \\ &\quad - \sum_{j}^{N_{\mathrm{el}}} \underbrace{\left(\int \varphi_j^*(\boldsymbol{x}_j) \frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \varphi_i(\boldsymbol{x}_j) \mathrm{d}\boldsymbol{x}_j \right) \varphi_j(\boldsymbol{x}_i)}_{\hat{K}_{ij}\varphi_i(\boldsymbol{x}_i)} \\ &= \varepsilon_i \varphi_i(\boldsymbol{x}_i) \end{aligned} \tag{3.18}$$

The Hartree-Fock equations constitute a self-consistency problem; the solution to the equations, i.e., the orbitals, depends on the orbitals in turn. One approach to solve this self-consistency problem is to expand the single-electron spin orbitals into a suitable basis set and solve the equations based on an initial guess for the Slater determinant or an initial guess for the effective potential. Using the solutions the equation can be set up and solved again. This procedure is repeated until the solution does not change anymore, i.e., when self consistency is reached.

[3] This can be seen if the exchange integral is explicitly carried out: $\int \mathrm{d}\boldsymbol{x} = \int \mathrm{d}r^3 \sum_\sigma$, where $\boldsymbol{r}$ are the spatial and σ the spin coordinates. φ_i can be separated into a spatial and a spin part $\varphi_i(\boldsymbol{x}) = \Lambda_i(\boldsymbol{r}) \cdot \chi_i(\sigma)$ with $\sum_\sigma \chi_j^*(\sigma)\chi_i(\sigma) = \delta_{ij}$, i.e., the exchange integral is zero if φ_i and φ_j are not associated with the same spin state.

[4] The orbitals can be chosen to fulfil Eq. 3.18. In principle, ε is a matrix. For details, see e.g., Ref. [136].

The Hartree-Fock energy in terms of the eigenvalues ε_i reads:

$$E^{\mathrm{HF}} = \sum_{i}^{N_{\mathrm{el}}} \varepsilon_i - \frac{1}{2} \sum_{i}^{N_{\mathrm{el}}} \sum_{j}^{N_{\mathrm{el}}} (J_{ij} - K_{ij}) \quad , \tag{3.19}$$

where J_{ij} and K_{ij} are the Coulomb and exchange integrals, respectively, defined in Eq. 3.16. The eigenvalues ε_i can be interpreted in terms of (approximate) ionization energies (Koopmans' theorem[165]). Assuming that the remaining orbitals do not change upon removing one electron from orbital i, the ionization energy I can be calculated as $I_i \stackrel{!}{=} E_i^{\mathrm{HF}}(N_{\mathrm{el}} - 1) - E^{\mathrm{HF}}(N_{\mathrm{el}}) \approx -\varepsilon_i$.

3.4 Beyond Hartree-Fock theory: electron correlation

By construction, the Hartree-Fock method yields the best approximation to the many-body wave function solution of the SGE based on a single-determinant ansatz. The exchange integral couples electrons with the same spin state. However, also electrons with different spin states are correlated via the non-local two-body Coulomb operator $\frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|}$. Due to the single-determinant ansatz in the Hartree-Fock method, the electron "sees" only the mean field of the other electrons. Quantum correlation (with the exception of Pauli correlation) is not captured. In fact, the quantum-chemical definition of the correlation energy is the difference between the true ground-state energy E_0 associated with the correct many-body wave function and the Hartree-Fock energy: $E_{\mathrm{corr}} = E_0 - E^{\mathrm{HF}} = E_{\mathrm{corr}}$[135, 136].

To capture the electronic correlations, one has to go beyond a single-determinant ansatz. As the Hartree-Fock method usually captures about 99% of the total energy[136], this is often done by using the Hartree-Fock solution as a starting point. If the Hartree-Fock problem is solved by expanding the orbitals in a basis set of P (linear independent) basis functions, one obtains N_{el} occupied and $(P - N_{\mathrm{el}})$ unoccupied single-electron spin orbitals. Based on the Hartree-Fock wave function, i.e., the ground-state wave function which involves all occupied orbitals, one can construct determinants that involve unoccupied states, termed excitations. A Slater determinant, where one electron is excited from an occupied to an unoccupied state is called a *single excitation* or *single* in short and is usually referred to by the letter *S*. A determinant with two electrons being promoted to unoccupied states is a *double excitation*, *double* or *D*. Analogously, one defines triple excitations (triples, *T*), quadruple excitations (quadruples, *Q*) and so on and so forth. The number of possible determinants involving excited states grows combinatorially with $\binom{P}{N_{\mathrm{el}}}$. If all possible determinants are included in the ansatz for the many-body wave function the method is referred to as full configuration interaction (CI). In the limit of a complete basis set, an ansatz involving all possible determinants yields the true many-body wave function. However, due to the combinatorial explosion, in practice normally a truncated CI version is used, i.e., only certain excitations are included. CIS refers to an ansatz, where single excitations are involved, in CISD singles and doubles are included and so on. The advantage of CI is that it is variational at each truncation level, i.e., the best wave function within the given ansatz always yields the lowest energy. However, truncated CI methods are not size extensive, which means that the energy does not scale linearly with the system size (number of particles). This is, e.g., a problem for the calculation of binding energies.

Another method to go beyond a single-determinant approach is to add correlation on top of the Hartree-Fock solution in a perturbative way. This was performed first by Møller and Plesset in 1934[166] and will be covered in the next section.

3.4.1 MØLLER-PLESSET PERTURBATION THEORY

If the Hamiltonian $\hat{\mathcal{H}}$ of the underlying problem differs only by a small perturbation from a Hamiltonian $\hat{\mathcal{H}}^0$, where the solution is known, one can split the full Hamiltonian $\hat{\mathcal{H}}$ into two parts:

$$\hat{\mathcal{H}} = \hat{\mathcal{H}}^0 + \lambda\Delta\hat{\mathcal{H}} \quad , \tag{3.20}$$

where $\lambda\Delta\hat{\mathcal{H}}$ is small. Furthermore, $\hat{\mathcal{H}}^0\Phi_i^{(0)} = E_i^{(0)}\Phi_i^{(0)}$, with $E_i^{(0)}$ and $\Phi_i^{(0)}$ denoting the unperturbed eigenvalues and eigenstates of $\hat{\mathcal{H}}^0$. The eigenvalues and eigenstates of $\hat{\mathcal{H}}$ can then be expanded in terms of λ

$$E_i = E_i^{(0)} + \lambda E_i^{(1)} + \lambda^2 E_i^{(2)} + \dots \quad , \tag{3.21}$$

$$\Phi_i = \Phi_i^{(0)} + \lambda\Phi_i^{(1)} + \lambda^2\Phi_i^{(2)} + \dots \quad . \tag{3.22}$$

One finds that

$$E_i^{(1)} = \left\langle \Phi_i^{(0)} \middle| \Delta\hat{\mathcal{H}} \middle| \Phi_i^{(0)} \right\rangle \quad , \tag{3.23}$$

$$E_i^{(2)} = \sum_{j\neq i} \frac{\left| \left\langle \Phi_j^{(0)} \middle| \Delta\hat{\mathcal{H}} \middle| \Phi_i^{(0)} \right\rangle \right|^2}{E_i^{(0)} - E_j^{(0)}} \quad . \tag{3.24}$$

The sum in Eq. 3.24 runs over all eigenstates of $\hat{\mathcal{H}}^0$. Perturbation theory is covered in all standard quantum-mechanics textbooks. For further details see, e.g., Ref. [135, 167].

The idea of Møller and Plesset in 1934[166] was to define the reference Hamiltonian as the sum of Fock operators $\hat{\mathcal{H}}^0 = \sum_i^{N_{\mathrm{el}}} \hat{\mathcal{F}}_i$, where the Fock operators were defined in Eq. 3.18. The actual Hamiltonian of the system is then $\hat{\mathcal{H}} = \hat{\mathcal{H}}^0 + \Delta\hat{\mathcal{H}}$ with

$$\Delta\hat{\mathcal{H}} = \frac{1}{2}\sum_i^{N_{\mathrm{el}}}\sum_{j\neq i}^{N_{\mathrm{el}}} \frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} - \sum_i^{N_{\mathrm{el}}}\sum_j^{N_{\mathrm{el}}} \left(\hat{J}_{ij} - \hat{K}_{ij}\right) \quad , \tag{3.25}$$

where $\hat{K}_{ij}$ and $\hat{J}_{ij}$ are the exchange and Coulomb operators defined in Eq. 3.18. The unperturbed eigenstates of $\hat{\mathcal{H}}^0$ are the Hartree-Fock Slater determinants with the ground-state wave function $\Phi_0^{(0)} = \Phi^{\mathrm{HF}}$ and the ground-state energy $E_0^{(0)} = \sum_i^{N_{\mathrm{el}}} \varepsilon_i$, where the sum runs over all occupied orbitals. The unperturbed energy plus the first-order correction corresponds to the Hartree-Fock energy:

$$E_{\mathrm{MP1}} = E_0^{(0)} + E_0^{(1)} = \sum_i^{N_{\mathrm{el}}} \varepsilon_i + \frac{1}{2}\sum_i^{N_{\mathrm{el}}}\sum_j^{N_{\mathrm{el}}} (J_{ij} - K_{ij}) - \sum_i^{N_{\mathrm{el}}}\sum_j^{N_{\mathrm{el}}} (J_{ij} - K_{ij}) = E^{\mathrm{HF}} \quad , \tag{3.26}$$

which is easily seen by comparing to Eq. 3.19. The second-order correction to the energy involves all eigenstates of $\hat{\mathcal{H}}^0$, i.e., all possible determinants that can be constructed from the Hartree-Fock orbitals involving all possible excitations. Basically, these are all determinants that would be

used in a full CI calculation. However, taking the Hartree-Fock determinant as the reference state in Eq. 3.24, all matrix elements that involve excitations higher than second order are zero as $\Delta\hat{\mathcal{H}}$ is a two-body operator and the Hartree-Fock orbitals are orthonormal. Single excitations do not contribute according to Brillouin's theorem[135]. Thus, the second-order correction to the energy includes only double excitations and reads

$$E_0^{(2)} = \sum_{\nu}^{\text{virtual}} \sum_{\mu<\nu}^{\text{virtual}} \sum_{i}^{\text{occ}} \sum_{j<i}^{\text{occ}} \frac{|\langle \Phi_0^{(0)} | \Delta\hat{\mathcal{H}} | \Phi_0^{ij\mu\nu} \rangle|^2}{E_0^{(0)} - E_0^{ij\mu\nu}} \quad , \tag{3.27}$$

with $\Phi_0^{ij\mu\nu}$ denoting a determinant where two electrons have been promoted from occupied states i and j to the virtual states μ and ν. The sum $E_0^{(0)} + E_0^{(1)} + E_0^{(2)}$ is called the MP2 energy, E_{MP2}. If P again denotes the number of basis functions, the computational effort of MP2 scales as P^5. It is the most popular correlated method as it scales relatively well compared to other methods (e.g., coupled-cluster theory, which will be explained below) and typically accounts for 80-90% of the correlation energy[136]. However, one drawback of the MP2 method is that it relies on the quality of the approximation of the Hartree-Fock wave function to the real many-body wave function. If HF does not yield a good approximation, MP2 will fail as well.

3.4.2 Coupled-cluster theory

Another method to introduce correlation beyond the Hartree-Fock level is coupled-cluster theory. Originally formulated for problems in nuclear physics[168], it has been used in the realm of quantum chemistry since the mid 1960s[169]. Its truncated version CCSD(T) (explained in more detail below) is often referred to as the "gold standard" of quantum chemistry as it yields very high accuracy, while still being computationally feasible with a scaling of P^7[170–172]. In fact, based on CCSD(T) calculations chemical accuracy or even subchemical accuracy, i.e., errors lower than ≈1 kcal/mol or 43 meV, can be obtained for the interaction energies of molecules with usual system sizes of up to ≈ 30 light atoms[171]. CCSD(T) calculations are often employed for benchmarks[173, 174] and would be the ultimate goal for the (large) molecules dealt with in the present thesis. However, the unfavorable scaling makes it infeasible for the system sizes treated in this work (108 to 440 atoms). Nevertheless, a brief account is given in the following.

The wave function in coupled-cluster theory reads[136]

$$\Phi^{\text{CC}} = \mathrm{e}^{\hat{T}} \Phi^{\text{HF}} \quad , \tag{3.28}$$

where Φ^{HF} is the Hartree-Fock Slater determinant and $\hat{T}$ is the cluster operator (not to be confused with the kinetic-energy operator), which is composed of a series of operators $\hat{T}_n$

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \hat{T}_4 + \cdots + \hat{T}_N \quad . \tag{3.29}$$

The operators $\hat{T}_n$ create all possible excitations up to a certain order N. $\hat{T}_1$ generates single

excitations, $\hat{T}_2$ generates double excitations and so on:

$$\begin{aligned} \hat{T}_1\Phi^{\mathrm{HF}} &= \sum_{\nu}^{\text{virtual}} \sum_{i}^{\text{occ}} t_i^{\nu} \Phi^{i\nu} \quad , \\ \hat{T}_2\Phi^{\mathrm{HF}} &= \sum_{\nu}^{\text{virtual}} \sum_{\mu<\nu}^{\text{virtual}} \sum_{i}^{\text{occ}} \sum_{j<i}^{\text{occ}} t_{ij}^{\nu\mu} \Phi^{ij\nu\mu} \quad , \\ &\vdots \end{aligned} \tag{3.30}$$

where t denote the excitation coefficients. The exponential function can be expanded yielding

$$\mathrm{e}^{\hat{T}} = \hat{I} + \hat{T}_1 + \left(\hat{T}_2 + \frac{1}{2}\hat{T}_1^2\right) + \left(\hat{T}_3 + \hat{T}_2\hat{T}_1 + \frac{1}{6}\hat{T}_1^3\right) + \cdots \quad , \tag{3.31}$$

where the terms producing the same order of excitations are grouped together. The first term reproduces the Hartree-Fock wave function and the second term generates all single excitations. The terms in the first bracket generate double excitations, where $\hat{T}_2$ produces connected double excitations and $\hat{T}_1^2$ disconnected double excitations, and so on. By generating the wave function in this way, coupled-cluster theory and also all truncated versions of $\hat{T}$ become size extensive, i.e., the computed energy scales properly with the system size.

In practice, only truncated versions of the cluster operator $\hat{T}$ are computationally feasible. The method based on $\hat{T} = \hat{T}_1 + \hat{T}_2$ is called CCSD and scales with P^6. CCSDT, i.e., $\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3$, scales with P^8. Often the triple excitations are treated perturbatively, referred to as CCSD(T), which reduces the scaling to P^7. For further details, the reader is referred to Refs. [135, 170, 175].

3.5 Density-functional theory

The quantum-chemistry methods discussed in the previous sections focus on the wave function as the central quantity. As the name density-functional theory (DFT) implies, here, the fundamental quantity is the electronic density, which is defined as

$$n(\boldsymbol{r}) = \left\langle \Phi \left| \sum_{i=1}^{N} \delta(\boldsymbol{r} - \boldsymbol{r}_i) \right| \Phi \right\rangle \quad . \tag{3.32}$$

The electron density is real-valued and positive. It depends on three spatial coordinates, which is a large simplification compared to the many-body wave function Φ, which is a complex function of $4N$ coordinates (spatial and spin coordinates). The formal foundation of DFT is the Hohenberg-Kohn theorem[176]. It contains two statements:

1. The ground-state electron density $n_0(\boldsymbol{r})$ uniquely defines the external potential $v^{\mathrm{ext}}(\boldsymbol{r})$, except for an additive constant ($v^{\mathrm{ext}}(\boldsymbol{r})$ corresponds to the electron-nuclei interaction $-\sum_I^{N_{\mathrm{at}}} \frac{Z_I}{|\boldsymbol{r}-\boldsymbol{R}_I|}$ plus possible other external fields). This implies that $n_0(\boldsymbol{r})$ also defines the many-body wave function (ground and excited states) and thus, that all observables of the system are unique functionals of the ground-state density.

2. The energy of the system can be written as a functional of the density $E = E[n] = F_{\mathrm{HK}}[n] + \int v^{\mathrm{ext}}(\boldsymbol{r}) n(\boldsymbol{r}) \mathrm{d}\boldsymbol{r}$ for any external potential. $F_{\mathrm{HK}}[n] = T[n] + E_{\mathrm{ee}}[n]$ is a *universal*

functional, which contains the kinetic-energy functional $T[n]$ and the electron-electron interaction energy functional $E_{\mathrm{ee}}[n]$. The ground-state energy E_0 is the global minimum of $E[n]$, which is obtained by the exact ground-state density n_0, i.e., $E[n] \geq E_0 = E[n_0]$.

Proofs for these theorems can be found in the literature, e.g., in Ref. [137]. The original proofs by Hohenberg and Kohn[176] were conducted for systems with non-degenerate ground states. However, it can be shown that the Hohenberg-Kohn theorem is also valid for systems with degenerate ground states[177, 178]. The second part of the theorem states that there is a variational principle for $E[n]$ and thus provides a recipe how to obtain the ground-state density. For any variation $\delta n(\boldsymbol{r})$

$$\delta \left\{ E[n] - \mu \left(\int n(\boldsymbol{r}) \mathrm{d}\boldsymbol{r} - N_{\mathrm{el}} \right) \right\} = 0 \qquad , \tag{3.33}$$

where μ is a Langrange parameter that ensures the conservation of the particle number N_{el}. However, the universal functional $F[n]$ is not known such that Eq. 3.33 does not provide a practical solution. The most widely used approach, which allows for a practical use of the Hohenberg-Kohn theorem, was introduced in 1965 by Kohn and Sham[179] and will be discussed in the next section.

3.5.1 Kohn-Sham equations

The basic idea of the Kohn-Sham ansatz[179] is to map the interacting system of electrons onto an auxiliary system of non-interacting electrons with the same electronic density. The solution of the SGE for a non-interacting system of particles at $T = 0\,\mathrm{K}$ is a Slater determinant of one-electron orbitals φ_i with the electron density

$$n(\boldsymbol{r}) = \sum_i^{N_{\mathrm{el}}} |\varphi_i|^2 \qquad . \tag{3.34}$$

The kinetic energy for this non-interacting system is known and reads

$$T_{\mathrm{s}}[n] = -\sum_i^{N_{\mathrm{el}}} \left\langle \varphi_i \left| \frac{\nabla^2}{2} \right| \varphi_i \right\rangle \qquad , \tag{3.35}$$

where the subscript s stands for *single* particles. The complete ansatz from Kohn and Sham for the energy functional is

$$E[n] = T_{\mathrm{s}}[n] + \int v^{\mathrm{ext}}(\boldsymbol{r}) n(\boldsymbol{r}) \mathrm{d}\boldsymbol{r} + E^{\mathrm{H}}[n] + E^{\mathrm{XC}}[n] \qquad . \tag{3.36}$$

$E^{\mathrm{H}}[n]$ is the Hartree-energy term, which describes the classical Coulomb interaction of two charge distributions

$$E^{\mathrm{H}} = \frac{1}{2} \int \int \frac{n(\boldsymbol{r}) n(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} \mathrm{d}\boldsymbol{r} \mathrm{d}\boldsymbol{r}' \qquad . \tag{3.37}$$

The exchange-correlation functional

$$E^{\mathrm{XC}} = F_{\mathrm{HK}}[n] - T_{\mathrm{s}}[n] - E^{\mathrm{H}}[n] \tag{3.38}$$

captures everything that is not accounted for by approximating the kinetic-energy functional as the kinetic-energy functional of non-interacting particles and by approximating the electron-electron interaction with the Hartree energy. This means the exchange-correlation energy contains the kinetic correlation, the exchange energy (arising from Pauli's principle), the correlation energy, and the self-interaction correction. The latter is the error in the Hartree energy term that arises by the electrons interacting with themselves, which is easily seen when considering a one-electron system: even for a one-electron system, the Hartree energy would incorrectly yield a non-zero contribution. Hartree-Fock theory, on the other hand, is self-interaction free as the self-interaction term arises in the exchange term with the opposite sign and thus exactly cancels with the self-interaction term in the Hartree energy. Applying the variational principle to the Kohn-Sham energy functional (see Eq. 3.33) yields

$$\frac{\delta T_\mathrm{s}[n]}{\delta n(\boldsymbol{r})} + v^\mathrm{eff}(\boldsymbol{r}) = \mu \qquad , \tag{3.39}$$

with the effective potential $v^\mathrm{eff}(\boldsymbol{r})$:

$$v^\mathrm{eff}(\boldsymbol{r}) = v^\mathrm{ext}(\boldsymbol{r}) + \int \frac{n(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}\mathrm{d}\boldsymbol{r}' + v^\mathrm{XC}(\boldsymbol{r}) \qquad , \tag{3.40}$$

where the exchange-correlation potential is defined as $v^\mathrm{XC}(\boldsymbol{r}) = \frac{\delta E^\mathrm{XC}[n]}{\delta n(\boldsymbol{r})}$. As $T_s[n]$ is the kinetic energy of a system of non-interacting particles, this equation corresponds to the case of single particles moving in an effective potential $v^\mathrm{eff}(\boldsymbol{r})$. The single-particle orbitals φ_i satisfy one-particle Schrödinger equations

$$\left\{-\frac{\nabla^2}{2} + v^\mathrm{eff}(\boldsymbol{r})\right\}\varphi_i(\boldsymbol{r}) = \varepsilon_i\varphi_i(\boldsymbol{r}) \qquad , \tag{3.41}$$

which are known as the Kohn-Sham equations. The Kohn-Sham equations are *effective* single-particle equations as $v^\mathrm{eff}(\boldsymbol{r})$ depends on the electron density, i.e., they constitute a self-consistency problem similar to the Hartree-Fock equations (Eq. 3.18).

After solving the Kohn-Sham equations, the density of the interacting system can be calculated using Eq. 3.34 as the auxiliary non-interacting system was chosen to have the same density as the interacting system. The Kohn-Sham energy in terms of the eigenvalues reads:

$$E_\mathrm{KS}[n] = \sum_i^{N_\mathrm{el}} \varepsilon_i - E^\mathrm{H} - \int v^\mathrm{XC}(\boldsymbol{r})n(\boldsymbol{r})\mathrm{d}\boldsymbol{r} + E^\mathrm{XC}[n] \qquad . \tag{3.42}$$

However, the problem that remains is that the exact expression for the exchange-correlation potential is unknown and approximations have to be found.

3.5.2 Approximations to the exchange-correlation functional

Since the beginnings of Kohn-Sham DFT there has been a great effort to find the best approximation to the exchange-correlation functional. Perdew classified this zoo of functionals in a Jacob's ladder picture[180], where the Hartree approach is located on earth and the functional that yields results with chemical accuracy (i.e., with errors that are smaller than ≈1 kcal/mol or 43 meV) placed in heaven. The different rungs of the ladder are constituted by functionals

with increasing levels of complexity. However, one should bear in mind that a higher level of complexity does not necessarily imply a more accurate description. There can be functionals on a higher rung that perform worse for a certain set of systems than a functional on a lower rung.

The functionals on one rung itself can be divided into two philosophical directions: non-empirical and empirical. In the former, free parameters are chosen in a way that satisfies physical constraints, while in the latter the free parameters are fitted to reference data.

3.5.2.1 LOCAL-DENSITY APPROXIMATION

The first rung of Perdew's ladder is the local-density approximation (LDA), which was already proposed in the original paper by Kohn and Sham[179]. The idea behind this approximation is to reduce the problem of the unknown exchange-correlation functional to a well-known model system, the homogeneous electron gas (HEG), as a starting point. One can rewrite the exchange-correlation (XC) energy

$$E^{\mathrm{XC}}[n(\boldsymbol{r})] = \int \varepsilon^{\mathrm{XC}}[n]\, n(\boldsymbol{r}) \mathrm{d}\boldsymbol{r} \quad , \tag{3.43}$$

in terms of the energy density per particle $\varepsilon^{\mathrm{XC}}[n]$.

In the LDA the system is divided into bins, where the electron density is assumed to be constant. In all bins the exchange-correlation energy density is then taken to be the energy density of the homogeneous electron gas with the corresponding electron density. In the limit of infinitesimal small bins this reads:

$$E^{\mathrm{XC}}_{\mathrm{LDA}}[n(\boldsymbol{r})] = \int \varepsilon^{\mathrm{XC}}_{\mathrm{HEG}}(n)\, n(\boldsymbol{r}) \mathrm{d}\boldsymbol{r} \quad . \tag{3.44}$$

The exchange-correlation energy of the homogeneous electron gas can be divided into an exchange and a correlation part:

$$\varepsilon^{\mathrm{XC}}_{\mathrm{HEG}} = \varepsilon^{\mathrm{C}}_{\mathrm{HEG}} + \varepsilon^{\mathrm{X}}_{\mathrm{HEG}} \quad . \tag{3.45}$$

The exchange part is known analytically[181, 182] with $\varepsilon^{\mathrm{X}}_{\mathrm{HEG}}[n] \propto n^{1/3}$. For the correlation part the low-density limit was determined by Wigner[183] and the high-density limit was determined by Gell-Mann and Brueckner[184]. For the intermediate range, very accurate quantum Monte Carlo data exists from Ceperley and Alder [185]. Different types of LDA parametrizations basically differ in the way how this data was interpolated. The most common parametrizations are the Perdew-Zunger[186], Perdew-Wang[187], and Vosko-Wilk-Nusair[188] forms of LDA.

LDA generally performs well for "well-behaved" solids, i.e., covalently bonded, ionic or metallic systems. This is partly due to an error-cancellation effect: While the exchange energy is typically underestimated, the correlation energy is typically overestimated in LDA[181]. Generally, LDA tends to overbind, i.e., lattice parameters are too small and cohesive energies are too large[136]. Approximating the exchange-correlation energy at each point in space by the exchange-correlation energy of a homogeneous electron gas presumes that the density of the system varies only slowly. For systems with rapidly varying densities, i.e., for instance molecules or atoms, LDA fails to correctly describe their properties. The next generation of exchange-correlation functionals, which occupy the second rung of Perdew's Jacob's ladder,

tries to account for this.

3.5.2.2 Generalized gradient approximation

The most evident way to improve on LDA is to take into account the information of the inhomogeneity of the density. The first effort in that direction was the gradient-expansion approximation (GEA) usually written as

$$\varepsilon^{\mathrm{XC}} = \varepsilon^{\mathrm{XC}}_{\mathrm{HEG}}[n(\boldsymbol{r})] \cdot F^{\mathrm{XC}}[n(\boldsymbol{r}), \nabla n(\boldsymbol{r}), \dots] \quad , \tag{3.46}$$

where F^{XC} is the so-called *enhancement factor*, which represents an expansion in terms of the gradient and higher-order derivatives. However, high-order expansion coefficients are hard to calculate and lower-order expansions were found to not necessarily improve on LDA. On the contrary, they often yield worse results because they violate physical constraints [136, 189–193]. It turned out that choosing a general type of F^{XC} as a function of $n(\boldsymbol{r})$ and $\nabla n(\boldsymbol{r})$ instead of a Taylor expansion yields much better results[136]. These types of approximations are referred to as generalized gradient approximations (GGAs). The GGA-type functionals occupy the second rung of Perdew's Jacob's ladder. The generalized form of F^{XC} offers flexibility for parametrization. Correspondingly, there exist many different GGA functionals. The first GGA functional was proposed in 1981 by Langreth and Mehl[194]. In the physics community the most widely used GGA is the PBE functional[15], which is named after its developers Perdew, Burke, and Ernzerhof. Its parameters were determined by imposing physical constraints, i.e., it is non-empirical. On the other hand, B88, named after its developer Becke and the year when it was proposed (1988)[195], is a formulation of the exchange energy where the parameters were fitted to Hartree-Fock calculations. It is very often used together with the correlation energy proposed by Lee, Yang, and Parr (LYP), which then goes by the name BLYP[196].

Due to the dependence on the gradient of the density $\nabla n(\boldsymbol{r})$, GGA functionals are often referred to as *semi-local*. In general, they make up for many of the deficiencies of LDA. Cohesive energies, atomization energies, and lattice parameters are improved, where the latter are typically slightly overestimated. Additionally, GGAs yield better results for the energetics of hydrogen-bonded systems[197, 198].

3.5.2.3 Hybrid functionals

In Hartree-Fock theory, the exchange energy is described exactly and the self-interaction is completely cancelled. However, the Hartree-Fock approach lacks the correlation energy and adding correlation corrections on top of Hartree Fock (see Section 3.4) has an unfavorable scaling with system size. On the other hand, one critical point of the semi-local exchange-correlation functionals in DFT is the insufficient cancellation of the self-interaction error. It is thus a promising route to couple Hartree Fock theory and density-functional approximations in order to reduce the self-interaction error present in the latter[199]. Those efforts result in the class of the so-called hybrid functionals, where a certain fraction of exact exchange in the spirit of Hartree-Fock is mixed into the exchange-correlation functional. In contrast to Hartree-Fock theory, where the Hartree-Fock orbitals are used, in DFT the exact-exchange (EX) integral (see Eq. 3.16) is written in terms of the Kohn-Sham orbitals. Becke[199] showed that the construction principle of hybrid functionals can be rationalized based on the adiabatic-connection formula[200–203],

which gives an exact description for the exchange-correlation energy. Approximating the latter yielded the ansatz[204]:

$$E_{\text{hybrid}}^{\text{XC}} = \alpha E^{\text{EX}} + (1-\alpha)E_{\text{DFA}}^{\text{X}} + E_{\text{DFA}}^{\text{C}} \quad , \tag{3.47}$$

where DFA stands for density-functional approximation. The best choice for α should depend on the system. However, the best single value for most molecules was determined to be $\alpha = 0.25$ based on perturbation-theory considerations[204]. When using the PBE functional[15] in Eq. 3.47, one arrives at the PBE0 hybrid functional[205, 206].

The DFT workhorse hybrid functional in the chemistry community is B3LYP[207, 208], where the exchange-correlation energy is expressed as follows:

$$E_{\text{B3LYP}}^{\text{XC}} = \alpha_0 E^{\text{EX}} + (1-\alpha_0)E_{\text{LDA}}^{\text{X}} + \alpha_1(E_{\text{B88}}^{\text{X}} - E_{\text{LDA}}^{\text{X}}) + (1-\alpha_2)E_{\text{VWN}}^{\text{C}} + \alpha_2 E_{\text{LYP}}^{\text{C}} \quad . \tag{3.48}$$

The parameters $\alpha_0 = 0.20$, $\alpha_1 = 0.72$, and $\alpha_2 = 0.81$ were determined by fitting to a database of atomization energies, proton affinities, ionization potentials, and total atomic energies[207, 209, 210]. $E_{\text{B88}}^{\text{X}}$ denotes the exchange energy functional proposed by Becke in 1988[195], $E_{\text{LYP}}^{\text{C}}$ is the GGA-type correlation energy by Lee, Yang, and Parr[196], and $E_{\text{VWN}}^{\text{C}}$ is the parametrization of the LDA correlation energy by Vosko, Wilk, and Nusair[188].

3.5.3 Dispersion corrections to the XC-functional approximations

Van der Waals (vdW)[5] or dispersion interactions are ubiquitous in nature and, as discussed in Section 2.1, it is crucial to take them into account when studying peptides and proteins[211]. They arise through quantum-mechanical fluctuations in the electron density of the atoms, which lead to instantaneous dipole moments (and higher-order multipole moments). The instantaneous multipole moments on one atom induce multipole moments on a second atom in turn. The interaction of these multipoles results in an attractive force, known as London dispersion forces in honor of Fritz London[40]. This is the definition of van der Waals interactions that is common in physics and, as mentioned in Section 2.1, which we will use here. One has to bear in mind, though, that in chemistry the term "van der Waals" forces refers not only to the dispersion interactions, but includes also permanent dipole-dipole interactions and permanent dipole-induced dipole interactions.

In a classical picture, the electric field of a dipole $\boldsymbol{p}_1$ decreases with $\boldsymbol{E}_1 \propto \boldsymbol{p}_1/R^3$, i.e., the dipole moment of an induced dipole in an atom at position R is $\boldsymbol{p}_2 = \alpha \boldsymbol{E_1} \propto \alpha \boldsymbol{p}_1/R^3$. The potential energy of the first dipole in the field of the second dipole, i.e., the interaction energy between the two dipoles, then reads $E = -\boldsymbol{p}_1 \cdot \boldsymbol{E}_2 \propto -\alpha p_1^2/R^6$. The magnitude of the interaction depends on the polarizability α of the atoms, which is the proportionality factor between the induced dipole moment of an atom and the external field causing the dipole moment.

In fact, the term proportional to R^{-6} is the first term in an expansion of the dispersion energy

[5] Named after the Dutch physicist Johannes Diderik van der Waals.

between two atoms in terms of the interatomic distance R[40, 212][6]

$$E_{\text{disp}} = -\sum_{i=6,8,10,\ldots}^{\infty} \frac{C_i}{R^i} \quad , \tag{3.49}$$

where C_i denote the dispersion coefficients. In practice, often only the first term is kept, which is (normally) the dominant term and determines the long-range behavior. In the correct quantum-electrodynamical description, the polarizability becomes frequency dependent and the C_6^{AB} coefficient for the dispersion interaction between two atoms A and B reads[213]:

$$C_6^{AB} = \frac{3}{\pi} \int_0^{\infty} \alpha_A(i\omega)\alpha_B(i\omega)\mathrm{d}\omega \quad . \tag{3.50}$$

The accuracy of DFT with respect to the description of dispersion interactions depends greatly on the XC functional. Along with the tendency of LDA to overbind, LDA yields too small binding distances and too high binding energies for van der Waals-bonded systems, while semi-local and hybrid functionals tend to yield purely repulsive behavior or at least underbind[214–218]. Despite differences in the exact description and performance, it is well established that present-day semi-local functionals cannot describe the R^{-6} decay of the long-range tail of the dispersion interactions correctly[211].

There are several approaches to make up for this deficiency. Various (hybrid) meta-GGAs (where besides gradient corrections also the kinetic-energy density is considered) developed in the group of Truhlar were parametrized to implicitly account for dispersion interactions[219–222]. They partially depend on more than 30 parameters. However, while they partly mimic dispersion interactions for small separation distances between the atoms, they miss the correct description of the long-range tail[223]. Another approach pursued by Langreth, Lundqvist and co-workers is to directly construct a non-local correlation energy functional, known as vdW-DF[224] and an improved version vdW-DF2[225].

A widespread approach to correct present approximations to the exchange-correlation functional for the long-range tail of van der Waals interactions is to use pairwise approaches of the form[16, 211, 226–235]:

$$E_{\text{disp}} = -\frac{1}{2} s \sum_{A,B} f_{\text{damp}}(R_{AB}, R_A^0, R_B^0) \frac{C_6^{AB}}{R_{AB}^6} \quad , \tag{3.51}$$

where the energy correction E_{disp} is then added to the DFT energy in an *a posteriori* fashion. The sum runs over all atom pairs AB and the factor $1/2$ corrects for double counting. R_{AB} is the distance between the atoms and R_A^0 and R_B^0 are the van der Waals radii of the corresponding atoms. Sometimes an overall scaling factor s is also used. In most cases, the damping function $f_{\text{damp}}(R_{AB}, R_A^0, R_B^0)$ is chosen such that the expression goes to zero for small R in order to avoid singularities of R^{-6} and to match the long-range vdW interaction with the short-range contributions in the functional. Different schemes mostly differ in the shape of the damping function and the way to determine the C_6^{AB} coefficients.

Most of the methods employing pairwise corrections are purely empirical and rely on fixed C_6 coefficients irrespective of the environment of the atoms[226–230, 234], with the DFT-D2 scheme

[6] Fritz London described the dispersion energy based on second-order perturbation theory, using a multipole expansion for the perturbation potential.

of Grimme[230] being one of the most widely-used approaches. However, the environment of an atom crucially influences its polarizability. The C_6 coefficient for carbon, for instance, can vary by as much as about 50% for different hybridization states sp, sp^2, and sp^3[227]. Recently, Grimme and co-workers have suggested a new scheme called DFT-D3[231], where atom-pairwise C_6 coefficients are calculated based on first principles (time-dependent DFT). They take system dependency into account by applying a concept of (fractional) occupation numbers[231]. However, the dispersion coefficients do not depend on the electronic structure. In this thesis, we use a method, where the C_6^{AB} coefficients explicitly depend on the electronic density. It was proposed by Tkatchenko and Scheffler in 2009[16] and we will refer to it as the TS scheme. A short account on this method is given in the following.

3.5.3.1 TS SCHEME

Starting from the Casimir-Polder integral (Eq. 3.50) one can derive[16]:

$$C_6^{AB} = \frac{2C_6^{AA}C_6^{BB}}{\frac{\alpha_B^0}{\alpha_A^0}C_6^{AA} + \frac{\alpha_A^0}{\alpha_B^0}C_6^{BB}} \quad , \tag{3.52}$$

where α_A^0 and α_B^0 are the static polarizabilities of atoms A and B, respectively. Based on this formula, heteronuclear C_6^{AB} coefficients can be calculated from the knowledge of their homonuclear counterparts.

The effective coefficients $C_6^{\mathrm{eff},AB}$ in a specific environment are calculated based on the values for the free atoms $C_6^{\mathrm{free},AA}$ via

$$\begin{aligned} C_6^{\mathrm{eff},AA} &= \left(\frac{V^{\mathrm{eff},A}}{V^{\mathrm{free},A}}\right)^2 C_6^{\mathrm{free},AA} \\ &= \left(\frac{\int r^3 n^{\mathrm{eff},A}(\boldsymbol{r})\mathrm{d}\boldsymbol{r}}{\int r^3 n^{\mathrm{free},A}(\boldsymbol{r})\mathrm{d}\boldsymbol{r}}\right)^2 C_6^{\mathrm{free},AA} \quad . \end{aligned} \tag{3.53}$$

The $C_6^{\mathrm{free},AA}$ coefficients and the static polarizabilities of free atoms are taken from the database of Chu and Dalgarno[236]. The effective density $n^{\mathrm{eff},A}(\boldsymbol{r})$ is obtained through Hirshfeld partitioning[237]:

$$n^{A,\mathrm{eff}}(\boldsymbol{r}) = n(\boldsymbol{r})\frac{n^{A,\mathrm{free}}(\boldsymbol{r})}{\sum_B n^{B,\mathrm{free}}(\boldsymbol{r})} \quad , \tag{3.54}$$

where the sum runs over all atoms B in the molecule. For the evaluation of the van der Waals energy, Eq 3.51 is used (with no overall scaling factor s). The damping function employed takes the form:

$$f_{\mathrm{damp}}(R_{AB}, R_{AB}^0) = \frac{1}{1+\exp\left[-d\left(\frac{R_{AB}}{s_R R_{AB}^0} - 1\right)\right]} \quad . \tag{3.55}$$

As mentioned earlier, $R_{A/B}^0$ are the van der Waals radii with $R_{AB}^0 = R_A^0 + R_B^0$. The effective van der Waals radius of an atom in a molecule can be obtained from its free-atom van der Waals radius via:

$$R^{0,\mathrm{eff}} = \left(\frac{V^{\mathrm{eff}}}{V^{\mathrm{free}}}\right)^{1/3} R^{0,\mathrm{free}} \quad . \tag{3.56}$$

The parameter d was set to 20 as it was found to have only a minor influence on the results in the range between 12 and 45. The parameter s_R controls the distance R_{AB}, at which the damping

function approaches zero, and hence defines the onset of the dispersion correction. The best value of s_R thus depends on the functional that is employed. It was determined for several exchange-correlation functionals using the S22 database[173].[7] The latter contains the accurate binding energies of 22 non-covalently bonded dimers based on CCSD(T) calculations extrapolated to the complete basis set limit. The 22 dimers are sorted into groups with predominant vdW-bonded character, predominant H-bonded character and "mixed" complexes. Recently, Marom *et al.*[223] assessed the performance of several XC functionals for these dimers with and without including vdW interactions based on the TS scheme. For all functionals tested, the inclusion of vdW interactions (TS scheme) improved the mean absolute error (MAE) to CCSD(T) for all three groups of dimers. The performance of the PBE functional[15] coupled to the TS-scheme (PBE+vdW) was explicitly benchmarked against CCSD(T) energy differences for 32 conformations of the alanine di- and tetrapeptide (see supplementary information of Ref. [44] and Ref. [17]). It could be shown that the PBE+vdW functional yields very good results for such systems, with a MAE of only 18 meV for the tested conformers.

3.5.3.2 Many-body van der Waals interactions

Obviously, pairwise schemes, such as the TS method, lack a description of non-additive many-body effects that go beyond the pairwise contributions. As mentioned earlier, the influence of the local environment on the polarizabilities is taken into account in the TS scheme by involving the ground-state electronic density through Hirshfeld partitioning. However, the polarizability of an atom is also influenced by the fluctuating dipoles originating at atom sites located at larger distances (electrostatic screening). Recently, Tkatchenko and co-workers[238, 239] proposed a method, here referred to as MBD@rsSCS or MBD* for short, that accounts both for many-body dispersion contributions and screening effects. This is achieved by modelling the atoms in the molecule as a collection of spherical quantum harmonic oscillators (QHOs), which are coupled to each other via dipole-dipole interactions [coupled fluctuating-dipole model (CFDM)[240]]. The Hamiltonian for this model system reads[238]:

$$\mathcal{H} = -\frac{1}{2}\sum_{p=1}^{N_{\mathrm{at}}} \nabla^2_{\boldsymbol{\chi}_p} + \frac{1}{2}\sum_{p=1}^{N_{\mathrm{at}}} \omega_p^2 \chi_p^2 + \sum_{p>q}^{N_{\mathrm{at}}} \omega_p \omega_q \sqrt{\alpha_p \alpha_q}\, \boldsymbol{\chi}_p \mathcal{T}_{pq} \boldsymbol{\chi}_q \quad , \tag{3.57}$$

where $\boldsymbol{\chi}_q = \sqrt{m_q}\boldsymbol{\xi}_q$ with $\boldsymbol{\xi}_q$ describing the displacement of the QHO q from equilibrium and $m_q = 1/[\alpha_q \omega_q^2]$. The key ingredients are the characteristic excitation frequencies ω_p, the polarizabilities α_p, and $\mathcal{T}_{pq}$, a dipole-dipole interaction tensor, which we will address in more detail below. After diagonalizing the Hamiltonian, the many-body dispersion (MBD) energy can be obtained via:

$$E_{\mathrm{MBD}} = \frac{1}{2}\sum_{i=1}^{3N_{\mathrm{at}}} \sqrt{\lambda_i} - \frac{3}{2}\sum_{p=1}^{N_{\mathrm{at}}} \omega_p \quad , \tag{3.58}$$

where λ_i denote the eigenvalues of the Hamiltonian.

As mentioned earlier in this chapter, the adiabatic-connection fluctuation-dissipation (ACFD) theorem[200–203, 241] gives an exact expression for the exchange-correlation energy. One of the most popular approximations to evaluate the correlation energy in this framework is the random-phase approximation (RPA)[242]. In fact, in can be shown that for the model system of

[7] For PBE s_R is 0.94.

QHOs coupled via a dipole-dipole potential the correlation energy of ACFD-RPA corresponds to the energy expression in Eq. 3.58[243]. From this, we can see that the Hamiltonian in Eq. 3.57 captures screening effects as well as many-body energy contributions.

Local or semi-local DFT exchange-correlation functionals already efficiently account for short-range correlation. In order not to double count short-range correlation, in the MBD@rsSCS method a range-separation approach is used ("rs" stands for range separated). This is realized by range-separating the dipole-dipole interaction tensor $\mathcal{T}$ into a long-range part $\mathcal{T}_{\mathrm{LR}}$ and a short-range part $\mathcal{T}_{\mathrm{SR}}$, where in the many-body Hamiltonian (Eq. 3.57) only the long-range part is employed. In this way, the many-body Hamiltonian will include long-range screening, but lack short-range screening effects. To account also for short-range screening effects, short-range screened polarizabilities $\alpha_p^{\mathrm{rsSCS}}$ (and characteristic excitation frequencies) are obtained, which are then used as input in the many-body Hamiltonian. This is done by again modelling each atom in the molecule as a spherical QHO and employing the self-consistent screening (SCS) equations from classical electrodynamics[244–246]:

$$\alpha_p^{\mathrm{rsSCS}}(i\omega) = \alpha_p^{\mathrm{TS}}(i\omega) + \alpha_p^{\mathrm{TS}}(i\omega) \sum_{q \neq p}^{N_{\mathrm{at}}} \mathcal{T}_{\mathrm{SR},pq} \alpha_q^{\mathrm{rsSCS}}(i\omega) \qquad , \tag{3.59}$$

where $\alpha_p^{\mathrm{TS}}(i\omega)$ denotes the frequency-dependent polarizability obtained from the TS scheme, which already accounts for hybridization effects[16]. The positions of the atoms (QHOs) are denoted by $\boldsymbol{r}_q$ and $\boldsymbol{r}_p$ with $r_{pq} = |\boldsymbol{r}_p - \boldsymbol{r}_q|$. By employing a short-range only dipole-dipole interaction tensor $\mathcal{T}_{\mathrm{SR}}$, the polarizablities $\alpha_p^{\mathrm{rsSCS}}(i\omega)$ capture only short-range screening. The characteristic excitation frequencies $\omega_p^{\mathrm{rsSCS}}$ are also obtained from the SCS equations described above.[8]

The short-range part of the dipole-dipole interaction tensor is given by

$$\mathcal{T}_{\mathrm{SR},pq} = (1 - f(r_{pq}))\, \mathcal{T}_{pq} \qquad , \tag{3.60}$$

where the dipole-dipole interaction tensor is defined as $\mathcal{T}_{pq} = \nabla_{\boldsymbol{r}_q} \bullet \nabla_{\boldsymbol{r}_p} W(r_{pq})$. $W(r_{pq}) = \mathrm{erf}[r_{pq}/(\sqrt{2}R)]/r_{pq}$ is the Coulomb potential for the interaction of two spherical Gaussian charge distributions at distance r_{pq}, where $R = \sqrt{R_p^2 + R_q^2}$ with $R_p = (\sqrt{2/\pi}\alpha_p^{\mathrm{TS}}/3)^{1/3}$ being the width of the Gaussian function. The function $f(r_{pq})$ is the Fermi-type damping function as used also in the TS approach (see Eq. 3.55). The parameter d in Eq. 3.55 is fixed to 6, while s_R is determined separately for each exchange-correlation functional by minimizing energy differences with respect to the S66×8 database[174]. In principle, $\mathcal{T}_{\mathrm{LR}}$ is defined as $\mathcal{T}_{\mathrm{LR}} = \mathcal{T} - \mathcal{T}_{\mathrm{SR}}$. However, $\mathcal{T}$ is frequency dependent,[9] which is not computationally efficient. As only the long range is described here, one can approximate $\mathcal{T}_{\mathrm{LR}}$ as the product of the damping function and the dipole-dipole interaction tensor of two point dipoles:

$$\mathcal{T}_{\mathrm{LR}} = f(r_{pq}) \frac{-3 r_{pq}^a r_{pq}^b + r_{pq}^2 \delta_{ab}}{r_{pq}^5} \qquad , \tag{3.61}$$

[8] In more detail, the self-consistently screened characteristic excitation frequencies are calculated from the C_6^{rsSCS} coefficients, which are obtained by integrating the Casimir-Polder integral (see Eq. 3.50) using α^{rsSCS} (see Refs. [16, 238]).
[9] The interaction potential is a function of the Gaussian width R, which depends on the polarizability, which is in turn frequency dependent.

where the indices a and b denote the Cartesian components of $\boldsymbol{r}_{pq}$.

The evaluation of the MBD@rsSCS long-range correlation energy can now be summarized in three steps:

1. In the first step, the polarizabilities are obtained in the TS scheme.
2. Then, the short-range (SR) range-separated self-consistently screened polarizabilites α^{rsSCS} are obtained using the SCS procedure defined in Eq. 3.59.
3. Using α^{rsSCS} and the long-range dipole-dipole interaction tensor $\mathcal{T}_{\mathrm{LR}}$ one can then evaluate the many-body long-range correlation energy using Eq. 3.58.

The performance of the MBD@rsSCS method (MBD* for short) coupled with the PBE[15] and PBE0[205, 206] exchange-correlation functionals (PBE+MBD*, PBE0+MBD*) was recently benchmarked for peptides by Rossi and co-workers[247]. Additionally, the TS scheme was assessed as well (PBE+vdW, PBE0+vdW). For the benchmarks, two test cases were addressed. The first one was a set of 73 conformers of three-residue peptides, for which accurate CCSD(T) energy differences exist in the literature[44, 248, 249]. The second test case was the larger and experimentally extensively studied Ac-Phe-Ala_5-Lys(H^+) peptide. For this peptide the presence of four different conformers and tentatively their relative abundances have been experimentally established[250, 251]. For the latter, the conformer-selective infrared-ultraviolet (IR-UV) double resonance technique was used, which will be discussed in more detail in Section 7.1. Turning to the first test case first, conformers of Gly-Phe-Ala (GFA), Gly-Gly-Phe (GGF), Phe-Gly-Gly (FGG)[248], and Ac-Ala_3-NMe[44, 249] were assessed yielding three conclusions[247]: (1) The inclusion of dispersion corrections (both TS and MBD*) improves the performance of both PBE and PBE0. (2) For the peptides that contain a phenylalanine (Phe) residue, the PBE0 functional corrected for dispersion interactions (both TS and MBD*) performs better than the corresponding dispersion-corrected PBE functional, while for Ac-Ala_3-NMe the performance is similar. (3) The performance of the TS scheme and the MBD* method are very similar. However, many-body effects are expected to become more important with increasing system size.

For the second test case, the peptide Ac-Phe-Ala_5-Lys(H^+), Rossi *et al.* found that the PBE0+MBD* functional including zero-point energy corrections comes closest to explaining the experimental findings of all methods tested including a recent study by Xie *et al.*[252], which assessed 19 different semi-local and hybrid DFT exchange-correlation functionals.

This points to PBE0+MBD* being the most reliable functional for the peptide systems considered in this thesis. However, relaxation of 10^3-10^4 structures for systems with 108-220 atoms, as needed for our conformational searches, is not computationally feasible with PBE0 as it involves the calculation of the exchange integral (see Section 3.5.2). Furthermore, the forces for the MBD* correction are only available in a finite-difference approach at present. For these reasons, we employ PBE+vdW for the production calculations in this thesis. However, the limitations and accuracy of PBE+vdW for the specific systems investigated in this work are assessed using targeted calculations with PBE+MBD*, PBE0+vdW, and PBE0+MBD* for selected conformers.

3.6 Numeric atom-centered orbitals: FHI-aims

The central problem of Kohn-Sham DFT is to solve the Kohn-Sham equations (Eq. 3.41). For this, the Kohn-Sham orbitals φ_i are commonly expanded in an appropriate basis set $\{\phi_i\}$:

$$\varphi_i(\boldsymbol{r}) = \sum_j c_{ji}\phi_j(\boldsymbol{r}) \quad , \tag{3.62}$$

where c_{ji} denote the expansion coefficients. With the Hamiltonian $\hat{h}^{\mathrm{KS}} = -1/2\nabla^2 + v^{\mathrm{eff}}$, the Kohn-Sham equations can be written in the form of a generalized eigenvalue problem:

$$\sum_j h_{ij}c_{jl} = \varepsilon_l \sum_j s_{ij}c_{jl} \tag{3.63}$$

with the Hamiltonian matrix elements

$$h_{ij} = \int \varphi_i^*(\boldsymbol{r})\hat{h}^{\mathrm{KS}}\varphi_j(\boldsymbol{r})\mathrm{d}\boldsymbol{r} \tag{3.64}$$

and the matrix elements of the overlap matrix s_{ij}

$$s_{ij} = \int \varphi_i^*(\boldsymbol{r})\varphi_j(\boldsymbol{r})\mathrm{d}\boldsymbol{r} \quad . \tag{3.65}$$

A common choice for the basis functions are plane waves, used for instance in VASP[253] or CASTEP[254]. Another option are localized basis functions. Gaussian-type orbitals, e.g., are a preferred choice due to their convenient analytical properties and are used in a number of programs including NWChem[255] and TURBOMOLE[256]. The FHI-aims program package[257], which is the DFT code used in this thesis, employs numeric atom-centered orbitals (NAOs). FHI-aims is an all-electron/full-potential code that can treat both cluster-type and periodic systems on equal footing. The NAOs take the form:

$$\phi_i(\boldsymbol{r}) = \frac{u_i(r)}{r}Y_{lm}(\Omega) \quad . \tag{3.66}$$

$Y_{lm}(\Omega)$ are the spherical harmonics and $u_i(r)$ is numerically tabulated. The latter is chosen to satisfy Schrödinger-like radial equations:

$$\left[-\frac{1}{2}\frac{d^2}{dr^2} + \frac{l(l+1)}{r^2} + v_i(r) + v_{\mathrm{cut}}(r)\right] u_i(r) = \varepsilon_i u_i(r) \quad , \tag{3.67}$$

which are solved on a dense logarithmic radial grid. The potential $v_i(r)$ defines the shape of $u_i(r)$ and is thus called the defining potential. The free choice of $v_i(r)$ renders $u_i(r)$ very flexible. Specifically, in FHI-aims free-hydrogen like ($v_i(r) = Z^{\mathrm{eff}}/r$ with Z^{eff} denoting the charge) and self-consistent free-atom and free-ion (doubly-positive) radial potentials are used. Gaussian-type functions can be employed as well. This is extremely valuable for direct comparisons with the results of other DFT codes that use Gaussian basis sets. The potential v_{cut} is the so-called cut-off potential, which ensures that $u_i(r)$ is exactly zero beyond a certain cut-off radius r_{cut}. It takes

the form:

$$v_{\mathrm{cut}}(r) = \begin{cases} 0 & r \leqslant r_{\mathrm{onset}} \quad , \\ s \cdot \exp\left(-\frac{w}{r-r_{\mathrm{onset}}}\right) \cdot \frac{1}{(r-r_{\mathrm{cut}})^2} & r_{\mathrm{onset}} < r < r_{\mathrm{cut}} \quad , \\ \infty & r \geqslant r_{\mathrm{cut}} \quad . \end{cases} \tag{3.68}$$

The parameter s denotes a global scaling factor. Beginning with an onset at r_{onset}, the cutoff potential smoothly approaches infinity over a range $w = r_{\mathrm{cut}} - r_{\mathrm{onset}}$.

The solution of the Kohn-Sham eigenvalue problem (Eq. 3.63) requires many numerical integration steps, e.g., Eqs. 3.64 and 3.65. In `FHI-aims`, the integrand is broken down into atom-centered fragments by using atom-centered partition functions[258]. The integration of each fragment is then carried out individually on spherical atom-centered integration shells [257, 259]. The specific choice of the basis functions in `FHI-aims` allows for an efficient scaling of these integrations. As the basis functions vanish exactly beyond a radius r_{cut}, these grid-based operations (e.g., Eqs. 3.64 and 3.65) scale with $O(N)$ in the limit of large system sizes (where N denotes the system size). Another advantage of the basis set choice is that due to the possibility to choose the defining potential to be the potential of spherically-symmetric free atoms, the (spherically-symmetric) free atom can be described exactly with only a small number of basis functions. These are the occupied orbitals of the atom, called the minimal basis. This is beneficial since the orbital shape close to the core does not change much even if the atoms bind and is thus described almost exactly as well.

When using overlapping atom-centered basis functions, the so-called basis set superposition error (BSSE) can arise through the overlap of basis functions centered at different atom sites. When, e.g., considering atomization energies

$$\Delta E^{\mathrm{atm}} = E^{\mathrm{compound}} - \sum_{i}^{N_{\mathrm{at}}} E^{\mathrm{atom},i} \quad , \tag{3.69}$$

the full compound is described by a larger basis set than the individual atoms. This can improve the energy for the full system compared to the free-atom energies, yielding wrong values for ΔE^{atm}. In `FHI-aims`, atomization energies (for DFT and non-spinpolarized spherical atoms) do not suffer from BSSE as the free atoms are described exactly and any further basis functions would not lower their energy. Molecular fragments, though, are not described exactly. However, the fragmentation BSSE, i.e., the BSSE arising by comparing energies of different fragments such as binding energies, is very small for DFT calculations using reasonably converged basis sets[257]. While DFT calculations only concern the occupied orbitals, explicitly correlated methods such as MP2 involve sums over unoccupied states (cf. Eq. 3.27). In principle, they have to be summed up to infinity, which presents a problem when using a necessarily limited basis set. This leads to a slow convergence of energy (differences) with basis-set size and to large BSSEs even when large basis sets are employed. Using the standard `FHI-aims` basis sets, Ren *et al.*[260] found that energy differences (e.g., binding energies) converge reasonably well with increasing basis set size *if* a counterpoise correction is employed. In a counterpoise correction[261], the BSSE is *a posteriori* removed by recalculating the energies of the fragments with the full basis set used to calculate the energy of the whole system. Alternatively, Zhang *et al.* recently constructed NAO basis sets that are suitable to converge total energies when using explicitly correlated methods such as MP2.

For a more detailed description, we will here focus on the standard FHI-aims basis sets. In order to construct suitable and accurate basis sets for all elements in the periodic table, an iterative procedure is used[257]. Starting from the minimal free-atom basis a single basis function from a large pool of "candidate" functions is added. Then the LDA total-energy error for a number of dimers at various separation distances is evaluated. The additional basis function, which gives the largest improvement to the energy, is permanently added to the basis set. This procedure is then repeated. In this way, hierarchical basis sets (for 102 elements in the periodic table) have been constructed, which are organized into different tiers (levels) called *tier1* to *tier4*.

In order to reach convergence of the target properties, not only the basis set has to be chosen sufficiently accurately, but also the other computational parameters have to be set properly. Distributed with the FHI-aims program package are a set of pre-constructed computational defaults, categorized as *light*, *tight*, and *really tight* settings. For each element, these settings define defaults for the size of the basis sets (in terms of the different tiers), but also specify the integration grid and the accuracy of the calculation of the Hartree potential.[10] Within each default setting the basis-set size can be systematically increased or decreased by systematically adding or removing tiers. Light settings allow for an initial assessment of energy hierarchies and geometries, while tight settings should be used for "final" results.

In order to perform geometry relaxations, calculate normal-mode frequencies based on finite differences or perform molecular dynamics (MD) simulations, total-energy derivatives (forces) are needed:

$$\boldsymbol{F}_I(\{\boldsymbol{R}_I\}) = -\frac{\partial}{\partial \boldsymbol{R}_I} V_{\mathrm{BO}}(\{\boldsymbol{R}_I\}) \quad . \tag{3.70}$$

The forces acting on a specific nucleus that originate from the electric field of the electrons and the other nuclei are called Hellman-Feynman forces[262, 263]

$$\boldsymbol{F}_I^{\mathrm{HF}} = \sum_{J \neq I}^{N_{\mathrm{at}}} \frac{Z_I Z_J}{|\boldsymbol{R}_I - \boldsymbol{R}_J|^3}(\boldsymbol{R}_I - \boldsymbol{R}_J) - \int n(\boldsymbol{r}) Z_I \frac{\boldsymbol{R}_I - \boldsymbol{r}}{|\boldsymbol{R}_I - \boldsymbol{r}|^3} \mathrm{d}\boldsymbol{r} \quad . \tag{3.71}$$

Due to the finite, atom-centered basis set and other approximations used in FHI-aims, two classes of correction terms to the Hellman-Feynman forces have to be taken into account. One correction term is arising from the truncation of the multipole expansion of the electronic density used for the calculation of the Hartree potential. Another correction term is due to the dependence of the basis functions on the atomic positions. They 'move' with $\boldsymbol{R}_i$, which gives rise to the so-called Pulay forces[264]. Additionally, for GGAs a further correction term has to be taken into account originating from the derivative of the density gradient. For more details and the exact expressions implemented in FHI-aims, the interested reader is referred to Refs. [257, 265].

[10] The Hartree potential is computed based on a multipole expansion of the electron density. The highest angular momentum that is taken into account in this series determines the accuracy of the calculation and can be set explicitly by the user.

4 EXPLORING ENERGY LANDSCAPES

After having discussed methods for describing the potential-energy surface (PES) in the previous chapter, this chapter is devoted to how molecules actually move on this surface and how it can be sampled. We first give a description of molecular dynamics (MD) simulations and different thermostats and then describe the sampling techniques used in this thesis.

4.1 MOLECULAR-DYNAMICS SIMULATIONS

In a usual MD simulation, the nuclei are treated as classical point particles, which move on the potential-energy surface (PES) $V(\{\boldsymbol{R}_I\})$, where $\{\boldsymbol{R}_I\}$ denote the spatial coordinates of the nuclei (see Section 3.2.1). The formulas discussed in the following are general with respect to the nature of $V(\{\boldsymbol{R}_I\})$. It can be the Born-Oppenheimer (BO) PES, but it can also refer to any empirical energy function such as a force field.

The motion of the nuclei follows Newton's equations of motion:

$$M_I \ddot{\boldsymbol{R}}_I = -\nabla_I V(\{\boldsymbol{R}_I\}) = \boldsymbol{F}_I \quad , \tag{4.1}$$

where M_I is the mass of nucleus I and $\boldsymbol{F}_I$ is the force acting on nucleus I. In order to obtain the trajectory that the nuclei follow, one has to integrate these equations. Starting from a position $\boldsymbol{R}_I(t)$ for nucleus I at time t, the position of this nucleus at time $t+\Delta t$ can be written as a Taylor expansion in terms of Δt[266]:

$$\begin{aligned} \boldsymbol{R}_I(t+\Delta t) &= \boldsymbol{R}_I(t) + \dot{\boldsymbol{R}}_I(t)\Delta t + \frac{1}{2}\ddot{\boldsymbol{R}}_I(t)\Delta t^2 + \frac{1}{3!}\dddot{\boldsymbol{R}}_I(t)\Delta t^3 + \mathcal{O}(\Delta t^4) \\ &= \boldsymbol{R}_I(t) + \boldsymbol{v}_I(t)\Delta t + \frac{1}{2M_I}\boldsymbol{F}_I(t)\Delta t^2 + \frac{1}{3!}\dddot{\boldsymbol{R}}_I(t)\Delta t^3 + \mathcal{O}(\Delta t^4) \quad . \end{aligned} \tag{4.2}$$

Truncating this expansion after the second order yields the Euler algorithm. However, due to large errors ($\mathcal{O}[\Delta t^3]$) it is not used in practice for MD simulations. A better way is to write down the analogous expansion for $\boldsymbol{R}_I(t-\Delta t)$:

$$\begin{aligned} \boldsymbol{R}_I(t-\Delta t) &= \boldsymbol{R}_I(t) - \dot{\boldsymbol{R}}_I(t)\Delta t + \frac{1}{2}\ddot{\boldsymbol{R}}_I(t)\Delta t^2 - \frac{1}{3!}\dddot{\boldsymbol{R}}_I(t)\Delta t^3 + \mathcal{O}(\Delta t^4) \\ &= \boldsymbol{R}_I(t) - \boldsymbol{v}_I(t)\Delta t + \frac{1}{2M_I}\boldsymbol{F}_I(t)\Delta t^2 - \frac{1}{3!}\dddot{\boldsymbol{R}}_I(t)\Delta t^3 + \mathcal{O}(\Delta t^4) \quad , \end{aligned} \tag{4.3}$$

and to sum up both equations

$$\begin{aligned} \boldsymbol{R}_I(t+\Delta t) + \boldsymbol{R}_I(t-\Delta t) &= 2\boldsymbol{R}_I(t) + \frac{1}{M_I}\boldsymbol{F}_I(t)\Delta t^2 + \mathcal{O}(\Delta t^4) \\ \Leftrightarrow \boldsymbol{R}_I(t+\Delta t) &= 2\boldsymbol{R}_I(t) - \boldsymbol{R}_I(t-\Delta t) + \frac{1}{M_I}\boldsymbol{F}_I(t)\Delta t^2 + \mathcal{O}(\Delta t^4) \quad . \end{aligned} \tag{4.4}$$

This yields the Verlet algorithm, where the error for $\boldsymbol{R}_I(t+\Delta t)$ goes with $\mathcal{O}(\Delta t^4)$. In this scheme, the velocities are not needed to calculate the new positions of the nuclei. However, they can be evaluated by subtracting Eq. 4.2 and Eq. 4.3 from each other:

$$\begin{aligned} \boldsymbol{R}_I(t+\Delta t) - \boldsymbol{R}_I(t-\Delta t) &= 2\boldsymbol{v}_I(t)\Delta t + \mathcal{O}(\Delta t^3) \\ \Leftrightarrow \boldsymbol{v}_I(t) &= \frac{\boldsymbol{R}_I(t+\Delta t) - \boldsymbol{R}_I(t-\Delta t)}{2\Delta t} + \mathcal{O}(\Delta t^2) \quad . \end{aligned} \tag{4.5}$$

In order to evaluate $\boldsymbol{R}_I(t+\Delta t)$ and $\boldsymbol{v}_I(t)$ in the Verlet algorithm, one needs to know the positions of the nuclei at $t-\Delta t$. However, this is not known for the starting point $t = t_0$. An alternative algorithm is the "velocity Verlet algorithm"[267], which is the one implemented in `FHI-aims`. It is equivalent to the standard Verlet algorithm (see, e.g., Ref. [266]), but it does not need any knowledge about times $t-\Delta t$ in order to obtain the positions $\boldsymbol{R}_I$ and velocities at time $t+\Delta t$. The update of the coordinates is the same as in the Euler algorithm, i.e.,

$$\boldsymbol{R}_I(t+\Delta t) = \boldsymbol{R}_I(t) + \boldsymbol{v}_I(t)\Delta t + \frac{1}{2M_I}\boldsymbol{F}_I(t)\Delta t^2 \tag{4.6}$$

and the velocities are calculated via:

$$\boldsymbol{v}_I(t+\Delta t) = \boldsymbol{v}_I(t) + \frac{\boldsymbol{F}_I(t+\Delta t) + \boldsymbol{F}_I(t)}{2M_I}\Delta t \quad . \tag{4.7}$$

Both Verlet algorithms are time reversible, i.e., when changing the time increment from Δt to $-\Delta t$ the trajectory is traced backward in time. In contrast, the Euler algorithm is not time reversible. It only becomes time reversible in the limit of an infinitesimal small time step.

The choice of the time step is crucial for the accuracy of the simulations, where the largest reasonable value is limited by the vibration with the largest oscillation period (or highest frequency). Thus, the lighter the atoms the system contains, the smaller the time step generally has to be chosen. For molecules that contain hydrogen atoms, as it is the case for the systems considered in this thesis, the time step has to be of the order of $\Delta t = 1$ fs. We will discuss this, as well as the many other practical issues determining the accuracy of an (*ab initio*) MD simulation, in more detail in Chapter 6.

When evolving the trajectories based on Newton's equations of motion, the energy and the momenta are conserved (apart from numerical inaccuracies). This corresponds to a simulation in the microcanonical ensemble (NVE), where also the number of particles N stays constant and the volume V – if it is possible to define one – is kept fixed. However, experiments are often performed under conditions, where not the energy, but different thermodynamic variables such as the temperature T or the pressure p are held constant. This corresponds to other statistical ensembles. The ensemble where the number of particles N, the volume V and the temperature T do not change (NVT) is known as the canonical ensemble. In the isothermal-isobaric ensemble (NPT) N, T, and the pressure P are constant. In order to carry out simulations

in these ensembles the system has to be coupled to a (heat or/and a pressure) bath. Methods for performing simulations in the canonical ensemble will be discussed in the following section.

4.1.1 Molecular-dynamics simulations in the canonical ensemble

Evolving the trajectories of the particles of the system by integrating Newton's equations of motion corresponds to a simulation in the microcanonical ensemble. In this subsection, we shall describe several methods (called thermostats) that enable MD simulations in the canonical ensemble, i.e., a simulation where the temperature T and the number of particles N do not change. Here we focus on the thermostats used in this thesis, while a more detailed account can be found in textbooks such as Ref. [266].

According to the equipartition theorem[268] the kinetic energy in a canonical ensemble is equally distributed over the momentum coordinates, each taking on average $k_{\mathrm{B}}T/2$. For the average kinetic energy[1] $\langle K \rangle$ of the molecule, it thus follows

$$\frac{3}{2} N_{\mathrm{at}} k_{\mathrm{B}} T = \left\langle \sum_{I}^{N_{\mathrm{at}}} \frac{M_I}{2} \boldsymbol{v}_I^2 \right\rangle = \langle K \rangle \qquad . \tag{4.8}$$

This relation can be used to calculate the instantaneous kinetic temperature T_K during a simulation by

$$T_K = \frac{1}{3 N_{\mathrm{at}} k_{\mathrm{B}}} \sum_{I}^{N_{\mathrm{at}}} M_I \boldsymbol{v}_I^2 \qquad , \tag{4.9}$$

where the probability density for the velocity $\boldsymbol{v}_I$ of particle I is described by the Maxwell-Boltzmann distribution:

$$P(\boldsymbol{v}_I) = \left(\frac{M_I}{2\pi k_{\mathrm{B}} T} \right)^{3/2} \exp\left(-\frac{M_I \boldsymbol{v}_I^2}{2 k_B T} \right) \qquad . \tag{4.10}$$

The instantaneous kinetic temperature fluctuates during a simulation in the canonical ensemble. Methods that maintain the instantaneous kinetic temperature as a constant during the simulation or the simpler velocity rescaling methods, such as the Berendsen thermostat[266, 269, 270], are not able to sample a canonical distribution.

One approach that generates a canonical ensemble is the Andersen thermostat[266, 271]. Here, the nuclei undergo stochastic collisions with the heat bath. In practice, this consists of three repeating steps.

1. The trajectories of all particles are evolved in time for Δt by integrating Newton's equations of motion.

2. Particles that are to undergo a collision are chosen. The probability for a particle to be selected is $\nu \cdot \Delta t$, where ν is the collision frequency, which is set by the user.

3. A new velocity is assigned to each particle that was selected to undergo a collision with the heat bath. The new velocity is drawn from a Maxwell-Boltzmann distribution at the target temperature T. All other particles remain unaffected. Within the next time interval all particles are evolved in time by integrating Newton's equations of motion again.

[1] Here we denote the kinetic energy with K in order to avoid confusion with the temperature T.

It can be shown that this procedure in fact generates a canonical distribution[271]. Unlike simulations in the *NVE* ensemble, which conserve the total energy, simulations with the Andersen thermostat do not have a conserved quantity. Such a conserved quantity is beneficial as it can be used to monitor the accuracy of the simulation (and check, e.g., if an appropriate time step was chosen).

One approach for performing deterministic and time-reversible simulations in the canonical ensemble is the method known as the Nosé-Hoover thermostat[272, 273]. In this approach, the Lagrangian (or the Hamiltonian) of the system is extended by introducing an additional degree of freedom. The equations of motion (in Hoover's formulation) read[272–274]:

$$\dot{\boldsymbol{R}}_I = \frac{\boldsymbol{p}_I}{M_I} \tag{4.11}$$

$$\dot{\boldsymbol{p}}_I = -\nabla_I V(\{\boldsymbol{R}_I\}) - \frac{\boldsymbol{p}_I p_\eta}{Q} \tag{4.12}$$

$$\dot{p}_\eta = \sum_{I=1}^{N_\mathrm{at}} \frac{\boldsymbol{p}_I^2}{M_I} - 3N_\mathrm{at} k_\mathrm{B} T \tag{4.13}$$

$$\dot{\eta} = \frac{p_\eta}{Q} \quad . \tag{4.14}$$

M_I is the mass of nucleus I and N_at is the number of atoms. The coordinate η is a fictitious degree of freedom and p_η denotes its conjugated momentum. Q can be understood as an effective mass associated with a fictitious oscillator, which couples to the system. It acts as a (negative or positive) friction to the momenta $\boldsymbol{p}_I$. From the Nosé-Hoover equations of motion follows a conserved quantity:

$$\mathcal{H}_\mathrm{NH} = V(\{\boldsymbol{R}_I\}) + \sum_I^{N_\mathrm{at}} \frac{\boldsymbol{p}_I^2}{2M_I} + \frac{p_\eta^2}{2Q} + 3N_\mathrm{at} k_\mathrm{B} T \eta \quad . \tag{4.15}$$

Nosé and Hoover[272, 273] showed that evolving the trajectories of this extended system by integrating its equation of motion samples a canonical ensemble for the original system[272, 273] provided that the dynamics are ergodic. A system is said to be ergodic if, in the limit of infinite time, the average of a quantity over time is equivalent to the average of this quantity over the phase space. In large (ergodic) systems, the Nosé-Hoover approach yields very good results and is one of the most widely used thermostats.

The oscillator mass and its corresponding frequency are related via $Q = 3N_\mathrm{at} k_\mathrm{B} T/\omega^2$. In order for the system to couple efficiently to the thermostat, the frequency of the oscillator should be chosen within a range where the system has (preferably delocalized) vibrational modes[274]. However, if the oscillator is in resonance with a specific localized very harmonic mode of the system, the simulation can get stuck in a small part of the phase space[275]. It has been shown that for too small or too harmonic systems the Nosé-Hoover dynamics suffer from ergodicity problems so that the canonical distribution is not generated[273–275]. A possible solution is to use a chain of Nosé-Hoover thermostats[274].

An alternative approach that does not exhibit the ergodicity problem, is a thermostat suggested by Bussi, Donadio, and Parrinello[276], also referred to as the BDP thermostat. This thermostat uses a stochastic rescaling of the velocities. The algorithm consists of four steps[276]:

1. The trajectories are evolved by integrating Newton's equations of motion.

2. The instantaneous kinetic energy K is calculated.

3. The instantaneous kinetic energy is evolved for one time step Δt using a stochastic dynamics that contains a velocity rescaling term and a Gaussian white-noise term.

4. In the last step, the velocities are rescaled by a factor α in order to match the value of the kinetic energy determined in the previous step.

The square of the rescaling factor α reads:

$$\begin{aligned} \alpha^2 &= \exp(-\Delta t/\tau) + \frac{\overline{K}}{N_f K}[1-\exp(-\Delta t/\tau)]\left(R_1^2 + \sum_{i=2}^{N_f} R_i^2\right) \\ &\quad +2\exp(-\Delta t/[2\tau])\sqrt{\frac{\overline{K}}{N_f K}[1-\exp(-\Delta t/\tau)]}R_1 \quad . \end{aligned} \tag{4.16}$$

The parameters R_i are independent random numbers drawn from a Gaussian distribution with variance 1, τ is the relaxation time of the thermostat and $\overline{K}$ is the average kinetic energy of a Maxwell-Boltzmann distribution at the target temperature. N_f is the number of degrees of freedom of the system (usually $3N_{\mathrm{at}}$). It can be shown that this thermostat samples the canonical ensemble. Remarkably, despite the random numbers involved, there is a quantity that is conserved during the simulation. The efficiency of the thermostat is rather insensitive to the value of the relaxation parameter τ (it should be chosen between 20–50 times the time step used). In this thesis, we used both the Nosé-Hoover and the BDP thermostat for our production runs.

4.2 Sampling techniques

Exploring the energy landscape of peptides is a demanding problem as they are very flexible molecules. As discussed in Chapter 2, there are two backbone torsional degrees of freedom per residue. This leads to an exponential growth of the conformational space with peptide length. Sampling the conformational space of peptides of the size treated in this thesis (108-440 atoms) is practically infeasible using only first-principles methods.[2]

For this reason, we pursue a two-step strategy where we start with a broad sampling of the conformational space using a force field and then follow up with a local refinement using density-functional theory (DFT) (PBE+vdW). Our methods of choice to scan the structure space are basin hopping and replica-exchange molecular dynamics (REMD), which will be discussed in more detail in the following sections. There may be more sophisticated methods (e.g., by introducing biasing potentials and collective variables, see, e.g., Refs. [277–283]), which sample the conformational space more efficiently. However, they introduce a bias to the system, while the beauty of the REMD approach is that it induces no bias at all. The main goal here is to sample the structure space as extensively as possible and find as many structure candidates as possible in the first step that are then relaxed with DFT in the second step. This is to reduce a force-field bias and not to miss any relevant conformation as it is well known that energy hierarchies and energy differences of peptide conformers can deviate between different force fields[13, 14] and also between different flavors of DFT.

[2]Consider, e.g., that one MD time step for our 108 atom system using *tight* computational settings takes about 55 s on 256 cores of the "aims" cluster at the Garching Computing Center (Intel Xeon octacore nodes).

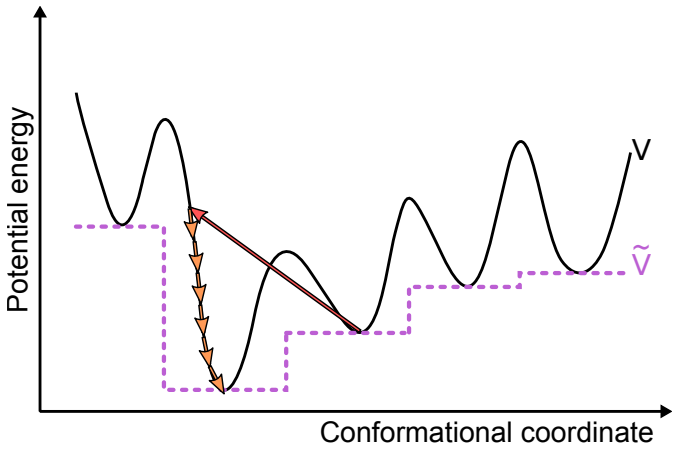


Figure 4.1: Schematic representation of the transformed energy landscape $\tilde{V}(\{\boldsymbol{R}_i\})$ sampled in the basin-hopping algorithm in comparison to the original landscape $V(\{\boldsymbol{R}_i\})$ in one dimension. The long red arrow represents a trial move followed by local geometry optimization (small yellow arrows).

The following subsections give a general explanation of the basin-hopping algorithm and the REMD method, where in Section 4.2.3 a practical example is given. All the details about the specific structure searches performed for the peptides studied in this work are described in Chapters 8 and 10, respectively.

4.2.1 Basin-hopping algorithm

In the basin-hopping approach[130, 284–286] the PES $V(\{\boldsymbol{R}_i\})$ is transformed to

$$\tilde{V}(\{\boldsymbol{R}_i\}) = \min\left[V(\{\boldsymbol{R}_i\})\right] \quad , \tag{4.17}$$

where $\{\boldsymbol{R}_i\}$ denote the positions of the nuclei. For each conformation $\{\boldsymbol{R}_i\}$ a geometry optimization (indicated by "min") is performed, such that the energy of each conformation is mapped onto the energy of the geometry to which it relaxes. In this way, the actual PES is transformed into a set of terraces, as illustrated in Fig. 4.1 for one dimension. Each plateau extends around a local minimum over the range of the basin of attraction, i.e., over the space, where all conformations relax into this specific minimum. Instead of the original PES, in the basin-hopping approach the transformed landscape is sampled. Starting from a specific geometry, the first step is to perform a so-called trial move, in which the coordinates of the initial geometry are perturbed (e.g., by a random displacement of the nuclear positions[130, 284, 287]). This is indicated by the long red arrow in Fig. 4.1. Then local optimization is performed as illustrated by the small yellow arrows. In the classic algorithm, a Metropolis Monte Carlo criterion is employed to judge if the new local minimum is *accepted* or *rejected*. The move is accepted with the probability $\min(1,\exp\{-(V_{\mathrm{new}} - V_{\mathrm{initial}})/[k_{\mathrm{B}}T_{\mathrm{eff}}]\})$, where the effective temperature T_{eff} has to be chosen appropriately. If it is accepted, the next trial move is performed starting from the new geometry. Otherwise another trial move based on the previous geometry is attempted. Other than the Metropolis criterion, acceptance criteria based on a threshold have also been employed[287, 288]. In this approach, all new geometries with $V_{\mathrm{new}} < V_{\mathrm{initial}}$ are accepted. If $V_{\mathrm{new}} > V_{\mathrm{initial}}$, the new conformation is accepted if $V_{\mathrm{new}} - V_{\mathrm{initial}}$ does not exceed a certain cutoff.

The basin-hopping approach effectively removes the barriers associated with transition states between local minima, which enhances inter-basin transitions. In a canonical MD simulation the system can only pass from one local minimum to the next by passing through the transition-state region. In the basin-hopping algorithm, however, the system can pass from one local minimum

to the next via all possible connecting paths, i.e., there are many more possibilities to reach a certain minimum, which again enhances the sampling efficiency. The system can hop directly from one local minimum to the next giving rise to the name of the algorithm – *basin hopping*.

For peptides, a trial move based on a random displacement of the nuclear positions is not the most efficient choice. As changes in bond angles and bond lengths involve relatively large energetic changes, an intuitive way to perform the trial move is to change the torsional angles of the peptide. Here, we use the implementation of the basin-hopping algorithm in the TINKER program package[155]. In this implementation, the Hessian matrix with respect to the torsional degrees of freedom is diagonalized and from that, the torsional eigenvectors are obtained. A search for local minima is carried out using each torsional eigenvector separately in turn starting from the one with the largest eigenvalue, until a user-specified number of torsional eigenvectors have been searched. The trial move is carried out in small steps along (and opposite to) the direction of the chosen torsional eigenvector until the energy in subsequent steps decreases. Then a geometry optimization is performed. In order to decide if the new geometry is accepted or rejected, a threshold approach is used. The geometry is accepted and used as a starting geometry for further searches if the energy of the new conformation lies within an energy window Δ above the lowest-energy conformer that has been found up to this point in the search. The size of the energy window Δ has to be specified by the user with typical values of 25-50 kcal/mol (about 1-2 eV)[28, 289]. In order to determine if the new conformation has been found already, the energy of the latter is compared to the energies of all previously found structures. If the energies differ by less than ϵ, where ϵ is given by the user, the geometries are considered to be the same. The algorithm stops if the structure space along the torsional distortions has been searched starting from all minima that were located and no new minima were found.

4.2.2 Replica-exchange molecuar dynamics (REMD)

The replica-exchange method[290–294] was originally proposed in the framework of Monte Carlo simulations[290]. However, it can be similarly formulated in terms of MD[293], where it goes by the name of replica-exchange molecular dynamics (REMD). In the following, we shall give a short introduction into the algorithm of REMD. The basic idea is to perform MD simulations of several copies (*replicas*) of the same system. Each copy is simulated simultaneously in the canonical ensemble, but (usually) at a different temperature. For this reason, the replica-exchange method is also referred to as parallel tempering. A schematic representation of the algorithm is given in Fig. 4.2. At high temperatures, barriers can be overcome more easily and the trajectory is less likely to get trapped in a local minimum basin, while at low temperatures the local structure space can be sampled very accurately. Combining both advantages, the nuclear coordinates of two replicas, which are propagated at different temperatures, are exchanged after specific intervals of simulation time. In this way, the different copies of the system "walk" through the temperature space. If the system was stuck in a local minimum at a low temperature, it may be able to overcome the barrier when swapping to a higher temperature. In the opposite situation, a basin can be sampled more accurately when the temperature of the replica is switched to a lower temperature.

