

# Going Clean: Structure and Dynamics of Peptides in the Gas Phase and Paths to Solvation

Carsten Baldauf<sup>1</sup>

<sup>1</sup>Fritz Haber Institute, Faradayweg 4–6, 14195 Berlin, Germany

E-mail: baldauf@fhi-berlin.mpg.de

Mariana Rossi<sup>2</sup>

<sup>2</sup> University of Oxford, South Parks Road, OX1 3QZ Oxford, UK

E-mail: mariana.rossi@chem.ox.ac.uk

**Abstract.** The gas phase is an artificial environment for biomolecules that has gained much attention both experimentally and theoretically due to its unique characteristic of providing a clean room environment for the comparison between theory and experiment. In this review we give an overview mainly on first-principles simulations of isolated peptides and the initial steps of their interactions with ions and solvent molecules: a bottom up approach to the complexity of biological environments. We focus the accuracy of different methods to explore the conformational space, the connections between theory and experiment regarding collision cross section evaluations and (anharmonic) vibrational spectra, and the challenges faced in this field.

## Contents

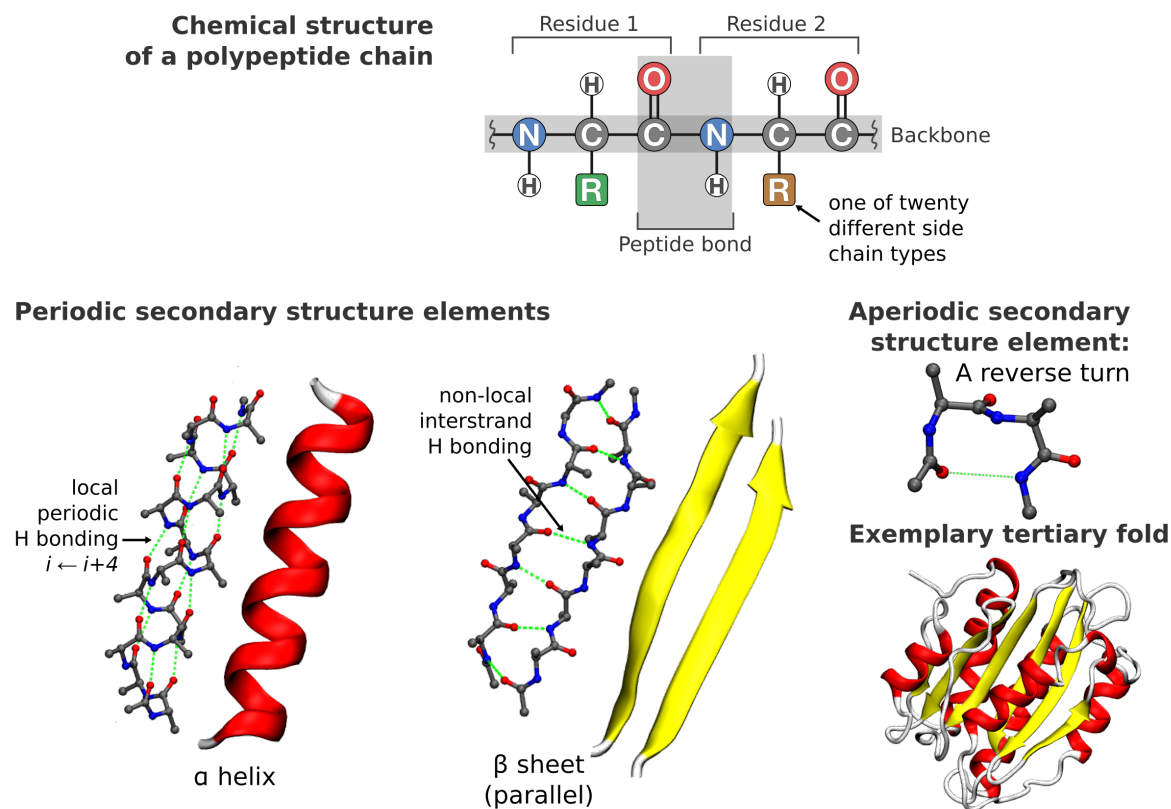
<b>1</b>	<b>Biomolecules in the Gas Phase</b>	<b>2</b>
1.1	Polymeric Biomolecules in the Gas Phase . . . . .	3
1.2	Experimental Techniques Probing Conformation and Dynamics in the Gas Phase . . . . .	5
1.2.1	Ion Mobility-Mass Spectrometry . . . . .	6
1.2.2	Vibrational Spectroscopy . . . . .	7
<b>2</b>	<b>Potential-Energy Surfaces for Peptides</b>	<b>11</b>
2.1	Accuracy of the Potential-Energy Surface . . . . .	11
2.2	Sampling the PES Connecting to First-Principles Methods . . . . .	16
<b>3</b>	<b>How Can Theory Predict Structure and Dynamics?</b>	<b>18</b>

<b>4 Theory-Experiment Comparison – Computation of Physical Observables</b>	<b>22</b>
4.1 Collision Cross Sections . . . . .	23
4.2 Vibrational Spectra . . . . .	26
4.3 Towards First-Principles Free Energies . . . . .	35
<b>5 Challenges Towards Solvation</b>	<b>37</b>
<b>6 Conclusions</b>	<b>39</b>
<b>7 Acknowledgements</b>	<b>42</b>
<b>8 References</b>	<b>42</b>

## 1. Biomolecules in the Gas Phase

In spite of bearing little resemblance to biological environments, the experimental and theoretical study of biomolecules in the gas phase has been steadily gaining importance in the past decades, especially among physical scientists. Pioneer experimental studies starting in the late 90s encompassing all main groups of biomolecules [1, 2, 3, 4, 5, 6, 7, 8, 9] were able to show that much physical insight on structure formation and dynamics of these molecules can be gained from transferring them to the gas phase. The reason is that the gas phase offers clean conditions, under which theory and experiment can meet on equal footing and can follow a stepwise bottom-up approach towards the full complexity of the real biological environment. The reduced size of the systems allows their treatment with a range of theoretical methods that rely on approaches to solving the quantum mechanical Schrödinger equation, usually referred to as “first-principles” methods. These methods are typically much more accurate than empirical models – but due to the intrinsic approximations in them, it is also *a priori* unclear how well they are actually able to describe the structure and dynamics of biomolecules in the gas phase. In this synergistic combination, experiments can serve as a benchmark for testing how appropriate the theoretical treatment of these complex systems is, while theory can be employed to give a physical interpretation to experiments.

In this review, we give a brief survey of the current state of the field regarding the study of, in particular, peptides in the gas phase. We focus on the theoretical side of this field, summarizing what is the current state-of-the-art with respect to accuracy of such calculations, the systems sizes they can treat, what is their predictive power, and where there is room for improvement – basing a good portion of it on works in which where the authors were involved. We also give special attention to the dynamical nature of these molecules, and the importance of grasping at least local entropic, anharmonic and temperature effects. There will be less of a focus on long time scale dynamics of these molecules, which involve large conformational rearrangements (e.g. folding). We choose to concentrate on local dynamics because these span time scales currently



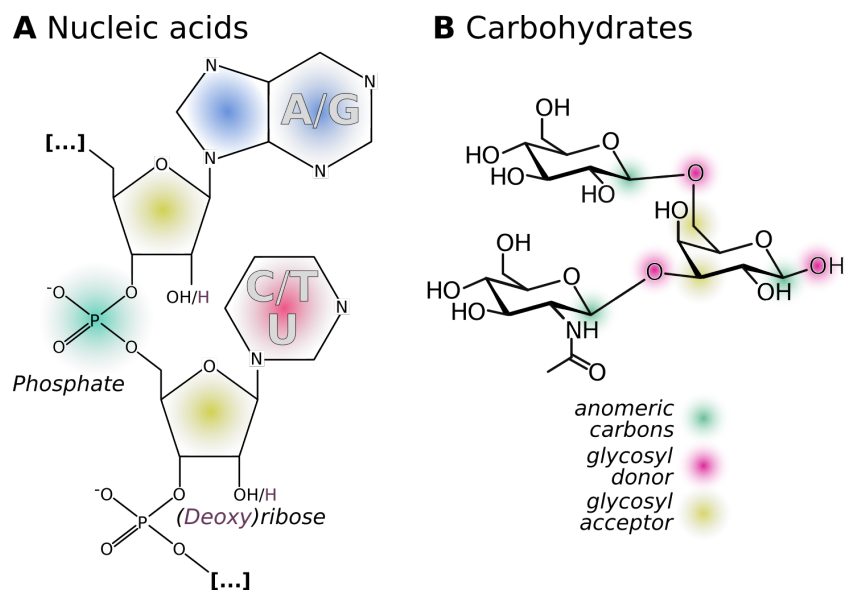
**Figure 1.** Overview over the structure levels of proteins with the chemical structure of a peptide chain, periodic and aperiodic secondary structure elements, and an example of a tertiary protein fold. The three-dimensional structure examples are taken from PDB-ID 3PPY [10].

accessible to first principles potentials. Also they can be connected to most state-of-the-art experiments available in the literature for medium-sized peptides.

### 1.1. Polymeric Biomolecules in the Gas Phase

There are three main classes of biomolecular oligomers and polymers, namely peptides and proteins (see Figure 1), nucleic acids (Figure 2A) and carbohydrates (Figure 2B). Below we briefly describe each of them, with a stronger focus on peptides and proteins, which will be the main subject of this review.

Peptides and proteins make up the machinery of life and are involved in virtually all of its manifestations, from comparably small signaling peptides to gigantic protein complexes. A peptide or protein is a linear chain (oligomer) of amino acids (residues) that are linked by so-called peptide bonds (see Figure 1, top). Peptide bonds are formed between the amino group and carboxylic acid group of two building blocks. In addition, amino acids carry a side chain ‘**R**’ of differing chemical functionality. The sequence of the different amino acid side chains **R** is called primary structure and defines the structure and dynamics of the peptide or protein. Oligomers beyond a certain length (from about



**Figure 2.** Schematic representations of **A** nucleic acids and **B** carbohydrates.

50 amino acids on) that are able to form distinct structural motifs are called proteins. Structure formation at the level of peptides (secondary structure) is mainly dependent on the conformational properties of the monomers and backbone hydrogen bonding. In larger oligomers, i.e. in proteins, side chain interactions and packing gain importance and govern tertiary structure formation. These larger proteins or even ensembles thereof can be studied in isolation as well, see, for example, a recent review by Carol Robinson [11].

Among other biological functions, nucleic acids are the carriers of genetic information. In an organism, a sequence of nucleotides in deoxyribose nucleic acids (DNA) can be transcribed into ribose nucleic acids (RNA) that then serves as template for the stepwise linkage of the amino acids into peptide or protein chain. They feature a sugar-phosphate backbone with nucleobases connected to the (deoxy)ribose moieties (see Figure 2A for a pictorial representation of the different groups). Structure formation is mainly triggered via intermolecular hydrogen bonding between specific pairs of bases (base pairing) in case of DNA or intramolecular base pairing in case of RNA. A recent review by J. Abi-Ghanem and V. Gabelica [12] may serve as an entry point in the literature about nucleic acids in the gas phase. A. Arcella *et al.* [13] investigate DNA in the gas phase by combining ion-mobility mass spectrometry and extensive classical molecular dynamics (MD) and *ab initio* molecular dynamics (AIMD) simulations. They describe rich dynamics of DNA that quickly loses memory of its solution structure in the gas phase and explores a large conformational space. Of special interest is the observation of protons hopping between phosphates of the DNA backbone that was seen in AIMD simulations.

Polymeric carbohydrates serve as nutrition and energy source or as structural scaffolds. They can also be linked to proteins acting as recognition molecules and



possibly playing a role in protein folding. Most of the known carbohydrates are composed of around 20 different monosaccharide units connected to each other by what is called the glycosidic bond or linkage (see Figure 2B). In contrast to the backbone of peptides or the backbone of nucleic acids, carbohydrates are not necessarily composed as a linear chain. The building blocks have one donor (the anomeric C) but multiple acceptors for glycosidic bonds, such that branched structures can be realized. In addition, due to chirality, glycosidic bonds can be formed in two chiral (enantiomeric) forms ( $\alpha$  or  $\beta$ ). These contributions result in a diversity of possible topologies of carbohydrates that surpasses the number of possible sequences in nucleic acids and peptides by orders of magnitude, even with relatively small numbers of building blocks [14]. The significant conformational degrees of freedom are rotations around the single bonds of the glycosidic linkages and the conformation of the monosaccharide rings.

The main focus in this review will be on secondary structure stabilization and dynamics in peptides containing from a few to some tens of amino acids. These motifs are shown in Figure 1. Briefly there are three main elements of secondary structure, namely, helices, pleated-sheets, and turns. The turns are regarded as non-periodic motifs, while helices and sheets are regarded as periodic, in the sense that a repeating unit can be defined, allowing for a characterization based on pairs of torsional angles. The nomenclature given to the helices depend on the hydrogen bonding pattern that arise from their constituting residues (amino acids). The most famous types, the  $\alpha$  and the  $3_{10}$  helices are characterized by H-bonds between residue  $i$  to  $i + 4$  and residue  $i$  to  $i + 3$ , respectively. Sheets are H-bonded between them and characterized as parallel and anti-parallel depending on the relative orientation of their peptide chains. Finally turns are necessary motifs to reverse the propagation of sheets and helices, so that compact structures can be formed. It is not necessary for a H-bond to form in order for the motif to be characterized as a turn, but many do form through the formation of H-bonds. The most common type is known as the  $\beta$  turn, which causes a  $180^\circ$  change in the propagation direction.