REMD is a generalized ensemble approach[293]. Let $\{\boldsymbol{p}_I\}$ denote the momenta of our system and $\{\boldsymbol{R}_I\}$ the nuclear positions, where $I = 1, \ldots, N_{\mathrm{at}}$, and N_{at} denotes the number of atoms in

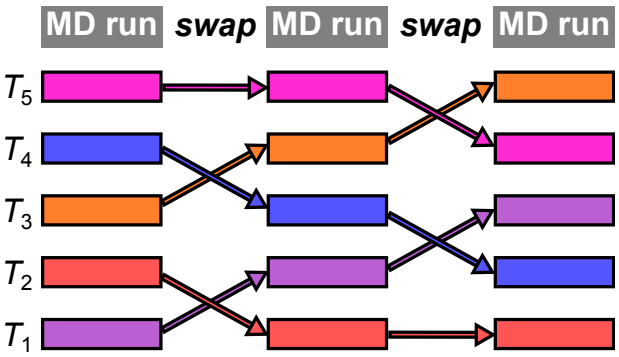


Figure 4.2: Schematic representation of the replica-exchange molecular dynamics technique. Five identical copies of the system are propagated in time via MD simulations at five different temperatures. After a certain number of MD steps swaps between pairs of replicas are performed. In this way the replicas are shifted through the temperature space. (For reasons of simplicity, we assume that all swap attempts are accepted in this illustration.)

the system. The Hamiltonian of the system can then be written as

$$\mathcal{H}(\{\boldsymbol{R}_I\},\{\boldsymbol{p}_I\}) = K(\{\boldsymbol{p}_I\}) + V(\{\boldsymbol{R}_I\}) \quad , \tag{4.18}$$

where $V(\{\boldsymbol{R}_I\})$ denotes the potential energy and

$$K(\{\boldsymbol{p}_I\}) = \sum_{I=1}^{N_{\mathrm{at}}} \frac{\boldsymbol{p}_I^2}{2m_I} \tag{4.19}$$

is the kinetic energy of the system. We simulate M *non-interacting* replicas n ($n = 1,\ldots,M$) of our system at M different temperatures T_m with $m = 1,\ldots,M$. There thus exists a bijective mapping

$$n = n(m) \equiv f(m) \quad . \tag{4.20}$$

A specific 'state' in the generalized ensemble can then be described as

$$X = (x_1^{n(1)}, x_2^{n(2)}, \ldots, x_M^{n(M)}) \quad , \tag{4.21}$$

where the subscript indicates the temperature and the superscript stands for the replica. The coordinates $x_m^{[n(m)]}$ for a specific replica are defined as

$$x_m^{[n(m)]} = (\{\boldsymbol{R}_I\}^{[n(m)]}, \{\boldsymbol{p}_I\}^{[n(m)]})_m \quad . \tag{4.22}$$

Let us now exchange two replicas in the generalized ensemble. Assume that we exchange replicas u and v that are propagated at temperatures k and l, respectively:

$$X = (\ldots, x_k^{[u]}, \ldots, x_l^{[v]}, \ldots) \longrightarrow X' = (\ldots, \hat{x}_k^{[v]}, \ldots, \hat{x}_l^{[u]}, \ldots) \quad . \tag{4.23}$$

When swapping the replicas, the information about the nuclear positions and momenta are exchanged, i.e.,

$$\left\{ \begin{array}{lcllcl} x_k^{[u]} & \equiv & (\{\boldsymbol{R}_I\}^{[u]}, \{\boldsymbol{p}_I\}^{[u]})_k & \longrightarrow & \hat{x}_k^{[v]} & \equiv & (\{\boldsymbol{R}_I\}^{[v]}, \{\hat{\boldsymbol{p}}_I\}^{[v]})_k \\ x_l^{[v]} & \equiv & (\{\boldsymbol{R}_I\}^{[v]}, \{\boldsymbol{p}_I\}^{[v]})_l & \longrightarrow & \hat{x}_l^{[u]} & \equiv & (\{\boldsymbol{R}_I\}^{[u]}, \{\hat{\boldsymbol{p}}_I\}^{[u]})_l \end{array} \right. \quad . \tag{4.24}$$

However, as indicated by the caret, the momenta of the replicas have to be adapted to the new temperature. The easiest way to account for the change in temperature is to simply scale the

momenta in the following way:

$$\begin{cases} \hat{\boldsymbol{p}}_I^{[v]} & \equiv & \sqrt{\frac{T_k}{T_l}}\boldsymbol{p}_I^{[v]} & , \\ \hat{\boldsymbol{p}}_I^{[u]} & \equiv & \sqrt{\frac{T_l}{T_k}}\boldsymbol{p}_I^{[u]} & . \end{cases} \tag{4.25}$$

This transformation ensures that the relation between the average of the kinetic energy and the temperature,

$$\left\langle \sum_{i=1}^{M} \frac{\boldsymbol{p}_I^2}{2m_I} \right\rangle_T = \frac{3}{2} N_{\mathrm{at}} k_{\mathrm{B}} T \qquad , \tag{4.26}$$

is preserved.

The replicas u and v depend on the temperatures l and k via the bijective mapping function f. The latter has to be updated after the exchange, i.e., f becomes f':

$$\begin{cases} u & = & f(k) & \longrightarrow & v & = & f'(k) \\ v & = & f(l) & \longrightarrow & u & = & f'(l) \end{cases} \qquad . \tag{4.27}$$

For the exchange process to converge towards an equilibrium distribution, the number of transitions $\mathcal{N}(X \to X')$ from state X to X' has to equal the number of transitions $\mathcal{N}(X' \to X)$ from X' to X (detailed balance):

$$\mathcal{N}(X \to X') = \mathcal{N}(X' \to X) \qquad . \tag{4.28}$$

$\mathcal{N}(X \to X')$ can be written as the product of the transition probability $w(X \to X')$ to go from X to X' and the weight factor $W_{\mathrm{REMD}}(X)$ of the state X in the generalized ensemble leading to

$$W_{\mathrm{REMD}}(X) w(X \to X') = W_{\mathrm{REMD}}(X') w(X' \to X) \tag{4.29}$$

As the replicas are non-interacting, the weight factor can be defined as the product of the Boltzmann factors of the individual replicas.

$$W_{\mathrm{REMD}} = \exp\left(- \sum_{m=1}^{M} \frac{1}{k_{\mathrm{B}} T_m} \mathcal{H}(\{\boldsymbol{R}_I\}, \{\boldsymbol{p}_I\})_m^{[n(m)]} \right) \qquad . \tag{4.30}$$

With this we obtain

$$\begin{aligned}
\frac{w(X \to X')}{w(X' \to X)} &= \frac{W_{\mathrm{REMD}}(X')}{W_{\mathrm{REMD}}(X)} \\
&= \exp\Bigg(-\frac{1}{k_{\mathrm{B}}T_k}\left[K(\{\hat{\boldsymbol{p}}_I\}^{[v]}) + V(\{\boldsymbol{R}_I\}^{[v]}) - K(\{\boldsymbol{p}_I\}^{[u]}) - V(\{\boldsymbol{R}_I\}^{[u]})\right] \\
&\quad -\frac{1}{k_{\mathrm{B}}T_l}\left[K(\{\hat{\boldsymbol{p}}_I\}^{[u]}) + V(\{\boldsymbol{R}_I\}^{[u]}) - K(\{\boldsymbol{p}_I\}^{[v]}) - V(\{\boldsymbol{R}_I\}^{[v]})\right]\Bigg) \\
&= \exp\Bigg(-\frac{1}{k_{\mathrm{B}}T_k}\left[\frac{T_k}{T_l}K(\{\boldsymbol{p}_I\}^{[v]}) + V(\{\boldsymbol{R}_I\}^{[v]}) - K(\{\boldsymbol{p}_I\}^{[u]}) - V(\{\boldsymbol{R}_I\}^{[u]})\right] \\
&\quad -\frac{1}{k_{\mathrm{B}}T_l}\left[\frac{T_l}{T_k}K(\{\boldsymbol{p}_I\}^{[u]}) + V(\{\boldsymbol{R}_I\}^{[u]}) - K(\{\boldsymbol{p}_I\}^{[v]}) - V(\{\boldsymbol{R}_I\}^{[v]})\right]\Bigg) \\
&= \exp\Bigg(-\frac{1}{k_{\mathrm{B}}T_k}\left[V(\{\boldsymbol{R}_I\}^{[v]}) - V(\{\boldsymbol{R}_I\}^{[u]})\right] - \frac{1}{k_{\mathrm{B}}T_l}\left[V(\{\boldsymbol{R}_I\}^{[u]}) - V(\{\boldsymbol{R}_I\}^{[v]})\right]\Bigg) \\
&= \exp(-\Delta) \quad ,
\end{aligned} \tag{4.31}$$

where

$$\Delta \equiv [\beta_l - \beta_k]\left(V(\{\boldsymbol{R}_I\}^{[u]}) - V(\{\boldsymbol{R}_I\}^{[v]})\right) \text{ with } \beta_m = \frac{1}{k_{\mathrm{B}}T_m} \quad . \tag{4.32}$$

Eq. 4.31 can be satisfied by using the well-known Metropolis acceptance criterion[295]

$$w(X \to X') = \min(1, \exp[-\Delta]) \quad . \tag{4.33}$$

In this way, the acceptance probability is 1 if the replica that is simulated at the higher temperature has a smaller potential energy. In the opposite case, the acceptance ratio decreases exponentially with the difference between the two potential energies.

In practice, an REMD run consists of two types of steps that are continuously repeated (as illustrated in Fig. 4.2):

1. Each replica is *simultaneously* and *independently* propagated in time by an MD simulation at a certain temperature in the canonical ensemble.

2. After a specific number of time steps, pairs of replicas are subjected to a swap attempt that is accepted with the probability given by the Metropolis criterion (Eq. 4.33). Without loss of generality, we can assume that the temperatures are sorted in ascending order. In practice, it is most efficient to only choose pairs of replicas with neighboring temperatures for the swap attempt as the acceptance probability decreases exponentially with the difference of β_l and β_k. This is realized by alternatingly choosing temperature pairs m and $m+1$ with even and odd m in subsequent exchange attempts (this is also illustrated in Fig. 4.2).

4.2.3 REMD FOR $(NO_3)^{-1}(HNO_3)$ + H_2O

While in the previous section the general framework of REMD was explained, in this section, we demonstrate some practical issues using monohydrated nitrate-nitric acid $(NO_3)^{-1}(HNO_3)$+H_2O as a test case[296]. We show how a conformational search using (*ab initio*) REMD can be performed and analyze its performance.

One advantage of the REMD approach is that one does not have to make *a priori* assumptions about the system. In the basin-hopping method, for instance, one has to define a proper random

move of the system. As discussed in Section 4.2.1, moves along torsional normal modes are well suited for peptides. However, this does not necessarily apply to $(NO_3)^{-1}(HNO_3)$+H_2O. In any case, one would first need to postulate suitable torsional angles, along which the molecule should be rotated. This means that when defining such a random move or, e.g., a merging protocol as needed in a genetic algorithm[280], one has to make *a priori* assumptions about the system. One needs to estimate, what kind of moves are "reasonable" (e.g., changes in the relative orientation of the two nitrate moieties) and which are not (e.g., disruption of the nitrate moiety). In the REMD approach, this is not the case – in principle, the nitrate moiety would be allowed to be destroyed (although, of course, it does not happen). The REMD run automatically restricts its search space to the physically relevant space. On the other hand, the disadvantage of the REMD approach might be that one may have to simulate longer than with a more carefully tuned algorithm, in which all "reasonable" moves are hard-coded and all "unreasonable" ones forbidden by hand.

For the REMD based structure search of the monohydrated nitrate-nitric acid cluster, we used the PBE+vdW (see Section 3.5.3) functional and employed 16 replicas. To initialize the REMD trajectories, four starting structures were used, which are depicted in Fig. 4.3. They were obtained from chemical intuition and local optimization by Nadja Heine and Knut Asmis from the Molecular Physics Department of the Fritz Haber Institute. Each starting structure was used to initialize four replicas. After every 2 ps of REMD simulation time, snapshots of all of the 16 replicas were relaxed with PBE+vdW using *tight* computational settings. As each replica was simulated for 30 ps, we obtained 240 PBE+vdW-relaxed structures in total. In order to identify the relevant structure types, these 240 geometries were sorted into families of similar structures. For this, we used a clustering approach based on interatomic distances. In this approach, for each structure a list containing the interatomic distances between all atom pairs is created. The comparison of the structures is conducted based on these distance matrices. If all distances r_{ij} of one structure do not deviate from another structure by more than $(\alpha \cdot\ r_{ij})$, the structures are considered to be similar and sorted into the same family. This criterion for the comparison ensures that larger interatomic distances are allowed to deviate more than small interatomic distances. This is reasonable as changes in the positions of nuclei that are close together changes the overall structure more than similarly large changes in nuclear positions that are far away. We determined $\alpha = 0.01$ to be a good parameter: choosing a value twice as large, $\alpha = 0.02$, yielded the same result, while choosing it half the size led to a differentiation of visually identical families. With $\alpha = 0.01$ we found 12 families, where the lowest PBE+vdW representative of each family is displayed in Fig. 4.3. The structures are sorted from 00 to 11 according to their PBE+vdW energy, which is given in Tab. 4.1 together with the zero-point energy (ZPE) corrected values. The ZPE is the energy of the quantum-mechanical ground state of the nuclei, which lies energetically above the classical ground state (where classical refers to the treatment of the nuclei). It will be explained in more detail in Chapter 5. As illustrated in Tab. 4.1, the ZPE corrections change the energy hierarchy. As the REMD runs rely on a classical treatment of the nuclei, they cannot account for the differences in the ZPE corrections of different conformers. If quantum nuclear effects become particularly important for a specific part of the search space, classical REMD does not capture that. This should be kept in mind, although it might be good enough for the sampling of the structure space in the present case.

During an MD simulation, structure families that are connected to each other by a rotation of

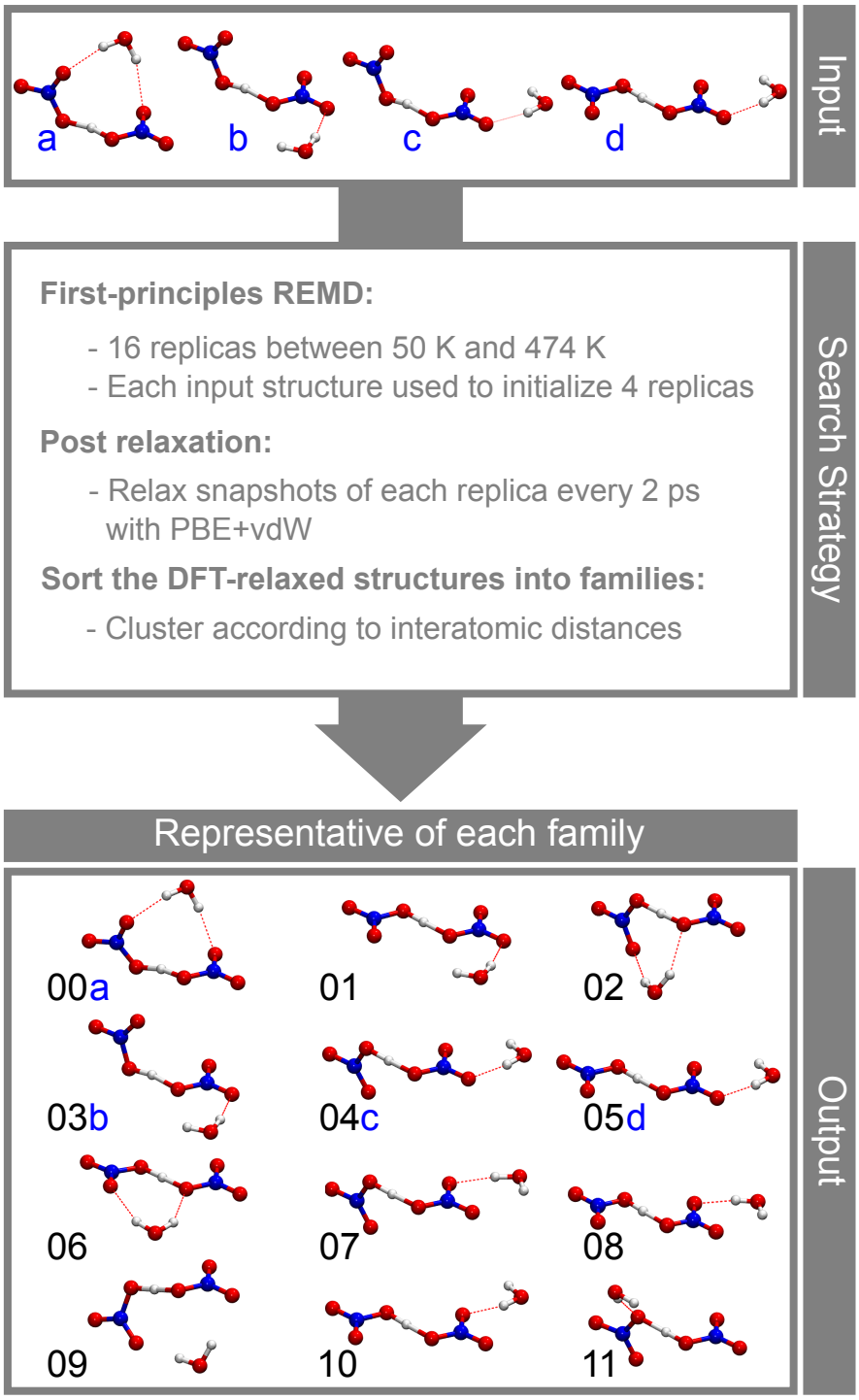


Figure 4.3: $(NO_3)^{-1}(HNO_3)+H_2O$: Input structures used as starting geometries for the REMD-based search (courtesy of Nadja Heine and Knut Asmis), details of the search strategy, and structures that drop out of the search. Input structures are labelled a, b, c, and d, while output structures are labelled 00 to 11. The output structure that corresponds to the respective initial structure is labelled accordingly with a, b, c, or d. The output structures are sorted according to their PBE+vdW energy hierarchy (see Tab. 4.1). All structures are aligned with respect to one nitrate moiety for a better comparison and for an easier differentiation of in-plane and out-of-plane arrangements of the nitrate moieties. Structure 08, e.g., is in plane, while 07 is the corresponding out-of-plane geometry.

Table 4.1: Relative PBE+vdW and zero-point corrected PBE+vdW energy differences for all families. The energies are relative to family 00. Additionally, the group assigned to each family is given (see Fig. 4.4).

		Energy differences (meV)	
family	group	PBE+vdW	PBE+vdW, ZP corrected
00	–	0.0	0.0
01	C	5.6	36.7
02	A	8.5	4.2
03	C	8.6	41.0
04	B	8.7	42.3
05	B	13.8	44.6
06	A	13.9	14.4
07	B	15.8	49.9
08	B	18.1	49.8
09	–	39.1	51.1
10	–	87.0	102.0
11	–	94.8	109.2

one nitrate moiety constantly interconvert. For this reason, we assign them to the same group. This is illustrated in Fig. 4.4, which shows sketches of the structure representatives of all families 00 to 11. All structures are aligned along one nitrate moiety, which allows for a better comparison between them. We distinguish the structures into *out-of-plane* geometries, where the planes of the two nitrate moieties are orthogonal to each other, and *in-plane* geometries, where the two nitrate moieties lie in the same plane. The structures that are arranged next to each other in one group are related by a rotation of one nitrate moiety, changing an in-plane geometry to the corresponding out-of-plane geometry. The lowest-energy family 00 is symmetric with respect to the position of the water molecule relative to the nitrate ions. In an in-plane arrangement of the nitrate ions, the oxygen atoms, which are H-bonded to the water molecule, would be too far away. This is why no corresponding in-plane geometry for family 00 exists. Families 02 and 06, on the other hand, are related by a rotation of one nitrate moiety from out-of-plane to in-plane and form group A. Families 04 and 05 and families 07 and 08 are likewise connected by a rotation of one nitrate ion. Additionally, families 05 and 08 and families 04 and 07 are related by a switch of the hydrogen bond of the water molecule from one oxygen of the nitrate moiety to the other one. These four families form group B. Group C contains family 01 and 03, which are again connected by an in-plane to out-of-plane rotation of one nitrate moiety.

Just as for family 00, there cannot exist an in-plane analogue of family 09. Family 10 should have an out-of-plane correspondent and also family 11 should have an in-plane correspondent. However, we do not find them in our structure search. Family 10 and 11 are about 90 meV higher in energy than the lowest-energy structure representative of family 00. This is why those structures might not be sampled in the (relatively short) REMD based search of 480 ps length in total (30 ps for each of the 16 replicas).

In order to examine the performance of our REMD simulation, the first question to be addressed is if enough replicas were chosen. The lowest temperature that we chose was 50.0 K and the highest temperature was 474.4 K. These two temperatures form the boundaries of our REMD simulation. If we used only these two temperatures, the swap probability (Eq. 4.33) between the two replicas would basically vanish as the gap between the two temperatures is too high. In order to enhance the efficiency of the REMD run, we have to find reasonable

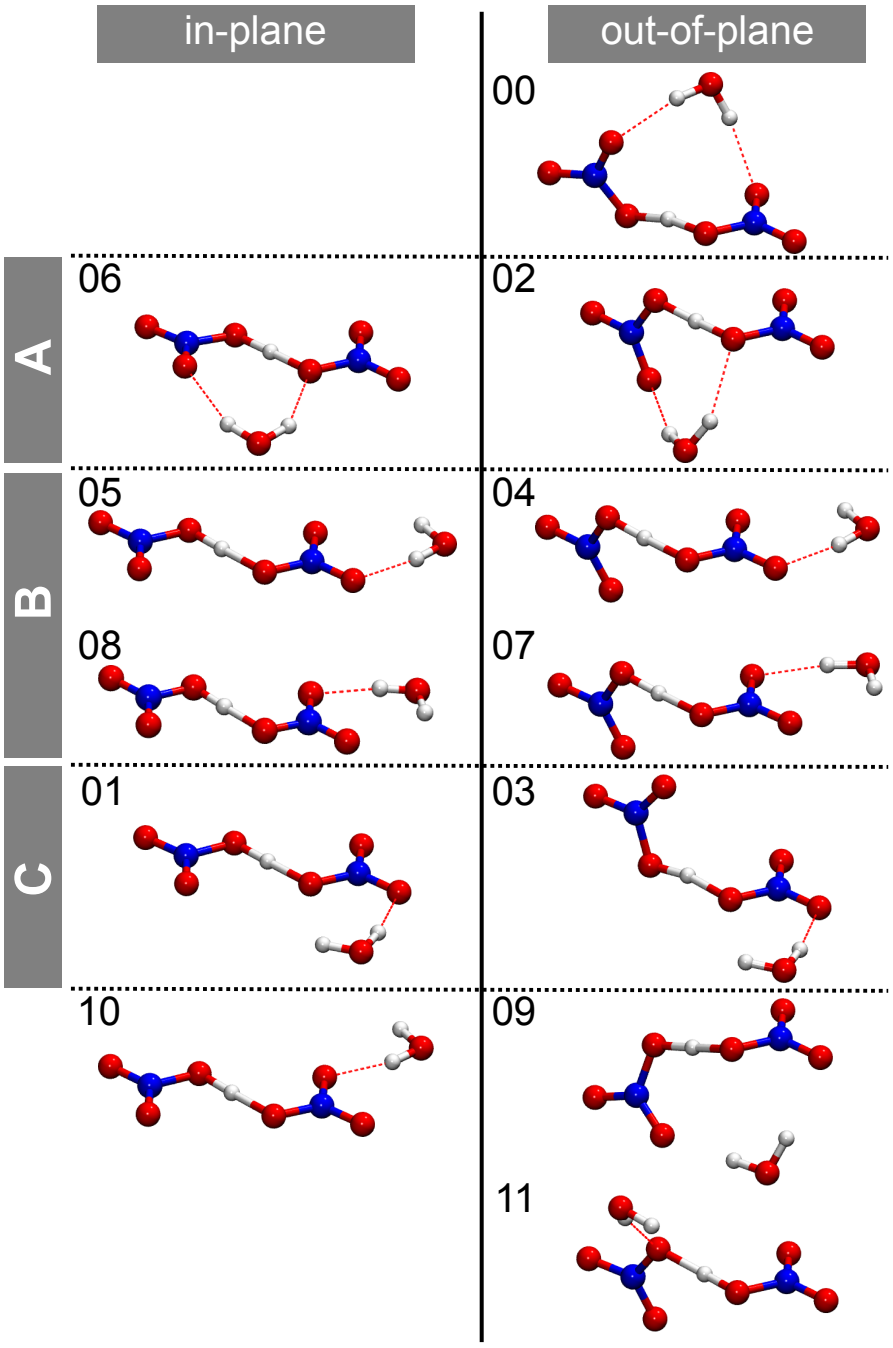


Figure 4.4: Sketches of the representative structures of family 00 to 11, where the labelling corresponds to the PBE+vdW energy hierarchy (see Tab. 4.1). The structures are sorted into geometries where the nitrate moieties are in-plane and geometries where the nitrate moieties are out-of-plane. Furthermore, the structures that are related to each other by rotation of one nitrate moiety or only a small change of the position of the water molecule are sorted into groups A, B, and C.

Table 4.2: Acceptance ratios for swap attempts of replicas simulated at adjacent temperatures for all neighboring temperature pairs.

Temperature pair			Acceptance ratio
50.0 K	$\leftrightarrow$	58.1 K	0.64
58.1 K	$\leftrightarrow$	67.5 K	0.68
67.5 K	$\leftrightarrow$	78.4 K	0.73
78.4 K	$\leftrightarrow$	91.1 K	0.67
91.1 K	$\leftrightarrow$	105.9 K	0.73
105.9 K	$\leftrightarrow$	123.0 K	0.68
123.0 K	$\leftrightarrow$	142.9 K	0.71
142.9 K	$\leftrightarrow$	166.0 K	0.65
166.0 K	$\leftrightarrow$	192.9 K	0.68
192.9 K	$\leftrightarrow$	224.1 K	0.62
224.1 K	$\leftrightarrow$	260.3 K	0.66
260.3 K	$\leftrightarrow$	302.5 K	0.68
302.5 K	$\leftrightarrow$	351.4 K	0.67
351.4 K	$\leftrightarrow$	408.3 K	0.60
408.3 K	$\leftrightarrow$	474.4 K	0.46

temperatures where to place intermediate replicas. *Enough* replicas are chosen if a sufficiently high swap acceptance ratio is obtained, where the swap efficiency should be larger than 0.1[293]. However, the larger the acceptance ratio the better. A perfect Monte Carlo move would have an acceptance ratio of 100%; this would be the most efficient way of sampling. In order to increase the acceptance ratio in REMD one has to choose more replicas, which increases the computational cost. For our specific REMD test case for the hydrated nitrate-nitric acid cluster the acceptance ratio is on average 0.66, where the lowest ratio is 0.46 as listed in Tab. 4.2. This shows that we used enough replicas.

The second question that needs to be addressed is if the temperatures were arranged in a reasonable fashion. An optimal arrangement would yield uniform acceptance ratios as this allows the replicas to perform a random walk in the temperature space. While also different distribution schemes have been proposed[297], most commonly, the temperatures in REMD simulations are distributed according to a geometric distribution. It can be shown that this kind of distribution of temperatures leads to uniform acceptance probabilities in the limit of a constant heat capacity[298–300]. In our test case we used a geometric distribution with $T_{m+1}/T_m \approx 1.16$. As shown in Tab. 4.2 the acceptance ratios are relatively uniform (all around $0.6 - 0.7$) except for the temperature pair 408.3 K $\leftrightarrow$ 474.4 K, where the acceptance ratio is 0.46. We shall analyze this further below.

The acceptance probability of a certain swap attempt depends on the difference between the potential energies of the two respective replicas (see Eq. 4.33). Figure 4.5 shows the probability-density distribution of the potential energy for the simulations at all of the different temperatures. There is sufficient overlap between the distributions leading to sufficiently high acceptance probabilities, as already discussed. However, Fig. 4.5 also shows that the potential-energy distribution of the highest temperature at 474.4 K shows two peaks. Investigating this issue further, we find that one replica looses its water molecule when being simulated at 474.4 K. This process happens after about 15 ps of simulation time. Afterwards, the replica mostly stays at 474.4 K. As the structure where the water molecule is very far away from the nitrate-nitric acid complex has a higher potential energy, we observe the second peak in the probability-density

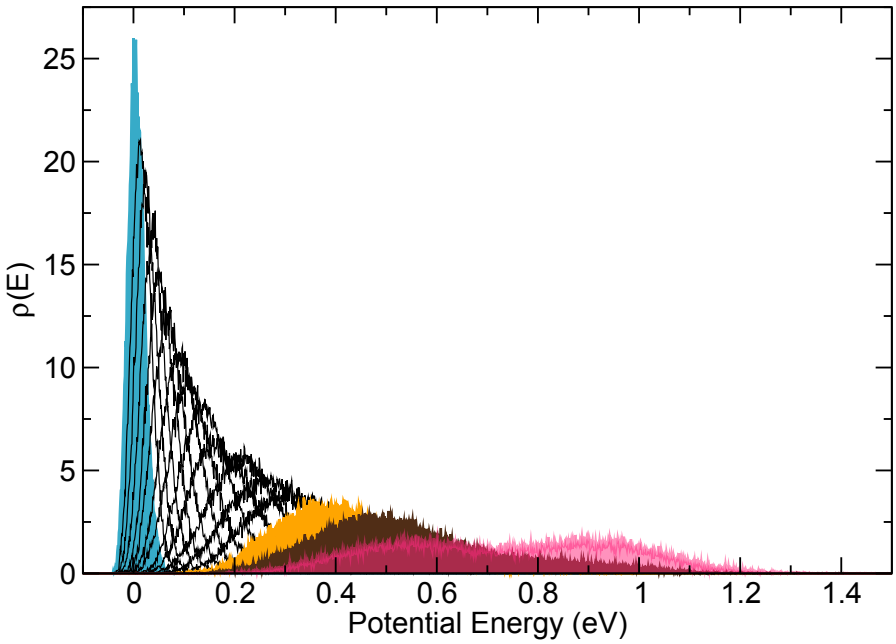


Figure 4.5: Probability-density distributions of the potential energy obtained from the simulations at all different temperatures. The leftmost distribution corresponds to the lowest temperature 50.0 K and the rightmost distribution corresponds to 474.4 K. The distributions were obtained by calculating histograms (bin width = 0.001 eV) of the potential energies taken at each time step of the respective simulation and normalizing them. The potential energy is given relative to the potential energy with the highest probability for 50.0 K.

distribution of the simulation at 474.4 K. The overlap with the distribution at the neighboring lower temperature reduces due to the two peak structure and so does the acceptance ratio. If, however, the replica is still successfully swapped to a lower temperature, it is very likely to swap back at the next attempt. This can be illustrated by the exchange charts shown in Fig. 4.6. The plots show the simulation temperature for each individual replica as a function of REMD simulation time. Figure 4.6a shows all replicas up to a simulation time of 5 ps, while Fig. 4.6b shows the random walk of three selected replicas up to 30 ps. The cyan colored replica starts at a temperature of 260.3 K. It walks through the temperature space, where at about 15 ps it arrives and mostly stays at 474.4 K. It looses its water molecule and gets stuck in the high-temperature regime. In contrast, the red-colored replica walks much more freely through the temperature space and the blue-colored replica even manages to touch both temperature boundaries within 12 ps. The only acceptance ratio that deviates from a uniform distribution can be explained by the system loosing its water molecule. Returning to the original question we can thus state that the temperatures were distributed in a reasonable way.

The next question that needs to be examined is if the temperature boundaries were chosen in a sensible fashion. While the lowest temperature is determined by the temperature that we are interested in, the highest temperature has to be chosen in such a way that the system does not get trapped in local minima. For the lowest temperature we chose 50.0 K. REMD samples the free-energy surface (FES) rather than the potential-energy surface (PES), i.e., our structure search is biased towards local minima of the FES. However, ultimately, we aim at a comparsion of calculated infrared (IR) spectra (at finite temperature) to experimental IR spectra that were conducted at similarly low (but finite) temperatures. Thus, the bias of the REMD search might

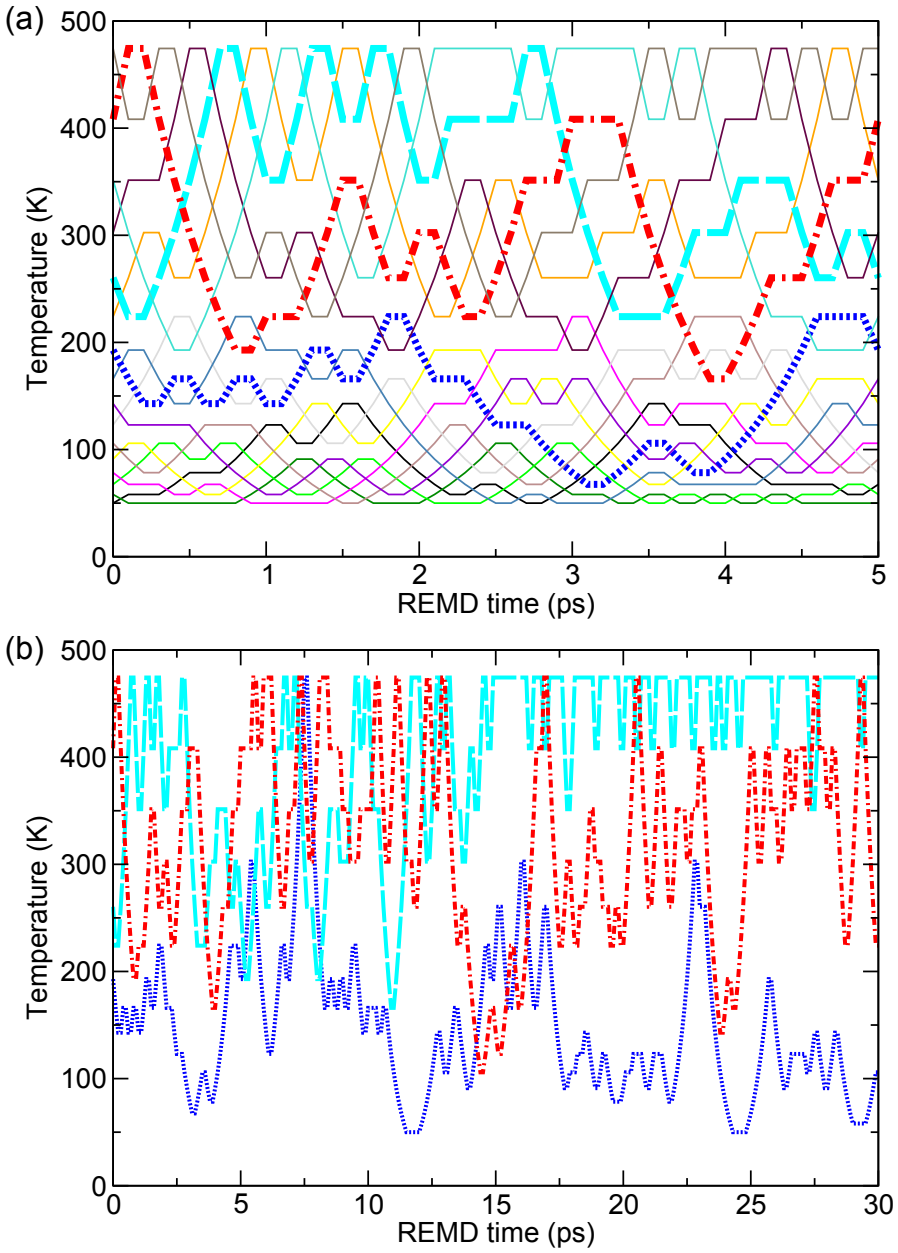


Figure 4.6: Exchange charts shown (a) for all replicas up to 5 ps of simulation time and (b) for selected replicas up to 30 ps.

even be favorable.

For the highest temperature we chose 474.4 K. As we saw earlier, it does not make sense to choose higher temperatures as the hydrated cluster already dissociates, i.e., looses its water molecule, at 474.4 K on an REMD time scale of 15 ps.

A further parameter in REMD simulations is the exchange attempt frequency (EAF). In our simulations we used an EAF of $10\,\mathrm{ps}^{-1}$. Although even larger frequencies might yield a higher efficiency[301], here we aimed at a compromise between efficiency and the computational cost arising from each swap attempt.

In summary, we can state that we used enough replicas, that our temperatures were reasonably distributed and that we used a sensible temperature range. However, another critical issue is if the REMD simulation time was sufficiently long. When we consider the exchange chart in Fig. 4.6b, we see that, e.g., the red-colored replica does not touch both temperature boundaries within 30 ps. The blue-colored replica touches both boundaries, but does not perform a *round trip*, i.e., does not go up-down-up or down-up-down. In fact, only two replicas perform a full round trip within 30 ps. In order to properly converge the REMD run with respect to coast-to-coast transitions, much longer simulation times are needed. However, our purpose here was to use the REMD run in order to find local minima of the PES (by local geometry optimization of snapshots of the replicas). By comparison of calculated IR spectra of the structures found by this search technique to experiment, we see that we were able to find the relevant conformers (not shown here). Additionally, the structures that turn out to be most relevant (number 06 and number 02, see Fig. 4.3) were not present in the initial structure pool, i.e., searching by intuition for the correct structures of systems even as small as our test case (12 atoms) might not always be sufficient.

5 MOLECULAR VIBRATIONS

After having discussed methods for describing and sampling the potential-energy surface (PES) in chapters 3 and 4, this chapter focuses on how one can describe and probe molecular vibrations. We discuss how infrared (IR) spectra can be calculated in the harmonic approximation and from molecular dynamics (MD) simulations and what kind of information they may reveal about the structure of the peptide under consideration. Additionally, we will address two approximations to the free energy, namely the harmonic-oscillator approximation and the rigid-rotor approximation. This chapter intends to give an overview, while further details can be found in textbooks, such as Refs. [266, 268, 302].

5.1 HARMONIC-OSCILLATOR APPROXIMATION

In a classical picture, the nuclei move on the Born-Oppenheimer (BO) PES $V_{\mathrm{BO}}(\{\boldsymbol{R}_I\})$ following Newton's equations of motion

$$m_I \ddot{\boldsymbol{R}}_I = \boldsymbol{F}_I = -\frac{\partial V_{\mathrm{BO}}(\{\boldsymbol{R}_I\})}{\partial \boldsymbol{R}_I} \quad , \tag{5.1}$$

where $V_{\mathrm{BO}}(\{\boldsymbol{R}_I\})$ is defined as:

$$\begin{aligned} V_{\mathrm{BO}}(\{\boldsymbol{R}_I\}) &= V_{\mathrm{nn}}(\{\boldsymbol{R}_I\}) + E^{\mathrm{e}}(\{\boldsymbol{R}_I\}) \\ &= \frac{1}{2}\sum_{I}^{N_{\mathrm{at}}}\sum_{J\neq I}^{N_{\mathrm{at}}} \frac{Z_I Z_J}{|\boldsymbol{R}_I - \boldsymbol{R}_J|} + E^{\mathrm{e}}(\{\boldsymbol{R}_I\}) \quad . \end{aligned} \tag{5.2}$$

V_{nn} is the nuclei-nuclei interaction and $E^{\mathrm{e}}(\{\boldsymbol{R}_I\})$ the electronic energy at a given conformation of the nuclei $\{\boldsymbol{R}_I\}$. We here refer to the BO PES $V_{\mathrm{BO}}(\{\boldsymbol{R}_I\})$ since we use it in this work. However, as before, V may likewise be an empirical potential-energy function such as a force field.

Let $\{R_{i,0}\}$ denote the positions of the nuclei in a local minimum conformation of the PES, where the index i runs over all $3N_{\mathrm{at}}$ degrees of freedom. It is advantageous to define new coordinates $\tilde{q}_i = (R_i - R_{i,0})$. The potential can then be expanded around this local minimum conformation in terms of small values of $\tilde{q}_i$:

$$V_{\mathrm{BO}}(\{\tilde{q}_i\}) = V_{\mathrm{BO}}(0) + \sum_{i}^{3N_{\mathrm{at}}} \left(\frac{\partial V_{\mathrm{BO}}}{\partial \tilde{q}_i}\right)_0 \cdot \tilde{q}_i + \frac{1}{2}\sum_{i}^{3N_{\mathrm{at}}}\sum_{j}^{3N_{\mathrm{at}}} \left(\frac{\partial^2 V_{\mathrm{BO}}}{\partial \tilde{q}_i \partial \tilde{q}_j}\right)_0 \cdot \tilde{q}_i \tilde{q}_j + \ldots \tag{5.3}$$

The first term is only a constant offset, while the second term vanishes for a local minimum conformation as the forces are zero. Thus, the first relevant term is the third one. Omitting

higher orders of the expansion yields the *harmonic approximation*. Within this approximation the equations of motion for the coordinates $\{\tilde{q}_i\}$ read:

$$m_i \ddot{\tilde{q}}_i = -\frac{\partial V_{\mathrm{BO}}}{\partial \tilde{q}_i} = -\sum_j^{3N_{\mathrm{at}}} \left(\frac{\partial^2 V_{\mathrm{BO}}}{\partial \tilde{q}_i \partial \tilde{q}_j} \right)_0 \cdot \tilde{q}_j \qquad . \tag{5.4}$$

It is advantageous to define mass-weighted coordinates $q_i = \sqrt{m_i}\tilde{q}_i$. With this, the equations of motion become:

$$\ddot{q}_i = -\sum_j^{3N_{\mathrm{at}}} \left(\frac{\partial^2 V_{\mathrm{BO}}}{\partial q_i \partial q_j} \right)_0 \cdot q_j \qquad . \tag{5.5}$$

Let us now make the ansatz $\boldsymbol{q} = \boldsymbol{A}\cos(\omega t + \phi)$. This is the equation of a harmonic oscillator, where all atoms oscillate with the same frequency and phase. Only the amplitude of the oscillation depends on the coordinate. With this, Eq. 5.5 can be transformed to an eigenvalue problem:

$$\omega^2 \boldsymbol{q} = \underline{\underline{\mathbb{H}}}\boldsymbol{q}, \qquad \underline{\underline{\mathbb{H}}} := \left(\frac{\partial^2 V_{\mathrm{BO}}}{\partial q_i \partial q_j} \right)_0 = \begin{pmatrix} \frac{\partial^2 V_{\mathrm{BO}}}{\partial q_1 \partial q_1} & \frac{\partial^2 V_{\mathrm{BO}}}{\partial q_1 \partial q_2} & \cdots & \frac{\partial^2 V_{\mathrm{BO}}}{\partial q_1 \partial q_{3N_{\mathrm{at}}}} \\ \frac{\partial^2 V_{\mathrm{BO}}}{\partial q_2 \partial q_1} & \frac{\partial^2 V_{\mathrm{BO}}}{\partial q_2 \partial q_2} & \cdots & \frac{\partial^2 V_{\mathrm{BO}}}{\partial q_2 \partial q_{3N_{\mathrm{at}}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 V_{\mathrm{BO}}}{\partial q_{3N_{\mathrm{at}}} \partial q_1} & \frac{\partial^2 V_{\mathrm{BO}}}{\partial q_{3N_{\mathrm{at}}} \partial q_2} & \cdots & \frac{\partial^2 V_{\mathrm{BO}}}{\partial q_{3N_{\mathrm{at}}} \partial q_{3N_{\mathrm{at}}}} \end{pmatrix}_0 \qquad , \tag{5.6}$$

where $\underline{\underline{\mathbb{H}}}$ denotes the Hessian matrix in mass-weighted coordinates. Solving Eq. 5.6 involves finding the eigenvectors $\boldsymbol{A}_n$ and eigenvalues ω_n^2 of $\underline{\underline{\mathbb{H}}}$. The eigenvectors $\boldsymbol{A}_n$ of $\underline{\underline{\mathbb{H}}}$ describe the amplitudes of the different coordinates for a state, where all atoms oscillate with the same frequency ω and phase ϕ. Such a state is called a *normal mode* and the time evolution of the coordinates is described by

$$\boldsymbol{q}(t) = \boldsymbol{A}_n \cos(\omega_n t + \phi_n) \qquad , \tag{5.7}$$

where $n = 1, \ldots, 3N_{\mathrm{at}}$. The system has $3N_{\mathrm{at}}$ degrees of freedom, where three degrees of freedom describe translations of the center of mass and three degrees describe rigid rotations. Thus, six eigenvalues are zero,[1] where the other $(3N_{\mathrm{at}} - 6)$ describe vibrational states of the molecule with positive eigenvalues ω_n^2. If the system is not in a local-minimum conformation, but in a saddle point, this will result in the occurrence of negative eigenvalues ω_n^2 (or imaginary frequencies ω_n). If we define $\underline{\underline{A}} = (\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_{3N_{\mathrm{at}}})$, where we assume the eigenvectors to be normalized, we can perform a coordinate transformation:

$$\boldsymbol{Q} = \underline{\underline{A}}^T \boldsymbol{q} \qquad , \tag{5.8}$$

where $\boldsymbol{Q} = (Q_i)$ are called the normal coordinates. This choice of coordinates is particularly useful as the potential energy (in the harmonic approximation) decouples in these coordinates:

$$V_{\mathrm{BO}} = \frac{1}{2}\boldsymbol{q}^T \underline{\underline{\mathbb{H}}}\boldsymbol{q} = \frac{1}{2}\boldsymbol{q}^T (\underline{\underline{A}}\underline{\underline{A}}^T)\underline{\underline{\mathbb{H}}}(\underline{\underline{A}}\underline{\underline{A}}^T)\boldsymbol{q} = \frac{1}{2}\boldsymbol{Q}^T \begin{pmatrix} \omega_1^2 & 0 & \cdots & 0 \\ 0 & \omega_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \omega_{3N_{\mathrm{at}}}^2 \end{pmatrix} \boldsymbol{Q} = \frac{1}{2}\sum_i^{3N_{\mathrm{at}}} \omega_i^2 Q_i^2 \tag{5.9}$$

[1] For linear molecules only five eigenvalues are zero.

The kinetic energy[2] in terms of the normal coordinates reads

$$K = \frac{1}{2}\sum_{i}^{3N_{\mathrm{at}}} \dot{Q}_i^2 \qquad . \tag{5.10}$$

In the next step, we shall move from the classical to a quantum-mechanical description of vibrations. In the Born-Oppenheimer approximation, the total wave function of a molecule decouples into an electronic part Φ_{e} and a nuclear part. The latter can further be written as a product of the translational wave function Φ_{T}, the rotational wave function Φ_{R}, and the vibrational wave function Φ_{V}:[3]

$$\Psi = \Phi_{\mathrm{e}}\Phi_{\mathrm{T}}\Phi_{\mathrm{R}}\Phi_{\mathrm{V}} \qquad . \tag{5.11}$$

In normal coordinates the vibrational Schrödinger equation (SGE) reads[302][4]:

$$-\frac{\hbar^2}{2}\sum_{i}^{3N_{\mathrm{at}}-6}\frac{\partial^2\Phi_{\mathrm{V}}}{\partial Q_i^2} + \frac{1}{2}\sum_{i}^{3N_{\mathrm{at}}-6}\omega_i^2 Q_i^2\Phi_{\mathrm{V}} = E_{\mathrm{V}}\Phi_{\mathrm{V}} \qquad , \tag{5.12}$$

where E_{V} denotes the energy associated with the vibration. As the potential is diagonal in the normal-mode coordinates the wave function can be written in a product ansatz:

$$\Phi_{\mathrm{V}} = \Phi(Q_1)\Phi(Q_2)\cdots\Phi(Q_{3N_{\mathrm{at}}-6}) \tag{5.13}$$

and the energy becomes $E_{\mathrm{V}} = E(1) + E(2) + \cdots + E(3N_{\mathrm{at}} - 6)$. The functions $\Phi(Q_i)$ then satisfy:

$$-\frac{\hbar^2}{2}\frac{\partial^2\Phi(Q_i)}{\partial Q_i^2} + \frac{1}{2}\omega_i^2 Q_i^2\Phi(Q_i) = E(i)\Phi(Q_i) \qquad . \tag{5.14}$$

This is the equation of the quantum-mechanical harmonic oscillator, for which the solutions are well known[167]. The energy for the harmonic oscillator with the normal frequency ω_i is

$$E_{n_i}(i) = \hbar\omega_i(n_i + \frac{1}{2}), \qquad n_i = 0, 1, \cdots \tag{5.15}$$

The variable i denotes the normal mode, while n_i is a quantum number defining the state of the harmonic oscillator. The total energy then amounts to

$$E = \sum_{i}^{3N_{\mathrm{at}}-6}\hbar\omega_i(n_i + \frac{1}{2}) \tag{5.16}$$

and the total wave function depends on all quantum numbers $n = \{n_i\}$.

The vibrational energy levels of a harmonic oscillator can be excited or de-excited by the absorption or emission of photons. Transitions between two levels can only occur if the photon energy matches the energy difference between those levels. In molecules, the energy associated with vibrational transitions corresponds to the energy of photons in the IR range. This enables to probe the vibrational modes of a molecule based on IR photon absorption (IR spectroscopy). In

[2]Here denoted as K in order to avoid confusion with the temperature T.
[3]Rotations and vibrations decouple only approximately. We here assume this approximation to be valid, where further details can be found in Ref. [302].
[4]We here explicitly write $\hbar$.

order to calculate the transition probability between two vibrational states one can consider the interaction of the dipole moment of the molecule with a weak electromagnetic field as a small perturbation to the system. Using Fermi's golden rule one finds that the transition probability between two states n and n' is proportional to the transition dipole moment $|(\boldsymbol{\mu})_{n,n'}|^2$[302], which is defined as:

$$(\boldsymbol{\mu})_{n,n'} = \int \Phi_n^* \boldsymbol{\mu} \Phi_{n'} \mathrm{d}\boldsymbol{Q} \quad , \tag{5.17}$$

where $\boldsymbol{\mu}$ is the dipole moment of the system and the integration runs over the whole configuration space. The dipole moment $\boldsymbol{\mu}$ can be expanded in terms of Q_i:

$$\boldsymbol{\mu} = \boldsymbol{\mu}(0) + \sum_i^{3N_{\mathrm{at}}-6} \left(\frac{\partial \boldsymbol{\mu}}{\partial Q_i}\right)_0 \cdot Q_i + \dots \quad , \tag{5.18}$$

where typically only the first two terms are kept (electrical harmonic approximation). As already the potential has been approximated, this is frequently called the double-harmonic approximation[303]. Based on this expansion, the transition dipole moment becomes

$$(\boldsymbol{\mu})_{n,n'} \approx \boldsymbol{\mu}(0) \int \Phi_n^* \Phi_{n'} \mathrm{d}\boldsymbol{Q} + \sum_i^{3N_{\mathrm{at}}-6} \left(\frac{\partial \boldsymbol{\mu}}{\partial Q_i}\right)_0 \cdot \int \Phi_n^* Q_i \Phi_{n'} \mathrm{d}\boldsymbol{Q} \quad . \tag{5.19}$$

This only takes non-zero values for specific cases, which gives rise to the well-known vibrational selection rules for harmonic oscillators. When taking into account that Φ_n and $\Phi_{n'}$ are products of the harmonic-oscillator wave functions and the harmonic-oscillator wave functions are orthonormal, the first term only contributes for $n = n'$, i.e., it does not affect the intensities of the vibrational spectrum. From the second term it is clear that only those transitions can arise where $n_i' = n_i \pm 1$ and all other oscillators have $n_k' = n_k$. This means that (in the double-harmonic approximation) only frequencies corresponding to normal-mode frequencies should appear in the IR spectrum. Additionally, only those normal frequencies will appear that are associated with vibrations that change the dipole moment of the system. For a more detailed account, the interested reader is referred to Ref. [302]. The integral absorption coefficient for the spectral line associated with the normal mode i is

$$\mathcal{I}_i = \frac{N_{\mathrm{A}}\pi}{3c} \left| \left(\frac{\partial \boldsymbol{\mu}}{\partial Q_i}\right)_0 \right|^2 \quad , \tag{5.20}$$

where N_{A} denotes Avogadro's number and c is the velocity of light. A derivation of this can be found, e.g., in Ref. [304].

5.1.1 Free energy in the harmonic oscillator-rigid rotor approximation

At physiological conditions, the actual surface that the peptide or protein explores during folding is the free-energy surface (FES). The free energy is the relevant thermodynamic quantity that describes the behavior of the system at finite temperature. As discussed in Section 2.5, we here concentrate on the Helmholtz free energy. If $Q(T)$ denotes the partition function of the canonical

ensemble, the Helmholtz free energy of a system can be calculated via:

$$F(T) = -k_{\mathrm{B}}T\ln[Q(T)] \quad , \tag{5.21}$$

where T is the temperature and k_{B} the Boltzmann factor. Here we give a brief account on how to evaluate the rotational and the vibrational contributions to the free energy in the harmonic oscillator-rigid rotor approximation.[5]

In the harmonic approximation, the vibrational partition function is a product of the partition functions of single harmonic oscillators with the normal frequencies ω_i:

$$q_{\mathrm{vib}} = \prod_{i=1}^{3N_{\mathrm{at}}-6}\sum_{n_i=0}^{\infty}\exp\left[-\frac{\hbar\omega_i}{k_{\mathrm{B}}T}\left(n_i+\frac{1}{2}\right)\right] = \prod_{i=1}^{3N_{\mathrm{at}}-6}\frac{\exp\left(-\frac{\hbar\omega_i}{2k_{\mathrm{B}}T}\right)}{1-\exp\left(-\frac{\hbar\omega_i}{k_{\mathrm{B}}T}\right)} \quad . \tag{5.22}$$

For the infinite sum, one can employ the rule for a geometric series such that the sum can be written in a closed form. According to Eq. 5.21 the vibrational contribution to the free energy in the harmonic approximation then reads

$$F_{\mathrm{vib}}(T) = \sum_{i=1}^{3N_{\mathrm{at}}-6}\left[\frac{\hbar\omega_i}{2}+k_{\mathrm{B}}T\ln(1-\mathrm{e}^{-\frac{\hbar\omega_i}{k_{\mathrm{B}}T}})\right] \quad . \tag{5.23}$$

At $T = 0\,\mathrm{K}$ this amounts to $\sum_{i=1}^{3N_{\mathrm{at}}-6}\frac{\hbar\omega_i}{2}$, which is referred to as the zero-point energy (ZPE). $F_{\mathrm{vib}}(T)$ can be expressed in terms of the entropy $S_{\mathrm{vib}}(T)$ and the internal energy $U_{\mathrm{vib}}(T)$ via

$$F_{\mathrm{vib}}(T) = U_{\mathrm{vib}}(T) - TS_{\mathrm{vib}}(T) \quad , \tag{5.24}$$

where

$$U_{\mathrm{vib}}(T) = k_{\mathrm{B}}T^2\frac{\partial\ln q_{\mathrm{vib}}}{\partial T} = \sum_{i=1}^{3N_{\mathrm{at}}-6}\left[\frac{\hbar\omega_i}{2}+\frac{\hbar\omega_i}{\exp\left(\frac{\hbar\omega_i}{k_{\mathrm{B}}T}\right)-1}\right] \quad . \tag{5.25}$$

The simplest example of a rigid-body rotor is a heteronuclear diatomic molecule. For this, the energy levels are given by[268]

$$\varepsilon_J = \frac{\hbar^2 J(J+1)}{2I} \text{ with } J = 0,1,2,\ldots \tag{5.26}$$

I denotes the moment of inertia and J is the angular-momentum quantum number. As the degeneracy of each level is $2J+1$ the rotational partition function reads

$$q_{\mathrm{rot}} = \sum_{J=0}^{\infty}(2J+1)\mathrm{e}^{-\frac{\hbar^2 J(J+1)}{2Ik_{\mathrm{B}}T}} = \sum_{J=0}^{\infty}(2J+1)\mathrm{e}^{-\frac{\Theta}{T}J(J+1)} \text{ with } \Theta = \frac{\hbar^2}{2Ik_{\mathrm{B}}} \quad . \tag{5.27}$$

For this sum there is no closed form as for the vibrational partition function. However, normally, $\Theta/T \ll 1$ for most molecules at room temperature such that the sum can be approximated by an

[5] The rotational and vibrational contributions to the partition function are to a very good approximation separable (for further details see Refs. [268, 302]).

integral

$$q_{\mathrm{rot}}(T) \approx \int_0^\infty (2J+1)\mathrm{e}^{-\frac{\Theta}{T}J(J+1)}\mathrm{d}J = \int_0^\infty \mathrm{e}^{-\frac{\Theta}{T}J(J+1)}\mathrm{d}[J(J+1)] = \frac{T}{\Theta} \quad . \tag{5.28}$$

In general, a non-linear polyatomic molecule has three principal moments of inertia I_A, I_B, and I_C, which determine its rotational properties. Assuming that $I_A \neq I_B \neq I_C$, the rotational partition function can be approximated in a manner similar to a heteronuclear diatomic molecule by[268]

$$q_{\mathrm{rot}}(T) = \frac{\pi^{1/2}}{\sigma}\left(\frac{2k_{\mathrm{B}}T}{\hbar^2}\right)^{3/2}(I_A I_B I_C)^{1/2} = \frac{\pi^{1/2}}{\sigma}\left(\frac{T^3}{\Theta_A\Theta_B\Theta_C}\right)^{1/2} \quad , \tag{5.29}$$

where σ is a symmetry factor describing the number of possible ways to rotate the system into itself again. Θ_X is the characteristic rotational temperature for each principal moment of inertia I_X, and is defined as $\Theta_X = \hbar^2/(2I_X k_{\mathrm{B}})$. A more detailed derivation can be found in the book by McQuarrie[268].

5.2 Infrared (IR) spectroscopy of proteins and peptides

As described in the previous section, IR spectroscopy is a tool to probe the vibrational modes of a peptide or protein based on IR photon absorption. The IR spectral region extends from 0.78 µm to 1000 µm, where the region between 0.78 µm and 2.5 µm is referred to as *near-infrared*, and the region between 2.5 µm and 50 µm is called *mid-infrared*, while the *far-infrared* region goes from 50 µm to 1000 µm[305]. In IR spectroscopy, the intensity is often plotted as a function of the wavenumber $\overline{\nu}$, which is defined as:

$$\overline{\nu} = \frac{1}{\lambda} \tag{5.30}$$

and is given in cm^{-1}. In this thesis, the mid-infrared region is the most interesting to us. It corresponds to a wavenumber range of 200 – 4000 cm^{-1}.

As discussed in Section 5.1, in the harmonic approximation, the vibrations of a molecule can be decoupled into a set of normal modes, where within each normal mode all atoms vibrate with the same frequency but different amplitudes. Often only a few amplitudes are significant, giving rise to modes that are localized at a small group of atoms[306]. Those localized vibrations can be differentiated into

- Stretching vibrations (change in bond length) and
- Bending vibrations (change in bond angle), where one distinguishes again between
  - *in-plane* bending, such as scissoring and rocking, where the atoms involved move within one plane, and
  - *out-of-plane* bending, such as wagging, twisting, and also umbrella motion, where the atoms that are involved move out of their equilibrium-position plane during the oscillation.

In the diatomic classical harmonic-oscillator model, the frequency ν of oscillation reads

$$\nu = \left(\frac{k}{\mu}\right)^{1/2} \frac{1}{2\pi} \quad , \tag{5.31}$$

where k denotes the force constant of the interaction between the two atoms and $1/\mu = 1/m_A + 1/m_B$ is the reduced mass of the system with the individual masses of the atoms m_A and m_B. From Eq. 5.31 it follows that the frequency of vibration is sensitive to the force constant and the reduced mass. It increases with increasing force constant and decreases with increasing mass.

Intra- and intermolecular interactions influence the force constant, making the frequency dependent on the environment. Thus, in a general picture, vibrational modes of a molecule become sensitive to, and thus reveal information about, its chemical composition, bond lengths, bond strengths, bond angles, hydrogen bonds, and the environment of the molecule[305, 306]. Hence, the vibrational modes depend on the conformation of the molecule or peptide so that the IR spectrum can give valuable information about its structure. For this reason, IR spectroscopy is a widely-used tool to probe the structure of molecules.

Many normal modes are localized in nature, i.e., the vibration involves only a few atoms, so that specific fragments of the protein or peptide can be probed individually. This may reveal information about the structure of this group of atoms and its environment. IR spectroscopy is able to directly probe the presence and strength of hydrogen bonds: generally, a stretching mode is shifted to lower wavenumbers when the atom involved is hydrogen bonded[305]. The reason for this is that the hydrogen bond weakens the restoring force. On the other hand, for a bending mode a hydrogen bond strengthens the restoring force, which induces a shift of the band to higher wavenumbers. A vibrational mode is IR active if the dipole moment changes upon the vibration (see Section 5.1). For this reason, generally all polar bonds yield a contribution to the IR spectrum[305]. This is an advantage of IR spectroscopy as it thus probes nearly all bonds. However, the larger the molecule becomes the more spectral lines occur and the harder it becomes to resolve certain features.

As discussed in Chapter 2, a peptide is a polyamide, where the monomeric units, the amino-acid residues, only differ in the side chains. The central repeating pieces of the peptide's backbone are the amide groups, characterized as R–C(=O)–N(H)-R' (where R and R' denote an organic group or a hydrogen atom). The amides show characteristic bands in the IR spectral region, which can yield valuable information about the peptide structure. They are[305, 306]:

- *Amide A and amide B modes:* ($\approx 3300\,\mathrm{cm}^{-1}$ and $\approx 3070\,\mathrm{cm}^{-1}$, respectively) The amide A band appears between $3310\,\mathrm{cm}^{-1}$ and $3270\,\mathrm{cm}^{-1}$. It is caused by the N–H stretching vibrations, where each mode is restricted to only one group of NH atoms and is thus very local. The frequency of vibration is sensitive to the presence and strength of hydrogen bonds. Due to the local nature of the vibration and the sensitivity towards H-bonds, the amide A band is very conformer sensitive. The amide B band appears between $3030\,\mathrm{cm}^{-1}$ and $3100\,\mathrm{cm}^{-1}$. It is rather low in intensity and results from a Fermi resonance[6] between the N–H stretching modes and another mode that has similar energy. In polypeptide

[6] If two modes have nearly identical energy and symmetry they can mix, which results in an enhanced splitting and a re-distribution of intensity between the two peaks.

helices, this is an overtone[7] of the amide II band.

- *Amide I* ($\approx 1650\,\mathrm{cm}^{-1}$) The amide I band occurs around $1650\,\mathrm{cm}^{-1}$ and is dominated by the stretching vibrations of the C=O groups. It also contains contributions from out-of-phase C–N stretching vibrations, N–H in-plane bendings, and C–C–N deformations. In contrast to the amide A band, the amide I band is formed by collective modes that involve not only one amide group, but several or all of them. The modes are barely influenced by the side chains, but depend on the backbone conformation and are especially sensitive to the secondary structure of the backbone.

- *Amide II* ($\approx 1550\,\mathrm{cm}^{-1}$) The amide II band occurs around $1550\,\mathrm{cm}^{-1}$. It is dominated by a coupling of in-plane N–H bendings and out-of-phase C–N stretching vibrations. Furthermore, it involves also smaller contributions from C–C and N–C stretching vibrations and in-plane C–O bendings. Just as the amide I band, the modes are rather collective and are only weakly influenced by the conformation and nature of the side chains. It has been shown that information about the amide II band alone can be sufficient to predict secondary structure[307].

- *Amide III* ($\approx$1200–1400 cm^{-1}) The amide III band occurs between $1200\,\mathrm{cm}^{-1}$ and $1400\,\mathrm{cm}^{-1}$. It arises through a mixture of localized and collective modes involving in-phase N–H in-plane bending and C–N stretching. Additionally, also C–O in-plane bending and C–C stretching vibrations play a role. In contrast to the amide I and amide II bands, which are barely influenced by the nature of the side chains, the side chains contribute to the amide III band. The amide III mode has been used for secondary-structure determination[308–310]. It is very sensitive to small changes in the secondary-structure conformation[311].

The position and shape of these characteristic modes reveal information about the structure of the peptide. The width of the peaks can give clues on the conformational freedom, where broader peaks indicate a higher flexibility. Especially the amide I band has been used for secondary-structure analysis (see reviews Refs. [305, 306]). This band is dominated by the C=O stretching vibrations. If the C=O groups are hydrogen bonded to N–H groups as present, e.g., in a helix, as a rule of thumb the amide I band is shifted to lower wavenumbers with respect to the non-hydrogen bonded case. This down-shifting increases with increasing helix length. On the other hand, the amide II band shifts to higher wavenumbers if the N–H groups are hydrogen bonded. Coupling of modes can induce relative peak shifting and splitting. The spacing between the amide I and amide II bands can reveal information, e.g., about the interaction of the C(=O) and N–H groups in the backbone[312].

5.3 Infrared (IR) spectra from the dipole time autocorrelation function

Calculating IR spectra in the double-harmonic approximation, as explained in Section 5.1, suffers from several problems[303, 312]. Obviously, the double-harmonic approximation does not take

[7] The selection rule of vibrational transitions of $\Delta n = \pm 1$ only strictly holds in the harmonic-oscillator model. In actual molecules, this selection rule is softened. If $\Delta n \neq \pm 1$ this gives rise to a so-called overtone, which is, however, lower in intensity than the transition associated with the fundamental frequency.

into account any effects due to the anharmonicity of the potential or higher-order corrections to the dipole moment. Furthermore, the spectra reflect the system at $T = 0\,\mathrm{K}$, while experimental spectra are recorded at finite temperatures. Peptides are very floppy molecules and can undergo structural changes at finite temperatures, which cannot be described in the double-harmonic approximation. A way to account for anharmonicities of the potential and conformational changes naturally is to calculate the IR spectrum based on the concept of time-correlation functions (see below) that are obtained from MD simulations. However, one has to keep in mind that in these approaches the nuclei are usually treated as classical particles. This means that, e.g., at room temperature (300 K) anharmonicities associated with vibrational modes below $200\,\mathrm{cm}^{-1}$, i.e., $\hbar\omega < k_\mathrm{B}T$, are described relatively well. On the other hand, the anharmonicities associated with higher wavenumbers are underestimated by the classical treatment of the nuclear motion[313, 314]. In order to approximate nuclear quantum time correlation functions, one can resort to techniques known as centroid molecular dynamics[315] or ring-polymer molecular dynamics[316], which are, however, computationally not feasible for the systems of interest in the present thesis (108 to 220 atoms). Here, we treat the nuclei as classical particles and employ an *a posteriori* quantum correction factor, as will be described below.

First, the concept of a time-correlation function will be outlined (see, e.g., Ref. [268]). Let A be a function of a system with the generalized spatial coordinates $\boldsymbol{q}$ and the generalized momenta $\boldsymbol{p}$:

$$A(t) = A\{\boldsymbol{p}(t), \boldsymbol{q}(t)\} = A(\boldsymbol{p}, \boldsymbol{q}; t) \qquad . \tag{5.32}$$

The classical time-correlation function is then defined as[268]

$$C(t) = \langle A(0) \cdot A(t) \rangle = \int \mathrm{d}\boldsymbol{q} \int \mathrm{d}\boldsymbol{p}\, A(\boldsymbol{p}, \boldsymbol{q}; 0) A(\boldsymbol{p}, \boldsymbol{q}; t) f(\boldsymbol{p}, \boldsymbol{q}) \qquad , \tag{5.33}$$

where $f(\boldsymbol{p}, \boldsymbol{q})$ denotes the equilibrium phase space distribution function and the integral runs over the whole phase space, i.e., $<>$ denotes an ensemble average. Specifically, this type of correlation function is called an *autocorrelation function* as it describes the correlation of $A(t)$ with itself. The quantum-mechanical analogue is generally defined as

$$C_{AA}^{\mathrm{qm}}(t) = \frac{\mathrm{Tr}\left[\exp(-\hat{\mathcal{H}}^0/\{k_\mathrm{B}T\})\hat{A}(0)\hat{A}(t)\right]}{\mathrm{Tr}\left[\exp(-\hat{\mathcal{H}}^0/\{k_\mathrm{B}T\})\right]} \qquad , \tag{5.34}$$

where $\hat{A}(t)$ is the time-dependent operator in the Heisenberg picture defined as

$$\hat{A}(t) = \exp\left(i\hat{\mathcal{H}}^0 t/\hbar\right)\hat{A}\exp\left(-i\hat{\mathcal{H}}^0 t/\hbar\right) \qquad . \tag{5.35}$$

In the following we will sketch the derivation of how to calculate the IR spectrum of a system from an MD trajectory, while a much more detailed account is given, e.g., in Ref. [268]. Let us first consider a small electric field $\boldsymbol{E}(t) = \boldsymbol{\epsilon} E_0 \cos(\omega t)$ acting on the system, where E_0 describes the amplitude and $\boldsymbol{\epsilon}$ denotes a unit vector pointing along the electric-field direction. The Hamiltonian $\hat{\mathcal{H}}^0$ of the system is then perturbed by

$$\lambda\hat{V}(t) = -\boldsymbol{E}(t) \cdot \hat{\boldsymbol{\mu}} \qquad , \tag{5.36}$$

where $\hat{\boldsymbol{\mu}}$ is the dipole-moment operator of the system. Employing time-dependent perturbation theory (up to first order) yields the famous Fermi golden rule, which describes the probability per unit time that the system goes from the initial state i to the final state f[268]:

$$P_{i\to f}(\omega) = \frac{\pi E_0^2}{2\hbar^2} \left|\langle f|\boldsymbol{\epsilon}\cdot\hat{\boldsymbol{\mu}}|i\rangle\right|^2 \left[\delta(\omega_{fi}-\omega)+\delta(\omega_{fi}+\omega)\right] \quad , \tag{5.37}$$

with $\omega_{fi} = \omega_f - \omega_i$. The first δ-function describes absorption and the second δ-function describes stimulated emission. For the rate of energy loss from radiation (E_{rad}) to the system it follows[268]:

$$-\frac{\mathrm{d}}{\mathrm{d}t}E_{\mathrm{rad}} = \sum_i \sum_f \rho_i \hbar\omega_{fi} P_{i\to f} \quad , \tag{5.38}$$

where ρ_i denotes the probability of the system to be in the initial state i. The absorption coefficient $\alpha(\omega)$, which describes the IR absorption spectrum, can be obtained by dividing the above expression by the incident flux of radiation. When assuming that $\rho_f = \rho_i \exp(-\hbar\omega_{fi}/[k_{\mathrm{B}}T])$ and interchanging the indices in the second delta function one finds (see Ref. [268] for more details):

$$\alpha(\omega) \quad \propto \quad \omega(1-\mathrm{e}^{-\frac{\hbar\omega}{k_{\mathrm{B}}T}}) \underbrace{\sum_f \sum_i \rho_i \left|\langle f|\boldsymbol{\epsilon}\cdot\hat{\boldsymbol{\mu}}|i\rangle\right|^2 \delta(\omega_{fi}-\omega)}_{=:G(\omega)} \tag{5.39}$$

When considering that $\delta(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{i\omega t}\mathrm{d}t$ one can rewrite $G(\omega)$ in the following way:

$$\begin{aligned} G(\omega) &= \frac{1}{2\pi}\sum_f\sum_i \rho_i \langle i|\boldsymbol{\epsilon}\cdot\hat{\boldsymbol{\mu}}|f\rangle\langle f|\boldsymbol{\epsilon}\cdot\hat{\boldsymbol{\mu}}|i\rangle \int_{-\infty}^{\infty} \exp\left[\left(\frac{E_f-E_i}{\hbar}-\omega\right)it\right]\mathrm{d}t \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{d}t\exp(-i\omega t)\sum_f\sum_i \rho_i \langle i|\boldsymbol{\epsilon}\cdot\hat{\boldsymbol{\mu}}|f\rangle \left\langle f|\boldsymbol{\epsilon}\underbrace{\exp(i\hat{\mathcal{H}}^0 t/\hbar)\cdot\hat{\boldsymbol{\mu}}\exp(-i\hat{\mathcal{H}}^0 t/\hbar)}_{\hat{\boldsymbol{\mu}}(t)}|i\right\rangle \quad . \end{aligned} \tag{5.40}$$

With $\sum_f |\, f\rangle\langle f\,| = 1$ and averaging $\boldsymbol{\epsilon}$ over all directions one arrives at

$$G(\omega) = \frac{1}{6\pi}\int_{-\infty}^{\infty} \underbrace{\sum_i \rho_i \langle i|\hat{\boldsymbol{\mu}}(0)\cdot\hat{\boldsymbol{\mu}}(t)|i\rangle}_{C^{\mathrm{qm}}_{\mu\mu}} \mathrm{e}^{-i\omega t}\mathrm{d}t \quad , \tag{5.41}$$

where $C^{\mathrm{qm}}_{\mu\mu}$ is the canonical quantum correlation function as defined in Eq. 5.34. Another well-known quantum correlation function is the Kubo-transformed correlation function[317]:

$$\tilde{C}_{AA} = \frac{1}{\beta\mathrm{Tr}\left[\exp(-\beta\hat{\mathcal{H}}^0)\right]}\int_0^\beta \mathrm{Tr}\left[\exp(-\{\beta-\lambda\}\hat{\mathcal{H}}^0)\hat{A}(0)\exp(-\lambda\hat{\mathcal{H}}^0)\hat{A}(t)\right]\mathrm{d}\lambda \quad , \tag{5.42}$$

where $\beta = 1/(k_{\mathrm{B}}T)$. The Fourier transforms of the two correlation functions are related via[318]

$$\int_{-\infty}^{\infty}\exp(-i\omega t)C^{\mathrm{qm}}_{AA}(t)\mathrm{d}t = \frac{\beta\hbar\omega}{1-\exp(-\beta\hbar\omega)}\int_{-\infty}^{\infty}\exp(-i\omega t)\tilde{C}_{AA}(t)\mathrm{d}t \quad . \tag{5.43}$$

The Kubo-transformed correlation function and the classical correlation function have the same symmetry properties[316, 318] suggesting that it is better to identify the classical correlation

function with $\tilde{C}_{AA}$ (rather than with C_{AA}^{qm}). In order to arrive at $G(\omega)$ (see Eq. 5.41) one then has to multiply the Fourier transform by

$$Q = \frac{\beta\hbar\omega}{1 - \exp(-\beta\hbar\omega)} \quad , \tag{5.44}$$

where Q is a so-called quantum correction factor. There have also other quantum correction factors been proposed[319], but Eq. 5.44 is the most widely used one and has been shown to yield the best results[303, 312, 313, 319–322]. With this the IR absorption coefficient becomes:

$$\alpha(\omega) \propto \omega^2 \int_{-\infty}^{\infty} \langle \boldsymbol{\mu}(t_0 = 0) \cdot \boldsymbol{\mu}(t) \rangle_{t_0} \, \mathrm{e}^{-i\omega t} \mathrm{d}t \quad . \tag{5.45}$$

In a molecular dynamics run the time zero t_0 can be set arbitrarily. As the system should be ergodic, one can replace the ensemble average by a time average denoted as $<>_{t_0}$, where each time step can be used as $t_0 = 0$ in order to calculate this average. This issue will be assessed in more detail in Chapter 6 again.

One can now exploit the properties of the Fourier transform regarding its time derivative. If the Fourier transform of a function $f(t)$ yields $g(\omega)$, for the Fourier transform of $\frac{\mathrm{d}}{\mathrm{d}t} f(t)$ it holds that $\mathcal{F}\left(\frac{\mathrm{d}}{\mathrm{d}t} f(t)\right) = i\omega g(\omega)$. With this, Eq. 5.45 can be rewritten in terms of the time-autocorrelation function of the time derivative of the dipole moment, which has proven numerically advantageous[323]:

$$\alpha(\omega) \propto \int_{-\infty}^{\infty} \langle \dot{\boldsymbol{\mu}}(0) \cdot \dot{\boldsymbol{\mu}}(t) \rangle_{t_0} \, \mathrm{e}^{-i\omega t} \mathrm{d}t \quad . \tag{5.46}$$

Furthermore, the autocorrelation function is a real quantity and symmetric in time so that the integral can be simplified to

$$\alpha(\omega) \propto \int_{0}^{\infty} \langle \dot{\boldsymbol{\mu}}(0) \cdot \dot{\boldsymbol{\mu}}(t) \rangle_{t_0} \cos(\omega t) \mathrm{d}t \quad . \tag{5.47}$$

This is the formula used to calculate IR spectra from MD simulations in this thesis.

Part II

Large polyalanine-based peptides: structure and spectroscopy

6 IR SPECTRA FROM *ab initio* MD

For the peptides studied in this work, we calculate infrared (IR) spectra from *ab initio* molecular dynamics (MD) simulations and compare them to experimental infrared multiphoton dissociation (IRMPD) data. In order to obtain meaningful and converged spectra many practical details need to be taken into account, which are assessed in this chapter. As a benchmark system, we use the peptide Ac-Ala$_4$-Lys(H^+). Specifically, we concentrate on the lowest-energy conformation obtained from a previous first-principles structure search (density-functional theory (DFT) with the PBE+vdW functional) by Rossi and co-workers[17, 28]. It is depicted in Fig. 6.1. The reasons for choosing this peptide as a test system are two fold. On the one hand, it is an alanine-based peptide containing a lysine residue, i.e., it is very similar to the other peptides studied in this thesis. On the other hand, it is relatively small (70 atoms) so that benchmark calculations become affordable. However, the simulations performed in this chapter are still expensive.[1] We assess the convergence of the spectra with increasing simulation time, number of runs, and sensitivity to different conformers. Additionally, we also study the influence of the time step used in the simulations. The IR spectra are calculated from the Fourier transform of the autocorrelation function of the dipole time derivative as explained in Section 5.3. We use the PBE+vdW functional (see Section 3.5.3) and the simulations are performed in the *NVE* ensemble with $\langle T \rangle = 300\,\mathrm{K}$ and *tight* computational settings. In principle, the IR spectra should be derived from constant-temperature MD simulations (canonical ensemble). However, in order to maintain a certain temperature, thermostats make changes to the atomic velocities, which can affect the dynamic correlations of the system, preventing the extraction of meaningful vibrational information. In the thermodynamic limit, the microcanonical (*NVE*) and the canonical ensemble (*NVT*) become identical. This means that for systems with a large number of degrees of freedom, as for the peptides studied here, it is a good approximation to derive the IR spectra from MD simulations in the *NVE* ensemble after a previous thermostatted equilibration of the system at the target temperature (here: 300 K). This is also the standard approach used in the literature.

6.1 WAVE-FUNCTION EXTRAPOLATION

In Born-Oppenheimer (BO) MD simulations, the forces acting on the nuclei have to be calculated at each time step

$$\boldsymbol{F}_I = -\nabla_I V_{\mathrm{BO}}(\{\boldsymbol{R}_I\}) \quad . \tag{6.1}$$

[1] Using a time step of 1 fs and the tightest accuracy settings for the self-consistency cycle (see Tab. 6.1), a simulation of 30 ps takes approximately ten days on 256 cores of the "aims" cluster at the Garching Computing Centre (Intel Xeon octacore nodes).

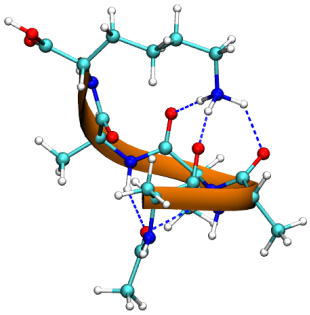


Figure 6.1: The Ac-Ala$_4$-Lys(H$^+$) conformer used as a test system here. This conformer is the lowest-energy conformer obtained from a previous first-principles structure search by Rossi and co-workers[17, 28].

For this, one has to find the self-consistent solution to the Kohn-Sham equations, i.e., a full self-consistent field (SCF) cycle has to be performed at each time step. To minimize the number of SCF iterations, and therewith the computational cost, one needs to find a good initial guess for the electronic degrees of freedom. One evident way is to extrapolate this guess from the optimized electronic degrees of freedom of previous time steps.[2] However, this approach typically suffers from a systematic long-term drift of the total energy, especially if just the optimized value from the last time step is used[325–328]. In short, the problem (discussed in depth by Niklasson and co-workers[328–330]) is that the SCF procedure is irreversible and self consistency is never fully reached, resulting in residual errors in the forces. If the starting point for the SCF procedure is not chosen in a time-reversible fashion, as obviously the case when the optimized value from the previous time step is used, this will break the time reversibility of the nuclear dynamics due to the error in the forces, resulting in energy-conservation problems[266, 326]. In order to overcome this problem, one needs to tighten the self-consistency criteria and/or use higher-order extrapolation methods[327].[3] In the following, we will present benchmarks for several sets of self-consistency criteria and extrapolation schemes to identify reasonable settings for the MD simulations needed in this thesis to derive IR spectra for the systems under study.

In a practical calculation, one has to decide first which quantity should be extrapolated. In `FHI-aims`[4] the contra-covariant density matrix is chosen, as suggested by Kühne *et al.*[331] The density matrix $\underline{\underline{P}}$ is defined as

$$\underline{\underline{P}} = \underline{\underline{C}}\,\underline{\underline{G}}\,\underline{\underline{C}}^T \quad . \tag{6.2}$$

The columns of the matrix $\underline{\underline{C}}$ are the coefficients of the Kohn-Sham eigenvectors as defined in Eqs. 3.62 and 3.63. The matrix $\underline{\underline{G}}$ is diagonal, with the diagonal containing the occupation numbers. If the density matrix is multiplied with the overlap matrix $\underline{\underline{S}}$ (see Eq. 3.65), one arrives at the contra-covariant density matrix $\underline{\underline{\mathcal{P}}}$

$$\underline{\underline{\mathcal{P}}} = \underline{\underline{P}}\,\underline{\underline{S}} \quad . \tag{6.3}$$

If $\underline{\underline{\mathcal{P}}}$ acts on $\underline{\underline{C}}$, it projects out the occupied subspace

$$\underline{\underline{\mathcal{P}}}\,\underline{\underline{C}} = \underline{\underline{C}}\,\underline{\underline{G}} \quad . \tag{6.4}$$

After each time step, $\underline{\underline{\mathcal{P}}}$ is constructed and stored. Based on the values at N preceding time steps

[2] Extrapolating the electronic degrees of freedom resembles to some extent the idea of Car-Parrinello (CP)MD. For further details, the interested reader is referred, e.g., to Ref. [324].
[3] Another option is to use a time-reversible extrapolation scheme, as proposed by Niklasson and co-workers[328–330]
[4] The wave-function extrapolation scheme was implemented by Jürgen Wieferink.

one can define an extrapolator $\underline{\underline{\tilde{\mathcal{P}}}}(t)$ that estimates the value of $\underline{\underline{\mathcal{P}}}$ at time t:

$$\underline{\underline{\tilde{\mathcal{P}}}}(t) := \sum_{m=1}^{N} B_m \underline{\underline{\mathcal{P}}}(t - mh) \qquad , \tag{6.5}$$

where h denotes the time step Δt and B_m is the extrapolation coefficient. All of the systems treated in this thesis have a well-defined gap that separates occupied and unoccupied states, so one can obtain the new (extrapolated) eigencoefficients by

$$\underline{\underline{\tilde{C}}}^{\mathrm{occ}}(t) = \underline{\underline{\tilde{\mathcal{P}}}}(t)\underline{\underline{C}}^{\mathrm{occ}}(t - h) = \sum_{m=1}^{N} B_m \underline{\underline{\mathcal{P}}}(t - mh)\underline{\underline{C}}^{\mathrm{occ}}(t - h) \tag{6.6}$$

and

$$\underline{\underline{\tilde{C}}}^{\mathrm{unocc}}(t) = \left[\underline{\underline{\mathbb{I}}} - \underline{\underline{\tilde{\mathcal{P}}}}(t)\right] \underline{\underline{C}}^{\mathrm{unocc}}(t - h) \qquad . \tag{6.7}$$

Afterwards the new eigenvectors have to be orthonormalized before calculating the charge density.