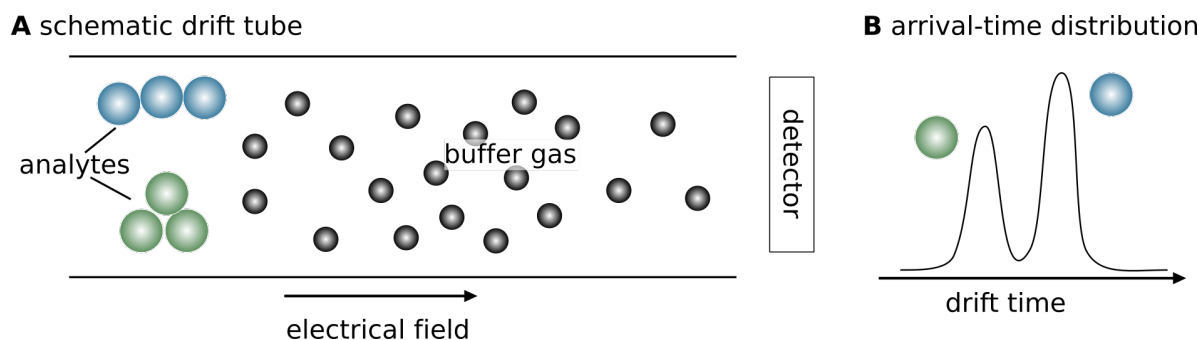
### *1.2. Experimental Techniques Probing Conformation and Dynamics in the Gas Phase*

The study of (bio)molecules in the gas phase has become more popular in the past decades mainly due to the development of experimental techniques in the late eighties, that can gently transfer intact biomolecules to the gas phase, like MALDI (Matrix-Assisted Laser-Desorption Ionization [15]) and ESI (Electro-Spray Ionization [16]), in combination with high-accuracy mass spectrometers [17, 18] or in molecular beams [19].

When dealing with peptides, it is possible to isolate secondary structure motifs in the gas phase, so that their “unperturbed” energy landscape and stabilizing intermolecular interactions can be carefully studied. The environmental effects can then be added in a controlled way, for example by the stepwise addition of water molecules to the polypeptide or by adding ions to the complexes. At the same time, these clean experiments in the gas phase allow to benchmark theoretical methods, at system sizes

that can be treated in a fully first-principles manner. There is much debate as to how biologically relevant the study of biomolecules in the gas phase actually is [20, 21, 22], since it is to be expected that due to the lack of solvent and hydrophobic/hydrophilic interactions one can actually stabilize different conformations in the gas phase. From a more physical perspective, it is undeniable though that in these experiments much insight about the fundamental stabilizing interactions can be gained, also encompassing what is the role of the protonation state and the first shells of solvation, or the interaction with only ions, ions and water, etc. This understanding can be certainly transferred to the more complex biological environment.

There are several reviews about studies of biomolecules (peptides, proteins, sugars, etc.) in the gas phase in the literature, from which we highlight only a few for the interested reader, namely Refs. [1, 18, 2, 22, 23, 6, 3, 24, 25, 26, 20, 21, 9, 8]. Below we give a brief overview of the main experimental techniques that yield quantities which can be connected to theoretical calculations that we will review in the next sections.



**Figure 3.** Schematic representation of an ion mobility spectrometry experiment.

*1.2.1. Ion Mobility-Mass Spectrometry* Mass spectrometry (MS) is a powerful gas-phase experimental technique that separates ionic clusters or molecular ions based on their mass-to-charge ratio ( $m/z$ ). With ion mobility (IM), or gas chromatography, charged molecules and clusters can be separated according to their different mobility in a buffer gas. Especially the combination of both techniques, ion mobility-mass spectrometry (IM-MS), first accomplished in 1962,[27] can allow for the separation and characterization of mixtures of compounds or conformers that would otherwise not be distinguishable. In the context of this review, we focus especially on the ability of IM-MS to investigate structural and dynamical properties of peptides.

In IM-MS experiments, an electric field drags the ions through a drift tube of a certain length. This drift tube is filled with a buffer gas (often He or N<sub>2</sub>) and collisions between buffer gas and ions slow down the ions depending on their shape and size. As a result, an arrival-time distribution (ATD) of  $m/z$  selected ions is measured by a detector. The arrival times can be transformed into collision cross sections (CCSs) by

the Mason-Schamp equation [28]

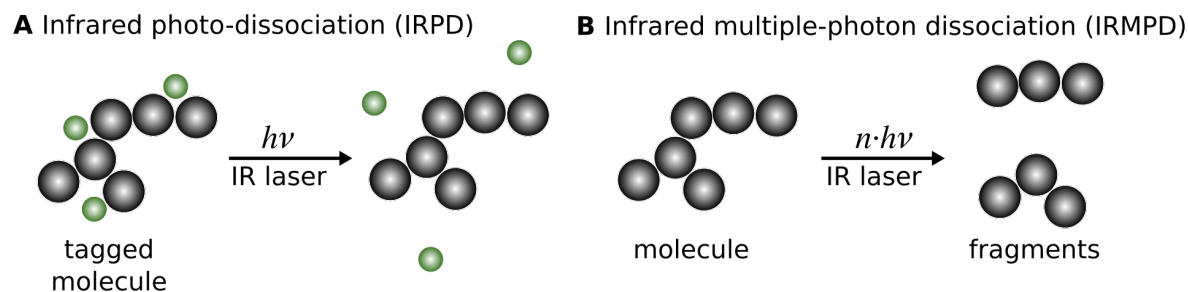
$$CCS = \frac{3ze}{16N} \left( \frac{2\pi}{\mu k_B T} \right)^{\frac{1}{2}} \frac{1}{K_0}, \quad (1)$$

where  $ze$  refers to the net charge of the system,  $\mu$  is the reduced mass of the ion and the buffer-gas particles (usually He atoms), and  $k_B$  is Boltzmann's constant. The reduced mobility  $K_0$  is the proportionality constant that relates drift velocity  $v_d$  and the electric field of the apparatus  $E$  following the relation  $v_d = K_0 E$ . The resulting CCS is a geometrical property of the molecule and it is ideally independent of the apparatus used.

A few examples of the use of IM-MS experiments to study structure and dynamics of peptides gas phase are

- M. Jarrold and coworkers have developed a high-temperature drift-tube instrument and studied polyalanine helices in the gas phase from room temperature to 725 K [29]. The surprising finding is that helical structures can be observed still at these high temperatures for peptides Ac-Ala<sub>15</sub>-Lys(H<sup>+</sup>). A. Tkatchenko *et al.* identified vdW interactions as the crucial stabilizing contribution in DFT-based molecular dynamics simulations [30], being essential to explain the high temperature stability of the helical structure observed in experiment.
- Based on ion-mobility measurements, K. Shelimov and M. Jarrold were able to show the unfolding and refolding of Cytochrome C in vacuum [31]. The folded *vs.* unfolded state is linked to different charge states with a folded to unfolded transition between charge states +5 and +7.
- By using a combination of IMS and MD simulation, G. von Helden and co-workers studied different combinations of *cis/trans* isomerization states of prolyl peptide bonds of ubiquitin [32]. CCS measurements and computations are sensitive enough to reveal the *cis* or *trans* conformation of a single peptide bond in a biological macromolecule.
- The group of D. Clemmer has played a leading role in devising drift-tube apparatus using them to investigate different kinetically-trapped conformations of, for example, Bradykinin [33, 34].
- D.H. Russel and co-workers have used a cryogenic drift tube at 80 K to investigate the structures of singly-protonated water clusters [35]. They were able to measure small (1 to 30 water molecules) and large clusters (31 up to about 120 water molecules) and to assign changes of H bonding upon loss of single water molecules from the clusters.

*1.2.2. Vibrational Spectroscopy* The low concentration of molecules in the gas phase renders it difficult to obtain vibrational spectra through absorption spectroscopy, the technique commonly used in the condensed phase. Instead, in what is called action spectroscopy, an intense tunable laser is employed that acts on a comparably small

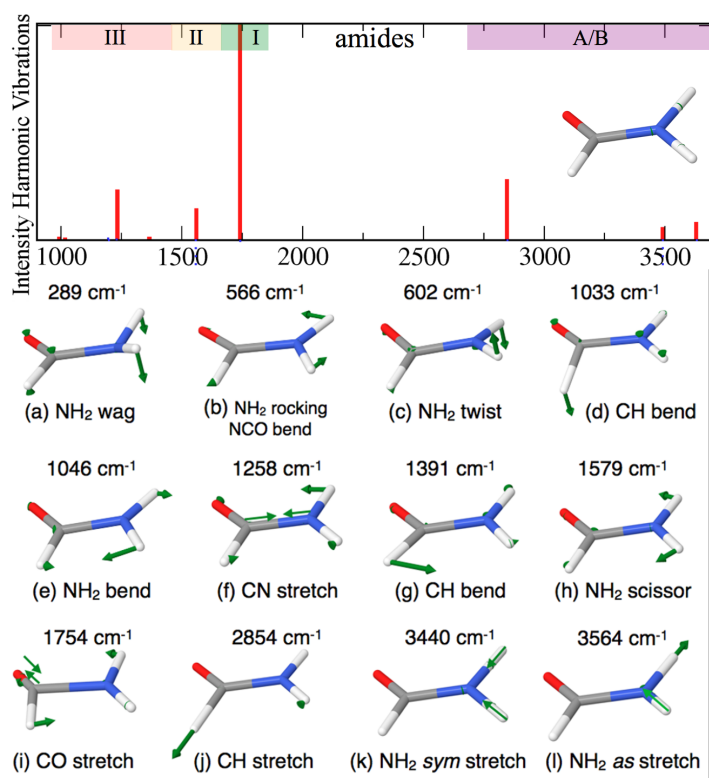


**Figure 4.** Vibrational action-spectroscopy techniques.

number of molecules is employed. When a resonance is encountered, due to the absorption of single or multiple photons, the sample dissociates or fragments and detection via mass spectrometry is possible. One can detect either the fragments or the depletion of the molecular beam. At this point, two types of action spectroscopy can be performed. One technique, commonly called infrared photo-dissociation IRPD (sketched in Figure 4A), is usually performed at lower temperatures and uses inert tags (e.g.  $\text{H}_2$ , Ar, Ne, etc.) on the target molecule that are released after the absorption of one or very few photons, due to the low binding energy of the tag. Another technique, called infrared multiple-photon dissociation IRMPD (sketched in Figure 4B) does not use any tag and simply measures the fragmentation of whole molecules due to the sequential absorption of at least a few tens of photons. For detailed reviews of the experimental techniques, we point the reader to Refs. [18, 36, 24, 37].

In both action spectroscopy techniques mentioned above, non-linear effects can arise due to the absorption of more than one photon. Therefore, different from absorption spectroscopy where the spectra can be safely approximated by a linear response theory, here it is not *a priori* clear that the measured vibrational spectra in this manner will allow a linear response modelling. Especially for IRMPD, where indeed many photons are absorbed sequentially, causing induced and spontaneous emission as well as energy redistribution among vibrational modes, it is clear that a linear response approximation may fail. It has been shown that while the line shape and intensity of the peaks can be strongly influenced by these non-linear absorption effects, the peak positions usually follow the ones calculated by linear response [38, 39] with slight red-shifts due to anharmonicities. It is found that especially for lower frequencies below  $1200\text{ cm}^{-1}$  some peaks are transparent to IRMPD (but not to IRPD), as was shown by the group of K. Asmis for microhydrated nitrate-nitric acid clusters [40] and bisulfate/sulfuric acid/water clusters [41]. The reason they propose is that the absorption of photons disrupts the hydrogen bond network of these systems and causes the modes to go out of resonance with the frequency of the laser. More specific comparisons regarding theoretical modeling and experimental IR spectra will be given in Section 4.2.

In the experimental studies, several different parts of the vibrational spectra can be probed, which are sensitive to different conformational properties: (i) The amide A/B regions, comprising localized CH and NH stretch vibrations above  $\approx 2500\text{ cm}^{-1}$ ,



**Figure 5.** Gas phase spectrum and normal modes of vibrations calculated with DFT-PBE functional for the formamide molecule. Amides I, II, III, and A/B regions are marked on top.

sensitive to the H-bonding pattern; (ii) the amide I (mainly collective CO stretch vibrations), amide II (mainly collective NH bend vibrations), and amide III (collective and localised CH and CN bend vibrations) regions between 2000 and 800  $\text{cm}^{-1}$ , sensitive to backbone conformation; and (iii) the “far-infrared” region, below 800  $\text{cm}^{-1}$ , which contains collective vibrations and is also sensitive to backbone conformation. While much focus has been given to the amide A/B and amide I and II regions in most studies, attention has been called to the amide III region in middle-sized polypeptides [42, 43] and to the far-infrared region in small polypeptides [44] as regions that can be used to differentiate conformations, if anharmonicities of the potential-energy surface are taken into account. As an illustration, we show these regions and the harmonic normal modes of vibrations calculated with the PBE exchange correlation functional for the formamide molecule in Fig. 5.

Gas phase investigations can be used to study distinct aspects of protein secondary structure, peptide bond properties, and aspects of microsolvation. In the following we discuss a few outstanding examples, all performed with double resonance techniques combining UV and IR radiation:

- K. Tanabe *et al.* have used UV/IR pump-probe experiments on clusters of acetanilide and water to investigate the motion of a single water molecule from the hydrogen bond acceptor (CO group) to the hydrogen bond donor (NH group)

- of a **peptide bond** [45].
- M. Gerhards and co-workers studied dimers of the short peptide Ac-Val-Tyr(Me)-NHMe in molecular beam experiments [46]. The combination of IR/UV double-resonance spectroscopy and simulated vibrational spectra (harmonic, B3LYP/cc-pVDZ) identifies the formation of an anti-parallel  $\beta$  **sheet-like structure**. The study shows that sheet-formation can be regarded as an intrinsic trend of peptides that is not necessarily linked to aqueous solution.
  - The group of T. Rizzo has a long-standing experience in UV/IR experiments on **helical peptides** Ac-Phe-Ala<sub>n</sub>-LysH<sup>+</sup> with a C terminal protonated Lys and a Phe residue as UV chromophore [47]. The helical pattern has been elucidated with a <sup>15</sup>N labeling technique. The C terminal capping motif that is present in the longer helices with  $n \geq 5$  has recently been shown to be present also in short peptides with  $n = 1$  [48]. These results confirm predictions about the helix onset made by M. Rossi *et al.* for very related systems [49]. These systems with the aromatic Phe side chain are a challenge to theory and will be discussed further on in this review.
  - Besides strands/sheets and helices, **turns** are the third main secondary structure motif in proteins. The group of M. Mons studied peptides Ac-Phe-NH<sub>2</sub>, Ac-Phe-Pro-NH<sub>2</sub>, and Ac-Pro-Phe-NH<sub>2</sub> from supersonic molecular beams with UV/IR double resonance spectroscopy [50]. The authors assign various turn types and indicate the dependence of Phe conformations on the neighboring residues.
  - M. Johnson and his group have most elegantly shown how gas-phase infrared spectroscopy of cold ionic **complexes** can be used to elucidate not only molecular structure, but also the way two molecules interact with each other [51]. They use site-specifically placed <sup>13</sup>C labels as conformational reporters. Difference spectra between the distinctly labeled systems allow for structural investigations of a single peptide ion and also complex formation through binding to sodium cations or with other molecules.
  - In order to directly estimate the energy barriers between different conformers, T. Zwier and co-workers developed a double resonance conformer selective pump and dump technique that excites molecules to a higher electronic level and then relaxes them back into a specific vibrational ground state [52]. With this approach the authors were able to **reconstruct the potential-energy surface** of tryptamine.
  - I. Compagnon and coworkers have carried out seminal work on the FELIX free electron laser on peptides in the gas-phase, for example looking at backbone preferences [53], internal proton transfer that can stabilize zwitterionic structures in the gas-phase [54], and microsolvation of amino acids [55]. More recently they have been looking at sugars in the gas phase [56], focusing especially on the issue of anharmonicities in vibrational modes.
  - The group of M. Lisy has a body of work based on IRPD regarding the influence of charge due to the interaction with metal ions and temperature on the conformational preferences of small biomolecules [57, 58].