In order to understand how the extrapolation is done, consider a function $f(t)$, for which its value at time t should be extrapolated based on the knowledge of its values at preceding time steps (f could be any component of $\underline{\underline{\mathcal{P}}}$). Let $\tilde{f}(t)$ be the extrapolator:

$$\tilde{f}(t) := \sum_{m=1}^{N} B_m f(t - mh) \qquad , \tag{6.8}$$

where h again denotes the time step Δt and B_m is the extrapolation coefficient. One can estimate the error of the extrapolation by expanding $f(t - mh)$ in a Taylor series around the extrapolated time t[332, 333]:

$$\begin{aligned} \Delta\tilde{f}(t) = \tilde{f}(t) - f(t) &= \sum_{m=1}^{N} B_m \sum_{k=0}^{\infty} \left(\frac{\mathrm{d}^k f}{\mathrm{d}t^k}\right)_t \frac{(-mh)^k}{k!} - f(t) \\ &= \sum_{k=0}^{\infty} \left(\frac{\mathrm{d}^k f}{\mathrm{d}t^k}\right)_t \frac{h^k}{k!} \underbrace{\left\{\sum_{m=1}^{N} B_m (-m)^k - \delta_{k,0}\right\}}_{A_k} \qquad . \end{aligned} \tag{6.9}$$

By solving a linear equation in the matrix $(-m)^k$ one can choose the B_m such that N values of A_k are zero. With decreasing time step h, the absolute error is minimized if the coefficients A_k with $k = 0, 1, \ldots, N-1$ are chosen to vanish. However, for the time reversibility of the error $\Delta\tilde{f}(t)$, only A_k with odd values of k are relevant. Thus, to enhance the time reversibility of the extrapolation one could also choose the first N coefficients A_k with odd k to be zero.

In the following we perform a series of benchmarks for different extrapolation schemes and different self-consistency accuracy settings. Convergence of the SCF cycle is said to be achieved if the specified quantities, such as total energy or sum of eigenvalues, varies between two subsequent iterations by less than the given convergence criterion. In `FHI-aims`, the accuracy of the convergence of the SCF cycle can be adjusted by the following keywords:

- `sc_accuracy_eev`: convergence criterion for the sum of eigenvalues (eV)

Table 6.1: Different sets of convergence criteria for the self-consistency cycle used for the test series in this chapter.

Criterion	Settings 0	Settings 1	Settings 2
`sc_accuracy_rho` (electrons)	10^{-3}	10^{-4}	10^{-5}
`sc_accuracy_eev` (eV)	10^{-2}	10^{-3}	10^{-4}
`sc_accuracy_etot` (eV)	10^{-5}	10^{-5}	10^{-6}
`sc_accuracy_forces` (eV/Å)	not checked	not checked	$5 \cdot 10^{-4}$

- `sc_accuracy_etot`: convergence criterion for the total energy (eV)
- `sc_accuracy_rho`: convergence criterion for the charge density; the convergence criterion refers to the volume-integrated root-mean square change of the charge density measured in electrons
- `sc_accuracy_forces`: convergence criterion for the maximum value of the forces (eV/Å).

For our tests, we used three different SCF settings, labelled as settings 0, 1, and 2. The according accuracy criteria are tabulated in Tab. 6.1. Set 0 has the lightest criteria and set 2 the tightest. We also employed different extrapolation schemes pno, where the name pno refers to an n point scheme, where the values of n previous time steps are used for the extrapolation. The order o specifies the value of k up to which all A_k are chosen to be zero. As the minimum value of k is zero, o can at most be $n-1$. The remaining degrees of freedom are used to choose A_k with odd k to vanish in order to enhance time reversibility. The choice p10 corresponds to using the one-particle coefficients of the previous time step to initialize the self-consistency cycle of the next.

As mentioned in the beginning of this chapter, we used the peptide Ac-Ala$_4$-Lys(H$^+$) as a benchmark system. All simulations were performed in the *NVE* ensemble after a previous equilibration of the molecule at 300 K. In all cases, *tight* computational settings were used.

6.1.1 SCF ACCURACY SETTINGS 0

Based on the SCF accuracy settings 0, we simulated four trajectories with the extrapolation schemes p10, p21, p31, and p32 up to a length of 3 ps using a time step of 1 fs. Table 6.2 lists the average number of SCF iterations needed to obtain convergence (based on the criteria of settings 0). When using the eigencoefficients from the last time step to initialize the SCF cycle (p10), the average amount of iterations needed to converge is 8.5. When including the last two time steps into the extrapolation (p21), this decreases to 8.0. Note that p21 is the same as p20. The next option is to take the converged values of three time steps for the extrapolation into account. There are two possibilities now, namely p32 and p31 (note again that p30 is the same as p31). While p32 should give the better estimate of the eigencoefficients, p31 yields an error in the forces that has a reduced time irreversibility. This is also confirmed by the results. Indeed, p32 yields a (slightly) lower number of SCF iterations that are needed for convergence than p31, namely 7.5 versus 7.6. The reduced time irreversibility of the error by using scheme p31 yields a smaller drift in the total energy as illustrated in Fig. 6.2, which is the smallest drift of all the tested schemes. Still the drift is about 70 meV for 3 ps or 23 meV/ps. This is much too high considering that we intend to perform simulations of about 30 ps length and that the drift should not exceed a few meV during the simulation as different conformational families are only

Table 6.2: Average number of SCF iterations per time step for SCF accuracy settings 0.

Extrapolation method	p10	p21	p31	p32
Av. SCF iterations per time step	8.4	8.0	7.6	7.5

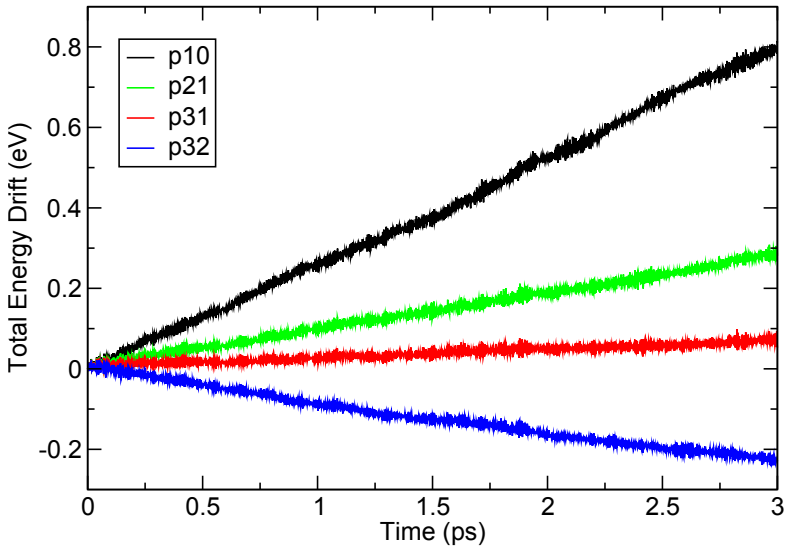


Figure 6.2: Energy drift for a 3 ps *NVE* run of Ac-Ala$_4$-Lys(H$^+$) using SCF accuracy settings 0.

separated by the order of 10 meV[17, 28]. However, just using the eigencoefficients from the last SCF cycle (p10) yields a drift of about 800 meV already after 3 ps.

We here compare the number of SCF iterations needed for the different extrapolation schemes to achieve the same level of convergence. If one would like to compare the number of SCF iterations needed to achieve the same average energy drift, the result would obviously differ from Tab. 6.2. In order to achieve the same drift with a p10 extrapolation as with p31 this would need more SCF iterations than the average 8.4 iterations given in Tab. 6.2. As a rough estimate it would be less than 10.8 (see the following subsection).

6.1.2 SCF ACCURACY SETTINGS 1

As we have found the extrapolation scheme p31 to be the most efficient, we will in the following concentrate on a comparison between p10, i.e., using the wave function of the previous time step as input for the next SCF cycle, and p31. We now use settings 1 (see Tab. 6.1) as the accuracy settings for the SCF cycle and again perform simulations in the *NVE* ensemble of 3 ps length with a time step of 1 fs. As can be seen from Tab. 6.3, employing scheme p31 reduces the average number of iterations per SCF cycle by about 1 compared to p10, from 10.8 to about 9.8. Fig. 6.3 shows the evolution of the total energy for both p31 and p10. According to a linear fit to the data, the drift for p10 is 6 meV/ps. This is smaller than the drift we found for p31 based on the less accurate SCF accuracy settings 0 (23 meV/ps, see previous subsection). With SCF accuracy

Table 6.3: Average number of SCF iterations per time step for SCF accuracy settings 1.

Extrapolation method	p10	p31
Av. SCF iterations per time step	10.8	9.8

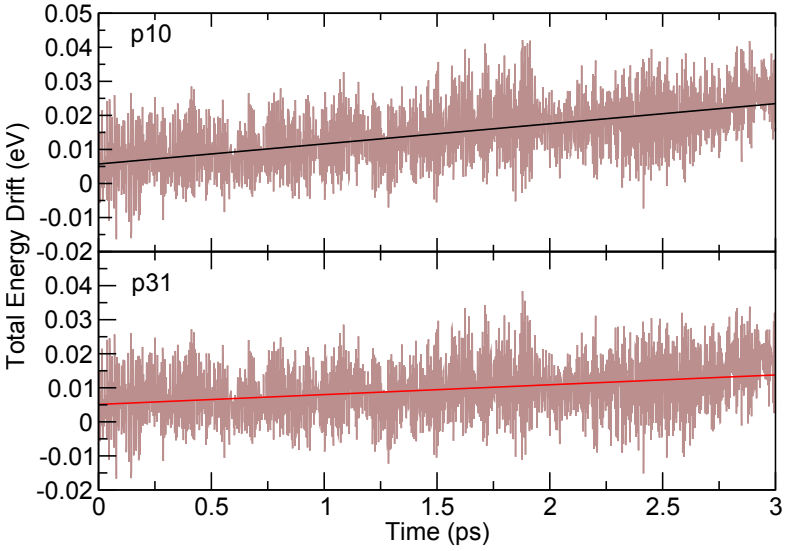


Figure 6.3: Energy drift for a 3 ps *NVE* run using SCF accuracy settings 1. The straight lines indicate a linear fit to the data.

Table 6.4: Average number of SCF iterations per time step for SCF accuracy settings 2.

Extrapolation method	p10	p31
Av. SCF iterations per time step	13.8	12.9

settings 1, however, the scheme p31 yields a smaller drift than p10, namely about 3 meV/ps. Still, for a run of 30 ps length we would thus have to expect a drift of about 90 meV. In the lowest 100 meV regime of Ac-Ala$_4$-Lys(H$^+$) there are at least five different structure types[17, 28]. The second-lowest structure type has an energy difference compared to the lowest-energy conformer of about 10 meV so that a drift of the order of or less than 10 meV would be desirable.

6.1.3 SCF ACCURACY SETTINGS 2

When increasing the accuracy of the SCF cycle settings even more to settings 2 (see Tab. 6.1), we still find the extrapolation scheme p31 to need approximately one less iteration to achieve convergence than p10. This is illustrated in Tab. 6.4. As depicted in Fig. 6.4, the drift for the extrapolation scheme p10 is basically the same as for p31. A linear fit to the data yields a drift of about 1 meV/ps. In order to see how the drift evolves, we increased the run using p31 to a simulation time of 40 ps. Figure 6.5 illustrates the drift in the total energy and the evolution of the temperature, which oscillates around 300 K. Although the energy oscillates with an amplitude of about 30 meV, the average drift is about or even less than 10 meV. As we mentioned earlier, different conformational families of Ac-Ala$_4$-Lys(H$^+$) are separated by the order of 10 meV[17, 28]. A drift of about 10 meV or less for a 40 ps simulation is hence considered to be sufficiently small. We will thus use SCF accuracy settings 2 in the following and throughout this thesis to calculate IR spectra from *ab initio* MD runs.

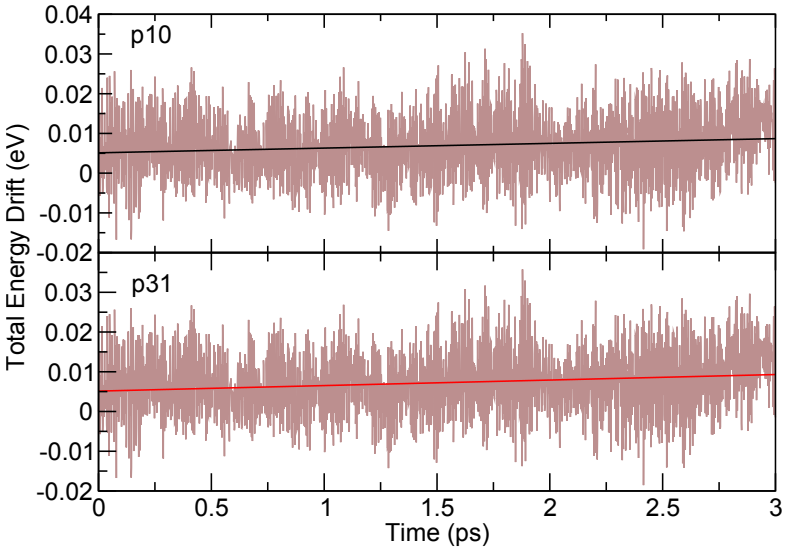


Figure 6.4: Energy drift for a 3 ps *NVE* run using SCF accuracy settings 2. The straight lines indicate a linear fit to the data.

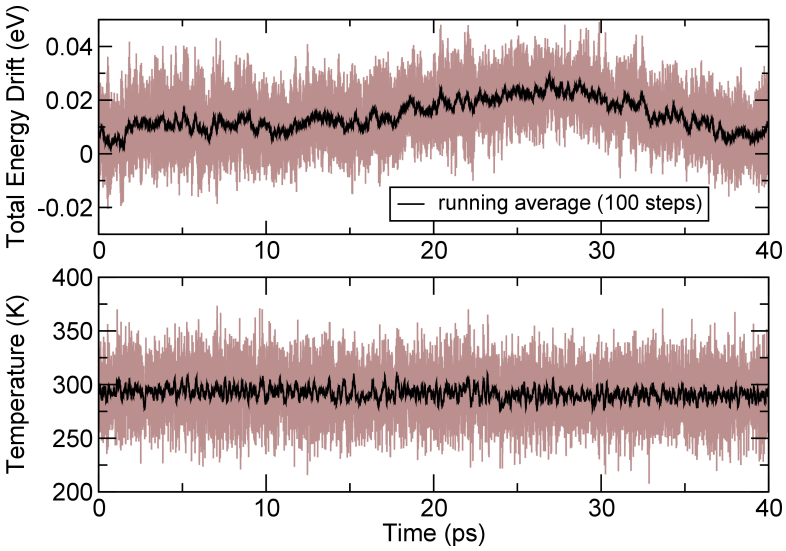


Figure 6.5: Evolution of the total energy and the instantaneous kinetic temperature (see Eq. 4.9) as a function of simulation time for an *NVE* simulation for Ac-Ala$_4$-Lys(H^+) with $\langle T \rangle = 300$ K. We used the SCF accuracy settings 2 and p31 as the extrapolation scheme. The brown points are the actual data points, while the black lines illustrate running averages over 100 fs.

6.2 PENDRY RELIABILITY FACTOR

In the present chapter, we aim to study the convergence of the IR spectrum of Ac-Ala$_4$-Lys(H$^+$) with respect to different parameters, such as the length or the time step of the simulation. For this, we need to compare different IR spectra to each other and judge their similarity. Even though the human eye is capable to capture many different features at the same time, a visual comparison is still subjective. Furthermore, the eye is often (unconsciously) attracted by single outstanding differences such as large differences in intensity or a missing or split peak[334] rather than by smaller shifts in the peak positions[335]. It would thus be desirable to have an unambiguous and quantitative measure of the agreement between two spectra, where it is known how exactly the comparison is performed and which features are taken into account or assigned particular weight. Such *reliability factors* (R-factors) are widely used in the context of X-ray diffraction and low-energy electron diffraction (LEED)[334–337]. They are numbers calculated by a given formula or prescription that tries to account for the differences and similarities between two datasets. Thus, the R-factors give a measure for the degree of correspondence (reliability). As there are many features that have to be taken into account and that can be weighed differently in the comparison, such as agreement of absolute/relative intensity, peak positions, shoulders and so on, there are different possibilities for constructing an R-factor[334–337]. With the reasons being explained in the following, in this thesis, we use the reliability factor proposed by John Pendry[337], which is widely used in the context of X-ray diffraction and LEED, but has also been successfully employed for IR spectroscopy[17, 338]. The Pendry R-factor (R_P, or just referred to as R-factor in this thesis) places emphasis on the positions of the peaks rather than on the peak intensities. This feature is particularly desirable for us as the experimental spectra to which we compare our theoretical spectra in this work are measured by IRMPD (see Chapters 9 and 12). The multiple-photon absorption process can affect the IR intensities (as will be discussed in Section 7.1.2), whereas the peak positions should match. Furthermore, due to its insensitivity to absolute peak intensities, R_P attributes equal weight to deviations of the peak positions in high-intensity regions and low-intensity regions of the spectrum. In this way, both discrepancies in the high-intensity amide I and II bands and in the less intense, but as discussed in Section 5.2 also structure sensitive[308–311], amide III band are accounted for equally by R_P.

Pendry achieved the special focus on the peak positions rather than on the intensity by comparing two spectra based on auxiliary functions defined as

$$Y(\overline{\nu}) = \frac{L^{-1}}{L^{-2} + V_0^2} \qquad \text{with} \qquad L(\overline{\nu}) = \frac{1}{I(\overline{\nu})} \frac{\mathrm{d}I}{\mathrm{d}\overline{\nu}} \quad . \tag{6.10}$$

$I(\overline{\nu})$ is the intensity as a function of the wavenumber $\overline{\nu}$ and V_0 is the approximate half width of the peaks. When comparing two spectra $I_1(\overline{\nu})$ and $I_2(\overline{\nu})$, the value of the Pendry reliability factor is then calculated via:

$$R_\mathrm{P} = \frac{\int (Y_1 - Y_2)^2 \mathrm{d}\overline{\nu}}{\int (Y_1^2 + Y_2^2) \mathrm{d}\overline{\nu}} \quad . \tag{6.11}$$

This equation yields $R_\mathrm{P} = 0$ for perfect correlation (agreement) between the spectra, $R_\mathrm{P} = 1$ if there is no correlation, and $R_\mathrm{P} = 2$ if the spectra are perfectly anticorrelated. The idea behind this prescription is the following[337]. One can think of the spectrum $I(\overline{\nu})$ to be approximately

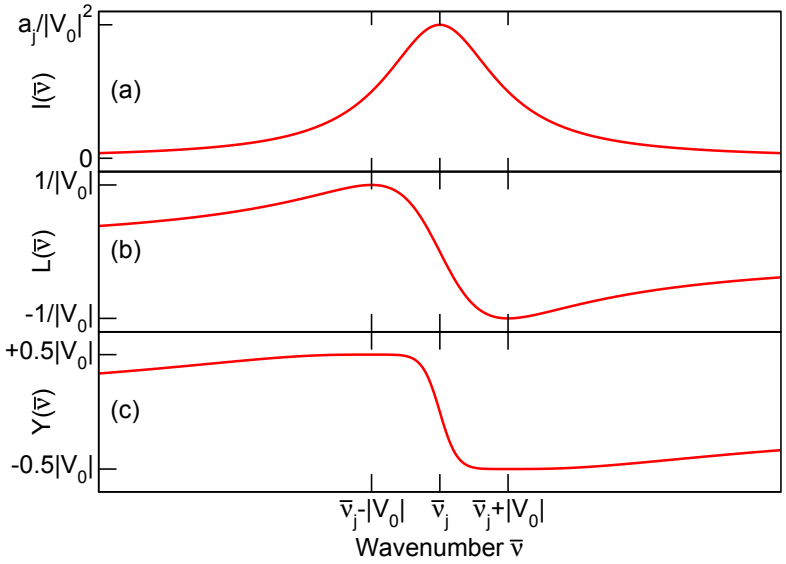


Figure 6.6: Illustration of a Lorentzian peak at position $\overline{\nu}_j$ (a), its logarithmic derivative L (b), and the Pendry Y function (c).

composed of a sum of Lorentzian peaks:

$$I \simeq \sum_j \frac{a_j}{(\overline{\nu} - \overline{\nu}_j)^2 + V_0^2} \quad , \tag{6.12}$$

where a_j/V_0^2 defines the amplitude of the Lorentzian peak j and V_0 is the half width of the peak.[5] The logarithmic derivative of the intensity,

$$L(\overline{\nu}) = \frac{\mathrm{d}I}{\mathrm{d}\overline{\nu}}/I \simeq \sum_j \frac{-2(\overline{\nu} - \overline{\nu}_j)}{(\overline{\nu} - \overline{\nu}_j)^2 + V_0^2} \quad , \tag{6.13}$$

has extremes of amplitude

$$L = \pm 1/|V_0| \qquad \text{at} \qquad \overline{\nu} = \overline{\nu}_j \mp |V_0| \quad . \tag{6.14}$$

This is illustrated in Fig. 6.6(b) for one model Lorentzian peak. If the different peaks are widely spaced, L is thus insensitive to the amplitudes. If the peaks lie close to each other, L retains some sensitivity. A direct comparison of the L functions of two spectra has the problem that the L function diverges if the intensity goes to zero, i.e., zeroes in the intensities gain too much emphasis. Instead, Pendry suggested to use the Y-function as defined in Eq. 6.10 for the comparison as it gives similar weights to zeroes and peaks in the spectra. If $L = \pm 1/|V_0|$, i.e., if $\overline{\nu} = \overline{\nu}_j \mp |V_0|$, Y becomes $Y = \pm 1/2|V_0|$ as illustrated in Fig. 6.6(c) for the example of a Lorentzian peak.

In this chapter, we use the Pendry R-factor to compare the wavenumber region between 1000 and 1800 cm^{-1} as this is the region considered by the experiments.

[5] In the experimental IRMPD spectrum the natural bandwidth is further broadened by mechanisms such as the multiple-photon absorption process and the finite bandwidth of the laser, which will be explained in Section 7.1.2. Thus, a Lorentzian peak shape is not necessarily the best shape to use, but sufficient for what is needed here.

6.2.1 RIGID SHIFTS ALONG THE WAVENUMBER AXIS

Vibrational frequencies calculated in the harmonic approximation usually overestimate the experimental values, due to the neglect of anharmonicity and inaccuracies of the theoretical approach[339]. This is commonly accounted for by employing scaling factors for the theoretical spectra, which depend on the wavenumber region, but also on the density-functional approximation used[340]. When including anharmonicity in the theoretical spectra by calculating them from *ab initio* MD simulations, *rigid* (but not variable) shifts still occur between theory and experiment[17, 303, 338]. These are most probably due to systematic mode softening caused by the generalized gradient approximation (GGA) functional and by the neglect of nuclear quantum effects[17]. For this reason, the Pendry R-factor between the experimental and theoretical spectrum is normally calculated including the rigid shift Δ of the theoretical spectrum along the wavenumber axis that yields the best agreement with experiment. In this chapter, we only compare theoretical spectra to theoretical spectra, which is why we do not include a rigid shift here.

6.2.2 CALCULATION OF IR SPECTRA

We now turn to the calculation of IR spectra from *ab initio* MD simulations. As a benchmark system, we again use the peptide Ac-Ala$_4$-Lys(H$^+$). As mentioned earlier in this chapter, for this system, a conformational search based on DFT (PBE+vdW functional) has been previously performed by Rossi and co-workers[17, 28]. In their analysis, all structures were sorted into conformational families according to their hydrogen-bonding connection pattern.[6] The lowest-energy member of each family was chosen as the representative of each family. We here concentrate on the lowest-energy family with its representative depicted in Fig. 6.1. The IR spectra are based on simulations of this conformational family in the *NVE* ensemble using the PBE+vdW functional. Prior to the *NVE* run, the peptides are equilibrated at 300 K by performing thermostatted runs at this target temperature for at least 3 ps. During the *NVE* runs the instantaneous kinetic temperature (see Eq. 4.9) thus oscillates around 300 K as illustrated in Fig. 6.5. All calculations are performed using *tight* computational settings (cf. Section 3.6). As explained in Section 5.3, we calculate the IR spectra by taking the Fourier transform of the dipole time-derivative autocorrelation function as given in Eq. 5.47. For this purpose, the time derivative of the dipole is calculated from finite differences. From an MD simulation consisting of N time steps, we thus obtain $N-1$ values for $\dot{\boldsymbol{\mu}}$ due to the finite difference evaluation of the time derivative. All of these $N-1$ times can be used as time zero to compute the average in $\langle \dot{\boldsymbol{\mu}}(0) \cdot \dot{\boldsymbol{\mu}}(t) \rangle$. For $t = 0$, $\langle \dot{\boldsymbol{\mu}}(0) \cdot \dot{\boldsymbol{\mu}}(0) \rangle$ will be the average of $N-1$ data points. For $t = 1$ there will be only $N-2$ possibilities of $\dot{\boldsymbol{\mu}}(0) \cdot \dot{\boldsymbol{\mu}}(1)$ and so on. The available statistics decreases with time t. For $\dot{\boldsymbol{\mu}}(0) \cdot \dot{\boldsymbol{\mu}}(N-1)$ there will be only one data point available. In order to reduce the noise that this decrease in statistics produces, we cut the autocorrelation function after $t = N/2$. The autocorrelation function is then padded with zeroes to increase the resolution of the Fourier transform. One example is illustrated in Fig. 6.7(a), which shows $\langle \dot{\boldsymbol{\mu}}(0) \cdot \dot{\boldsymbol{\mu}}(t) \rangle$ obtained from a 40 ps *NVE* run with $\langle T \rangle = 300$ K for Ac-Ala$_4$-Lys(H$^+$) (cf. Fig. 6.5). The autocorrelation function is cut after half the time (20 ps) and then padded with zeroes. It does not decay to zero perfectly

[6] A hydrogen bond was considered to be present if the distance between a hydrogen atom and an acceptor oxygen was less than 2.5 Å.

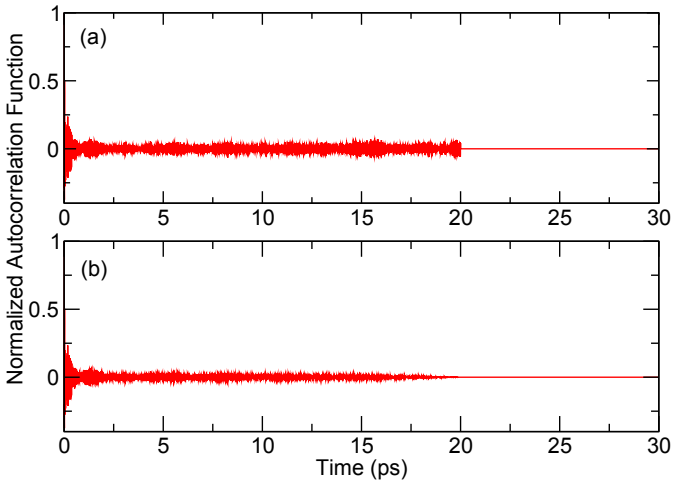


Figure 6.7: (a) Normalized dipole time-derivative autocorrelation function obtained from a 40 ps *NVE* run with $\langle T \rangle = 300\,\mathrm{K}$ for Ac-Ala$_4$-Lys(H$^+$) and (b) multiplied by a triangular windowing function.

(this would only happen after much longer simulation times), which produces noise when calculating the Fourier transform. This is a well-known problem in signal processing and solved by the help of so-called windowing functions. We here multiply the autocorrelation function by a triangular windowing function as illustrated in Fig. 6.7(b) (see also the PhD thesis of Mariana Rossi[17] for other windowing functions).

6.2.3 Influence of the convolution

The top panel of Fig. 6.8 shows the result of the raw Fourier transform of the dipole time-derivative autocorrelation function for the *NVE* run at $\langle T \rangle = 300\,\mathrm{K}$ of 40 ps length discussed in the previous subsection (cf. Fig. 6.7 and Fig. 6.5). The Pendry R-factor is sensitive to small kinks in the spectra and the raw Fourier transform is very noisy. To smooth it, we convolute the raw spectrum with a Gaussian

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad . \tag{6.15}$$

The other reason for the convolution is that we normally aim at a comparison with experimental data (although for the test system Ac-Ala$_4$-Lys(H$^+$), we do not have experimental IR spectra available). In experiment, the peaks are broadened by the multiple-photon absorption process, but also due to the finite bandwidth of the laser (see Section 7.1.2). The full width at half maximum (FWHM) of the laser is about 1–2% of the wavenumber[341–343], corresponding to a $\sigma = 0.004$–0.008 times the wavenumber. For this reason, we convolute the spectra with a Gaussian with a variable width depending on the wavenumber. Apart from the raw spectrum, Fig. 6.8 also shows the spectra with convolutions between $\sigma = 0.001 \cdot \overline{\nu}$ and $\sigma = 0.1 \cdot \overline{\nu}$. We use a convolution of $\sigma = 0.005 \cdot \overline{\nu}$ for the rest of our benchmarks as the spectrum still retains the important features and is sufficiently smooth. Of course, $\sigma = 0.003 \cdot \overline{\nu}$ retains a bit more resolution. However, close to 1000 cm^{-1} is is still quite wiggly and, as we shall see in Chapters 9 and 12, we found $\sigma = 0.005 \cdot \overline{\nu}$ to give similar resolution to that obtained in experiment.

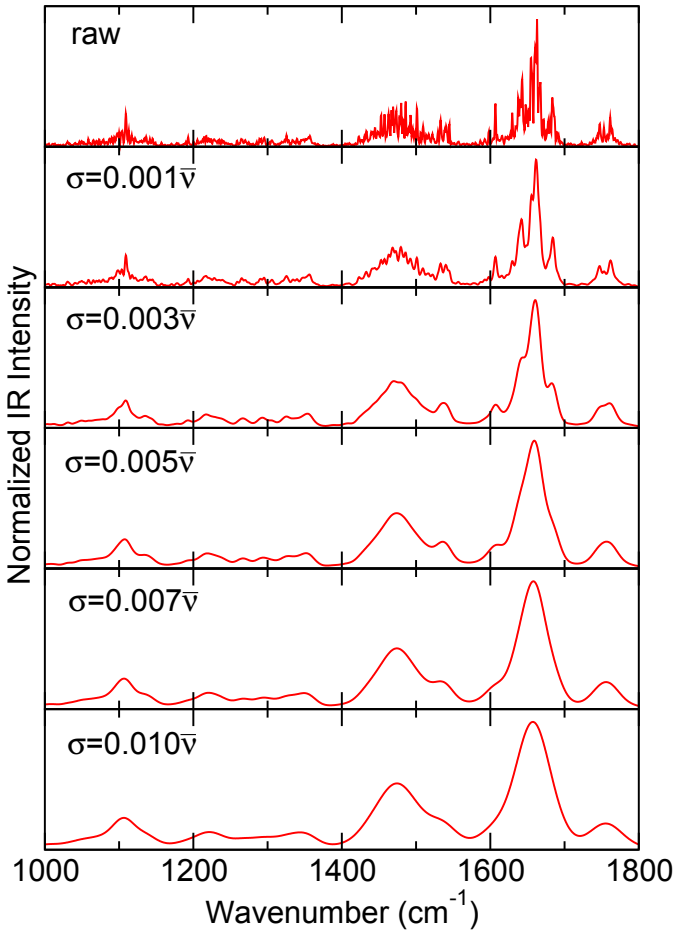


Figure 6.8: Spectrum of Ac-Ala$_4$-Lys(H$^+$) after convolution with a Gaussian of different variable width σ. The spectrum was obtained from an *NVE* run of 40 ps length with a time step of 1 fs and $\langle T \rangle = 300\,\mathrm{K}$.

6.3 Convergence of spectra and sensitivity for different conformers

As a next step, we analyze the convergence of the theoretical spectrum of the lowest-energy conformational family of Ac-Ala$_4$-Lys(H$^+$) with increasing simulation time. Fig. 6.9a) shows the spectra obtained from simulations of 5, 15, 25, 35, and 40 ps length using a time step of 0.75 fs. Using the spectrum obtained after 40 ps as the reference, we calculated the Pendry R-factors of all other spectra. While the spectrum obtained after 5 ps has an R-factor of 0.20, the spectrum after 35 ps deviates only very little with an R-factor of 0.01. Already after 25 ps the peaks look rather converged, which is confirmed by an R-factor of 0.04. In order to analyze the local structural stability of the peptide during the MD run, we plot the evolution of the hydrogen-bonding network with simulation time in Fig. 6.9b). Essentially, it does not change, but stays stable during the run.

In order to calculate the Pendry R-factor one has to choose a reasonable value for V_0, the approximate half widths of the peaks. We here use 10 cm^{-1}, which is the value used throughout this thesis. A Gaussian function with a variable broadening of about $\sigma = 0.005 \cdot \bar{\nu}$ has peak widths of about 5-10 cm^{-1} between 1000 and 2000 cm^{-1}, i.e., choosing V_0 of that order should provide a good estimate. In fact, the R-factor is rather insensitive to this quantity (see Ref. [17]). We demonstrate this based on one example in Fig. 6.10, where we calculate the R-factor between two spectra using different values for V_0.

As discussed in Section 5.3, we assume our peptide simulations to be ergodic, i.e., it should not matter if the dipole time autocorrelation function is averaged over the ensemble or over time. To assess this, we performed several shorter (10 ps) simulations initialized from different starting points. These starting points differ in their exact geometry and their velocities, but all of them belong to the same H-bonding family. We then averaged the autocorrelation functions obtained from the different runs before calculating the IR spectrum. The IR spectra obtained from one, four, eight, and twelve runs, respectively, are shown in Fig. 6.11a). In the same figure, the corresponding R-factors are given, taking the spectrum obtained from twelve runs (12x10 ps) as the reference. After four runs, the spectrum is rather converged with an R-factor of less than 0.1. The IR spectrum derived from a single trajectory depicted in Fig. 6.11a) has an R-factor of 0.18. This is only one example with the R-factors of the IR spectra corresponding to the other eleven single trajectories vary between 0.14 and 0.23.

For a direct comparison, we also show the spectrum obtained from one 25 ps long simulation in Fig. 6.11b) again (labelled as 1x25 ps). The R-factor with respect to the 12x10 ps spectrum is 0.10, which is rather small. From a visual comparison the peak positions also match well. The main difference is that for the 1x25 ps spectrum the amide I band is split, while for the 12x10 ps it is not. Given the good agreement of the two spectra, we conclude that in order to calculate the spectrum of a single conformational family one long (25 ps) MD run should yield decent results. This is also confirmed by observations made for IR spectra derived from force-field MD simulations for the helical peptide Ac-Ala$_{19}$-Lys(H$^+$). Taking the spectrum derived from an *NVE* trajectory of 1 ns lengths as a reference, we see that after 25 ps the spectrum is essentially converged with $R_{\mathrm{P}} = 0.07$ (see Appendix A.1).

It is now important to see how the spectrum of a different conformational family compares to this. For this, we chose the second lowest-energy hydrogen-bonding family from the search by

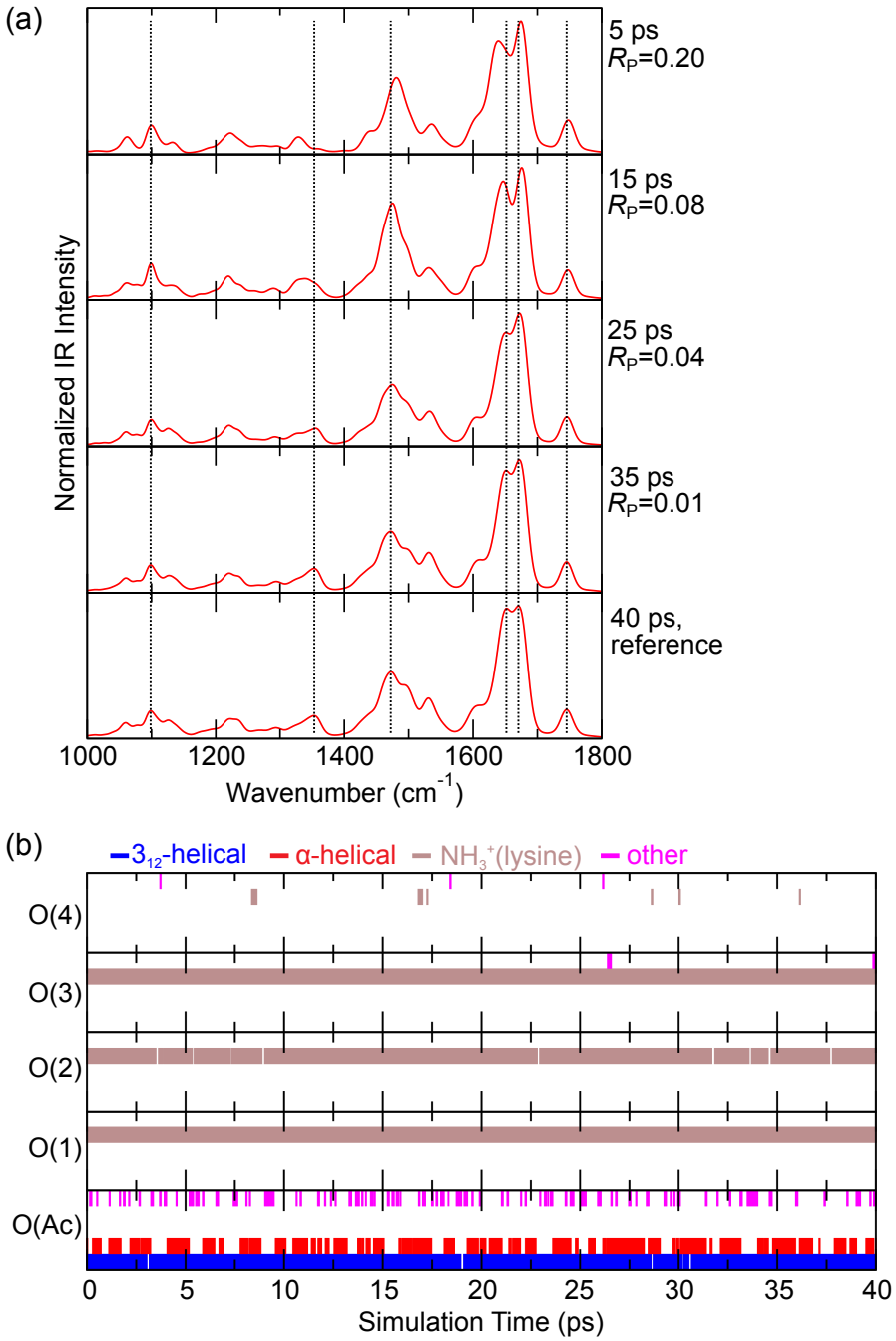


Figure 6.9: a) Convergence of the spectrum for Ac-Ala$_4$-Lys(H$^+$) with the time of the MD trajectory (PBE+vdW, microcanonical ensemble with $\langle T \rangle = 300\,\mathrm{K}$). The spectrum is calculated using a time step of 0.75 fs. The spectrum obtained after 40 ps is taken as the reference for the calculation of the R-factors with the spectra obtained after intermediate simulation times. The spectra are not shifted. Dotted lines serve as a guide to the eye for the peaks of the reference spectrum. b) Hydrogen-bond network evolution with simulation time. Each graph represents the hydrogen-bonding connection for the given oxygen atom. O(Ac) is the oxygen of the acetyl, while O(i) specifies the carbonyl oxygen of the corresponding alanine residue i. The color of the bar denotes the type of hydrogen bond that is formed, namely 3_{10}-helical (blue), α-helical (red), a hydrogen bond to the N-terminal lysine NH_3^+ group (brown), or "other" (pink).

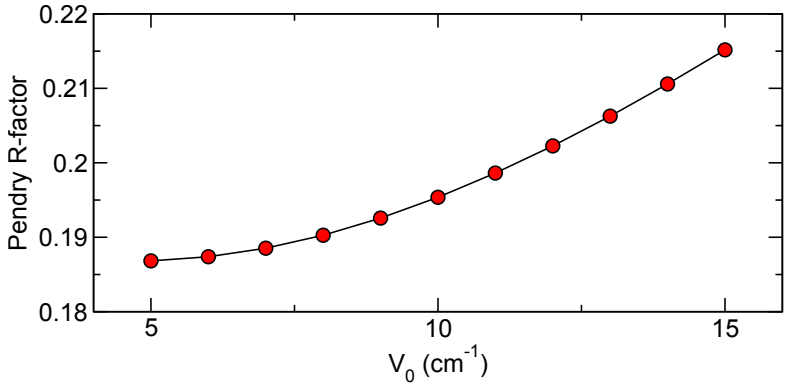


Figure 6.10: Plot of the Pendry R-factor as a function of V_0 demonstrating that the choice of V_0 has minimal influence on R_P. The R-factor is calculated between the two spectra obtained after 5 ps and 40 ps of simulation time from Fig. 6.9a).

Rossi and co-workers[17, 28] with its structure representative depicted next to Fig. 6.11c). The energy difference to the structure representative of family 1, which is depicted next to Fig. 6.11a), is $\Delta E = 10\,\mathrm{meV}$. In family 1, the carbonyl oxygen atoms of the alanine residues 1, 3, and 4 are hydrogen bonded to the lysine NH_3^+ group, while in family 2 it binds to the acetyl oxygen and the carbonyl oxygens of alanine residues 2 and 4. Furthermore, the carbonyl oxygen of the first alanine residue is hydrogen-bonded to the nitrogen atom of the fourth alanine residue. These differences in the conformation of family 1 and 2 are also reflected in their IR spectra. The IR spectrum of family 2 has, compared to the reference spectrum of family 1 [Fig. 6.11(a)], a Pendry R-factor of 0.35. Most importantly, this is significantly higher than the R-factor between the (1x25 ps)- and the (12x10 ps)-spectrum for family 1 ($R_P = 0.1$). Also in a visual comparison the spectra from family 1 and 2 are different. This can be seen in Fig. 6.11 by the vertical lines drawn into the spectra as a guide to the eye, highlighting the peak positions of the reference spectrum for family 1. In summary, the differences between the spectra of the two families are thus both manifested in the comparison based on the Pendry R-factor and on a visual impression.

6.4 Comparison of different time steps

Another issue to be addressed is the time step of the simulation. The smaller the time step, the more accurate the simulation will be. However, with a smaller time step one needs more steps to achieve a given target MD simulation time, which is computationally more expensive. In order to compare the influence of the time step on the IR spectra, we calculated IR spectra from *NVE* simulations of 25 ps length using a time step of 1 fs, 0.75 fs and 0.5 fs. Figure 6.12 illustrates all three spectra together with their R-factors with reference to the spectrum obtained from the simulation with 0.5 fs. The R-factors are 0.18 for a time step of 1 fs and 0.15 for a time step of 0.75 fs. This is a reasonable agreement considering that the spectra obtained from one 25 ps and after 12 runs of 10 ps length have an R-factor of the same order ($R_P = 0.1$). Also the peak positions as indicated by the dotted lines in Fig. 6.12 match rather well. However, the amide-I peak (at around 1650 cm^{-1}) is split for the spectrum obtained from the simulation with time steps of 0.75 fs and 0.5 fs. In the case of the spectrum based on a 0.5 fs time step also the amide-II peak at around 1470 cm^{-1} is split compared to the spectrum obtained from simulations with

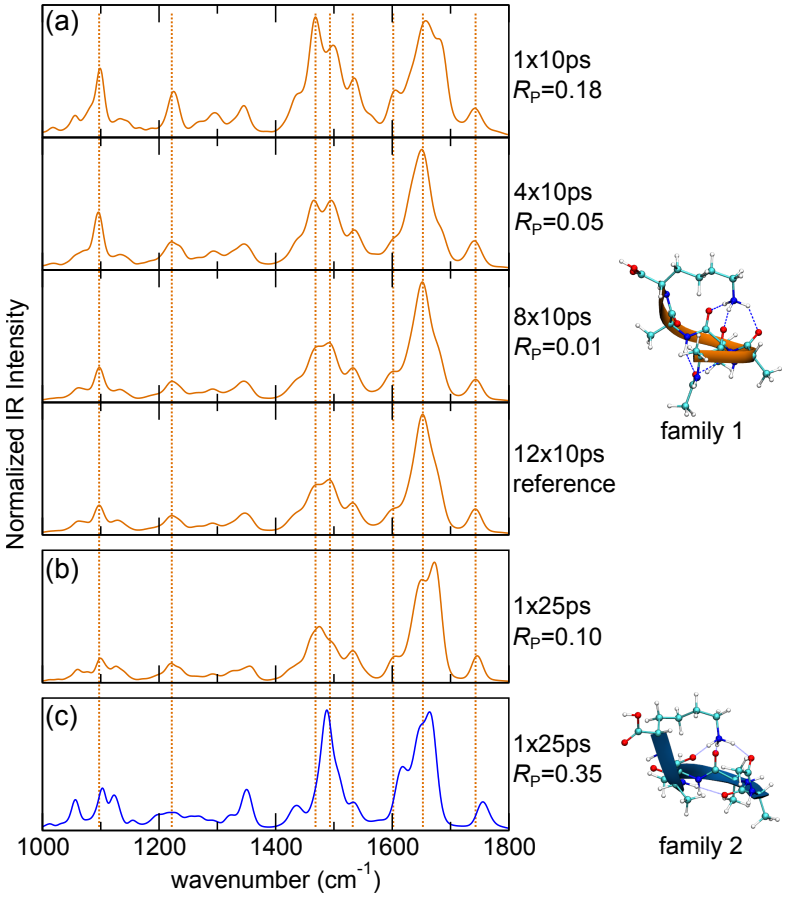


Figure 6.11: a) IR spectra obtained from the dipole time autocorrelation function after averaging over several *ab initio* MD runs of 10 ps length (microcanonical ensemble, PBE+vdW, $\langle T \rangle$ =300 K). All spectra refer to the conformational family 1 illustrated at the right side of the plot, which is the same family that was used for all previous tests. The differences between the spectra are quantified based on the Pendry reliability factor. b) IR spectrum derived from a single 25 ps long MD run. c) IR spectrum for a second family 2 depicted next to the plot. The spectra are not shifted. Vertical lines serve as a guide to the eye to illustrate the peak positions of the reference spectrum for family 1. The time step used was 0.75 fs in all cases.

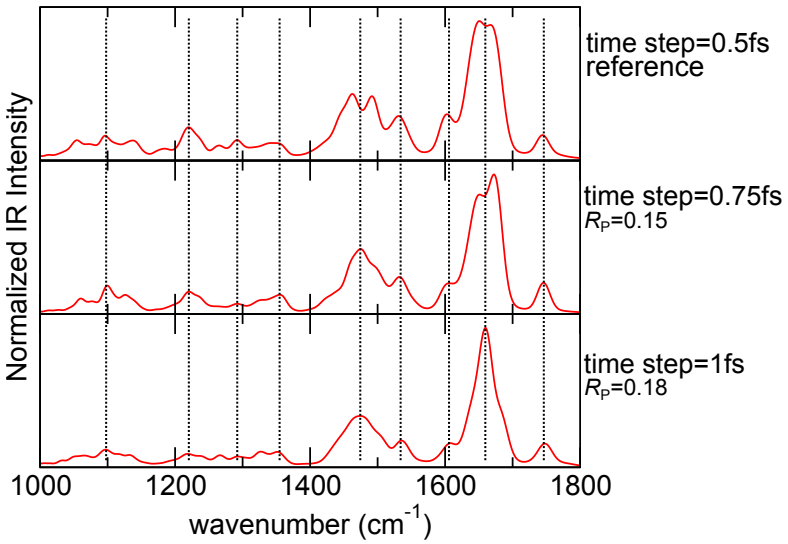


Figure 6.12: IR spectra obtained from 25 ps *ab initio* MD runs using different time steps (microcanonical ensemble, PBE+vdW, $\langle T \rangle = 300$ K). The differences between the spectra are quantified based on the Pendry reliability factor. The spectra are not shifted.

time steps of 0.75 fs and 1.0 fs time steps. However, the minimum between the split peaks always lies directly on top of the non-split peaks. A clear reason why the peaks are split or not cannot be given. It might be an artifact of the time step. On the other hand, it might also be a problem of the level of convergence achieved – for instance, the amide I band of the spectrum for a time step of 0.75 fs obtained from 12x10 ps is not split while the one obtained from a single run of 25 ps length is split (cf. Fig. 6.11).

6.5 Summary

In this chapter, we performed a series of benchmark tests for the calculation of IR spectra from *ab initio* MD simulations. As a model system we used the alanine-based peptide Ac-Ala$_4$-Lys(H$^+$). In the first step, we tested different SCF accuracy settings and wave-function extrapolation schemes. We found that the drift in the total energy is small enough for accuracy settings 2 (see Tab. 6.1), i.e., if the volume-integrated root-mean square change of the charge density is less than 10^{-5} electrons, the sum of eigenvalues is converged up to 10^{-4} eV, the total energy up to 10^{-6} eV and the forces up to $5 \cdot 10^{-4}$ eV/Å. We have shown how it is possible to judge the agreement between two spectra employing a quantitative measure of agreement, namely the Pendry reliability factor[337]. Based on comparisons using the Pendry R-factor and also visual impression, we investigated the convergence of the IR spectra for a single conformational family for increasing simulation times. We found that after about 25 ps the peaks should be sufficiently converged. Moreover, we calculated the IR spectrum based on the average of the autocorrelation function over an increasing number of shorter (10 ps) simulation runs. The R-factor between the IR spectrum obtained from 12x10 ps runs, i.e., 120 ps in total, and from one single run of 25 ps length is very small ($R_\mathrm{P} = 0.1$). It should thus yield decent results to calculate the IR spectrum from one long (25 ps) run. Moreover, we see that the R-factor of the IR spectrum for a different conformational family is significantly higher, namely $R_\mathrm{P} = 0.35$.

7 Proteins and peptides in the gas phase

In the present thesis, we study unsolvated alanine-based peptides. For this reason, we here give a short account on the motivation for gas-phase studies of peptides and proteins and the experimental techniques relevant to this thesis. We further present an overview of the work on alanine-based peptides in the literature with a particular focus on gas-phase studies. Specifically, we deal with the peptides Ac-Ala$_{19}$-Lys + H^+ and Ac-Lys-Ala$_{19}$ + H^+ in this thesis. In Section 7.3, thus, a detailed account on previous work on this peptide system is given, which is intended as a motivation and an introduction to Chapter 8, where the actual results from this work are presented.

Naturally, proteins and peptides occur and are active in aqueous solution. Thus, the gas phase is not an obvious environment in which to study their properties[344]. Both intermolecular (e.g., peptide-water) and intramolecular interactions act together and influence the structure and function of peptides. In this respect, gas-phase studies can be understood as a "reductionist" approach, aiming at a more thorough understanding of the intramolecular interactions when the solvent is stripped away[345]. Investigating the isolated peptide enables a direct assessment of its intrinsic features. Complementary, the gas phase also allows to study solvation effects in a precise way as it is possible, e.g., to examine peptides attached to a single (or more) water molecules ("microsolvation")[289, 346–348]. In this way, the hydration process can be studied in a stepwise fashion as a function of water molecules adsorbed. Gas-phase studies can address questions such as[345]: What role do the intramolecular interactions play for structure formation in proteins and peptides and what part does the solvent play? Is the folded structure preserved under conditions that differ from the natural (solvated) environment, i.e., how (much) does the solution structure change in the gas phase? Knowledge of the latter can offer valuable information about how robust protein and peptide structure is against changes of the environment. Furthermore, it is possible to choose sizes of isolated systems small enough so that they can be treated on a fully first-principles level[17, 289, 349], which facilitates accurate theory-experiment benchmarks. Due to the development of suitable experimental techniques (see the following section), the last two decades have seen an increasing interest in gas-phase studies[27, 303, 312, 344, 345, 348–356].

7.1 Experimental techniques

In order to study biomolecules under isolated conditions, they have to be transferred into the gas phase first. However, proteins and peptides are non-volatile, i.e., they do not evaporate easily. Evaporation techniques involving heat most often lead to decomposition of the biomolecule

necessitating special "softer" vaporization techniques. Pulsed laser desorption methods, e.g., can be used for this, but are limited to rather small biomolecules (1 kDa[1])[350]. A major breakthrough in this respect was the development of electrospray ionization (ESI)[357].[2] This technique allowed J. B. Fenn and his co-workers to transfer molecules with weights of 130 kDa into the gas phase without decomposition, thereby extending the realm of mass spectrometry from small and medium-sized molecules to large molecules. For his work, John B. Fenn was awarded the Nobel Prize in 2002. In ESI, a solution containing the protein of interest is pumped through a thin metal capillary, which is kept at a high voltage (typically several kV). Depending on the polarity of this voltage, positive or negative ions can be produced, respectively. At the end of the capillary the solution is transformed into a fine spray of charged droplets. The solvent molecules in the droplets evaporate aided by an inert nebulizing gas (e.g., N_2) until the size of the droplet reaches the so-called Rayleigh limit[352]. At this point, Coulombic repulsion of the charges within the droplet overcomes the surface tension, which results in fission of the droplet. The resulting smaller droplets undergo the same process iteratively, until finally gas-phase ions are formed. However, the final process of how the gas-phase ions are ultimately produced is not completely understood and two different models exists[360, 361]. A more detailed explanation can be found, e.g., in Ref. [352].

In the following, we shall introduce several experimental methods that can be used to investigate the properties of vapor-phase ions. For this, we concentrate on the techniques relevant to this thesis. All those experimental set-ups are coupled to a mass spectrometer used to separate the gas-phase ions according to their mass over charge ratio.

7.1.1 Ion-mobility mass-spectrometry (IM-MS)

Insight about the structure of the vapor-phase ions can be gained by ion mobility-mass spectrometry (IM-MS)[362]. In this technique, the peptide ions are guided through a drift tube filled with a buffer gas (typically He) under the influence of a weak electric field. The arrival time at the end of the tube depends on the average collision cross section (CCS) with the buffer gas, which is inherently linked to the overall structure. Peptide ions that adopt a compact conformation encounter less collisions than more extended (e.g., helical) structures and thus travel faster through the drift tube. The larger the average CCS, the more collisions the peptide ion will undergo and the later it will reach the end of the tube. This way, IM-MS is able to separate ions according to their structure. The most simple experimental setup consists of an (electrospray) source and a drift tube, which is linked to a mass spectrometer and finally a detector[350]. At the detector, the arrival time distribution (ATD) of the mass-selected ions is measured. The CCS, Ω, can be inferred from the arrival time via[26, 363, 364]:

$$\Omega = \frac{(18\pi)^{1/2}}{16}\left(\frac{1}{m}+\frac{1}{m_\mathrm{b}}\right)^{1/2}\frac{ze}{(k_\mathrm{B}T)^{1/2}}\frac{t_\mathrm{D}E}{L\rho} \quad , \tag{7.1}$$

where L denotes the length of the drift tube, E is the electric field and t_D is the drift time. The charge of the ion is denoted by ze and m and m_b are the masses of the ion and the buffer gas

[1] 1 Da=1 amu

[2] Another important vaporization technique is matrix-assisted laser desorption/ionization (MALDI)[358, 359]. However, the experiments to which we compare our theoretical results in this thesis employed ESI, which is why we focus on the explanation of this technique here.

atoms, respectively. The parameter ρ is the number density of the buffer gas. These measured CCSs can be compared to calculated CCSs for specific conformations of the peptide ion in order to deduce structural information. The calculation of CCSs for a given molecular geometry is explained in Section 7.1.1.1.

7.1.1.1 Calculation of collision cross sections

In order to calculate the average CCS of a molecule, the following orientationally-averaged collision integral has to be evaluated[363, 365]:

$$\Omega(T) = k \int_0^\infty \mathrm{d}\epsilon \int_0^\pi \mathrm{d}\alpha \; f(\epsilon, T)\sigma(\epsilon, \alpha)(1 - \cos(\alpha)) \qquad , \tag{7.2}$$

where k denotes a normalization constant. The angle α describes the scattering angle, i.e., the angle between the velocity vector before and after a scattering process between the molecule and a buffer gas atom. The integral averages over all possible kinetic energies ϵ, where $f(\epsilon, T)$ is the probability density of the kinetic energy, and over all possible scattering angles α, where $\sigma(\epsilon, \alpha)$ describes the probability density of a specific scattering angle given a kinetic energy ϵ. In fact, $\sigma(\epsilon, \alpha)$ is difficult to evaluate as it has to be averaged over all possible collision geometries. For the collision of two hard spheres with radii r_1 and r_2 the cross section (Eq. 7.2) becomes $\Omega = \pi(r_1 + r_2)^2$. In the case of real molecules, however, it has to be evaluated numerically. The most direct approach to solve the problem is to calculate molecular dynamics (MD) trajectories of the peptide guided by the underlying scattering potential. From such a simulation the scattering angle can be directly inferred. The cross section can then be obtained by averaging over a sufficiently large amount of collision geometries. In the trajectory method (TJM) model by Jarrold and co-workers[366, 367] this ansatz is implemented by approximating the scattering potential by a sum of two-body potentials that include a Lennard-Jones term and ion-induced dipole interactions. The latter term accounts for the interaction between the charge of the atom in the peptide ion and the dipole that it induces in the helium buffer gas atom. The parameters of this model potential are fitted to experimental values and the method is implemented in the MOBCAL program[368], which is used in this thesis. For very large systems the TJM model becomes computationally very expensive (approx. 10^5 trajectories are needed to reach converged results for the cross sections). In this case, the scattering potential is often approximated by a sum of hard-sphere potentials, which is referred to as exact hard-sphere scattering (EHSS)[369]. For the EHSS calculations performed in this thesis we used a self-written program by Gert von Helden working in the Molecular Physics Department of the Fritz Haber Institute. Another approach to obtain the average cross section is the projection approximation (PA)[364]. Here, the problem of evaluating the orientationally-averaged collision integral (Eq. 7.2) is simplified to the calculation of two-body collision integrals. These two-body collision integrals between a buffer-gas (helium) atom and each atom of the peptide ion are evaluated based on a potential that includes a Lennard-Jones term and ion-induced dipole interactions [(12,6,4)-potential]. The values for these integrals are tabulated and can be found in Ref. [370]. Based on the collision integral Ω one can define a collision radius in analogy to a hard sphere problem as

$$R_{\mathrm{coll}} = \sqrt{\frac{\Omega}{\pi}} \tag{7.3}$$

for each atom in the peptide ion. The strategy to calculate the total cross section is as follows. Based on the structure of the peptide ion, a sphere is drawn around the position of each atom, with a radius corresponding to the respective R_{coll}. Then, this three-dimensional collection of spheres is perpendicularly projected onto a random plane in space. The area that this projection amounts to is the cross section corresponding to this specific plane. In order to calculate the projection area, circles are drawn on the plane corresponding to the shadow that each sphere casts. Then, a quadratic area A is selected, which covers all circles, i.e., the whole projection. Randomly, points on the area A are chosen. If a specific point lies within one or more circles it is counted as a hit, otherwise as a miss. In the end, the fraction of number of hits over the number of tries multiplied with the size of area A yields the projection area, i.e., the cross section for the chosen plane. These steps are repeated for further randomly chosen planes until the average of the cross section converges within given error limits[364]. In this thesis, we use the PA method as implemented in the program distributed with Ref. [364].

7.1.2 Gas-phase spectroscopy

As discussed in Section 5.2, infrared (IR) spectroscopy can yield valuable information about the structure of peptides. However, due to the low density of mass-selected ions in the gas phase, it is not possible to use absorption spectroscopy techniques to record the IR spectrum. One possibility is to resort to so-called *action spectroscopy* methods. Here, not the intensity reduction of the incoming photon beam due to absorption of photons by the ions is measured, but the response of the ions to the absorption of the photons. This can be, for instance, photon or electron emission, or the fragmentation of the ion[355, 371].

7.1.2.1 Infrared multiphoton dissociation (IRMPD)

The absorption of many photons can lead to a dissociation of the peptide ion if an IR-active resonance frequency of the ion is hit. By measuring the fragmentation (or the depletion of the parent signal) using a mass spectrometer the IR spectra of mass-selected gas-phase ions can be reconstructed. This technique goes by the name of infrared multiphoton dissociation (IRMPD). For the interested reader, an extensive review of IRMPD spectroscopy can be found in Ref. [371]. For IRMPD spectroscopy, the laser has to meet two important requirements. First, it has to be sufficiently tunable in the IR range. Secondly, it has to have a sufficiently high power to enable action spectroscopy. For these purposes free-electron lasers (FELs) prove to work well[371]. The first IRMPD spectrum recorded with a FEL was published in 2000 by Oomens *et al.*[372]. As mentioned earlier, the IRMPD mechanism relies on the fragmentation of the ion after absorbing 50-100 photons. Obviously, it is not *a priori* clear that this technique leads to spectra that are comparable to single-photon absorption spectra. If the laser is resonant with a normal mode frequency of the molecular ion, the first photon will transfer the molecular ion from its vibrational ground state to its first excited vibrational state. In a harmonic picture, all energy levels are equidistant. Thus, the second photon would transfer the ion to its second vibrational state and so on. The molecular ion would climb the vibrational energy ladder by subsequently absorbing photons of the same frequency. However, the potential is not harmonic. Due to anharmonicities, the energy differences between the vibrational levels are not equidistant. After a few photons, the laser would already be out of resonance, hindering the absorption of more

photons. This is called the *anharmonicity bottleneck*[371]. However, if the molecule is sufficiently large and thus, has a large density of states, intramolecular vibrational redistribution (IVR)[373] can occur. By this process, the energy of the absorbed photon is distributed over the bath of vibrational states mediated through anharmonic coupling. This ensures that the molecular ion absorbs each photon at the same vibrational (ground-state) level. After the absorption of typically 50-100 photons the energy crosses the dissociation threshold and the molecular ion fragments. In this way IR spectra that are similar to single-photon absorption spectra can be obtained. However, the multiple-photon absorption process can lead to a broadening of the bands and affect relative band intensities[355, 371]. The finite bandwidth of the excitation laser leads to a further broadening of the bands. The experimental IRMPD spectra presented in this thesis were recorded at the Free Electron Laser for Infrared eXperiments (FELIX) facility in the Netherlands[374]. The bandwidth [full width at half maximum (FWHM)] of this laser is about 1-2% of the corresponding wavenumber[341–343]). If not stated otherwise, the measurements were performed by Stephan Warnke, Kevin Pagel, Frank Filsinger, Peter Kupser, and Gert von Helden from the Molecular Physics Department in the Fritz Haber Institute, Berlin.

7.1.2.2 Messenger technique

Another variant of action spectroscopy is the so-called messenger technique[375], where the ion to be studied is tagged by a rare-gas atom. As the rare-gas atom is assumed to be only very weakly bound to the ion through van der Waals (vdW) interactions, it should not (significantly) influence the structure of the ion. For the same reason, one single photon is sufficient to dissociate the complex. This enables the measurement of the single-photon IR spectrum of the ion, as the dissociation can be detected with a mass-spectrometer. However, such measurements can only be performed at rather low temperatures as otherwise the complex is not stable.

7.1.2.3 IR-UV double resonance

Another technique used for the spectroscopy of biomolecular ions in the gas phase is infrared-ultraviolet (IR-UV) double resonance[349, 356]. The advantage of double-resonance techniques is that they enable the measurement of isomer-selected spectra. One strategy is to employ a peptide sequence containing a UV chromophore, e.g., phenylalanine, tyrosine, or tryptophan residues. The UV spectrum can be measured by photo-fragmentation, as normally a fraction of the peptide ions elevated by a UV photon to an excited electronic state will dissociate (if an appropriate system is chosen, see Ref. [356] for a discussion). The conformer-selective IR spectrum is then measured in the following way. First, the UV spectrum of the ions is recorded, which may give information about conformer-specific peaks[250, 251, 356]. One can then fix the UV laser at a certain frequency, where the spectrum shows a resonance, which corresponds to the electronic excitation of a specific conformer. Typically 200 ns[356] before the UV photo-fragmentation signal is measured, an IR laser pulse is fired at the ions. If the IR frequency is in resonance with an IR-active vibrational transition of the UV-selected conformer, the vibrational ground state population will be depleted. This will show as a dip in the UV photo-fragmentation signal. The IR spectrum can then be reconstructed by scanning the IR laser through the IR range. Of course, this approach is only possible if the ions in the vibrationally excited state do not absorb UV photons at the same frequency (or at least at a lower efficiency) as the ions in the ground state.

7.2 ALANINE-BASED PEPTIDES AND HELIX FORMATION

The work presented in this thesis focuses on isolated polyalanine-based peptides. Polyalanine has been used much as a model system to investigate secondary-structure formation in peptides, in particular helices. There is also an increasing amount of studies on the properties of these peptides in the gas phase. In order to put the work performed here into context, in this section we give a short overview of the field (with special regard to gas-phase studies).

The α-helix is the most prominent secondary-structure building block with about one third of the amino acids in natural proteins being involved in this structural motif[376]. Crystal structures of proteins, which have been extensively studied since the late 1950's[377], revealed that the frequency of occurrence of the different amino acids in α-helices is not distributed equally[378–381]. For instance, alanine is found frequently in helical segments, while glycine is only rarely present. On average, an α-helix in natural proteins comprises eleven residues[382]. However, initial experiments in the late 1960's on short fragments from helical regions of proteins (less than 20 amino acids) did not show helix formation[382–384]. This was in agreement with studies by Scheraga and co-workers. Based on helix-coil transition curves of random copolymers[3], their work indicated that short peptides should not form helices[385–387]. This finding led to the assumption that helices could not serve as independent structure elements during protein folding. However, in 1982, Baldwin and co-workers[388], following up on work by Brown and Klee[389], proved helix formation in a 13-residue peptide extracted from the protein ribonuclease A. Thus, this work showed that helices can serve as folding intermediates after all, triggering a renewed interest in the investigation of stabilizing mechanisms of helices.

In 1987, Marqusee and Baldwin[390] demonstrated the first formation of an α-helix in a designed peptide in water.[4] Their peptides of choice were alanine based with glutamic and lysine residues inserted for reasons of solubility. They rationalized the helix formation in these peptides by charge-charge interactions of the side chains of the negatively charged glutamic acid and the positively charged lysine. To test for helix formation when stabilizing side chain interactions are absent, Baldwin and co-workers[391] investigated alanine-based peptides with only either glutamic or lysine residues inserted. The occurrence of helices in these peptides let them deduce that alanine itself has a high helix-forming potential, which was also confirmed in follow-up studies[382, 392]. In fact, even in peptides with 13 consecutive alanine residues, α-helices were found[393] revealing alanine to be a true intrinsic helix former. Many studies were carried out determining the intrinsic helical tendencies of the different amino acids in solution (helix propensities)[394–398]. All agree that alanine has the highest (or at least a very high) helix propensity. This may be explained by its small side chain (CH_3 group) so that, compared to other amino acids, the loss in side-chain entropy due to helix formation is smaller[399, 400]. The amino acid with the lowest helix propensity is glycine[395] having no side chain at all (only a hydrogen atom, see Fig. 2.3). This can be understood by a different line of reasoning. Due to the non-existing side chain, glycine has a larger conformational freedom in the non-helical state than the other amino acids, leading to a larger entropic destabilization of the helical state[399].

Apart from the intrinsic helix propensities of the amino acids, the interactions between

[3] A copolymer is a polymer that consists of two different monomer units. Scheraga and co-workers used hydroxybutyl- or hydroxypropyl-glutamine as the "host" residue and then built copolymers of the chosen host residue with any of the 20 amino acids in turn (referred to as "guest" residue).

[4] In the literature, the design of artificial peptide sequences is frequently referred to as *de novo* design.

different amino acids also play a role in helix formation. As mentioned above, interactions of charged side chains can stabilize helices[388, 390]. Another important influencing factor are helix dipole-charge interactions[39, 392, 397, 401, 402]. As discussed in Chapter 2, the peptide units exhibit a dipole of approximately 3.5 Debye[39]. In the α-helix, about 97% of the dipole moment of the peptide units point along the helical axis[403] leading to a dipole of the helix that amounts to about 3.5 Debye times the number of residues involved. The direction of this dipole points from the C- to the N-terminus, i.e., there is a net negative charge at the C-terminus and a net positive charge at the N-terminus. Hence, positively charged residues at the C-terminus or negatively charged residues at the N-terminus stabilize helices[404]. On the other hand, charges close to the pole with the same sign of the dipole charge lead to a destabilization. This is in agreement with the frequency of occurrence of charged residues found at the termini of helices in natural proteins[405]. Negatively charged residues are indeed often found at the N-terminus of helices, while positively charged residues occur more at the C-terminus.

Ooi and co-workers[406, 407] investigated the importance of charge-helix dipole interactions for helix stability. They designed different types of peptides that each consisted of two blocks of 20 amino acids. In the first case the two blocks were formed by alanine (Ala_{20}) and glutamic acid (Glu_{20}) and in the second case they were formed by alanine (Ala_{20}) and lysine (Lys_{20}). For each set of blocks they designed two peptides – one where the block of charged residues was attached to the N-terminus of the alanine block and one where it was attached at the C-terminus of the alanine sequence (Glu_{20}: negatively charged, Lys_{20}: positively charged). They observed stabilization of the polyalanine helix when the charged block was located at the helix pole with opposite charge (Ala_{20}-Lys_{20}-Phe, Glu_{20}-Ala_{20}-Phe) and destabilization when the charged block was close to the pole with equal sign of charge (Lys_{20}-Ala_{20}-Phe, Ala_{20}-Glu_{20}-Phe).[5]

In fact, in a landmark experiment from 1998, Jarrold and co-workers[25] could show that a protonated lysine residue incorporated at the C-terminus of a polyalanine sequence enables the formation of stable α-helices in the *gas phase*. In particular, positively charged peptides of the form Ac-Ala_n-Lys(H^+) with $n = 5 \ldots 19$ were studied using IM-MS[408, 409]. The N-terminus of these peptides was capped with an acetyl group (Ac) so that the amine group of the lysine side chain becomes the preferred protonation site[26]. The stabilization of the helical structure by the protonated lysine residue arises from two factors. As illustrated in Fig. 7.1, for Ac-Ala_{19}-Lys(H^+) the positive charge located at the lysine residue interacts electrostatically favorably with the helix dipole. Furthermore, in an ideal α-helix the four backbone carbonyl oxygens at the C-terminal end of the helix are not involved in hydrogen bonds. Due to its flexible side chain, the protonated lysine amine group is able to cap them, which thus stabilizes the helical structure. In Fig. 7.1, e.g., the protonated lysine amine group coordinates to three of the four dangling backbone carbonyl groups. Based on the comparison between measured and calculated CCSs obtained from force-field MD simulations, it was concluded that Ac-Ala_n-Lys(H^+) with $n \geq 7$ forms helices in the gas phase at room temperature. Following up on that, Jarrold and co-workers performed many further studies of polyalanine-based peptides in the gas phase[26, 27, 346, 347, 410–416]. For the particular example of Ac-Ala_{15}-Lys(H^+) they showed that the α-helix is extremely stable even up to high temperatures[415]. The peptide stays almost completely helical up to 450 °C (725 K), which is its dissociation limit. This is in contrast to α-helices in solution, where peptides hardly show any helical content at temperatures of 70 °C[390, 391]. This extreme stability of

[5] The phenylalanine (Phe) residue was needed as a marker during the peptide synthesis process.

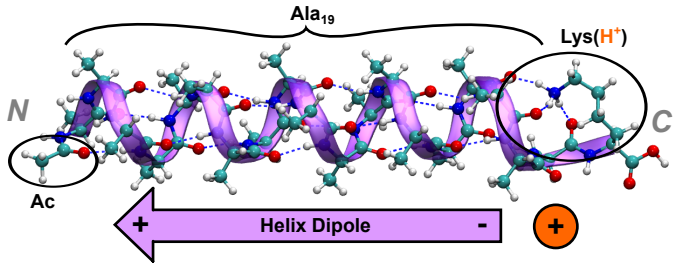


Figure 7.1: Helical structure of Ac-Ala$_{19}$-Lys(H$^+$). The alignment of the peptide units along the helical axis leads to a helix macrodipole pointing from the C- to the N-terminus indicated by the purple arrow. The charge located at the lysine residue interacts electrostatically favorable with the helix dipole.

Ac-Ala$_{15}$-Lys(H$^+$) in the gas phase has recently been confirmed by Tkatchenko *et al.*[44] based on first-principles calculations. They employed density-functional theory (DFT) with the PBE[15] functional corrected for vdW interactions[16] (PBE+vdW, see Section 3.5.3). The results reveal that accounting for vdW interactions is essential to explain the high-temperature stability of the helix. Laskin and co-workers even found that Ac-Ala$_{15}$-Lys(H$^+$) is helical when adsorbed on self-assembled monolayer (SAM) surfaces[417, 418].

As discussed above, the protonated lysine residue located at the C-terminus of peptides Ac-Ala$_n$-Lys(H$^+$) leads to a stabilization of the α-helix due to favorable interactions of the charge with the helix dipole. Consequently, helical conformation is destabilized when the lysine is located at the N-terminus [Ac-Lys(H$^+$)-Ala$_n$]. With only small contributions from helices and helical dimers, Jarrold and co-workers found these peptides to predominantly form compact globular structures[26, 27]. We will go into this in more detail in Section 7.3.

As mentioned in the beginning of this chapter, in the gas phase, it is possible to investigate peptides attached to a single or several water molecules ("microhydration"). Using water adsorption experiments under equilibrium conditions, Jarrold and co-workers studied the hydration propensities of Ac-Lys(H$^+$)-Ala$_{20}$ and Ac-Ala$_{20}$-Lys(H$^+$)[346]. In such experiments (see also work by Bowers and co-workers[419, 420]), an IM-MS drift-tube apparatus is used, where a (constant) partial pressure of water vapor is imposed on the tube. By measuring the intensities of the unsolvated and singly-solvated peptides in the mass spectrometer the equilibrium constant for the adsorption of the first water molecule can be determined. Jarrold and co-workers found that the hydration propensity of Ac-Lys(H$^+$)-Ala$_{20}$ [compact] is higher than for Ac-Ala$_{20}$-Lys(H$^+$) [helical][346]. In fact, they did not detect water adsorption for the helical peptide at all. Based on these findings, in a follow-up study[347] using the same technique, they estimated the helical onset for the series Ac-Ala$_n$-Lys(H$^+$) as a function of n. Analyzing the water adsorption propensity of Ac-Ala$_n$-Lys(H$^+$) versus Ac-Lys(H$^+$)-Ala$_n$ points to the conclusion that Ac-Ala$_n$-Lys(H$^+$) should at least be helical for $n \geq 8$. This is consistent with results reported by Bowers and co-workers for the same peptide series using the same technique[419]. Recent work by Chutia *et al.*[289] for Ac-Ala$_8$-Lys(H$^+$) and Ac-Ala$_5$-Lys(H$^+$) using a DFT-based analysis indicates that the preferred monohydration site for both peptides is the lysine amine group. The calculated finite-temperature equilibrium constants for water adsorption agree well with the experimental values[347, 419]. Moreover, it could be shown that the difference in water adsorption propensity between Ac-Ala$_5$-Lys(H$^+$) [non-helical] and Ac-Ala$_8$-Lys(H$^+$) [helical] is mostly due to a small change in the vibrational contribution to the

free energy (harmonic approximation) leading to a drop in adsorption energy for the helical structure of about 1 kcal/mol (or 0.04 eV). In order to study the helical onset of Ac-Ala$_n$-LysH$^+$ at a first-principles level of theory, Rossi *et al.*[28] performed conformational searches based on DFT with the PBE+vdW functional for n = 4-8. Free-energy conformational hierarchies (in the harmonic oscillator-rigid rotor approximation) indicate that the α-helix is the dominant structure at 300 K already for $n = 6$. However, other more compact conformations are still close in energy. For $n = 8$, on the other hand, all other conformations are significantly higher in free energy than the α-helix so that the α-helix should be the only conformation that is observed at room temperature.

If the N-terminus of the polyalanine peptide is not capped by an acetyl group and no lysine is incorporated in the sequence, the N-terminal amine group has the highest proton affinity[410]. Thus, the charge in protonated polyalanine (Ala$_n$H$^+$) is most probably located at the N-terminus, which would destabilize helical conformations. Indeed, Jarrold and co-workers observed mostly compact structures for Ala$_n$H$^+$ and Gly$_n$H$^+$, $n = 3 \ldots 20$, based on a comparison of measured and calculated CCSs. The calculations were performed for selected structures obtained from force-field based MD simulations[410, 414]. The structures are self-solvated globules, i.e., the peptide chain obtains a conformation that shields ("solvates") the charge as well as possible. Similar observations were made by Bowers and co-workers for protonated polyglycine sequences (GlyH$^+$)[421]. For the larger protonated polyalanine sequences (e.g., $n = 17$) Jarrold and co-workers found also small helical regions in the self-solvated globules. Apart from the self-solvated globules they also found evidence of pure helical structures[414]. As discussed above, for a helical structure to be stable, the proton needs to be located at the backbone carbonyl groups close to the C-terminus. This is, however, not the preferred protonation site (which would be the N-terminal amine group), but this might be overcome by the energetic preference to form a helix if the peptide is large enough[414]. At elevated temperatures (around 450 K) the helix and globule rapidly interconvert. As in the helix the proton should be located close to the C-terminus and in the globule close to the N-terminus, this interconversion of structures indicates a mobility of the proton, i.e., it is able to move freely along the backbone. When the proton is exchanged for an alkali metal ion (Li$^+$, Na$^+$, K$^+$, Cs$^+$, Rb$^+$) no globular structures, but only rigid helices are formed[411, 416]. This is a result of the metal ion coordinating to the C-terminus.

The CCSs obtained by IM-MS measurements give only a measure of the overall shape of the peptide. As discussed in Section 5.2, a method that should yield higher structural resolution is IR spectroscopy. IR spectroscopy has been employed by several groups to study neutral peptides in the gas phase as reviewed in Ref. [349]. For instance, Mons and co-workers performed a series of studies of small peptides containing only a few residues based on IR-UV spectroscopy[422–429] (see Section 7.1.2). In a bottom-up approach starting from peptides with only one residue up to four residues they benchmarked the sensitivity of IR spectra in the NH stretch region (amide A, see Section 5.2) with special regard to hydrogen bonding. They studied the intrinsic local conformational preference of the backbone of individual amino acids, the effect of the side chains and the influence from neighboring residues with special focus on secondary structure. In a three-residue peptide they demonstrated the presence of a 3_{10}-helix in the gas phase[423, 425]. Rizzo and co-workers[250, 251, 356] studied the larger, protonated peptides Ac-Phe-Ala$_5$-LysH$^+$, Ac-Phe-Ala$_{10}$-Lys(H$^+$), and Ac-Lys(H$^+$)-Phe-Ala$_{10}$. The phenylalanine (Phe) residue provides the necessary chromophore to permit IR-UV spectroscopic studies. They compared the IR

spectra in the NH stretch region obtained at about 10 K with calculated harmonic IR spectra based on DFT and the B3LYP functional. From this they could infer that Ac-Phe-Ala_5-$LysH^+$ and Ac-Phe-Ala_{10}-$Lys(H^+)$ form helices, while Ac-$Lys(H^+)$-Phe-Ala_{10} adopts a globular structure in agreement with the expectation from Jarrold and co-workers' work[25–27] based on IM-MS.

Martens *et al.* studied sodiated polyalanine sequences Ala_nNa^+, n = 8–12. As discussed above, preceding work by Jarrold and co-workers[411, 416] indicates that these form helices at least for all sequences with $n > 12$. Martens *et al.* showed that the IRMPD spectra for the peptides in the NH stretch region can be used to determine the helical onset more exactly to occur at $n = 9$. Vaden *et al.* employed IRMPD in the NH stretch region to study the series of protonated polyalanine peptides Ala_nH^+ (n = 3, 4, 5, and 7) in the gas phase. They compared the experimental data to calculated harmonic IR spectra for conformations predicted with DFT (B3LYP). In this way, they could corroborate the globular nature of these peptides in the gas phase predicted by Jarrold and co-workers based on IM-MS before[410, 414] (discussed above). In follow-up studies for the same peptide series, Gaigeot and co-workers[430, 431] demonstrated that IR spectra derived from DFT-based MD simulations yield a much better match with experiment than the harmonic ones as they naturally account for vibrational anharmonicities (within the classical-nuclei approximation) and conformational fluctuations[303, 312].

In previous work from Rossi and co-workers[17, 338], IR spectra derived from first-principles MD simulations were calculated for Ac-Ala_n-$LysH^+$ with $n = 5, 10, 15$. The theoretical spectra were compared to IRMPD data measured in the wavenumber region between 1000 and 1800 cm^{-1} (amide I, II, III) by the group of Gert von Helden working in the Molecular Physics department of the Fritz Haber Institute. The degree of similarity was assessed with a quantitative measure (Pendry reliability factor, explained in Chapter 6). In agreement with the findings from Jarrold and co-workers[25, 347], Ac-Ala_{15}-$LysH^+$ and Ac-Ala_{10}-$LysH^+$ were found to form α-helices in the gas phase, while Ac-Ala_5-$LysH^+$ is represented by a mixture of conformations at room temperature.

7.3 Ac-Ala_{19}-Lys + H^+ vs. Ac-Lys-Ala_{19} + H^+

In the previous section, a general overview of studies on (isolated) polyalanine-based peptides was given. In the present thesis, we particularly concentrate on the two peptides Ac-Ala_{19}-Lys + H^+ and Ac-Lys-Ala_{19} + H^+. In this section, we thus give a more detailed description of the work performed previously by Jarrold and co-workers that is directly relevant for this peptide system. In particular, part of the work of this thesis was performed in a collaboration with the group of Gert von Helden working in the Molecular Physics department of the Fritz Haber Institute. Some of their experimental results are shown in this section serving as an introduction to the subsequent chapters, where the actual work of this thesis is presented.

As mentioned briefly in the previous section, in 1998, Jarrold and co-workers demonstrated helix formation in protonated alanine-based polypeptides in the *gas phase*[25, 26]. They designed a peptide series, where the N-terminus of a polyalanine sequence was capped with an acetyl group and at the C-terminus a lysine residue was inserted: Ac-Ala_n-Lys + H^+, with $n = 5 - 19$. As the N-terminus is capped by the acetyl group (Ac), the lysine amine group is the preferred protonation site[26]. The positive charge located at the lysine residue interacts electrostatically favorably with the helix dipole and, additionally, the protonated amine group forms hydrogen

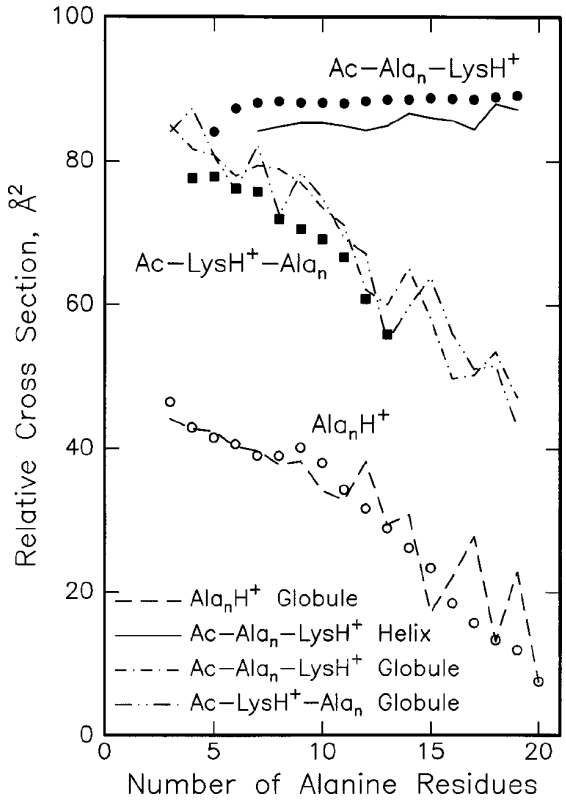


Figure 7.2: Relative collision cross sections (CCSs) for Ac-Ala$_n$-Lys + H$^+$ (filled black circles), Ac-Lys-Ala$_n$ + H$^+$ (filled black squares), and Ala$_n$ + H$^+$ (empty circles) as a function of the number of alanine residues n measured by Jarrold and co-workers. The lines indicate the values of the calculated CCSs for the corresponding peptides. The relative CCSs are given as $\Omega_{\text{av}} - 14.50\,\text{Å}^2 \cdot n$, where Ω_{av} is the CCS, n denotes the number of alanine residues, and 14.50 Å^2 is the calculated value of the average CCS per residue of an ideal polyalanine α-helix. Reprinted with permission from Ref. [26]. Copyright 1999 American Chemical Society. URL pointing to the article: http://dx.doi.org/10.1021/ja983996a.

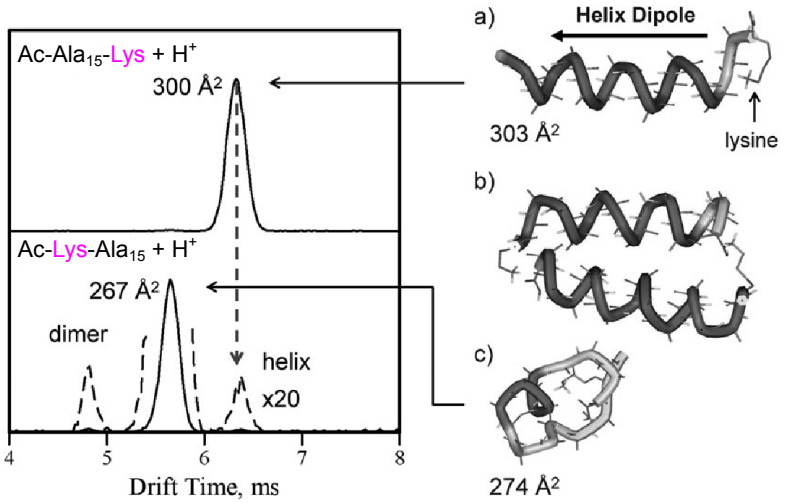


Figure 7.3: Left side: Drift-time distributions measured for Ac-Ala$_{15}$-Lys + H$^+$ and Ac-Lys-Ala$_{15}$ + H$^+$ at 300 K by Jarrold and co-workers. Right side: Snapshots taken from force-field based MD simulations for Ac-Ala$_{15}$-Lys + H$^+$ [a): helix] and for Ac-Lys-Ala$_{15}$ + H$^+$ [b): helical dimer, c): compact globular structure]. Reproduced from Ref. [27] with permission from the PCCP Owner Societies. Labels were adjusted to match the nomenclature used in this thesis. URL pointing to the article: http://dx.doi.org/10.1039/b612615d.

bonds with the dangling carbonyl oxygen atoms at the C-terminal end of the helix. In this way, the lysine at the C-terminus stabilizes a helical conformation as illustrated in Fig. 7.1. Figure 7.2 shows the relative CCSs measured by Jarrold and co-workers[26] for Ac-Ala$_n$-Lys + H^+ as a function of n using IM-MS. The relative scale is given by $\Omega_{\mathrm{av}} - 14.50\,\text{Å}^2\,n$, where n is the number of alanine residues and Ω_{av} is the average CCS in units of Å^2. The value 14.50 Å^2 is the average CCS per residue determined for an ideal polyalanine α-helix[25, 26]. In this idealized model, the relative CCSs for a helical conformation of the peptide should not depend on the number of residues and lie on a horizontal line in Fig. 7.2. In fact, for $n \gtrsim 7$ the measured relative CCSs for Ac-Ala$_n$-Lys + H^+ lie on a horizontal line, i.e., they form helices. On the other hand, if the lysine residue is located at the N-terminus (Ac-Lys-Ala$_n$ + H^+) a helical conformation is destabilized. In agreement with this, the relative CCSs shown in Fig. 7.2 decrease with peptide length, which points to more compact structures.

Figure 7.3 directly compares the ATDs (drift times) measured by Jarrold and co-workers[27] for Ac-Ala$_{15}$-Lys + H^+ and Ac-Lys-Ala$_{15}$ + H^+. The upper panel shows the ATD for Ac-Ala$_{15}$-Lys + H^+, which exhibits only one single peak. According to their previous findings this corresponds to a helical conformation. Subplot a) in Fig. 7.3 shows a helical snapshot taken from a force-field based MD simulation. The lower panel of Fig. 7.3 illustrates the ATD for Ac-Lys-Ala$_{15}$ + H^+. There is one main peak and two small peaks (scaled by a factor of 20 in the picture), which flank the dominant peak. The small peak at the right side of the main peak appears at the same position as the peak of the Ac-Ala$_{15}$-Lys + H^+ helix. It is thus consistent with a helical conformation. The appearance of the main peak at smaller drift times points to more compact structures. Based on force-field MD simulations Jarrold and co-workers associated this peak with compact globular structures, where the peptide chain is wrapped around to self-solvate the charge. A structure snapshot taken from the simulation is shown in subplot c) of Fig. 7.3. Notably, the overall compact structure still contains a small helical fragment.

The mass spectrometer in the IM-MS setup separates the peptides according to their mass over charge ratio, i.e., protonated monomers and doubly-protonated dimers cannot be distinguished. With the help of force-field based MD simulations and analysis of the IM-MS and mass spectrometer data[26, 27], Jarrold and co-workers associated the small peak at the left side of the main peak in the ATD of Ac-Lys-Ala$_{15}$ + H^+ (see Fig. 7.3) with helical dimers. The most probable conformation of such a dimer is a "head-to-toe" arrangement of two helical monomers as illustrated in subplot b) of Fig. 7.3. In such a structure, the protonated lysine amine group of one helical monomer interacts with the C-terminus of the other helical monomer by forming hydrogen bonds with the dangling carbonyl oxygens. Furthermore, the helix dipoles of the two monomers are in antiparallel alignment, which is electrostatically favorable. Jarrold and co-workers further argue that the helical monomers (small peak at the right side of the dominant peak) arise through a dissociation of those dimers, upon which the proton of the lysine residue of one monomer is transferred to the C-terminus of the other helical monomer. In this way, the positive charge located at the C-terminus stabilizes a helical conformation.

Depending on the precise experimental set-up, the amount of dimers present in the ATD measured by Jarrold and co-workers varies[26, 27]. They found that when decreasing the injection energy of the peptide ions into the drift tube the amount of dimers observed increases. Upon entering the drift tube, the ions are heated by collisions with the buffer gas before being thermalized, which can lead to conformational changes. Thus, the dimers might be the dominant

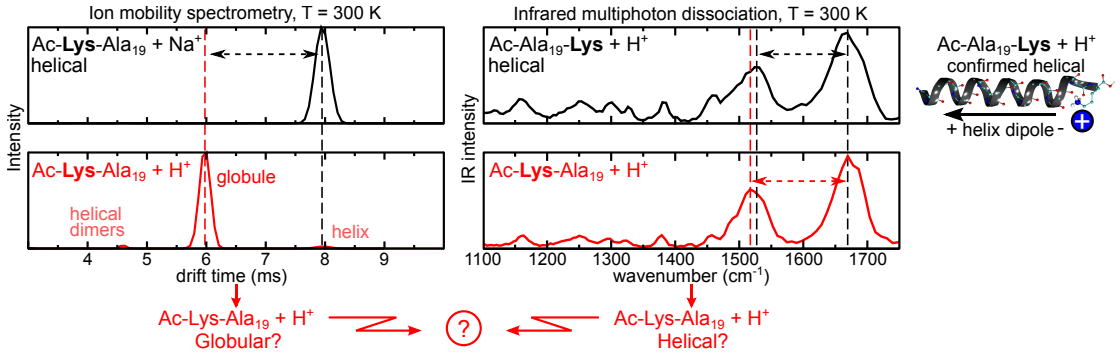


Figure 7.4: Collected experimental data for Ac-Ala$_{19}$-Lys + H$^+$ and Ac-Lys-Ala$_{19}$ + H$^+$/Na$^+$, which were performed by Stephan Warnke, Peter Kupser, Kevin Pagel, and Gert von Helden working in the Molecular Physics Department of the Fritz Haber Institute. Left panel: Ion mobility-mass spectrometry (IM-MS) drift time distributions for Ac-Lys-Ala$_{19}$ + Na$^+$ (helical[411, 432]) versus Ac-Lys-Ala$_{19}$ + H$^+$. Right panel: Infrared multiphoton dissociation (IRMPD) spectra for Ac-Ala$_{19}$-Lys + H$^+$[343] (helical[17, 25–28, 338]) versus Ac-Lys-Ala$_{19}$ + H$^+$. The experimental data result in a conformational puzzle: While IM-MS points to mostly compact, globular conformations for Ac-Lys-Ala$_{19}$ + H$^+$, the similarity of the IRMPD spectrum to the helical Ac-Ala$_{19}$-Lys + H$^+$ would imply helical structures.

conformation before entering the drift tube, but dissociate upon collisional heating when entering the drift tube. This means that the dimers are either present in solution or form during the electrospray process. The latter is considered more probable[414] as during the electrospray process (discussed in Section 7.1) the ions enter the gas phase from droplets that evaporate. If the concentration of the peptide in solution is high, the formation of dimers during this process becomes more probable. This was demonstrated by IM-MS measurements performed by Stephan Warnke, Gert von Helden, and Kevin Pagel and shown in Appendix A.3.

As mentioned earlier, in this thesis we concentrate on the two peptides Ac-Ala$_{19}$-Lys + H$^+$ and Ac-Lys-Ala$_{19}$ + H$^+$. Figure 7.4 shows the experimental data for this system collected by Stephan Warnke, Peter Kupser, Kevin Pagel, and Gert von Helden as part of a collaboration for the work presented in this thesis. The left panel illustrates room-temperature IM-MS drift-time distributions for Ac-Lys-Ala$_{19}$ + Na$^+$ and Ac-Lys-Ala$_{19}$ + H$^+$. The results are analogous to the findings from Jarrold and co-workers[27] (Fig. 7.3) for Ac-Ala$_{15}$-Lys + H$^+$ versus Ac-Lys-Ala$_{15}$ + H$^+$. For Ac-Lys-Ala$_{19}$ + Na$^+$ only one peak is observed, which we associate with a helical conformation as sodiated polyalanine peptides have been shown to form helices before[411, 432]. Just as for Ac-Lys-Ala$_{15}$ + H$^+$, for Ac-Lys-Ala$_{19}$ + H$^+$ three distinct peaks are found, where the largest peak is flanked by two peaks with significantly smaller intensity. As the position of the small peak at the right side of the dominant peak coincides with the helical peak for Ac-Lys-Ala$_{19}$ + Na$^+$, we associate it with a helical conformation. Just as Jarrold and co-workers we associate the main peak appearing at smaller drift times with more compact structures labelled as "globule" in the figure. Similarly, we assign the small peak at the left side of the dominant peak to helical dimers.

The right panel of Fig. 7.4 shows room-temperature IRMPD measurements for Ac-Ala$_{19}$-Lys + H$^+$ in the wavenumber range between 1100 and 1750 cm^{-1}, which has been confirmed to be helical by various studies as discussed in the previous section[17, 25–28, 338]. The IR spectrum is compared to the IR spectrum of Ac-Lys-Ala$_{19}$ + H$^+$, which is expected to be predominantly globular according to the IM-MS results. As discussed in Chapter 6 such differences in the conformations should affect the IR spectra. Surprinsingly, however, the IR spectra of Ac-Ala$_{19}$-

Lys + H^+ and Ac-Lys-Ala_{19} + H^+ are very similar and show only subtle differences such as the shift of the amide II band by about $10\,cm^{-1}$. The peak positions of the amide I bands are almost on top of each other. Given this similarity, the IRMPD spectra would rather point to helical conformations for Ac-Lys-Ala_{19} + H^+, while the IM-MS measurements suggest mainly compact globular conformations resulting in a conformational "puzzle" for Ac-Lys-Ala_{19} + H^+. In the following chapters, we aim at an understanding of this apparent mismatch based on a first-principles level of theory.

8 First-principles structure predictions for Ac-Ala$_{19}$-Lys + H$^+$ vs. Ac-Lys-Ala$_{19}$ + H$^+$

In this work, we study the peptide system Ac-Ala$_{19}$-Lys + H$^+$ and Ac-Lys-Ala$_{19}$ + H$^+$. As discussed in the previous chapter, the positively charged lysine residue at the C-terminus of Ac-Ala$_{19}$-Lys + H$^+$ stabilizes an α-helical conformation due to a favorable electrostatic interaction with the helix dipole[25]. The question is now what happens if this charged residue is shifted to the N-terminus (Ac-Lys-Ala$_{19}$ + H$^+$)? In this case, the charge would rather destabilize an α-helical conformation. However, the two gas-phase experiments discussed in Section 7.3 give seemingly different answers. In the ion mobility-mass spectrometry (IM-MS) experiments, predominantly compact globular structures ("globules") are found[26, 27]. However, the infrared multiphoton dissociation (IRMPD) spectra for Ac-Ala$_{19}$-Lys + H$^+$ and Ac-Lys-Ala$_{19}$ + H$^+$ are very similar, suggesting that Ac-Lys-Ala$_{19}$ + H$^+$ might form helices after all (cf. Fig. 7.4 in the previous chapter). We here set out in order to solve this seeming experimental mismatch using a first-principles screening effort, unprecedented for a flexible system of that size (220 atoms).

In the work of Jarrold and co-workers, the structures of Ac-Lys-Ala$_{19}$ + H$^+$ were generally characterized as being "globular", while no clear structure assignment was performed[26, 27]. Here we assess to what extent the structure space of this 20-residue peptide can be predicted on a *quantitative* level. For this, several difficulties have to be overcome. First of all, the peptides have a huge conformational space infeasible to be sampled using a pure first-principles approach alone. On the other hand, an accurate method needs to be used for describing the potential-energy surface (PES), especially as different structures for Ac-Lys-Ala$_{19}$ + H$^+$ might exist close in energy[26, 27]. In previous work for the smaller peptides Ac-Ala$_n$-Lys(H$^+$), $n = 4-8$, force-field basin-hopping searches followed by a relaxation of conformers with the PBE+vdW functional have proven as a valuable tool[17, 28]. However, such basin-hopping approaches were found to reach their limits for system sizes of about 100 atoms and are thus not feasible for the peptides aimed to study here. Furthermore, what sets Ac-Lys-Ala$_{19}$ + H$^+$ apart from all peptides that have been studied in our group before is that there exists no *a priori* knowledge about its structural preferences. While for Ac-Ala$_n$-Lys(H$^+$) it is at least known that the construction principle supports helix formation, the Ac-Lys-Ala$_{19}$ + H$^+$ peptide involves the problem that *a priori* it is not known how the most probable structure might look like and many completely different structures might co-exist. Thus, a new search approach has to be developed.

8.1 Assessment of conformational search strategy

We start our structure search with a *global sampling* of the conformational space using a force field. For this, we use replica-exchange molecular dynamics (REMD). The beauty of this approach is on the one hand that it does not employ a bias to the system and on the other hand that it naturally restricts the search space to the physically meaningful part. For a *local refinement*, we subsequently relax thousands of structures with PBE+vdW. While the force-field step generates diversity of structures, with the PBE+vdW relaxations, we pin point a more reliable energy hierarchy. However, the PBE+vdW conformers still include a force-field bias as they are relaxed from force-field guesses. In order to find the lowest-energy PBE+vdW structure of the respective basin of the PES we follow up with *ab initio* REMD structure searches for the lowest-energy structures to refine the local structural environment.

By the local sampling step (huge amount of PBE+vdW relaxations with subsequent *ab initio* REMD) we aim to make our search approach as independent as possible from the choice of the specific force field (here OPLSAA).

As discussed in Section 3.5, density-functional theory (DFT) is in principle an exact theory, but the exchange-correlation functional has to be approximated. It is thus an important question, which functional yields the most reliable results for the system under study. For the lowest-energy PBE+vdW structures identified in our structure search, we investigate the influence of different approximations to the exchange-correlation functional in Section 8.6. As a validation, we compare the structure predictions to the experimental IM-MS and IRMPD data in Chapter 9. The results dramatically highlight the importance of an accurate description of the energetics with small systematic errors being able to lead to completely different predictions.

It is important to make sure that our search strategy reliably identifies the lowest-energy conformers of the peptide under study. For this reason, we carefully assess all steps of our search strategy one after the other in the following subsections. As the steps for Ac-Ala$_{19}$-Lys + H$^+$ and Ac-Lys-Ala$_{19}$ + H$^+$ are the same and it is already known that isolated Ac-Ala$_{19}$-Lys + H$^+$ forms α-helices[25], we focus the description on Ac-Lys-Ala$_{19}$ + H$^+$. Apart from the general conformational search for Ac-Lys-Ala$_{19}$ + H$^+$ discussed in the following, we performed two independent searches to specifically find model conformers of helical dimers and monomers. As discussed extensively in Section 7.3, Jarrold and co-workers proposed the presence of small amounts of such helices based on IM-MS data[26, 27]. The respective structure searches will be addressed later in this Chapter (Section 8.4).

8.1.1 Global sampling of the conformational space

As mentioned earlier, we used REMD simulations to broadly sample the conformational space of Ac-Lys-Ala$_{19}$ + H$^+$. They were carried out with the GROMACS program package[433]. We employed 16 replicas in the temperature range between 300 K and 915 K with a time step of 0.5 fs and a swapping attempt frequency of 2 ps^{-1}. All replicas were initialized with ideal α-helices.

For a first analysis, we let all replicas run for 100 ns. The REMD simulations yield a huge amount of structures, basically one geometry for each time step and for each replica, i.e., 3,200,000,000 structures in our case. It is thus necessary to identify the most important structure types. Snapshots of a molecular dynamics (MD) simulation taken after subsequent time steps hardly vary. We thus extract snapshots only after every 2 ps. Furthermore, we first concentrate

only on the 300 K trajectory resulting in 50,000 structures. This pool of structure snapshots is then separated into clusters of similar conformations (based on their root mean square deviation (RMSD)). For this, we use the cluster algorithm by Daura *et al.*[99] implemented in the GROMACS program package[433]. The first step of this algorithm is to calculate an RMSD matrix, i.e., the RMSD for each geometry with all other geometries. The RMSD is evaluated using

$$\mathrm{RMSD} = \sqrt{1/M \sum_{i=1}^{N} m_i (\boldsymbol{x}_i^1 - \boldsymbol{x}_i^2)^2} \quad , \tag{8.1}$$

where $M = \sum_{i=1}^{N} m_i$ and N denotes the number of atoms taken into account – for our calculations that were the backbone atoms plus the nitrogen and carbon atoms of the lysine side chain. The vectors $\boldsymbol{x}^1$ and $\boldsymbol{x}^2$ are the positions of equivalent atoms and m_i is their corresponding mass. After evaluating the RMSD matrix, the number of neighbors for each structure is determined. Two structures are considered to be neighbors if their RMSD value is lower than a given cut-off criterion. The conformer with the largest amount of neighbors forms – together with its neighbors – the first cluster. In an analogous way, the second cluster is determined from the remaining structures and so on. For each cluster the midpoint structure is calculated. It is defined as the structure that has the lowest average RMSD to all other structures of the cluster.

In order to determine a reasonable cut-off criterion, we clustered our pool of structures with a series of different cut-off values (0.25 Å, 0.5 Å, 0.75 Å, and 1.0 Å). The results are presented in Fig. 8.1. For the 50 largest clusters, the plot illustrates the relative number of structures comprised by each of it. In addition, the total number of clusters and the number of the largest clusters, which account for 85 % of all structures are given. Among the 50,000 structures present in the pool, a cut-off criterion of 0.25 Å yields 49,326 clusters, where the largest 41,826 clusters contain 85 % of the structures. A cut-off criterion of 0.5 Å leads to 6117 clusters, where the 891 largest clusters account for 85 % of all structures. Even larger cut-offs of 0.75 Å and 1 Å yield 1456 and 750 clusters, respectively. Here we decided to use a cut-off criterion of 0.5 Å in the following. On the one hand, it leads to a measurable reduction in the amount of structures, but on the other hand, it is still relatively small so that no important structure types should be overlooked.

In order to see how the conformational space is sampled with advancing REMD simulation time, we clustered the structure snapshots of the 300 K trajectory (taken after every 2 ps) after different steps of simulation time. The number of clusters as a function of simulation time is given in Tab. 8.1, which shows that it is only converging slowly. In order to analyze the energetic spread of the clusters, we chose the largest clusters that contained 85% of all structures and relaxed their midpoint structures with OPLSAA. Figure 8.2 shows the relative energies of the relaxed midpoint structures as a function of the simulation time. The lowest-energy conformation found by this strategy already appears after 200 ns. We take this as an indicator that a simulation time of more than twice as much should be sufficient so that we base our further analysis on the results obtained after 500 ns of REMD simulation time (per trajectory). Note that a simulation time of 500 ns per trajectory corresponds to a total simulation of even 8 µs. To identify the energy hierarchy of the 20,877 clusters (see Tab. 8.1) obtained after this simulation time we relaxed the midpoint structure of each of them with OPLSAA. This yielded [based on an RMSD and energy comparison (10^{-6} eV)] 9,620 different structures for Ac-Lys-Ala_{19} + H^+.

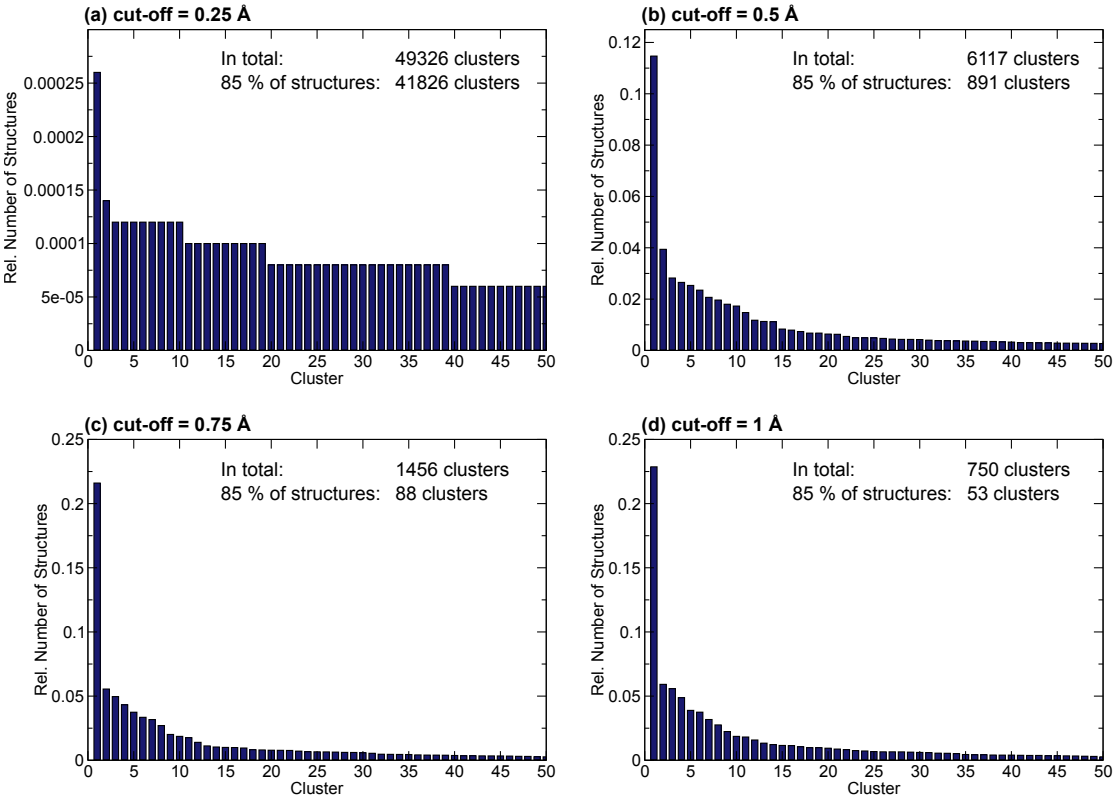


Figure 8.1: Results of clustering approaches for structures of the peptide Ac-Lys-Ala$_{19}$ + H$^+$ based on the algorithm described in Ref. [99]. The pool of structures (50,000) was comprised of snapshots of the force-field based REMD simulations up to a length of 100 ns. The plots shows the relative number of structures contained in the 50 largest clusters, which were obtained using a cut-off criterion of (a) 0.25 Å, (b) 0.5 Å, (c) 0.75 Å, and (d) 1.0 Å. In addition, the plots give the total number of clusters and the number of the largest clusters, which comprise 85% of all structures.

Table 8.1: Number of clusters obtained for Ac-Lys-Ala$_{19}$ + H$^+$ from the snapshots (taken every 2 ps) of the 300 K REMD trajectory after the given time of simulation. For the clustering approach the algorithm by Daura *et al.*[99] was used with a cut-off criterion of 0.5 Å. In addition, the number of the largest clusters that comprise 85% of all structures are given.

Time (ns)	N_{total}	$N_{85\,\%}$
2	305	305
10	911	225
20	1733	407
30	2594	625
40	3324	770
50	3903	831
75	5083	873
100	6117	891
200	9240	902
300	12321	976
400	16279	1278
500	20788	1726

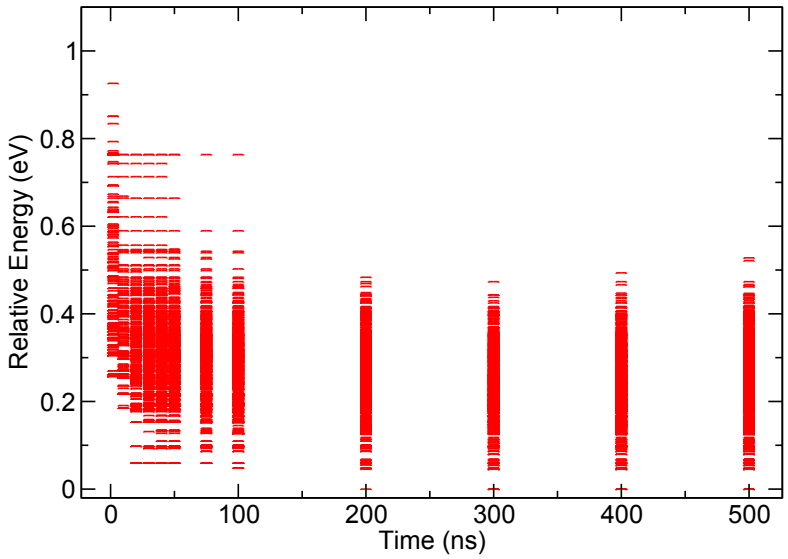


Figure 8.2: Plot of the relative force-field energy of the relaxed (OPLSAA) midpoint structures of each cluster obtained after the given simulation time. The clustering was carried out for the snapshots (taken every 2 ps) from the 300 K trajectory of the REMD simulation performed for Ac-Lys-Ala_{19} + H^+. The cut-off was 0.5 Å. The OPLSAA relaxations were performed for the largest clusters that comprised 85% of the structures. The energies are given relative to the lowest-energy structure found.

8.1.2 Local refinement

We now arrived at the second step of our search strategy, namely the local refinement with DFT. We relaxed the 1,026 lowest-energy OPLSAA conformers with PBE+vdW (OPLSAA energy range: 0.26 eV) and *light* settings. As a general remark, all DFT calculations discussed here are performed with the "Fritz Haber Institute *ab initio* molecular simulation" (`FHI-aims`) code[257]. Structure relaxations based on DFT are always performed in two steps. First, the structures are relaxed with *light* computational settings and afterwards the lowest-energy structures are further relaxed with *tight* computational settings, where the forces are converged down to $5 \cdot 10^{-3}$ eV/Å. This *light* $\rightarrow$ *tight* cascade saves much computation time as one self-consistent field (SCF) iteration with *light* settings takes about 10 % of the CPU time of an iteration with *tight* settings (for this particular system). All energies that are explicitly reported or discussed in this thesis are calculated with *tight* computational settings unless stated otherwise.

Figure 8.3 shows the energy hierarchies obtained with the force-field and the PBE+vdW functional (*light* and *tight* computational settings). Between OPLSAA and PBE+vdW (*light* settings) significant rearrangements occur. On the other hand, the changes between *light* and *tight* settings are very small. Thus, it is sufficient to relax only the lowest-energy regime with *tight* computational settings.

In the following, we analyze the correlation between the force field and PBE+vdW (*light* settings) more closely. This allows us to assess the possibility whether any relevant structure type might be overlooked by not relaxing with PBE+vdW all other conformers that have even higher OPLSAA energies. As a first step, we relaxed a second batch of OPLSAA conformers with PBE+vdW, choosing them from the remaining pool of structures in intervals of 50 according to the force-field energy hierarchy. The PBE+vdW and OPLSAA energy hierarchies are illustrated in Fig. 8.4. In the figure, the OPLSAA conformers are ranked according to their relative force-field

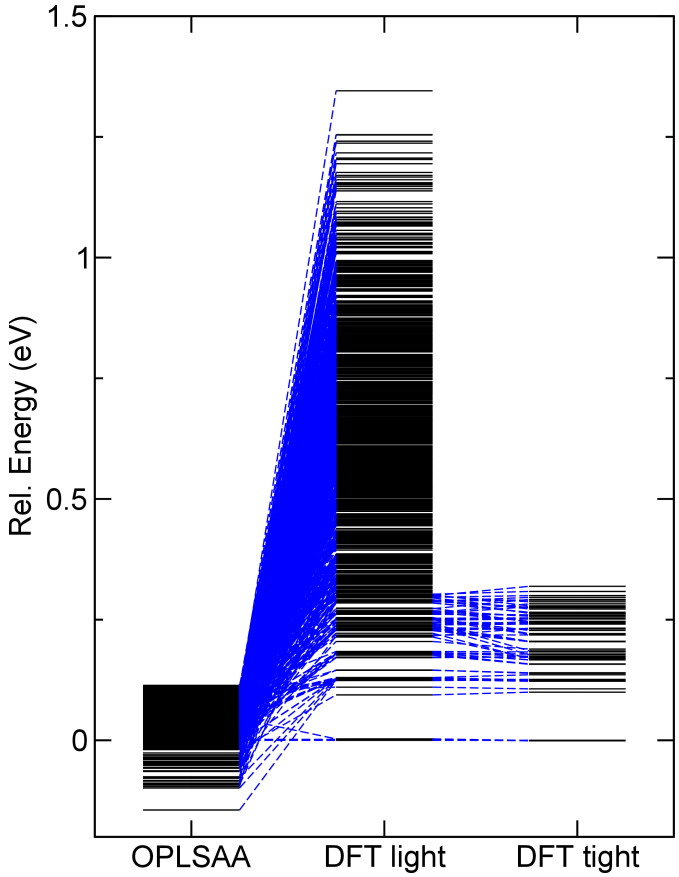


Figure 8.3: Energy hierarchies of Ac-Lys-Ala$_{19}$ + H$^+$ (black horizontal bars) obtained with the OPLSAA force field and with PBE+vdW *light* and *tight* computational settings. The energies are given relative to the lowest-energy PBE+vdW minimum structure.

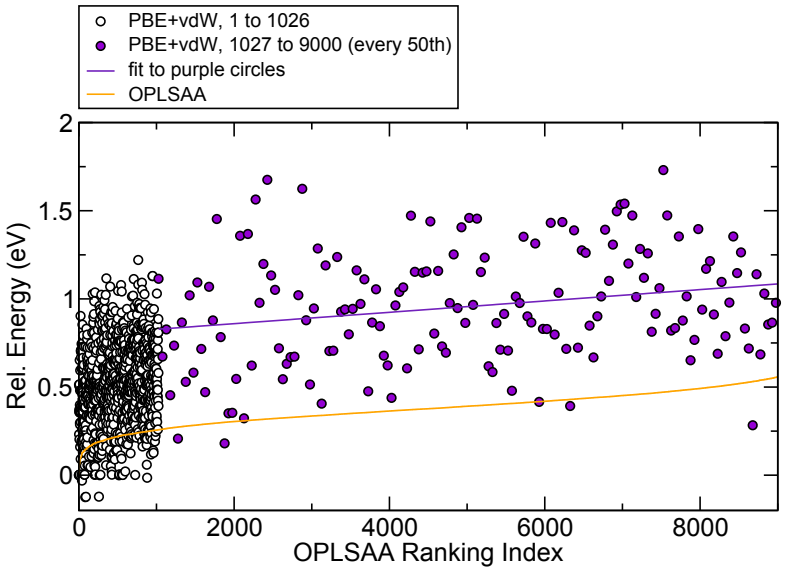


Figure 8.4: Correlation between the OPLSAA (orange line) and the PBE+vdW energy hierarchies (circles) for the structures obtained from the Ac-Lys-Ala_{19} + H^+ REMD trajectory run at 300 K. Each circle represents a PBE+vdW minimum relaxed from the force-field minimum with the corresponding ranking index. White circles: PBE+vdW conformers relaxed from force-field conformers with ranking indices 1 to 1026. Purple circles: PBE+vdW conformers relaxed from force-field conformers with ranking indices between 1027 and 9000 taken in intervals of 50. Purple line: linear fit to the data represented by the purple circles. All energies are given relative to the energy of the lowest-energy force-field conformer [ranking index 1] or the PBE+vdW minimum following from the relaxation of the latter.

energy, i.e., the lowest-energy structure is assigned index 1, the second lowest-energy structure is assigned index 2 and so on. The orange line in Fig. 8.4 shows the relative force-field energy of the conformers as a function of the ranking index. The white circles represent the outcome of the PBE+vdW relaxations from the OPLSAA conformers with ranking indices 1 to 1026 and the purple circles represent the same for the second batch of PBE+vdW relaxations (intervals of 50). All energies are given relative to the structure with ranking index 1. First of all, one has to take into account that the offset between the data sets is determined by the conformer that is chosen as the reference. This is arbitrary and the important point is to what extent the PBE+vdW data (circles) resemble the shape of the OPLSAA data (line). For the first batch (indices 1 to 1026, white circles) hardly any correlation between the OPLSAA and the PBE+vdW results is visible. However, taking the whole plot into account, especially the larger ranking indices, there clearly is some weak correlation. More specifically, a linear fit to the second batch of PBE+vdW data (purple line) shows an ascending tendency with increasing ranking index similar to the OPLSAA data curve. Of course, there is a large scatter, but it seems at least justified to weakly trust the OPLSAA force field energy hierarchy (for large energy differences). Another point that is reassuring with respect to (weakly) trusting the force-field hierarchy is that the lowest-energy PBE+vdW structure is found relatively early, corresponding to the OPLSAA ranking index of 75.

In order to see how similar the structures in the low-energy DFT regime are, we sorted them into clusters in the same way as described earlier (with a cut-off criterion of 1 Å). This is illustrated in Fig. 8.5a. All clusters with members found within 200 meV of the lowest-energy PBE+vdW structure are displayed by colored circles, where the same color denotes that the structures belong to the same cluster. The lowest-energy PBE+vdW structure belongs to the

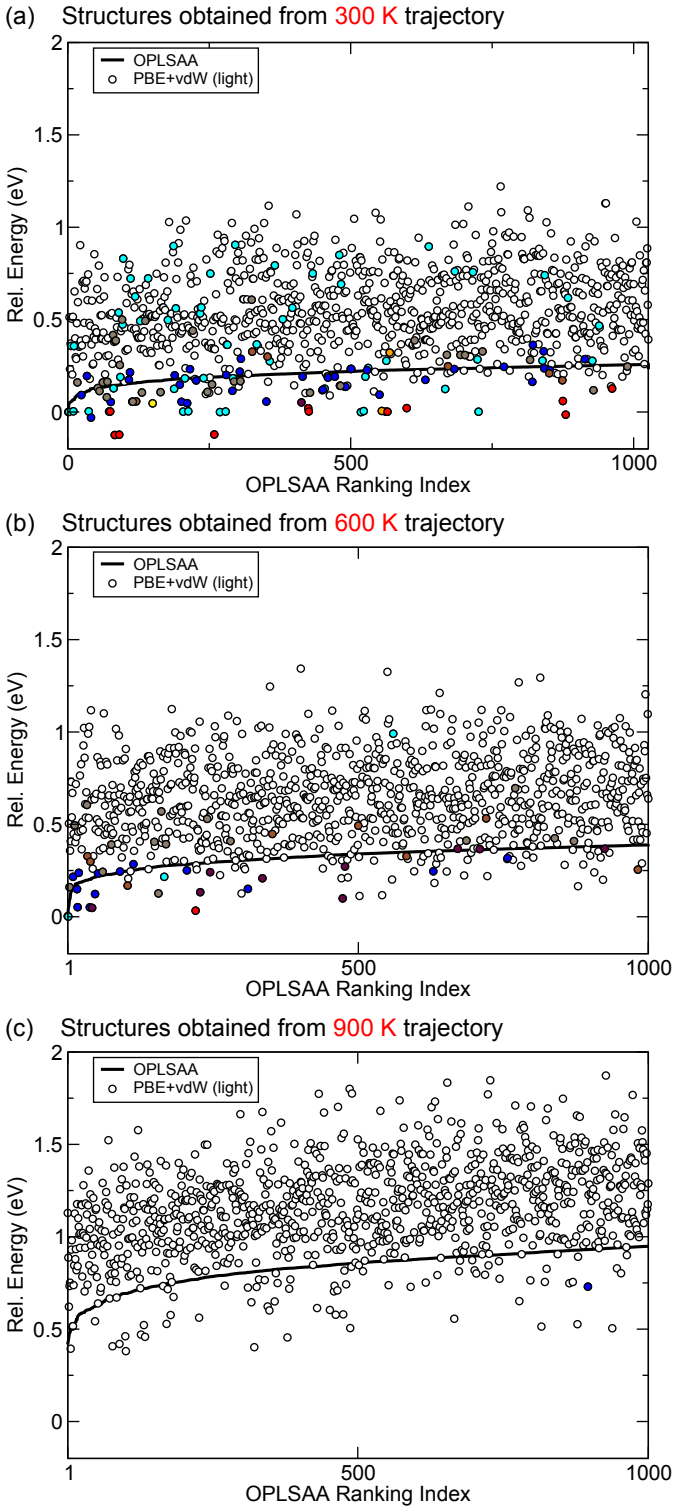


Figure 8.5: Correlation between the OPLSAA (black straight line) and the PBE+vdW energy hierarchies (circles) for the structures obtained from the Ac-Lys-Ala$_{19}$ + H$^+$ REMD trajectories run at 300 K (a), 600 K (b), and 900 K (c). Each circle represents a PBE+vdW minimum relaxed from the force-field minimum with the corresponding ranking index. All energies are given relative to the energy of the lowest-energy force field conformer [ranking index 1 in subplot (a)] or the PBE+vdW minimum following from the relaxation of the latter. All PBE+vdW structures were clustered with the algorithm given in Ref. [99] and a cut-off criterion of 1 Å. The clusters with members that have energies within 200 meV of the lowest-energy PBE+vdW minimum are displayed with colored circles, where the same color denotes that the structures belong to the same cluster.

"red" cluster (OPLSAA ranking index of 75). The good news here is that all low-energy clusters (except for the orange one, for which the first member was found at an index of 555) were found within ranking indices of less than 500, i.e., less than half of the ranking indexes we took into account. As a conclusion, we consider it very likely that the lowest-energy structure types have been identified, although it is impossible to exclude that there exist large errors in the force field possibly causing the problem that a structure type low in PBE+vdW energy might be overlooked after all.

Another issue to be addressed is that so far we only considered the structure snapshots taken from the 300 K trajectory, while in principle we have 16 trajectories available. In order to check how much additional information we can infer from the other trajectories (or how important they are), we applied the same clustering and relaxation strategy to the trajectories that were run at 600 K and 900 K. For both cases, we relaxed the 1,000 lowest-energy OPLSAA conformers. The results are displayed in Fig. 8.5b) and c), respectively. The OPLSAA ranking index for the different trajectories is chosen according to the force-field hierarchy of the OPLSAA structures obtained from the corresponding trajectory. The energies, however, are all given relative to the lowest-energy OPLSAA structures obtained in total (ranking index 1 of the results for 300 K) and its corresponding DFT structure. Neither from the structures obtained from the 600 K trajectory nor from the 900 K trajectory, the lowest-energy PBE+vdW structure obtained from the 300 K trajectory was found. In addition, the lowest-energy structures obtained from the 600 K trajectory all belong to clusters that had been found from the 300 K trajectory as well. From the 900 K trajectory we rather find only very high-energy conformers. We take this as an indication that it should be sufficient to concentrate on the 300 K trajectory.

8.1.2.1 First-principles REMD simulations

In the next step of our strategy, we follow up with DFT-based REMD simulations for the four lowest-energy PBE+vdW structure types. For this, we employed 16 replicas in the temperature range between 300 K and 623 K. Although larger swap attempt frequencies might be even more efficient[301], given the computational cost involved with each attempt, we here chose a swap attempt frequency of 1 per 100 time steps (time step: 1 fs). After each ps of REMD simulation time, all replicas were relaxed with PBE+vdW. Figure 8.6 shows a particular case, where we found many structures that were lower in energy than the starting geometry by significant energy differences (up to $\approx$ 250 meV). All replicas were initialized by the initial geometry depicted at the top of Fig. 8.6 and whose energy is taken as the reference zero. The plot shows the energies of all relaxed replicas as a function of the simulation time. The lowest-energy structure that is found is likewise depicted at the top of Fig. 8.6 and labelled as C2 (the reason for this labelling will become clear below). C2 is more than 200 meV lower in energy than the initial geometry. Overall the two structures are very similar (RMSD of 0.6 Å). However, small rearrangements occurred close to the termini and the turn. As a conclusion, this means that despite the limited time scales computationally feasible for *ab initio* REMD, the approach is able to lead to refinements of the local structural environment along with lower-energy structures.

In order to further analyze the sampling ability of the *ab initio* REMD run, we plotted the Ramachandran plot of the dihedral-angle pairs of all relaxed replicas obtained after a given simulation time (Fig. 8.7). The first plot (00 ps) shows the dihedral angle pairs for the initial geometry, while with increasing simulation time the sampled dihedral angle space increases.

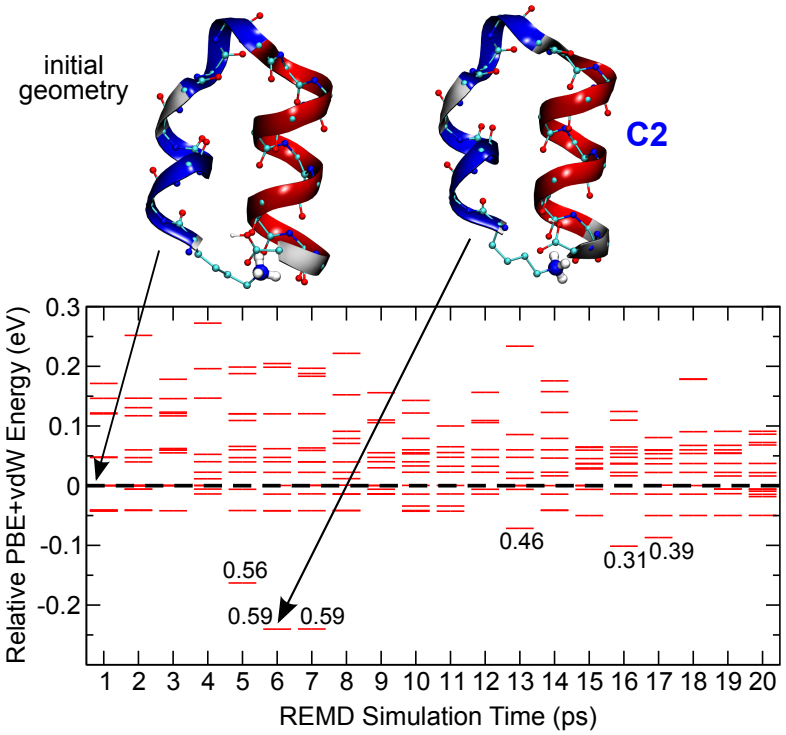


Figure 8.6: DFT-based REMD simulation, where lower-energy structures than the initial geometry were found. All replicas were initialized by the "initial geometry". After each ps all 16 replicas were relaxed with PBE+vdW (*light* computational settings). The energies of the relaxed replicas (red bars) are given relative to the energy of the initial geometry. For selected structures, the RMSD (in Å) with respect to the initial geometry is written at the corresponding bar.

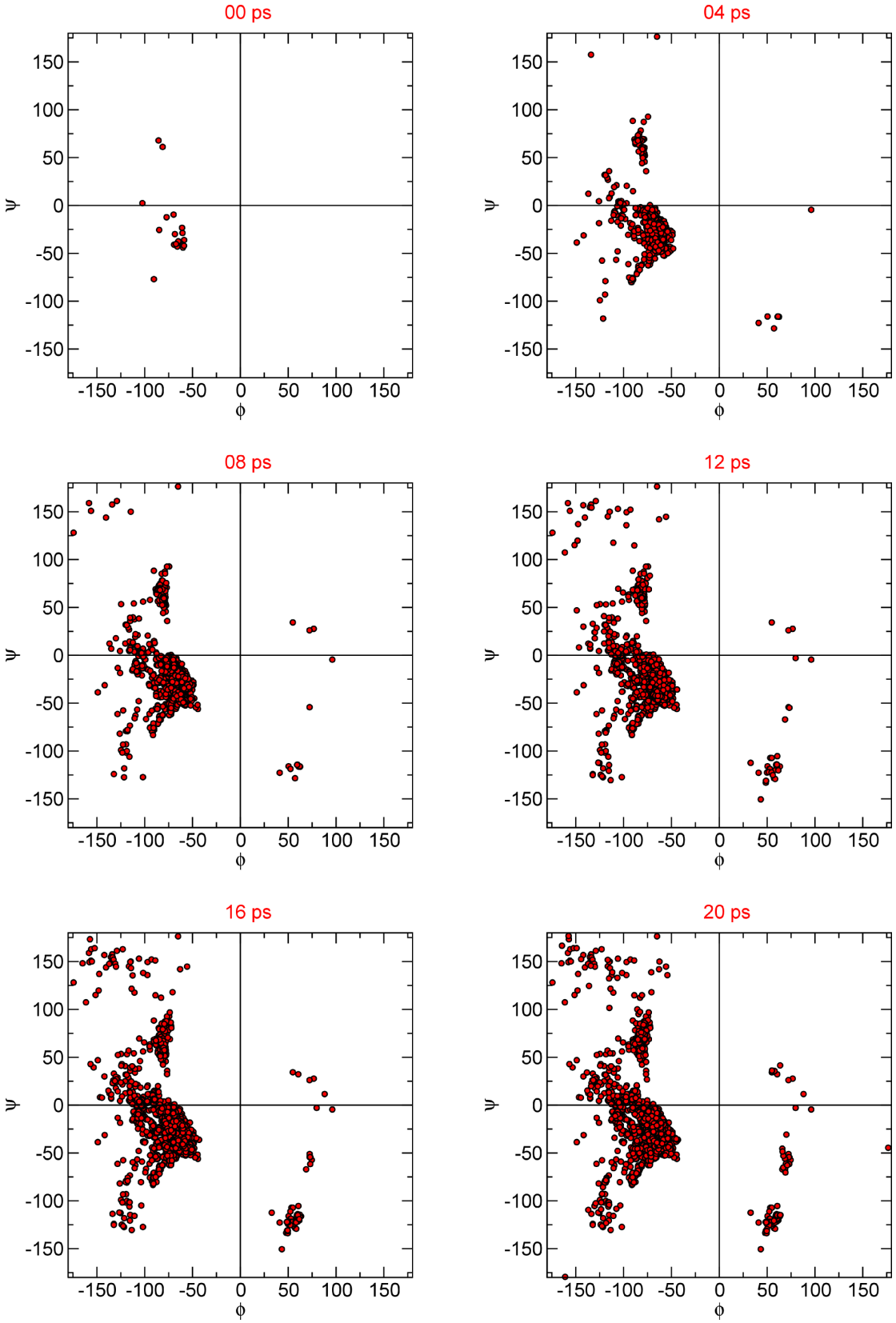


Figure 8.7: Ramachandran plots of PBE+vdW relaxed replicas obtained from the DFT-based REMD search for Ac-Lys-Ala_{19} + H^+ (see Fig. 8.6). The plots show the dihedral-angle pairs for all structures obtained up to the given simulation time.

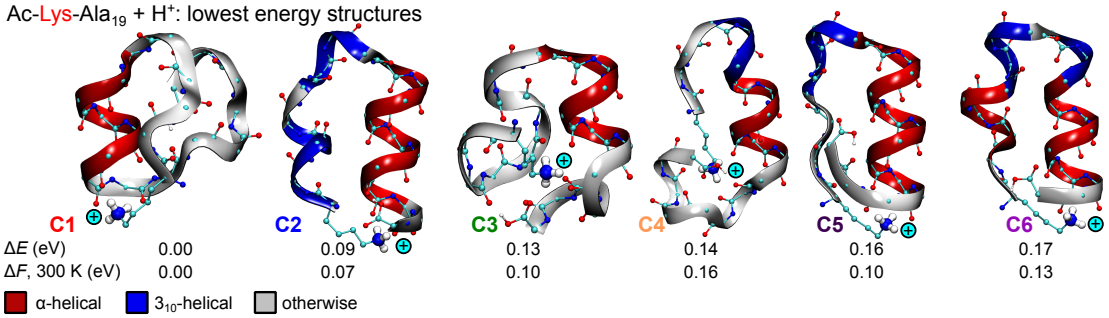


Figure 8.8: Structure types obtained from our structure search for Ac-Lys-Ala$_{19}$ + H$^+$ monomers. The ribbon color denotes the helix type, namely α-helical (red), 3_{10}-helical (blue), or otherwise (grey). The structures are labelled C1 to C6 according to their energy hierarchy. Their relative energies and free energies (at 300 K) are given below the structural representations. All energies are given relative to C1. The free energies are calculated using the harmonic oscillator-rigid rotor approximation (see Section 5.1.1). The "plus" sign indicates the positive charge located at the lysine residue.

Table 8.2: Sum of helical hydrogen bonds (α- and 3_{10}) for all structure types C1 to C6 of Ac-Lys-Ala$_{19}$ + H$^+$ in comparison to an ideal α-helix. A hydrogen bond is defined to be present if the distance between a hydrogen atom and an acceptor oxygen is less than 2.5 Å and if the angle $\sphericalangle$(O,H,N) is larger than 150°.

C1	C2	C3	C4	C5	C6	α-helix
5	13	5	8	10	11	17

8.2 Structure classification

In total, we obtained more than 4000 PBE+vdW structures for Ac-Lys-Ala$_{19}$ + H$^+$ from our conformational search. Most notably, the peptide is large enough so that the structures are composed of more than one secondary structure element. For a classification, we focus on these elements, particularly on helical hydrogen bonds (α-helical or 3_{10}-helical hydrogen bonds, or otherwise). We find six structure prototypes within the lowest 170 meV of the global minimum. They are depicted in Fig. 8.8 together with their energies, where the α-helical parts are highlighted by red ribbons and the 3_{10}-helical parts are color coded by blue ribbons. According to their energy hierarchy the structures are labelled C1 to C6. Relative free energies calculated in the harmonic oscillator-rigid rotor approximation at 300 K (see Section 5.1.1) are also given in Fig. 8.8.

The C1 structure type contains an α-helical part, approximately in the middle of the peptide chain. The ends of the strand are arranged in an antiparallel fashion. C2 consists of an α-helical and a 3_{10}-helical part connected by a turn. C3 also contains an α-helical fragment, where the N-terminal part of the peptide chain forms a loop. For C4, the whole peptide chain forms a loop comprising an α-helical section with a 3_{10}-helical part at its N-terminal end. C5 is comprised of an α-helical segment and a 2_7-strand connected by a turn. Finally, C6 contains two α-helical segments with some 3_{10}-helical hydrogen bonds at the turn connecting the two helices. The number of helical (both α and 3_{10}) hydrogen bonds for each structure are given in Tab. 8.2 in comparison to an ideal α-helix. From this follows that C1 and C3 are the least helical structure types, while C2 has the highest helical content. All structure types share a common stabilization motif. The lysine side chain with its protonated amine group wraps around the peptide and caps the C-terminal part (negative end) of the α-helical segment[26].

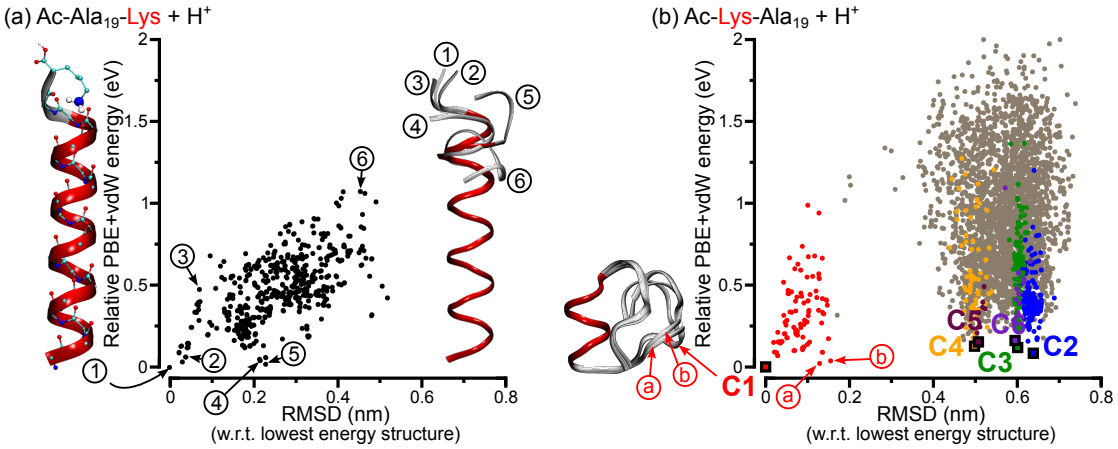


Figure 8.9: Relative PBE+vdW energies as a function of the RMSD (with respect to the lowest-energy conformation) for all structures obtained in the structure searches for Ac-Ala$_{19}$-Lys + H$^+$ (a) and Ac-Lys-Ala$_{19}$ + H$^+$ (b). (a) Ac-Ala$_{19}$-Lys + H$^+$: Each PBE+vdW minimum of the potential-energy surface (PES) is depicted by a black circle. The global minimum (denoted as 1) is shown at the left side of the plot. At the right side of the plot the backbone ribbon representation of six example structures, labelled as 1 to 6, are illustrated. The backbone atoms of the residues Ac to Ala$_{14}$ are fitted to the global minimum (structure 1). (b) Ac-Lys-Ala$_{19}$ + H$^+$: Each PBE+vdW minimum of the PES is depicted by a brown circle. The structure type representatives C1 to C6 are highlighted by squares. All structures with an RMSD of less than 1.6 Å with respect to one of these structure types are marked in the corresponding color. On the left side of the plot the lowest-energy structure C1 is shown together with the structure examples *a* and *b*. The backbone atoms of residues Ala$_6$ to Ala$_{11}$ (α-helical segment) of structures *a* and *b* are fitted onto C1.

8.3 Energy landscapes: Ac-Lys-Ala$_{19}$ + H$^+$ vs. Ac-Ala$_{19}$-Lys + H$^+$

Figure 8.9(b) illustrates the relative PBE+vdW energy of all conformers found in the structure search for Ac-Lys-Ala$_{19}$ + H$^+$ as a function of the RMSD with respect to the global minimum (C1). The structure types C1 to C6 are very close in energy, while they are very distant in their structure (measured by the RMSD). Three conformers of the C1 structure type are exemplified and compared to each other at the left side of the plot. The large gap in conformational space separating C1 from the other structures (based on the RMSD) could indicate the presence of an energy barrier.

For Ac-Ala$_{19}$-Lys + H$^+$ with the protonated lysine residue located at the C-terminus, the scenario is completely different (cf. Fig. 8.9a). All structures are α-helical. The lowest-energy structure is depicted on the left side of the plot. It is an ideal α-helix, where the lysine side chain caps the dangling carbonyl oxygens close to the C-terminus. In order to demonstrate that all structures are basically α-helical, Fig. 8.9(a) shows the backbone ribbons of six exemplary structures at the right side of the plot aligned on top of each other. All structures are α-helical with only slight deviations close to the C-terminus. The relative energy rises with increasing RMSD, which lets us identify Ac-Ala$_{19}$-Lys + H$^+$ as a structure seeker with only one folding funnel. As mentioned earlier in this chapter, the structure search for Ac-Ala$_{19}$-Lys + H$^+$ was analogous to the one for Ac-Lys-Ala$_{19}$ + H$^+$. Just as for Ac-Lys-Ala$_{19}$ + H$^+$, we started from a perfect α-helical conformation and carried out an OPLAA-based REMD simulation with 16 replicas. The total simulation time was again 8 µs. From the clustering approach of the 300 K trajectory we obtained 464 conformers, which we all relaxed with PBE+vdW (OPLSAA energy

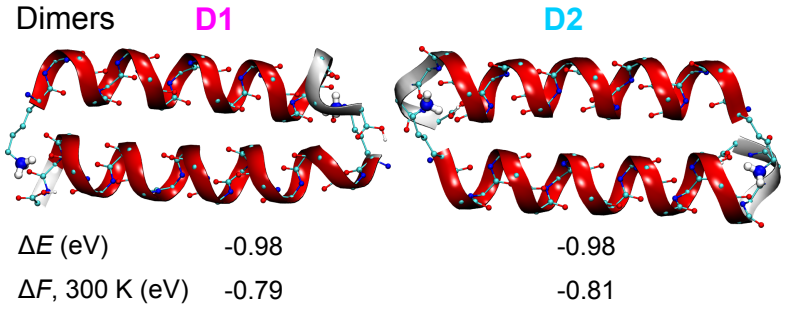


Figure 8.10: Structural representations of the two lowest-energy helical dimers found in the conformational search. Red ribbons indicate an α-helical conformation. For both structures the energies and free energies at 300 K (per monomer) are given with respect to C1 (cf. Fig. 8.8). The free energies are calculated in the harmonic oscillator-rigid rotor approximation.

range: 0.55 eV). Following up on that, we carried out a DFT-based REMD simulation for the lowest-energy DFT conformer obtained from the preceding PBE+vdW relaxation of the OPLSAA conformers.

8.4 Helical models: Ac-Lys-Ala$_{19}$ + H$^+$

As discussed in Section 7.3, the IM-MS data for Ac-Lys-Ala$_{19}$ + H$^+$ in Fig. 7.4 reveal that apart from compact monomers there are small amounts of Ac-Lys-Ala$_{19}$ + H$^+$ dimers and helical monomers present in the experiment. For this reason, we also performed structure searches for helical monomers and helical dimers.

8.4.1 Helical dimers: Ac-Lys-Ala$_{19}$ + H$^+$

The structure search for helical dimers was analogous to the search for the Ac-Lys-Ala$_{19}$ + H$^+$ monomers described in Section 8.1. As initial geometry for the force field-based REMD simulations, we chose an ideal coiled-coil structure[434] from the protein data bank (structure ID: 1SER, where we used only the coiled-coil part of the structure). With this structure we initialized 22 replicas and ran the REMD simulation for 200 ns (total simulation time: 4.4 µs). Just as before, we extracted snapshots of the 300 K trajectory (every 2 ps) and clustered them with the algorithm described above and a cut-off criterion of 0.5 Å. Relaxing the midpoint structure of each cluster with OPLSAA resulted in 2,180 conformers, from which we relaxed the 96 lowest-energy ones with PBE+vdW (OPLSAA energy range: 0.20 eV). The dimer conformations differ slightly in the terminations, the angle between the two helical axes, and the shift between the helical monomers (along the helix axis). The two lowest-energy PBE+vdW dimers are depicted in Fig. 8.10. For both dimers, the energy per monomer is lower than the energy of C1, the lowest-energy monomer. However, dimer formation depends on the partial pressure of the monomers (see Eq. 2.3), which is low and thus makes dimer formation (in the gas phase) rather unlikely. As discussed in Section 7.3, if dimers are formed, they most probably form during the electrospray process depending on the precise experimental conditions. In the IM-MS experiment shown in Fig. 7.4, a small amount of dimers is present. However, for the IRMPD set-up the experimentalists explicitly checked for the existence of helical dimers and concluded that they are not populated to a measurable extent. For this, they utilized a mixture of equal

amounts of Ac-Lys-Ala$_{19}$ + H$^+$ and For-Lys-Ala$_{19}$ + H$^+$, where in the latter case the N-terminus is capped by a formyl (For) group instead of an acetyl (Ac) group. Due to the different masses of For and Ac, the corresponding monomers can be distinguished by mass spectrometry. On the other hand, monomers of the same type and (doubly charged) dimers of the same monomer cannot be distinguished (same mass over charge ratio). However, mixed Ac-Lys-Ala$_{19}$ + H$^+$/For-Lys-Ala$_{19}$ + H$^+$ dimers would show a separate peak in the spectrum so that the existence of such dimers can be directly tested. If Ac-Lys-Ala$_{19}$ + H$^+$/Ac-Lys-Ala$_{19}$ + H$^+$ dimers are formed in experiment, Ac-Lys-Ala$_{19}$ + H$^+$/For-Lys-Ala$_{19}$ + H$^+$ dimers should form as well. When electrospraying the mixture at the same experimental conditions as used to measure the infrared (IR) spectrum of Ac-Lys-Ala$_{19}$ + H$^+$, exclusively isolated monomers were observed. This led to the conclusion that helical dimers should most probably not contribute to the IR spectrum of Ac-Lys-Ala$_{19}$ + H$^+$ shown in Fig. 7.4.

8.4.2 Helical monomers: Ac-Lys-Ala$_{19}$ + H$^+$

For the helical monomers, the proton is most likely associated with a carbonyl oxygen close to the C-terminus as this allows a favorable electrostatic interaction of the charge with the helix dipole[26, 27]. However, it is not *a priori* clear where it is exactly located. Hence, it is necessary to allow the proton to hop between different positions during the conformational search. As most force fields, including the OPLSAA force field, cannot describe bond breaking, a different search strategy as used for the dimers and monomers has to be employed. Since we look for overall helical structures, the search space is rather narrow and we sampled it with pure DFT (PBE+vdW) based simulations. As starting geometries for the *ab initio* REMD simulation, we chose three structures where the proton is located at different carbonyl oxygen atoms. In order to construct the latter, we used an ideal α-helical structure with a straight lysine side chain as a template. We then manually placed the proton at the carbonyl oxygen of the 16*th*, the 17*th*, and the 18*th* alanine residue, respectively, and followed up with a geometry optimization. These three initial structures are depicted on the left side of Fig. 8.11. In total, we employed 18 replicas in the temperature range between 300 K and 688 K, where the starting geometry for each replica was alternately chosen from those three conformations. All other parameters were the same as for the other *ab initio* REMD simulations. As before, after each ps of simulation time all replicas were relaxed with PBE+vdW. Figure 8.11 shows their relative energies as a function of simulation time relative to the lowest-energy initial structure. The lowest-energy structure found in this conformational search is depicted at the right side of Fig. 8.11. It is about 370 meV lower in PBE+vdW energy than the lowest-energy initial structure. This example shows that, despite the limited time scale accessible, *ab initio* REMD is able to lead to reasonable rearrangements of the structure. The lysine side chain is bent to interact with the acetyl group, while it was straight in the initial conformers. Moreover, the proton is located at a position where the C-terminal carbonyl group can interact with it.

For a comparison, we also performed an *ab initio* REMD simulation started from a helical conformation with the proton located at the N-terminal lysine residue. Figure 8.12 compares lowest-energy helix found for this type (H$^+$ at Lys) and for the case where the proton is located close to the C-terminus. As indicated in the figure, the location of the proton at the N-terminus leads to an unfavorable interaction of the charge with the helix dipole. This agrees with the high energy of 2.88 eV relative to the lowest-energy compact monomer C1. The helical conformation

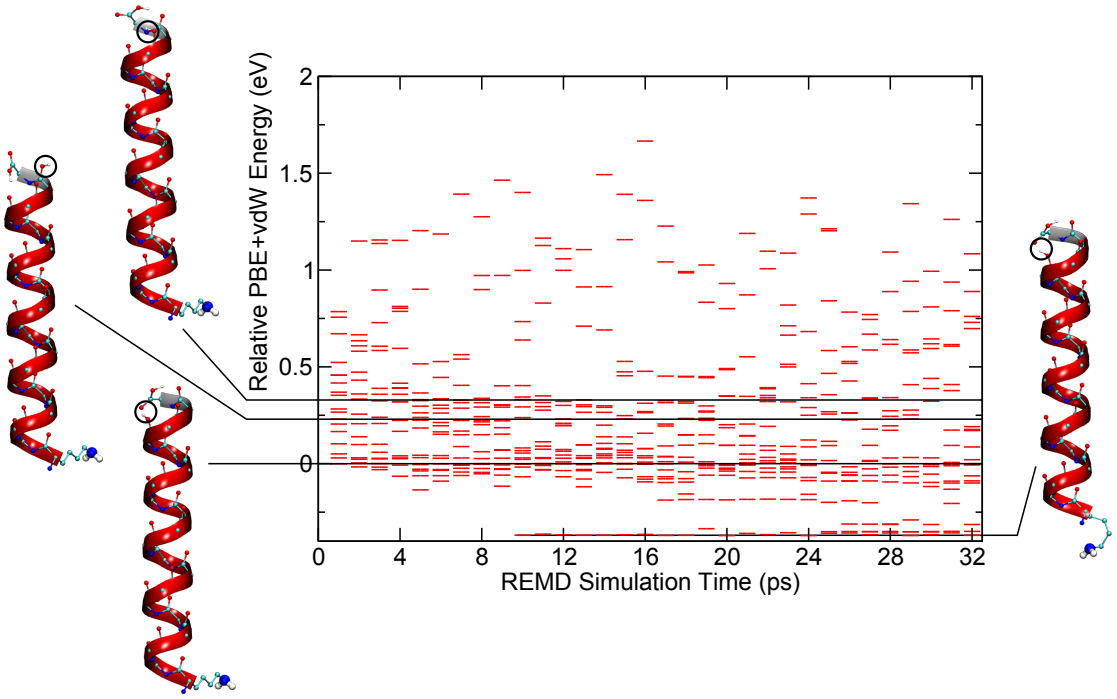


Figure 8.11: *Ab initio* REMD simulation for Ac-Lys-Ala$_{19}$ + H$^+$ based on PBE+vdW. At the left side of the plot the three helices that were used to initialize the replicas are shown. The starting geometry for each of the 18 replicas was alternately chosen from the three initial helical structures. After each ps all replicas were relaxed with PBE+vdW. The plot shows their energies (red bars) relative to the lowest-energy initial helix as a function of the simulation time. The black lines denote the energies of the respecting conformer. The lowest-energy conformer found in this search is depicted at the right side of the plot. In the structural representation the position of the proton is highlighted by a black circle.

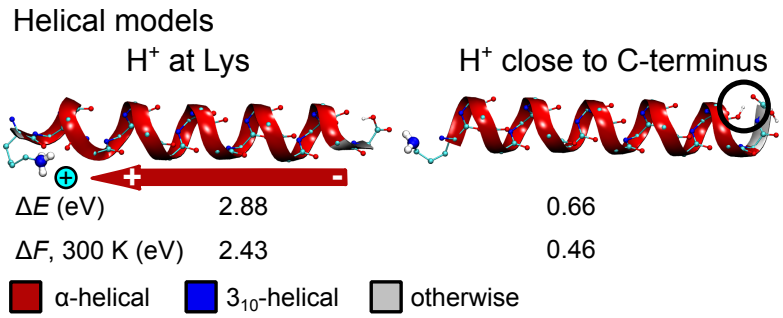


Figure 8.12: Structural representations of the lowest-energy helical models for Ac-Lys-Ala_{19} + H^+, both for the case where the proton is located at the N-terminal lysine (left) and close to the C-terminus (right). Red ribbons indicate an α-helical conformation. For both structures the energies and free energies at 300 K are given with respect to C1 (cf. Fig. 8.8). The free energies are calculated in the harmonic oscillator-rigid rotor approximation.

with the proton located close to the C-terminus has a lower energy (0.66 eV), but it is still relatively high implying that this helix type would not be populated in experiment. On the other hand, we find a small helix contribution in the IM-MS measurements (see Section 7.3). This issue will be discussed further in the subsequent sections.

8.5 Role of a higher-level force field

Compared to force-field calculations, DFT calculations are much more expensive. From a general point of view, DFT is a higher level of theory than force fields. However, one question that arises is whether the DFT (PBE+vdW in this case) calculations are really needed or if a force field could actually yield the same results (despite the generally lower level of theory). Considering the comparison between the energy hierarchies obtained with the OPLSAA[143] force field and and PBE+vdW for Ac-Lys-Ala_{19} + H^+ (see Fig. 8.3), we clearly see that OPLSAA and PBE+vdW give completely different answers. However, the OPLSAA force field is just one out of many force-field parametrizations. In fact, it was recently shown that the Amber99sb[153] and the AmoebaPro04[154] force fields perform relatively well for a benchmark set of CCSD(T) energies of 27 Ac-Ala_3-NMe conformers[247]. They yielded similar mean absolute errors (MAEs) as the DFT PBE and PBE0 functionals (*without* van der Waals (vdW) corrections). On the other hand, the OPLSAA[143] and Charm22[435] force fields performed significantly worse. These results for a peptide similar to the one considered in this work (Ac-Lys-Ala_{19} + H^+) suggest to test at least the Amber99sb and the AmoebaPro04 force fields. As we base our study in the gas phase, the peptide of interest, Ac-Lys-Ala_{19} + H^+, is protonated at the C-terminus, which is usually not the case for peptides in solution (water). For this reason, most force fields, including Amber99sb and AmoebaPro04, lack parameters for the protonated C-terminus. Here we face one of the limitations of force fields, namely that their range of validity is determined by the training set employed for the parametrization process. However, the most recent parametrization of the Amoeba force field (AmoebaPro13[154]), as implemented in the version 6.2 of the TINKER program[155], includes parameters for the protonated C-terminus. Apart from that, the class of Amoeba force fields are higher-level force fields as they are polarizable, i.e., they do not use the fixed-charge model employed by most common force fields (including the Amber force fields). For these reasons, we concentrated on the AmoebaPro13 force field here and re-calculated the

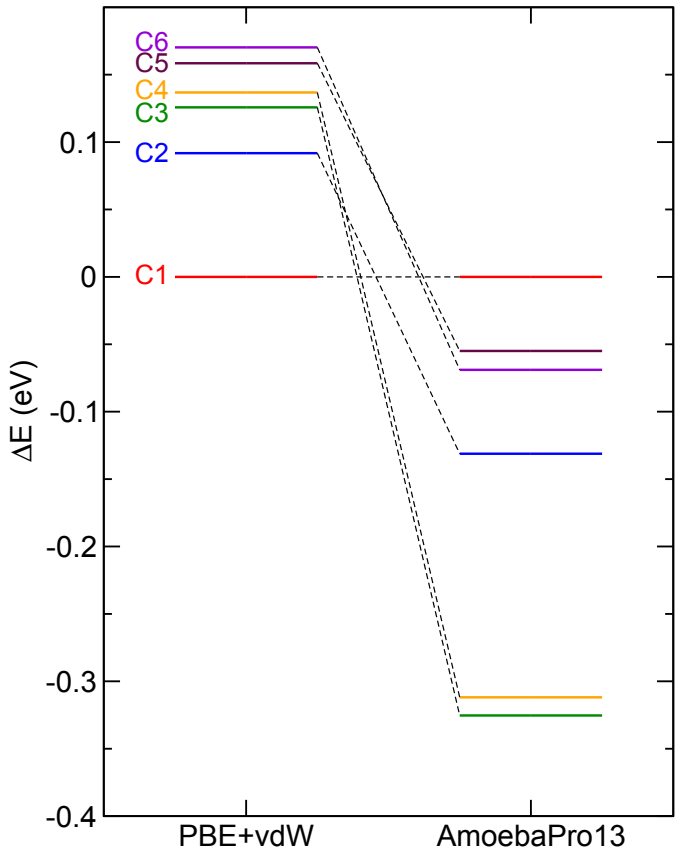


Figure 8.13: Energy hierarchies (horizontal bars) obtained with the PBE+vdW functional and the AmoebaPro13 force field for the Ac-Lys-Ala$_{19}$ + H$^+$ structure types C1 to C6. All energies are given relative to C1 and the dashed lines serve as a guide to the eye.

energy hierarchies of the most important structure types C1 to C6 using this force field. In principle, it would be interesting to consider also the helical structure with the proton located close to the C-terminus as this structure was shown to be populated (to a small extent) in the IM-MS measurements (cf. Fig. 7.4). However, we again face the problem that the corresponding parameters for such a structure are missing in the force field so that we focus only on C1 to C6. We also relaxed the structures in order to compare the energies of the corresponding local PES minima. However, the relaxation changed the structures relatively little with RMSDs of less than 0.45 Å, where the exact RMSD for all conformers can be found in Tab. A.1 in Appendix A. The energy hierarchies for C1 to C6 obtained with AmoebaPro13 and PBE+vdW are compared in Fig. 8.13. The energy hierarchies change quite significantly. While C1 is the lowest-energy conformer in PBE+vdW, it becomes the highest-energy one in AmoebaPro13. On the other hand, AmoebaPro13 predicts C3 to be the most probable conformer, closely followed by C4. Notably, according to AmoebaPro13, C3 is more than 300 meV lower in energy than C1. From this comparison we can conclude that with AmoebaPro13 we would have definitely not arrived at the same answer than with PBE+vdW. Since either method may have errors, it is important to remember that we cannot state with full certainty which one is correct from this comparison alone. That said, we take PBE+vdW as the trusted method as (i) it is the higher level of theory (first principles) and (ii) in the aforementioned benchmark[247] against CCSD(T) data it performed clearly better than AmoebaPro04 (the performance of AmoebaPro04 was similar to the bare PBE

functional without a vdW correction). We will see below that different DFT functionals change the energy hierarchy as well, but the scatter is much narrower. The structure predictions will be critically assessed by comparison to experimental fingerprints in the next chapter.

8.6 Impact of different functionals and dispersion corrections

A common question that arises in the context of DFT is how much the result is actually affected by the approximation to the DFT exchange-correlation functional that is used. Here we assess this problem by testing the influence of different functionals. As discussed in Section 3.5.3.2, a recent benchmark by Rossi *et al.*[247] found that the PBE0+MBD* functional yields excellent results for Ac-Phe-Ala$_5$-Lys(H$^+$). Of all functionals tested, including a recent study of 19 different semi-local and hybrid DFT exchange-correlation functionals by Xie *et al.*[252], PBE0+MBD* comes closest to explaining the experimental results for Ac-Phe-Ala$_5$-Lys(H$^+$), namely the co-existence of specific conformers and their relative abundances. For this reason, we here concentrate on an assessment of the PBE[15] and the PBE0[205, 206] functionals. As it is well established that for peptides it is important to take vdW interactions into account[44, 338, 436], we include in all cases a long-range vdW dispersion correction. Specifically, we employed two schemes, which were discussed in detail in Section 3.5.3. The first one is the pairwise TS scheme[16], denoted as "+vdW" in the functional description, and the second one is the recently developed many-body correction scheme[238, 239] MBD@rsSCS or MBD* for short. This means that we assessed in total four functionals: PBE+vdW, which we used for our conformational search, PBE+MBD*, PBE0+vdW, and PBE0+MBD*. In principle, we would like to compare the energy hierarchies of all conformers that we have found with our PBE+vdW based search with all other functionals. However, especially for the PBE0+vdW and PBE0+MBD* functionals this would be too expensive. Thus, for a proof of concept, we concentrated on PBE+MBD* first and re-calculated (including relaxation) the energies of all PBE+vdW conformers obtained from the force-field search (> 1000 conformers) with PBE+MBD*. Within the lowest 200 meV, we did not find any conformer different from C1 to C6. For this reason, we focused for a further analysis only on C1 to C6. Additionally, we also included the helical conformer with the proton located close to the C-terminus, as this is also seen to be present (to a small extent) in the IM-MS experiments (see Fig. 7.4). In order to compare local minima of the corresponding PES, we also relaxed all structures with the respective functionals. For the PBE-based functionals, we used *tight* computational settings[257], while for the PBE0-based functionals we used (for reasons of computational feasibility) *LVL-intermediate* settings[1] and then followed up on this by single-point calculations with full *tight* computational settings. The structural changes upon relaxation are marginal in all cases, with RMSD values of less than 0.1 Å. The detailed RMSD values for all conformers are listed in Tab. A.1 in Appendix A. The energy hierarchies obtained with PBE+MBD*, PBE+vdW, PBE0+vdW, and PBE0+MBD* are illustrated in Fig. 8.14, where the results for PBE+vdW, i.e., the reference data obtained from the search, are highlighted with a grey background. Starting from the reference PBE+vdW data and exchanging the pairwise

[1]These are in principle *tight* computational settings, but with a smaller basis set. The basis set includes tier1 and the first basis function from tier2[257] together with additional basis functions for the auxiliary basis set, which is used to expand the basis products needed for the evaluation of the exchange integral.

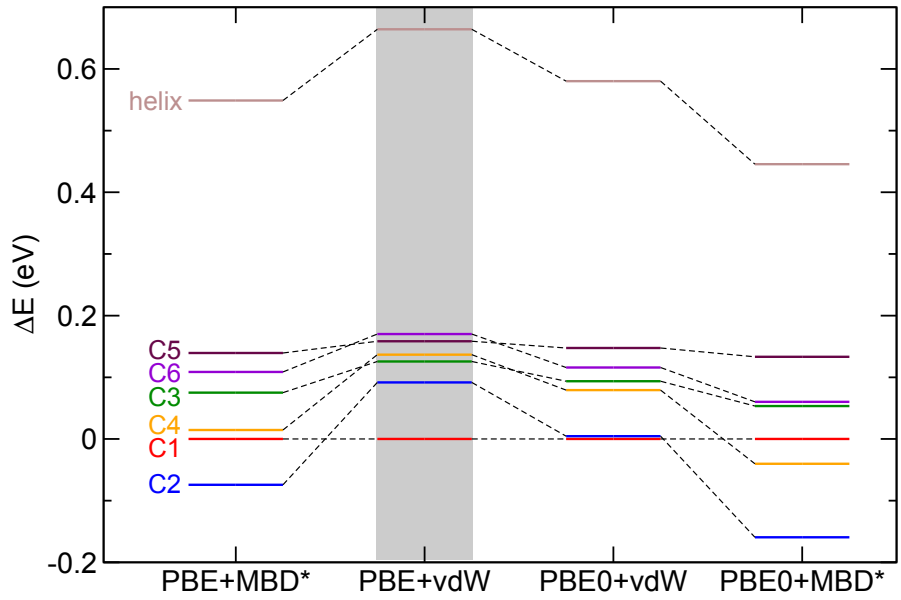


Figure 8.14: Energy hierarchies (horizontal bars) obtained with the PBE+MBD*, PBE+vdW, PBE0+vdW, and PBE+MBD* functional for the Ac-Lys-Ala$_{19}$ + H$^+$ structure types C1 to C6 and the helical model with the proton located at the C-terminus. All energies are given relative to C1. The dashed lines serve as a guide to the eye and the PBE+vdW reference data from the original search are highlighted with a grey background.

vdW correction for MBD* (PBE+MBD*), we see that C2 emerges as the most stable conformer (instead of C1 for PBE+vdW). The change in energy with respect to C1 is about 175 meV. A similar effect is also seen for the pure helix. When going from PBE+vdW to PBE0+vdW we see again that C2 is stabilized, being only 5 meV higher in energy than C1. The pure helix is stabilized as well. Moving then to PBE0+MBD* the stabilization mechanisms for C2 by PBE0 and by MBD* act together so that C2 emerges as the most stable conformer and, furthermore, becomes significantly separated from the second-lowest conformer C4 by 120 meV. Notably, the conformer C3, which was particularly stable in the AmoebaPro13 force field, is not predicted to be dominant by any of the functionals tested, where all of the functionals should be more accurate than the AmoebaPro13 force field.

In a next step, we also included rotational and vibrational free-energy contributions at 300 K (rigid rotor-harmonic oscillator approximation, see Section 5.1.1). For this, we did not recalculate the normal mode frequencies at the PBE0 or the MBD* levels due to the computational cost involved,[2] but employed the results from PBE+vdW. The free-energy hierarchies obtained with the different functionals are illustrated in Fig. 8.15. The picture stays relatively similar, with C2 being now even more separated from all other conformers in PBE0+MBD*, namely by about 160 meV. While PBE+vdW would rather predict a conformational ensemble with several structure types co-existing, PBE0+MBD* points to C2 as the single dominant conformer. We have to keep in mind that energy differences of about 200 meV with different functionals are very small considering the size of the system (220 atoms). In fact, an accuracy of less than 1 meV/atom is a challenge to any applicable approximate electronic-structure method. However, although small compared to the error bars of the method, 200 meV is indeed thermodynamically relevant,

[2] As mentioned in Section 3.5.3.2, the forces for the MBD* correction are only available in a finite-difference approach at present.

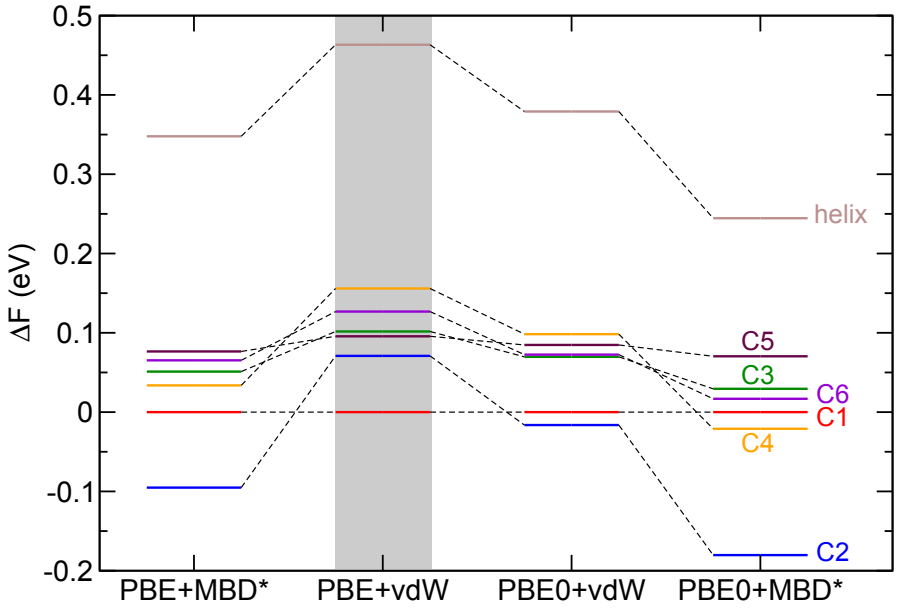


Figure 8.15: Free-energy hierarchies (horizontal bars) obtained with the PBE+MBD*, PBE+vdW, PBE0+vdW, and PBE+MBD* functional for the Ac-Lys-Ala$_{19}$ + H$^+$ structure types C1 to C6 and the helical model with the proton located at the C-terminus. The free energies are calculated based on the harmonic oscillator-rigid rotor approximation and a temperature of 300 K. Both the vibrational and the rotational contributions to the free energy are obtained from the PBE+vdW results and not recomputed at the other levels of theory. All energies are given relative to C1. The dashed lines serve as a guide to the eye and the PBE+vdW reference data from the original search are highlighted with a grey background.

demonstrating again the (conformational) challenge that we here face. In the following chapter, we will critically assess the structural predictions by comparing to experimental fingerprints.