- T. Vaden, L. Snoek, and coworkers have measured IRMPD spectra of a variety of peptides in the gas phase, also performing extensive structural searches involving density functional theory. They have, for example, studied the  $\text{Ala}_n\text{H}^+$ ,  $n=3,4,5,7$  series of peptides [59] in the amide A/B region concluding that these peptides form mostly globular structures at larger sizes, despite the high propensity of the Ala amino acid to form helices. They have also looked at peptide sequences relevant to amyloid formation, showing that even if the isolated structure of Ac-VQIVYK-NHMe is folded, the simple interaction with another monomer in the gas phase seems to be energetically favorable enough to trigger a conformational change and  $\beta$ -sheet aggregation [60].

Vibrational spectroscopy techniques can also be combined with ion mobility-mass spectrometry. A first example of, in that case, electronic spectroscopy of mobility selected peptides was published by T. Rizzo and coworkers [61]. They use a Field Asymmetric waveform Ion Mobility Spectrometry (FAIMS) setup combined with UV photofragment spectroscopy in order to decompose the electronic spectrum of doubly-protonated bradykinin in a conformer-specific manner. Also L. Voronina and T. Rizzo demonstrate how to use a combination of ion-mobility selection and cold-ion spectroscopy to study kinetically trapped conformers of triply-protonated bradykinin [62]. G. von Helden, K. Pagel, and co-workers have used ion mobility in order to separate conformers of protonated benzocaine and to record vibrational spectra [63].

The spectral resolution can be improved by measuring vibrational spectra of cold species. This can, for example, be realized in cold traps that are utilized in IR/UV double resonance experiments where changes of UV fragmentation yield are recorded as a function of IR excitations [64, 65]. Alternatively, ions can be embedded in liquid He nanodroplets [66] and therewith cooled to an equilibrium temperature of about 0.4 K. Employing such a setup, G. von Helden and co-workers have measured vibrational spectra of the short peptide leucine-enkephalin [67].

## 2. Potential-Energy Surfaces for Peptides

### 2.1. Accuracy of the Potential-Energy Surface

The potential-energy surface (PES) of a system is often defined as an energy function of the coordinates that tells how energy changes with respect to a change in any atomic position. This definition assumes an adiabatic separation of the electronic and nuclear degrees of freedom (known as the Born-Oppenheimer approximation). Moreover, it is usually (but not necessarily) also connected to the assumption that nuclei are classical particles. Even if both of these assumptions can break down in many situations (some of them discussed in the next sections), they are, in most cases, a good approximation or at least a good starting point to map the energy profile of a system.

When dealing with biomolecules in general, the challenging aspect is that the potential energy surfaces (PES) are often far from simple: due to the existence of

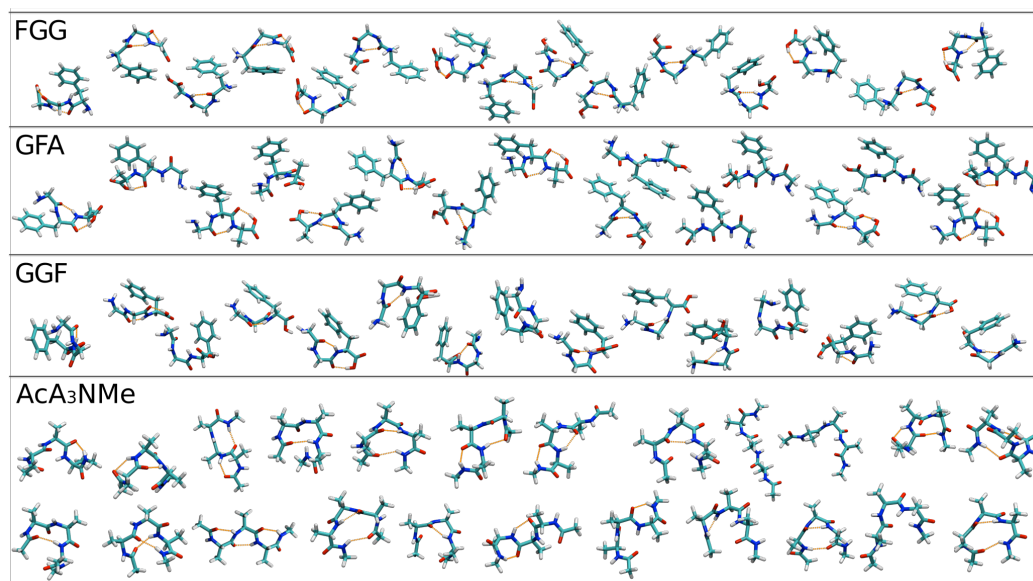
several soft and anharmonic degrees of freedom these PES tend to have several different local minima – all of which will contribute to the partition function and thus define the thermodynamical properties of the system. If this PES is not rigorously described also all thermodynamic properties and structural preferences of the system will not be reliable. Especially the amount of anharmonic degrees of freedom make most harmonic approximations fail for these systems.

Perhaps the most popular way of evaluating PES are the so-called force fields. Force fields are parametrized empirical energy functions that represent the energy of a given system in terms of the sum of qualitatively different interactions. In the case of molecules (and especially peptides) the different contributions are separated into bonded interactions (e.g. potentials for bond lengths, bond angles, and torsions) and non-bonded interactions (e.g. van der Waals and electrostatics). For all of these terms, the functional form is physically motivated but arbitrary, and the parameters are fitted to either experimental data or quantum chemistry methods. The advantages of such an approach is that energy evaluations are computationally cheap. Therefore, these methods (if used in combination with smart sampling techniques [68]) typically allow enough statistical sampling to enable the evaluation of thermodynamical properties and to treat system sizes that can bear more connection to biological size- and time-scales with respect to more accurate methods that are too computationally expensive. If used with caution, these potentials can yield good physical insight on the structure and dynamics of biomolecules. However, it is more and more clear that their performance in many situations is far from optimal. Especially regarding polypeptides, recent literature has shown that force fields have several limitations when compared and benchmarked against higher level quantum chemistry methods. Relative energies between different peptide conformations are not well reproduced [69, 70, 71, 72] and differ quite drastically between different force fields. Regarding the interaction of peptides with ions, force fields have been shown to yield even poorer energetics with respect to high level theoretical benchmark data [73, 74], even when especially tailored parameters and polarizable potentials are used. More recently, a study has shown that kinetic models derived from converged simulations based on different non-polarizable force fields largely disagree [75].

The desired solution would be to describe the potential-energy surface (PES) at least as accurately as possible for the electronic degrees of freedom – which would mean to use methods like full configuration interaction (full CI), coupled cluster with a high enough order of excitations (e.g. with single, double, and perturbative triple excitations CCSD(T)), or quantum Monte Carlo (QMC). These methods are considered the gold standard of quantum chemistry, and do indeed provide a very accurate description of potential energy surfaces, but of course, are very costly to compute. Even if they can be used for benchmarking purposes it is not computationally feasible to routinely use them for PES exploration and the simulation of other physical properties.

A good compromise can be found among the wave function based methods, for example with Møller-Plesset perturbation theory (MP2) or coupled-cluster methods





**Figure 6.** Conformers of the peptides Phe-Gly-Gly (FGG), Gly-Phe-Ala (GFA), Gly-Gly-Phe (GGF), and Ac-Ala<sub>3</sub>-NMe (AcA<sub>3</sub>NMe) used for energy benchmark calculations appearing in references [83] and [84].

with lower-order excitations, e.g. singles and doubles (CCSD). A promising route is to use approximations like the domain based local pair natural orbital coupled cluster method with single-, double-, and perturbative triple excitations (DLPNO-CCSD(T)) [76]. The method is described as efficient enough to perform rather accurate coupled cluster calculations even for relatively large molecules with hundreds of atoms. However, some of the approximations must be carefully balanced [77]. It is typically computationally cheaper to use electronic density based methods like density-functional theory (DFT). DFT, with its approximate exchange correlation functionals, is arguably the best compromise between cost and accuracy in the market of electronic structure theory methods. Its advantage is that it allows one to treat molecules of sizes up to few thousand of atoms and reach time scales of hundreds of picoseconds in its most optimized implementations (Big-DFT [78], ONETEP[79], FHI-aims[80], CASTEP[81], CP2K[82], etc.).

It is well known that results from DFT can depend on the choice of exchange-correlation functional. However, since the theory itself is based on the first principles of quantum mechanics, it is possible to obtain accurate results as long as one ensures that the chosen functional can describe the relevant physical properties of the system. For example, most standard DFT functionals lack, by construction, long range van der Waals dispersion. It is, however, now widely accepted that these interactions have a critical impact on the structure [85, 70, 86, 83, 73, 74, 49] and dynamics [30, 85] of peptides, especially for the larger sizes. It becomes thus almost mandatory to include these interactions in the most accurate manner in DFT calculations of peptides in any type of environment, and several schemes for including these corrections have been

**Table 1.** Mean absolute error and maximum error for the energy hierarchies of 16 conformers of Gly-Phe-Ala (GFA), 15 conformers of Gly-Gly-Phe (GGF), 15 conformers of Phe-Gly-Gly (FGG), and 27 conformers of Ac-Ala<sub>3</sub>-NMe, compared to CCSD(T) reference data from Refs. [83, 84]. Values for the mean-absolute errors (MAE) and maximal errors (Max.) are reported in meV (in parentheses: converted to kcal/mol).

	PBE	PBE+vdW <sup>TS</sup>	PBE+MBD	PBE0	PBE0+vdW <sup>TS</sup>	PBE0+MBD
<b>FGG</b>						
MAE	43(1.0)	37(0.8)	36(0.8)	35(0.8)	23(0.5)	23(0.5)
Max.	160(3.7)	59(1.4)	88(2.0)	132(3.0)	38(0.9)	59(1.4)
<b>GFA</b>						
MAE	53(1.2)	32(0.7)	44(1.0)	40(0.9)	17(0.4)	25(0.6)
Max.	108(2.5)	88(2.0)	76(1.7)	89(2.0)	72(1.7)	61(1.4)
<b>GGF</b>						
MAE	48(1.1)	36(0.8)	40(0.9)	38(0.9)	26(0.6)	28(0.6)
Max.	143(3.3)	99(2.3)	84(1.9)	119(2.7)	78(1.8)	66(1.5)
<b>Ac-Ala<sub>3</sub>-NMe</b>						
MAE	55(1.3)	21(0.5)	22(0.5)	54(1.2)	18(0.4)	20(0.5)
Max.	131(3.0)	72(1.7)	66(1.5)	132(3.0)	47(1.1)	54(1.2)
	OPLS-aa	Amber99sb	Charmm22	AmoebaPro04		
<b>Ac-Ala<sub>3</sub>-NMe</b>						
MAE	108(2.5)	42(1.0)	91(2.1)	53(1.2)		
Max.	246(5.7)	86(2.0)	271(6.2)	112(2.6)		
<b>GGF</b>						
MAE				91(2.1)		
Max.				606(14.0)		

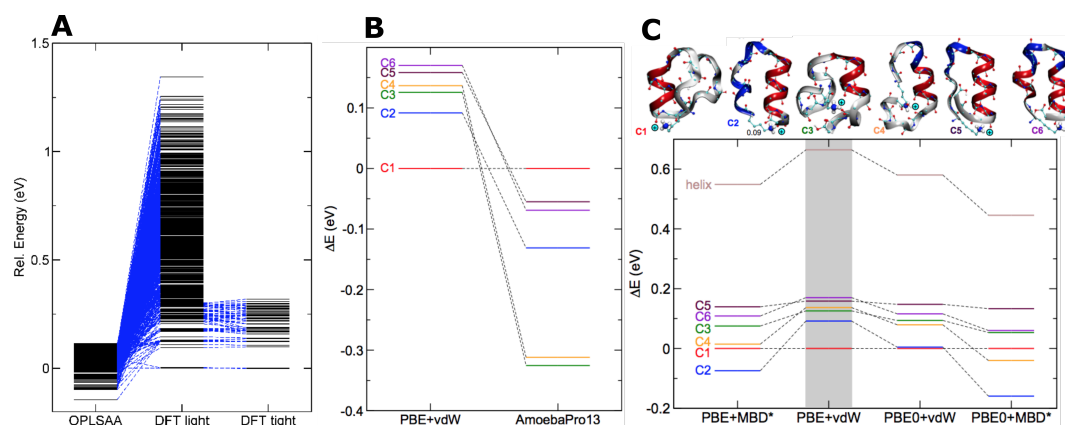
proposed in the last decade, which were nicely reviewed in [87]. Also, the inclusion of Hartree-Fock exchange can mitigate the self-interaction/delocalization problem of DFT and substantially change the strength of H bonds, the description of polarizability, or barriers for conformational dynamics. What is more prudent to avoid in DFT is to blindly use different types of functionals without any kind of physical reasoning or benchmarks.