For all of the tested functionals, the helix has an energy that would not point to a measurable helix population in experiment. Still, a small amount of helices is observed in the IM-MS measurements (cf. Fig. 7.3). This discrepancy could arise for different reasons. First of all, we might not have identified the lowest-energy helix in our *ab initio* REMD based structure search. As discussed in Section 8.4.2, the helix looks reasonable, but it is possible that there could be a helical structure that is even lower in energy. As we saw in Fig. 8.6, even small structural rearrangements can already lead to changes of the energy of the order of 100 meV. Another possibility is an error in the functionals. However, we consider this to be rather remote given the large predicted energy differences. Finally, the reason could be due to the experiment itself. As suggested by Jarrold[27], it is likely that the helices originate from dissociation of dimers. If the energetic barrier is high, they might be trapped in this local minimum.

8.7 Summary

In this chapter, we presented a conformational search for the two peptides Ac-Lys-Ala$_{19}$ + H$^+$ and Ac-Ala$_{19}$-Lys + H$^+$. For this, we relied on a combined force field-DFT approach, zooming into the relevant structure space with increasing accuracy. We first created a huge pool of force-field conformers based on REMD simulations with total simulation times of 8 µs. In order to identify different structure types, we then clustered snapshots of the trajectories according to their RMSD. Subsequently, thousands of structures were optimized with DFT (PBE+vdW).

For the four most important structure types of Ac-Lys-Ala$_{19}$ + H$^+$, we followed up with REMD simulations based on DFT with the PBE+vdW functional. Despite the limited simulation time lengths feasible for peptides of this size (220-440 atoms), we showed that such short ($\approx$ 20-30 ps) REMD simulations can lead to a refinement of the structure and the energy. While the structures stayed overall very similar during the simulation, we observed rearrangements of the hydrogen-bonding network, especially close to the termini.

Within the lowest 170 meV (PBE+vdW) we could identify six different structure types for Ac-Lys-Ala$_{19}$ + H$^+$, labelled as C1 to C6. We tested the effect of the PBE and PBE0 functional coupled with two different vdW corrections schemes on the energy hierarchy. Both the PBE0 functional and the many-body dispersion correction MBD* lead to a stabilization of C2 with respect to C1. When considering rotational and vibrational free-energy contributions at 300 K, C2 is separated by 160 meV from all other conformers. On the other hand, PBE+vdW rather predicts a conformational ensemble with different structures co-existing. The influence of the AmoebaPro13 force field was also tested, which predicts C3 to be the most stable conformer. In contrast to that, none of the (in principle more accurate) DFT functionals finds C3 to be particularly stable. In the following chapter, we will assess the structure predictions by comparing to experimental fingerprints.

9 Connecting to experiment

In this chapter, we will connect the first-principles structure predictions for Ac-Lys-Ala_{19} + H^+ discussed in the previous chapter to experimental fingerprints. For this, we have both collision cross sections (CCSs) obtained from ion mobility-mass spectrometry (IM-MS) measurements and infrared multiphoton dissociation (IRMPD) spectra available. The measurements were performed by Stephan Warnke, Kevin Pagel, Peter Kupser, and Gert von Helden working in the Molecular Physics Department of the Fritz Haber Institute.

9.1 Ion-mobility mass-spectrometry

As discussed in Section 7.1.1, the arrival times measured in the IM-MS experiments can be converted to CCSs. Those can then be compared to CCSs calculated for the predicted structures. As described in Section 7.1.1.1, there are different methods to calculate CCSs, namely the projection approximation (PA)[364], the trajectory method (TJM)[366, 367], and exact hard-sphere scattering (EHSS)[369]. We calculated CCSs for the lowest-energy Ac-Lys-Ala_{19} + H^+ compact monomers C1 to C6 and the helical models of monomers and dimers (see previous chapter) with all three methods. The results are listed in Tab. 9.1 together with the experimental CCSs.

All methods yield qualitatively the same trend, i.e., the dimers have the largest CCSs, and the compact monomers have the lowest CCSs. In detail, we find that PA yields smaller CCSs than TJM, while EHSS yields larger ones. The main reason for this is that EHSS overestimates multiple scattering events of the molecule with the buffer gas atom, which slows the molecule down more than a single scattering event. On the other hand, PA does not account for multiple scattering events[365–368]. For smaller molecules this effect becomes less important and the numbers obtained with PA and TJM should thus become more similar. This is also what we observe for the smaller peptides (90 – 108 atoms) considered in Part III/Chapter 12 of this thesis. The description used in TJM is the most accurate one and is thus generally understood to be most reliable[26, 365]. This is why we will focus on these values for the comparison to experiment. C1 and C3, the compact conformers with the lowest helical content (cf. Tab. 8.2 and Fig. 8.8), yield very similar CCSs that are 16 and 17 $Å^2$ lower than the experimental one. The compact structure types with the largest helical content C2, C4, C5, and C6 (cf. Tab. 8.2 and Fig. 8.8) have also very similar CCSs. They perfectly match the experimental value with deviations of less than 1 %.

In the IM-MS experiments, the width of the peak assigned to the compact monomers is very narrow (cf. Fig. 7.4). In fact, as can be seen in the same figure, the width is even a bit smaller than the width of the peak for Ac-Lys-Ala_{19} + Na^+, which is expected to reflect the presence of only one (helical) conformer. Such a narrow peak width for Ac-Lys-Ala_{19} + H^+ can arise

Table 9.1: Collision cross sections (CCSs) for Ac-Lys-Ala_{19} + H^+ helical dimers (cf. Fig. 8.10), compact monomers C1 to C6 (cf. Fig. 8.8), and helical monomer models (H^+ located close to the C-terminus or at the N-terminal lysine, cf. Fig. 8.12). We calculated the CCSs with the projection approximation (PA)[364], the trajectory method (TJM)[366, 367], and exact hard sphere scattering (EHSS)[369]. All methods were discussed in Section 7.1.1.1. For PA we used a standard deviation of 0.2 % and for EHSS 0.5 %. For TJM we employed the Hirshfeld[237] charges of the PBE density and 500,000 trajectories per structure, which resulted in standard deviations of less than 1 %. The experimental values were measured by Stephan Warnke, Kevin Pagel, and Gert von Helden working in the Molecular Physics Department of the Fritz Haber Institute. All CCSs are given in units of $Å^2$.

	Helical dimers		Compact monomers						Helical models	
	D1	D2	C1	C2	C3	C4	C5	C6	H^+ near C-term.	H^+ at Lys
PA	528	520	299	315	295	316	312	314	367	371
TJM	571	561	308	325	307	326	323	326	367	373
EHSS	593	587	318	336	323	340	337	343	392	396
Exp.	—(a)		324						371	

(a) intensity too low for a reliable CCS determination

for different reasons. One scenario is the presence of only one conformer or an ensemble of co-existing conformers, which have essentially the same CCS. The third scenario is that many structures with different CCSs rapidly interconvert on the time scale of the measurement so that only a single peak corresponding to the average CCS is observed (see, e.g., Ref. [413]). As we cannot determine barriers, we cannot assess the second possibility here. However, such a scenario could potentially match all of the different predictions by the different methods tested in Section 8.6. Apart from that, we can state the following facts and conclusions. C1 and C3 (group A) have essentially the same CCSs and the same applies to C2, C4, C5, and C6 (group B). If both a structure of group A and a structure of group B were co-existing (and not rapidly interconverting) in the experimental beam to a measurable extent, one would observe two peaks in the arrival time distribution (ATD).[1] As detailed above, the comparison of measured and calculated CCSs points to group B, i.e., C2, C4, C5, and C6. However, we cannot tell from the experiment if all structure types of group B co-exist or if there is one dominant conformation.

We can now turn to the structure predictions of the PBE+vdW, PBE+MBD*, PBE0+vdW, and PBE0+MBD* functionals and the AmoebaPro13 force field discussed in Section 8.6. AmoebaPro13 predicts C3 and C4 to be the two dominant conformers, with very similar energies. A co-existence of both conformers would yield two peaks in the experiment, which is in disagreement with the actual observation (one peak). PBE+vdW predicts C1 to be the most probable conformer, while the experiment rather points to C2, C4, C5, or C6. Additionally, PBE+vdW also predicts other conformers to contribute, e.g., C2, which would yield two peaks in the experiment (there is only one though). Based on similar considerations, neither the PBE+MBD* nor the PBE0+vdW predictions match the experiment. PBE0+MBD* predicts C2 to be basically the only conformer that should be present at 300 K. This scenario would be in agreement with the IM-MS measurements.

[1] This was also confirmed by Kevin Pagel (Private Communication, 2014).

9.2 IR SPECTROSCOPY

While the CCS gives an overall measure of the shape of the peptide, the infrared (IR) spectrum would be expected to be more sensitive to the actual conformation (as discussed in Section 5.2). For this reason, we now turn to a comparison of our structure predictions to fingerprints from IRMPD spectroscopy. As a quantitative measure of agreement between the spectra, we again employ the Pendry reliability factor[337], which we have introduced in Chapter 6. We will also refer to it as the R-factor or R_{P} in the following. As the Pendry reliability factor is sensitive to small kinks or wiggles in the spectra, the experimental raw data has to be smoothed before being compared to the theoretical spectra. In order not to oversmooth the spectra, we first splined the raw data on a grid with a spacing of $2\,\mathrm{cm}^{-1}$. Afterwards the spectra were smoothed twice using a three-point formula

$$\tilde{y}_n = \frac{y_{n-1} + 2y_n + y_{n+1}}{4} \quad . \tag{9.1}$$

Subsequently, the spectra were splined on a fine numerical grid with a spacing of $0.5\,\mathrm{cm}^{-1}$ to perform the R-factor calculations. The raw data compared to the smoothed data for the measurements of both Ac-Lys-Ala$_{19}$ + H$^+$ and Ac-Ala$_{19}$-Lys + H$^+$ are displayed in Fig. 9.1. Most importantly, the figure shows that no features were lost during the smoothing process. As discussed in Chapter 6, the R-factor is calculated including a rigid shift Δ_x of the theoretical spectrum along the wavenumber axis. This shift most probably reflects systematic mode softening due to the exchange-correlation (XC) functional and the neglect of nuclear quantum effects. Additionally, we here also allow for a shift Δ_y along the normalized intensity axis to account for possible offsets in experiment. We here note that for the experiment of Ac-Lys-Ala$_{19}$ + H$^+$ we have two data sets available, which look very similar although they originate from completely different measurement cycles. We discuss the spectrum with the better resolution here in the main text, while all details for the second spectrum can be found in Appendix A. The comparisons with the theoretical spectra yield essentially the same results for both experimental spectra.

In a first step, we calculated IR spectra in the harmonic approximation. For this, we considered the Ac-Lys-Ala$_{19}$ + H$^+$ helical dimers D1 and D2 (cf. Fig. 8.10), the helical models with both the proton located at the N-terminal lysine or close to the C-terminus (cf. Fig. 8.12), and the compact monomers C1 to C6. For comparison, we also calculated the harmonic IR spectrum for the lowest-energy α-helical conformation of Ac-Ala$_{19}$-Lys + H$^+$. The spectra are illustrated in Fig. 9.2. We both show the wavenumber region between 1100 and $1750\,\mathrm{cm}^{-1}$ and the region between 2000 and $4000\,\mathrm{cm}^{-1}$. The dotted lines indicate the positions of the amide I and II peaks for the α-helical Ac-Ala$_{19}$-Lys + H$^+$. As discussed in Section 7.3 (cf. Fig. 7.4), in the experimental spectra the amide II peak of Ac-Lys-Ala$_{19}$ + H$^+$ is red-shifted by about $10\,\mathrm{cm}^{-1}$ compared to the spectrum for Ac-Ala$_{19}$-Lys + H$^+$. In the comparison of the calculated harmonic spectra of the Ac-Lys-Ala$_{19}$ + H$^+$ conformers with the helical Ac-Ala$_{19}$-Lys + H$^+$, we see that the amide II peak of Ac-Ala$_{19}$-Lys + H$^+$ agrees with the peak position of the amide II band for both the helical dimers and the helical monomer with the proton close to the C-terminus. In contrast, for all compact monomers C1 to C6 of Ac-Lys-Ala$_{19}$ + H$^+$, the amide II peak is red-shifted compared to Ac-Ala$_{19}$-Lys + H$^+$. This agrees with the IM-MS observation (cf. Fig. 7.4) that predominantly compact monomers are present in experiment.

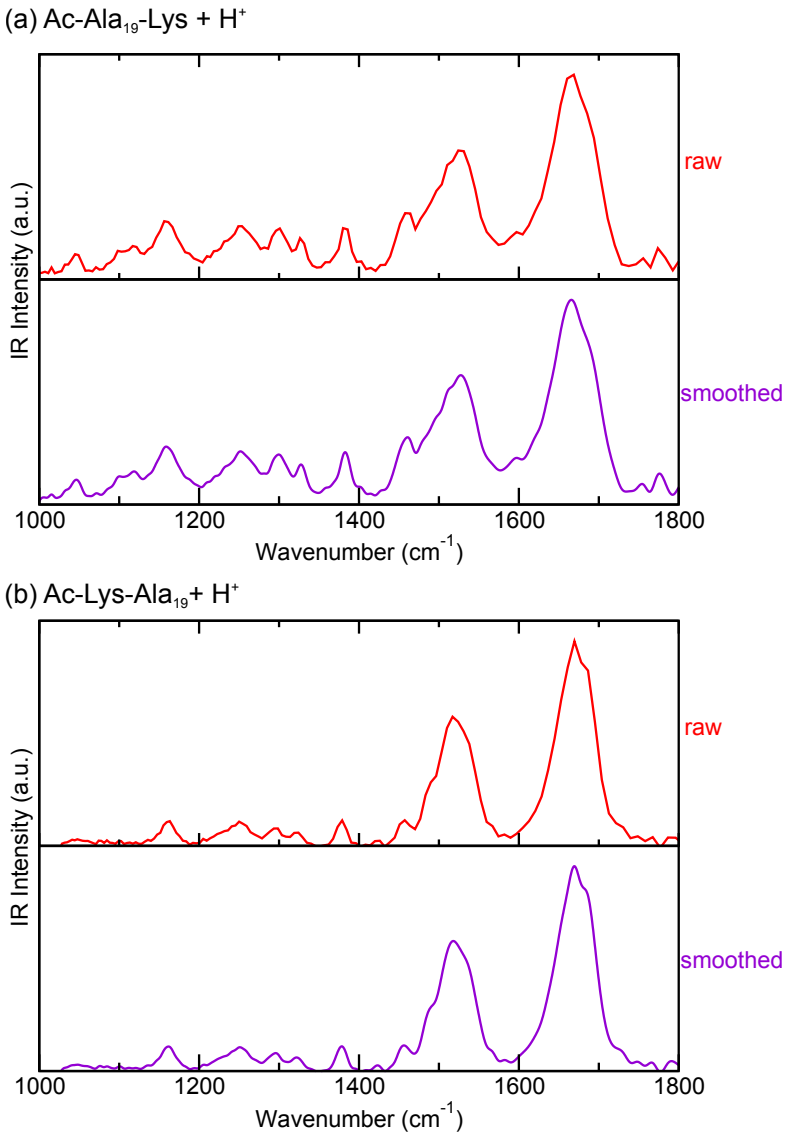


Figure 9.1: Comparison of raw and smoothed experimental IRMPD spectra for (a) Ac-Ala_{19}-Lys + H^+ and (b) Ac-Lys-Ala_{19} + H^+. The spectra were measured by Stephan Warnke, Kevin Pagel, Peter Kupser, and Gert von Helden working in the Molecular Physics Department of the Fritz Haber Institute.

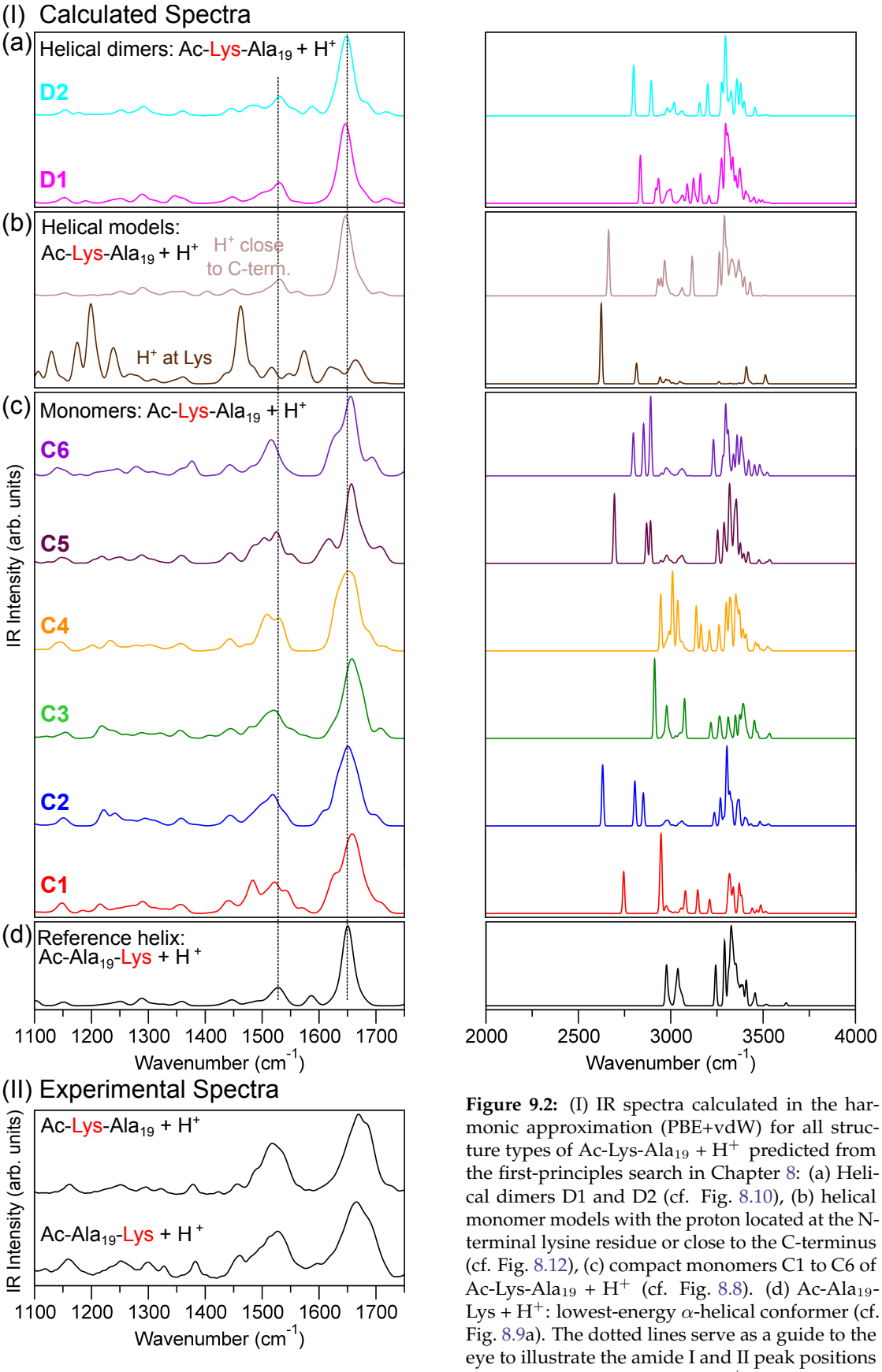


Figure 9.2: (I) IR spectra calculated in the harmonic approximation (PBE+vdW) for all structure types of Ac-Lys-Ala$_{19}$ + H$^+$ predicted from the first-principles search in Chapter 8: (a) Helical dimers D1 and D2 (cf. Fig. 8.10), (b) helical monomer models with the proton located at the N-terminal lysine residue or close to the C-terminus (cf. Fig. 8.12), (c) compact monomers C1 to C6 of Ac-Lys-Ala$_{19}$ + H$^+$ (cf. Fig. 8.8). (d) Ac-Ala$_{19}$-Lys + H$^+$: lowest-energy α-helical conformer (cf. Fig. 8.9a). The dotted lines serve as a guide to the eye to illustrate the amide I and II peak positions for the helical Ac-Ala$_{19}$-Lys + H$^+$. All spectra are convoluted with a Gaussian function with a variable width of $\sigma = 0.5\,\%$ of the wavenumber in the region between 1100 and 1750 cm^{-1} and with a constant convolution of $\sigma = 5$ cm^{-1} between 2000 and 4000 cm^{-1}. The spectra are not shifted nor scaled. The intensity is normalized to the highest peak in the respective wavenumber region. (II) Experimental spectra.

Table 9.2: Pendry reliability factors (R_P) calculated between the harmonic IR spectra (PBE+vdW) of the different structure types of Ac-Lys-Ala$_{19}$ + H$^+$ and the experimental spectrum measured for Ac-Lys-Ala$_{19}$ + H$^+$ in the wavenumber region between 1130 and 1736 cm^{-1}. In addition, the rigid shifts Δ_x along the wavenumber axis and Δ_y along the normalized intensity axis are given. The same is shown for the lowest-energy (PBE+vdW) α-helical conformer of Ac-Ala$_{19}$-Lys + H$^+$ and the experimental spectrum of Ac-Ala$_{19}$-Lys + H$^+$. For the R-factor calculation the harmonic stick spectra were convoluted with a Gaussian function with a variable width of $\sigma = 0.5\%$ of the wavenumber.

Conformation	R_P	Δ_x (cm^{-1})	Δ_y
C1	0.47	13.0	0.045
C2	0.32	11.5	0.025
C3	0.39	10.5	0.025
C4	0.28	16.0	0.015
C5	0.37	12.0	0.010
C6	0.38	13.0	0.025
D1	0.48	10.5	0.020
D2	0.52	9.0	0.000
Ac-Lys-Ala$_{19}$ + H$^+$, helix, (H$^+$ near C-term.)	0.58	8.0	0.005
Ac-Lys-Ala$_{19}$ + H$^+$, helix, (H$^+$ at N-term. Lys)	0.76	21.0	0.000
Ac-Ala$_{19}$ + H$^+$, helix	0.42	12.0	0.010

Except for the spectrum of the helical monomer with the proton located at the N-terminal lysine, the spectra for the different conformations of Ac-Lys-Ala$_{19}$ + H$^+$ look overall relatively similar at a first glance. However, upon closer inspection there are considerable differences reflected by shifted or split peaks. This is also confirmed by the R-factors calculated between the theoretical and the experimental spectra, which are shown in Tab. 9.2. As Ac-Ala$_{19}$-Lys + H$^+$ is without doubt α-helical, we take its R-factor of 0.42 between experiment and theory as a reference. All helical models of Ac-Lys-Ala$_{19}$ + H$^+$ have R-factors that are higher than 0.42. The compact monomers generally show a better agreement than the helical dimers and helical monomers (however, this changes for the anharmonic spectra as discussed later in the text). Except for C1, all R-factors are lower than 0.42. C4 and C2 give the best agreement with R-factors of 0.32 and 0.28, respectively.

As expected (and discussed in Section 5.2) the IR spectra are even more structure sensitive in the wavenumber region between 2000 and 4000 cm^{-1}, which is also depicted in Fig. 9.2. In this regime, very localized hydrogen stretching modes are probed, making this region very conformer sensitive. However, no experimental data are available in this range to compare to.

The spectra discussed so far were calculated based on the harmonic approximation. In the next step, we follow up with the computation of IR spectra derived from molecular dynamics (MD) simulations with $\langle T \rangle = 300$ K. Benchmarks presented in Chapter 6 showed that a long run of 25 ps length should yield reliable results. Due to the high computational cost,[2] we only chose the most important structure types to calculate the anharmonic IR spectra, namely the compact monomers C1 to C4 of Ac-Lys-Ala$_{19}$ + H$^+$ and the helical model with the proton located close to the C-terminus.

As before, we convoluted the spectra with a Gaussian function. For this, we found a variable width of $\sigma = 0.5\,\%$ of the wavenumber to reflect the broadening present in experiment best. Of

[2]Such a simulation takes approximately 57 days on 384 cores of the thnec cluster (hexa-core Intel Westmere X5650 processors).

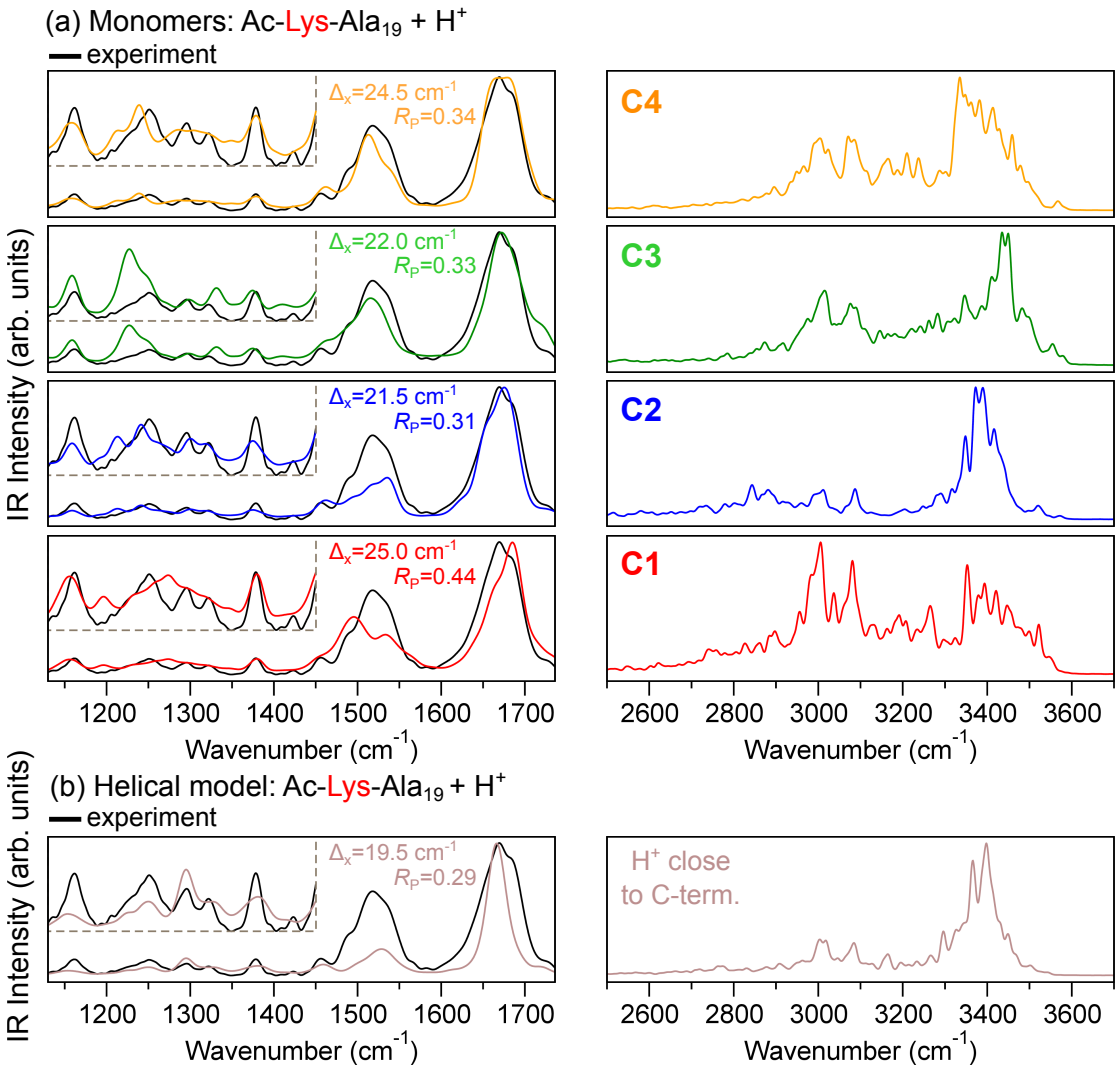


Figure 9.3: IR spectra derived from PBE+vdW MD simulations (in the *NVE* ensemble with $\langle T \rangle = 300\,\mathrm{K}$, time step: 1 fs) for the Ac-Lys-Ala$_{19}$ + H$^+$ conformers C1 to C4 (a) and the helical model with the proton located at the C-terminus (b). All spectra are convoluted with a Gaussian function with a variable width of $\sigma = 0.5\,\%$ of the wavenumber in the region between 1130 and 1736 cm^{-1} and with a constant convolution of $\sigma = 5\,\mathrm{cm}^{-1}$ between 2500 and 3700 cm^{-1}. The intensity is normalized to the highest peak in the respective wavenumber region. For the wavenumber region between 1130 and 1736 cm^{-1} the Pendry reliability factor (R_P) is given with the corresponding rigid shift Δ_x. The theoretical spectra are shifted accordingly. The black lines denote the experimental data. The insets show the wavenumber region between 1130 and 1450 cm^{-1}.

course, the exact convolution is to a certain extent a matter of taste, but different convolutions would not change the general picture. The anharmonic IR spectra are compared to the experimental spectrum in the wavenumber region between 1130 and 1736 cm^{-1} at the left side of Fig. 9.3. In the plots, the corresponding Pendry reliability factors are given and the theoretical spectra are shifted rigidly by Δ_x, which is given in the plots as well. The shifts Δ_y are not shown in the figure, but detailed in Appendix A.5. As with the harmonic IR spectra, the anharmonic spectra feature a considerable variation of the peak positions and peak shapes. In agreement with the results for the harmonic IR spectra, C1 gives a rather poor match with the experimental spectrum compared to the other conformers. In contrast, the conformer with the best match (based on the R-factor) is the helical model with the proton located at the C-terminus ($R_\mathrm{P} = 0.29$). C2, C3, and C4 yield similarly good R-factors ($R_\mathrm{P} = 0.31$, 0.33, and 0.34, respectively). However, the peak width of the experimental amide I band is matched much better by the compact monomers than by the helical model.

The wavenumber region between 2500 and 3700 cm^{-1} is plotted on the right side of Fig. 9.3. As mentioned above, for this region there are no experimental data available. As already seen for the harmonic spectra (cf. Fig. 9.2), this region is more structure sensitive. The predicted IR intensity for the helical monomer is rather concentrated around 3400 cm^{-1}, while for C1, C3, and C4 it is more spread out. The spectrum of C2, on the other hand, is very similar to the helical monomer. This probably results from the high helical content of C2 (cf. Tab. 8.2 and Fig. 8.8).

Based on this, we can conclude as follows. In the wavenumber region between about 1000 to 2000 cm^{-1}, IR spectroscopy cannot safely distinguish between helical and compact conformations. This might also be the reason for the similarity of the experimental IRMPD spectra for Ac-Ala_{19}-Lys + H^+ and Ac-Lys-Ala_{19} + H^+ discussed in Section 7.3 (cf. Fig. 7.4). However, there is some structure sensitivity in the IR spectra. Most likely, e.g., C1 is not present (or only to a small extent) in experiment. This is in agreement with the IM-MS results. For C3, on the other hand, the situation is not that clear. While the IM-MS measurements suggest that C3 is not present in experiment, the agreement of the calculated IR spectrum with the experimental one is rather good. However, this does not mean that C3 has to be present in experiment – other conformers show a similarly good agreement of their IR spectrum with experiment and IR spectroscopy might not be structure sensitive enough to reveal the differences.

We can now again turn to the structure predictions by the different methods tested in Section 8.6. PBE+vdW would predict C1 as the dominant conformer, which is in disagreement with the rather poor similarity of the anharmonic IR spectrum with the experimental one. PBE0+vdW and PBE+MBD* both favor C2, but only for PBE0+MBD* C2 should be the only dominant conformer. This would be in accord with the good agreement of the calculated IR spectrum for C2 with experiment. AmoebaPro13 predicts C3 and C4 to be almost equally stable, where both calculated spectra yield a good agreement with experiment.

9.3 Summary

In this chapter, we connected the first-principles theoretical predictions for Ac-Lys-Ala_{19} + H^+ to experimental fingerprints. We compared CCSs determined using IM-MS measurements to calculated CCSs. The computation of the CCSs employs empirical potentials, while the geometries of the structure types were determined based on PBE+vdW. From the comparison of

the experimental and theoretical CCSs we can infer that the least helical conformers C1 and C3 should not be the most probable structure candidates. On the other hand, the calculated CCSs for C2, C4, C5, and C6 perfectly match the experimental values.

Using the Pendry reliability factor, we compared IR spectra derived from MD simulations with the experimental IR spectrum for Ac-Lys-Ala_{19} + H^+ in the wavenumber region between 1130 and 1736 cm^{-1}. The spectra of C2, C3, C4 and the helical model yield a similarly good agreement with experiment. However, the helical model should hardly be populated in experiment according to the IM-MS data. We thus find that IR spectroscopy cannot reliably differentiate between fully helical and more compact conformers in the region between 1000 and 2000 cm^{-1}. The spectral similarity of the experimental IRMPD spectrum of the helical Ac-Ala_{19}-Lys + H^+ and Ac-Lys-Ala_{19} + H^+ may thus also be attributed to a lack of sufficient structure sensitivity. However, there is a level of resolution where we can draw conclusions, namely that C1 should not be present, as the agreement of the experimental and the calculated IR spectra for C1 is rather poor. This is also in agreement with the CCS comparisons.

The wavenumber region between 2500 and 3700 cm^{-1} proves to be more structure sensitive. However, there are no experimental data available to compare to.

We critically assessed the structural predictions by the AmoebaPro13 force field and the PBE+vdW, PBE+MBD*, PBE0+vdW, and PBE0+MBD* functionals by comparing to the experimental fingerprints. With respect to the IM-MS measurements, all structure predictions could potentially agree with a scenario of rapidly interconverting conformers, which would lead to only one peak in the experimental ATD corresponding to the average CCS. In a scenario of a co-existing conformers according to their relative free energies, only the PBE0+MBD* functional yields a prediction that matches the IM-MS data. PBE0+MBD* predicts C2 to be the only dominant conformer at room temperature (300 K), which also matches the IRMPD data. In summary, both experiments and theory (PBE0+MBD*) could thus agree on C2 as "the" outstanding conformer. PBE0+MBD* is also the method that yielded the best results for Ac-Phe-Ala_5-Lys(H^+) in a recent benchmark by Rossi *et al.*[247].

Part III

Dealing with conformational flexibility: homologous peptides

10 First-principles structure predictions for Ac-β^2hAla$_6$-Lys(H$^+$)

After having addressed the conformational challenge of a large natural 20-residue peptide in Part II of this thesis, in Part III we concentrate on non-natural β-peptides with an artificially increased structure space. From a biochemical point of view, β-peptides are interesting as they could be used as artificial modulators of protein function (see Section 2.4 for a detailed discussion). As discussed in Chapter 2, compared to the natural α-amino acids, β-amino acids have one additional methylene (CH_2) group in the backbone (see Fig. 10.1a for a direct comparison). This backbone extension of β-peptides leads to one additional torsional degree of freedom per residue (see Section 2.4 and Fig. 2.14) yielding in turn an even more complex conformational space.

In the following chapters, we investigate the influence of such an increased backbone flexibility on the structure space by comparing a β-peptide to its related α-peptidic sequence. As discussed extensively in previous chapters, the peptide series Ac-Ala$_n$-Lys(H$^+$) with $n \simeq$ 6-19, forms α-helices in the gas phase[17, 25–28, 338, 347]. This is due to a favorable interaction of the charge at the lysine residue with the helix dipole and a hydrogen bond capping of the dangling C-terminal backbone carbonyl oxygens by the lysine NH_3^+ group. Here we transfer this design principle to β-peptides in order to examine whether still helical structures are enforced despite the increased conformational flexibility. For this, we exchange the alanine amino acid for its equivalent β-amino acid, β^2hAla, which stands for (R)-β-aminoisobutyric acid.[1] Specifically, we concentrate on Ac-β^2hAla$_6$-Lys(H$^+$). We address the challenge of the artificially increased conformational space by performing two independent first-principles structure search strategies with particular regard not to overlook any helical conformers.

In Chapter 11, we then contrast the structure spaces and conformational preferences of Ac-β^2hAla$_6$-Lys(H$^+$) and its related α-peptide Ac-Ala$_6$-Lys(H$^+$). To critically assess how far we can push the PBE+vdW level of theory for increasing flexibility, we analyze the impact of different exchange-correlation functionals and compare the structural predictions to experimental "fingerprints" from infrared multiphoton dissociation (IRMPD) and ion mobility-mass spectrometry (IM-MS) measurements in Chapter 12. These were performed by Stephan Warnke, Gert von Helden, and Kevin Pagel working at the Molecular Physics Department of the Fritz Haber Institute, Berlin. The energy hierarchies and conformers for the natural peptide Ac-Ala$_6$-Lys(H$^+$)

[1] As discussed in Section 2.4, the nomenclature β^2 indicates that the methyl group is substituted at the C_β atom [see Fig. 10.1a)].

are based on previous work performed in our group by Mariana Rossi[17, 28]. However, all additional results based on the latter [collision cross sections (CCSs) or infrared (IR) spectra] are the work of this thesis.

10.1 Equivalent H-bonding patterns in helices of α- and β-peptides

As discussed in Section 2.4, helix types with hydrogen bonds (or H-bonds) pointing along the sequence direction, in opposite direction or alternating direction ("mixed" helices) have been found in β-peptides[18, 19, 69, 70, 72–74, 77, 79–81]. Due to the location of the charge at the lysine residue close to the C-terminus[26], our design should strongly favor helices with H-bonds, and thus a helix dipole, pointing in opposite sequence direction similar to helices in natural α-peptides. Helices of α- and β-peptides with equivalent H-bonding patterns are directly compared in Fig. 10.1c. Hydrogen bonds between the backbone carbonyl oxygen of residue i and the NH group of residue $i+5$ lead to a π-helix in α-peptides with 16 atoms in the hydrogen-bonded pseudocycles (cf. Fig. 10.1b). In a β-peptide the equivalent H-bonding pattern leads to hydrogen-bonded pseudocycles that contain 20 atoms. For this reason the corresponding helix is denoted as H20-helix. The H12-helix in β-peptides has the H-bonding pattern $i \leftarrow i+3$, resembling the 3_{10}-helix of α-peptides. The most prominent helix in α-peptides, the α-helix has the H-bonding pattern $i \leftarrow i+4$. The equivalent pattern in a β-peptide leads to the H16-helix. Surprisingly, despite many efforts[69, 71, 72, 77, 79, 90, 105, 437, 438] there has been no evidence for the existence of the H16-helix. There are only hints that stem from a Hartree-Fock study[74] and diffraction experiments on nylon-3 derivatives[62, 111].

The color code introduced in Fig. 10.1, namely green for π- or H20-helices, red for α- or H16-helices, and blue for 3_{10}- or H12-helices will be used throughout the rest of this thesis. Non-helical structures are color-coded in grey. For reasons of clarity, non-polar hydrogens are omitted in structural representations.

10.2 Assessing the conformational space of Ac-β^2hAla$_6$-Lys(H$^+$)

In order to sample the conformational space of Ac-β^2hAla$_6$-Lys(H$^+$) we again adopt a two-step strategy, where we first create a large conformational pool based on force field (OPLSAA) simulations and then carry out a refinement based on density-functional theory (DFT) using the PBE+vdW functional[15, 16]. In fact, we perform two individual searches of this kind. The first strategy is based on a series of basin-hopping searches, which will be addressed in Section 10.2.3, while the second strategy relies on replica-exchange molecular dynamics (REMD), and is covered in Section 10.2.5. In order to judge the performance of the two search strategies we subsequently compare the outcome of the two approaches. As mentioned above, we are especially interested in the potential of β-peptides to form helical structures. For this reason, we take particular care not to overlook helical structures in our conformational searches.

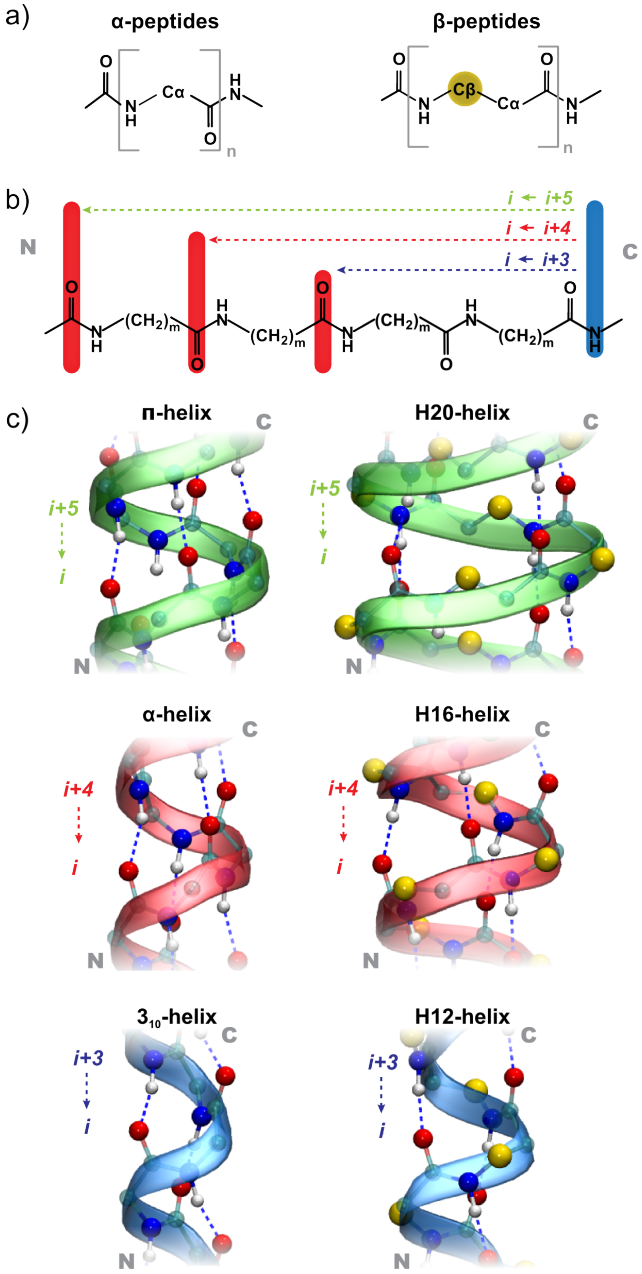


Figure 10.1: a) Schematic representation of the chemical formula of α- and β-peptides. They differ by one methylene group (CH_2) in the backbone per amino-acid residue. b) Scheme of possible hydrogen bond patterns for α-peptides ($m = 1$) and β-peptides ($m = 2$). c) Sketch of a π-helix ($i \leftarrow i + 5$) in α-peptides and the equivalent H20-helix ($i \leftarrow i + 5$) in β-peptides, an α-helix and a H16-helix ($i \leftarrow i + 4$), and a 3_{10}-helix and a H12-helix ($i \leftarrow i + 3$). Non-polar hydrogens are omitted for clarity. The extra carbon atom in β-peptides is highlighted in yellow. The color code, namely green for H20-helices, red for H16-helices, and blue for H12-helices is used throughout the rest of this thesis.

10.2.1 Details of conformational analysis

As in Part II of this thesis, DFT-based structure optimizations are always performed in two steps. First, the structures are relaxed with *light* computational settings and afterwards the lowest-energy structures are further relaxed with *tight* computational settings[257], where the forces were converged down to $5 \cdot 10^{-3}$ eV/Å. All energies that are explicitly reported or discussed are calculated with *tight* computational settings unless stated otherwise.

For the analysis of the conformers found in our structure searches, we sort the conformers into families according to their hydrogen-bond pattern. A hydrogen bond (H-bond) is defined to be present if the distance between the hydrogen and acceptor is less than 2.5 Å. From each family, the member with the lowest energy is selected as the representative of the family so that energy hierarchies or structure representations given for a specific family always refer to the corresponding representative member. Predominantly, we separate hydrogen-bond families into H12-helices, H16-helices, H20-helices or miscellaneous. Structural representations of these families are displayed with blue, red, green, or gray ribbons, according to the color code introduced earlier in this chapter. In order to decide to which group a family belongs, we take the corresponding hydrogen-bond pattern as a basis. However, for instance, an H-bond between a residue $i+5$ and i (H20-helical) can be involved in a helical twist, but may also be involved in a different kind of structure element such as a turn. Only multiple H-bonds of the same helical character formed in a row are a clear indicator of a corresponding helical structure. However, as our peptide of choice, Ac-β^2hAla$_6$-Lys(H$^+$), only contains six β^2hAla residues and one lysine residue, only a small number of helical hydrogen bonds are possible at all. This is illustrated in Fig. 10.2. When not taking into account the oxygen of the acetyl group, there are only two possibilities for H20-helical hydrogen bonds. These two possible hydrogen bonds are highlighted in pink in Fig. 10.2a. For a H16 helix there are three possible H-bonds (see Fig. 10.2b) and for a H12-helix there are four possible hydrogen bonds (see Fig. 10.2c) also highlighted in pink. In order to identify helical structure families, we picked all families that had at least one H20-helical bond or two H16- or H12-helical bonds, when not taking into account the acetyl group. For these helical candidates we performed a visual check in order to finally judge if they are helices or not. The possible involvement of the oxygen of the acetyl group in a hydrogen bond is not included in the above considerations as it is at the terminus of the peptide, which is the most flexible part.

10.2.2 Assessment of the OPLSAA force field

As a first step in both conformational search strategies we perform a global sampling of the structure space based on force-field simulations in order to generate input structures that are relaxed with DFT in the subsequent step. As in Part II of this thesis, we employed the OPLSAA force field[143]. However, the force field lacks explicit parameters for the additional CH_2 group present in the β-amino acid. To account for this, we here adopted the CH_2 parameters given for alkanes in the force field. In order to ensure that the force field with the additional parameters for the CH_2 group yields reasonable results, we performed a series of tests. For this, we used the β-peptide Ac-β^2hAla-NMe as a benchmark system, which is illustrated in Fig. 10.3. It contains two peptide bonds with the corresponding torsional angles ω_1 and ω_2. Between the two peptide bonds, there are three rotatable backbone bonds, with the corresponding dihedral angles labelled

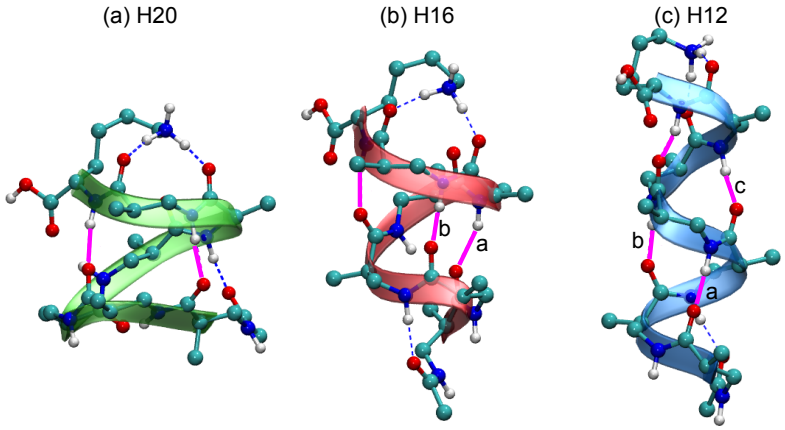


Figure 10.2: H20 (a), H16 (b), and H12 (c) helical conformations of Ac-β^2hAla$_6$-Lys(H$^+$) exhibiting all corresponding helical hydrogen bonds that are possible within this small peptide (highlighted in pink). Additionally, for the H16-helix and the H12-helix hydrogen bonds in the "middle" of the helix are labelled a and b, and a, b, and c, respectively.

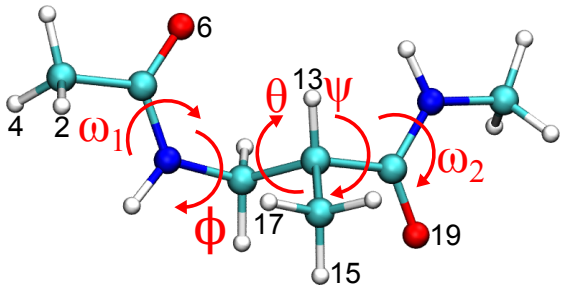


Figure 10.3: Structural representation of the Ac-β^2hAla-NMe conformer that was used as the initial point to create the dihedral-energy plots. The dihedral angles are marked accordingly. The numbers denote the atom number of the specified atom.

as ϕ, θ, and ψ, as it was explained in Chapter 2 and Fig. 2.14. Starting from the geometry depicted in Fig. 10.3, we changed the value of the torsional angles from $-180°$ to +180° in steps of 5° for one torsion at a time, leaving the other dihedral angles at their initial values. For each structure we calculated the single-point energies both using the OPLSAA force field and with PBE+vdW (*light* computational settings). Figure 10.4a shows the OPLSAA and PBE+vdW energies as a function of the dihedral angle ϕ. At $\phi = 30°$, atoms number 2 and 6 (cf. Fig. 10.3) come very close as illustrated in the upper panel of Fig. 10.4a. While the relative DFT energy rises to about 4.5 eV, the force-field energy has a maximum at about 1327 eV. In the dihedral-energy plot for θ (Fig. 10.4b) there are three maxima in the relative energy of the force field. Each of them coincides with steric clashes of two atoms as depicted in the upper panel of Fig. 10.4b. The DFT energy curve responds much more softly to these interferences than the force field energy. One reason for this is that in the OPLSAA force field, the partial charges are fixed and no charge redistribution, i.e., screening, is possible. The behavior of the force-field energy curve and the DFT energy curve for the dihedral energy plot of ψ is relatively similar (Fig. 10.4c), although again the DFT curve is much smoother.

In general, the dihedral-energy plots for the angles ϕ, θ, and ψ show the same trends. Although they differ in the details, they agree on the global minima. Due to the partial double-bond character, the peptide bond is planar and the corresponding torsional angle ω adopts either angles around $0°$ (cis) or around $-180/180°$ (trans) (see also Chapter 2). We thus expect the energy as a function of ω to show minima at $0°$ and $-180/180°$ and maxima in between. As illustrated in Fig. 10.5b, for ω_2 (cf. Fig. 10.3) we find exactly this behavior for both the force field and PBE+vdW. For ω_1 (cf. Fig 10.3) we find it for PBE+vdW, too, while the force-field energy curve exhibits two additional maxima at positions where again two atoms come very close (as depicted in the upper panel of Fig. 10.5a). As a test, we calculated the same dihedral-energy plot for ω_1 based on a different initial geometry of Ac-β^2hAla-NMe where no clashes between atoms occur upon the change of ω_1. As illustrated in Fig. 10.6, in this case we find the same smooth behavior of the OPLSAA curve as in Fig. 10.5b for ω_2.

All previous tests were performed for Ac-β^2hAla-NMe. As a next step, we performed a consistency check for Ac-β^2hAla$_6$-Lys(H$^+$). For this, we chose the lowest 100 OPLSAA structures (found in the basin-hopping search, which will be explained in the following section) and relaxed them with PBE+vdW (*light* settings). Figure 10.7 shows histogram plots of the relative frequency of occurrence of the different values for the dihedral angles ω, ϕ, θ, and ψ. The distributions for the force-field structures and the relaxed PBE+vdW structures are relatively similar.

In summary, although we find that the force-field and PBE+vdW results disagree in the details, they seem to give similar trends. With this caveat in mind, we will use the augmented force field to create a (large) pool of structure candidates for relaxation with DFT. As it is crucial to ensure that the most important conformers for Ac-β^2hAla$_6$-Lys(H$^+$) are identified by our conformational search, we will independently assess the two search strategies in the following and compare their outcome.

10.2.3 Search strategy 1: basin hopping

This section is devoted to carefully assess the capabilities and limitations of the first strategy we used to explore the conformational space of Ac-β^2hAla$_6$-Lys(H$^+$). As discussed earlier in this chapter, we adopt a two-step approach. In the first step, we perform a global sampling of the

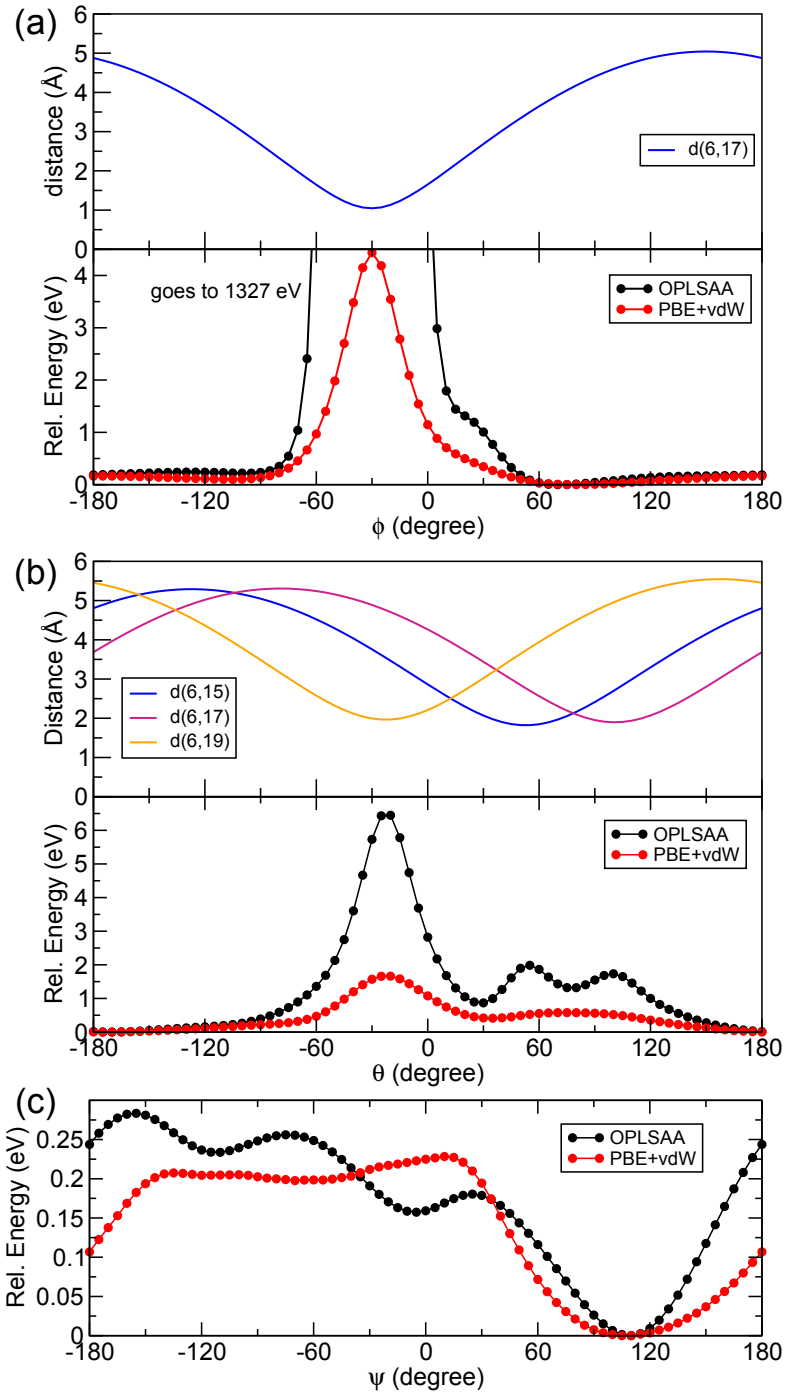


Figure 10.4: OPLSAA and PBE+vdW (*light* computational settings) energies as a function of the torsional angles (a) ϕ, (b) θ, and (c) ψ for the Ac-β^2hAla-NMe reference structure shown in Fig. 10.3. All energies are given with respect to the conformation with the lowest energy. The upper panels in (a) and (b) illustrate the distance d(n,m) between atoms n and m when varying the corresponding torsional angle. The atom numbers are specified in Fig. 10.3.

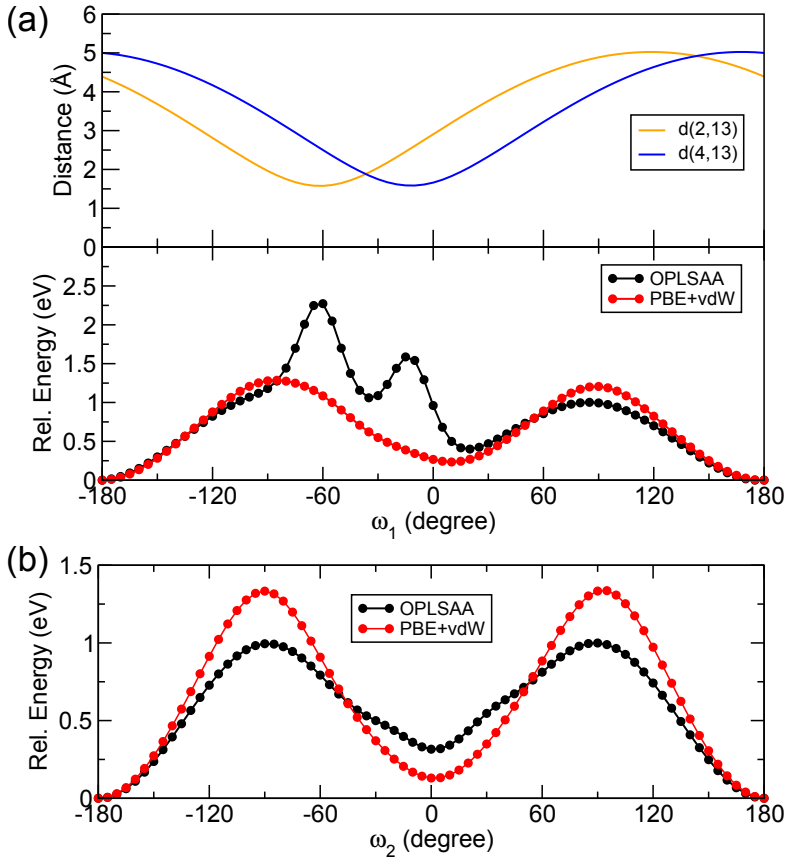


Figure 10.5: OPLSAA and PBE+vdW (*light* computational settings) energies as a function of the torsional angles (a) ω_1 and (b) ω_2 for the Ac-β^2hAla-NMe reference structure shown in Fig. 10.3. All energies are given with respect to the conformation with the lowest energy. The upper panel in (a) illustrates the distance d(n,m) between atoms n and m when varying ω_1. The atom numbers are specified in Fig. 10.3.

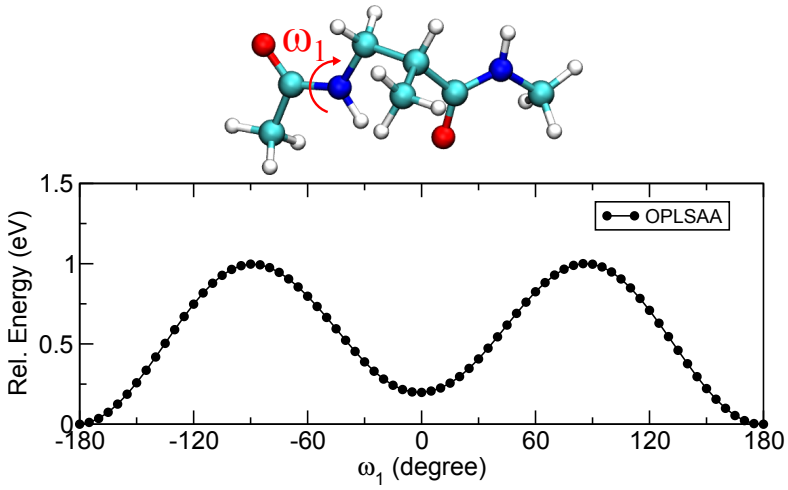


Figure 10.6: OPLSAA energies as a function of the torsional angle ω_1 for the Ac-β^2hAla-NMe reference structure that is shown in the upper part of the figure. All energies are given with respect to the conformation with the lowest energy.

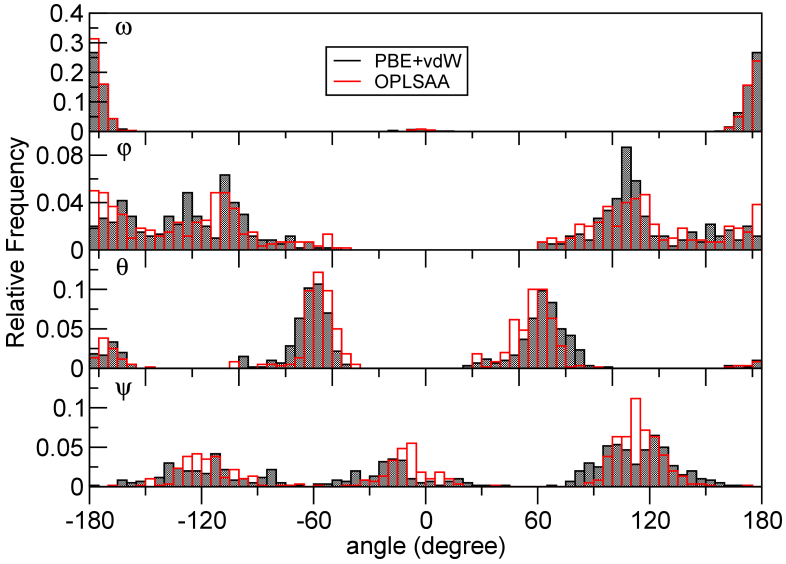


Figure 10.7: Normalized dihedral-angle distribution for the 100 lowest-energy OPLSAA conformers (red-bordered bars) resulting from the basin-hopping search (discussed in Section 10.2.3). The grey-shaded bars show the same for the conformers after being relaxed with PBE+vdW (*light* computational settings).

conformational space with a series of basin-hopping searches using the OPLSAA force field. In the second step, thousands of force-field conformers are relaxed with DFT using the PBE+vdW functional.

The basin-hopping technique employed here for Ac-β^2hAla$_6$-Lys(H$^+$) was already discussed in Section 4.2.1. Three parameters need to be specified:

- The number of torsional eigenvectors (search directions) N_{modes}, along which the system is perturbed in order to search for new minima,
- an energy threshold ϵ, which serves to decide if two structures are considered to be the same or not, and
- an energy cut-off value Δ above which all newly found minima are discarded.

For the latter we chose Δ =50 kcal/mol (2.2 eV), which should cover the relevant energy range. In order to test the convergence of the search with respect to the number of torsional eigenvectors taken into account, we subsequently increased the number of search directions with ϵ set to 10^{-4} kcal/mol. We used search directions between 2 and 35 and all runs were started from the same initial structure. This initial structure was the global minimum found in a preceding initial basin-hopping search attempt. All searches also find this initial structure to be the lowest-energy structure. However, the total number of conformers and also the number of conformers in the lowest-energy regime (2, 5, 8, and 10 kcal/mol) differ and converge only slowly with the number of search directions as illustrated in Tab. 10.1. The number of conformers found is huge ($\sim 10^6$). With 35 search directions we find about 15,000 conformers in the lowest 8 kcal/mol (0.35 eV) regime. For the α-peptide Ac-Ala$_6$-Lys(H$^+$) there were only about 1,000[17]. This increase in structures is due to the higher flexibility of the backbone of Ac-β^2hAla$_6$-Lys(H$^+$). We rank all conformers that were found according to their relative force-field energy, with the lowest-energy conformer having index 1, the second-lowest index 2 and so on. Figure 10.8

Table 10.1: Convergence behavior of the number of conformers found in basin-hopping searches with increasing numbers of search directions N_{modes}.

N_{modes}	N_{conf}	N_{conf} in lower			
		2 kcal/mol	5 kcal/mol	8 kcal/mol	10 kcal/mol
2	201851	15	222	1168	2690
5	272677	77	1092	5432	11449
10	298217	101	1807	9337	18627
12	304033	109	2015	10534	20510
15	308482	115	2156	11535	22172
20	313335	120	2408	12753	24163
25	316165	118	2491	13499	25352
30	317682	119	2536	13904	26004
35	318705	120	2593	14234	26543

shows the relative energy of all conformers found as a function of the ranking index for all different number of search directions used. The curves slowly converge for increasing number of search directions. Notably, the shape of the curves follow a "funnel plus tail" behavior, where the funnel incorporates approximately the lowest 8 kcal/mol energy regime as indicated in the figure by the pink line. The tail has a slope of approximately 10^{-4} kcal/mol, which is the energy-convergence threshold ϵ used to decide if two conformers are considered to be the same or not. This means that the tail regime is a sort of continuum region of conformers, where we find approximately one structure per ϵ interval. This implies in turn that there might be other structures that are overlooked by the search as they have the same (on the order of ϵ) force-field energy as a conformer found at an earlier stage. To analyze this, we attempted to decrease ϵ to 10^{-5} kcal/mol using 35 search directions. However, this run did not finish within a reasonable amount of time.[2] We stopped it after it had found 2,669,697 conformers and checked the lowest-energy regime. The lowest-energy conformer found is consistent with the searches performed with $\epsilon = 10^{-4}$ kcal/mol. In the lowest 2 kcal/mol regime 79 conformers were found, i.e., less than with the converged (35 search directions) search for $\epsilon = 10^{-4}$ kcal/mol. All of these 79 conformers were also found by the search with $\epsilon = 10^{-4}$ kcal/mol. This observation reassures us that decreasing the energy-convergence threshold ϵ would only find new conformers in the "tail" region and not in the funnel region, which is the one of more interest to us. A similar observation was made for basin-hopping searches for the α-peptide Ac-Ala$_4$-LysH$^+$[17].

10.2.3.1 Unconstrained basin-hopping search

After having assessed the performance and convergence of the basin-hopping algorithm for Ac-β^2hAla$_6$-Lys(H$^+$) in the previous subsection, we will now concentrate on the analysis of the conformers that were actually found. For this, we focus on the run based on 35 torsional modes. The outcome of the search is presented in Fig. 10.9 giving the OPLSAA energy of each conformer that was obtained as a function of the order in which it was found by the algorithm. All conformers are denoted by a black dot. As mentioned in the previous section, the lowest-energy structure found is the structure that was used to initialize the basin-hopping

[2] After five days of computation time on 128 cores of the aims cluster at the Garching Computing Centre, the search had found 2,669,697 conformers already. However, it had only performed the normal mode search for 1,097,303 of these implying that it was far away from finishing. This amount of conformers was already one order of magnitude larger than the number found in the corresponding search with $\epsilon = 10^{-4}$ kcal/mol (cf. Tab. 10.1).

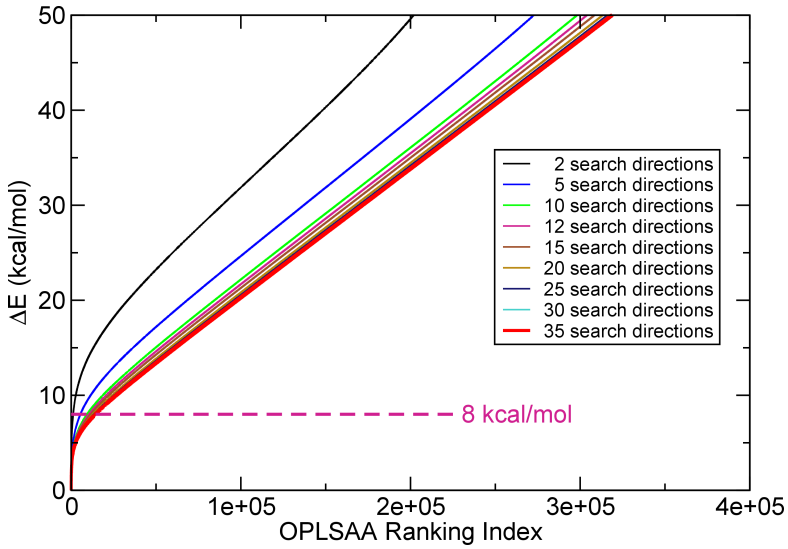


Figure 10.8: Relative OPLSAA energy of the conformers found for Ac-β^2hAla$_6$-Lys(H$^+$) in basin-hopping searches with different number of search directions as a function of the ranking index. The ranking index follows the energy hierarchy of the conformers, with the lowest-energy structure being assigned index 1, the second lowest structure being assigned index 2 and so on. The energies are given relative to the global minimum (which was the same for all search directions).

search. Nevertheless, close to the end of the search a similarly low energy basin is explored (see Fig. 10.9). The number of conformers found is so high and they lie so densely that they form partly a completely black area. This is in principle positive as we attempt to sample the conformational space as broadly as possible. However, it poses a challenge for the selection of the most relevant conformational types. In the second step of the search, we follow up with a relaxation of thousands of OPLSAA conformers with DFT. We aim to cover all conformers in the lowest 8 kcal/mol, as this regime forms the relevant "funnel" part of the OPLSAA hierarchy as was illustrated in Fig. 10.8. However, this part still consists of 14,234 conformers. In order to pick the most important structure representatives for relaxation with DFT, we sorted those structures into clusters according to their root mean square deviation (RMSD). For this, we used the GROMACS program package[433], the clustering algorithm suggested by Daura *et al.*[99] and a cut-off criterion of 0.05 nm. In total, we obtained 6,490 clusters out of the 14,234 structures from the lowest 8 kcal/mol regime. For all of these clusters we relaxed the lowest-energy force-field representative with PBE+vdW (*light* computational settings). While we most certainly do not miss any relevant structure type by employing this clustering algorithm, it can yet lead to a distortion of the energy hierarchy as the PBE+vdW relaxation of any other cluster member than the lowest-energy force field member might yield a lower-energy PBE+vdW structure. However, given the huge amount of conformers there is little way around employing the clustering algorithm at this point. Nevertheless, we have to keep this uncertainty factor in mind.

Additionally, we chose 500 conformers equally distributed among the remaining 304,471 conformers that were higher in force-field energy than 8 kcal/mol (the "tail" regime) and relaxed them with PBE+vdW (*light* computational settings). The reason for this was to compare the predictions of the force field and DFT and see if conformers that are predicted as being high in energy by the force field become relevant in DFT.

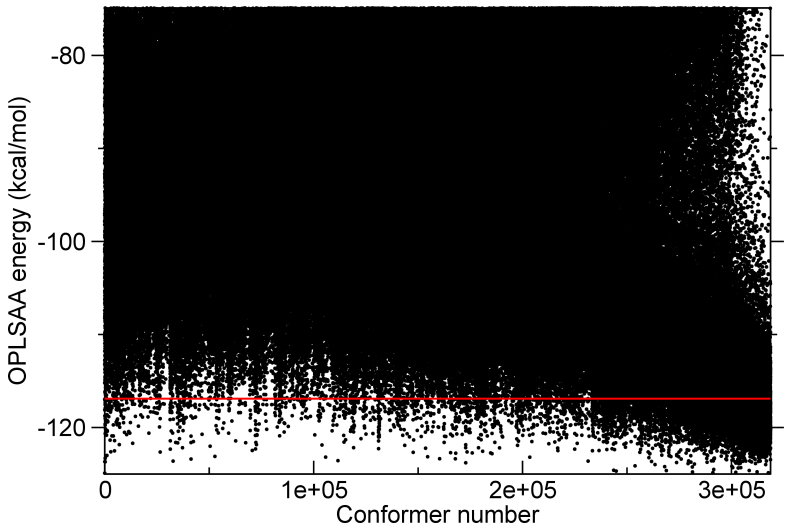


Figure 10.9: Outcome of the basin-hopping search for Ac-β^2hAla$_6$-Lys(H$^+$) using 35 search directions, $\epsilon = 10^{-4}$ kcal/mol and $\Delta = 50$ kcal/mol. Each dot represents one force-field minimum that was found in the search, where the y-axis gives its energy and the x-axis shows the number of the conformer, ranked according to the order in which it was found by the search. The red line marks the lowest 8 kcal/mol (0.35 eV) regime.

Figure 10.10 illustrates the outcome of this relaxation process with DFT. The conformers are ordered according to the force-field energy hierarchy such that the lowest-energy OPLSAA minimum is assigned index 1, the second lowest is assigned index 2 and so on, as described earlier. The red line shows the "funnel plus tail" shape of the force-field curve. Each DFT minimum is assigned the ranking index of the force-field conformer from which it originates and is represented by a dot. For a perfect correlation between force field and PBE+vdW, the dots should lie on top of the red line. This is obviously not the case. However, there is clearly some correlation. When considering the purple dots, which correspond to the PBE+vdW relaxations performed for the OPLSAA conformers beyond the lowest 8 kcal/mol regime, we see that although the dots show a broad scatter, the average energy increases with increasing ranking index, following the OPLSAA trend. This means that loosely trusting the force-field hierarchy in terms of large energy differences is reasonable here. When considering the lowest 8 kcal/mol regime of the force-field hierarchy, which contains 14,234 conformers, there can hardly be seen any correlation between the force-field and the PBE+vdW energy hierarchies. Nevertheless, the lowest PBE+vdW conformer was found by relaxation of the force-field conformer with a ranking index of 741, i.e., coming relatively early in the OPLSAA ranking index. Within the statistics that can be assessed by our combined force field-DFT methodology, we thus conclude as follows. The lowest-energy PBE+vdW conformer identified here is likely the lowest-energy one if the weak observed correlation between the force field and DFT (PBE+vdW) holds up to higher energies. However, a completely definitive assessment would require a different search strategy, directly based on DFT, which is presently not available to our knowledge. The number of torsional degrees of freedom alone is 24, i.e., a simple grid search by discretization is not possible. It would be very interesting to pursue the problem with more sophisticated, direct DFT based methods such as a genetic algorithm. Strategies in this direction are currently being pursued in our group, but go beyond the scope of the present thesis.

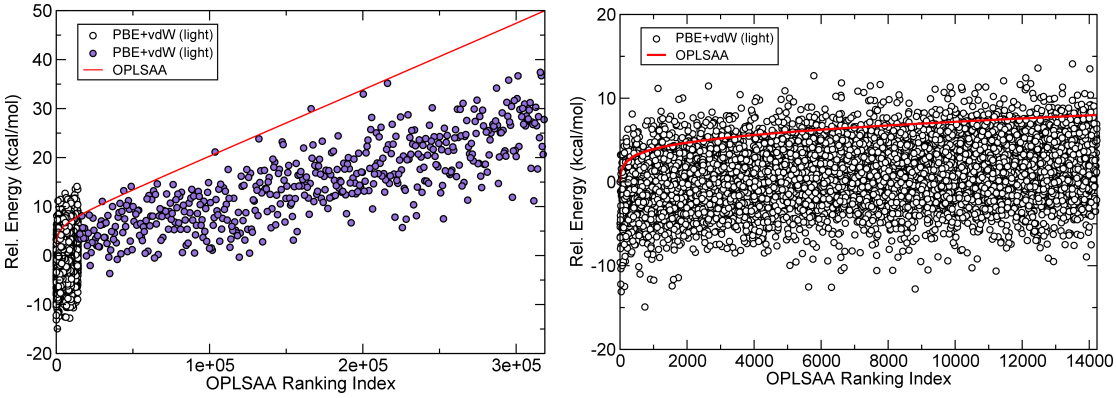


Figure 10.10: Correlation between the OPLSAA and the PBE+vdW energy hierarchies. The red line illustrates the relative force field energy as a function of the ranking index, which is chosen according to the OPLSAA energetic ordering of the conformers. Each dot represents a PBE+vdW minimum relaxed from the force field minimum with the corresponding ranking index. All energies are given relative to the energy of the lowest-energy force field conformer (OPLSAA) or the PBE+vdW minimum following from the relaxation of the latter. The white dots denote PBE+vdW relaxed conformers originating from the lowest 8 kcal/mol force-field regime ("funnel" region) and the purple dots denote conformers relaxed from higher-energy force-field minima. The right panel represents a zoom into the "funnel" region.

The PBE+vdW relaxations discussed above were all performed with *light* computational settings. Following up on that, we relaxed all PBE+vdW conformers found within 400 meV of the lowest PBE+vdW minimum with *tight* computational settings[257]. Figure 10.11 shows a comparison of the energy hierarchies obtained with the force field, PBE+vdW *light* computational settings, and PBE+vdW *tight* computational settings. The energies are all relative to the energy of the conformer that was found to be the lowest-energy minimum in PBE+vdW (both the same for *light* and *tight* computational settings). While the changes in the hierarchy between the force field and PBE+vdW (*light*) are significant, the changes between *light* and *tight* settings are very small. Most conformers show changes of less than 30 meV. Only ten conformers show changes between 30 meV and 60 meV and there is one conformer, whose energy changed by about 160 meV when optimizing with *tight* settings. The reason for this spike is that the structure relaxed to a different minimum with *tight* settings.

In order to analyze the different types of structures that were found in the preceding search, we sorted the PBE+vdW conformers into families according to their hydrogen-bond pattern. The scheme used for labelling the different acceptors and donors for hydrogen bonds is introduced in Fig. 10.12. It shows a picture of the fully extended structure of Ac-β^2hAla$_6$-Lys(H$^+$), with the different backbone carbonyl oxygens and N(H) groups numbered in ascending order from the N- to the C-terminus. The carbonyl oxygen of the acetyl is labelled as O(Ac) and the one of the C-terminus as O(COOH). Apart from the N(H) groups of the backbone there are four other hydrogen atoms that may be involved in a hydrogen bond. We label the hydrogen of the COOH group of the C-terminus as "COOH" and collect the hydrogen atoms of the lysine NH_3^+ group under the label "NH_3^+". The extra methylene groups that distinguish Ac-β^2hAla$_6$-Lys(H$^+$) from Ac-Ala$_6$-Lys(H$^+$) are highlighted with orange rectangles in Fig. 10.12. As already mentioned in Section 10.2.1, we consider a hydrogen bond to be present if the distance between a hydrogen atom and a possible acceptor is less than 2.5 Å. This is a rather broad criterion so that we might count hydrogen bonds that would not be considered a hydrogen bond. However, for us it is more important not to miss a potential hydrogen bond. As an example, we sometimes find

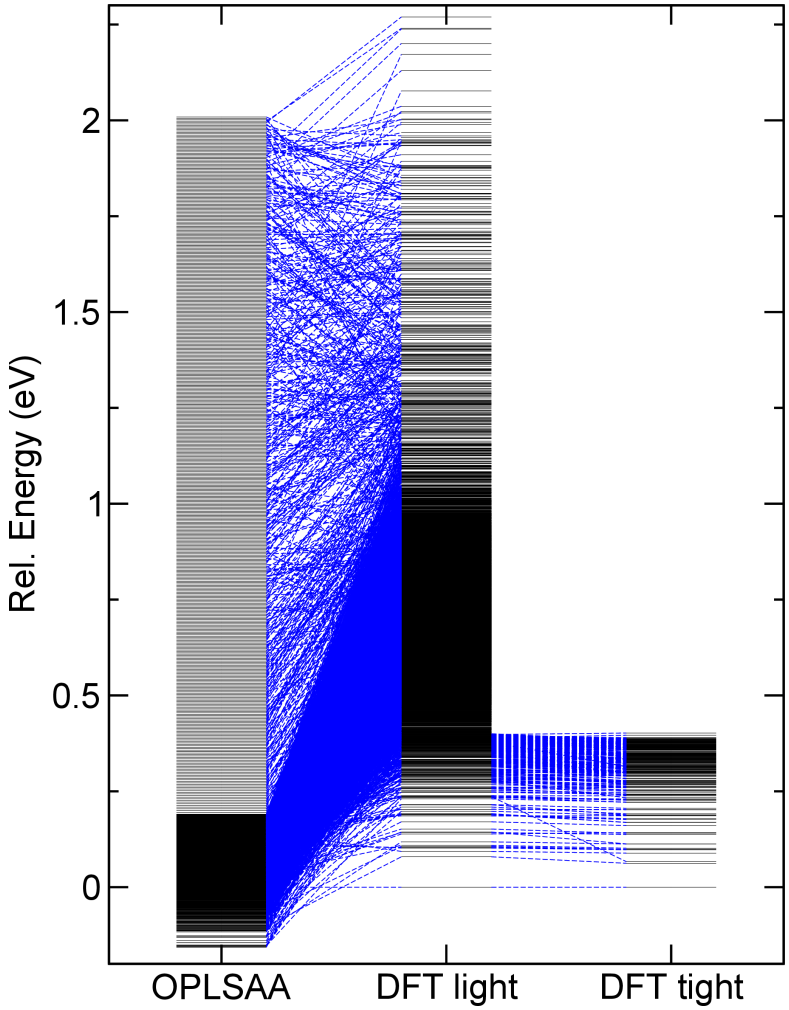


Figure 10.11: Energy hierarchies of Ac-β^2hAla$_6$-Lys(H$^+$) (black horizontal bars) obtained with the OPLSAA force field and from relaxations with PBE+vdW *light* and *tight* computational settings. The energies are given relative to the lowest PBE+vdW minimum structure.

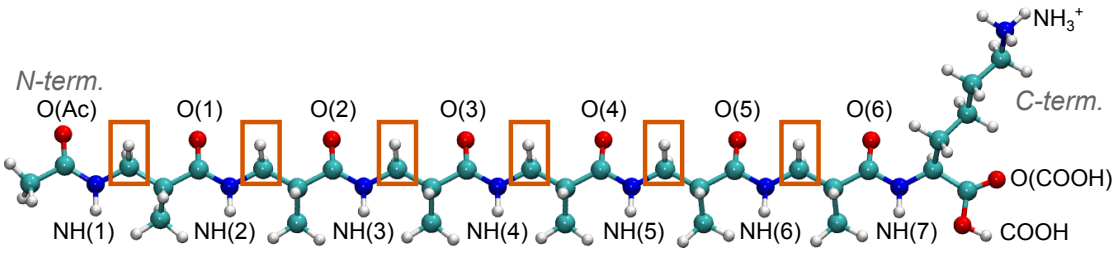


Figure 10.12: Representation of the fully extended structure of Ac-β^2hAla$_6$-Lys(H$^+$) with the labels used for the possible hydrogen bond acceptors and donors. The extra methylene groups that distinguish Ac-β^2hAla$_6$-Lys(H$^+$) from Ac-Ala$_6$-Lys(H$^+$) are highlighted with orange rectangles. In this representation non-polar hydrogens are not omitted (unlike most of the other structural representations shown in this chapter).