As an example of the type of accuracy that can be reached with state-of-the-art DFT methods nowadays, we show in Table 1 mean absolute errors and maximum errors on relative energies for three-residue peptides, shown in Fig. 6 (FGG, GFA, GGF, Ac-Ala<sub>3</sub>-NMe), of DFT functionals with respect to CCSD(T) reference benchmark data. We test a generalized gradient exchange correlation functional (PBE [88]) and include both a pairwise van der Waals correction with  $C_6$  coefficients that depend on the electronic density [89] (vdW<sup>TS</sup>), and another that includes both electrostatic screening and many body effects up to infinite order through a coupled fluctuating dipole model [90, 91] (MBD@rsSCS, which we here call MBD). We also test a hybrid exchange correlation functional with these corrections, namely PBE0 [92]. For comparison, we also calculate the same relative energies with popular non-polarizable force fields (OPLS-aa [93],

Amber99sb [94], Charmm22 [95, 96]) and the polarizable force field AmoebaPro04 [97, 98]. Augmenting DFT approaches with a correction for long-range van der Waals interactions leads to energy estimates that agree very well with CCSD(T) calculations, which is evident by low mean-absolute errors (MAE) and low maximal errors. For example PBE0+MBD yields MAEs of only up to 28 meV (0.6 kcal/mol) and a maximal error of 66 meV (1.5 kcal/mol). The force fields tested here and the bare functionals alike give higher MAE and also higher maximal errors that severely limit their predictive power.

In order to illustrate how such errors can impact larger polypeptides, the experimental benchmark helix-forming peptide Ac-Phe-Ala<sub>5</sub>-LysH<sup>+</sup> is ideal. From very accurate conformer selective UV-IR double resonance experiments in the gas-phase by J. Stearns and coworkers [47], it was established that four conformers are present in the experimental beam, which have been satisfactorily assigned to helix-forming structures, based on the similarity of their harmonic IR spectra to the measured ones. A subsequent study [99] considered 19 density functionals, plus Hartree-Fock and MP2 methods, finding that the spread of the relative energies of these four conformers could vary by around 0.15 eV for these methods. None of the functionals considered included long-range van der Waals interactions. Further studies on the same system by M. Rossi and coworkers [70] considered a larger pool of conformers coming from an extensive first-principles scan of the PES of this peptide. Based on the benchmarks shown above, the authors found that indeed when considering the energy hierarchies at the PBE0+MBD level and (harmonic) zero point energy contributions on this system, the four conformers observed in experiment are indeed predicted to be the ones with lowest energies. The spread of their energy differences is also consistent with what is estimated from experiment ( $\approx 50$  meV), and within the estimated error bars, such that the detailed energy hierarchy between them cannot be safely predicted by any DFT method. Interestingly, Ref. [70] finds that the relative abundances for different conformers observed in experiment are better explained by a kinetic trapping from higher temperatures.

Finally for even larger peptides, where the experimental data is also not so conclusive, small energy differences can be even more important as the conformational landscape can get more convoluted. We take as an example the 20-residue peptide Ac-Lys-Ala<sub>19</sub>-H<sup>+</sup>, studied in Ref. [100] by F. Schubert, the authors of this review, and coworkers. We show in Fig. 7 (data reproduced from [100] and [101]) in panel A the comparison between the force field relative energies for thousands of conformers predicted by the OPLS-aa force field, and relative energies of the same conformers when further relaxed with PBE+vdW<sup>TS</sup> “light settings” (smaller basis sets and integration grids in the FHI-aims [80] code) and “tight settings” (larger basis sets and integration grids). The scatter is huge, spanning up to 1.5 eV in DFT for conformers that were 0.5 eV apart in OPLS-aa. We also show in Fig. 7B and C, for a set of selected conformers of this molecule the comparison between the energy hierarchies of PBE+vdW<sup>TS</sup> and the AmoebaPro13 force field [102], and the comparison between different the functionals

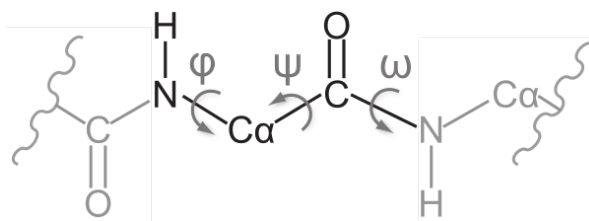


**Figure 7.** Ac-Lys-Ala<sub>19</sub>-LysH<sup>+</sup> reproduced and adapted from [100] and [101].

and van der Waals corrections discussed above. The energy differences between the functionals are much smaller than the difference comparing to AmoebaPro13. Many body van der Waals dispersion do indeed have an impact in molecules of this size, which in this case also improves agreement to experimental data, as discussed further in Sec. 4.2.

## 2.2. Sampling the PES Connecting to First-Principles Methods

The degrees of freedom (DOF) that define a PES are the positions of all atoms of the molecules expressed in, for example, Cartesian space, internal coordinates, etc. For molecules, one can often simplify that (reducing the number of DOF) by assuming a fixed configuration of the molecular system (basically assuming that covalent bonds do not break). As a consequence, an internal coordinate system consisting of bond lengths, bond angles, and torsion angles can be used to describe a molecule's structural (conformational) space. Since bond lengths and bond angles typically vary around a single equilibrium value, torsion angles are often the most descriptive internal coordinates for a molecular system. An exploration of a molecule's potential energy surface must sample the space defined by the combination of all its torsional degrees of freedom. For a typical peptide molecule with three backbone torsion angles per residue and further torsions in the side chain, the problem easily gets too large for a systematic grid-based enumeration of possible points on the PES. A single alanine building block in a peptide chain has three torsional DOF (see Figure 8: the torsions  $\phi$  and  $\psi$  represent rotations around single bonds and the peptide bond torsion angle  $\omega$  adopts *cis* or *trans* conformations). Assuming a grid of 60 degrees for discretization of the single-bond rotations yields  $6 \times 6 \times 2 = 72$  conformations to test for a single building block. For a chain of  $N$  building blocks this number virtually explodes already for short peptides with  $72^N$ . A variety of strategies has been developed and employed to explore these conformational spaces connecting to first principles methods. Below, we will give a rough definition and some examples of them.



**Figure 8.** Backbone torsion angles of a prototypical amino acid building block embedded in a peptide chain.

- Systematic searches can be performed by discretization of the involved degrees of freedom with sufficiently fine grids. All combinations of torsion angles are either subject to a single point energy calculation or serve as starting point for local geometry optimizations. Such an approach is well applicable to small molecular systems, e.g. dipeptides. With a more “target-oriented” objective, also bigger systems can be studied in a systematic way, if only a particular region of the search space is of interest. An example is the search for all possible helical structures in homologous peptides, i.e. peptides which have their backbones extended by methylene units. With the aim of finding such periodic and hydrogen bonded structures, the same combination of backbone torsion angles is applied to all subunits and only geometries that are (i) clash free and (ii) feature a backbone hydrogen-bonding pattern of interest are considered [103, 104, 105].
- Systematic searches can easily be performed for monomers. The knowledge gained in this way can then be combined in the creation of starting structures for longer oligomers of the respective building block(s). This approach has been successfully employed for example to  $\beta$ -peptides, which are homologous peptides with an addition of one methylene unit. [106, 107, 108, 105].
- Parallel-tempering or replica-exchange molecular dynamics (REMD) can substantially enhance the sampling of conformational space in comparison to standard MD simulations [109, 110, 111, 112, 113, 114]. REMD requires only limited human interaction and no definition of collective variable or alike. Robust protocols exist for a wide range of simulation programs. Several copies (a.k.a. replicas) are simulated in parallel by means of MD simulations at different temperatures. At predefined intervals, pairs of replicas with neighboring temperatures are eventually swapped based on a Metropolis criterion. The individual copies traverse a wide temperature range and can overcome barriers.
- Basin hopping [115] reduces the PES to attraction basins centered on local minima. In contrast to REMD, moves on the landscape do not follow realistic pathways. The basic algorithm starts with a structure guess and a local optimization to the next local minimum. A perturbation of coordinates generates a new starting point for a geometry optimization that leads to the next minimum. This sequence of coordinate perturbation and local optimization is repeated until a convergence criterion is met.

Frequently used implementations are for example in the programs TINKER [116] or GMIN [117].

- Genetic algorithms (GAs) are frequently used for global structure search and optimization of chemical compounds [118, 119, 120]. They use a ‘survival of the fittest’ concept. Starting from a population of random solutions, genetic operations are applied and energy-optimal solutions are selected. GAs use the accumulated information to explore promising regions of conformational space. Examples are the program foldaway by M. Damsbo *et al.* [121] and the program Fafoom [122, 123] that can employ first-principles techniques.

A complete sweep of the potential-energy surface with any of the above mentioned methods is anything but trivial. All methods require parameter choices that have to be made by the respective user as well as a careful selection of the energy function to be used. While force fields offer low computational costs and therefore allow for a more exhaustive sampling of the PES, the results can suffer from the systematic energy errors that were discussed in the previous section. First-principles methods offer a description of the energetics that is unbiased by empirical parameters, but that may demand far more computational resources. Clever combinations of search techniques and stepwise increase of accuracy can be a way out that, however, requires experience. In the next section, we will review some of these combination methods.

### 3. How Can Theory Predict Structure and Dynamics?

As presented in the last sections, several benchmark works have shown that force fields may not be accurate enough to predict quantitative energy differences between peptide conformations in the gas phase. However, as also mentioned in the previous section, the high dimensionality of the potential-energy surface renders the direct exploration with first-principles potentials an elusive task. Therefore, theoretical studies that aim to explore the PES of larger polypeptides (and biomolecules in general) with first principles methods tend to follow an overall similar work flow [124, 125, 49, 73, 70, 100, 126, 127, 128, 129].

The general aim is to balance a broad sampling of conformers and an accurate description of the energetics with the available computer power. We exemplify this work flow in general below, exemplifying it by the technique followed in Ref. [100], which we believe to be among the largest current computational efforts to study the conformational space of alanine based polypeptides from first principles. The work flow is also schematically represented in Fig. 9.

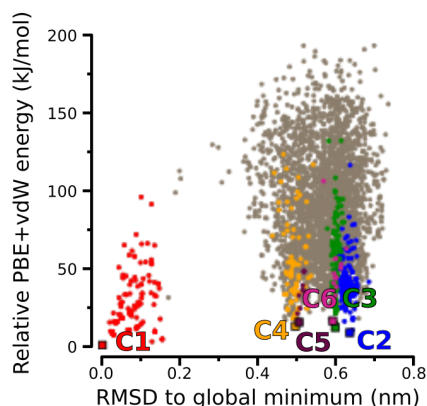
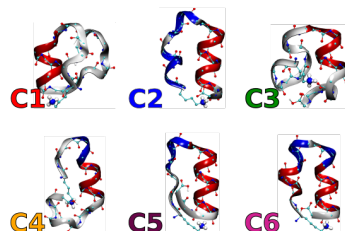
The **first step** involves a thorough enumeration of different conformers using a force field. These conformers are commonly local minima in the force field, found by different sampling techniques, like basin hopping, replica exchange, genetic algorithms, or any other sampling method. The idea is to perform a global and thorough exploration of structure space. For example in Ref. [100] replica-exchange molecular dynamics

**Target** is the peptide Ac-LysH<sup>+</sup>-Ala<sub>19</sub>

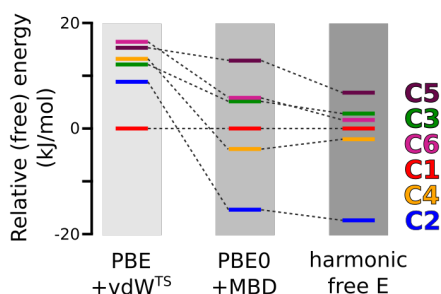
**Exploration** of a large 'search space' to find minima with **force field** and **ab initio** replica-exchange MD

**Clustering** of conformers:  
 - Energy ordering  
 - RMSD clustering  
 - Machine learning of descriptors

**Overview** of the PES at a reasonable level of theory

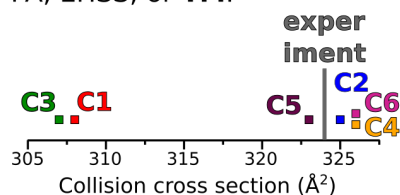


**Refinement** by applying higher levels of theory and free energy estimations

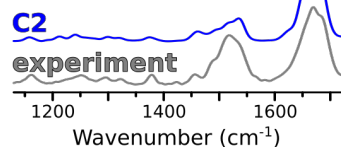


**Comparison** to experiment by computing physical observables

Computing CCS by means of PA, EHSS, or TM.



Simulating vibrational spectra, harmonic, from AIMD, or VSCF



**Figure 9.** Typical steps followed by theoretical studies regarding structure search and prediction based on first principles methods.

(REMD) simulations were performed with the OPLS-aa force field [93, 130] with 16 replicas for a total of 500 ns per replica. From these simulations conformations at each 2 ps were considered to generate an overall set of conformations. The less reliable the PES is at this step, the more conformers will have to be considered in the second step.