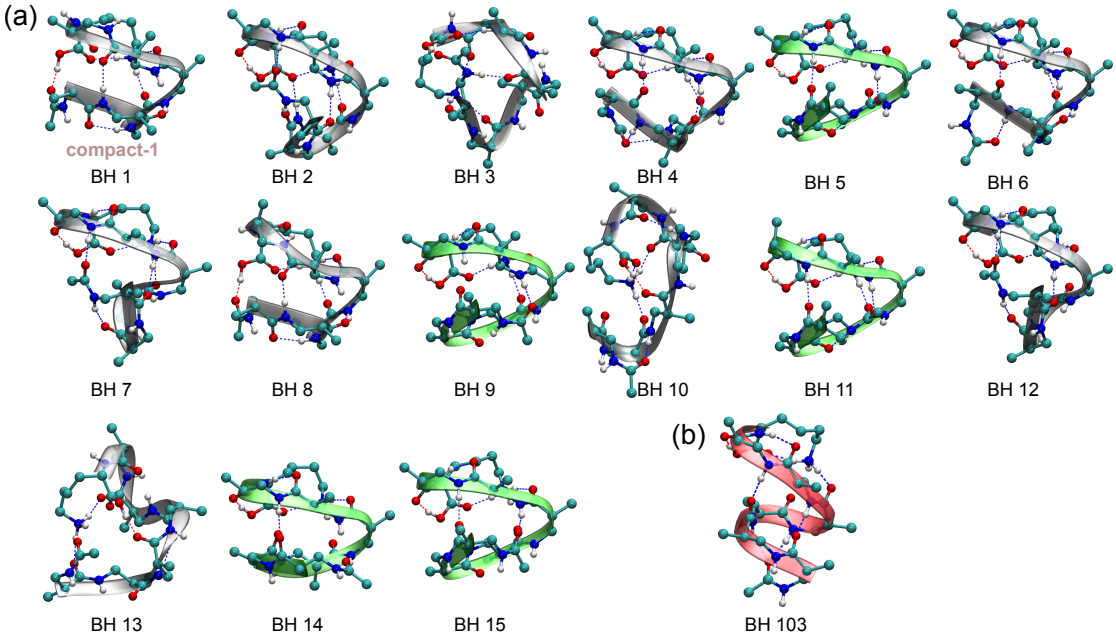
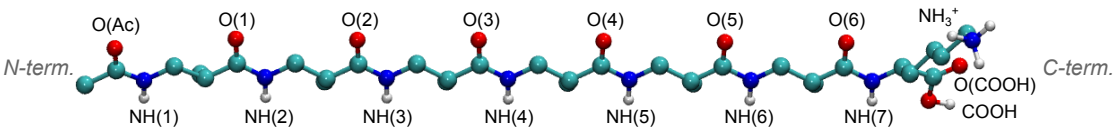


Figure 10.13: Hydrogen-bond families for Ac-β^2hAla$_6$-Lys(H$^+$) up to 200 meV above BH 1 (a) and the lowest energy H16-helical family BH 103 (b). Their detailed hydrogen-bond network is given in Tab. 10.2. H20-helical families are displayed with green ribbons, while H16-helical families are shown with red ribbons.

a distance of less then 2.5 Å between NH(7) and O(COOH) that would not be considered a hydrogen bond due to the geometrical arrangement. However, keeping this information in our clustering approach gives additional structural information on the orientation of the C-terminus.

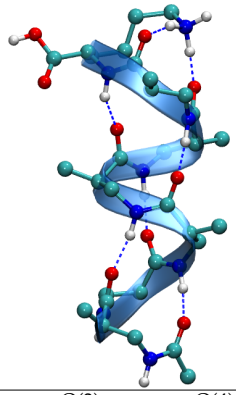
From the 271 conformers that were relaxed with *tight* computational settings, we determined 190 H-bonding families. Table 10.2 lists the hydrogen-bond network of all families within 200 meV of the lowest-energy family. The families are ordered according to their energy hierarchy and labelled as BH (basin hopping) 1, 2, and so on. The relative energies given in Tab. 10.2 refer to the lowest-energy member of each family chosen as the representative member, which are depicted in Fig. 10.13. Family BH 1 is very compact. We also denote it as "compact-1" here, which will be relevant in the following chapters. It contains a turn comprising a 12-membered hydrogen bond. The two ends of the turn are linked by a further hydrogen bond. This means that this structure is similar to a β-hairpin, which is found in natural α-peptides (see Section 2.3). The first H20-helical family is BH 5. Altogether, we find five H20-helical families within the first 15 families (up to 200 meV). However, we do not find any H16-helix nor H12-helix families. The lowest H16-helical family is BH 103, depicted in Fig. 10.13b, with an energy of 340 meV relative to BH 1 (compact-1). We do not find any H12-helical family within the DFT conformers. For this reason, we checked if any of the 318,705 conformers found by the force field was a H12-helix exhibiting the H12-helical hydrogen bonds labelled a, b, and c in Fig. 10.2c. In fact, there was one single match with an OPLSAA energy of 45.38 kcal/mol (1.97 eV). Due to its high force-field energy we did not relax this conformer with PBE+vdW. As discussed earlier, in this high-energy region of the force-field landscape, we find one conformer in each interval ϵ. This implies that the conformational space is very dense in the high-energy regime. If there is more than one structure in the interval ϵ, only the first structure found is stored and all others are discarded.

Table 10.2: Detailed hydrogen-bond networks of the families found within the lowest 200 meV. For each oxygen acceptor atom the corresponding donor groups are given. Additionally, the hydrogen-bond pattern of BH 103, the lowest-energy H16-helix is reported. Relative energies for all families are given with respect to BH 1/compact-1. H12-, H16- and H20-helical hydrogen bonds are highlighted in the respective color code.



BH	O(Ac)	O(1)	O(2)	O(3)	O(4)	O(5)	O(6)	O(COOH)	ΔE (eV)
1[a]	COOH	NH(1) NH(3)	NH(4) NH(5) (H12)	NH_3^+	NH_3^+	NH(2)	free	NH_3^+	0.000
2	NH(6)	NH(5) (H16)	NH(1) NH(3)	NH_3^+	NH_3^+	NH(7)	COOH	NH(3) NH_3^+	0.062
3	COOH	NH_3^+	NH_3^+	NH(3)	NH(1)	NH(4)	NH_3^+	NH(6)	0.089
4	NH(2) NH(3) (H12)	NH(6) (H20)	NH(5) (H12)	NH_3^+	NH_3^+	NH(7)	COOH	NH_3^+	0.098
5	NH(1) NH(5) (H20)	NH(6) (H20)	NH(1)	NH_3^+	NH_3^+	NH(7)	COOH	NH_3^+	0.102
6	NH(2) NH(3) (H12)	NH(6) (H20)	NH(4) NH(5) (H12)	NH_3^+	NH_3^+	NH(7)	COOH	NH_3^+	0.113
7	NH(6)	NH(4) (H12) NH(5) (H16)	NH(1)	NH_3^+	NH_3^+	NH(7)	COOH	NH_3^+	0.113
8	COOH	NH(1) NH(3)	NH(4) NH(5) (H12)	NH_3^+	NH_3^+	NH(2)	NH(6)	NH_3^+	0.142
9	NH(5) (H20)	NH(6) (H20)	NH(1)	NH_3^+	NH_3^+	NH(7)	COOH	NH_3^+	0.169
10	NH(3) (H12)	free	NH_3^+	COOH	NH(4)	NH_3^+	NH(5)	NH_3^+	0.176
11	NH(1) NH(4) (H16) NH(5) (H20)	NH(6) (H20)	NH(1)	NH_3^+	NH_3^+	NH(7)	COOH	NH_3^+	0.179
12	NH(6)	NH(5) (H16)	NH(1)	NH_3^+	NH_3^+	NH(7)	COOH	NH_3^+	0.186
13	NH_3^+	NH_3^+	NH(2)	COOH	NH(3)	NH_3^+	NH(5)	NH_3^+	0.188
14	NH(4) (H16) NH(5) (H20)	NH(6) (H20)	free	NH_3^+	NH_3^+	NH(7)	COOH	NH_3^+	0.189
15	NH(5) (H20)	NH(6) (H20)	free	NH_3^+	NH_3^+	NH(7)	COOH	NH_3^+	0.191
103	NH(3) (H12) NH(4) (H16)	NH(5) (H16)	NH(6) (H16)	NH_3^+	NH_3^+	NH(7)	COOH	NH_3^+	0.342

[a] compact-1



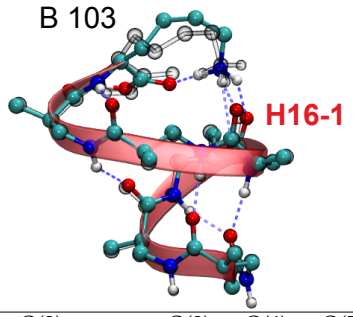
Name	O(Ac)	O(1)	O(2)	O(3)	O(4)	O(5)	O(6)	O(COOH)	ΔE (eV)
H12-1	NH(3) (H12)	NH(4) (H12)	NH(5) (H12)	NH(6) (H12)	NH(7) (H12)	NH_3^+	NH_3^+	free	0.299

Figure 10.14: Structure representation and hydrogen-bond network of the lowest-energy (PBE+vdW) conformer found in the constrained H12-helical basin hopping search. The structure is labelled as H12-1. The energy is given relative to BH 1/compact-1 (see Tab. 10.2 and Fig. 10.9).

In order to analyze more carefully the existence of H12-helices for Ac-β^2hAla$_6$-Lys(H$^+$), we performed a further basin-hopping search starting from the H12-helical conformer shown in Fig. 10.2c. During this search we constrained the H-bonds labelled as a, b, and c in the picture. We did not constrain the helical H-bonds at the acetyl group nor at the C-terminus, as we wanted to keep the termini flexible to sample the H12-helical space as broadly as possible.

10.2.3.2 Constrained basin hopping: H12

From the outcome of the basin-hopping search with constrained H12-helical hydrogen bonds, we relaxed the 1000 lowest-energy force-field minima with PBE+vdW. The lowest-energy structure is depicted in Fig. 10.14 together with its hydrogen-bond network. We label it as H12-1. It is about 300 meV higher in energy than the lowest-energy family, BH 1/compact-1 (see Tab. 10.2 and Fig. 10.9). In order to be able to compare the (energy of) force-field structures of the constrained search and the unconstrained search, we have to relax the conformers from the constrained search without constraints. In order to see if H12-1 was overlooked in the initial unconstrained basin hopping search, we optimized the OPLSAA minimum, which yielded H12-1 without constraints. This lead to only marginal differences in the structure so that we can assume that this optimized OPLSAA structure would result in H12-1 when relaxed with PBE+vdW. When evaluating the OPLSAA energy we find that there is another structure found in the unconstrained basin-hopping search with the exact same energy (within a precision of $\epsilon = 10^{-4}$ kcal/mol). However, this structure is completely different from H12-1. In summary, we cannot tell if the force-field minimum of H12-1 was found in the unconstrained basin-hopping search and discarded because another structure with the same energy had been found at an earlier stage of the search, or, if it was not found at all. However, based on the fact that the respective minimum exists on the (unconstrained) force field potential-energy surface (PES), the first option is more likely.



name	O(Ac)	O(1)	O(2)	O(3)	O(4)	O(5)	O(6)	O(COOH)	ΔE (eV)
H16-1	NH(3) (H12) NH(4) (H16)	NH(5) (H16)	NH(6) (H16)	NH_3^+	NH_3^+	NH(7)	COOH	NH_3^+	0.285

Figure 10.15: Structure representation and hydrogen-bond network of the lowest-energy (PBE+vdW) conformer found in the constrained H16-helical basin hopping search (opaque and colored). The structure is labelled as H16-1. It is superimposed by family BH 103 found in the unconstrained search (transparent gray), which has the same hydrogen-bond network, but shows slight deviations close to the C-terminus and of the lysine side chain. The hydrogen-bond network is listed together with the energy of H16-1 relative to BH 1/compact-1 (see Tab. 10.2 and Fig. 10.9).

10.2.3.3 Constrained basin hopping: H16

As we already found H20-helical families in the low-energy regime of the outcome of the unconstrained basin-hopping search, we did not perform any additional constrained basin-hopping search for this kind of helices. However, the lowest-energy H16-helical family is only number 103 with an energy $\Delta E = 0.342$ eV relative to BH 1/compact-1. For this reason, we analyzed the existence of lower-energy H16-helical conformational families by performing a basin-hopping search where we constrained the H-bonds labelled a and b in Fig. 10.2b. From the outcome of this search we relaxed the lowest 1000 structures with PBE+vdW. The lowest-energy PBE+vdW H16-helical conformer that resulted from this optimization is depicted in Fig. 10.15 and labelled H16-1. It is about 60 meV lower in energy than BH 103. However, they have the same hydrogen-network pattern, which is shown in Fig. 10.15. BH 103 is superimposed on H16-1 in Fig. 10.15. Predominantly, they differ by a slight rearrangement of the lysine side chain. An obvious question is why we did not find this structure in the unconstrained basin-hopping search. In fact, we find that the initial OPLSAA minimum of the constrained search that resulted in H16-1 upon relaxation with PBE+vdW was found in the unconstrained basin-hopping search as well. However, as discussed earlier, we have clustered the force-field structures and only optimized representatives of each cluster with PBE+vdW. For this reason, we did not optimize this particular structure. In fact, the structure that was picked from that particular cluster and relaxed with PBE+vdW was the structure that resulted in family BH 103, the lowest energy H16-helical family found in the unconstrained search.

10.2.4 Summary

In this section, we presented a two-step conformational search for the β-peptide Ac-β^2hAla$_6$-Lys(H$^+$). We first created a huge conformational pool based on force-field simulations and then followed this by relaxing thousands of conformers with DFT (PBE+vdW). The force-field conformational searches were based on a series of basin-hopping runs. We performed one unconstrained search and two searches where we constrained H12-helical hydrogen bonds

and H16-helical hydrogen bonds, respectively. Due to the increased backbone flexibility of Ac-β^2hAla$_6$-Lys(H$^+$) compared to its related α-peptide Ac-Ala$_6$-Lys(H$^+$) the conformational space of Ac-β^2hAla$_6$-Lys(H$^+$) is much denser. In the unconstrained search we found about 15,000 conformers in the lowest 8 kcal/mol regime compared to approximately 1,000 that were found for Ac-Ala$_6$-Lys(H$^+$)[17]. We tackled this challenge by clustering those 15,000 structures in order to identify different structure types. This resulted in 6490 clusters, which we relaxed with DFT (PBE+vdW). In fact, through this clustering process we should not miss any specific structure type, but we might lose individual conformers. The latter has been illustrated based on the outcome of the H16-helical constrained search. The lowest-energy structure found (H16-1) has the same hydrogen-bond network as the lowest-energy H16-helical family (BH 103) found in the unconstrained search. However, we missed H16-1 in the unconstrained search because the initial force-field structures, from which the respective PBE+vdW relaxation were started, belonged to the same cluster.

By comparing the force-field and PBE+vdW energy hierarchies, especially for conformers in the "continuum" regime beyond 8 kcal/mol, we found a weak correlation between the force field and DFT, so that it is likely that any structure that would relax into a low-energy PBE+vdW structure is captured within the lowest 8 kcal/mol. However, we cannot rule out large systematic errors in the force field. H12-helices, e.g., are not sampled within the lowest 8 kcal/mol of the force field, although when comparing PBE+vdW energies they are relatively similar to H16-helices, which are in fact sampled. H12-helices have a very high force-field energy (higher than the 8 kcal/mol funnel regime) with energies in the "continuum" region.

In the next section, we compare the performance of the basin-hopping based conformational searches to a different search strategy based on REMD, which we have already used in a similar fashion in Part II of this thesis.

10.2.5 Search strategy 2: Replica-exchange MD

As in Part II of this thesis, here we also perform a conformational search based on REMD and compare the results to the basin-hopping based structure search. This search is a four-step process:

1. We first sample the conformational space using REMD simulations with the augmented OPLSAA force field (see Section 10.2.2).
2. Then we extract snapshots of the simulated trajectories and sort them into clusters according to their RMSD.
3. Each cluster representative is relaxed with DFT (PBE+vdW).
4. In the end, we follow up with DFT-based REMD simulations that we initialize with the three lowest-energy (PBE+vdW) H12-, H16-, and H20-helices, respectively, to find the lowest-energy structures of the corresponding type.

For the first step, we employed 16 replicas in the temperature range between 300 K and 915 K. Each replica was simulated for 500 ns, i.e., the total simulation time amounted to 8 µs. We then extracted snapshots (interval: 2 ps) of the 300 K trajectory and sorted them into clusters according to their RMSD using the method described in Ref. [99] and a cut-off criterion of 0.05 nm. This yielded 22,554 clusters. We relaxed the 'midpoint' structure of all of the clusters with OPLSAA, where the midpoint structure of a cluster is the structure with the lowest average RMSD to all other structures in the cluster. We distinguish between structures using an energy threshold of 10^{-5} eV yielding 4,629 conformers, which we relaxed with PBE+vdW.

The PBE+vdW relaxations were again performed in two steps. First, all 4,629 conformers were optimized with *light* computational settings. Then the lowest-energy structures were further relaxed with *tight* computational settings. Figure 10.16 shows a comparison of the energy hierarchies obtained with the OPLSAA force field, with PBE+vdW *light* computational settings and *tight* computational settings. As before, we find large deviations between the OPLSAA and the PBE+vdW *light* energy hierarchies. The changes between PBE+vdW *light* and *tight* settings are minimal (less than 40 meV in all cases).

In order to analyze the different structure types found in this search attempt more carefully, we again sorted all DFT structures in families according to their hydrogen-bond network. The families are ordered according to their energy hierarchy and labelled as RE (replica exchange) 1, 2, and so on. The structure representatives of RE 1–11 are shown in Fig. 10.17(a) with their hydrogen-bond network given in Tab. 10.3. Their energies relative to BH 1/compact-1 (the lowest-energy conformer found in the previous basin-hopping based conformational search) is also listed in Tab. 10.3. The first apparent finding is that RE 1 is a H20-helix and about 188 meV higher in energy than BH 1/compact-1. When we compare the hydrogen-bond network and the energy of RE 1 to the families that were found in the basin-hopping based search series, we see that it corresponds to BH 14 (see Tab. 10.2). However, this H20-helix is not the lowest-energy H20-helix that was found in the basin-hopping based search.

Considering H16-helices, we find three such families within 120 meV of RE 1. The first one is RE 7 and corresponds to H16-1, the lowest-energy H16-helix that was found in the basin-hopping based search series (see Fig. 10.15). The lowest-energy H12-helical families are RE 15, 24, and

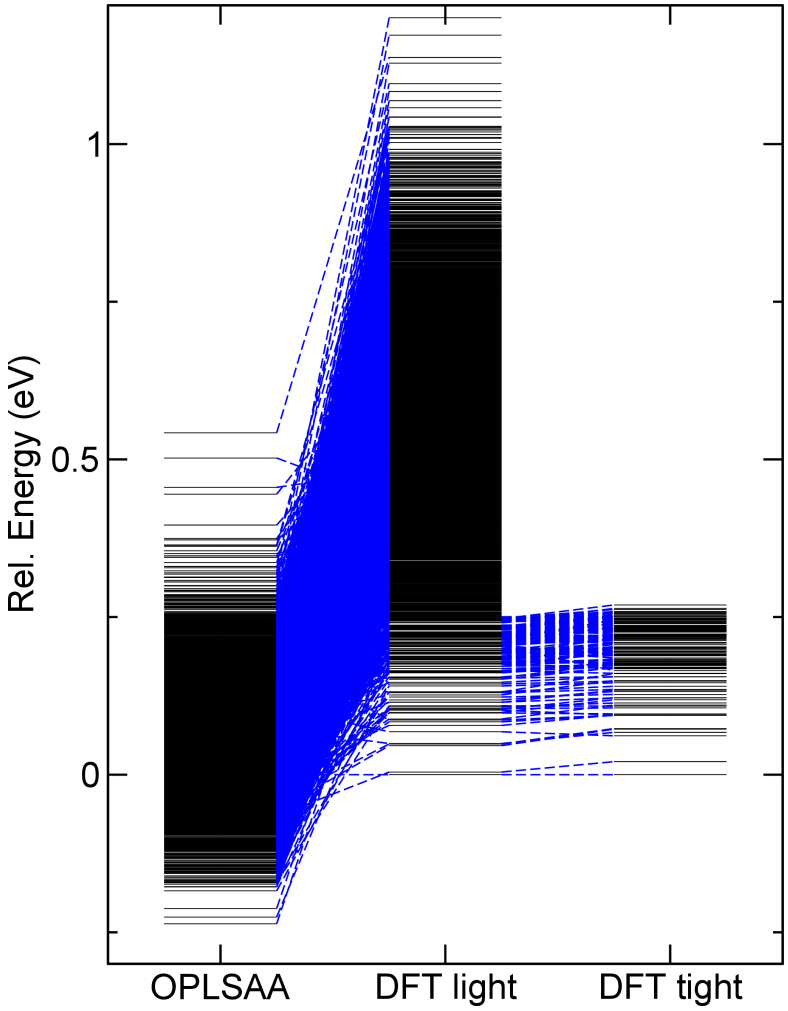


Figure 10.16: Energy hierarchies of Ac-β^2hAla$_6$-Lys(H$^+$) conformers (black horizontal lines) obtained with the OPLSAA force field and from relaxations with the PBE+vdW functional and *light* versus *tight* computational settings. The energies are given relative to the lowest-energy PBE+vdW conformer that was found in the replica-exchange search attempt.

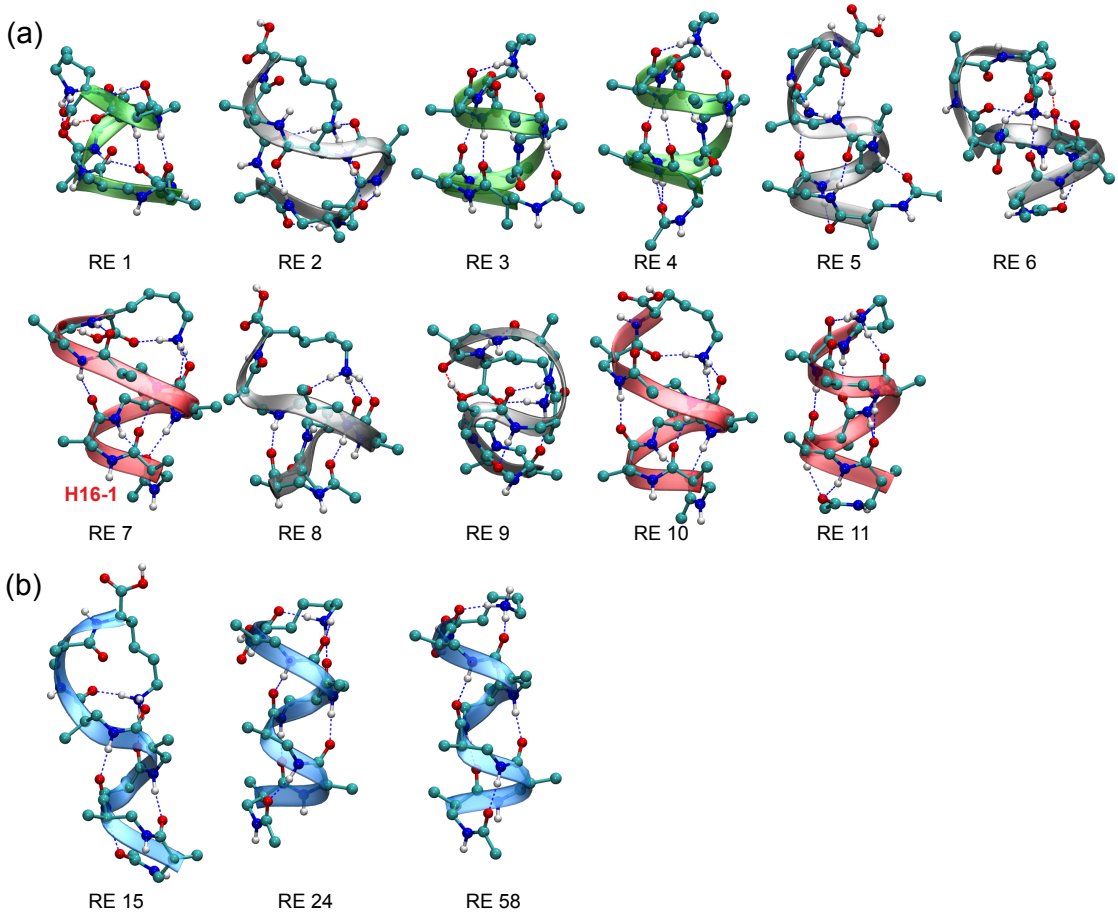
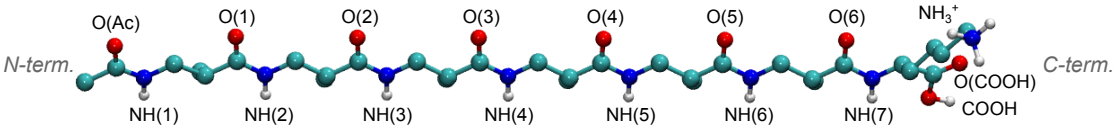


Figure 10.17: Hydrogen-bond families for Ac-β^2hAla$_6$-Lys(H$^+$) found with the REMD-based search strategy: (a) RE 1–11 and (b) RE 15, 24, and 58. Note that RE-7 corresponds to H16-1. Their detailed hydrogen-bond network is given in 10.3. H20-helical families are displayed with green ribbons, while H16-helical families are shown with red ribbons and H12-helices have blue ribbons.

58. They are depicted in Fig. 10.17(b) and their hydrogen-bond network is given in Tab. 10.3. However, the lowest-energy H12-helix found in the previous basin-hopping based search, namely H12-1, is lower in energy than these H12-helices and has a different hydrogen-bond network.

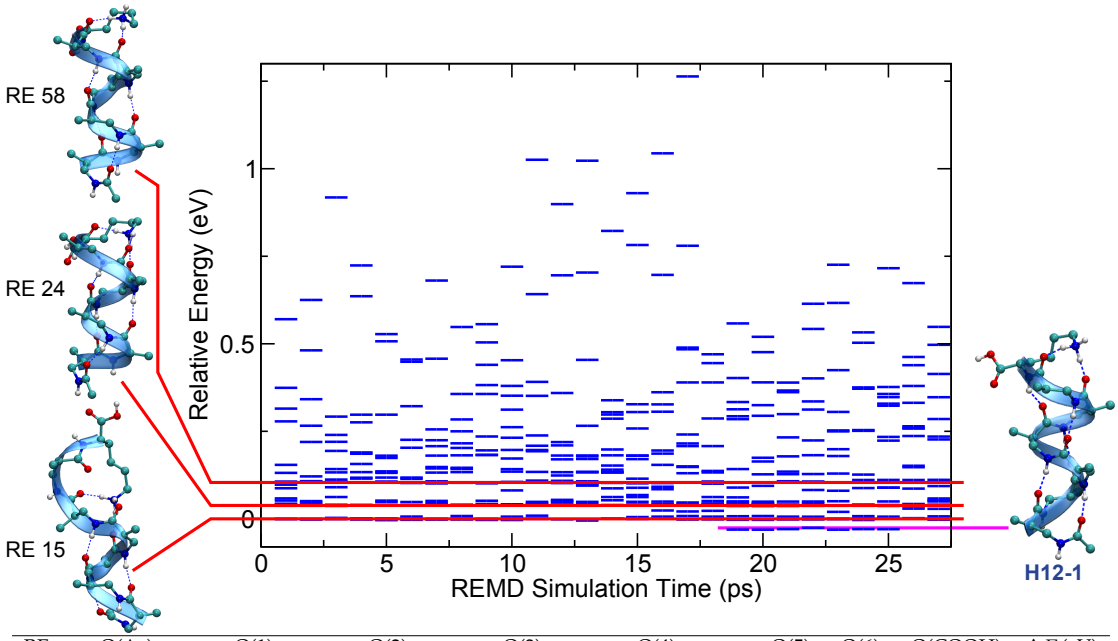
In order to eliminate a possible bias from the force field with respect to the helical structures, we follow up with DFT-based (PBE+vdW) REMD simulations. We performed three REMD runs, one initialized with three H16-helices (RE 7, 10, and 11), one initialized with three H20-helices (RE 1, 3, and 4), and another one initialized with three H12-helices (RE 15, 24, and 58). For each of these runs, we employed 18 replicas in the temperature range between 300 and 687 K. The starting geometry for each replica in a given REMD run was alternately chosen from the three initial helical structures used for that run. The total simulation times amounted to 486 ps (H12), 576 ps (H16), and 558 ps (H20). After each ps all replicas were relaxed with DFT (PBE+vdW) and *light* computational settings leading to 486 (H12), 576 (H16), and 558 (H20) conformers, respectively. As always, the lowest-energy conformers were further relaxed with *tight* computational settings afterwards.

Table 10.3: Detailed hydrogen-bond networks of the lowest-energy hydrogen bond families found based on the REMD strategy. For each oxygen-acceptor atom the corresponding donor groups are given. Additionally, the hydrogen-bond patterns of RE 15, 24, and 58 are given. These are the lowest-energy H12-helical families found by the REMD strategy. The energies are given relative to BH 1/compact-1, the lowest-energy family found with the basin-hopping based conformational search strategy. H12-, H16- and H20-helical hydrogen bonds are highlighted in the respective color code.



RE	O(Ac)	O(1)	O(2)	O(3)	O(4)	O(5)	O(6)	O(COOH)	ΔE (eV)
1	NH(4) (H16) NH(5) (H20)	NH(6) (H20)	free	NH_3^+	NH_3^+	NH(7)	COOH	NH_3^+	0.188
2	NH_3^+	NH(3)	NH(4) NH(5) (H12)	NH_3^+	NH_3^+	NH(2)	NH(6)	NH(7)	0.209
3	NH(5) (H20)	NH(6) (H20)	NH(7) (H20)	free	NH_3^+	NH_3^+	NH_3^+	NH(4) NH(7)	0.260
4	NH(2) NH(3) (H12)	NH(6) (H20)	NH(7) (H20)	free	NH_3^+	NH_3^+	NH_3^+	NH(4) NH(7)	0.261
5	NH(4) (H16)	NH(3)	NH_3^+	NH_3^+	NH(2)	NH_3^+	NH(5)	NH(7)	0.282
6	NH(2) NH(3) (H12)	NH_3^+	NH_3^+	COOH	free	NH_3^+	free	NH(5)	0.283
7[a]	NH(3) (H12) NH(4) (H16)	NH(5) (H16)	NH(6) (H16)	NH_3^+	NH_3^+	NH(7)	COOH	NH_3^+	0.285
8	NH(5) (H20)	NH(6) (H20)	free	NH_3^+	NH_3^+	NH(3) NH_3^+	free	NH(7)	0.297
9	free	NH_3^+	NH(4)	NH(5) NH(6) (H12)	NH_3^+	NH(7)	COOH	NH_3^+	0.299
10	NH(3) (H12) NH(4) (H16)	NH(5) (H16)	NH(6) (H16)	NH_3^+	NH_3^+	free	NH_3^+	NH(7)	0.302
11	NH(2) NH(3) (H12)	NH(4) (H12) NH(5) (H16)	NH(6) (H16)	NH(7) (H16)	NH_3^+	NH_3^+	NH_3^+	NH(7)	0.307
15	NH(3) (H12)	NH(4) (H12)	NH(5) (H12)	NH_3^+	NH_3^+	NH_3^+	free	NH(7)	0.323
24	NH(3) (H12)	NH(4) (H12)	NH(5) (H12)	NH(6) (H12)	NH_3^+	NH_3^+	NH_3^+	NH(7)	0.362
58	NH(3) (H12)	NH(4) (H12)	NH(5) (H12)	NH(6) (H12)	free	NH_3^+	NH_3^+	NH(7)	0.423

[a] H16-1



RE	O(Ac)	O(1)	O(2)	O(3)	O(4)	O(5)	O(6)	O(COOH)	ΔE (eV)
15	NH(3) (H12)	NH(4) (H12)	NH(5) (H12)	NH_3^+	NH_3^+	NH_3^+	free	NH(7)	0.323
24	NH(3) (H12)	NH(4) (H12)	NH(5) (H12)	NH(6) (H12)	NH_3^+	NH_3^+	NH_3^+	NH(7)	0.362
58	NH(3) (H12)	NH(4) (H12)	NH(5) (H12)	NH(6) (H12)	free	NH_3^+	NH_3^+	NH(7)	0.423
H12-1	NH(3) (H12)	NH(4) (H12)	NH(5) (H12)	NH(6) (H12)	NH(7) (H12)	NH_3^+	NH_3^+	free	0.296

Figure 10.18: Energy (red horizontal bars) of all replicas relaxed after each ps of REMD simulation time. The energy is given relative to the lowest-energy input conformer (RE 15). The structures on the left side of the plot are the structures used as initial structures for the REMD run (RE 15, RE 24, RE 58) meaning that the starting geometry for each of the 18 replicas was alternately chosen from these three initial helical structures. The structure depicted on the right side is the lowest-energy structure found in this search, which is the already known H12-helical structure H12-1 that we have found in the basin-hopping based search strategy before. The table gives the hydrogen-bond network of all depicted families together with their energy relative to BH 1/compact-1.

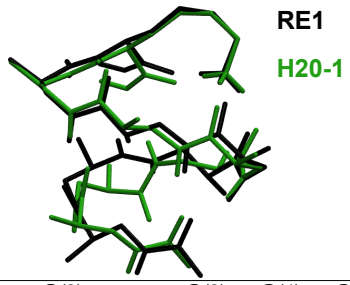
10.2.5.1 *Ab initio* REMD

We start with an analysis of the *ab initio* REMD run initialized with the H12-helical families RE 15, 24, and 58 (cf. Fig. 10.17b). Figure 10.18 shows the energy of all relaxed replicas after each ps of simulation time relative to RE 15. In fact, we find a family, that has a lower energy than RE 15, 24 and 58. Its hydrogen-bond network and relative energy is detailed in the table in Fig. 10.18. In fact, this family corresponds to H12-1, the lowest-energy H12-helical family that was found in the basin-hopping based series of conformational searches.

In a similar search initialized with the H16-helices RE 7, 10, and 11, we did not find any lower-energy conformers than the initial structures.

Initializing an REMD run with the H20-helical families RE 1, 3, 4 we find a lower energy structure, named H20-1. It has the same hydrogen-bond network as RE 1, deviating slightly in the backbone structure as can be seen by the superimposed arrangement of RE 1 and H20-1 in Fig. 10.19. H20-1 is about 60 meV lower in energy than RE 1. Still, H20-1 has a different hydrogen-bond network and is about 20 meV higher in energy than the lowest H20-helix BH 5 found in the basin-hopping search.

Figure 10.20 shows the hydrogen-bond network evolution of the 300 K trajectory during all



RE/name	O(Ac)	O(1)	O(2)	O(3)	O(4)	O(5)	O(6)	O(COOH)	ΔE (eV)
1	NH(4) (H16) NH(5) (H20)	NH(6) (H20)	free	NH_3^+	NH_3^+	NH(7)	COOH	NH_3^+	0.188
3	NH(5) (H20)	NH(6) (H20)	NH(7) (H20)	free	NH_3^+	NH_3^+	NH_3^+	NH(4) NH(7)	0.260
4	NH(2) NH(3) (H12)	NH(6) (H20)	NH(7) (H20)	free	NH_3^+	NH_3^+	NH_3^+	NH(4) NH(7)	0.261
H20-1	NH(4) (H16) NH(5) (H20)	NH(6) (H20)	free	NH_3^+	NH_3^+	NH(7)	COOH	NH_3^+	0.125

Figure 10.19: Superimposed structures of the newly found conformer H20-1 (green) and RE 1 (black). The table gives the hydrogen-bond network of all H20-helical families RE 1, RE 3, and RE4 used as input structures for the *ab initio* REMD simulation that produced H20-1. The energies are given relative to BH 1/compact-1.

three *ab initio* REMD simulations. The families from which the REMD trajectories were initialized are displayed on the left side of the plots. On the right side of the plots are snapshots of the structures taken at the simulation time as specified in the picture. For each oxygen starting from O(Ac) to O(6) the H-bonding connection is plotted. The labels used for the oxygens are the same as explained in Fig. 10.12. For simplicity, we concentrate on H12-, H20-, and H16-helical H-bonds and on H-bonds to the NH_3^+ terminus. As throughout this thesis, a hydrogen bond is considered to be present if the distance O$\cdots$H is smaller than 2.5 Å. If a specific H-bond is present at a given simulation time, this is indicated by a bar, where the different types of hydrogen bonds are represented by a different color and a different position of the bar as specified in the figure caption. As apparent from Fig. 10.20a, in the first-principles REMD simulation initialized solely from H12-helical conformations, H16-helices are also found in the 300 K trajectory. When initializing only from H16-helices, H20- and H12-helices are also found (Fig. 10.20b), and in Fig. 10.20c it is clear that when initializing from H20-helices, H16-helices are found. In REMD, individual molecular dynamics (MD) trajectories are propagated at different temperatures with a periodic swap attempt for the replicas. Therefore, Fig. 10.20 does not show a continuous MD trajectory. However, it shows that initially exclusively conformers of one type can co-exist with the other types in a canonical ensemble and that the barriers are evidently not insurmountable although these runs do not allow us to quantify how high they are.

10.3 Summary

We performed two individual strategies to search the conformational space of the β-peptide Ac-β^2hAla$_6$-Lys(H^+) on a first-principles level. The first one was based on a series of basin-hopping runs (strategy 1), and the other one was based on a series of REMD simulations (strategy 2). Both strategies started with a global sampling of the structure space based on the OPLSAA force

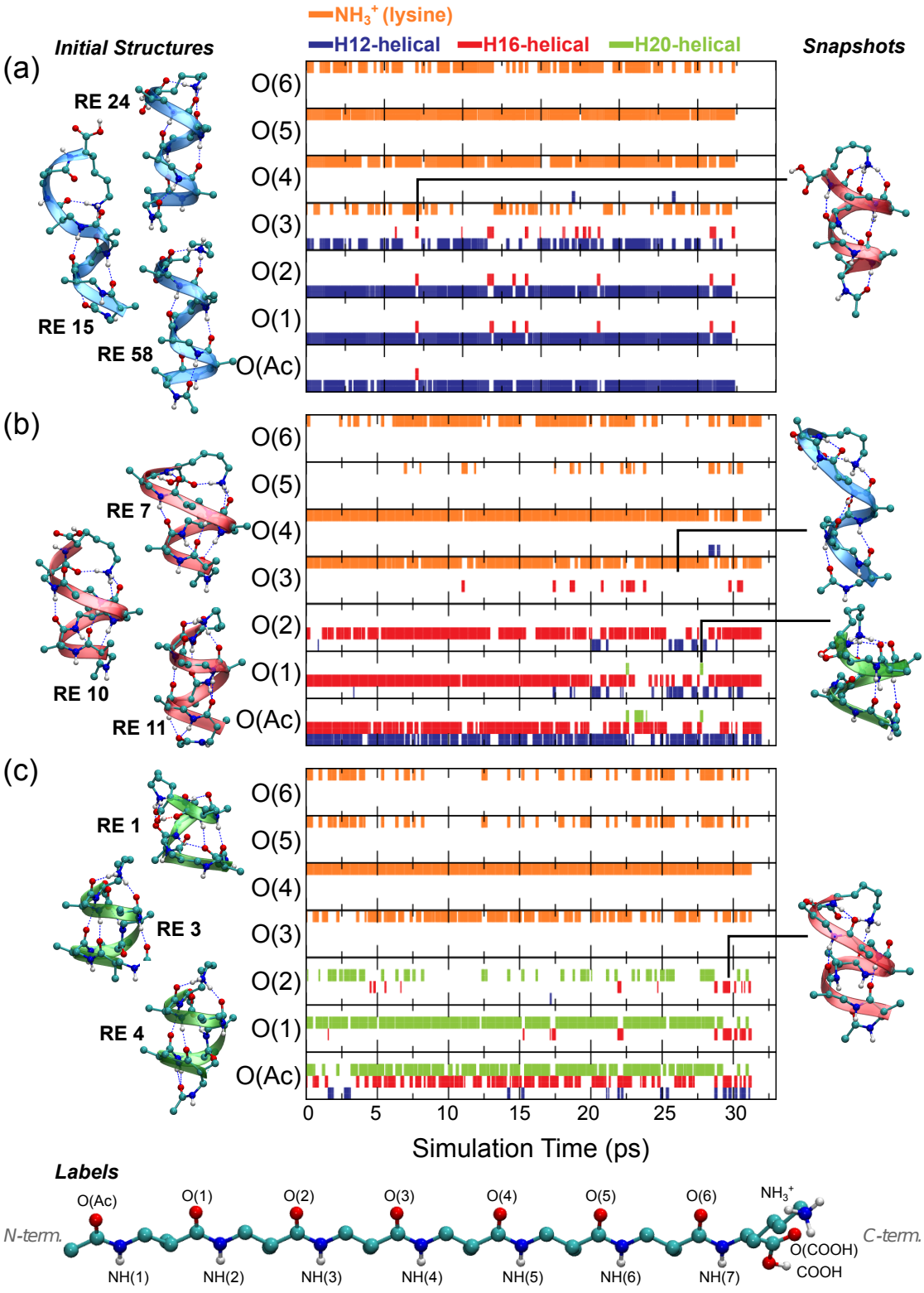


Figure 10.20: Hydrogen-bond network evolution of the 300 K trajectory of the DFT-based REMD run started from H12-helices (a), H16-helices (b), and H20-helices (c). Each graph shows the H-bonding connections for one oxygen, where the oxygens run from O(Ac) to O(6) as explained at the bottom. Each hydrogen bond is represented by a bar, where the color and the position denote the type of H-bond. Blue bars at the bottom of a graph denote an H12-helical bond, red bars occupy the second quarter of a plot and denote an H16-helical bond, green bars in the third quarter of a graph represent an H20-helical bond, and brown bars in the fourth quarter of a graph denote hydrogen bonds formed with the lysine NH_3^+ group. The starting geometry for each replica in a given REMD run was alternately chosen from the three initial helical structures displayed at the left side of the respective plot. On the right side of each plot snapshots of structures taken at the specified time intervals are displayed.

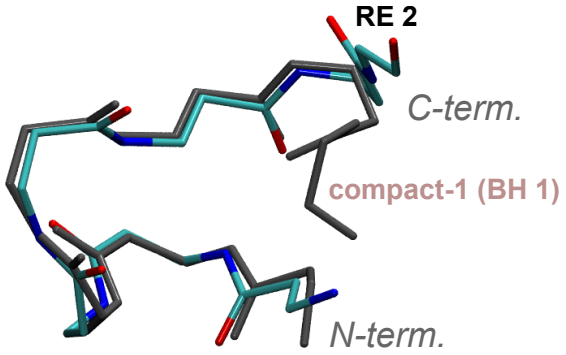


Figure 10.21: Superposition of BH 1/compact-1 (grey) and RE 2 (colored). Hydrogen atoms and side chains are omitted for clarity.

field and followed up with a refinement using DFT, specifically the PBE+vdW functional.

In order to compare the outcome of the two searches, we will focus on two aspects, namely the performance with respect to helical structures and the difference in the lowest-energy structure that was found. With respect to H12-helices and H16-helices we found the same lowest-energy helical families with both methods. This makes it very likely that we have in fact found the lowest-energy H16- and H12-helices. For the H20-helices the situation is more complicated. The *ab initio* REMD search (strategy 2) initialized with H20-helices was not long enough to find the lowest-energy H20-helix found with strategy 1. However, the lowest-energy H20-helix that was found in the *ab initio* REMD run, named H20-1, was not found in the basin-hopping based search in turn. We have to note though that we did not perform a special basin-hopping search with constrained H20-helical hydrogen bonds. In fact, H20-1 will turn out to be the H20-helical structure motif with the lowest free energy (at 300 K) in the next chapter.

Considering the lowest-energy structure that was found with the two strategies, the basin-hopping based search found BH 1/compact-1, which is a conformer similar to a β-hairpin. On the other hand, with the REMD based search we found RE 1, a H20-helix, to be the lowest-energy structure although 188 meV higher in energy than compact-1. When considering the next family in the energy hierarchy established by the REMD search, RE 2, we see that this family is very similar to BH 1/compact-1. A superimposed picture of both families is given in Fig. 10.21, which shows that the two structures predominantly deviate in the terminations. In fact, when not considering the terminations, the backbone RMSD between the two structures is only 0.6 Å. For reasons of computational cost we did not perform an explicit check. However, it seems quite certain that an *ab initio* REMD simulation initialized from RE 2, would lead to BH 1/compact-1.

11 Conformational preferences: Ac-β^2hAla$_6$-Lys(H^+) vs. Ac-Ala$_6$-Lys(H^+)

After assessing the huge conformational space of Ac-β^2hAla$_6$-Lys(H^+) using two different first-principles based search techniques in the previous chapter, we here bring together the results in order to analyze the conformational preferences of Ac-β^2hAla$_6$-Lys(H^+). Subsequently, we perform a comparison to the equivalent natural peptide Ac-Ala$_6$-Lys(H^+)[17, 28].

11.1 Results for Ac-β^2hAla$_6$-Lys(H^+)

From the previous searches of the conformational space for Ac-β^2hAla$_6$-Lys(H^+), we obtained 14,739 PBE+vdW relaxed conformations with *light* computational settings. In detail, these resulted from:

1. Strategy 1 (Section 10.2.3)
   - unconstrained basin-hopping search: 6,490
   - constrained basin-hopping search for H16 helices: 1,000
   - constrained basin-hopping search for H12 helices: 1,000
2. Strategy 2 (Section 10.2.5)
   - force-field based replica-exchange molecular dynamics (REMD): 4,629
   - density-functional theory (DFT)-based REMD initialized by H12 helices: 486
   - DFT-based REMD initialized by H16 helices: 576
   - DFT-based REMD initialized by H20 helices: 558

All conformers within an energy window of 400 meV of the global minimum were relaxed with PBE+vdW *tight* settings.[1] In order to analyze the different conformational types we again sort all conformers into families according to their hydrogen-bonding network. Within each family the lowest-energy member is chosen as the representative of this family. As before, all energies or structural representations that are reported here for a specific family always refer to this corresponding representative.

[1] In fact, all conformers up to at least 450 meV that did not result from the unconstrained basin-hopping search were relaxed with *tight* settings.

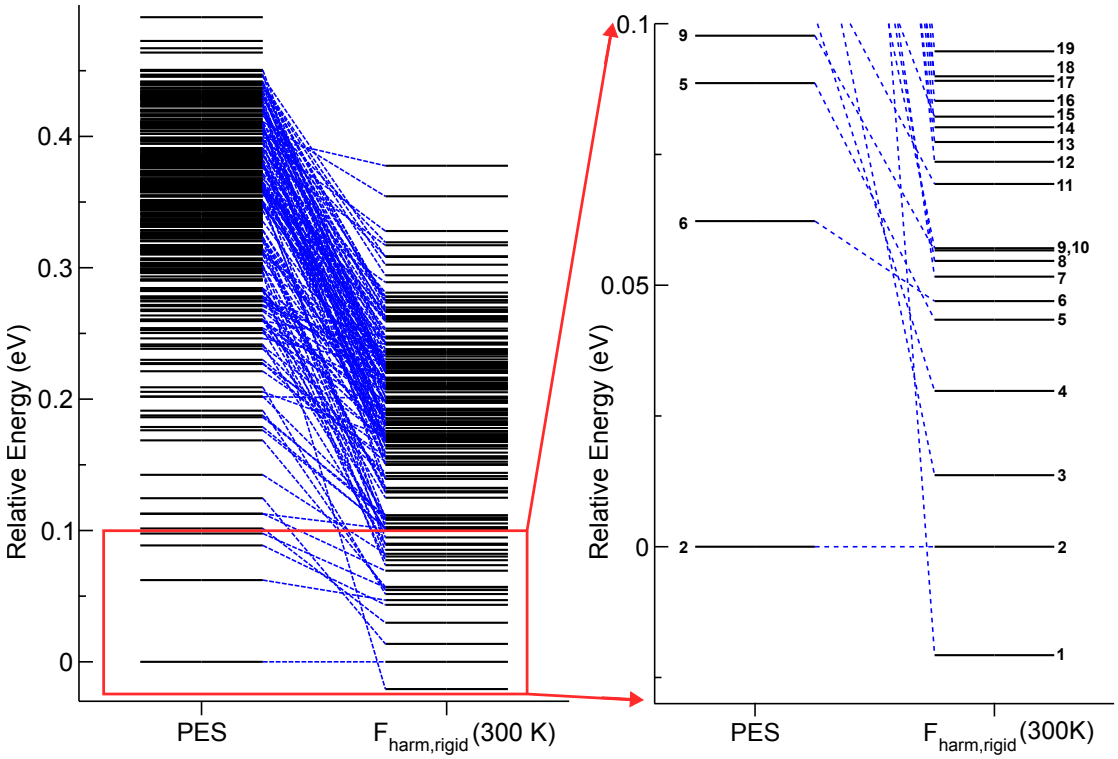


Figure 11.1: Energy hierarchies (horizontal bars) of the hydrogen-bonding families found for Ac-β^2hAla$_6$-Lys(H$^+$) based on the PBE+vdW functional and *tight* computational settings. The horizontal bars represent the lowest-energy representative of each family. All of those structures represent a local minimum on the PBE+vdW potential-energy surface (PES). In addition to the PES hierarchy, the energy hierarchy after adding corrections for the free energy at 300 K in the harmonic oscillator-rigid rotor approximation are shown ($F_{\mathrm{harm,rigid}}$). At the right side an inset of the energy hierarchies is plotted including all families up to 100 meV from the global minimum of the PES. The families are numbered according to the free-energy hierarchy.

As discussed extensively in previous chapters, the actual quantity that guides the behavior of the system at finite temperature is the free energy. Using the harmonic oscillator-rigid rotor approximation (Section 5.1.1), we calculated the vibrational and rotational contributions to the free energy at 300 K for the hydrogen-bonding families found for Ac-β^2hAla$_6$-Lys(H$^+$).[2] The normal modes of vibration are calculated from a finite-differences approach. A plot that demonstrates the convergence of the modes with respect to the parameters we used for the calculation can be found in Appendix B.1 (see also Ref. [28]).

Figure 11.1 compares the energy hierarchies of all hydrogen-bonding families of Ac-β^2hAla$_6$-Lys(H$^+$) without and with vibrational and rotational contributions to the free energy at 300 K. Here we label the families according to the free-energy hierarchy. Structural representations of the families are illustrated in Fig. 11.2, while their hydrogen-bonding networks and relative free energies are given in Tab. 11.1. The inset of the energy-hierarchy plot at the right side of Fig. 11.1 shows the lowest 100 meV regime above the global minimum of the PES. The global minimum of the PES is family 2, which was also called "compact-1" (see Chapter 10), a compact structure

[2] Apart from the families that resulted from the unconstrained basin-hopping search, we calculated the free-energy corrections for all families within at least 400 meV of the global minimum of the potential-energy surface (PES). For families that resulted from the unconstrained basin-hopping search the energy window was at least 280 meV plus families extended like an H16 or H12-helix.

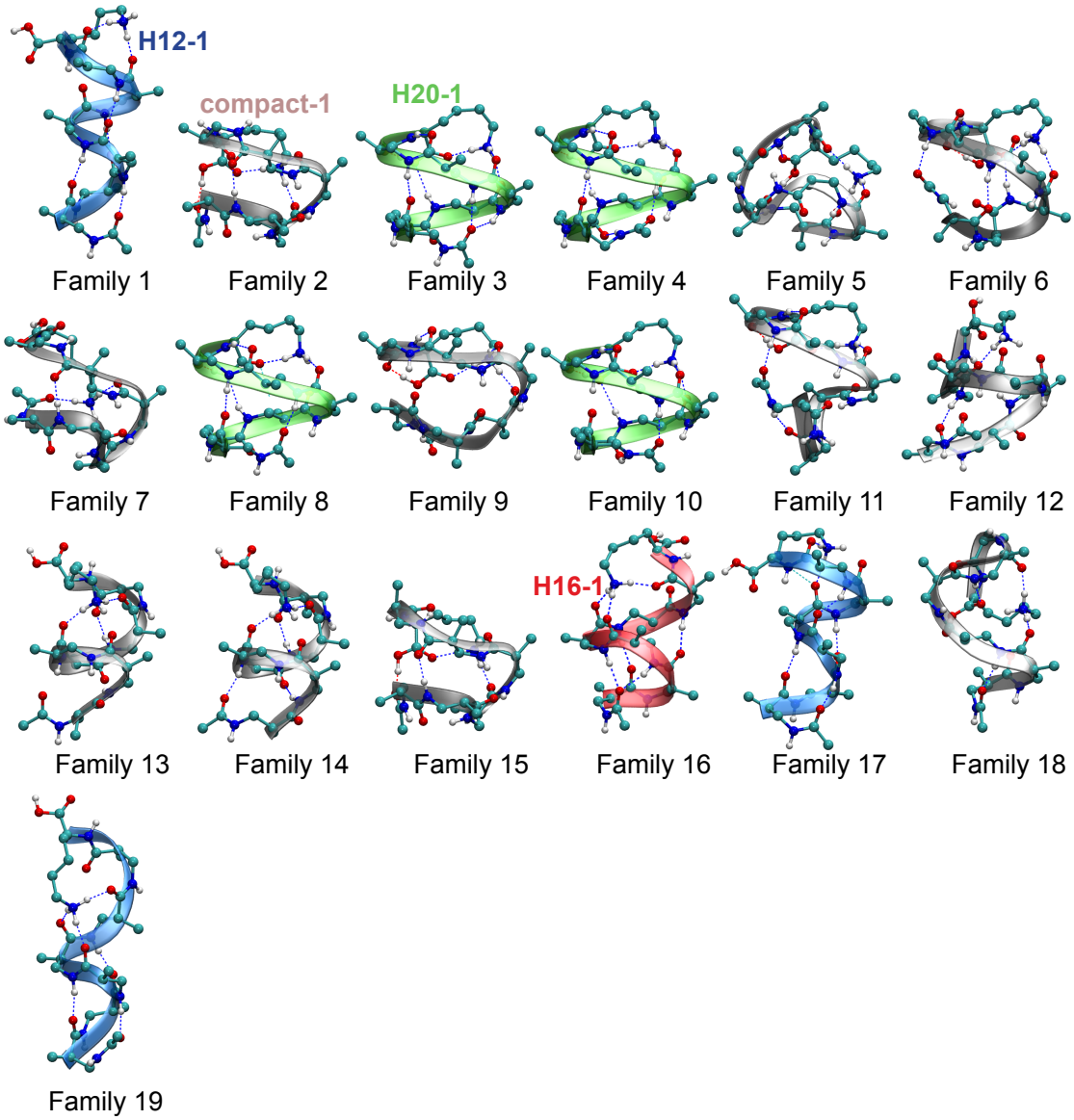
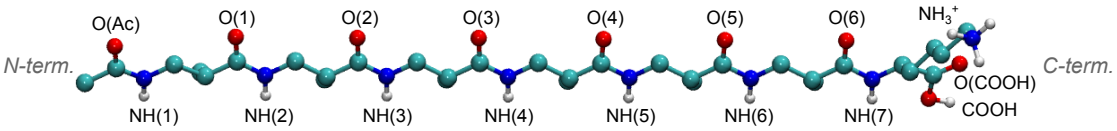


Figure 11.2: Structure representations of the hydrogen-bonding families of Ac-β^2hAla$_6$-Lys(H$^+$) including all families with a free energy of up to 100 meV above the global minimum of the PES (family 2/compact-1). The families are numbered according to the free-energy hierarchy (300 K, harmonic oscillator-rigid rotor approximation). H12-helices are depicted with blue ribbons, H20-helices with green ribbons and H16-helices with red ribbons. In addition to the family number the lowest-energy family of each helix type and compact family is labelled as H12-1, compact-1, H20-1, and H16-1.

Table 11.1: Hydrogen-bond networks of the families found within 100 meV of family 2/compact-1. The families are numbered according to the free-energy hierarchy (300 K, harmonic oscillator-rigid rotor approximation). In addition to the family number the lowest-energy family of each helix type and compact family is labelled as H12-1, compact-1, H20-1, and H16-1. For each oxygen acceptor atom the corresponding donor groups are listed. The free energies are given relative to family 2/compact-1.



F	O(Ac)	O(1)	O(2)	O(3)	O(4)	O(5)	O(6)	O(COOH)	$\Delta F_{\mathrm{harm,rigid}}$ (eV)
1[a]	NH(3) (H12)	NH(4) (H12)	NH(5) (H12)	NH(6) (H12)	NH(7) (H12)	NH_3^+	NH_3^+	free	-0.021
2[b]	COOH	NH(1) NH(3)	NH(4) NH(5) (H12)	NH_3^+	NH_3^+	NH(2)	free	NH_3^+	0.000
3[c]	NH(4) (H16) NH(5) (H20)	NH(6) (H20)	free	NH_3^+	NH_3^+	NH(7)	COOH	NH_3^+	0.014
4	NH(1) NH(5) (H20)	NH(6) (H20)	NH(1)	NH_3^+	NH_3^+	NH(7)	COOH	NH_3^+	0.030
5	COOH	NH_3^+	NH_3^+	NH(3)	NH(1)	NH(4)	NH_3^+	NH(6)	0.043
6	NH(6)	NH(5) (H16)	NH(1) NH(3)	NH_3^+	NH_3^+	NH(7)	COOH	NH(3) NH_3^+	0.047
7	NH_3^+	NH(3)	NH(4) NH(5) (H12)	NH_3^+	NH_3^+	NH(2)	NH(6)	NH(7)	0.052
8	NH(5) (H20)	NH(6) (H20)	free	NH_3^+	NH_3^+	NH(7)	COOH	NH_3^+	0.055
9	NH(2) NH(3) (H12)	NH(6) (H20)	NH(5) (H12)	NH_3^+	NH_3^+	NH(7)	COOH	NH_3^+	0.057
10	NH(5) (H20)	NH(6) (H20)	NH(1)	NH_3^+	NH_3^+	NH(7)	COOH	NH_3^+	0.057
11	NH(6)	NH(4) (H12) NH(5) (H16)	NH(1)	NH_3^+	NH_3^+	NH(7)	COOH	NH_3^+	0.069
12	NH(5) (H20)	NH(6) (H20)	free	NH_3^+	NH_3^+	NH(3) NH_3^+	free	NH(7)	0.074
13	NH(4) (H16)	NH(3)	NH_3^+	NH_3^+	NH(2)	NH_3^+	NH(5)	NH(7)	0.077
14	NH(4) (H16)	free	NH_3^+	NH_3^+	NH(2)	NH_3^+	NH(5)	NH(7)	0.080
15	COOH	NH(1) NH(3)	NH(4) NH(5) (H12)	NH_3^+	NH_3^+	NH(2)	NH(6)	NH_3^+	0.082
16[d]	NH(3) (H12) NH(4) (H16)	NH(5) (H16)	NH(6) (H16)	NH_3^+	NH_3^+	free	NH_3^+	NH(7)	0.085
17	NH(3) (H12)	NH(4) (H12)	NH(5) (H12)	NH(6) (H12)	NH_3^+	NH_3^+	NH_3^+	NH(7)	0.089
18	NH(3) (H12)	free	NH_3^+	COOH	free	NH_3^+	free	NH(5) NH_3^+	0.090
19	NH(3) (H12)	NH(4) (H12)	NH(5) (H12)	NH_3^+	NH_3^+	NH_3^+	free	NH(7)	0.095

[a] H12-1
[b] compact-1
[c] H20-1
[d] H16-1

similar to a β-hairpin in α-peptides. The other three families found within the lowest 100 meV of the PES are rather compact structures as well.

Measured again from the global minimum of the PES, we find 19 families within the lowest 100 meV of the free-energy regime, where there were only four for the PES. While Family 2/compact-1 is the global minimum of the PES, it is only the second-lowest minimum regarding free energies. The global minimum is formed by the H12-helix family 1, which we called H12-1 in Chapter 10. It is about about 21 meV lower in free energy than family 2/compact-1. Family 3 is a H20-helix, which we named H20-1 earlier. The first H16-helical family is family 16, which was referred to as H16-1 in Chapter 10.

While the horizontal bars in Fig. 11.1 show the (free) energies relative to family 2/compact-1, the blue lines indicate the changes in the energy hierarchies given by $\Delta F_{\mathrm{vib(harmonic\ oscillator)}} + \Delta F_{\mathrm{rot(rigid\ rotor)}}$. In fact, family 1/H12-1 is stabilized by 317 meV relative to family 2/compact-1. For H20-1 the stabilization amounts to 111 meV and for H16-1 it is 217 meV. When analyzing this further, we find that the rigid-rotor contribution is minimal (up to 10 meV). This means, that the stabilization of H12-1, H20-1, and H16-1 relative to family 2/compact-1 originates from the vibrational contribution to the free energy (in the harmonic oscillator approximation). A detailed list of the different contributions for all families can be found in Tab. B.1 in Appendix B.2.

11.2 Comparison of Ac-β^2hAla$_6$-Lys(H$^+$) and Ac-Ala$_6$-Lys(H$^+$)

Stabilization of helical motifs by vibrational free energy compared to more compact structures is already known in the literature[17, 28, 413, 439]. Recently, this phenomenon was analyzed by Mariana Rossi from our group for the natural-peptide series Ac-Ala$_n$-LysH$^+$, $n = 4, \ldots, 8$[17, 28]. The energy hierarchies of the conformational families found for Ac-Ala$_6$-Lys(H$^+$)[17, 28] are shown in a comparison to the hierarchies of Ac-β^2hAla$_6$-Lys(H$^+$) in Fig. 11.3. Just as for Ac-β^2hAla$_6$-Lys(H$^+$), the families of Ac-Ala$_6$-Lys(H$^+$) are labelled according to their free-energy hierarchy. The global minimum of the PES of Ac-Ala$_6$-Lys(H$^+$), family 2, is a non-helical motif. In the lowest 100 meV regime from family 2, there are four families present. This is the same for Ac-β^2hAla$_6$-Lys(H$^+$). Considering the lowest 100 meV of the free-energy regime (seen from the global minimum of the PES), for Ac-Ala$_6$-Lys(H$^+$) again only the same four families are present. By adding vibrational and rotational corrections to the free energy at 300 K the α-helical family 1 is stabilized relative to the non-helical family 2 by about 108 meV, where again the rigid-rotor contributions are minimal (4 meV).

The stabilization we find in the case of Ac-β^2hAla$_6$-Lys(H$^+$) for H12-1 relative to family 2/compact-1 is much larger (317 meV). Another difference between Ac-β^2hAla$_6$-Lys(H$^+$) and Ac-Ala$_6$-Lys(H$^+$) is that for Ac-β^2hAla$_6$-Lys(H$^+$) the low free-energy region is much more densely populated than for Ac-Ala$_6$-Lys(H$^+$). However, for Ac-Ala$_6$-Lys(H$^+$) the free-energy corrections were calculated only for a few family representatives[17, 28], i.e., there might be more than four families in the low free-energy regime. Helical families, which were found to be stabilized most[17, 28] by vibrational free-energy contributions, were all included though.

In order to understand the origin of the stabilization of helical motifs by vibrational free energy and why the effect is more pronounced in Ac-β^2hAla$_6$-Lys(H$^+$) than in Ac-Ala$_6$-Lys(H$^+$),

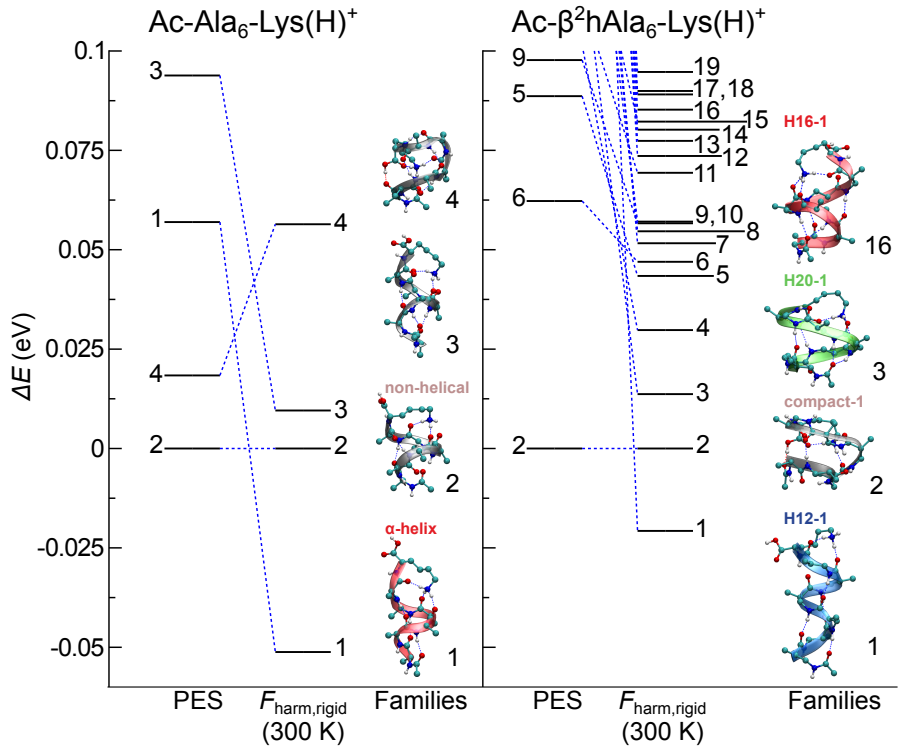


Figure 11.3: Energy hierarchies (horizontal bars) of the low-energy hydrogen-bonding families for Ac-Ala$_6$-Lys(H$^+$) (left) and Ac-β^2hAla$_6$-Lys(H$^+$) (right) obtained with the PBE+vdW functional and *tight* computational settings. The horizontal bars represent the lowest-energy representative of each family. All structures represent a local minimum on the PBE+vdW potential-energy surface (PES). In addition to the hierarchy of the PES, the energy hierarchy after adding corrections to the free energy at 300 K in the harmonic oscillator-rigid rotor approximation are shown ($F_{\mathrm{harm,rigid}}$). All energies are given with respect to the global minimum of the PES. Structural representations and labels of the families is shown, where the latter are numbered according to the free-energy hierarchy. The position of the structures is not directly related to the energy axis (y-axis), which is only connected to the horizontal bars. The data and geometries for Ac-Ala$_6$-Lys(H$^+$) are taken from Refs. [17, 28].

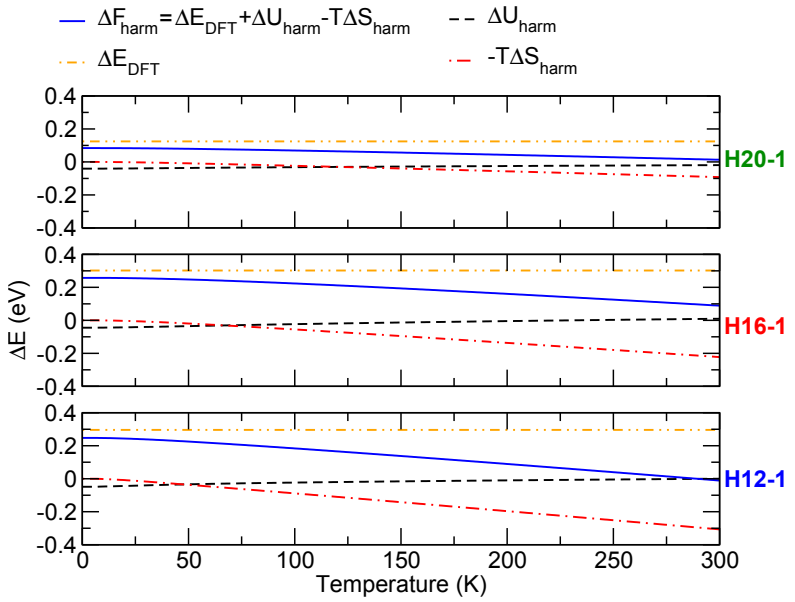


Figure 11.4: Contributions to the vibrational free energy as a function of temperature T for H20-1, H16-1, and H12-1. All energies are given relative to compact-1 (cf. Fig. 11.2). The harmonic free energy ΔF_{harm} as detailed in Eq. 11.1 is represented by a solid blue line, the PBE+vdW energy differences ΔE_{DFT} are plotted with a dot-dash-dot orange line, and the dashed black line shows the differences of the internal vibrational energy $\Delta U_{\mathrm{vib}}(T)$. The entropy contribution $-T\Delta S_{\mathrm{vib}}$ is plotted with a red dot-dash line.

we perform a similar analysis as in Ref. [28] for the peptide series Ac-Ala$_n$-LysH$^+$, $n = 4, \ldots, 8$. As a first step, we analyze the different contributions to the free energy in the harmonic oscillator approximation:

$$F_{\mathrm{harm}} = E_{\mathrm{DFT}} + \underbrace{F_{\mathrm{vib(harmonic\ oscillator)}}}_{U_{\mathrm{harm}}(T) - TS_{\mathrm{harm}}(T)} \quad . \tag{11.1}$$

It contains the PBE+vdW total energy E_{DFT}, the internal vibrational energy $U_{\mathrm{vib}}(T)$ (see Eq. 5.25), which includes the zero-point energy, and the entropy contribution $-TS_{\mathrm{vib}}(T)$. All these individual contributions are shown in Fig. 11.4 for H12-1, H20-1, and H16-1 as energy differences, taking compact-1 as the reference. A negative slope corresponds to a stabilization of the respective family relative to the compact family 2/compact-1. From Fig. 11.4 it follows that with increasing temperature all families H12-1, H16-1, and H20-1 are monotonically stabilized by vibrational free energy compared to compact-1. For H12-1, we observe a crossover of the lowest-energy structures at about 290 K. The dominant part of the stabilization of the helical conformations over the compact one is taken over by the entropy term $-T\Delta S_{\mathrm{vib}}$. The zero-point energy ($U_{\mathrm{vib}}(T)$ at $T = 0$ K) yields also a stabilization of the helical motifs, but only a minor contribution. This is the same result that has been observed for helical motifs for the peptide series Ac-Ala$_n$-LysH$^+$, $n = 4, \ldots, 8$ before[17, 28].

A further result of the analysis for Ac-Ala$_n$-LysH$^+$, $n = 4, \ldots, 8$ in Ref. [17, 28] is that this stabilization of helical motifs arises predominantly by the presence of lower-frequency vibrational modes. We can analyze this for Ac-β^2hAla$_6$-Lys(H$^+$) by considering the temperature-dependent contribution of each vibrational mode to the vibrational free energy at temperature T

$$Q_i = k_{\mathrm{B}}T \ln\left[1 - \exp\left(-\hbar\omega_i/(k_{\mathrm{B}}T)\right)\right] \quad , \tag{11.2}$$

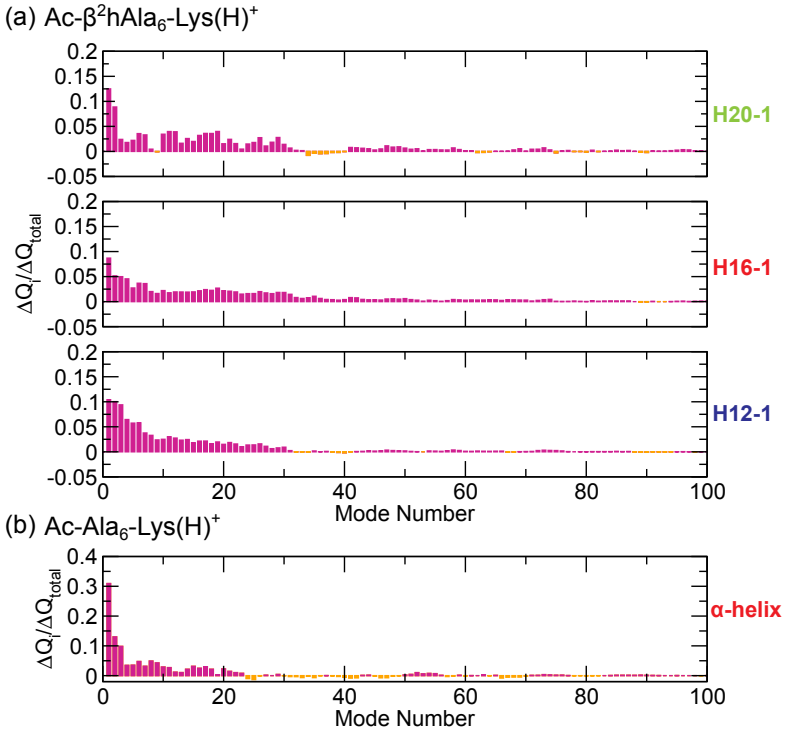


Figure 11.5: Temperature-dependent contribution of each individual vibrational mode to the vibrational free energy at 300 K as a function of the mode number for (a) Ac-β^2hAla$_6$-Lys(H$^+$) and (b) Ac-Ala$_6$-Lys(H$^+$). Q_i is defined in Eq. 11.2 and the difference Δ refers to compact-1 (family 2) for Ac-β^2hAla$_6$-Lys(H$^+$) and to the non-helical family 2 for Ac-Ala$_6$-Lys(H$^+$). Positive values are indicated by pink bars and negative values are indicated by orange bars. The maximum frequency plotted for Ac-β^2hAla$_6$-Lys(H$^+$) is similar for all families (around 620–630 cm^{-1}). For Ac-Ala$_6$-Lys(H$^+$) it is about 570 cm^{-1}. The data for Ac-Ala$_6$-Lys(H$^+$) are taken from Ref. [28].

with ω_i denoting the vibrational frequency of the mode number i (where the modes are sorted in order of increasing frequency). The temperature-dependent part of the free-energy difference is the sum over all individual contributions $\Delta Q_{\text{total}} = \sum_i \Delta Q_i$. In Fig. 11.5 we plot for each mode the ratio $\Delta Q_i / \Delta Q_{\text{total}}$ so that the sum over all individual contributions is 1. Pink bars denote a stabilization of the corresponding conformer, while orange bars denote a destabilization over the reference family. This is compact-1 for Ac-β^2hAla$_6$-Lys(H$^+$) (Fig. 11.5, subplot a) and the non-helical family 2 for Ac-Ala$_6$-Lys(H$^+$) (Fig. 11.5, subplot b). A stabilization arises if the frequency ω_i of the respective family given at the right side of the plot is lower than the corresponding frequency of the reference family. The largest part of the stabilization arises from the smallest mode numbers. This means that the helical families have in general lower vibrational frequencies than the more compact reference families. When comparing the plots for the H12-1, H16-1, and H20-1 families of Ac-β^2hAla$_6$-Lys(H$^+$) (Fig. 11.5, subplot a) to the picture for Ac-Ala$_6$-Lys(H$^+$) (Fig. 11.5, subplot b), the stabilization is delocalized over a larger amount of modes for all Ac-β^2hAla$_6$-Lys(H$^+$) families. This is also an effect that was observed for larger peptides such as Ac-Ala$_7$-LysH$^+$ and Ac-Ala$_8$-LysH$^+$ in Ref. [28], so it might be due to the increased length of the β-peptide Ac-β^2hAla$_6$-Lys(H$^+$) compared to the α-peptide Ac-Ala$_6$-Lys(H$^+$). Figure 11.5 implies that the lowest vibrational mode is already an indicator of the stabilization of the corresponding family with respect to more compact conformers (for the PBE+vdW functional). The lowest vibrational modes for the selected families of Ac-β^2hAla$_6$-

Table 11.2: Lowest vibrational modes for selected families of Ac-β^2hAla$_6$-Lys(H$^+$) and Ac-Ala$_6$-Lys(H$^+$) obtained with the PBE+vdW functional. The data for Ac-Ala$_6$-Lys(H$^+$) are taken from Ref. [28].

Ac-β^2hAla$_6$-Lys(H$^+$)	Lowest vib. mode (cm^{-1})
compact	30
H12-1	10
H16-1	17
H20-1	21

Ac-Ala$_6$-Lys(H$^+$)	Lowest vib. mode (cm^{-1})
non-helical	20
α-helical	8

Lys(H$^+$) and Ac-Ala$_6$-Lys(H$^+$) are given in Tab. 11.2. For Ac-β^2hAla$_6$-Lys(H$^+$) we find that the more elongated the structure is (H20 $\rightarrow$ H16 $\rightarrow$ H12), the smaller is its lowest vibrational mode. This reflects the trend of the amount of stabilization that the specific family experiences compared to compact-1.

For Ac-Ala$_6$-Lys(H$^+$) the lowest vibrational mode of the α-helix is at about 8 cm^{-1}, while the non-helical, more compact, family has its lowest vibrational mode at 20 cm^{-1}. When comparing the values for H12-1 [Ac-β^2hAla$_6$-Lys(H$^+$)] and the α-helix [Ac-Ala$_6$-Lys(H$^+$)] we find that the lowest vibrational modes are very similar, namely 10 cm^{-1} versus 8 cm^{-1}. However, the values for the more compact families differ by 10 cm^{-1}. While for the non-helical, more compact Ac-Ala$_6$-Lys(H$^+$) family it is 20 cm^{-1}, for the β-peptide family compact-1 it is 30 cm^{-1}. This suggests that the changes in the energy hierarchies when including vibrational corrections to the free energy are larger for H12-1 [Ac-β^2hAla$_6$-Lys(H$^+$)] than for the α-helix [Ac-Ala$_6$-Lys(H$^+$)] because the reference family chosen has higher vibrational modes.

11.3 Impact of different functionals and dispersion corrections

For the structure search of Ac-β^2hAla$_6$-Lys(H$^+$) the PBE+vdW functional was employed. In order to investigate how far we can push the PBE+vdW level of theory for the problem of increasing flexibility we will analyze the influence of different methods in this subsection. In contrast to the Ac-Lys-Ala$_{19}$ + H$^+$ case, we do not test different force fields here as we would need to find suitable parameters for the additional methylene groups in the backbone of the β-amino acids. This would distort the results in the first place already. For this reason, we here focus only on a comparison of the PBE and PBE0 exchange-correlation functionals coupled with the pairwise van der Waals (vdW) correction scheme ("+vdW") and the many-body scheme ("+MBD*) (see Section 3.5.3). Furthermore, we concentrate on the conformers found within 100 meV of compact-1 (the global PES minimum obtained with PBE+vdW) of the PBE+vdW free-energy hierarchy at 300 K. This is illustrated in the second column of Fig. 11.6, which is highlighted in gray. Just as in Section 8.6 for the Ac-Lys-Ala$_{19}$ + H$^+$ peptide, we relaxed the structures with the respective method leading to only marginal structural changes (root mean square deviation (RMSD) of less than 0.02 nm). The free-energy hierarchies obtained with the different functionals are displayed in Fig. 11.6. As before, we did not recompute the vibrational frequencies, but employed the results obtained with the PBE+vdW functional. The different hierarchies do not show a clear trend. As discussed in the preceding section, H12-1 is the most stable conformer obtained with PBE+vdW. Moving to PBE0+vdW, this stabilization is even increased with respect to the other conformers. However, for PBE+MBD* compact-1 is the most

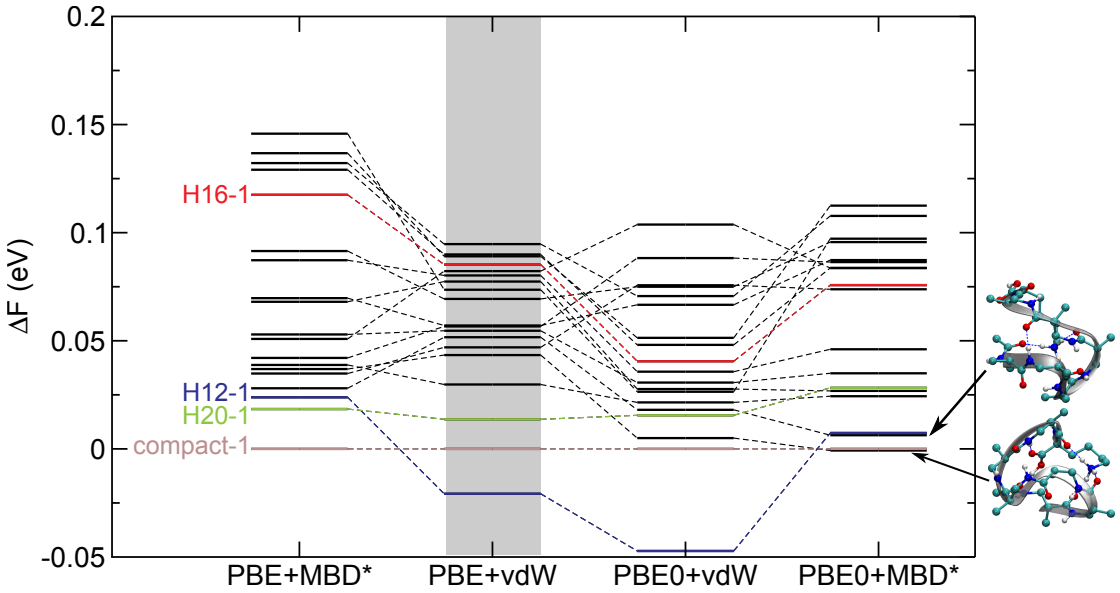


Figure 11.6: Free-energy hierarchies (horizontal bars) obtained with the PBE+MBD*, PBE+vdW, PBE0+vdW, and PBE+MBD* functional for Ac-β^2hAla$_6$-Lys(H$^+$) conformers. Specifically, we concentrated on the 19 conformers within 100 meV (free energy) of compact-1 obtained with PBE+vdW from the original search. The free energies are calculated based on the harmonic oscillator-rigid rotor approximation and a temperature of 300 K. Both the vibrational and the rotational contributions to the free energy are obtained from the PBE+vdW results and not recomputed at the other levels of theory. All energies are given relative to compact-1. The dashed lines serve as a guide to the eye and the PBE+vdW reference data from the original search are highlighted with a grey background.

stable conformer and for PBE0+MBD* it is the structure representative of family 5 (cf. Tab. 11.1 and Fig. 11.2), which is depicted next to the hierarchy plot in Fig. 11.6. We will critically assess the different structure predictions by a comparison to experimental fingerprints in the following chapter.

11.4 Summary

In this chapter, we characterized and summarized the outcome of our extensive first-principles conformational search using the PBE+vdW functional for Ac-β^2hAla$_6$-Lys(H$^+$), which was described in the previous chapter. The lowest-energy structure compact-1 is very similar to a β-hairpin in α-peptides. Including harmonic vibrational (and rigid-rotor rotational) corrections to the free energy at 300 K substantially changes the energy hierarchy, yielding the H12-helix (H12-1) as the most stable structure. We showed that just as in α-peptide helices[17, 28], the stabilization of helical families compared to more compact motifs by vibrational free energy can be retraced to overall lower vibrational modes.

We also tested the influence of different exchange-correlation functionals on the free-energy hierarchy (PBE+vdW, PBE+MBD*, PBE0+vdW, PBE0+MBD*). However, the results do not show a clear trend, but different conformers are predicted to be the most stable one. We will assess the structure predictions in the following chapter by comparing to experimental ion mobility-mass spectrometry (IM-MS) data and infrared multiphoton dissociation (IRMPD) spectra.

12 CONNECTING TO EXPERIMENT

In this chapter, we will perform a comparison of the first-principles structure predictions for both Ac-Ala$_6$-Lys(H$^+$) and Ac-β^2hAla$_6$-Lys(H$^+$) with experimental data. For this, we have both ion mobility-mass spectrometry (IM-MS) measurements and infrared multiphoton dissociation (IRMPD) spectra available. All measurements were performed by Stephan Warnke, Gert von Helden, and Kevin Pagel working in the Molecular Physics department of the Fritz Haber Institute.

The infrared (IR) spectra presented in this chapter are derived from molecular dynamics (MD) simulations in the *NVE* ensemble using the PBE+vdW functional and *tight* computational settings. Subsequently they were convoluted with a Gaussian function with a variable width of $\sigma = 0.5\,\%$ of the wavenumber in order to account for broadening effects in experiment. The simulation length was 25 ps in all cases and performed with a time step of 0.75 fs.[1] Initially, the molecules were equilibrated at 300 K by performing thermostated runs at 300 K for at least 5 ps.