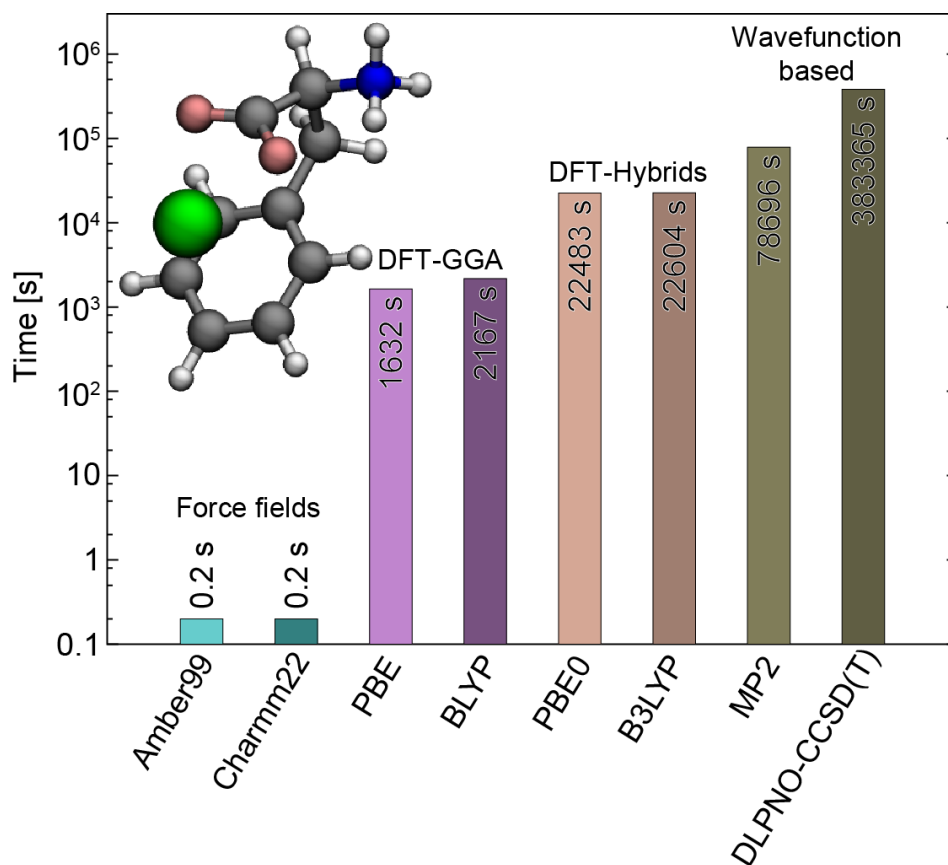
The **second step** is choosing which conformers from the force field sampling will be considered for the treatment with higher level methods (e.g. density-functional theory or other quantum chemistry methods). The conformers can be ranked by

energy from lowest to highest. As described above, there can be large possible errors related to the force fields. The discrepancy between empirical and first principles descriptions is highlighted, for example, in Figure 7A. Many conformers (hundreds to thousands, depending on the system’s characteristics) should be considered, otherwise low-energy conformers may be completely missed. Alternatively, conformers can be sorted by structural criteria in order to generate a pool of candidate structures that is as diverse as possible for investigation. Examples are clustering algorithms based on the root-mean-square deviation (RMSD) of Cartesian coordinates (e.g. in Ref. [100]) or sorting of structures according to hydrogen-bonding patterns (e.g. in Ref. [103, 43]). Other descriptors for structural similarity can, for example, be found by using machine learning methods similar to the ones presented in Refs. [131, 132, 133]. The chosen conformers are typically fully optimized with higher-level methods. Especially the local geometry optimization of force-fields minima with first-principles methods can involve large conformational changes that may lead to new local minima, which are not present in the force field. In Refs. [125, 49, 73, 70, 100, 126] we could highlight the importance of considering a large pool of conformers: Considering only a couple of tens of conformers would have led to missing many of the relevant conformations discussed in these papers. The discrepancy in relative energies from FF and DFT illustrated in Figure 7A also raises the question if all relevant local minima can be located from simply re-relaxing the force-field conformers. As a means of ameliorating the situation, it is possible to introduce a **third step**, a local first-principles sampling. In Ref. [100], for example,  $16 \times 20$  ps *ab initio* REMD simulations were performed and the most stable conformer (C2) of the study was only found in this refinement step.

After that step, one can continue increasing the accuracy for a subset of the conformers from the previous step. The conformers can again be clustered and a new smaller set can be chosen according to the same criteria as in the first step or others. The accuracy can be increased either by increasing numerical settings of the calculations (basis sets, grids, etc.) or by going to even higher level theoretical methods. In Ref. [100] both were done, by going to a higher numerical accuracy as well as using computationally more expensive (and often more accurate) hybrid DFT functionals, and many-body van der Waals dispersion corrections [90]. Other works have also used MP2 and CCSD(T) methods for smaller systems in this step [128].

In order to exemplify the range of computational costs of different methods, we present in Figure 10 timings that were measured for a comparably small system, namely phenylalanine with a  $\text{Ca}^{2+}$  cation. Please note that the accuracy level of the DFT (*really\_tight* settings mean a very large basis and very fine integration grid) and the wavefunction calculations (with 3-4 extrapolation to the complete basis-set limit) are chosen rather high compared to what one would perform as standard calculation. The specific timings for each method can vary considerably when using different (smaller or larger) basis sets, when using different codes, or when treating larger and denser systems. The nominal scaling with system size  $N$  is for DFT  $N^3$ , for MP2  $N^5$ , and for CCSD(T)  $N^7$ . In all cases however, developments are ongoing to reduce the respective scaling





**Figure 10.** Timings for typical single point calculations of conformers of phenylalanine with  $\text{Zn}^{2+}$ . Standard protein force fields (Amber 99 and Charmm22) were computed with Tinker [116]. DFT calculations in the generalized gradient approximation (PBE and BLYP) and with hybrid functionals (PBE0 and B3LYP) were done with FHI-aims [80] (including pairwise Tkatchenko-Scheffler van der Waals correction and *really\_tight* computational settings). Wavefunction-based calculations (MP2 and DLPNO-CCSD(T) [76]) were performed with the Orca code [134] using Ahlrich’s basis sets for a 3-4 extrapolation to the complete-basis-set limit. The timings for the DFT calculations include force evaluations. The timings for the wavefunction calculations include both steps, the triple- and quadruple- $\zeta$  calculations. If the calculations were running in parallel (DFT and wavefunction), the real timings were multiplied with the number of cores. Please note, the numbers are meant to give a rough qualitative idea about the range of timings that can be expected with different methods. Different codes, settings, systems, and computer infrastructures will result in quantitatively different timings.

by the use of smart algorithms [79, 135, 76]. Nevertheless, the timings presented in Figure 10 are good guidelines for what to expect in computational cost when increasing accuracy.

Having finished with a smaller subset of the most-likely structure candidates, it is desirable to connect to more physical quantities than a simple scan of the potential energy surface. Free energies and thermodynamic properties at realistic/experimental conditions can be explored at this step, either by performing anharmonic free energy

evaluations with a method of choice (steered dynamics, metadynamics, umbrella sampling, replica exchange, etc.) or at least considering these contributions in the harmonic approximation. If the system is too large, again it becomes unfeasible to calculate more accurate anharmonic quantities with a higher-level electronic-structure method, such that the harmonic approximation remains as the last resort. Its predictive power, though, has to be critically assessed for these soft and flexible systems.

With the low (free) energy conformers at hand, the connection to experiment can be established by computing physical observables. In the present work we focus especially on collision cross sections that are experimentally derived from IM-MS (see section 1.2.1) and on vibrational spectra (see section 1.2.2). Other possible quantities of interest are electronic spectra, neutron scattering data, or any other experimental technique that is the most applicable to the environment where the biomolecule is measured in experiment.

Another important application of first-principles based conformational searches are studies that compare properties across chemical space. An example is the search for essentially all conformers of 20 proteinogenic amino acids alone and interacting with either of the cations  $\text{Ca}^{2+}$ ,  $\text{Ba}^{2+}$ ,  $\text{Sr}^{2+}$ ,  $\text{Cd}^{2+}$ ,  $\text{Pb}^{2+}$ , and  $\text{Hg}^{2+}$  [127]. As a result, one obtains comparable data for sets of compounds and/or complexes, generated on equal footing with respect to the search technique and the employed energy function. Based on such grounds, physical observables can be computed and compared across chemical compound space. The workflow employed by Ropo and coworkers [127] starts from a force field based structure search (Tinker scan [116] with the OPLS-AA force field [93]) and the relaxation with DFT-PBE+vdW. Again, it is necessary to refine the search results with a local first-principles search step. The bias from the initial treatment with empirical potentials can only be compensated by *ab initio* REMD simulations. The multi-step search procedure yielded an essentially unbiased first-principles data set of more than 45,000 stationary points on the PESs of the different molecular systems. The data can be used as a starting point for, e.g., the parameterization of empirical potentials, comparisons of properties like cation binding strength across chemical space, or as input for spectra calculations. The data is available from the website <http://aminoaciddb.rz-berlin.mpg.de> and from the NOMAD repository [136].

#### 4. Theory-Experiment Comparison – Computation of Physical Observables

A major challenge when performing simulations is to match the experimental conditions in a simulation setup. An effort on both ends is needed. Experimental conditions should be well controlled and the data recorded precise and sharp and the system size and character that is considered in the simulation should be as realistic as computationally feasible. The gas phase is an excellent environment in this respect, where it is possible to simulate physical observables on a very similar footing with experiments.

In the next section we focus on the calculation of collision cross sections and

vibrational spectra. In addition, there are several optical spectroscopy techniques that can probe also electronic excitations and dynamics of excited states, connected to UV and visible probes. For example, in the UV-IR pump-probe experiments mentioned above, the UV laser induces electronic excitations that can be used to select different conformers. Reviews and perspectives of such optical spectroscopies in the gas phase, applied to peptides and other biomolecules can be found in Refs. [23, 8, 137, 9]. R. Antoine and P. Dugourd report the possibility of recording electron photo-detachment following electronic excitation in negatively charged peptides to obtain gas-phase optical spectra for large systems (even proteins), since this process does not suffer from limitations brought by energy redistribution into vibrational modes with system size and is less convoluted than a vibrational spectrum for large systems [137]. Theoretical modelling of electronic excited states and the resulting processes and dynamics is a major challenge, since it requires the use of time-dependent or explicitly correlated electronic structure techniques [138, 139, 140] that can treat excited states. These are very computationally expensive if compared to ground state techniques and have many further limitations included in the approximations, such that their application to large biomolecular systems is still limited, but growing fast.

#### 4.1. Collision Cross Sections

From the Cartesian coordinates of conformers that result from a structure search for a particular molecular ion, it is possible to compute CCS values. The underlying collisions of the ion with the buffer-gas atoms (e.g. He) or molecules (e.g. N<sub>2</sub>) can be modeled including different levels of detail. We will review here the three most-commonly used methods, the projection approximation [141], the exact hard-sphere scattering [142], and the trajectory method [143].

The projection approximations, or in short PA [141], takes the shape of the molecule into account, modelling the interaction between ion and buffer-gas particles by means of Lennard-Jones and charge-dipole interactions. The averaged collision cross section in the PA approximation ( $CCS_{PA}$ ) is calculated by using the collision parameters  $\theta$ ,  $\phi$ , and  $\gamma$  as well as the minimal impact parameter  $b_{min}$  as follows:

$$CCS_{PA} = \frac{1}{4\pi^2} \int_0^{2\pi} d\theta \int_0^{\pi} d\phi \sin\phi \int_0^{2\pi} d\gamma \pi b_{min}^2 \quad (2)$$

In practice,  $b_{min}$  is tabulated as atom-wise impact parameters, in a simplified view they are stored as up-scaled atomic radii. The CCS value for a given molecular conformation is computed numerically by: (i) projecting the atoms of the molecule onto a randomly chosen plane, (ii) drawing the collision radii around positions of the nuclei, and (iii) repeatedly selecting random points within an area  $A$  enclosing the projected molecule. Out of step (iii), a CCS value for a planar orientation  $N$  is computed following the formula  $CCS_N = (h/t)*A$ , where  $h$  is the number of hits within the projected outline of the molecule and  $t$  is the number of overall tries. Steps (i) to (iii) are repeated

for different planes and an average CCS value out of values  $CCS_N$  is computed until convergence to a given threshold is reached. PA is shown to work well especially for largely convex molecules.

PA neglects scattering events as well as multiple collisions between buffer-gas particles and the ion. However, such effects are especially pronounced for concave molecular surfaces where certain surface areas can be shielded by parts of the molecule, while in others multiple collisions may occur. The projection-superposition approximation (PSA) aims to compensate for this with a shape factor that accounts for the concavity of a molecule [144]. Alternatively, scattering and multiple-collision effects can be considered by regarding ion and buffer-gas particles as hard-spheres. The exact hard-sphere scattering (EHSS) approach [142] explicitly follows the trajectory of a He atom that is shot at the molecule or cluster through all possible collisions until it leaves the molecule/cluster for good. Here, the scattering angle  $\chi$  (the angle between the trajectories before and after a collision event between the molecular ion and a buffer-gas particle) is computed as a function of the collision parameters  $\theta$ ,  $\phi$ , and  $\gamma$  and the impact parameter  $b$  for multiple collision geometries and thus an average  $CCS_{EHSS}$  can be obtained:

$$CCS_{EHSS} = \frac{1}{4\pi^2} \int_0^{2\pi} d\theta \int_0^{\pi} d\phi \sin\phi \int_0^{2\pi} d\gamma \times \int_0^{\infty} db 2b(1 - \cos \chi(\theta, \phi, \gamma, b)) \quad (3)$$

The trajectory method (TM) models one extra bit of the physics defining the drift of an ion through a buffer gas, namely long-range interactions between the drifting ion and the buffer gas. The importance of this contribution depends on the polarizability of the buffer gas, which is for example stronger in  $N_2$  than in He, and on the charge distribution in the (molecular) ion. The charge(s) of the drifting ion induces dipoles in the buffer gas atoms altering its drift velocity without ‘physical contact’ [143].

$$CCS_{TM} = \frac{1}{4\pi^2} \int_0^{2\pi} d\theta \int_0^{\pi} d\phi \sin\phi \int_0^{2\pi} d\gamma \times \int_0^{\infty} db 2b(1 - \cos \chi(\theta, \phi, \gamma, b)) \times \left( \frac{\mu}{k_B T} \right)^3 \int_0^{\infty} dg e^{-\mu g^2 / 2k_B T} g^5 \quad (4)$$

In addition to the symbols explained above, the reduced mass  $\mu$  and the relative velocity  $g$  are being used. The interaction between the ion and the buffer-gas particles is modeled by two terms: a Lennard-Jones 12-6 potential and a term that accounts for the

interaction between the charge (distribution) of the ion and the charge-induced dipole of the buffer-gas particle. This treatment considers differences in polarizability between buffer gases, for example between He and N<sub>2</sub>.