12.1 ION MOBILITY-MASS SPECTROMETRY

We will first concentrate on the experimental IM-MS data. The analysis given here was performed by Stephan Warnke who measured the arrival time distributions (ATDs) for both Ac-Ala$_6$-Lys(H$^+$) and Ac-β^2hAla$_6$-Lys(H$^+$). If the molecule in the IM-MS set-up featured only one conformation, the width of the peak in the ATD of the IM-MS experiment would be determined by diffusion (and the width of the initial pulse). In this case, the peak shape $\Phi(t)$ can be written as[363]:

$$\begin{aligned}\Phi(t) &= \int \mathrm{d}t' \left\{ \frac{C}{\sqrt{D(t-t')}} \left(v_\mathrm{d} + \frac{L}{(t-t')} \right) \right. \\ &\left. \exp\left[\frac{-(L - v_\mathrm{d}(t-t'))^2}{4D(t-t')} \right] P(t') \right\} \quad , \end{aligned} \tag{12.1}$$

with C being a constant and D the diffusion coefficient. L is the length of the drift tube and v_d describes the average drift velocity. $P(t')$ describes the shape of the ion cloud when it enters the drift tube at whose end the ATD is recorded. It is a rectangle pulse with a length of 100 µs. The measured ATDs for Ac-Ala$_6$-Lys(H$^+$) and Ac-β^2hAla$_6$-Lys(H$^+$) are given in Fig. 12.1 A and B, respectively (blue lines). The theoretical curves calculated based on Eq. 12.1 are given by the red lines. The measured widths are only slightly larger than the ones obtained with Eq. 12.1. This suggests that there is either (i) only one single conformer type present in experiment, or

[1] Such a simulation for Ac-β^2hAla$_6$-Lys(H$^+$) took about 22 days on 256 cores of the "aims" cluster (Intel Xeon octacore nodes) at the Garching Computing Centre.

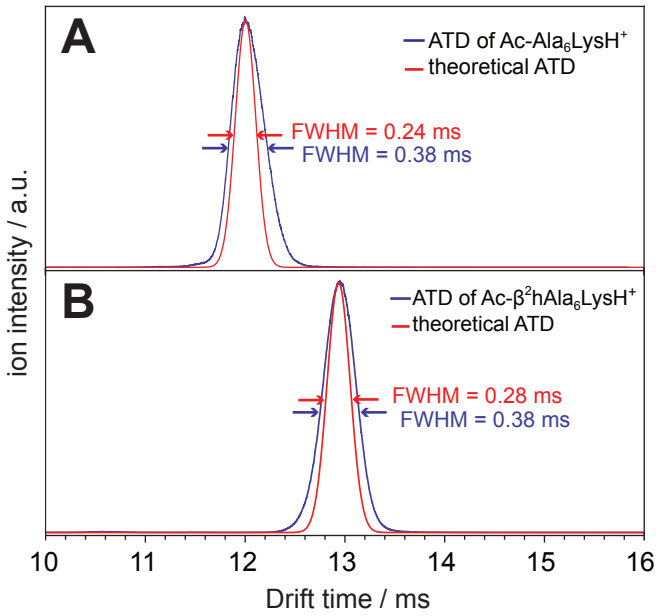


Figure 12.1: Measured arrival time distributions (ATDs) (blue lines) from drift tube IM-MS measurements and theoretical curves (red lines) according to Eq. 12.1 for (A) Ac-Ala_6-Lys(H^+) and (B) Ac-β^2hAla_6-Lys(H^+). Courtesy of Stephan Warnke.

(ii) an ensemble of (distinctly) different conformers that have essentially all the same collision cross section (CCS), or (iii) a rapid interconversion between different conformers so that all ions migrate through the drift tube with the same time-averaged CCS[413].

When considering the experimental ATD peak positions for Ac-Ala_6-Lys(H^+) versus Ac-β^2hAla_6-Lys(H^+), we see that the peak for Ac-β^2hAla_6-Lys(H^+) is located at larger drift times (CCS) than Ac-Ala_6-Lys(H^+), which is consistent with the backbone extension of Ac-β^2hAla_6-Lys(H^+).

12.2 Infrared multiphoton dissociation (IRMPD) spectra

We now turn to a general assessment of the IR spectra of the considered system followed by an analysis of the experimental IRMPD spectra. As explained in Chapter 5, each normal mode, i.e., each peak in the harmonic IR spectrum, can be assigned to a particular vibration of parts of the molecule. In order to see how the additional methylene groups (CH_2) of β-peptides show in the IR spectrum, we performed a detailed normal-mode analysis of the H12-helical Ac-β^2hAla_6-Lys(H^+) family H12-1 between 1000 and 1800 cm^{-1} depicted in Fig. 12.2. The vibrations associated with the additional CH_2 groups mix with the vibrations of the CH_3 and CH groups (at the Cα carbon), which occur approximately between 1000 and 1450 cm^{-1}. The C(=O)O(H) group at the C-terminus of H12-1 is not involved in a hydrogen bond, but dangling and the corresponding C=O stretching mode lies at 1758 cm^{-1}. As was discussed in more detail in Section 5.2, stretching modes appear at smaller wavenumbers if the atom involved is hydrogen bonded, since the H-bond reduces the restoring force. This is the case for H20-1 (cf. Fig. 11.3), where the C=O stretching mode of the C-terminus is indeed shifted to lower wavenumbers. In

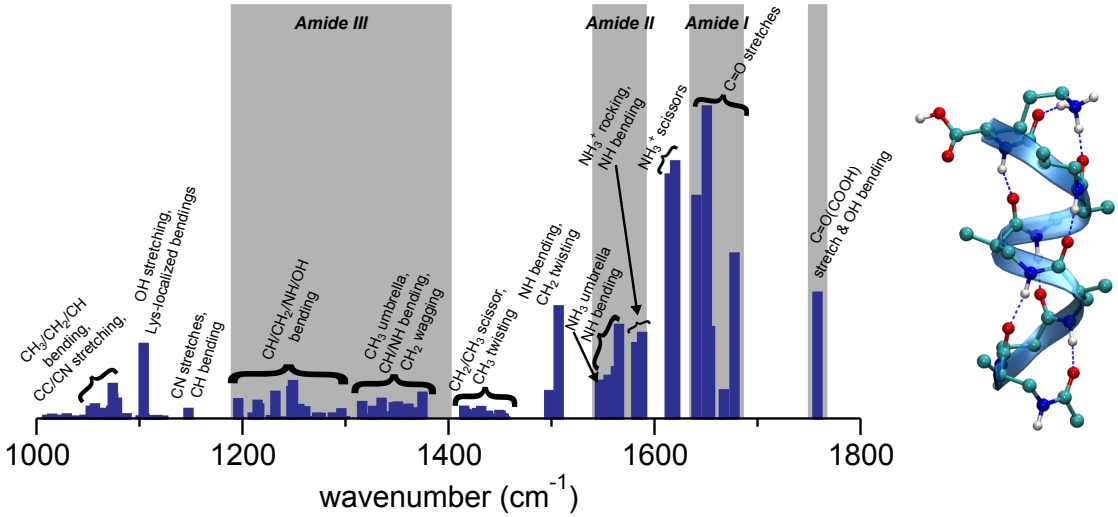


Figure 12.2: Detailed normal-mode analysis (left) of the harmonic IR spectrum of the Ac-β^2hAla$_6$-Lys(H$^+$) family H12-1 (right) between 1000 and 1800 cm^{-1}. The amide I, II, and III bands and the vibration localized at the non-hydrogen bonded C-terminus are highlighted with a gray background.

fact, for H20-1 it is shifted to 1677 cm^{-1} and mixes with the amide-I band (not explicitly shown here).

The experimental IR spectra of Ac-Ala$_6$-Lys(H$^+$) and Ac-β^2hAla$_6$-Lys(H$^+$) are depicted in Fig. 12.3. At first glance they look very similar, especially the amide I ($\sim$1670 cm^{-1}) and amide II ($\sim$1520 cm^{-1}) bands, which are most sensitive to the secondary structure (see discussion in Section 5.2). However, the peak positions in the region between 1100 cm^{-1} and 1400 cm^{-1} are very different. The intensity in this region is much smoother for Ac-β^2hAla$_6$-Lys(H$^+$) suggesting that there are more modes in this regime than for Ac-Ala$_6$-Lys(H$^+$). This agrees with the harmonic normal-mode analysis, which showed that the peak positions arising from the additional CH$_2$ groups in the Ac-β^2hAla$_6$-Lys(H$^+$) backbone occur in this wavenumber region. Furthermore, we see that in both experimental spectra there is a small peak at the right side of the amide I band, indicating a dangling C(=O)O(H) group at the C-terminus.

As the R-factor is very sensitive to small kinks and the experimental spectra are very wiggly beyond 1720 cm^{-1} [especially the one for Ac-Ala$_6$-Lys(H$^+$)], we only take into account the wavenumber region between 1100 and 1720 cm^{-1} for the computation of R-factors in the following. However, this means that the peak indicating a dangling C(=O)O(H) group is not considered in the R-factor.

The experimental spectra were smoothed using the same procedure used and explained in Chapter 9. A comparison of the raw experimental data to the smoothed spectra can be found in the Appendix B.3.

12.3 α-Peptide Ac-Ala$_6$-Lys(H$^+$)

We will start our analysis with Ac-Ala$_6$-Lys(H$^+$). For this, we concentrate on the α-helix and the non-helical conformer depicted on the right side of Fig. 12.4. These are the two most stable structures at 300 K predicted by PBE+vdW as discussed in Chapter 11. Figure 12.4 illustrates the experimental IM-MS data in comparison with the CCSs for both families calculated with

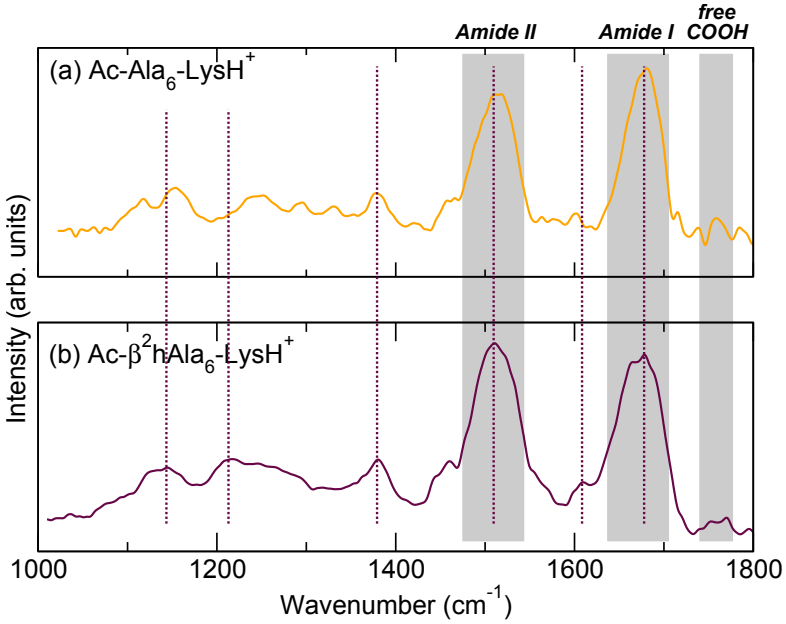


Figure 12.3: Experimental (smoothed) IRMPD spectra for (a) Ac-Ala_6-Lys(H^+) and (b) Ac-β^2hAla_6-Lys(H^+). Dotted lines serve as a guide to the eye to illustrate the peak positions of Ac-β^2hAla_6-Lys(H^+).

Table 12.1: Comparison of measured collision cross sections (CCSs) and calculated CCSs based on the projection approximation (PA)[364] and the trajectory method (TJM)[366–368]. As both PA and TJM rely on random sampling, repeated runs can lead to slightly different results. For PA we used an accuracy of 0.2%. For the TJM we used Hirshfeld charges[237] of the PBE densities and 500,000 trajectories per single structure leading to standard deviations of about 1%.

	Ac-β^2hAla_6-Lys(H^+)				Ac-Ala_6-Lys(H^+)	
	H12	compact	H20	H16	α-helix	non-helical
PA (Å^2)	203	183	182	191	180	172
TM (Å^2)	204	182	182	193	181	171
Exp. (Å^2)		190			180	

the projection approximation (PA) method. The values obtained with the trajectory method (TJM)[366–368] approach are essentially the same (see Tab. 12.1). The theoretical value for the α-helix perfectly matches the experimental peak, while the non-helical family has a smaller CCS. Due to the perfect match of the α-helical CCS of Ac-Ala_6-Lys(H^+) with experiment and since the larger peptides of the series Ac-Ala_n-Lys(H^+) are firmly confirmed to be α-helical[17, 25, 26, 28, 338], we here conclude that most likely there is mainly the α-helix present in experiment. This would also be in agreement with the free-energy hierarchy at 300 K obtained with PBE+vdW. Moreover, as discussed in Section 12.1, the narrow experimental peak width in the ATD would be in accord with the presence of only one structure as well.

Apart from the CCSs, also the comparison of the measured and calculated (anharmonic) IR spectra points to the α-helix, as illustrated in Fig. 12.5. Both the visual impression and also the comparison based on the R-factor gives a good match with experiment – and a better one than the non-helical family.

As a summary, we can thus state that for the peptide Ac-Ala_6-Lys(H^+) both the PBE+vdW (free) energy hierarchy, the IM-MS measurements, and the IRMPD spectra agree on the same result, namely that the α-helix should be the dominant conformer at room temperature.

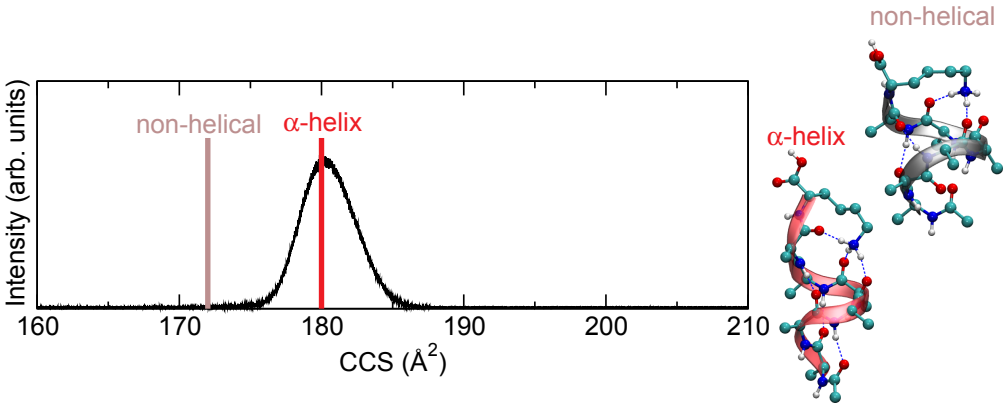


Figure 12.4: Experimental (black curve) and calculated collision cross sections (CCSs) for the non-helical and the α-helical family of Ac-Ala$_6$-Lys(H$^+$). The bars indicate the CCSs calculated using the projection approximation (PA) for the relaxed PBE+vdW structures.

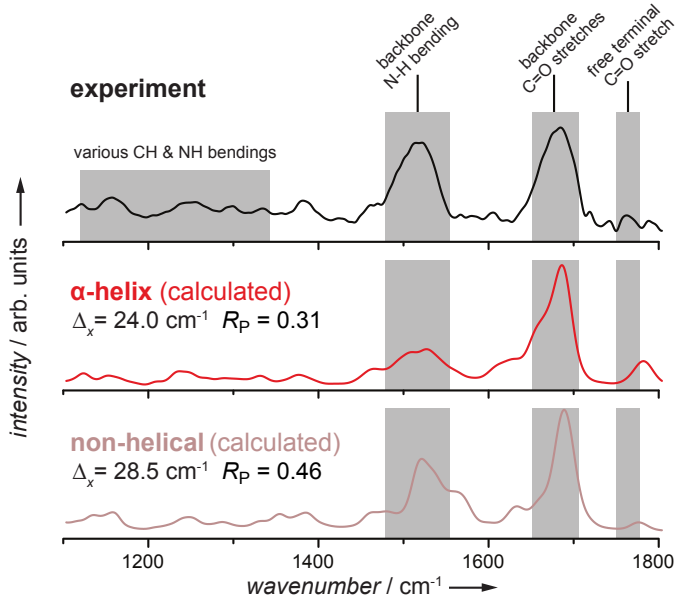


Figure 12.5: Experimental IRMPD spectra (black, smoothed) for Ac-Ala$_6$-Lys(H$^+$) compared to the calculated IR spectra for the non-helical family and the α-helix. The IR spectra are derived from MD simulations with $\langle T \rangle = 300\,\mathrm{K}$ and thus include anharmonic effects. They are based on the PBE+vdW functional and *tight* computational settings with a time step of 0.75 fs. The Pendry reliability factor together with the rigid shift Δ_x are given. The theoretical spectra are shifted accordingly. The rigid shifts Δ_y are not included here (they are listed in Appendix B.4).

12.4 β-PEPTIDE AC-β^2HALA$_6$-LYS(H$^+$)

We will now perform an analogous analysis as above for Ac-β^2hAla$_6$-Lys(H$^+$). For this, we will first concentrate on H12-1, compact-1, and H20-1, which were, in the given order, the most probable structure candidates at 300 K suggested by PBE+vdW and the harmonic oscillator-rigid rotor approximation. Additionally, we will also consider H16-1, which is the equivalent helix to the α-helix with ΔF about 106 meV compared to H12-1 (PBE+vdW). The corresponding structure predictions are again illustrated in Fig. 12.6. For all these structures we calculated the CCS using the PA method and the TJM approach (see Tab. 12.1). The two methods yield essentially the same results for this system size (Tab. 12.1). As the PA method is computationally cheaper, calculations for further structures (see below) were performed with the PA method.

The vertical bars in Fig. 12.6 give the theoretical CCSs obtained with the PA method for the hydrogen-bonding families compact-1, H12-1, H16-1, and H20-1. In agreement with the expectation from a visual impression of the structures, H20-1 and compact-1 have similar CCSs, while the CCS of the more elongated family H16-1 is larger and H12-1, which is the most extended family, has the largest CCS. Inaccuracies in the model to calculate the CCSs (see Section 7.1.1.1) could induce shifts against the "true" values so that the real structure family would have a calculated CCS differing from the measured one. However, as for Ac-Ala$_6$-Lys(H$^+$) (90 atoms) the calculated CCS for the α-helix agrees very well with the experimental peak, we expect the model to give also reasonable results for the similar-sized peptide Ac-β^2hAla$_6$-Lys(H$^+$) (108 atoms). While both compact-1 and H20-1 have lower CCSs than the experimental value and H12-1 has a much larger value, the CCS calculated for H16-1 matches the experiment very well.

We complement the IM-MS measurements by experimental IRMPD fingerprints as shown in Fig. 12.7. For a comparison, we again calculated vibrational IR spectra from MD simulations, which thus include anharmonic effects (in the classical-nuclei approximation). The results for H12-1, H16-1, H20-1, and compact-1 are likewise displayed in Fig. 12.7. Both from the visual impression and the R-factors, compact-1 and H12-1 do not match the experimental spectrum at all. This coincides with the fact that the calculated CCSs did not match the experiment, either. On the other hand, the spectra for H20-1 and H16-1 give a rather similar match with experiment. However, compared to the α-helix for Ac-Ala$_6$-Lys(H$^+$), the R-factors are worse. The C-terminal C(=O)O(H) group of H20-1 is involved in hydrogen bonds (cf. Fig. 12.6) and in fact does not show the characteristic peak of a dangling C(=O)O(H) as indicated to be present in experiment. However, H16-1 has a free C-terminus and shows this peak. Indeed, H12-1 also has this peak, but is however already ruled out by the overall comparison of the spectra and the CCS criterion. Hence, for this selection of conformational types, the IM-MS data and the IR spectra would be most consistent with a H16-helix (H16-1).

12.4.1 DISCUSSION

So far, we have only assessed four selected hydrogen-bonding families of Ac-β^2hAla$_6$-Lys(H$^+$). We have to keep in mind, however, that the low free-energy regime is densely populated and also other conformations might be important. For this reason, we also calculated the CCSs for all 163 hydrogen-bonding families of Ac-β^2hAla$_6$-Lys(H$^+$), for which we had calculated the free energy at 300 K. Additionally, we also calculated the R-factor of the corresponding convoluted harmonic spectrum with experiment. The (harmonic) R-factor versus the difference between the

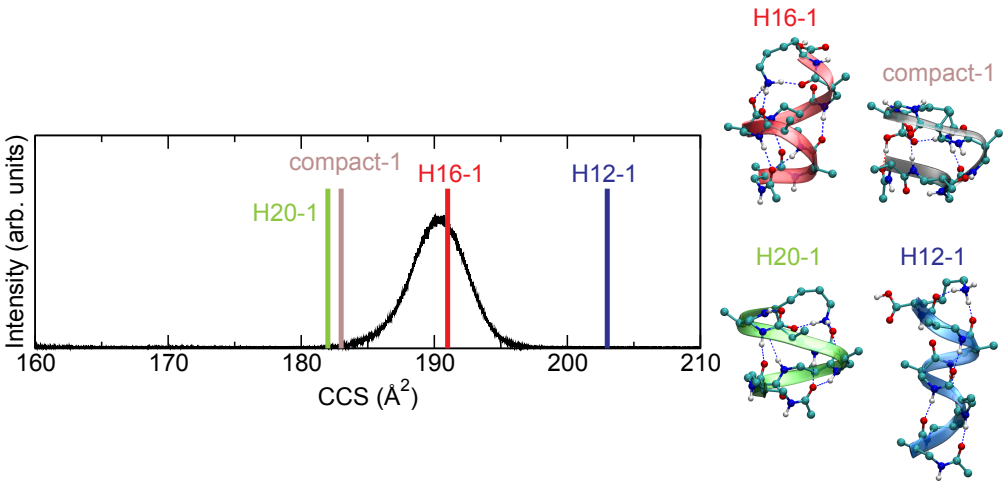


Figure 12.6: Experimental (black curve) and calculated collision cross sections (CCSs) for the selected families of Ac-β^2hAla$_6$-Lys(H$^+$). The bars indicate the CCSs calculated using the projection approximation (PA) for the relaxed PBE+vdW structures.

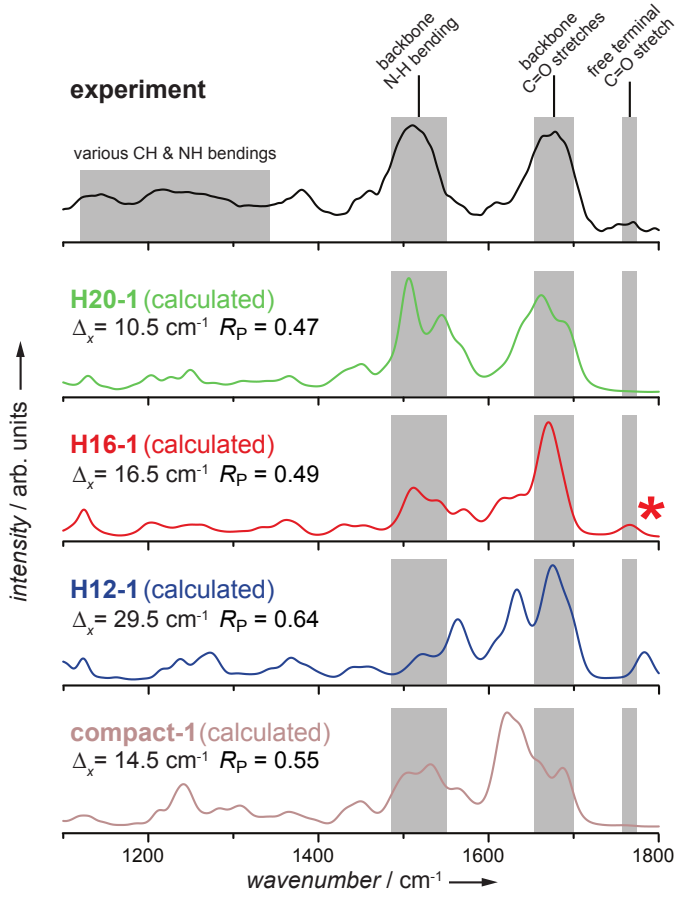


Figure 12.7: Experimental IRMPD spectra (black, smoothed) for Ac-β^2hAla$_6$-Lys(H$^+$) compared to the calculated IR spectra for H12-1, H16-1, H20-1, and compact-1. The IR spectra are derived from MD simulations with $\langle T \rangle = 300\,\mathrm{K}$ and thus include anharmonic effects. They are based on the PBE+vdW functional and *tight* computational settings with a time step of 0.75 fs. The Pendry reliability factor together with the rigid shift Δ_x are given. The theoretical spectra are shifted accordingly. The rigid shifts Δ_y are not included here (they are listed in Appendix B.4). The red star marks the peak associated with the dangling C=O stretch of the C-terminus.

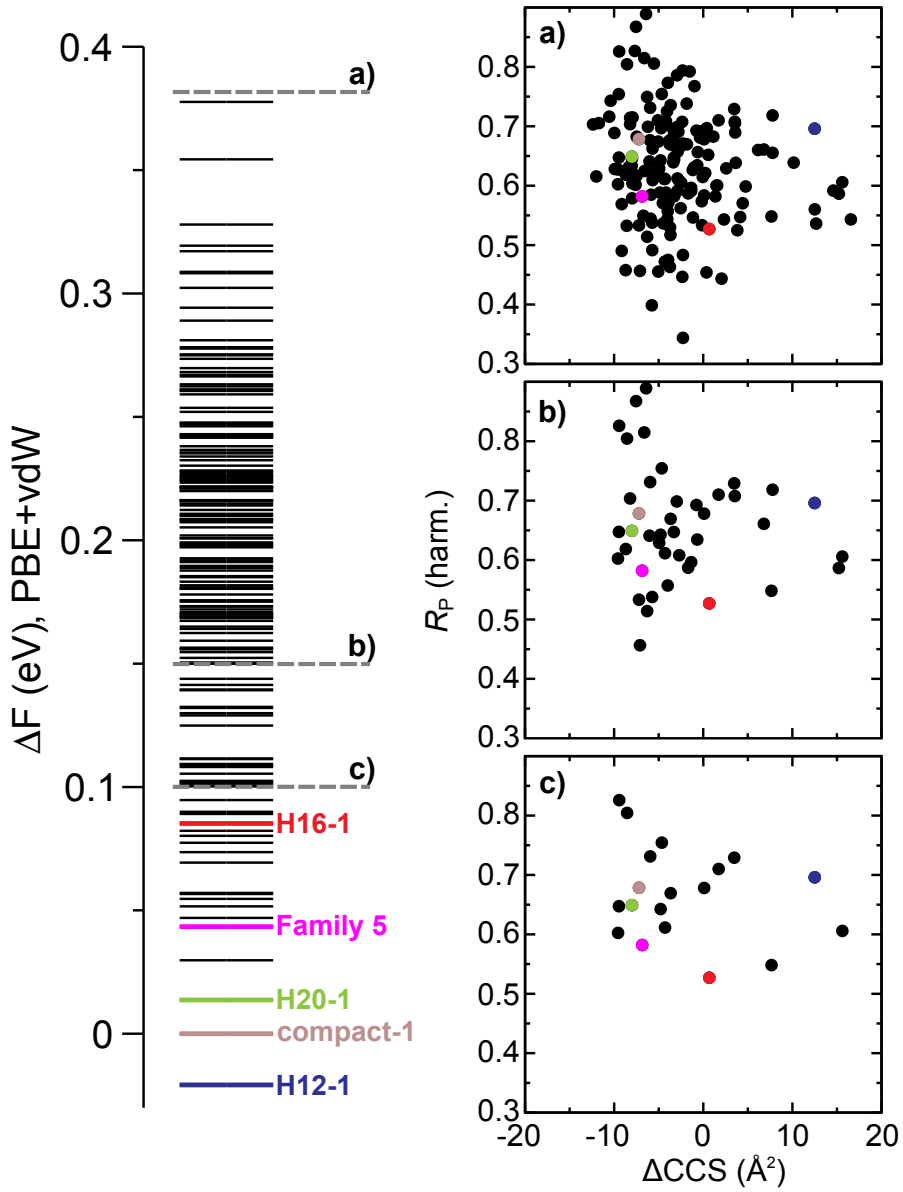


Figure 12.8: PBE+vdW free-energy hierarchy (300 K, harmonic oscillator-rigid rotor approximation) obtained for Ac-β^2hAla$_6$-Lys(H$^+$) (see Chapter 11). The plots show the harmonic R-factor of the conformers (y-axis) against the deviation of the calculated CCS from the measured one (x-axis). The plots comprise all structures up to a given free-energy threshold indicated by the dashed grey line in the hierarchy. H12-1, compact-1, H20-1, H16-1, and Family 5 are highlighted in color.

measured and the computed CCS for all structures in different free-energy regimes are shown in Fig. 12.8. H12-1, compact-1, H20-1, and H16-1 are highlighted in color. The subplot c) in Fig. 12.8 comprises all structures, for which we also re-calculated the free-energy hierarchies with different functionals in Section 11.3. The structure predicted to be most stable at 300 K by PBE0+MBD* (Family 5, see Section 11.3) is highlighted in color as well. The R-factor of the harmonic and the anharmonic spectrum for one structure type can differ quite significantly. The harmonic R-factor for H20-1, e.g., is significantly worse than the harmonic R-factor for H16-1, while the R-factors are very similar when deriving the IR spectrum from MD simulations (see Fig. 12.7). For this reason, we take the harmonic R-factor rather as a consistency criterion and rely more on the CCS comparison. We here note again that the deviation between the experimental CCS and the calculated value for the "correct" structure can depend on the size of the system[365]. On the other hand, for the calculated CCS for the "correct" Ac-β^2hAla$_6$-Lys(H$^+$) structure no deviation from the experimental value is expected due to the excellent agreement for Ac-Ala$_6$-Lys(H$^+$) (see Section 12.3). When taking all 163 conformational types into account, many structures have calculated CCSs matching well the experimental one (Fig. 12.8a), i.e., are potential structure candidates. When only considering the lowest 100 meV regime of compact-1 (19 structures), H16-1 emerges as the structure with the best (harmonic) R-factor. It yields furthermore a perfect match with respect to the CCS as already observed in the previous section (12.4). However, there are other structures that yield a similarly good match. Family 5, though, the structure that was predicted by PBE0+MBD* does not have a CCS in agreement with experiment and thus does rather not represent a possible structure candidate.

As discussed in Section 12.1, the narrow peak width of the IM-MS experiments is consistent with different scenarios. One scenario would be an ensemble of structures that have essentially the same CCSs. In fact, we find many conformers that yield a good match with the experimental CCS. However, there are also many (low-energy) conformers that do not yield a good match with the experimental CCS. We thus consider it rather unlikely that, of all the others, only those conformers co-exist that happen to have the same CCS. Another possibility for the narrow peak width is an ensemble of rapidly interconverting conformations (with respect to the drift time scale [$\sim$10 ms]). The most probable interconversion would happen between the different helix types H12-1, H16-1, and H20-1 as for this only the hydrogen bonds have to be shifted from one residue to the next. We also see such conversions happen during our *ab initio* replica-exchange molecular dynamics (REMD) simulations (cf. Fig. 10.20). A rapid interconversion between H12-1, H16-1, and H20-1 would yield an average CCS, which is in agreement with the experimental value. However, mixing the (anharmonic) spectra for H12-1, H16-1, and H20-1 yields the best agreement with the experimental spectrum (based on the R-factor) for 0% H12-1, 50% H16-1, and 50% H20-1 ($R_\mathrm{P} = 0.42$). This rather contradicts a mixture of H12-, H16-, and H20-helices, where the contribution of H12-1 would be needed to obtain an average CCS consistent with the experimental CCS. However, we cannot exclude that there is a rapid interconversion between other (different) conformers. Another explanation for the narrow peak width in the IM-MS measurements is the presence of only one conformer type. H16-1 yields a very good match of the calculated CCS with experiment and the calculated (anharmonic) IR spectrum yields also a possible match with the experimental spectrum. The H16 helix, H16-1, is thus a possible match to both experiments and (within reason) to all tested functionals (PBE+vdW, PBE+MBD*, PBE0+vdW and PBE0+MBD*). We point out that the agreement is within reason, because

neither method predicts H16-1 to be the most probable structure at 300 K. The agreement would be "within reason" as it is within the expected error bars of our methods – resolving (free) energy differences of less than 1 meV per atom is a challenge for all currently available density-functional theory (DFT) exchange-correlation functionals. Moreover, none of these functionals predicts any structure to be the lowest in free energy that yields a CCS or an IR spectrum that is in agreement with the experiments. PBE+MBD* predicts compact-1, PBE+vdW H12-1, PBE0+vdW also H12-1, and PBE0+MBD* predicts Family 5 and compact-1 to be equally stable. None of these conformers have a CCS that is close to the experimental value and the matches of the (anharmonic) IR spectra for compact-1 and H12-1 with experiment are very poor. On the other hand, there could be also a conformer apart from the H16-helix that yields similarly good agreement with both experiments, which, however, could not be inspected more closely due to the huge amount of structures that were found. Despite the functional accuracy, we are also limited by the harmonic approximation to the free energy here. The changes in the hierarchy that the vibrational contribution to the free energy induces are huge, namely more than 300 meV of H12-1 with respect to compact-1, meaning that also its errors will have a significant impact. In summary, we have clearly reached the limits of DFT in combination with the harmonic approximation to the free energy here.

13 Conclusions and outlook

Understanding the physical mechanism that connects the amino-acid sequence of a peptide or protein to its structure is an ongoing and interdisciplinary challenge. In the present work an unprecedented first-principles structure-screening effort has been performed, yielding important insights into the capabilities and limitations of first-principles methods for predicting peptide structure. Two systems were chosen that present a challenge with respect to length and flexibility, respectively. The first one was a 20-residue peptide system, *large* enough to partially form a tertiary structure (Ac-Ala_{19}-Lys + H^+ versus Ac-Lys-Ala_{19} + H^+). The second system was the β-peptide Ac-β^2hAla_6-Lys(H^+), which features increased backbone *flexibility* due to its backbone elongation. This leads to an even more complex conformational space compared to natural α-peptides. In order to critically assess structure predictions by different density-functional theory (DFT) exchange-correlation functionals, a comparison to gas-phase ion mobility-mass spectrometry (IM-MS) measurements and gas-phase infrared multiphoton dissociation (IRMPD) spectra was performed.

One main objective of this work was to develop an efficient and accurate search strategy to sample the huge conformational space of the respective peptides. As a first step, global sampling of the conformational space was achieved by performing replica-exchange molecular dynamics (REMD) simulations using the OPLSAA force field. Then, snapshots of the REMD trajectories were clustered and thousands of structure representatives were relaxed with DFT using the PBE[15] functional coupled with a pairwise correction for the long-range tail of the van der Waals (vdW) dispersion interactions[16] (PBE+vdW). For the lowest-energy structure candidates, *ab initio* REMD simulations using the PBE+vdW functional were performed to refine the local structural environment. It has been shown that such *ab initio* REMD simulations can indeed lead to rearrangements of the hydrogen-bonding network and yield lower-energy structures.

It is well established that the peptide series Ac-Ala_n-Lys(H^+) ($n \simeq$ 6-19) forms α-helices in the gas phase[17, 25–28, 338]. The α-helical conformation is stabilized because the protonated lysine caps the dangling carbonyl oxygens close to the C-terminus and the charge interacts electrostatically favorably with the helix dipole. In fact, the structure search employed in this work exclusively found α-helical conformations for Ac-Ala_{19}-Lys(H^+), which only deviate close to the termini. Ac-Ala_{19}-Lys(H^+) could thus be identified as a structure seeker. For Ac-Lys(H^+)-Ala_{19} the protonated lysine residue is located at the N-terminus, which would consequently destabilize a helical conformation. Indeed, the conformations that were identified here are more compact than a straight helix, but still contain helical segments[27, 410]. The PBE+vdW functional predicts a conformational ensemble with several conformers being relatively close in energy, but different in three-dimensional structure. Six conformational types could be identified in the lowest-energy regime (170 meV). In order to critically assess the PBE+vdW prediction,

targeted calculations for those six conformers were performed with different functionals. For this, the PBE and the hybrid PBE0 functional coupled with a pairwise vdW[16] correction scheme (PBE+vdW, PBE0+vdW) and a many-body scheme[238, 239] (PBE+MBD*, PBE0+MBD*) were chosen. PBE0+MBD* emerged as the most reliable functional in a recent benchmark for the similar alanine-based peptide Ac-Phe-Ala_5-Lys(H^+)[247]. For Ac-Lys(H^+)-Ala_{19}, PBE0+MBD* changes the energy hierarchy given by PBE+vdW and predicts one single conformer to be dominant. In order to connect to experimental data, collision cross sections (CCSs) were calculated and infrared (IR) spectra were derived from molecular dynamics (MD) simulations using the PBE+vdW functional. The latter account for conformational flexibility and anharmonicities within the classical-nuclei approximation. In fact, the PBE+vdW predictions are not in agreement with the experimental results, while the conformer (denoted as C2) predicted by PBE0+MBD* matches both experiments. This conformer contains an α-helical and a 3_{10}-helical part, which are connected by a turn. The lysine side chain is bent to interact with the negative end of the helix dipole of the α-helical segment. This means that alanine, which is a known helix former, still tries to retain its α-helical preference even if the lysine is located at the nominally wrong end for a single helix.

The IM-MS experiments not only predict the existence of mainly compact conformers, but also a small amount of (most likely) helical conformers, where the proton is most probably located close to the C-terminus. A structure search for low-energy conformations for this type requires the explicit freedom for the proton to hop between the residues. This cannot be accomplished by the OPLSAA force field, and neither by the majority of other force fields, as they do not allow for bond breaking. Apart from that, it is not clear what force-field parameters should be used for the description of the proton sitting close to the C-terminus. Hence, a purely first-principles (PBE+vdW) structure search was employed using REMD. It could be shown that structures were found that are significantly lower in energy (370 meV) than the structures used to initialize the search. The lowest-energy structure obtained adopts a reasonable conformation with the lysine chain bent to interact with the acetyl group and the proton being located at a position where the C-terminal carbonyl group can interact with it. However, for all of the tested functionals (including PBE0+MBD*) the helix has a rather high energy, which would indicate that it should not be present to a measurable extent in experiment, even though it is. This might be a problem of the functional – although we consider this possibility rather remote given the large energy difference that is predicted. Another reason for the discrepancy could be the search technique employed, i.e., with a longer simulation run an even lower-energy structure might have been found. On the other hand, Jarrold and co-workers[27] suggest that the helices most likely originate from dissociation of dimers. Thus, if the barrier is high enough, they might be trapped in this local minimum.

In contrast to natural α-peptides, β-peptides feature one additional methylene group per residue in the backbone, yielding increased flexibility and thus an even more complex conformational space. Within this work, it was investigated how far the limits of PBE+vdW can be pushed for such increased flexibility. The design principle of the helical series "Ac-Ala_n-Lys(H^+)" was used as a template in order to derive a β-peptide that might have a helical conformational preference in the gas phase as well. For this, the alanine residues were exchanged for the equivalent β-amino-acid residues and Ac-β^2hAla_6-Lys(H^+) was chosen as a test case. In order to investigate the performance of the REMD based conformational search strategy, the outcome

was compared to a second, independent, search approach. It consisted of a basin-hopping search with the (augmented) OPLSAA force field and further similar basin-hopping searches, where the structure was constrained to a helix. The force-field structures were then clustered to identify the most important structure types, followed by relaxation of thousands of cluster representatives with DFT (PBE+vdW). Both sampling approaches yielded a relatively similar performance and led to about 15,000 PBE+vdW structures in total.

Although the design principle of Ac-β^2hAla$_6$-Lys(H$^+$) was derived from the Ac-Ala$_n$-Lys(H$^+$) series, which exhibits a helical preference in the gas phase, no helical structure was found in the low-energy regime (PBE+vdW, $\sim$100 meV), but only very compact structures. In order to connect to the experimental conditions (room temperature), vibrational and rotational free energies in the harmonic oscillator-rigid rotor approximation were calculated. This favors helices dramatically over the more compact structures, so that the H12-helix becomes the most stable structure at 300 K in PBE+vdW. As already found for α-peptides[17, 28, 413, 439], this stabilization originates from the fact that helices have much softer vibrational modes. In a direct comparison of the free-energy hierarchies of Ac-β^2hAla$_6$-Lys(H$^+$) versus its natural α-peptidic analogue Ac-Ala$_6$-Lys(H$^+$), the stabilization of helices compared to the lowest-energy structure is more pronounced. This is due to the reference conformer, which has higher vibrational frequencies for Ac-β^2hAla$_6$-Lys(H$^+$) than for Ac-Ala$_6$-Lys(H$^+$). Moreover, the low free-energy regime of Ac-β^2hAla$_6$-Lys(H$^+$) is much more densely populated than for Ac-Ala$_6$-Lys(H$^+$). This is in agreement with the expectation due to the more flexible backbone of Ac-β^2hAla$_6$-Lys(H$^+$). For Ac-Ala$_6$-Lys(H$^+$), PBE+vdW in combination with the harmonic oscillator-rigid rotor approximation, predicts the α-helix to be the most probable structure at room temperature (300 K)[28]. Both the calculated CCS and the (anharmonic) IR spectrum yield a good match with the experimental data, i.e., both the PBE+vdW prediction and the two experiments are in line and point to the α-helix. The situation is different for Ac-β^2hAla$_6$-Lys(H$^+$). Here, PBE+vdW in combination with the harmonic oscillator-rigid rotor approximation to the free energy predicts the H12-helix to be the most probable structure at room temperature followed by a very compact structure and a H20-helix. However, both the H12-helix and the compact structure do not match either experiment. The H20-helix has a calculated CCS that is different from the experimental value as well, but the (anharmonic) IR spectrum yields a possible match to experiment. None of the functionals tested (PBE+vdW, PBE+MBD*, PBE0+vdW, PBE0+MBD*) predicts a most-stable conformer that is compatible with both experimental results. The H16-helix yields a possible match to both experiments. However, it is not the lowest free-energy conformer in either method, but it is found within a reasonable (free) energy range (100 meV of the lowest free-energy conformer). The H16-helix is a unique structure as it is the equivalent to the α-helix in natural peptides and it has not been reliably predicted or experimentally found before. However, it cannot be excluded that there are other conformers that yield a similarly good agreement with both experiments. Here, the limits of the method are eventually reached, most likely due to the approximation to the vibrational free energy, which induces dramatic changes between the total-energy hierarchy and the free-energy hierarchy at 300 K.

In summary, the work of this thesis represents a step forward in the pursuit of reliable peptide secondary-structure prediction using first-principles electronic-structure methods. The results presented here highlight the advances of first-principles methods to address the conformational challenge of peptides. However, they also illustrate the current limitations, pointing out direc-

tions for future efforts, but also pitfalls. It is not enough to consider only a few "reasonable" conformers, even if the peptide is intuitively designed to adopt a certain structure [e.g., a helix as the case for Ac-β^2hAla$_6$-Lys(H$^+$)]. Accurate high-level benchmark data [e.g., CCSD(T)] would be highly desirable in order to pinpoint the potential-energy surface (PES) in the first place. This is not feasible at present, but an important objective for future developments. Another important issue is the determination of reliable anharmonic free energies, especially revealed by the study of the β-peptide. Again, at present, this is not feasible for the systems considered here, but it is a route that needs to be pursued.

Appendices

A EXTRA DETAILS FOR PART II

A.1 IR SPECTRA DERIVED FROM FORCE-FIELD MD SIMULATIONS

For the helical peptide Ac-Ala$_{19}$-Lys(H$^+$), we calculated infrared (IR) spectra from OPLSAA molecular dynamics (MD) simulations. First, we equilibrated the molecule at 300 K with the help of a short (10 ps) thermostatted run. Then, we performed simulations in the NVE ensemble using the same settings as employed for the *ab initio* MD simulations in Chapter 6 (time step of 1 fs). IR spectra were derived from the trajectory for different lengths and convoluted with a Gaussian function with a variable width of $\sigma = 0.005$ times the wavenumber. They are shown in Fig. A.1, together with the corresponding R_P-factor, taking the spectrum for 1 ns as a reference. We see that with $R_\mathrm{P} = 0.07$, the spectrum is already converged after about 25 ps.

A.2 RMSD OF STRUCTURES RELAXED WITH DIFFERENT METHODS

For the most important structural types of Ac-Lys-Ala$_{19}$ + H$^+$ that we found in our PBE+vdW based structure search (discussed in Chapter 8) we re-calculated the energy hierarchies with the PBE0+vdW, the PBE0+MBD*, and the PBE+MBD* functionals and the AmoebaPro13 force field. For this, we also relaxed the structures with the respective method (see Section 8.6). As detailed in Tab. A.1, the structural changes upon relaxation are very small, especially for the different density-functional theory (DFT) functionals.

Table A.1: Root mean square deviation (RMSD), in nm, between the structure relaxed with the PBE+vdW functional and relaxed with other methods for all conformational types of Ac-Lys-Ala$_{19}$ + H$^+$. For the case of the helical structure, the proton is located close to the C-terminus. All atoms except for hydrogen atoms were considered.

	RMSD w.r.t. PBE+vdW structure (nm)			
Conformer	AmoebaPro13	PBE+MBD*	PBE0+MBD*	PBE0+vdW
C1	0.033	0.003	0.007	0.005
C2	0.039	0.005	0.007	0.005
C3	0.036	0.005	0.006	0.006
C4	0.041	0.005	0.010	0.006
C5	0.043	0.004	0.008	0.005
C6	0.040	0.005	0.006	0.005
helix	–	0.002	0.006	0.005

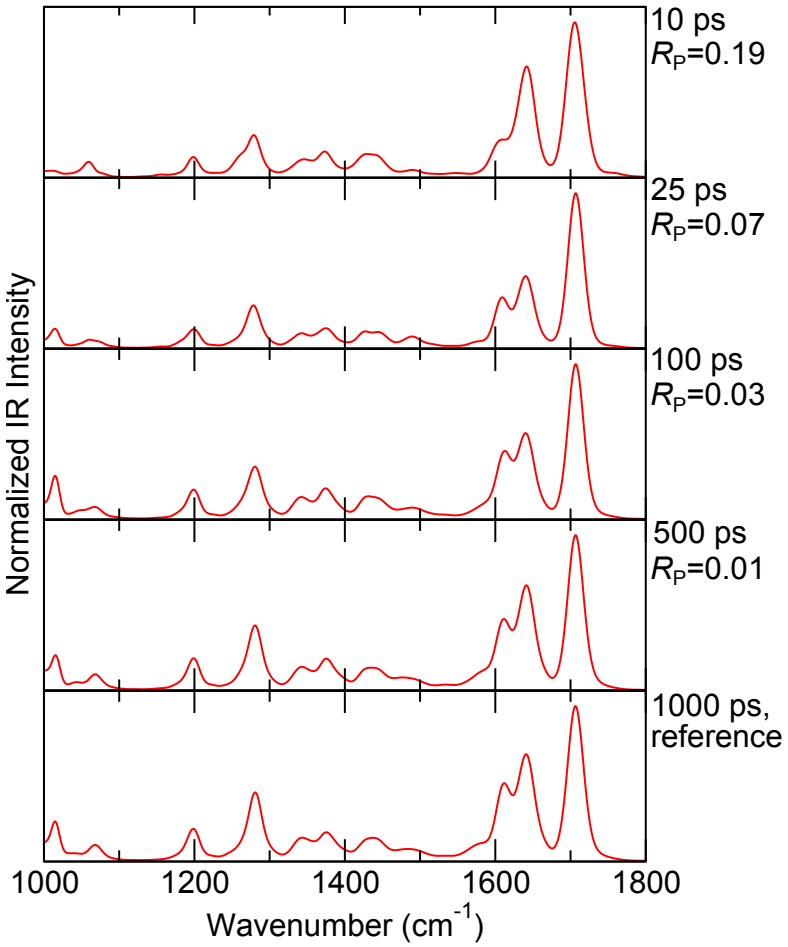


Figure A.1: Ac-Ala$_{19}$-Lys(H$^+$): Convergence of IR spectrum, calculated from OPLSAA MD simulations in the NVE ensemble with $\langle T \rangle = 300\,\mathrm{K}$, with simulation time. The spectrum obtained after 1000 ps is taken as the reference for the calculation of the Pendry reliability factors. No shift of the wavenumber axis is employed.

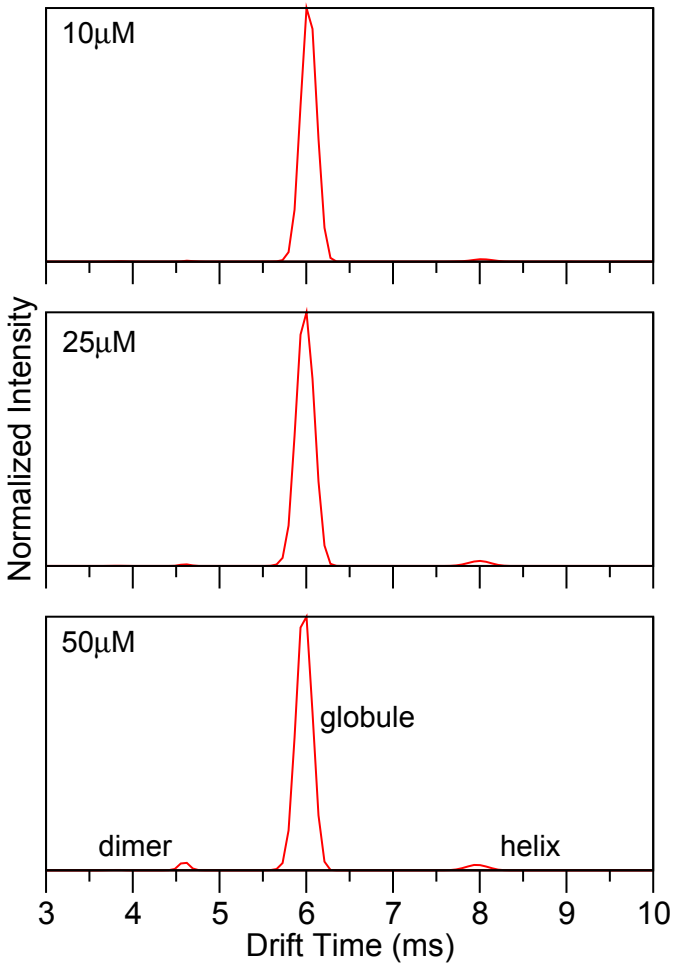


Figure A.2: Normalized arrival time distributions (ATDs) measured for Ac-Lys-Ala$_{19}$ + H$^+$ for different peptide concentrations in the solution that was electrosprayed (10 µM, 25 µM, and 50 µM). The measurements were performed by Stephan Warnke, Gert von Helden, and Kevin Pagel working in the Molecular Physics Department of the Fritz Haber Institute.

A.3 Ion-mobility mass-spectrometry measurements for Ac-Lys-Ala$_{19}$ + H$^+$

Figure A.2 shows arrival time distributions (ATDs) of Ac-Lys-Ala$_{19}$ + H$^+$ measured using ion mobility-mass spectrometry (IM-MS). Three different concentrations of the peptide in the solution that was electrosprayed were used, namely 10 µM, 25 µM, and 50 µM. The amount of dimers observed increases with increasing concentration. This is expected as during the electrospray process the ions enter the gas phase from evaporating droplets. The higher the concentration of the peptide in the droplets, the higher is the possibility that dimers are formed.

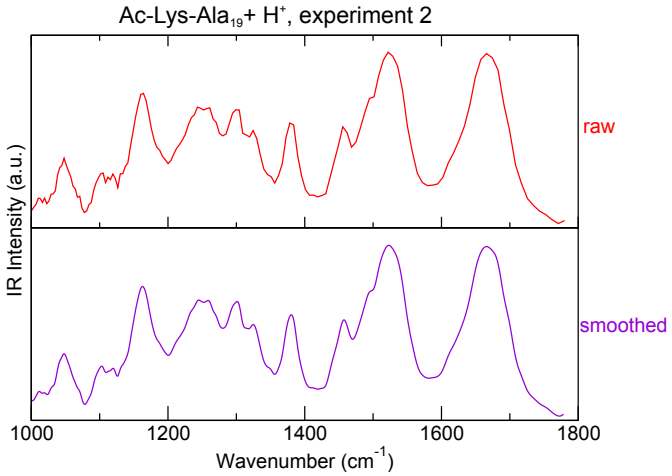


Figure A.3: Comparison of raw and smoothed experimental IRMPD spectrum of Ac-Ala$_{19}$-Lys + H$^+$. This spectrum was obtained from a separate measurement to that discussed in the main text.

A.4 Second experimental IR spectrum for Ac-Lys-Ala$_{19}$ + H$^+$

The infrared multiphoton dissociation (IRMPD) spectrum of Ac-Lys-Ala$_{19}$ + H$^+$ was measured twice in two completely independent measurement cycles. The one with the better resolution is discussed in the main text. The second spectrum is shown in Fig. A.3. The measurements were performed by Peter Kupser and Gert von Helden working in the Molecular Physics Department of the Fritz Haber Institute. The spectrum is also published in the Ph.D. thesis of Peter Kupser[343]. As discussed extensively in the main text, we base the quantitative comparison of theoretical and experimental spectra on the Pendry reliability factor[337]. As it is sensitive to small kinks the experimental spectrum has to be smoothed. A comparison of the experimental raw data and the smoothed spectrum for the result of the second measurement cycle of Ac-Lys-Ala$_{19}$ + H$^+$ is depicted in Fig. A.3. The smoothing procedure used is the same one that was detailed in Chapter 9.

A.5 Comparison of IR spectra

Table A.2 lists the Pendry reliability factors[337] calculated between both experimental spectra for Ac-Lys-Ala$_{19}$ + H$^+$ and the harmonic spectra for the Ac-Lys-Ala$_{19}$ + H$^+$ structure types. In addition, the rigid shifts along the wavenumber axis Δ_x and the normalized intensity axis Δ_y are given. The comparison of the experimental spectra with the IR spectra derived from MD simulations with $\langle T \rangle = 300\,\mathrm{K}$ is shown in Tab. A.3. Qualitatively, the comparisons with the two experimental spectra yield the same results. The Pendry reliability factors were calculated in the wavenumber range between 1130 and 1736 cm^{-1}.

Table A.2: Pendry reliability factors between the harmonic IR spectra for different structure types of Ac-Lys-Ala$_{19}$ + H$^+$ and the two experimental spectra for the wavenumber range between 1130 and 1736 cm^{-1}. In addition, the rigid shifts of the theoretical spectra along the wavenumber axis (Δ_x) and along the normalized intensity axis (Δ_y) are given.

	experiment 1			experiment 2		
conformation	R_P	Δ_x (cm^{-1})	Δ_y	R_P	Δ_x (cm^{-1})	Δ_y
C1	0.47	13.0	0.045	0.40	14.0	0.115
C2	0.32	11.5	0.025	0.27	15.0	0.090
C3	0.39	10.5	0.025	0.42	13.0	0.100
C4	0.28	16.0	0.015	0.31	17.5	0.085
C5	0.37	12.0	0.010	0.39	13.5	0.080
C6	0.38	13.0	0.025	0.32	12.5	0.090
D1	0.48	10.5	0.020	0.52	12.5	0.080
D2	0.52	9.0	0.000	0.50	12.0	0.060
Ac-Lys-Ala$_{19}$ + H$^+$, helix, (H$^+$ near C-term.)	0.58	8.0	0.005	0.57	11.5	0.080
Ac-Lys-Ala$_{19}$ + H$^+$, helix, (H$^+$ at N-term. Lys)	0.76	21.0	0.000	0.73	23.5	0.005

Table A.3: Pendry reliability factors between the anharmonic IR spectra for different structure types of Ac-Lys-Ala$_{19}$ + H$^+$ and the two experimental spectra for the wavenumber range between 1130 and 1736 cm^{-1}. The theoretical spectra are derived from PBE+vdW MD simulations with $\langle T \rangle = 300$ K. In addition, the rigid shifts of the theoretical spectra along the wavenumber axis (Δ_x) and along the normalized intensity axis (Δ_y) are given.

	experiment 1			experiment 2		
conformation	R_P	Δ_x (cm^{-1})	Δ_y	R_P	Δ_x (cm^{-1})	Δ_y
C1	0.44	25.0	0.000	0.31	25.0	0.025
C2	0.31	21.5	0.000	0.23	22.5	0.025
C3	0.33	22.0	0.000	0.33	24.5	0.050
C4	0.34	24.5	0.000	0.24	25.0	0.035
Ac-Lys-Ala$_{19}$ + H$^+$, helix, (H$^+$ near C-term.)	0.29	19.5	0.000	0.27	20.0	0.020

B EXTRA DETAILS FOR PART III

B.1 CONVERGENCE OF VIBRATIONAL NORMAL MODES

The normal modes of vibration are calculated based on a finite-difference approach. We here test the numerical convergence of these modes for the parameters chosen. Figure B.1a shows the frequency as a function of the mode number for our default choices, namely *tight* computational settings, a displacement length δ of 0.0025 Å and a maximal force of less than 10^{-3} eV/Å. Figure B.1b,c, and d illustrate the differences of the vibrational frequencies when varying the displacement length δ (Fig. B.1b,c,d), the radial integration grid density (Fig. B.1c), and the force convergence criterion used for the optimization of the structures (Fig. B.1d). All calculations are performed with the PBE+vdW functional. Varying the default choices in the way shown in Fig. B.1 produces changes in the frequencies of less than 2 cm^{-1} in all cases.

B.2 ANALYSIS OF ENERGETIC CONTRIBUTIONS

In this thesis, free energies are calculated in the rigid rotor-harmonic oscillator approximation as explained in Section 5.1.1 with

$$F_{\mathrm{harm,rigid}} = E_{\mathrm{DFT}} + \underbrace{F_{\mathrm{vib,harm}}}_{U_{\mathrm{vib,harm}} - TS_{\mathrm{vib,harm}}} + F_{\mathrm{rot,rigid}} \quad , \tag{B.1}$$

where E_{DFT} denotes the total energy, $F_{\mathrm{rot,rigid}}$ is the rotational contribution to the free energy in the rigid-rotor approximation, and $F_{\mathrm{vib,harm}}$ is the vibrational contribution to the free energy in the harmonic-oscillator approximation. The latter is split here into the harmonic internal energy $U_{\mathrm{vib,harm}}$ and the entropic part $-TS_{\mathrm{vib,harm}}$, where S denotes the entropy and T the temperature. Table B.1 lists those individual contributions for the representative structures of all hydrogen-bonding families found for Ac-β^2hAla$_6$-Lys(H$^+$) in Chapters 10 and 11 with a free energy $F_{\mathrm{harm,rigid}}$ of less than 100 meV seen from the reference family 2 (also named compact-1). The representative structure of each family was always chosen as the one with the lowest energy among all members. The free energies are calculated at $T = 300$ K.

B.3 IRMPD SPECTRA: RAW DATA VS. SMOOTHED DATA

As the Pendry R-factor is sensitive to small kinks or wiggles in the spectra, the experimental spectra were smoothed before being compared to the calculated spectra. For this, the same procedure as used in Part II was employed. In order not to oversmooth the spectra they were

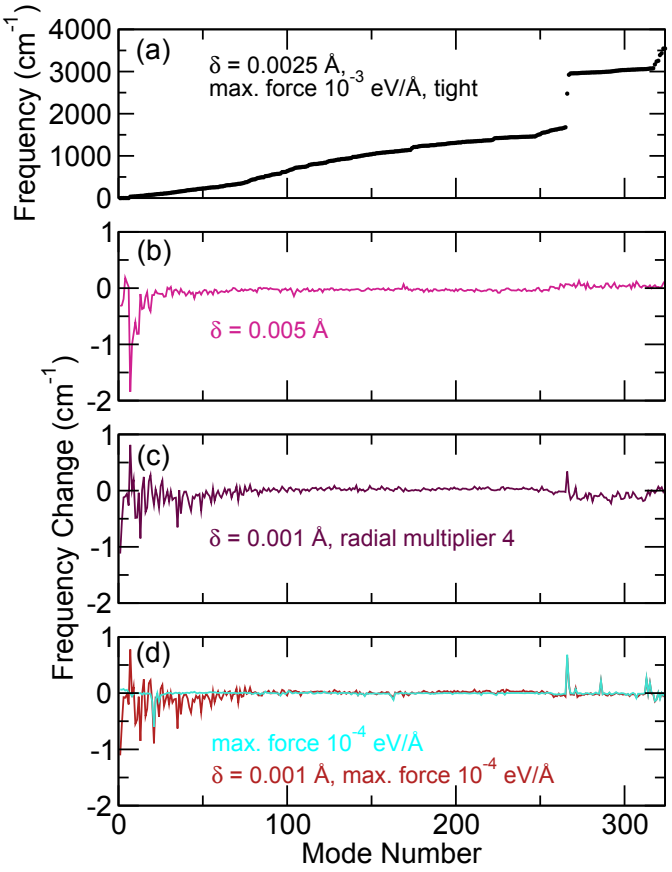


Figure B.1: Convergence of the normal mode frequencies calculated based on finite-differences. (a) Frequency as a function of the mode number for our default choices as given in the plot ($\delta = 0.0025\,$Å, max. force criterion used for relaxation of less than 10^{-3} eV/Å and *tight* computational settings). The plots (b), (c), and (d) show the changes in the frequencies with respect to the frequencies shown in (a) when varying the parameters as specified in the corresponding plot.

Table B.1: Individual contributions to the free energy calculated in the harmonic oscillator-rigid rotor approximation as detailed in Eq. B.1 with $T = 300$ K. The first column indicates the number of the family. The contributions are given for all hydrogen-bonding families found for Ac-β^2hAla$_6$-Lys(H$^+$) with free energies of less than 100 meV relative to the reference (Family 2/compact-1, see Chapter 11). All energies are given relative to this reference.

F	$\Delta F_{\mathrm{harm,rigid}}$ (eV)	ΔE_{DFT} (eV)	$\Delta F_{\mathrm{rot,rigid}} + \Delta F_{\mathrm{vib,harm}}$ (eV)	$\Delta F_{\mathrm{rot,rigid}}$ (eV)	$\Delta F_{\mathrm{vib,harm}}$ (eV)	$\Delta U_{\mathrm{vib,harm}}$ (eV)	$-T\Delta S_{\mathrm{vib,harm}}$ (eV)
1[a]	-0.021	0.296	-0.317	-0.010	-0.307	0.000	-0.307
2[b]	0.000	0.000	0.000	0.000	0.000	0.000	-0.000
3[c]	0.014	0.125	-0.111	0.000	-0.111	-0.019	-0.092
4	0.030	0.102	-0.072	0.000	-0.072	-0.006	-0.066
5	0.043	0.089	-0.045	-0.001	-0.044	-0.000	-0.044
6	0.047	0.062	-0.015	-0.001	-0.015	-0.018	0.003
7	0.052	0.209	-0.157	-0.004	-0.154	0.008	-0.162
8	0.055	0.191	-0.137	-0.000	-0.137	-0.013	-0.124
9	0.057	0.098	-0.041	-0.000	-0.041	-0.017	-0.023
10	0.057	0.169	-0.112	0.001	-0.112	-0.031	-0.082
11	0.069	0.113	-0.043	-0.002	-0.041	-0.005	-0.037
12	0.074	0.297	-0.223	-0.004	-0.219	-0.003	-0.216
13	0.077	0.282	-0.205	-0.005	-0.200	0.010	-0.210
14	0.080	0.320	-0.240	-0.005	-0.235	0.012	-0.248
15	0.082	0.142	-0.060	-0.001	-0.059	-0.013	-0.046
16[d]	0.085	0.302	-0.217	-0.004	-0.213	0.010	-0.223
17	0.089	0.362	-0.273	-0.007	-0.265	0.018	-0.284
18	0.090	0.242	-0.152	-0.002	-0.150	-0.002	-0.148
19	0.095	0.323	-0.229	-0.011	-0.217	0.016	-0.233

[a] H12-1
[b] compact-1, reference
[c] H20-1
[d] H16-1

first splined onto a grid with a width of 2 cm^{-1} before splined twice with a three-point formula:

$$\tilde{y}_n = \frac{y_{n-1} + 2y_n + y_{n+1}}{4} \quad . \tag{B.2}$$

Afterwards the spectra were splined onto a fine numerical grid (grid width: 0.5 cm^{-1}) to perform the R-factor calculation. A comparison of the raw experimental spectra to the smoothed ones are given in Fig. B.2 (a) for Ac-Ala$_6$-Lys(H$^+$) and (b) for Ac-β^2hAla$_6$-Lys(H$^+$).

B.4 Comparison of IR spectra based on the Pendry R-factor

Table B.2 gives the Pendry R-factors and the rigid shifts Δ_x and Δ_y included in the computation of R_{P} for selected families of Ac-Ala$_6$-Lys(H$^+$) and Ac-β^2hAla$_6$-Lys(H$^+$). Rigid shifts between experiment and calculated spectra along the wavenumber axis (x-axis) Δ_x most probably account for small systematic mode softening arising from the exchange-correlation (XC) functional and the neglect of quantum nuclear effects. To account for offsets in the experimental spectra, we additionally included a rigid shift along the (normalized) intensity axis (y-axis) Δ_y.

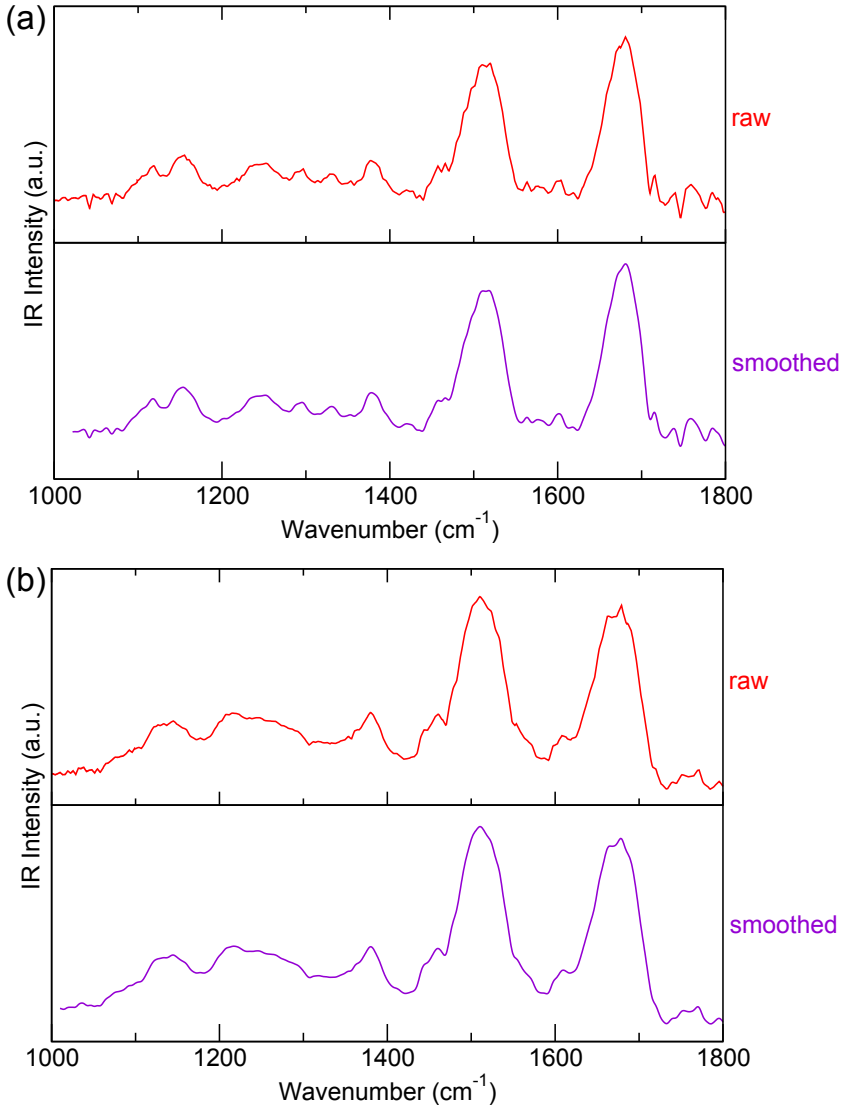


Figure B.2: Comparison of raw and smoothed experimental IRMPD spectra for (a) Ac-Ala_6-Lys(H^+) and (b) Ac-β^2hAla_6-Lys(H^+).

Table B.2: Pendry R-factors and rigid shifts Δ_x and Δ_y for Ac-Ala_6-Lys(H^+) and Ac-β^2hAla_6-Lys(H^+) hydrogen-bonding families.

Family	R_P	Δ_x (cm^{-1})	Δ_y
Ac-Ala_6-Lys(H^+)			
non-helical	0.46	28.5	0.140
α-helix	0.31	24.0	0.100
Ac-β^2hAla_6-Lys(H^+)			
compact	0.55	14.5	0.120
H12	0.64	29.5	0.215
H16	0.49	16.5	0.155
H20	0.47	10.5	0.220

Eidesstattliche Versicherung

Hiermit versichere ich, alle Hilfsmittel und Hilfen angeben und auf dieser Grundlage die Arbeit selbstständig verfasst zu haben. Die Arbeit wurde nicht schon einmal in einem früheren Promotionsverfahren angenommen oder als ungenügend beurteilt.

CURRICULUM VITÆ

For reasons of data protection,
the curriculum vitæ is not included in the online version.

Acknowledgements

I would like to thank Matthias Scheffler for giving me the opportunity to carry out my Ph.D. studies in the very stimulating atmosphere of the Fritz Haber Institute. I had the chance to meet and discuss with many experts of the field, not to mention the easy access to computational facilities and time.

I would furthermore like to thank the members of the Ph.D. committee, specifically Roland Netz for being the second referee.

A big thank you goes to Volker Blum. Volker, thank you very much for closely supervising my work throughout the last four years. I have learned many things and I am grateful for your ideas and support.

Carsten, thank you very much for sharing your biochemical insights with me. I am grateful for all your advice and ideas.

Mariana, thank you for answering so many questions! You literally always had an open door for me. Muito obrigada!

Furthermore, I enjoyed collaborating with the experimentalists from the Molecular Physics Department of the FHI. I thank Kevin Pagel, Stephan Warnke, Gert von Helden, and also Nadja Heine and Knut Asmis. Thanks for all the fruitful discussions and your patient explanation of the experimental details.

I would also like to acknowledge all my present and former colleagues from the Theory Department. Thanks to all of you! Thanks, Luca, for helping me with the REMD framework, for proofreading parts of the manuscript and for sharing your insights into statistical mechanics. I would also like to thank Julia, Birgit, Gaby, Hanna, and Steffen for your support with all administrative questions (and beyond). A special thanks goes to the (bio)group for many fruitful discussions and feedback. Jürgen, thank you for your patience and explanations especially at the beginning of my Ph.D. Suchi and Matti, it was great to share the office with you for such a long time, almost three years. Adriana, thank you for all your kind support during the writing of this thesis and for all the nice scientific discussions and cafeteria lunches.

A special thanks goes to Anthony, for being a good friend, for proofreading, and also for sharing your insights into the English language (which or that?).

Guoxu, Claudi, and Norina – thank you so much; our girls team was great fun. Guoxu, thank you for making me familiar with Chinese culture, especially the delicious food.

Finally, I would like to thank my family and my friends outside the Institute, especially Rina and Andreas. A big thanks to my parents and my sister.

BIBLIOGRAPHY

[1] FREDDOLINO, P. L., HARRISON, C. B., LIU, Y., AND SCHULTEN, K. Challenges in protein-folding simulations. *Nat. Phys. 6*, 10 (2010), 751.

[2] DILL, K. A., AND MACCALLUM, J. L. The Protein-Folding Problem, 50 Years On. *Science 338*, 6110 (2012), 1042.

[3] LANE, T. J., SHUKLA, D., BEAUCHAMP, K. A., AND PANDE, V. S. To milliseconds and beyond: challenges in the simulation of protein folding. *Curr. Opin. Struct. Biol. 23*, 1 (2013), 58.

[4] ANFINSEN, C. B., HABER, E., SELA, M., AND WHITE, F. H. The Kinetics of Formation of Native Ribonuclease During Oxidation of the Reduced Polypeptide Chain. *Proc. Natl. Acad. Sci. U.S.A. 47*, 9 (1961), 1309.

[5] ANFINSEN, C. B. Principles that govern the folding of protein chains. *Science 181*, 4096 (1973), 223.

[6] ANFINSEN, C. B. On the possibility of predicting tertiary structure from primary sequence. In *New Perspectives in Biology (M. Sela, ed.)*. Elsevier Pub. Co., Amsterdam, The Netherlands, 1964, p. 42.

[7] DILL, K. A., AND CHAN, H. S. From Levinthal to pathways to funnels. *Nature Struct. Biol. 4*, 1 (1997), 10.

[8] BERRY, R. S., ELMACI, N., ROSE, J. P., AND VEKHTER, B. Linking Topography of Its Potential Surface with the Dynamics of Folding of a Protein Model. *Proc. Natl. Acad. Sci. U.S.A. 94*, 18 (1997), 9520.

[9] BOEHR, D. D., NUSSINOV, R., AND WRIGHT, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol. 5*, 11 (2009), 789.

[10] PANDE, V. S. Folding@home. http://folding.stanford.edu [Accessed in April 2013].

[11] VOELZ, V. A., BOWMAN, G. R., BEAUCHAMP, K., AND PANDE, V. S. Molecular Simulation of ab Initio Protein Folding for a Millisecond Folder NTL9(1-39). *J. Am. Chem. Soc. 132*, 5 (2010), 1526.

[12] SHAW, D. E., MARAGAKIS, P., LINDORFF-LARSEN, K., PIANA, S., DROR, R. O., EASTWOOD, M. P., BANK, J. A., JUMPER, J. M., SALMON, J. K., SHAN, Y., AND WRIGGERS, W. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science 330*, 6002 (2010), 341.

[13] BEST, R. B., BUCHETE, N.-V., AND HUMMER, G. Are Current Molecular Dynamics Force Fields too Helical? *Biophys. J. 95*, 1 (2008), L07.

[14] PENEV, E., IRETA, J., AND SHEA, J.-E. Energetics of Infinite Homopolypeptide Chains: A New Look at Commonly Used Force Fields. *J. Phys. Chem. B 112*, 22 (2008), 6872.

[15] PERDEW, J. P., BURKE, K., AND ERNZERHOF, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett. 77*, 18 (1996), 3865.

[16] TKATCHENKO, A., AND SCHEFFLER, M. Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Phys. Rev. Lett. 102*, 7 (2009), 073005.

[17] ROSSI, M. *Ab initio study of secondary structure formation in gas-phase peptides*. Ph.D. thesis, TU Berlin and Fritz-Haber-Institut der Max-Planck-Gesellschaft, 2011.

[18] SEEBACH, D., AND GARDINER, J. β-Peptidic Peptidomimetics. *Acc. Chem. Res. 41*, 10 (2008), 1366.

[19] PILSL, L. K. A., AND REISER, O. α/β-Peptide foldamers: state of the art. *Amino Acids 41*, 3 (2011), 709.

[20] FRACKENPOHL, J., ARVIDSSON, P. I., SCHREIBER, J. V., AND SEEBACH, D. The Outstanding Biological Stability of β- and γ-Peptides toward Proteolytic Enzymes: An In Vitro Investigation with Fifteen Peptidases. *ChemBioChem 2*, 6 (2001), 445.

[21] KRITZER, J. A., LUEDTKE, N. W., HARKER, E. A., AND SCHEPARTZ, A. A Rapid Library Screen for Tailoring β-Peptide Structure and Function. *J. Am. Chem. Soc. 127*, 42 (2005), 14584.

[22] KRITZER, J. A., LEAR, J. D., HODSDON, M. E., AND SCHEPARTZ, A. Helical β-Peptide Inhibitors of the p53-hDM2 Interaction. *J. Am. Chem. Soc. 126*, 31 (2004), 9468.

[23] MICHEL, J., HARKER, E. A., TIRADO-RIVES, J., JORGENSEN, W. L., AND SCHEPARTZ, A. In Silico Improvement of β^3-Peptide Inhibitors of p53•hDM2 and p53•hDMX. *J. Am. Chem. Soc. 131*, 18 (2009), 6356.

[24] ARMSTRONG, G. Computational chemistry: Binding better. *Nat. Chem. , doi: 10.1038/nchem.243* (2009).

[25] HUDGINS, R. R., RATNER, M. A., AND JARROLD, M. F. Design of Helices That Are Stable in Vacuo. *J. Am. Chem. Soc. 120*, 49 (1998), 12974.

[26] HUDGINS, R. R., AND JARROLD, M. F. Helix Formation in Unsolvated Alanine-Based Peptides: Helical Monomers and Helical Dimers. *J. Am. Chem. Soc. 121*, 14 (1999), 3494.

[27] JARROLD, M. F. Helices and Sheets in vacuo. *Phys. Chem. Chem. Phys. 9*, 14 (2007), 1659.

[28] ROSSI, M., SCHEFFLER, M., AND BLUM, V. Impact of Vibrational Entropy on the Stability of Unsolvated Peptide Helices with Increasing Length. *J. Phys. Chem. B 117*, 18 (2013), 5574.

[29] VOET, D., AND VOET, J. G. *Biochemistry*, 4 ed. John Wiley & Sons, 2011.

[30] CRICK, F. H. C. On protein synthesis. *Symp. Soc. Exp. Biol. 12* (1958), 138.

[31] SEWALD, N., AND JAKUBKE, H.-D. *Peptides: chemistry and biology*. Wiley-VCH, Weinheim, 2009.

[32] FISCHER, G. Chemical aspects of peptide bond isomerisation. *Chem. Soc. Rev. 29*, 2 (2000), 119.

[33] DUGAVE, C., AND DEMANGE, L. Cis-Trans Isomerization of Organic Molecules and Biomolecules: Implications and Applications. *Chem. Rev. 103*, 7 (2003), 2475.

[34] RAMACHANDRAN, G., RAMAKRISHNAN, C., AND SASISEKHARAN, V. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol. 7* (1963), 95.

[35] HO, B. K., THOMAS, A., AND BRASSEUR, R. Revisiting the Ramachandran plot: Hard-sphere repulsion, electrostatics, and H-bonding in the α-helix. *Protein Sci. 12*, 11 (2003), 2508.

[36] LOVELL, S. C., DAVIS, I. W., ARENDALL, W. B., DE BAKKER, P. I. W., WORD, J. M., PRISANT, M. G., RICHARDSON, J. S., AND RICHARDSON, D. C. Structure validation by Cα geometry: ϕ,ψ and Cβ deviation. *Proteins: Struct., Funct., Bioinf. 50*, 3 (2003), 437.

[37] PORTER, L. L., AND ROSE, G. D. Redrawing the Ramachandran plot after inclusion of hydrogen-bonding constraints. *Proc. Natl. Acad. Sci. U.S.A. 108*, 1 (2011), 109.

[38] COCK, P. Drawing Ramachandran (phi/psi) plots for Proteins. warwick.ac.uk/fac/sci/moac/people/students/peter_cock/r/ramachandran [Accessed in April 2014].

[39] HOL, W. G. The role of the α-helix dipole in protein function and structure. *Prog. Biophys. Mol. Biol. 45*, 3 (1985), 149.

[40] LONDON, F. Zur Theorie und Systematik der Molekularkräfte. *Z. Physik 63*, 3-4 (1930), 245.

[41] IRETA, J., NEUGEBAUER, J., SCHEFFLER, M., ROJO, A., AND GALVÁN, M. Density Functional Theory Study of the Cooperativity of Hydrogen Bonds in Finite and Infinite α-Helices. *J. Phys. Chem. B 107*, 6 (2003), 1432.

[42] GUO, H., AND KARPLUS, M. Solvent Influence on the Stability of the Peptide Hydrogen Bond: A Supramolecular Cooperative Effect. *J. Phys. Chem. 98*, 29 (1994), 7104.

[43] DANNENBERG, J. The Importance of Cooperative Interactions and a Solid-State Paradigm to Proteins: What Peptide Chemists Can Learn from Molecular Crystals. In *Advances in Protein Chemistry*, vol. 72. Academic Press, 2005, p. 227.

[44] TKATCHENKO, A., ROSSI, M., BLUM, V., IRETA, J., AND SCHEFFLER, M. Unraveling the Stability of Polypeptide Helices: Critical Role of van der Waals Interactions. *Phys. Rev. Lett. 106*, 11 (2011), 118102.

[45] LINDERSTRØM-LANG, K. U. *Lane Medical Lectures: Proteins and Enzymes*, vol. 6. Stanford University Press, 1952.

[46] COREY, R. B., AND PAULING, L. Molecular Models of Amino Acids, Peptides, and Proteins. *Rev. Sci. Instrum. 24*, 8 (1953), 621.

[47] HUMPHREY, W., DALKE, A., AND SCHULTEN, K. VMD – Visual Molecular Dynamics. *J. Mol. Graphics 14* (1996), 33.

[48] PAULING, L., COREY, R. B., AND BRANSON, H. R. The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. U.S.A. 37*, 4 (1951), 205.

[49] EISENBERG, D. The discovery of the α-helix and β-sheet, the principal structural features of proteins. *Proc. Natl. Acad. Sci. U.S.A. 100*, 20 (2003), 11207.

[50] PAULING, L., AND COREY, R. B. The Pleated Sheet, A New Layer Configuration of Polypeptide Chains. *Proc. Natl. Acad. Sci. U.S.A. 37*, 5 (1951), 251.

[51] VENKATACHALAM, C. M. Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units. *Biopolymers 6*, 10 (1968), 1425.

[52] LEWIS, P. N., MOMANY, F. A., AND SCHERAGA, H. A. Chain reversals in proteins. *Biochem. Biophys. Acta 303*, 2 (1973), 211.

[53] CHOU, P. Y., AND FASMAN, G. D. β-turns in proteins. *J. Mol. Biol. 115*, 2 (1977), 135.

[54] MÖHLE, K., GUSSMANN, M., AND HOFMANN, H.-J. Structural and energetic relations between β turns. *J. Comput. Chem. 18*, 11 (1997), 1415.

[55] APPELLA, D. H., CHRISTIANSON, L. A., KARLE, I. L., POWELL, D. R., AND GELLMAN, S. H. β-Peptide Foldamers: Robust Helix Formation in a New Family of β-Amino Acid Oligomers. *J. Am. Chem. Soc. 118*, 51 (1996), 13071.

[56] GELLMAN, S. H. Foldamers: A Manifesto. *Acc. Chem. Res. 31*, 4 (1998), 173.

[57] KOVACS, J., BALLINA, R., RODIN, R. L., BALASUBRAMANIAN, D., AND APPLEQUIST, J. Poly-β-L-aspartic Acid. Synthesis through Pentachlorophenyl Active Ester and Conformational Studies. *J. Am. Chem. Soc. 87*, 1 (1965), 119.

[58] BESTIAN, H. Poly-β-amides. *Angew. Chem., Int. Ed. 7*, 4 (1968), 278.

[59] SCHMIDT, V. E. Über optisch aktive Poly-β-amide. *Angew. Makromol. Chem. 14*, 1 (1970), 185.

[60] YUKI, H., OKAMOTO, Y., TAKETANI, Y., TSUBOTA, T., AND MARUBAYASHI, Y. Poly(β-amino acid)s. IV. Synthesis and conformational properties of poly(α-isobutyl-L-aspartate). *J. Polymer Sci. Polymer Chem. 16*, 9 (1978), 2237.

[61] FERNÁNDEZ-SANTÍN, J. M., AYMAMÍ, J., RODRÍGUEZ-GALÁN, A., MUÑOZ-GUERRA, S., AND SUBIRANA, J. A. A pseudo α-helix from poly(α-isobutyl-L-aspartate), a nylon-3 derivative. *Nature 311*, 5981 (1984), 53.