We note, though, that in principle all methods are designed to work with He as the buffer gas. When comparing to measurements made with, for example, N<sub>2</sub>, parameters going into the calculations have to be adapted. An overview about specific contributions to the collision cross section can be found in a paper by T. Wyttenbach *et al.* [145] where for a wide range of systems experimental and PSA-simulated CCS are compared. There are several programs described in the literature, which can be more or less straightforward to obtain. We list here only some of the more popular ones:

**MOBCAL** is developed in the group of M. Jarrold and incorporates PA, EHSS, and TM. It can be downloaded at [www.indiana.edu/~nano/software.html](http://www.indiana.edu/~nano/software.html).

**sigma** was developed in the group of M. Bowers and it computes CCSs according to the PA and EHSS method. It is available under this URL: [bowers.chem.ucsb.edu/theory\\_analysis/cross-sections/sigma.shtml](http://bowers.chem.ucsb.edu/theory_analysis/cross-sections/sigma.shtml).

**FHIsigma** is a spin-off of **sigma** by M. Wesemann and G. von Helden and comes with a graphical user interface. The program is available at: [sigma.fhi-berlin.mpg.de](http://sigma.fhi-berlin.mpg.de).

**IMPACT** is intended for structural proteomics applications and claims to compute extremely fast PA-CCSs [146]. The software is available at: [benesch.chem.ox.ac.uk/resources.html](http://benesch.chem.ox.ac.uk/resources.html).

The choice of method, for example between PA, PSA, EHSS, and TM, can be critical for the predictive power of the CCS calculation. Some examples are collected in table 2. Depending on the nature of the ionic cluster/complex or molecular ion under investigation, the alternative methods can agree, like in the case of two peptides from reference [126], where PA and TM give virtually the same results. But there are also examples where the methods give qualitatively different results. Different protonation states (protomers) of benzocaine exist that result in either the distribution of the positive charge over the molecule or in its localization at a protonated amino function [63]. In the experiment, both forms can be separated with a polarizable buffer gas (N<sub>2</sub>). In simulations, the CCSs computed with the PA are indistinguishable, while TM predicts distinct values for the protomers and allows an interpretation of experiment.

The interpretation of an experimental arrival-time distribution or of the derived CCS distribution is not unambiguous. The theoretical CCS of a single conformer represents a projection of the conformational degrees of freedom onto a single coordinate. As a consequence similar CCSs may still result from different structures. Also, in the experimental CCS, even a single sharp peak represents not only a projection of spatial coordinates, but also the dynamics of the molecular or cluster ion over the drift time. Consequently, measuring a single sharp peak can mean that either (i) there is only a single conformational family present in the ion cloud, (ii) there are multiple (more than one) conformational families present in the ion cloud that have the same CCS, or even (iii) the time average over multiple interconverting conformers for a single molecule is

**Table 2.** CCS values computed with PA or PSA and TM for different conformers/protomers of three molecules compared to the respective experiment-derived CCS.

Structure	CCS <sub>PA/PSA</sub> in Å <sup>2</sup>	CCS <sub>TM</sub> in Å <sup>2</sup>	CCS <sub>Exp</sub> in Å <sup>2</sup>
Ac-Ala <sub>6</sub> -Lys(H <sup>+</sup> ) from ref. [126]			
α helix	180	181	180
compact	172	171	
Ac-β <sup>2</sup> hAla <sub>6</sub> -Lys(H <sup>+</sup> ) from ref. [126]			
H12	203	204	190
H16	191	193	
H20	182	182	
compact	183	182	
Benzocaine from ref. [63]			
O-prot./trans	131	133	135
O-prot./gauche	132	133	
N-prot./trans	133	144	155
N-prot./gauche	130	144	

converged during the drift time and the measured CCS basically represents a converged average over the CCSs of the different structures. An example was shown in Ref. [126], where IMS data of a β peptide is interpreted to represent the interconversion between related helix types. In a sense, ion-mobility experiments, especially in conjunction with molecular simulations, can be used to deduce not only the structure of molecules, but also their dynamics.

#### 4.2. Vibrational Spectra

As mentioned in Section 1.2.2, several experiments probe the vibrational spectra of biomolecules in the gas phase. These spectra contain more detailed structural information than CCS experiments and simulations. However, especially for larger and more anharmonic systems, a comparison to theoretical simulations is necessary in order to interpret the experimental signal. Good reviews on several types of theoretical spectroscopy methods that can be used in connection to first principles potential-energy surfaces for biomolecules can be found, e.g., in [147, 148].

Theoretically, the “zeroth-order” way to model the vibrational properties of any system is the harmonic (or double harmonic) approximation. In this approximation a Taylor expansion of the Born-Oppenheimer potential with respect to displacements of nuclear coordinates is truncated on the second (quadratic) order and harmonic frequencies of vibrations are calculated for the problem of coupled harmonic oscillators with force constants corresponding to the second derivative of the potential [149]. From Fermi’s golden rule, it is known that the IR intensities are proportional to the square of the matrix elements of dipole-allowed transitions. One can thus

Taylor expand the dipole moment with respect to nuclear displacements, solve the quantum mechanical Hamiltonian in the harmonic approximation, and find the allowed transitions. By truncating the expansion of the dipole moment at first order, one arrives at the expressions for the so called “double harmonic” approximation. Not only this approximation does not contain any anharmonicities, it also does not allow any other transition beyond the fundamental ones. For Raman spectra, similar expressions can be calculated for the harmonic approximation relying on the estimation of matrix elements of allowed transitions from the polarizability tensor [149]. This type of approximation is frequently used for a first comparison of structural properties in connection with scaling factors that compensate for the complete lack of anharmonicities (both of the classical PES and connected to the quantum nature of the nuclei).

A fundamental problem with the harmonic approximation for the study of biomolecules is that these molecules can have very anharmonic potential-energy surfaces. A well known way to calculate IR transitions including anharmonicities is to relate Fermi’s Golden Rule to time correlation functions – a derivation found in many textbooks (e.g. Ref. [150]). One finds that the IR absorption spectrum can be written as the product of the frequency-dependent refractive index  $n(\omega)$  and the Beer-Lambert absorption coefficient  $\alpha(\omega)$  as

$$n(\omega)\alpha(\omega) = \frac{\pi\omega(1 - e^{-\beta\hbar\omega})}{3cV\hbar\epsilon_0} I_{\mu\mu}(\omega), \quad (5)$$

where  $\beta$  is the inverse temperature,  $V$  the volume,  $\epsilon_0$  the dielectric permittivity of vacuum,  $c$  the speed of light and  $I_{\mu\mu}(\omega)$  is the Fourier transform of the dipole auto-correlation function, here defined in the canonical ensemble  $C_{\mu\mu}(t) = \text{Tr}[e^{-\beta H} \boldsymbol{\mu}(0) \boldsymbol{\mu}(t)]/Z$ , where the partition function  $Z = \text{Tr}[e^{-\beta H}]$  and  $\boldsymbol{\mu}(t) = e^{iHt/\hbar} \boldsymbol{\mu} e^{-iHt/\hbar}$ . Since the correlation functions are usually approximated by classical (nuclei) or semi-classical dynamics, the correlation function that is in fact better approximated is the Kubo-transformed one, defined as  $\tilde{C}_{\mu\mu}(t)$ :

$$\tilde{C}_{\mu\mu}(t) = \frac{1}{\beta} \int_0^\beta C_{\mu\mu}^\lambda(t) d\lambda, \quad (6)$$

$$C_{\mu\mu}^\lambda(t) = \text{Tr} [e^{-(\beta-\lambda)H} \boldsymbol{\mu} e^{-\lambda H} \boldsymbol{\mu}(t)] / Z. \quad (7)$$

The Kubo transformed correlation has the same symmetries as a classical correlation function [151] and arises naturally in several approximate quantum dynamics schemes [152, 151]. The Fourier transform of the Kubo transformed time correlation  $\tilde{I}_{\mu\mu}(\omega)$  and the one of the canonical time correlation  $I_{\mu\mu}(\omega)$  are related by

$$I_{\mu\mu}(\omega) = \frac{\beta\hbar\omega}{1 - e^{-\beta\hbar\omega}} \tilde{I}_{\mu\mu}(\omega). \quad (8)$$

Thus, the commonly coined “quantum correction factor” [153, 154] arises naturally from the relationship of these two correlations. The expression that one usually

calculates for IR absorption is

$$n(\omega)\alpha(\omega) = \frac{\pi\beta\omega^2}{3cV\epsilon_0} \tilde{I}_{\mu\mu}(\omega) = \frac{\pi\beta\omega^2}{3cV\epsilon_0} \int dt e^{-i\omega t} \langle \boldsymbol{\mu}(0)\boldsymbol{\mu}(t) \rangle \quad (9)$$

where the brackets denote a time average, and  $\boldsymbol{\mu}(t)$  is generated by classical or approximate quantum dynamics for the nuclei. Similar expressions for Raman spectra can be found with respect to the autocorrelation functions of the polarizability tensor [155]. When classical dynamics (e.g. Born Oppenheimer *ab initio* molecular dynamics) is employed to approximate these auto-correlation functions only the anharmonicities of the underlying (classical) potential-energy surface are taken into account. The remaining discrepancies when comparing to benchmark experiments can be due to the lack of considering the quantum nature of the nuclei (which introduces what is sometimes referred to as quantum anharmonicities), the use of an approximate potential-energy surface, or sampling of the wrong (ensemble of) conformers – all of which can cause the spectra to change considerably, as discussed in more detail below.

Other techniques to obtain anharmonic vibrational spectra are, e.g. vibrational self consistent field (VSCF) and second order vibrational perturbation theory (VPT2). These methods and their applications to biomolecules have been reviewed by T. Roy and B. Gerber [156], and V. Barone and coworkers [157] recently. In both of them the quantum nuclear Hamiltonian is approximately solved either in a mean field approximation or a perturbation theory one, thus including quantum anharmonicities. However, the inclusion of temperature and explicit dynamics (where many conformations may be sampled) is not straightforward [158, 159], and the methods are expensive to treat very large molecules. An impressive recent work from a computational point of view was the application of VSCF-PT2 with the B3LYP functional to the spectra of two conformers of Gramicidin S, comparing to cold gas-phase IR-UV double resonant spectra, obtaining satisfactory agreement [160].

Even though the evaluation of IR and other vibrational spectra from autocorrelation functions has been popular for decades especially for condensed phase systems and empirical potentials, M. P. Gaigeot and coworkers have pioneered its use in connection to first-principles (DFT) potential energy surfaces and applying it to isolated and solvated small polypeptides [161, 162, 25, 163]. It is remarkable how well the simulated spectra based on a linear absorption regime (cf. Eq. 9) agree with those measured with the IRMPD technique. Great examples are spectra for  $\text{Ala}_2\text{H}^+$ ,  $\text{Ala}_3\text{H}^+$  that were derived from *ab initio* molecular dynamics simulations employing the BLYP functional [164, 165]. The authors observe that at room temperature the peptides interconvert between a few different structures and that these dynamics are important for the comparison with the IRMPD spectra. This type of studies serves also as an indirect probe of the dynamics. They also reported sensitivity to different conformations in the amide III regions for polyalanine peptides [25], and good structure selectivity and comparison to IR-UV IRMPD spectra in the far-infrared region for Ac-Phe-Gly-NH<sub>2</sub> and Ac-Phe-Ala-NH<sub>2</sub> [44]. This is very interesting, since vibrations in this lower wavenumber

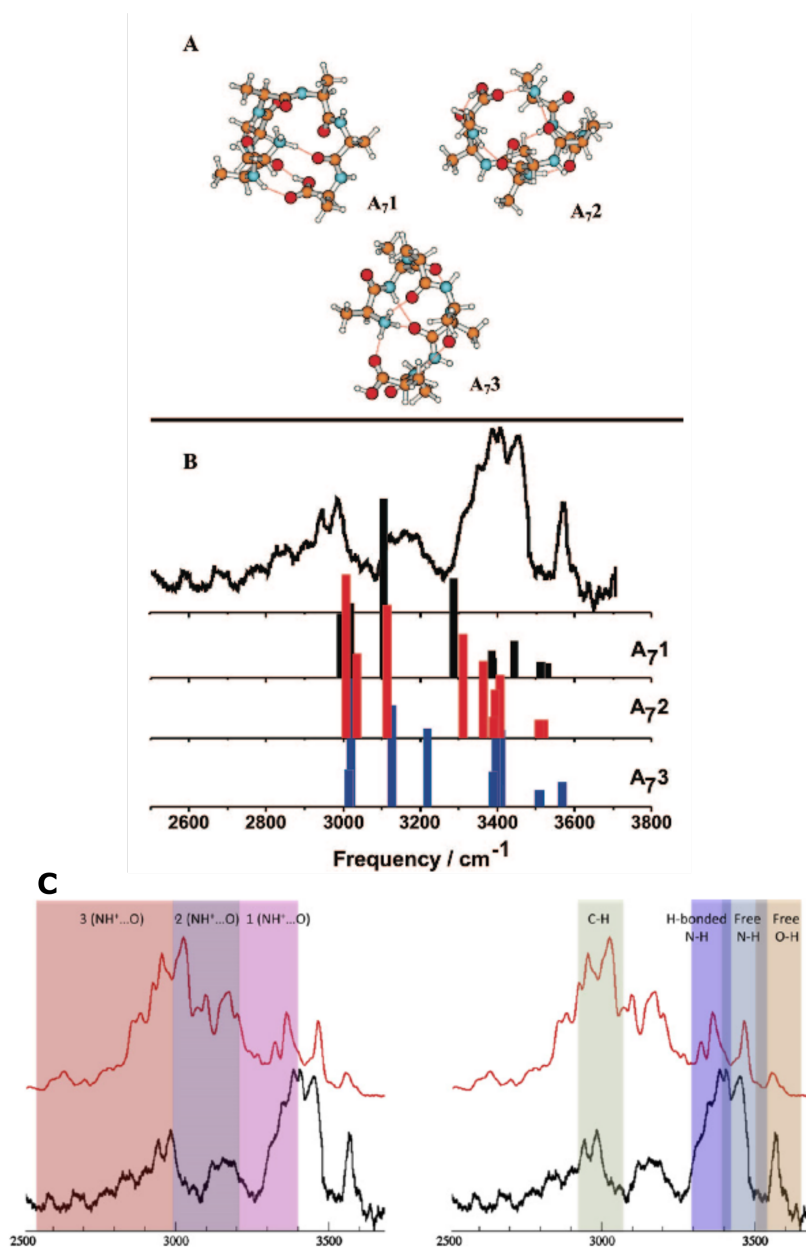


region are more classical in nature and can be more accurately represented by classical (*ab initio*) molecular dynamics, not requiring simulation techniques that incorporate quantum effects of the nuclei.

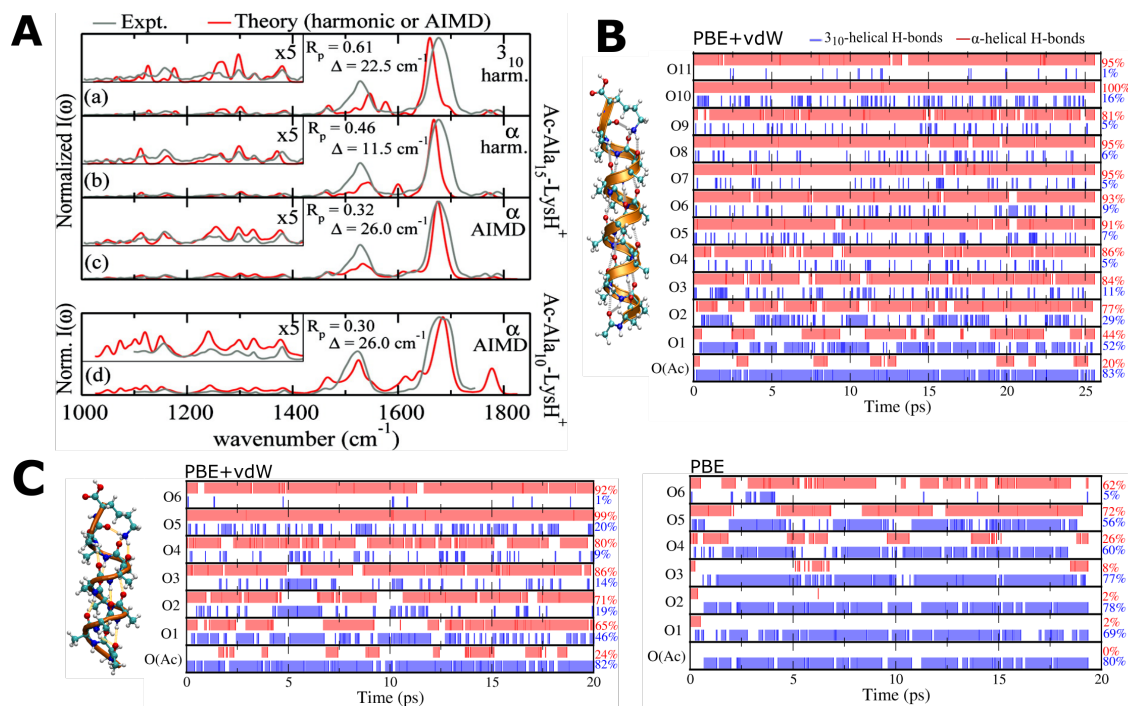
As an illustration of their work about the importance of anharmonicities in comparison to experiments, we highlight a larger peptide,  $\text{Ala}_7\text{H}^+$ , for which IRMPD spectra were measured by T. Vaden and coworkers [59]. In that study, T. Vaden and coworkers also performed extensive structural searches starting with a force field, then passing through a cascade of more accurate (standard) DFT functionals (until B3LYP), identifying conformational families, and finally performing single point calculations with MP2 for the energetically most favored conformers and calculating harmonic vibrations at the B3LYP level. The most likely globular structures, and the comparison of their harmonic IR spectra at the B3LYP level with the measured room temperature IRMPD spectrum is shown in Fig. 11 A and B (reproduced from Ref. [59]). M. P. Gaigeot and coworkers then took these structures and calculated IR spectra from *ab initio* molecular dynamics at the BLYP level and  $T=350\text{K}$  in Ref. [166]. The comparison between this anharmonic spectrum and the same experiment is shown in Fig. 11C, reproduced from Ref. [166]. It is immediately apparent that even if the agreement is not perfect, anharmonicities in this NH and CH stretch regions are necessary to reproduce the experimentally observed intensities below  $\approx 3100\text{ cm}^{-1}$ . The authors conclude that these structures adopt more globular conformations with the  $\text{NH}_3^+$  group self solvated within CO groups of the molecule. As will be shown below, the exact placement of the position of the simulated peaks with respect to experiment in the anharmonic case may be a fortuitous cancellation of errors, since the inclusion of van der Waals interactions can change considerably the dynamics of the molecule and inclusion of nuclear quantum effects cause large red shifts in this spectral region.

It is worth noting that intensities are typically not to be trusted when comparing theory and IRMPD experiments due to the strong non-linear effects expected in the multiple-photon absorption process. Attempts have been made by F. Calvo and coworkers to model specifically IRMPD [167] with all relevant dynamical effects, which can yield good results for small molecules albeit relying on some empirical modelling. Comparisons to IRPD would be interesting, since it is less prone to non-linearity in the lineshape and peak positions. However, the tag which is often used can also disturb the spectrum (as observed in Kr tagged gold clusters [168] and Ar tagged protonated water clusters [169]), and one is usually restrained to low temperatures due to the low binding energy of the tag. In most of the work present in the literature so far, it must be said, though, that the modelling of the IR spectra within linear response theory (including anharmonicity) has been able to provide important interpretations to vibrational signatures obtained from IRPD or IRMPD.

V. Blum and coworkers (including the authors of this review) have focused on the study of larger polypeptides, especially in the fundamental characterization of interactions governing structure formation and dynamics. For the benchmark series of helix-forming alanine based polypeptides  $\text{Ac-Ala}_n\text{-LysH}^+$  the authors have studied



**Figure 11.** A and B: Structures of Ala<sub>7</sub>H<sup>+</sup> and their corresponding harmonic IR spectra with the B3LYP functional, compared to the measured IRMPD, reproduced from Ref. [59]. Anharmonic IR spectrum (classical nuclei) with the BLYP functional (red) for the same molecule, compared to the experimental IRMPD spectrum (black), reproduced from Ref. [166].



**Figure 12.** A: Reproduced from Ref. [43]. Comparison between experimental [gray lines] and theoretical [red lines] (PBE+vdW functional) vibrational spectra, all normalized to 1 for the highest peak. (a,b) Ac-Ala<sub>15</sub>-LysH<sup>+</sup>: calculated spectra based on the harmonic approximation, for a 3<sub>10</sub>-helical (a) and an α-helical (b) local minimum of the potential-energy surface. (c) Ac-Ala<sub>15</sub>-LysH<sup>+</sup>: calculated spectrum from AIMD (including anharmonic effects), starting from an α-helix and α-helical in character throughout the simulation. (d) Same as panel c, for Ac-Ala<sub>10</sub>-LysH<sup>+</sup>. Pendry R-factors and rigid shifts Δ between measured and calculated spectra are included in each graph (calculated spectra are shifted by Δ for visual comparison). B: Illustration of the hydrogen bond network evolution of Ac-Ala<sub>15</sub>-LysH<sup>+</sup> during a PBE+vdW microcanonical simulation. On the right side of the plot, the ratios of α-helical and 3<sub>10</sub>-helical bonds observed during the simulation for each oxygen, labeled from N to C-terminus is shown. C: Illustration of the hydrogen bond network evolution of Ac-Ala<sub>10</sub>-LysH<sup>+</sup> during a PBE+vdW<sup>TS</sup> and a PBE microcanonical simulation (labels are the same as in B).

many different aspects related to secondary structure formation using DFT and *ab initio* molecular dynamics. Regarding the smaller members of this polypeptide series,  $n=4-8$ , the authors have reported that beyond the formation of stable H-bond chains with increasing  $n$ , an important contribution to helix stabilization comes from the vibrational entropy of very soft modes that are present in the helices but not in more compact structures [49]. Helices are predicted to be the most stable isolated structures in the gas phase starting at  $n=8$ , in agreement with experimental evidence from IMMS measurements [170].

For a more direct structural characterization, M. Rossi and coworkers have also calculated the (classical-nuclei) anharmonic IR spectra of  $n=5, 10$ , and  $15$ , and compared to experimental IRMPD measurements at room temperature [43]. In general, the

structural characterization of gas-phase peptides based on vibrational spectra requires an objective metric of agreement between simulation and experiment. To that end, Rossi and coworkers have employed the Pendry reliability factor  $R_P$  [171] in an implementation that was distributed with Ref. [172]. Since, as it was already discussed, the IRMPD spectra could have peak intensities that are distorted due to the absorption of many photons, a simple overall least squares fit for the intensities would not suffice for a comparison between theory and experiment. The Pendry R-factor, originally used in low energy electron diffraction experiments [171], addresses the need to match mainly peak positions, rather than the intensities. Given two continuous curves with intensities  $I_{exp}(\omega)$  and  $I_{th}(\omega)$ , this R-factor compares the renormalized logarithmic derivatives of the intensities, given by:

$$Y(\omega) = L^{-1}(\omega)/[L^{-2}(\omega) + W^2] \quad (10)$$

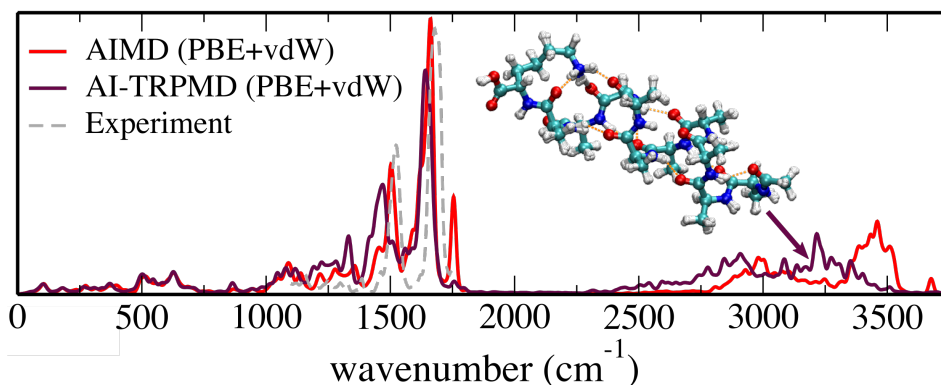
with  $L(\omega) = I'(\omega)/I(\omega)$ , and  $W$  approximately the half width of peaks in the spectra. The advantage is that the  $L$  functions have a sign inversion exactly where the maximum of the peak is, and if peaks are far enough apart, relative intensities are completely ignored, while if they are close together,  $L(\omega)$  is moderately sensitive. However, the  $L$  functions would be too sensitive to zeroes in the intensity, since the logarithmic derivatives would have singularities in this case. The  $Y$  function is a simple transformation of  $L$ , which avoids such singularities, by giving similar weights to maxima and zeroes in the intensities. The Pendry R-factor ( $R_P$ ) is then defined as:

$$R_P = \int d\omega (Y_{th} - Y_{exp})^2 / (Y_{th}^2 + Y_{exp}^2), \quad (11)$$

which leads in practice to values of  $R_P=0$  for perfect agreement,  $R_P=1$  for uncorrelated spectra, and  $R_P=2$  for complete anti-correlation.  $R_P$  is always defined with respect to a rigid shift  $\Delta$  between the two curves considered. A python script for the calculation of this and other reliability factors is available from Github.‡

We reproduce in Figure 12A the theoretical IR spectra obtained with DFT-PBE adding pairwise van der Waals corrections (PBE+vdW [89]) for helical structures of Ac-Ala<sub>10</sub>-LysH<sup>+</sup> and Ac-Ala<sub>15</sub>-LysH<sup>+</sup> compared to experiment. For  $n=15$  the comparison of the harmonic spectra of a helix containing mostly  $3_{10}$  helical H-bonds, another containing  $\alpha$ -helical H-bonds, and the anharmonic spectra obtained from Eq. 9 from PBE+vdW molecular dynamics shows (quantitatively) how the agreement to experiment increases in the anharmonic case. A Pendry reliability factor  $R_P$  of 0.32, obtained with respect to a rigid shift  $\Delta$  of the whole spectrum by 26 cm<sup>-1</sup> is an indication that the structure of this molecule is indeed the  $\alpha$ -helical one shown in Fig. 12B, where the lysine residue is completely self-solvated in the backbone carbonyl groups. Also in panel B, we show the H-bond dynamics of the molecule in the trajectory generating that spectrum, highlighting  $3_{10}$ - and  $\alpha$ -helical H bonds. Although fluctuations are observed,

‡ <https://github.com/mahrossi/r-factors>



**Figure 13.** Infrared absorption spectra of Ac-Ala<sub>10</sub>-LysH<sup>+</sup> calculated with ab initio molecular dynamics (AIMD-PBE+vdW) at 300K, with ab initio thermostatted ring polymer molecular dynamics[173] (TRPMD-PBE+vdW) at 300K, and the experimental IRMPD room-temperature spectrum from Ref. [43].

the molecule maintains a mostly  $\alpha$ -helical structure throughout. For Ac-Ala<sub>10</sub>-LysH<sup>+</sup> we also find a good agreement between the theoretical (anharmonic) and experimental IR spectrum for the  $\alpha$ -helix. Examining the dynamics of this molecule when switching off the vdW interactions, we can show in panel C that the structure becomes more extended, stabilizing a  $3_{10}$  helical motif, and worsening the agreement with experiment (shown only in Ref. [85]). This observation is also in line with a study of interplay between H-bond cooperativity and vdW contributions in polyalanine helices: H-bonds get systematically strengthened by vdW interactions, and the high temperature stability of Ac-Ala<sub>15</sub>-LysH<sup>+</sup> is increased, while at lower temperatures the lack of vdW interactions also stabilize a more extended  $3_{10}$ -helical structure [30].