[62] FERNÁNDEZ-SANTÍN, J. M., MUÑOZ-GUERRA, S., RODRÍGUEZ-GALÁN, A., AYMAMI, J., LLOVERAS, J., SUBIRANA, J. A., GIRALT, E., AND PTAK, M. Helical conformations in a polyamide of the nylon-3 family. *Macromolecules 20*, 1 (1987), 62.

[63] NAVAS, J. J., ALEMÁN, C., AND MUÑOZ-GUERRA, S. Effects of Aqueous and Organic Solvents on the Conformational Properties of the Helix-Forming α-Methyl-β-l-aspartamyl Residue. *J. Org. Chem. 61*, 20 (1996), 6849.

[64] LÓPEZ-CARRASQUERO, F., GARCÍA-ALVAREZ, M., NAVAS, J. J., ALEMÁN, C., AND MUÑOZ-GUERRA, S. Structural Study on Poly(β-l-aspartate)s with Short Alkyl Side Chains: Helical and Extended Crystal Forms. *Macromolecules 29*, 26 (1996), 8449.

[65] SEEBACH, D., OVERHAND, M., KÜHNLE, F. N. M., MARTINONI, B., OBERER, L., HOMMEL, U., AND WIDMER, H. β-Peptides: Synthesis by Arndt-Eistert homologation with concomitant peptide coupling. Structure determination by NMR and CD spectroscopy and by X-ray crystallography. Helical secondary structure of a β-hexapeptide in solution and its stability towards pepsin. *Helv. Chim. Acta 79*, 4 (1996), 913.

[66] SEEBACH, D., CICERI, P. E., OVERHAND, M., JAUN, B., RIGO, D., OBERER, L., HOMMEL, U., AMSTUTZ, R., AND WIDMER, H. Probing the Helical Secondary Structure of Short-Chain β-Peptides. *Helv. Chim. Acta 79*, 8 (1996), 2043.

[67] APPELLA, D. H., CHRISTIANSON, L. A., KLEIN, D. A., POWELL, D. R., HUANG, X. L., BARCHI, J. J., AND GELLMAN, S. H. Residue-based control of helix shape in beta-peptide oligomers. *Nature 387*, 6631 (1997), 381.

[68] WU, Y. D., AND WANG, D. P. Theoretical study on side-chain control of the 14-helix and the 10/12-helix of beta-peptides. *J. Am. Chem. Soc. 121*, 40 (1999), 9352.

[69] CHENG, R. P., GELLMAN, S. H., AND DEGRADO, W. F. β-Peptides: From Structure to Function. *Chem. Rev. 101*, 10 (2001), 3219.

[70] HILL, D. J., MIO, M. J., PRINCE, R. B., HUGHES, T. S., AND MOORE, J. S. A Field Guide to Foldamers. *Chem. Rev. 101*, 12 (2001), 3893.

[71] SAWADA, T., AND GELLMAN, S. H. Structural Mimicry of the α-Helix in Aqueous Solution with an Isoatomic $\alpha/\beta/\gamma$-Peptide Backbone. *J. Am. Chem. Soc. 133*, 19 (2011), 7336.

[72] GOODMAN, C. M., CHOI, S., SHANDLER, S., AND DEGRADO, W. F. Foldamers as versatile frameworks for the design and evolution of function. *Nat. Chem. Biol. 3*, 5 (2007), 252.

[73] HORNE, W. S., AND GELLMAN, S. H. Foldamers with Heterogeneous Backbones. *Acc. Chem. Res. 41*, 10 (2008), 1399.

[74] BALDAUF, C., AND HOFMANN, H.-J. Ab initio MO Theory – An Important Tool in Foldamer Research: Prediction of Helices in Oligomers of ω-Amino Acids. *Helv. Chim. Acta 95*, 12 (2012), 2348.

[75] ZHANG, D.-W., ZHAO, X., HOU, J.-L., AND LI, Z.-T. Aromatic Amide Foldamers: Structures, Properties, and Functions. *Chem. Rev. 112*, 10 (2012), 5271.

[76] JUWARKER, H., SUK, J.-M., AND JEONG, K.-S. Foldamers with helical cavities for binding complementary guests. *Chem. Soc. Rev. 38*, 12 (2009), 3316.

[77] MARTINEK, T., AND FÜLÖP, F. Peptidic foldamers: ramping up diversity. *Chem. Soc. Rev. 41*, 2 (2012), 687.

[78] CHOUTKO, A., AND VAN GUNSTEREN, W. F. Conformational Preferences of a β-Octapeptide as Function of Solvent and Force-Field Parameters. *Helv. Chim. Acta 96*, 2 (2013), 189.

[79] SEEBACH, D., HOOK, D. F., AND GLÄTTLI, A. Helices and other secondary structures of β- and γ-peptides. *Biopolymers 84*, 1 (2006), 23.

[80] BANERJEE, A., AND BALARAM, P. Stereochemistry of peptides and polypeptides containing omega amino acids. *Curr. Sci. 73*, 12 (1997), 1067.

[81] SEEBACH, D., BECK, A. K., AND BIERBAUM, D. J. The World of β- and γ-Peptides Comprised of Homologated Proteinogenic Amino Acids and Other Components. *Chem. Biodiversity 1*, 8 (2004), 1111.

[82] ARVIDSSON, P. I., RUEPING, M., AND SEEBACH, D. Design, machine synthesis, and NMR-solution structure of a β-heptapeptide forming a salt-bridge stabilised 3_{14}-helix in methanol and in water. *Chem. Commun.*, 7 (2001), 649.

[83] HART, S. A., BAHADOOR, A. B. F., MATTHEWS, E. E., QIU, X. J., AND SCHEPARTZ, A. Helix Macrodipole Control of β^3-Peptide 14-Helix Stability in Water. *J. Am. Chem. Soc. 125*, 14 (2003), 4022.

[84] BAQUERO, E. E., JAMES III, W. H., CHOI, S. H., GELLMAN, S. H., AND ZWIER, T. S. Single-Conformation Ultraviolet and Infrared Spectroscopy of Model Synthetic Foldamers: β-Peptides Ac-β^3-hPhe-NHMe and Ac-β^3-hTyr-NHMe. *J. Am. Chem. Soc. 130*, 14 (2008), 4784.

[85] BAQUERO, E. E., JAMES III, W. H., CHOI, S. H., GELLMAN, S. H., AND ZWIER, T. S. Single-Conformation Ultraviolet and Infrared Spectroscopy of Model Synthetic Foldamers: β-Peptides Ac-β^3-hPhe-β^3-hAla-NHMe and Ac-β^3-hAla-β^3-hPhe-NHMe. *J. Am. Chem. Soc. 130*, 14 (2008), 4795.

[86] JAMES III, W. H., MÜLLER, C. W., BUCHANAN, E. G., NIX, M. G. D., GUO, L., ROSKOP, L., GORDON, M. S., SLIPCHENKO, L. V., GELLMAN, S. H., AND ZWIER, T. S. Intramolecular Amide Stacking and Its Competition with Hydrogen Bonding in a Small Foldamer. *J. Am. Chem. Soc. 131*, 40 (2009), 14243.

[87] JAMES III, W. H., BAQUERO, E. E., SHUBERT, V. A., CHOI, S. H., GELLMAN, S. H., AND ZWIER, T. S. Single-Conformation and Diastereomer Specific Ultraviolet and Infrared Spectroscopy of Model Synthetic Foldamers: α/β-Peptides. *J. Am. Chem. Soc. 131*, 18 (2009), 6574.

[88] JAMES III, W. H., BAQUERO, E. E., CHOI, S. H., GELLMAN, S. H., AND ZWIER, T. S. Laser Spectroscopy of Conformationally Constrained α/β-Peptides: Ac-ACPC-Phe-NHMe and Ac-Phe-ACPC-NHMe. *J. Phys. Chem. A 114*, 3 (2010), 1581.

[89] DAURA, X., GADEMANN, K., SCHÄFER, H., JAUN, B., SEEBACH, D., AND VAN GUNSTEREN, W. F. The β-Peptide Hairpin in Solution: Conformational Study of a β-Hexapeptide in Methanol by NMR Spectroscopy and MD Simulation. *J. Am. Chem. Soc. 123*, 10 (2001), 2393.

[90] SOARES, T., CHRISTEN, M., HU, K., AND VAN GUNSTEREN, W. F. Alpha- and beta-polypeptides show a different stability of helical secondary structure. *Tetrahedron 60*, 35 (2004), 7775.

[91] GEE, P. J., AND VAN GUNSTEREN, W. F. Terminal-group effects on the folding behavior of selected beta-peptides. *Proteins: Struct., Funct., Bioinf. 63*, 1 (2006), 136.

[92] BARON, R., BAKOWIES, D., VAN GUNSTEREN, W. F., AND DAURA, X. β-Peptides with Different Secondary-Structure Preferences: How Different Are Their Conformational Spaces? *Helv. Chim. Acta 85*, 11 (2002), 3872.

[93] WANG, D., FREITAG, F., GATTIN, Z., HABERKERN, H., JAUN, B., SIWKO, M., VYAS, R., VAN GUNSTEREN, W. F., AND DOLENC, J. Validation of the GROMOS 54A7 Force Field Regarding Mixed α/β-Peptide Molecules. *Helv. Chim. Acta 95*, 12 (2012), 2562.

[94] GLÄTTLI, A., AND VAN GUNSTEREN, W. F. Are NMR-Derived Model Structures for β-Peptides Representative for the Ensemble of Structures Adopted in Solution? *Angew. Chem., Int. Ed. 43*, 46 (2004), 6312.

[95] SEEBACH, D., SCHREIBER, J. V., ABELE, S., DAURA, X., AND VAN GUNSTEREN, W. F. Structure and Conformation of β-Oligopeptide Derivatives with Simple Proteinogenic Side Chains: Circular Dichroism and Molecular Dynamics Investigations. *Helv. Chim. Acta 83*, 1 (2000), 34.

[96] GLÄTTLI, A., SEEBACH, D., AND VAN GUNSTEREN, W. F. Do Valine Side Chains Have an Influence on the Folding Behavior of β-Substituted β-Peptides? *Helv. Chim. Acta 87*, 10 (2004), 2487.

[97] GLÄTTLI, A., DAURA, X., BINDSCHÄDLER, P., JAUN, B., MAHAJAN, Y. R., MATHAD, R. I., RUEPING, M., SEEBACH, D., AND VAN GUNSTEREN, W. F. On the Influence of Charged Side Chains on the Folding–Unfolding Equilibrium of β-Peptides: A Molecular Dynamics Simulation Study. *Chem. Eur. J. 11*, 24 (2005), 7276.

[98] LIN, Z., TIMMERSCHEIDT, T. A., AND VAN GUNSTEREN, W. F. Using enveloping distribution sampling to compute the free enthalpy difference between right- and left-handed helices of a β-peptide in solution. *J. Chem. Phys. 137*, 6 (2012), 064108.

[99] DAURA, X., GADEMANN, K., JAUN, B., SEEBACH, D., VAN GUNSTEREN, W. F., AND MARK, A. E. Peptide Folding: When Simulation Meets Experiment. *Angew. Chem., Int. Ed. 38*, 1-2 (1999), 236.

[100] ZAGROVIC, B., GATTIN, Z., LAU, J. K.-C., HUBER, M., AND VAN GUNSTEREN, W. F. V. Structure and dynamics of two β-peptides in solution from molecular dynamics simulations validated against experiment. *Eur. Biophys. J. 37*, 6 (2008), 903.

[101] WU, Y.-D., LIN, J.-Q., AND ZHAO, Y.-L. Theoretical Study of β-Peptide Models: Intrinsic Preferences of Helical Structures. *Helv. Chim. Acta 85*, 10 (2002), 3144.

[102] MÖHLE, K., GÜNTHER, R., THORMANN, M., SEWALD, N., AND HOFMANN, H.-J. Basic conformers in β-peptides. *Biopolymers 50*, 2 (1999), 167.

[103] BALDAUF, C. *Secondary Structure Formation in Homologous Peptides*. Ph.D. thesis, Universität Leipzig, 2005.

[104] WU, Y.-D., AND WANG, D.-P. Theoretical Studies of β-Peptide Models. *J. Am. Chem. Soc. 120*, 51 (1998), 13485.

[105] BALDAUF, C., GÜNTHER, R., AND HOFMANN, H.-J. Mixed Helices—A General Folding Pattern in Homologous Peptides? *Angew. Chem., Int. Ed. 43*, 12 (2004), 1594.

[106] BALDAUF, C., GÜNTHER, R., AND HOFMANN, H.-J. Side-chain control of folding of the homologous α-, β-, and γ-peptides into "mixed" helices (β-helices). *Biopolymers 80*, 5 (2005), 675.

[107] SEEBACH, D., MATHAD, R. I., KIMMERLIN, T., MAHAJAN, Y. R., BINDSCHÄDLER, P., RUEPING, M., JAUN, B., HILTY, C., AND ETEZADY-ESFARJANI, T. NMR-Solution Structures in Methanol of an α-Heptapeptide, of a β^3/β^2-Nonapeptide, and of an all-β^3-Icosapeptide Carrying the 20 Proteinogenic Side Chains. *Helv. Chim. Acta 88*, 7 (2005), 1969.

[108] HETÉNYI, A., MÁNDITY, I. M., MARTINEK, T. A., TÓTH, G. K., AND FÜLÖP, F. Chain-Length-Dependent Helical Motifs and Self-Association of β-Peptides with Constrained Side Chains. *J. Am. Chem. Soc. 127*, 2 (2005), 547.

[109] THRELFALL, R., DAVIES, A., HOWARTH, N. M., FISHER, J., AND COSSTICK, R. Peptides derived from nucleoside β-amino acids form an unusual 8-helix. *Chem. Commun.*, 5 (2008), 585.

[110] SZOLNOKI, É., HETÉNYI, A., MARTINEK, T. A., SZAKONYI, Z., AND FÜLÖP, F. Self-association-driven transition of the β-peptidic H12 helix to the H18 helix. *Org. Biomol. Chem. 10*, 2 (2011), 255.

[111] LÓPEZ-CARRASQUERO, F., ALEMÁN, C., GARCÍA-ALVAREZ, M., DE ILARDUYA, A. M., AND MUÑOZ-GUERRA, S. Poly(α-butyl β-L-aspartate): A second alkoxycarbonyl nylon-3 derivative in helical conformation. *Macromol. Chem. Phys. 196*, 1 (1995), 253.

[112] SEEBACH, D., GADEMANN, K., SCHREIBER, J. V., MATTHEWS, J. L., HINTERMANN, T., JAUN, B., OBERER, L., HOMMEL, U., AND WIDMER, H. 'Mixed' β-Peptides: A unique helical secondary structure in solution. Preliminary communication. *Helv. Chim. Acta 80*, 7 (1997), 2033.

[113] RUEPING, M., SCHREIBER, J. V., LELAIS, G., JAUN, B., AND SEEBACH, D. Mixed β^2/β^3-Hexapeptides and β^2/β^3-Nonapeptides Folding to (P)-Helices with Alternating Twelve- and Ten-Membered Hydrogen-Bonded Rings. *Helv. Chim. Acta 85*, 9 (2002), 2577.

[114] URRY, D. W., GOODALL, M. C., GLICKSON, J. D., AND MAYERS, D. F. The Gramicidin A Transmembrane Channel: Characteristics of Head-to-Head Dimerized π(L,D) Helices. *Proc. Natl. Acad. Sci. U.S.A. 68*, 8 (1971), 1907.

[115] KOVACS, F., QUINE, J., AND CROSS, T. A. Validation of the single-stranded channel conformation of gramicidin A by solid-state NMR. *Proc. Natl. Acad. Sci. U.S.A. 96*, 14 (1999), 7910.

[116] ARNDT, H.-D., BOCKELMANN, D., KNOLL, A., LAMBERTH, S., GRIESINGER, C., AND KOERT, U. Cation Control in Functional Helical Programming: Structures of a D,L-Peptide Ion Channel. *Angew. Chem., Int. Ed. 41*, 21 (2002), 4062.

[117] GIANNIS, A., AND KOLTER, T. Peptidomimetics for Receptor Ligands—Discovery, Development, and Medical Perspectives. *Angew. Chem., Int. Ed. 32*, 9 (1993), 1244.

[118] IMAMURA, Y., UMEZAWA, N., OSAWA, S., SHIMADA, N., HIGO, T., YOKOSHIMA, S., FUKUYAMA, T., IWATSUBO, T., KATO, N., TOMITA, T., AND HIGUCHI, T. Effect of Helical Conformation and Side Chain Structure on γ-Secretase Inhibition by β-Peptide Foldamers: Insight into Substrate Recognition. *J. Med. Chem. 56*, 4 (2013), 1443.

[119] LEVINTHAL, C. How to fold graciously. *Mössbauer Spectroscopy in Biological Systems, Proceedings of a meeting held at Allerton house, Monticello, Illinois 67*, 41 (1969), 22.

[120] LEVINTHAL, C. Are there pathways for protein folding? *J. Chim. Phys. 65*, 1 (1968), 44.

[121] ZWANZIG, R., SZABO, A., AND BAGCHI, B. Levinthal's paradox. *Proc. Natl. Acad. Sci. U.S.A. 89*, 1 (1992), 20.

[122] DILL, K. A. Theory for the folding and stability of globular proteins. *Biochemistry 24*, 6 (1985), 1501.

[123] BALDWIN, R. L. Matching speed and stability. *Nature 369*, 6477 (1994), 183.

[124] BALDWIN, R. L. The nature of protein folding pathways: The classical versus the new view. *J. Biomol. NMR 5*, 2 (1995), 103.

[125] BRYNGELSON, J. D., ONUCHIC, J. N., SOCCI, N. D., AND WOLYNES, P. G. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins: Struct., Funct., Bioinf. 21*, 3 (1995), 167.

[126] KARPLUS, M. The Levinthal paradox: yesterday and today. *Fold. Des. 2*, 0 (1997), S69.

[127] CHAN, H. S., AND DILL, K. A. Protein Folding in the Landscape Perspective: Chevron Plots and Non-Arrhenius Kinetics. *Proteins: Struct., Funct., Bioinf. 30*, 1 (1998), 2.

[128] ŠALI, A., SHAKHNOVICH, E., AND KARPLUS, M. How does a protein fold? *Nature 369*, 6477 (1994), 248.

[129] BROOKS, C. L., ONUCHIC, J. N., AND WALES, D. J. Taking a Walk on a Landscape. *Science 293*, 5530 (2001), 612.

[130] WALES, D. *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses*. Cambridge University Press, 2003.

[131] KITTEL, C., AND KRÖMER, H. *Thermodynamik: mit 23 Tabellen und 140 Aufgaben*. Oldenbourg Verlag, 2001.

[132] NGUYEN, P. H., STOCK, G., MITTAG, E., HU, C., AND LI, M. S. Free energy landscape and folding mechanism of a β-hairpin in explicit water: A replica exchange molecular dynamics study. *Proteins: Struct., Funct., Bioinf. 61*, 4 (2005), 795.

[133] POULAIN, P., CALVO, F., ANTOINE, R., BROYER, M., AND DUGOURD, P. Competition between secondary structures in gas phase polyalanines. *Europhys. Lett. 79*, 6 (2007), 66003.

[134] DREIZLER, R. M., AND E.K.U, G. *Density Functional Theory: An Approach to the Quantum Many-Body Problem*. Springer-Verlag, 1991.

[135] SZABO, A., AND OSTLUND, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, new ed. Dover Pubn Inc, 1996.

[136] KOHANOFF, J. *Electronic Structure Calculations for Solids and Molecules: Theory and Computational Methods*. Cambridge University Press, 2006.

[137] MARTIN, R. M. *Electronic Structure: Basic Theory and Practical Methods*, 1 ed. Cambridge University Press, 2008.

[138] MONTICELLI, L., AND TIELEMAN, D. P. Force Fields for Classical Molecular Dynamics. In *Biomolecular Simulations*, vol. 924. Humana Press, Totowa, NJ, 2013, p. 197.

[139] JORGENSEN, W. L., MAXWELL, D. S., AND 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. 118*, 45 (1996), 11225.

[140] JORGENSEN, W. L., MADURA, J. D., AND SWENSON, C. J. Optimized intermolecular potential functions for liquid hydrocarbons. *J. Am. Chem. Soc. 106*, 22 (1984), 6638.

[141] JORGENSEN, W. L., AND TIRADO-RIVES, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc. 110*, 6 (1988), 1657.

[142] WEINER, S. J., KOLLMAN, P. A., NGUYEN, D. T., AND CASE, D. A. An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem. 7*, 2 (1986), 230.

[143] KAMINSKI, G. A., FRIESNER, R. A., TIRADO-RIVES, J., AND JORGENSEN, W. L. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B 105*, 28 (2001), 6474.

[144] CORNELL, W. D., CIEPLAK, P., BAYLY, C. I., GOULD, I. R., MERZ, K. M., FERGUSON, D. M., SPELLMEYER, D. C., FOX, T., CALDWELL, J. W., AND KOLLMAN, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc. 117*, 19 (1995), 5179.

[145] BROOKS, B. R., BRUCCOLERI, R. E., OLAFSON, B. D., STATES, D. J., SWAMINATHAN, S., AND KARPLUS, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem. 4*, 2 (1983), 187.

[146] BROOKS, B. R., BROOKS, C. L., MACKERELL, A. D., NILSSON, L., PETRELLA, R. J., ROUX, B., WON, Y., ARCHONTIS, G., BARTELS, C., BORESCH, S., CAFLISCH, A., CAVES, L., CUI, Q., DINNER, A. R., FEIG, M., FISCHER, S., GAO, J., HODOSCEK, M., IM, W., KUCZERA, K., LAZARIDIS, T., MA, J., OVCHINNIKOV, V., PACI, E., PASTOR, R. W., POST, C. B., PU, J. Z., SCHAEFER, M., TIDOR, B., VENABLE, R. M., WOODCOCK, H. L., WU, X., YANG, W., YORK, D. M., AND KARPLUS, M. CHARMM: The biomolecular simulation program. *J. Comput. Chem. 30*, 10 (2009), 1545.

[147] ZHU, X., LOPES, P. E. M., AND MACKERELL, A. D. Recent developments and applications of the CHARMM force fields. *Wiley Interdiscip. Rev. Comput. Mol. Sci. 2*, 1 (2012), 167.

[148] PONDER, J. W., WU, C., REN, P., PANDE, V. S., CHODERA, J. D., SCHNIEDERS, M. J., HAQUE, I., MOBLEY, D. L., LAMBRECHT, D. S., DISTASIO, R. A., HEAD-GORDON, M., CLARK, G. N. I., JOHNSON, M. E., AND HEAD-GORDON, T. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B 114*, 8 (2010), 2549.

[149] LINDORFF-LARSEN, K., MARAGAKIS, P., PIANA, S., EASTWOOD, M. P., DROR, R. O., AND SHAW, D. E. Systematic Validation of Protein Force Fields against Experimental Data. *PLoS ONE 7*, 2 (2012), e32131.

[150] CINO, E. A., CHOY, W.-Y., AND KARTTUNEN, M. Comparison of Secondary Structure Formation Using 10 Different Force Fields in Microsecond Molecular Dynamics Simulations. *J. Chem. Theory Comput. 8*, 8 (2012), 2725.

[151] BEAUCHAMP, K. A., LIN, Y.-S., DAS, R., AND PANDE, V. S. Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements. *J. Chem. Theory Comput. 8*, 4 (2012), 1409.

[152] LANGE, O. F., VAN DER SPOEL, D., AND DE GROOT, B. L. Scrutinizing Molecular Mechanics Force Fields on the Submicrosecond Timescale with NMR Data. *Biophys. J. 99*, 2 (2010), 647.

[153] HORNAK, V., ABEL, R., OKUR, A., STROCKBINE, B., ROITBERG, A., AND SIMMERLING, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Bioinf. 65*, 3 (2006), 712.

[154] SHI, Y., XIA, Z., ZHANG, J., BEST, R., WU, C., PONDER, J. W., AND REN, P. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput. 9*, 9 (2013), 4046.

[155] PONDER, J. W. Tinker - software tools for molecular design. http://dasher.wustl.edu/ffe/. Unless stated otherwise we used version 5.0 of the program and the versions of the force fields distributed with the package.

[156] BONOMI, M., BRANDUARDI, D., BUSSI, G., CAMILLONI, C., PROVASI, D., RAITERI, P., DONADIO, D., MARINELLI, F., PIETRUCCI, F., BROGLIA, R. A., AND PARRINELLO, M. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun. 180*, 10 (2009), 1961.

[157] SCHRÖDINGER, E. Quantisierung als Eigenwertproblem. *Ann. d. Physik 79*, 4 (1926), 361.

[158] DIRAC, P. A. M. The Quantum Theory of the Electron. *Proc. R. Soc. Lond. A 117*, 778 (1928), 610.

[159] DIRAC, P. A. M. A Theory of Electrons and Protons. *Proc. R. Soc. Lond. A 126*, 801 (1930), 360.

[160] BORN, M., AND OPPENHEIMER, R. Zur Quantentheorie der Molekeln. *Ann. d. Physik 84*, 20 (1927), 457.

[161] HARTREE, D. R. The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods. *Proc. Cambridge Phil. Soc. 24*, 01 (1928), 89.

[162] SLATER, J. C. The Self Consistent Field and the Structure of Atoms. *Phys. Rev. 32*, 3 (1928), 339.

[163] SLATER, J. C. Note on Hartree's Method. *Phys. Rev. 35*, 2 (1930), 210.

[164] FOCK, V. Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Z. Physik 61*, 1-2 (1930), 126.

[165] KOOPMANS, T. Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den einzelnen Elektronen eines Atoms. *Physica 1*, 1–6 (1934), 104.

[166] MØLLER, C., AND PLESSET, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev. 46*, 7 (1934), 618.

[167] SCHWABL, F. *Quantenmechanik (QM I)*. Springer DE, 2002.

[168] COESTER, F., AND KÜMMEL, H. Short-range correlations in nuclear wave functions. *Nucl. Phys. 17* (1960), 477.

[169] ČÍŽEK, J. On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods. *J. Chem. Phys. 45*, 11 (1966), 4256.

[170] BARTLETT, R. J., AND MUSIAŁ, M. Coupled-cluster theory in quantum chemistry. *Rev. Mod. Phys. 79*, 1 (2007), 291.

[171] RILEY, K. E., PITOŇÁK, M., JUREČKA, P., AND HOBZA, P. Stabilization and Structure Calculations for Noncovalent Interactions in Extended Molecular Systems Based on Wave Function and Density Functional Theories. *Chem. Rev. 110*, 9 (2010), 5023.

[172] ŘEZÁČ, J., AND HOBZA, P. Describing Noncovalent Interactions beyond the Common Approximations: How Accurate Is the "Gold Standard," CCSD(T) at the Complete Basis Set Limit? *J. Chem. Theory Comput. 9*, 5 (2013), 2151.

[173] JUREČKA, P., ŠPONER, J., ČERNÝ, J., AND HOBZA, P. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys. 8*, 17 (2006), 1985.

[174] ŘEZÁČ, J., RILEY, K. E., AND HOBZA, P. S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput. 7*, 8 (2011), 2427.

[175] BARTLETT, R. J. Coupled-cluster approach to molecular structure and spectra: a step toward predictive quantum chemistry. *J. Phys. Chem. 93*, 5 (1989), 1697.

[176] HOHENBERG, P., AND KOHN, W. Inhomogeneous Electron Gas. *Phys. Rev. 136*, 3B (1964), B864.

[177] LEVY, M. Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem. *Proc. Natl. Acad. Sci. U.S.A. 76*, 12 (1979), 6062.

[178] LEVY, M. Electron densities in search of Hamiltonians. *Phys. Rev. A 26*, 3 (1982), 1200.

[179] KOHN, W., AND SHAM, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev. 140*, 4A (1965), A1133.

[180] PERDEW, J. P., AND SCHMIDT, K. Jacob's ladder of density functional approximations for the exchange-correlation energy. *AIP Conf. Proc. 577*, 1 (2001), 1.

[181] FIOLHAIS, C., NOGUEIRA, F., AND MARQUES, M. A. L. *A Primer in Density Functional Theory*. Springer, Berlin, 2003.

[182] BURKE, K. *The ABC of DFT*. Department of Chemistry, University of California, Irvine, CA 92697, http://www.chem.uci.edu/~kieron/dftold2/literature.php [Accessed in June 2013], 2007.

[183] WIGNER, E. Effects of the electron interaction on the energy levels of electrons in metals. *Trans. Faraday Soc. 34*, 0 (1938), 678.

[184] GELL-MANN, M., AND BRUECKNER, K. A. Correlation Energy of an Electron Gas at High Density. *Phys. Rev. 106*, 2 (1957), 364.

[185] CEPERLEY, D. M., AND ALDER, B. J. Ground State of the Electron Gas by a Stochastic Method. *Phys. Rev. Lett. 45*, 7 (1980), 566.

[186] PERDEW, J. P., AND ZUNGER, A. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B 23*, 10 (1981), 5048.

[187] PERDEW, J. P., AND WANG, Y. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B 45*, 23 (1992), 13244.

[188] VOSKO, S. H., WILK, L., AND NUSAIR, M. Accurate spin-dependent electron liquid correlation energies for local spin-density calculations: a critical analysis. *Can. J. Phys. 58*, 8 (1980), 1200.

[189] MA, S.-K., AND BRUECKNER, K. A. Correlation Energy of an Electron Gas with a Slowly Varying High Density. *Phys. Rev. 165*, 1 (1968), 18.

[190] CAPELLE, K. A bird's-eye view of density-functional theory. *Braz. J. Phys. 36*, 4A (2006), 1318.

[191] LANGRETH, D. C., AND PERDEW, J. P. Theory of nonuniform electronic systems. I. Analysis of the gradient approximation and a generalization that works. *Phys. Rev. B 21*, 12 (1980), 5469.

[192] PERDEW, J. P. Accurate Density Functional for the Energy: Real-Space Cutoff of the Gradient Expansion for the Exchange Hole. *Phys. Rev. Lett. 55*, 16 (1985), 1665.

[193] PERDEW, J. P., AND RUZSINSZKY, A. Fourteen easy lessons in density functional theory. *Int. J. Quantum Chem. 110*, 15 (2010), 2801.

[194] LANGRETH, D. C., AND MEHL, M. J. Easily Implementable Nonlocal Exchange-Correlation Energy Functional. *Phys. Rev. Lett. 47*, 6 (1981), 446.

[195] BECKE, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A 38*, 6 (1988), 3098.

[196] LEE, C., YANG, W., AND PARR, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B 37*, 2 (1988), 785.

[197] LEE, C., CHEN, H., AND FITZGERALD, G. Structures of the water hexamer using density functional methods. *J. Chem. Phys. 101*, 5 (1994), 4472.

[198] HAMANN, D. R. H_2O hydrogen bonding in density-functional theory. *Phys. Rev. B 55*, 16 (1997), R10157.

[199] BECKE, A. D. A new mixing of Hartree–Fock and local density-functional theories. *J. Chem. Phys. 98*, 2 (1993), 1372.

[200] HARRIS, J. Adiabatic-connection approach to Kohn-Sham theory. *Phys. Rev. A 29*, 4 (1984), 1648.

[201] LANGRETH, D. C., AND PERDEW, J. P. The exchange-correlation energy of a metallic surface. *Solid State Comm. 17*, 11 (1975), 1425.

[202] LANGRETH, D. C., AND PERDEW, J. P. Exchange-correlation energy of a metallic surface: Wave-vector analysis. *Phys. Rev. B 15*, 6 (1977), 2884.

[203] GUNNARSSON, O., AND LUNDQVIST, B. I. Exchange and correlation in atoms, molecules, and solids by the spin-density-functional formalism. *Phys. Rev. B 13*, 10 (1976), 4274.

[204] PERDEW, J. P., ERNZERHOF, M., AND BURKE, K. Rationale for mixing exact exchange with density functional approximations. *J. Chem. Phys. 105*, 22 (1996), 9982.

[205] ERNZERHOF, M., AND SCUSERIA, G. E. Assessment of the Perdew–Burke–Ernzerhof exchange-correlation functional. *J. Chem. Phys. 110*, 11 (1999), 5029.

[206] ADAMO, C., AND BARONE, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys. 110*, 13 (1999), 6158.

[207] BECKE, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys. 98*, 7 (1993), 5648.

[208] STEPHENS, P. J., DEVLIN, F. J., CHABALOWSKI, C. F., AND FRISCH, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem. 98*, 45 (1994), 11623.

[209] GILL, P. M. W., JOHNSON, B. G., POPLE, J. A., AND FRISCH, M. J. An investigation of the performance of a hybrid of Hartree-Fock and density functional theory. *Int. J. Quantum Chem. 44*, S26 (1992), 319.

[210] CURTISS, L. A., RAGHAVACHARI, K., TRUCKS, G. W., AND POPLE, J. A. Gaussian-2 theory for molecular energies of first- and second-row compounds. *J. Chem. Phys. 94*, 11 (1991), 7221.

[211] GRIMME, S. Density functional theory with London dispersion corrections. *WIREs Comput. Mol. Sci. 1*, 2 (2011), 211.

[212] KAPLAN, I. G. *Intermolecular Interactions*. John Wiley & Sons. Ltd, 2006.

[213] CASIMIR, H. B. G., AND POLDER, D. The Influence of Retardation on the London-van der Waals Forces. *Phys. Rev. 73*, 4 (1948), 360.

[214] PÉREZ-JORDÁ, J. M., SAN-FABIÁN, E., AND PÉREZ-JIMÉNEZ, A. J. Density-functional study of van der Waals forces on rare-gas diatomics: Hartree–Fock exchange. *J. Chem. Phys. 110*, 4 (1999), 1916.

[215] PÉREZ-JORDÁ, J., AND BECKE, A. A density-functional study of van der Waals forces: rare gas diatomics. *Chem. Phys. Lett. 233*, 1–2 (1995), 134.

[216] KRISTYÁN, S., AND PULAY, P. Can (semi)local density functional theory account for the London dispersion forces? *Chem. Phys. Lett. 229*, 3 (1994), 175.

[217] HOBZA, P., ŠPONER, J., AND RESCHEL, T. Density functional theory and molecular clusters. *J. Comput. Chem. 16*, 11 (1995), 1315.

[218] KURITA, N., AND SEKINO, H. Ab initio and DFT studies for accurate description of van der Waals interaction between He atoms. *Chem. Phys. Lett. 348*, 1–2 (2001), 139.

[219] ZHAO, Y., AND TRUHLAR, D. G. A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *J. Chem. Phys. 125*, 19 (2006), 194101.

[220] ZHAO, Y., AND TRUHLAR, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Account 120*, 1-3 (2008), 215.

[221] ZHAO, Y., AND TRUHLAR, D. G. Density Functionals with Broad Applicability in Chemistry. *Acc. Chem. Res. 41*, 2 (2008), 157.

[222] ZHAO, Y., AND TRUHLAR, D. G. Exploring the Limit of Accuracy of the Global Hybrid Meta Density Functional for Main-Group Thermochemistry, Kinetics, and Noncovalent Interactions. *J. Chem. Theory Comput. 4*, 11 (2008), 1849.

[223] MAROM, N., TKATCHENKO, A., ROSSI, M., GOBRE, V. V., HOD, O., SCHEFFLER, M., AND KRONIK, L. Dispersion Interactions with Density-Functional Theory: Benchmarking Semiempirical and Interatomic Pairwise Corrected Density Functionals. *J. Chem. Theory Comput. 7*, 12 (2011), 3944.

[224] DION, M., RYDBERG, H., SCHRÖDER, E., LANGRETH, D. C., AND LUNDQVIST, B. I. Van der Waals Density Functional for General Geometries. *Phys. Rev. Lett. 92*, 24 (2004), 246401.

[225] LEE, K., MURRAY, É. D., KONG, L., LUNDQVIST, B. I., AND LANGRETH, D. C. Higher-accuracy van der Waals density functional. *Phys. Rev. B 82*, 8 (2010), 081101.

[226] ELSTNER, M., HOBZA, P., FRAUENHEIM, T., SUHAI, S., AND KAXIRAS, E. Hydrogen bonding and stacking interactions of nucleic acid base pairs: A density-functional-theory based treatment. *J. Chem. Phys. 114*, 12 (2001), 5149.

[227] WU, Q., AND YANG, W. Empirical correction to density functional theory for van der Waals interactions. *J. Chem. Phys. 116*, 2 (2002), 515.

[228] JUREČKA, P., ČERNÝ, J., HOBZA, P., AND SALAHUB, D. R. Density functional theory augmented with an empirical dispersion term. Interaction energies and geometries of 80 noncovalent complexes compared with ab initio quantum mechanics calculations. *J. Comput. Chem. 28*, 2 (2007), 555.

[229] GRIMME, S. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J. Comput. Chem. 25*, 12 (2004), 1463.

[230] GRIMME, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem. 27*, 15 (2006), 1787.

[231] GRIMME, S., ANTONY, J., EHRLICH, S., AND KRIEG, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys. 132*, 15 (2010), 154104.

[232] BECKE, A. D., AND JOHNSON, E. R. Exchange-hole dipole moment and the dispersion interaction. *J. Chem. Phys. 122*, 15 (2005), 154104.

[233] BECKE, A. D., AND JOHNSON, E. R. A density-functional model of the dispersion interaction. *J. Chem. Phys. 123*, 15 (2005), 154101.

[234] ORTMANN, F., BECHSTEDT, F., AND SCHMIDT, W. G. Semiempirical van der Waals correction to the density functional description of solids and molecular structures. *Phys. Rev. B 73*, 20 (2006), 205101.

[235] SATO, T., AND NAKAI, H. Density functional method including weak interactions: Dispersion coefficients based on the local response approximation. *J. Chem. Phys. 131*, 22 (2009), 224104.

[236] CHU, X., AND DALGARNO, A. Linear response time-dependent density functional theory for van der Waals coefficients. *J. Chem. Phys. 121*, 9 (2004), 4083.

[237] HIRSHFELD, F. L. Bonded-atom fragments for describing molecular charge densities. *Theoret. Chim. Acta 44*, 2 (1977), 129.

[238] TKATCHENKO, A., DISTASIO, R. A., CAR, R., AND SCHEFFLER, M. Accurate and Efficient Method for Many-Body van der Waals Interactions. *Phys. Rev. Lett. 108*, 23 (2012), 236402.

[239] AMBROSETTI, A., REILLY, A. M., DISTASIO, R. A., AND TKATCHENKO, A. Long-range correlation energy calculated from coupled atomic response functions. *J. Chem. Phys. 140*, 18 (2014), 18A508.

[240] DONCHEV, A. G. Many-body effects of dispersion interaction. *J. Chem. Phys. 125*, 7 (2006), 074713.

[241] KUBO, R. The fluctuation-dissipation theorem. *Rep. Prog. Phys. 29*, 1 (1966), 255.

[242] REN, X., RINKE, P., JOAS, C., AND SCHEFFLER, M. Random-phase approximation and its applications in computational chemistry and materials science. *J. Mater. Sci. 47*, 21 (2012), 7447.

[243] TKATCHENKO, A., AMBROSETTI, A., AND DISTASIO, R. A. Interatomic methods for the dispersion energy derived from the adiabatic connection fluctuation-dissipation theorem. *J. Chem. Phys. 138*, 7 (2013), 074106.

[244] THOLE, B. T. Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys. 59*, 3 (1981), 341.

[245] FELDERHOF, B. U. On the propagation and scattering of light in fluids. *Physica 76*, 3 (1974), 486.

[246] OXTOBY, D. W., AND GELBART, W. M. Collisional polarizability anisotropies of the noble gases. *Mol. Phys. 29*, 5 (1975), 1569.

[247] ROSSI, M., CHUTIA, S., SCHEFFLER, M., AND BLUM, V. Validation Challenge of Density-Functional Theory for Peptides - Example of Ac-Phe-Ala_5-$LysH^+$. *J. Phys. Chem. A 118*, 35 (2014), 7349.

[248] VALDES, H., PLUHÁČKOVÁ, K., PITONÁK, M., ŘEZÁČ, J., AND HOBZA, P. Benchmark database on isolated small peptides containing an aromatic side chain: comparison between wave function and density functional theory methods and empirical force field. *Phys. Chem. Chem. Phys. 10*, 19 (2008), 2747.

[249] DISTASIO, R. A., STEELE, R. P., RHEE, Y. M., SHAO, Y., AND HEAD-GORDON, M. An improved algorithm for analytical gradient evaluation in resolution-of-the-identity second-order Møller-Plesset perturbation theory: Application to alanine tetrapeptide conformational analysis. *J. Comput. Chem. 28*, 5 (2007), 839.

[250] STEARNS, J. A., BOYARKIN, O. V., AND RIZZO, T. R. Spectroscopic Signatures of Gas-Phase Helices: Ac-Phe-$(Ala)_5$-Lys-H^+ and Ac-Phe-$(Ala)_{10}$-Lys-H^+. *J. Am. Chem. Soc. 129*, 45 (2007), 13820.

[251] STEARNS, J. A., SEAIBY, C., BOYARKIN, O. V., AND RIZZO, T. R. Spectroscopy and conformational preferences of gas-phase helices. *Phys. Chem. Chem. Phys. 11*, 1 (2009), 125.

[252] XIE, Y., SCHAEFER, H. F., SILAGHI-DUMITRESCU, R., PENG, B., LI, Q.-S., STEARNS, J. A., AND RIZZO, T. R. Conformational Preferences of Gas-Phase Helices: Experiment and Theory Struggle to Agree: The Seven-Residue Peptide Ac-Phe-$(Ala)_5$-Lys-H^+. *Chem. Eur. J. 18*, 41 (2012), 12941.

[253] KRESSE, G., AND HAFNER, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B 47*, 1 (1993), 558.

[254] SEGALL, M. D., LINDAN, P. J. D., PROBERT, M. J., PICKARD, C. J., HASNIP, P. J., CLARK, S. J., AND PAYNE, M. C. First-principles simulation: ideas, illustrations and the CASTEP code. *J. Phys.: Condens. Matter 14*, 11 (2002), 2717.

[255] VALIEV, M., BYLASKA, E., GOVIND, N., KOWALSKI, K., STRAATSMA, T., VAN DAM, H., WANG, D., NIEPLOCHA, J., APRA, E., WINDUS, T., AND DE JONG, W. NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Comput. Phys. Commun. 181*, 9 (2010), 1477.

[256] TURBOMOLE V6.5, 2013. A development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from http://www.turbomole.com.

[257] BLUM, V., GEHRKE, R., HANKE, F., HAVU, P., HAVU, V., REN, X., REUTER, K., AND SCHEFFLER, M. Ab initio molecular simulations with numeric atom-centered orbitals. *Comput. Phys. Commun. 180*, 11 (2009), 2175.

[258] DELLEY, B. An all-electron numerical method for solving the local density functional for polyatomic molecules. *J. Chem. Phys. 92*, 1 (1990), 508.

[259] DELLEY, B. High order integration schemes on the unit sphere. *J. Comput. Chem. 17*, 9 (1996), 1152.

[260] REN, X., RINKE, P., BLUM, V., WIEFERINK, J., TKATCHENKO, A., SANFILIPPO, A., REUTER, K., AND SCHEFFLER, M. Resolution-of-identity approach to Hartree–Fock, hybrid density functionals, RPA, MP2 and GW with numeric atom-centered orbital basis functions. *New J. Phys. 14*, 5 (2012), 053020.

[261] BOYS, S., AND BERNARDI, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys. 19*, 4 (1970), 553.

[262] HELLMANN, H. Zur Rolle der kinetischen Elektronenenergie für die zwischenatomaren Kräfte. *Z. Physik 85*, 3-4 (1933), 180.

[263] FEYNMAN, R. P. Forces in Molecules. *Phys. Rev. 56*, 4 (1939), 340.

[264] PULAY, P. Ab initio calculation of force constants and equilibrium geometries in polyatomic molecules. *Mol. Phys. 17*, 2 (1969), 197.

[265] GEHRKE, R. *First-principles basin-hopping for the structure determination of atomic clusters*. Ph.D. thesis, FU Berlin and Fritz-Haber-Institut der Max-Planck-Gesellschaft, 2009.

[266] FRENKEL, D., AND SMIT, B. *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, 2001.

[267] SWOPE, W. C., ANDERSEN, H. C., BERENS, P. H., AND WILSON, K. R. 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. 76*, 1 (1982), 637.

[268] MCQUARRIE, D. A. *Statistical mechanics*. University Science Books, Mill Valley, Calif., 2000.

[269] BERENDSEN, H. J. C., POSTMA, J. P. M., VAN GUNSTEREN, W. F., DINOLA, A., AND HAAK, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys. 81*, 8 (1984), 3684.

[270] LEMAK, A. S., AND BALABAEV, N. K. On The Berendsen Thermostat. *Mol. Simulat. 13*, 3 (1994), 177.

[271] ANDERSEN, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys. 72*, 4 (1980), 2384.

[272] NOSÉ, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys. 81* (1984), 511.

[273] HOOVER, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A 31*, 3 (1985), 1695.

[274] MARTYNA, G. J., KLEIN, M. L., AND TUCKERMAN, M. Nosé–Hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys. 97*, 4 (1992), 2635.

[275] TOXVAERD, S., AND OLSEN, O. H. Canonical Molecular Dynamics of Molecules with Internal Degrees of Freedom. *Ber. Bunsenges. Phys. Chem. 94*, 3 (1990), 274.

[276] BUSSI, G., DONADIO, D., AND PARRINELLO, M. Canonical sampling through velocity rescaling. *J. Chem. Phys. 126*, 1 (2007), 014101.

[277] TORRIE, G., AND VALLEAU, J. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys. 23*, 2 (1977), 187.

[278] KIRKPATRICK, S., JR., D. G., AND VECCHI, M. P. Optimization by Simulated Annealing. *Science 220*, 4598 (1983), 671.

[279] GRUBMÜLLER, H. Predicting slow structural transitions in macromolecular systems: Conformational flooding. *Phys. Rev. E 52*, 3 (1995), 2893.

[280] DAMSBO, M., KINNEAR, B. S., HARTINGS, M. R., RUHOFF, P. T., JARROLD, M. F., AND RATNER, M. A. Application of evolutionary algorithm methods to polypeptide folding: Comparison with experimental results for unsolvated Ac-(Ala-Gly-Gly)$_5$-LysH$^+$. *Proc. Natl. Acad. Sci. U.S.A. 101*, 19 (2004), 7215.

[281] GOEDECKER, S. Minima hopping: An efficient search method for the global minimum of the potential energy surface of complex molecular systems. *J. Chem. Phys. 120*, 21 (2004), 9911.

[282] LAIO, A., AND GERVASIO, F. L. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep. Prog. Phys. 71*, 12 (2008), 126601.

[283] MARINARI, E., AND PARISI, G. Simulated Tempering: A New Monte Carlo Scheme. *Europhys. Lett. 19*, 6 (1992), 451.

[284] WALES, D. J., AND DOYE, J. P. K. Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. *J. Phys. Chem. A 101*, 28 (1997), 5111.

[285] DOYE, J. P. K., AND WALES, D. J. Thermodynamics of Global Optimization. *Phys. Rev. Lett. 80*, 7 (1998), 1357.

[286] DOYE, J. P. K., WALES, D. J., AND MILLER, M. A. Thermodynamics and the global optimization of Lennard-Jones clusters. *J. Chem. Phys. 109*, 19 (1998), 8143.

[287] GEHRKE, R., AND REUTER, K. Assessing the efficiency of first-principles basin-hopping sampling. *Phys. Rev. B 79*, 8 (2009), 085412.

[288] HOFFMANN, K. H., FRANZ, A., AND SALAMON, P. Structure of best possible strategies for finding ground states. *Phys. Rev. E 66*, 4 (2002), 046706.

[289] CHUTIA, S., ROSSI, M., AND BLUM, V. Water Adsorption at Two Unsolvated Peptides with a Protonated Lysine Residue: From Self-Solvation to Solvation. *J. Phys. Chem. B 116*, 51 (2012), 14788.

[290] SWENDSEN, R. H., AND WANG, J.-S. Replica Monte Carlo Simulation of Spin-Glasses. *Phys. Rev. Lett. 57*, 21 (1986), 2607.

[291] HUKUSHIMA, K., AND NEMOTO, K. Exchange Monte Carlo Method and Application to Spin Glass Simulations. *J. Phys. Soc. Jpn. 65* (1996), 1604.

[292] HANSMANN, U. H. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett. 281*, 1–3 (1997), 140.

[293] SUGITA, Y., AND OKAMOTO, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett. 314*, 1-2 (1999), 141.

[294] EARL, D. J., AND DEEM, M. W. Parallel tempering: Theory, applications, and new perspectives. *Phys. Chem. Chem. Phys. 7*, 23 (2005), 3910.

[295] METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H., AND TELLER, E. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys. 21*, 6 (1953), 1087.

[296] HEINE, N., YACOVITCH, T. I., SCHUBERT, F., BRIEGER, C., NEUMARK, D. M., AND ASMIS, K. R. Infrared Photodissociation Spectroscopy of Microhydrated Nitrate-Nitric Acid Clusters $NO_3^-(HNO_3)_m(H_2O)_n$. *J. Phys. Chem. A 118*, 35 (2014), 7613.

[297] PATRIKSSON, A., AND VAN DER SPOEL, D. A temperature predictor for parallel tempering simulations. *Phys. Chem. Chem. Phys. 10*, 15 (2008), 2073.

[298] KOFKE, D. A. On the acceptance probability of replica-exchange Monte Carlo trials. *J. Chem. Phys. 117*, 15 (2002), 6911.

[299] KOFKE, D. A. Erratum: "On the acceptance probability of replica-exchange Monte Carlo trials" [J. Chem. Phys. 117, 6911 (2002)]. *J. Chem. Phys. 120*, 22 (2004), 10852.

[300] SABO, D., MEUWLY, M., FREEMAN, D. L., AND DOLL, J. D. A constant entropy increase model for the selection of parallel tempering ensembles. *J. Chem. Phys. 128*, 17 (2008), 174109.

[301] SINDHIKARA, D. J., EMERSON, D. J., AND ROITBERG, A. E. Exchange Often and Properly in Replica Exchange Molecular Dynamics. *J. Chem. Theory Comput. 6*, 9 (2010), 2804.

[302] WILSON, E. B., DECIUS, J., AND CROSS, P. *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*, new ed. Courier Dover Publications, 2003.

[303] GAIGEOT, M.-P., MARTINEZ, M., AND VUILLEUMIER, R. Infrared spectroscopy in the gas and liquid phase from first principle molecular dynamics simulations: application to small peptides. *Mol. Phys. 105*, 19-22 (2007), 2857.

[304] NEUGEBAUER, J., REIHER, M., KIND, C., AND HESS, B. A. Quantum chemical calculation of vibrational spectra of large molecules—Raman and IR spectra for Buckminsterfullerene. *J. Comput. Chem. 23*, 9 (2002), 895.

[305] BARTH, A. Infrared spectroscopy of proteins. *Biochim. Biophys. Acta 1767*, 9 (2007), 1073.

[306] BARTH, A., AND ZSCHERP, C. What vibrations tell about proteins. *Q. Rev. Biophys. 35*, 4 (2002), 369.

[307] OBERG, K. A., RUYSSCHAERT, J.-M., AND GOORMAGHTIGH, E. The optimization of protein secondary structure determination with infrared and circular dichroism spectra. *Eur. J. Biochem. 271*, 14 (2004), 2937.

[308] FU, F.-N., DEOLIVEIRA, D. B., TRUMBLE, W. R., SARKAR, H. K., AND SINGH, B. R. Secondary Structure Estimation of Proteins Using the Amide III Region of Fourier Transform Infrared Spectroscopy: Application to Analyze Calcium-Binding-Induced Structural Changes in Calsequestrin. *Appl. Spectrosc. 48*, 11 (1994), 1432.

[309] CAI, S., AND SINGH, B. R. Identification of β-turn and random coil amide III infrared bands for secondary structure estimation of proteins. *Biophys. Chem. 80*, 1 (1999), 7.

[310] CAI, S., AND SINGH, B. R. A Distinct Utility of the Amide III Infrared Band for Secondary Structure Estimation of Aqueous Protein Solutions Using Partial Least Squares Methods. *Biochemistry 43*, 9 (2004), 2541.

[311] WEYMUTH, T., JACOB, C. R., AND REIHER, M. A Local-Mode Model for Understanding the Dependence of the Extended Amide III Vibrations on Protein Secondary Structure. *J. Phys. Chem. B 114*, 32 (2010), 10649.

[312] GAIGEOT, M.-P. Theoretical spectroscopy of floppy peptides at room temperature. A DFTMD perspective: gas and aqueous phase. *Phys. Chem. Chem. Phys. 12*, 14 (2010), 3336.

[313] SCHMITZ, M., AND TAVAN, P. Vibrational spectra from atomic fluctuations in dynamics simulations. II. Solvent-induced frequency fluctuations at femtosecond time resolution. *J. Chem. Phys. 121*, 24 (2004), 12247.

[314] BARONE, V. *Computational strategies for spectroscopy from small molecules to nano systems*. John Wiley & Sons, Hoboken, N.J., 2012.

[315] CAO, J., AND VOTH, G. A. The formulation of quantum statistical mechanics based on the Feynman path centroid density. II. Dynamical properties. *J. Chem. Phys. 100*, 7 (1994), 5106.

[316] CRAIG, I. R., AND MANOLOPOULOS, D. E. Quantum statistics and classical mechanics: Real time correlation functions from ring polymer molecular dynamics. *J. Chem. Phys. 121*, 8 (2004), 3368.

[317] KUBO, R. Statistical-Mechanical Theory of Irreversible Processes. I. General Theory and Simple Applications to Magnetic and Conduction Problems. *J. Phys. Soc. Jpn. 12*, 6 (1957), 570.

[318] HABERSHON, S., MANOLOPOULOS, D. E., MARKLAND, T. E., AND MILLER, T. F. Ring-Polymer Molecular Dynamics: Quantum Effects in Chemical Dynamics from Classical Trajectories in an Extended Phase Space. *Annu. Rev. Phys. Chem. 64*, 1 (2013), 387.

[319] RAMÍREZ, R., LÓPEZ-CIUDAD, T., KUMAR P, P., AND MARX, D. Quantum corrections to classical time-correlation functions: Hydrogen bonding and anharmonic floppy modes. *J. Chem. Phys. 121*, 9 (2004), 3973.

[320] AHLBORN, H., SPACE, B., AND MOORE, P. B. The effect of isotopic substitution and detailed balance on the infrared spectroscopy of water: A combined time correlation function and instantaneous normal mode analysis. *J. Chem. Phys. 112*, 18 (2000), 8083.

[321] GAIGEOT, M.-P., AND SPRIK, M. Ab Initio Molecular Dynamics Computation of the Infrared Spectrum of Aqueous Uracil. *J. Phys. Chem. B 107*, 38 (2003), 10344.

[322] IFTIMIE, R., AND TUCKERMAN, M. E. Decomposing total IR spectra of aqueous systems into solute and solvent contributions: A computational approach using maximally localized Wannier orbitals. *J. Chem. Phys. 122*, 21 (2005), 214508.

[323] AIDA, M., AND DUPUIS, M. IR and Raman intensities in vibrational spectra from direct ab initio molecular dynamics: D_2O as an illustration. *J. Mol. Struc-Theochem. 633*, 2–3 (2003), 247.

[324] MARX, D., AND HUTTER, J. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*. Cambridge University Press, 2009.

[325] PULAY, P., AND FOGARASI, G. Fock matrix dynamics. *Chem. Phys. Lett. 386*, 4–6 (2004), 272.

[326] HUTTER, J. Car–Parrinello molecular dynamics. *WIREs Comput. Mol. Sci. 2*, 4 (2012), 604.

[327] HERBERT, J. M., AND HEAD-GORDON, M. Accelerated, energy-conserving Born–Oppenheimer molecular dynamics via Fock matrix extrapolation. *Phys. Chem. Chem. Phys. 7*, 18 (2005), 3269.

[328] NIKLASSON, A. M. N., TYMCZAK, C. J., AND CHALLACOMBE, M. Time-Reversible Born-Oppenheimer Molecular Dynamics. *Phys. Rev. Lett. 97*, 12 (2006), 123001.

[329] NIKLASSON, A. M. N., TYMCZAK, C. J., AND CHALLACOMBE, M. Time-reversible ab initio molecular dynamics. *J. Chem. Phys. 126*, 14 (2007), 144103.

[330] NIKLASSON, A. M. N. Extended Born-Oppenheimer Molecular Dynamics. *Phys. Rev. Lett. 100*, 12 (2008), 123004.

[331] KÜHNE, T. D., KRACK, M., MOHAMED, F. R., AND PARRINELLO, M. Efficient and Accurate Car-Parrinello-like Approach to Born-Oppenheimer Molecular Dynamics. *Phys. Rev. Lett. 98*, 6 (2007), 066401.

[332] KOLAFA, J. Numerical Integration of Equations of Motion with a Self-Consistent Field given by an Implicit Equation. *Mol. Simulat. 18*, 3 (1996), 193.

[333] KOLAFA, J. Time-reversible always stable predictor–corrector method for molecular dynamics of polarizable molecules. *J. Comput. Chem. 25*, 3 (2004), 335.

[334] VAN HOVE, M., TONG, S., AND ELCONIN, M. Surface structure refinements of 2H-MoS_2, 2H-$NbSe_2$ and W(100)p(2 × 1)-O via new reliability factors for surface crystallography. *Surf. Sci. 64*, 1 (1977), 85.

[335] ZANAZZI, E., AND JONA, F. A reliability factor for surface structure determinations by low-energy electron diffraction. *Surf. Sci. 62*, 1 (1977), 61.

[336] BLUM, V., AND HEINZ, K. Fast LEED intensity calculations for surface crystallography using Tensor LEED. *Comput. Phys. Commun. 134*, 3 (2001), 392.

[337] PENDRY, J. B. Reliability factors for LEED calculations. *J. Phys. C: Solid State Phys. 13*, 5 (1980), 937.

[338] ROSSI, M., BLUM, V., KUPSER, P., VON HELDEN, G., BIERAU, F., PAGEL, K., MEIJER, G., AND SCHEFFLER, M. Secondary Structure of Ac-Ala_n-$LysH^+$ Polyalanine Peptides (n = 5,10,15) in Vacuo: Helical or Not? *J. Phys. Chem. Lett. 1*, 0 (2010), 3465.

[339] POPLE, J. A., SCHLEGEL, H. B., KRISHNAN, R., DEFREES, D. J., BINKLEY, J. S., FRISCH, M. J., WHITESIDE, R. A., HOUT, R. F., AND HEHRE, W. J. Molecular orbital studies of vibrational frequencies. *Int. J. Quantum Chem. 20*, S15 (1981), 269.

[340] SCOTT, A. P., AND RADOM, L. Harmonic Vibrational Frequencies: An Evaluation of Hartree-Fock, Møller-Plesset, Quadratic Configuration Interaction, Density Functional Theory, and Semiempirical Scale Factors. *J. Phys. Chem. 100*, 41 (1996), 16502.

[341] PAGEL, K., KUPSER, P., BIERAU, F., POLFER, N. C., STEILL, J. D., OOMENS, J., MEIJER, G., KOKSCH, B., AND VON HELDEN, G. Gas-phase IR spectra of intact α-helical coiled coil protein complexes. *Int. J. Mass Spectrom. 283*, 1-3 (2009), 161.

[342] KUPSER, P., PAGEL, K., OOMENS, J., POLFER, N., KOKSCH, B., MEIJER, G., AND HELDEN, G. V. Amide-I and -II Vibrations of the Cyclic β-Sheet Model Peptide Gramicidin S in the Gas Phase. *J. Am. Chem. Soc. 132*, 6 (2010), 2085.

[343] KUPSER, P. *Infrared spectroscopic characterization of secondary structure elements of gas-phase biomolecules*. Ph.D. thesis, FU Berlin and Fritz-Haber-Institut der Max-Planck-Gesellschaft, 2011.

[344] JARROLD, M. F. Unfolding, Refolding, and Hydration of Proteins in the Gas Phase. *Acc. Chem. Res. 32*, 4 (1999), 360.

[345] SIMONS, J. Good vibrations: probing biomolecular structure and interactions through spectroscopy in the gas phase. *Mol. Phys. 107*, 23-24 (2009), 2435.

[346] KOHTANI, M., AND JARROLD, M. F. The Initial Steps in the Hydration of Unsolvated Peptides: Water Molecule Adsorption on Alanine-Based Helices and Globules. *J. Am. Chem. Soc. 124*, 37 (2002), 11148.

[347] KOHTANI, M., AND JARROLD, M. F. Water Molecule Adsorption on Short Alanine Peptides: How Short Is the Shortest Gas-Phase Alanine-Based Helix? *J. Am. Chem. Soc. 126*, 27 (2004), 8454.

[348] WYTTENBACH, T., AND BOWERS, M. T. Hydration of biomolecules. *Chem. Phys. Lett. 480*, 1–3 (2009), 1.

[349] DE VRIES, M. S., AND HOBZA, P. Gas-Phase Spectroscopy of Biomolecular Building Blocks. *Annu. Rev. Phys. Chem. 58*, 1 (2007), 585.

[350] JARROLD, M. F. Peptides and Proteins in the Vapor Phase. *Annu. Rev. Phys. Chem. 51*, 1 (2000), 179.

[351] WEINKAUF, R., SCHERMANN, J.-P., DE VRIES, M. S., AND KLEINERMANNS, K. Molecular physics of building blocks of life under isolated or defined conditions. *Eur. Phys. J. D 20*, 3 (2002), 309.

[352] BENESCH, J. L. P., RUOTOLO, B. T., SIMMONS, D. A., AND ROBINSON, C. V. Protein Complexes in the Gas Phase: Technology for Structural Genomics and Proteomics. *Chem. Rev. 107*, 8 (2007), 3544.

[353] RUOTOLO, B. T., AND ROBINSON, C. V. Aspects of native proteins are retained in vacuum. *Curr. Opin. Chem. Biol. 10*, 5 (2006), 402.

[354] BENESCH, J. L. P., AND ROBINSON, C. V. Biological chemistry: Dehydrated but unharmed. *Nature 462*, 7273 (2009), 576.

[355] POLFER, N. C., AND OOMENS, J. Vibrational spectroscopy of bare and solvated ionic complexes of biological relevance. *Mass. Spectrom. Rev. 28*, 3 (2009), 468.

[356] RIZZO, T. R., STEARNS, J. A., AND BOYARKIN, O. V. Spectroscopic studies of cold, gas-phase biomolecular ions. *Int. Rev. Phys. Chem. 28*, 3 (2009), 481.

[357] FENN, J. B., MANN, M., MENG, C. K., WONG, S. F., AND WHITEHOUSE, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science 246*, 4926 (1989), 64.

[358] KARAS, M., AND HILLENKAMP, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem. 60*, 20 (1988), 2299.

[359] TANAKA, K. The Origin of Macromolecule Ionization by Laser Irradiation (Nobel Lecture). *Angew. Chem., Int. Ed. 42*, 33 (2003), 3860.

[360] DOLE, M., MACK, L. L., HINES, R. L., MOBLEY, R. C., FERGUSON, L. D., AND ALICE, M. B. Molecular Beams of Macroions. *J. Chem. Phys. 49*, 5 (1968), 2240.

[361] IRIBARNE, J. V., AND THOMSON, B. A. On the evaporation of small ions from charged droplets. *J. Chem. Phys. 64*, 6 (1976), 2287.

[362] UETRECHT, C., ROSE, R. J., VAN DUIJN, E., LORENZEN, K., AND HECK, A. J. R. Ion mobility mass spectrometry of proteins and protein assemblies. *Chem. Soc. Rev. 39*, 5 (2010), 1633.

[363] MASON, E. A., AND MCDANIEL, E. W. *Transport properties of ions in gases*. Wiley, New York, 1988.

[364] WYTTENBACH, T., HELDEN, G., BATKA, J. J., CARLAT, D., AND BOWERS, M. T. Effect of the long-range potential on ion mobility measurements. *J. Am. Soc. Mass Spectrom. 8* (1997), 275.

[365] BLEIHOLDER, C., WYTTENBACH, T., AND BOWERS, M. T. A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (I). Method. *Int. J. Mass Spectrom. 308*, 1 (2011), 1.

[366] MESLEH, M. F., HUNTER, J. M., SHVARTSBURG, A. A., SCHATZ, G. C., AND JARROLD, M. F. Structural information from ion mobility measurements: effects of the long-range potential. *J. Phys. Chem. 100*, 40 (1996), 16082.

[367] MESLEH, M. F., HUNTER, J. M., SHVARTSBURG, A. A., SCHATZ, G. C., AND JARROLD, M. F. Structural Information from Ion Mobility Measurements: Effects of the Long-Range Potential. *J. Phys. Chem. A 101*, 5 (1997), 968.

[368] MOBCAL - A Program to Calculate Mobilities. http://www.indiana.edu/~nano/software.html. Downloaded in April 2013.

[369] SHVARTSBURG, A. A., AND JARROLD, M. F. An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chem. Phys. Lett. 261*, 1–2 (1996), 86.

[370] VIEHLAND, L., MASON, E., MORRISON, W., AND FLANNERY, M. Tables of transport collision integrals for (n, 6, 4) ion-neutral potentials. *At. Data Nucl. Data Tables 16*, 6 (1975), 495.

[371] OOMENS, J., SARTAKOV, B. G., MEIJER, G., AND VON HELDEN, G. Gas-phase infrared multiple photon dissociation spectroscopy of mass-selected molecular ions. *Int. J. Mass Spectrom. 254*, 1–2 (2006), 1.

[372] OOMENS, J., ROIJ, A. J. A. V., MEIJER, G., AND HELDEN, G. V. Gas-Phase Infrared Photodissociation Spectroscopy of Cationic Polyaromatic Hydrocarbons. *Astrophys. J. 542*, 1 (2000), 404.

[373] LEHMANN, K. K., SCOLES, G., AND PATE, B. H. Intramolecular Dynamics from Eigenstate-Resolved Infrared Spectra. *Annu. Rev. Phys. Chem. 45*, 1 (1994), 241.

[374] VALLE, J. J., EYLER, J. R., OOMENS, J., MOORE, D. T., VAN DER MEER, A. F. G., VON HELDEN, G., MEIJER, G., HENDRICKSON, C. L., MARSHALL, A. G., AND BLAKNEY, G. T. Free electron laser-Fourier transform ion cyclotron resonance mass spectrometry facility for obtaining infrared multiphoton dissociation spectra of gaseous ions. *Rev. Sci. Instrum. 76*, 2 (2005), 023103.

[375] OKUMURA, M., YEH, L. I., MYERS, J. D., AND LEE, Y. T. Infrared spectra of the cluster ions $H_7O_3^+ \bullet H_2$ and $H_9O_4^+ \bullet H_2$. *J. Chem. Phys. 85*, 4 (1986), 2328.

[376] BARLOW, D., AND THORNTON, J. Helix geometry in proteins. *J. Mol. Biol. 201*, 3 (1988), 601.

[377] KENDREW, J. C., BODO, G., DINTZIS, H. M., PARRISH, R. G., AND WYCKOFF, H. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature 181* (1958), 662.

[378] GUZZO, A. V. The Influence of Amino Acid Sequence on Protein Structure. *Biophys. J. 5*, 6 (1965), 809.

[379] PROTHERO, J. W. Correlation between the distribution of amino acids and alpha helices. *Biophys. J. 6*, 3 (1966), 367.

[380] COOK, D. The relation between amino acid sequence and protein conformation. *J. Mol. Biol. 29*, 1 (1967), 167.

[381] CHOU, P. Y., AND FASMAN, G. D. Conformational parameters for amino acids in helical, β-sheet, and random coil regions calculated from proteins. *Biochemistry 13*, 2 (1974), 211.

[382] SCHOLTZ, J. M., AND BALDWIN, R. L. The Mechanism of α-Helix Formation by Peptides. *Annu. Rev. Biophys. Biomol. Struct. 21*, 1 (1992), 95.

[383] EPAND, R. M., AND SCHERAGA, H. A. The influence of long-range interactions on the structure of myoglobin. *Biochemistry 7*, 8 (1968), 2864.

[384] TANIUCHI, H., AND ANFINSEN, C. B. An experimental approach to the study of the folding of staphylococcal nuclease. *J. Biol. Chem. 244*, 14 (1969), 3864.

[385] VON DREELE, P. H., LOTAN, N., ANANTHANARAYANAN, V. S., ANDREATTA, R. H., POLAND, D., AND SCHERAGA, H. A. Helix-Coil Stability Constants for the Naturally Occurring Amino Acids in Water. II. Characterization of the Host Polymers and Application of the Host-Guest Technique to Random Poly(hydroxypropylglutamine-co-hydroxybutylglutamine). *Macromolecules 4*, 4 (1971), 408.

[386] SUEKI, M., LEE, S., POWERS, S. P., DENTON, J. B., KONISHI, Y., AND SCHERAGA, H. A. Helix-coil stability constants for the naturally occurring amino acids in water. 22. Histidine parameters from random poly[(hydroxybutyl)glutamine-co-L-histidine]. *Macromolecules 17*, 2 (1984), 148.

[387] SCHERAGA, H. A. Use of random copolymers to determine helix-coil stability constants of the naturally occuring amino acids. *Pure Appl. Chem. 50*, 4 (1978), 315.

[388] BIERZYNSKI, A., KIM, P. S., AND BALDWIN, R. L. A salt bridge stabilizes the helix formed by isolated C-peptide of RNase A. *Proc. Natl. Acad. Sci. U.S.A. 79*, 8 (1982), 2470.

[389] BROWN, J. E., AND KLEE, W. A. Helix-coil transition of the isolated amino terminus of ribonuclease. *Biochemistry 10*, 3 (1971), 470.

[390] MARQUSEE, S., AND BALDWIN, R. L. Helix stabilization by $Glu^{-}...Lys^{+}$ salt bridges in short peptides of de novo design. *Proc. Natl. Acad. Sci. U.S.A. 84*, 24 (1987), 8898.

[391] MARQUSEE, S., ROBBINS, V. H., AND BALDWIN, R. L. Unusually Stable Helix Formation in Short Alanine-Based Peptides. *Proc. Natl. Acad. Sci. U.S.A. 86*, 14 (1989), 5286.

[392] SCHOLTZ, J. M., YORK, E. J., STEWART, J. M., AND BALDWIN, R. L. A neutral, water-soluble, α-helical peptide: the effect of ionic strength on the helix-coil equilibrium. *J. Am. Chem. Soc. 113*, 13 (1991), 5102.

[393] SPEK, E. J., OLSON, C. A., SHI, Z., AND KALLENBACH, N. R. Alanine Is an Intrinsic α-Helix Stabilizing Amino Acid. *J. Am. Chem. Soc. 121*, 23 (1999), 5571.

[394] PADMANABHAN, S., MARQUSEE, S., RIDGEWAY, T., LAUE, T. M., AND BALDWIN, R. L. Relative helix-forming tendencies of nonpolar amino acids. *Nature 344*, 6263 (1990), 268.

[395] PACE, N. C., AND SCHOLTZ, M. J. A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins. *Biophys. J. 75*, 1 (1998), 422.

[396] CHAKRABARTTY, A., KORTEMME, T., AND BALDWIN, R. L. Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. *Protein Sci. 3*, 5 (1994), 843.

[397] HOROVITZ, A., MATTHEWS, J. M., AND FERSHT, A. R. α-Helix stability in proteins: II. Factors that influence stability at an internal position. *J. Mol. Biol. 227*, 2 (1992), 560.

[398] MOREAU, R. J., SCHUBERT, C. R., NASR, K. A., TÖRÖK, M., MILLER, J. S., KENNEDY, R. J., AND KEMP, D. S. Context-Independent, Temperature-Dependent Helical Propensities for Amino Acid Residues. *J. Am. Chem. Soc. 131*, 36 (2009), 13107.

[399] LYU, P. C., LIFF, M. I., MARKY, L. A., AND KALLENBACH, N. R. Side chain contributions to the stability of alpha-helical structure in peptides. *Science 250*, 4981 (1990), 669.

[400] CREAMER, T. P., AND ROSE, G. D. Side-chain entropy opposes α-helix formation but rationalizes experimentally determined helix-forming propensities. *Proc. Natl. Acad. Sci. U.S.A. 89*, 13 (1992), 5937.

[401] HOL, W. G. J., VAN DUIJNEN, P. T., AND BERENDSEN, H. J. C. The α-helix dipole and the properties of proteins. *Nature 273*, 5662 (1978), 443.

[402] SHOEMAKER, K. R., KIM, P. S., YORK, E. J., STEWART, J. M., AND BALDWIN, R. L. Tests of the helix dipole model for stabilization of α-helices. *Nature 326*, 6113 (1987), 563.

[403] WADA, A. The alpha-helix as an electric macro-dipole. *Adv. Biophys. 9* (1976), 1.

[404] CREIGHTON, T. E. Stability of alpha-helices. *Nature 326*, 6113 (1987), 547.

[405] BLAGDON, D. E., AND GOODMAN, M. Mechanisms of protein and polypeptide helix initiation. *Biopolymers 14*, 1 (1975), 241.

[406] IHARA, S., OOI, T., AND TAKAHASHI, S. Effects of salts on the nonequivalent stability of the α-helices of isomeric block copolypeptides. *Biopolymers 21*, 1 (1982), 131.

[407] TAKAHASHI, S., KIM, E.-H., HIBINO, T., AND OOI, T. Comparison of α-helix stability in peptides having a negatively or positively charged residue block attached either to the N- or C-terminus of an α-helix: The electrostatic contribution and anisotropic stability of the α-helix. *Biopolymers 28*, 5 (1989), 995.

[408] CLEMMER, D. E., AND JARROLD, M. F. Ion Mobility Measurements and their Applications to Clusters and Biomolecules. *J. Mass. Spectrom. 32*, 6 (1997), 577.

[409] DUGOURD, P., HUDGINS, R. R., CLEMMER, D. E., AND JARROLD, M. High-resolution ion mobility measurements. *Rev. Sci. Instrum. 68*, 2 (1997), 1122.

[410] HUDGINS, R. R., MAO, Y., RATNER, M. A., AND JARROLD, M. F. Conformations of Gly_nH^+ and Ala_nH^+ Peptides in the Gas Phase. *Biophys. J. 76*, 3 (1999), 1591.

[411] KOHTANI, M., KINNEAR, B. S., AND JARROLD, M. F. Metal-Ion Enhanced Helicity in the Gas Phase. *J. Am. Chem. Soc. 122*, 49 (2000), 12377.

[412] KINNEAR, B. S., AND JARROLD, M. F. Helix Formation in Unsolvated Peptides: Side Chain Entropy Is Not the Determining Factor. *J. Am. Chem. Soc. 123*, 32 (2001), 7907.

[413] KINNEAR, B. S., HARTINGS, M. R., AND JARROLD, M. F. The Energy Landscape of Unsolvated Peptides: Helix Formation and Cold Denaturation in Ac-$A_4G_7A_4$ + H^+. *J. Am. Chem. Soc. 124*, 16 (2002), 4422.

[414] KOHTANI, M., SCHNEIDER, J. E., JONES, T. C., AND JARROLD, M. F. The Mobile Proton in Polyalanine Peptides. *J. Am. Chem. Soc. 126*, 51 (2004), 16981.

[415] KOHTANI, M., JONES, T. C., SCHNEIDER, J. E., AND JARROLD, M. F. Extreme Stability of an Unsolvated α-Helix. *J. Am. Chem. Soc. 126*, 24 (2004), 7420.

[416] KOHTANI, M., JARROLD, M. F., WEE, S., AND O'HAIR, R. A. J. Metal Ion Interactions with Polyalanine Peptides. *J. Phys. Chem. B 108*, 19 (2004), 6093.

[417] WANG, P., AND LASKIN, J. Helical Peptide Arrays on Self-Assembled Monolayer Surfaces through Soft and Reactive Landing of Mass-Selected Ions. *Angew. Chem., Int. Ed. 47*, 35 (2008), 6678.

[418] HU, Q., WANG, P., AND LASKIN, J. Effect of the surface on the secondary structure of soft landed peptide ions. *Phys. Chem. Chem. Phys. 12*, 39 (2010), 12802.

[419] LIU, D., WYTTENBACH, T., AND BOWERS, M. T. Hydration of protonated primary amines: effects of intermolecular and intramolecular hydrogen bonds. *Int. J. Mass Spectrom. 236*, 1–3 (2004), 81.

[420] GAO, B., WYTTENBACH, T., AND BOWERS, M. T. Hydration of Protonated Aromatic Amino Acids: Phenylalanine, Tryptophan, and Tyrosine. *J. Am. Chem. Soc. 131*, 13 (2009), 4695.

[421] WYTTENBACH, T., BUSHNELL, J. E., AND BOWERS, M. T. Salt Bridge Structures in the Absence of Solvent? The Case for the Oligoglycines. *J. Am. Chem. Soc. 120*, 20 (1998), 5098.

[422] PLOWRIGHT, R. J., GLOAGUEN, E., AND MONS, M. Compact Folding of Isolated Four-Residue Neutral Peptide Chains: H-Bonding Patterns and Entropy Effects. *Chem. Phys. Chem. 12*, 10 (2011), 1889.

[423] BRENNER, V., PIUZZI, F., DIMICOLI, I., TARDIVEL, B., AND MONS, M. Spectroscopic Evidence for the Formation of Helical Structures in Gas-Phase Short Peptide Chains. *J. Phys. Chem. A 111*, 31 (2007), 7347.

[424] CHIN, W., PIUZZI, F., DIMICOLI, I., AND MONS, M. Probing the competition between secondary structures and local preferences in gas phase isolated peptide backbones. *Phys. Chem. Chem. Phys. 8*, 9 (2006), 1033.

[425] CHIN, W., PIUZZI, F., DOGNON, J.-P., DIMICOLI, I., TARDIVEL, B., AND MONS, M. Gas Phase Formation of a 3_{10}-Helix in a Three-Residue Peptide Chain: Role of Side Chain-Backbone Interactions as Evidenced by IR-UV Double Resonance Experiments. *J. Am. Chem. Soc. 127*, 34 (2005), 11900.

[426] BRENNER, V., PIUZZI, F., DIMICOLI, I., TARDIVEL, B., AND MONS, M. Chirality-Controlled Formation of β-Turn Secondary Structures in Short Peptide Chains: Gas-Phase Experiment versus Quantum Chemistry. *Angew. Chem., Int. Ed. 46*, 14 (2007), 2463.

[427] CHIN, W., MONS, M., DOGNON, J.-P., PIUZZI, F., TARDIVEL, B., AND DIMICOLI, I. Competition between local conformational preferences and secondary structures in gas-phase model tripeptides as revealed by laser spectroscopy and theoretical chemistry. *Phys. Chem. Chem. Phys. 6*, 10 (2004), 2700.

[428] CHIN, W., DOGNON, J.-P., PIUZZI, F., TARDIVEL, B., DIMICOLI, I., AND MONS, M. Intrinsic Folding of Small Peptide Chains: Spectroscopic Evidence for the Formation of β-Turns in the Gas Phase. *J. Am. Chem. Soc. 127*, 2 (2005), 707.

[429] CHIN, W., COMPAGNON, I., DOGNON, J.-P., CANUEL, C., PIUZZI, F., DIMICOLI, I., VON HELDEN, G., MEIJER, G., AND MONS, M. Spectroscopic Evidence for Gas-Phase Formation of Successive β-Turns in a Three-Residue Peptide Chain. *J. Am. Chem. Soc. 127*, 5 (2005), 1388.

[430] CIMAS, A., VADEN, T. D., DE BOER, T. S. J. A., SNOEK, L. C., AND GAIGEOT, M.-P. Vibrational Spectra of Small Protonated Peptides from Finite Temperature MD Simulations and IRMPD Spectroscopy. *J. Chem. Theory Comput. 5*, 4 (2009), 1068.

[431] SEDIKI, A., SNOEK, L. C., AND GAIGEOT, M.-P. N–H^+ vibrational anharmonicities directly revealed from DFT-based molecular dynamics simulations on the Ala_7H^+ protonated peptide. *Int. J. Mass Spectrom. 308*, 2–3 (2011), 281.

[432] MARTENS, J. K., COMPAGNON, I., NICOL, E., MCMAHON, T. B., CLAVAGUÉRA, C., AND OHANESSIAN, G. Globule to Helix Transition in Sodiated Polyalanines. *J. Phys. Chem. Lett. 3*, 22 (2012), 3320.

[433] HESS, B., KUTZNER, C., VAN DER SPOEL, D., AND LINDAHL, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput. 4*, 3 (2008), 435.

[434] CRICK, F. H. C. The packing of α-helices: simple coiled-coils. *Acta Crystallogr. 6*, 8 (1953), 689.

[435] MACKERELL, A. D., BASHFORD, D., BELLOTT, DUNBRACK, R. L., EVANSECK, J. D., FIELD, M. J., FISCHER, S., GAO, J., GUO, H., HA, S., JOSEPH-MCCARTHY, D., KUCHNIR, L., KUCZERA, K., LAU, F. T. K., MATTOS, C., MICHNICK, S., NGO, T., NGUYEN, D. T., PRODHOM, B., REIHER, W. E., ROUX, B., SCHLENKRICH, M., SMITH, J. C., STOTE, R., STRAUB, J., WATANABE, M., WIÓRKIEWICZ-KUCZERA, J., YIN, D., AND KARPLUS, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B 102*, 18 (1998), 3586.

[436] HUA, S., XU, L., LI, W., AND LI, S. Cooperativity in Long α- and 3_{10}-Helical Polyalanines: Both Electrostatic and van der Waals Interactions Are Essential. *J. Phys. Chem. B 115*, 39 (2011), 11462.

[437] BALDAUF, C., GÜNTHER, R., AND HOFMANN, H.-J. Helix Formation in α,γ- and β,γ-Hybrid Peptides: Theoretical Insights into Mimicry of α- and β-Peptides. *J. Org. Chem. 71*, 3 (2006), 1200.

[438] BASUROY, K., DINESH, B., SHAMALA, N., AND BALARAM, P. Structural Characterization of Backbone-Expanded Helices in Hybrid Peptides: $(\alpha\gamma)_n$ and $(\alpha\beta)_n$ Sequences with Unconstrained β and γ Homologues of L-Val. *Angew. Chem. 124*, 35 (2012), 8866.

[439] MA, B., TSAI, C.-J., AND NUSSINOV, R. A Systematic Study of the Vibrational Free Energies of Polypeptides in Folded and Random States. *Biophys. J. 79*, 5 (2000), 2739.

PUBLICATIONS RELATED TO THIS THESIS

- F. Schubert, M. Rossi, C. Baldauf, K. Pagel, S. Warnke, G. von Helden, F. Filsinger, P. Kupser, G. Meijer, M. Salwiczek, B. Koksch, M. Scheffler, and V. Blum, *Exploring the conformational preferences of 20-residue peptides in isolation: Ac-Ala$_{19}$-Lys + H$^+$ vs. Ac-Lys-Ala$_{19}$ + H$^+$ and the current reach of DFT*, in preparation
- F. Schubert, K. Pagel, M. Rossi, S. Warnke, M. Salwiczek, B. Koksch, G. von Helden, V. Blum, C. Baldauf, and M. Scheffler, *Native like helices in a specially designed β-peptide*, in preparation
- N. Heine, T. I. Yacovitch, F. Schubert, C. Brieger, D. M. Neumark, and K. R. Asmis, *Infrared Photodissociation Spectroscopy of Microhydrated Nitrate-Nitric Acid Clusters $NO_3^-(HNO_3)_m(H_2O)_n$*, *J. Phys. Chem. A* **118**, 35 (2014), 7613