The effect of the location of the charge and the peptide sequence was also studied for even larger alanine-based polypeptides, namely Ac-Ala<sub>19</sub>-LysH<sup>+</sup> and Ac-Lys-Ala<sub>19</sub>-H<sup>+</sup> [100]. Ac-Ala<sub>19</sub>-LysH<sup>+</sup> was seen to form helices, consistent with measured ion mobility cross sections. Ac-Lys-Ala<sub>19</sub>-H<sup>+</sup> presented cross sections consistent with more compact, globular conformers (as expected due to the unfavorable interaction of the charge with the possible helix macrodipole), but its IR spectrum was very similar to helical structures. Theoretical calculations could solve this puzzle: even if of a compact/globular nature, energetically favored conformers of Ac-Lys-Ala<sub>19</sub>-H<sup>+</sup> still retained a large helical content.

Here we take the opportunity to address a commonly adopted approximation in these simulations, namely that of performing dynamics considering classic nuclei. Hydrogen atoms, ubiquitous in these molecules, are quite quantum entities even at temperatures as high as room temperature. These effects are known to affect the structure and dynamics of condensed phase systems (especially water) [174, 175] and hydrogen bonds [176, 177]. A simulation technique that has been progressively gaining more attention to include nuclear quantum effects (NQE) beyond the harmonic approximation at least in non time-dependent observables is path integral molecular

dynamics (PIMD). This technique exploits an exact isomorphism between the statistical properties of a quantum system and that of a classical ring polymer, where each bead is a repetition of the original system, connected to each other by harmonic springs. A detailed explanation of this technique is beyond the scope of this manuscript, but good descriptions can be found in Refs. [178, 179]. This technique is especially suited to massively parallel architectures, since each replica of the system can be run in parallel given that there are enough CPUs available. For time-dependent observables, e.g. time correlation functions, the situation is much trickier, due to the difficulties of modelling true quantum dynamics. Also within path integral molecular dynamics there are a few approximations to time correlation functions that have been proposed, namely centroid molecular dynamics [180], ring polymer molecular dynamics [151], and thermostatted ring polymer molecular dynamics (TRPMD) [173]. Albeit approximate these methods can give reliable results especially for larger systems and/or extended systems [181], and are the only methods so far that can be applied on a more routine basis to realistic multi-dimensional systems. At room temperature, even for the most efficient of these methods, one must use several tens of replicas of the system, making these simulations still substantially more costly than their classical-nuclei counterparts.

We used TRPMD to calculate the IR spectrum of Ac-Ala<sub>10</sub>LysH<sup>+</sup>, shown in Fig. 13. We used the FHI-aims program package [80] in connection to the i-PI program [182] in order to perform the dynamics. We simulated 20 ps of TRPMD dynamics, starting from the thermalized  $\alpha$ -helical structure, a time step of 0.5 fs for the integration, 16 replicas of the system (beads), and light settings in FHI-aims for the PBE+vdW force evaluation. In Fig. 13 we compare the IR spectrum thus obtained with the AIMD-PBE+vdW spectrum (tight settings, without any shifts applied) and the IRMPD room temperature experimental spectrum already published in Ref. [43]. We observe that while for very low frequency modes the classical and quantum nuclei simulations agree pretty well, above 1000 cm<sup>-1</sup> most of the modes are softened (red-shifted) in the quantum case, something that becomes progressively more pronounced for all modes above 2500 cm<sup>-1</sup>. This observation is in line with the fact that higher frequency modes are more quantum in nature. Even if TRPMD is known to over-broaden the line-shapes [173], the red-shifts should be reliable, modulo the limitations of the DFT functional itself (lower barriers, softer H-bonds). As also shown in Fig. 13, this effect goes in the opposite direction of the experimental data, which is already slightly blue shifted from the classical nuclei simulation. This is an indication that the PBE+vdW functional itself is here at fault. We also know that in these systems, when calculating harmonic frequencies of vibration with, e.g., the PBE0+vdW functional, they are all blue shifted with respect to PBE+vdW. The over-softening of the modes is one more manifestation of the self-interaction problem. It seems, thus, that in order to get better agreement of peak positions with experiment in a fully anharmonic picture, one should perform a simulation with van der Waals corrected hybrid functionals (which are, unfortunately, considerably more expensive than standard generalized gradient ones) and include nuclear quantum effects.

So far, only studies of polypeptides in isolation have been discussed. As mentioned in the introduction, the gas phase is ideal not only due to its “clean room” conditions, but also to the fact that it is straightforward to control the gradual inclusion of “external agents”, as for example ions, metal cations and small metallic clusters, and solvent molecules, for example water. We dedicate a following section to the discussion of microsolvation. Here we briefly review the interaction with ions. Since the early 2000s, IMS experiments have pointed to the role of cations stabilizing helical structures in polyalanine peptides [183], and more recently evidence for helix stabilization has been established based on the measurement of gas phase IRMPD spectra in the Amide A/B range of sodiated polyalanine peptides of various sizes [184]. Through measurement of IR spectra, also the role of metal cations to stabilize the zwitterionic form of some amino acids in the gas phase has been studied [18].

We had a detailed look at the effect of small cations ( $\text{Li}^+$  and  $\text{Na}^+$ ) on the structure of prototypical turn-forming peptides Ac-Ala-Ala-Pro-Ala-NMe and Ac-Ala-Asp-Pro-Ala-NMe [73]. The different systems were investigated by means of theoretical and experimental vibrational spectroscopy. First of all it was evident that in the gas phase, the interaction of the peptide carbonyl groups with the strong positive charge of the cations enforces conformations on the backbone that would not be possible for the peptide alone. Furthermore, the preferred conformations differ depending on the cation. The comparison between experimental and simulated spectra revealed that multiple conformers co-exist and probably interconvert in the gas phase. Consequently, the computed spectra for individual conformers have to be mixed in order to match the spectra recorded in the experiments, but a good agreement is reached. One can raise the question of how relevant are these results in solution. Hints come from short *ab initio* MD simulations that were performed on energetically stable conformations of peptide-cation systems with a few dozens of waters. Within the time scales accessible, the interactions between the cation and the peptide backbone remained preferred over direct solvation of the cation by the water molecules.

#### 4.3. Towards First-Principles Free Energies

Even if the PES is really the basis for all thermodynamic quantities, the sole knowledge of the PES does not allow a direct connection with real-world physics. For equilibrium properties, what is really needed is a good estimate of the partition function from statistical mechanics and all thermodynamic quantities that can be derived from it, most importantly, free energies.

Unfortunately, estimating free energy values for biomolecules is not an easy task. The harmonic approximation for the free energy (discussed in many textbooks, e.g. [150]), is the most common approximation. The reason is that it is the only one feasible with more costly (e.g. first principles) potentials and for larger molecular sizes. Due to the anharmonic nature of these molecules, it is not guaranteed though that this approximation will be plausible even at relatively low temperatures.

In order to get vibrational contributions to the free energy it is possible to use, for example, the VSCF and VPT2 methods, already discussed in the last section. For small molecules, M. Basire and coworkers have developed a technique which relies on the estimation of microcanonical densities of states and partition functions, that gives access to temperature effects and relative populations connected to a second order vibrational perturbation theory [159, 158] approach. However for higher dimensional and flexible systems this technique becomes very challenging. Quasi-harmonic analysis, in which dynamics can be decomposed into principal components and entropies calculated from this decomposition can be used as an approximation, provided there is enough sampling, but again, they rely on a quasi-harmonic picture that is likely to fail in many situations.

We have shown in the previous sections that it is possible to extract, for example, vibrational spectra from first-principles molecular dynamics (MD) simulations. However, the estimation of (relative) free energies requires a sampling of the conformational space that can currently only be realized for rather small molecular systems with few well defined degrees of freedom [185]. For larger systems, with hundreds of atoms, it is a much larger (and close to impossible) effort to gather the required statistical sampling of conformational space in order to estimate these free energies. It is worth noting though that with smart algorithms and optimized codes these quantities are becoming accessible [186]. There are two main points to be addressed [187]: (i) The simulation has to be long enough to ensure that the time-average of the simulations resembles the ensemble average of the system and (ii) free energies from MD simulations require the definition of collective variables, that are not trivial to define. In the field of biomolecular simulations a variety of MD-based simulation techniques are being used to solve point (i), we only summarize some frequently used types here:

- A straightforward approach is the computation of long ( $\mu\text{s}$  to  $\text{ms}$  time-scale) trajectories. This idea brought to the extreme is the construction of dedicated hardware like the molecular-dynamics supercomputer Anton [188] that provides access to the kinetics and thermodynamics of, for example, protein folding [189, 190].
- Alternatively, many short MD trajectories can be combined by using Markov-chain models [191, 192, 193, 194, 195]. This approach is striking because it is inherently parallel and allows the use of distributed computational resources [196, 197].
- The necessity of either very long or large numbers of independent shorter MD simulations comes from the nature of the transitions between the different metastable states on the free-energy landscape of a given system. These transitions are often rare events and in order to obtain converged values, these events have to be observed sufficiently often. In order to enhance sampling and therewith shorten the required simulation times, multiple methods are available: replica-exchange MD, umbrella sampling [198, 199, 200, 201], metadynamics [202, 203], *etc.*

One or several collective variables are needed as degrees of freedom (DOF) that define the free-energy surface. In case of, e.g., umbrella sampling or meta-dynamics,



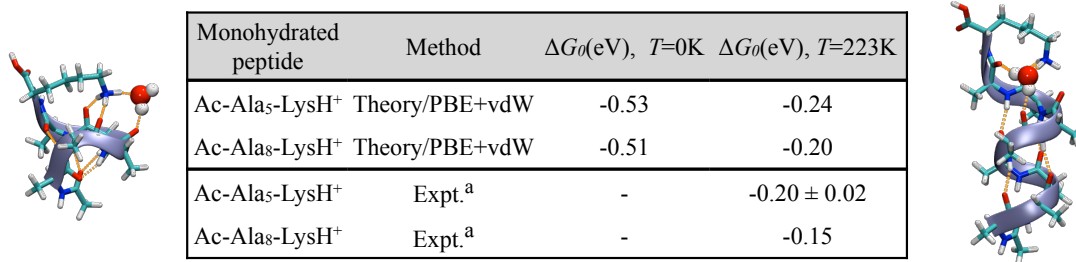
these collective variables have to be known *a priori*, while they can be defined *a posteriori* in non-biased MD simulations. Overall, it would be interesting to pursue methods that can be even more efficient in sampling, or methods that can reach convergence with a small amount of statistics.

## 5. Challenges Towards Solvation

A biomolecule immersed in a solvent presents three different qualitative types of interactions that need to be described. These are the intramolecular interactions, the biomolecule-solvent interactions, and the intra- and inter-molecular interactions of the solvent. The interactions between the biomolecule and the solvent and the influence of the collective interactions of the solvent on the biomolecule are the ones that will ultimately define the solvated state. It is important to note that the solvent is often not a simple homogeneous environment, but includes ions and other inhomogeneities that also need to be accurately captured. Studying biomolecules directly in solution has the drawback that the resulting measurements are quite congested by the amount of different interactions that play a role. It is thus desirable to build up the solvated state step by step, so that theory and experiment can work in synergy towards a consistent and reliable description of these molecules in solution.

Experimentally, regarding the solute-solvent and solvent-solvent interactions, perhaps the most detailed characterizations of physical properties are connected to mass spectrometry (MS), where it is possible to perform thermochemical equilibrium measurements [204] and, if connected to spectroscopy techniques, to measure also more detailed structural information. In these experiments, solvation with water molecules or ions (or both) can be investigated in a stepwise manner, such that the physical properties of the very first stages of solvation can be identified. For example, it is possible to measure equilibrium constants, binding enthalpies, and vibrational spectra that can be directly connected to calculations.

Using only IM-MS, thermochemical equilibrium properties and overall geometric information have been gathered for a range of biomolecules and the first stages of their interaction with the solvent (microsolvation) [205, 183, 206, 207, 208, 3, 209]. A review in this area can be found in [3]. More recently, also the measurement of vibrational spectra of mass selected species in the gas phase were able to probe more detailed conformational properties of clusters of solvent molecules [210, 211, 212, 213, 214, 215, ] or the first stages solvation especially of peptides [216], and sugars [217, 218, 6]. We highlight here two recent experimental works dealing with peptides to illustrate the state of the field. Impressive results have been reported by N. Nagornova and coworkers [216] on the microsolvation of Gramicidin S cooled to 12 K. By performing conformer selective double resonance IR-UV spectroscopy they were able to connect IR features to structural changes caused by the absorption of 1 to 15 water molecules. Another work by S. Warnke and coworkers [205] instead used ion mobility-mass spectrometry to show how crown-ethers can micro-solvate charged Lys side chains in cytochrome-C

Ac-Ala<sub>5</sub>-LysH<sup>+</sup> · H<sub>2</sub>OAc-Ala<sub>8</sub>-LysH<sup>+</sup> · H<sub>2</sub>O


Monohydrated peptide	Method	$\Delta G_0(\text{eV}), T=0\text{K}$	$\Delta G_0(\text{eV}), T=223\text{K}$
Ac-Ala <sub>5</sub> -LysH <sup>+</sup>	Theory/PBE+vdW	-0.53	-0.24
Ac-Ala <sub>8</sub> -LysH <sup>+</sup>	Theory/PBE+vdW	-0.51	-0.20
Ac-Ala <sub>5</sub> -LysH <sup>+</sup>	Expt. <sup>a</sup>	-	$-0.20 \pm 0.02$
Ac-Ala <sub>8</sub> -LysH <sup>+</sup>	Expt. <sup>a</sup>	-	-0.15

<sup>a</sup> Kohtani and Jarrold, *JACS* 126, 8454 (2004), values converted from  $K_1$  equilibrium constants that were read from Figure 2. Error bars from our estimate.

**Figure 14.** Calculated  $\Delta G_0(T)$  (in eV, and corresponding to a reference pressure of  $p_0 = 1.01325 \times 10^5 \text{ Pa} = 760 \text{ Torr}$ ) for monohydration of Ac-Ala<sub>5</sub>-LysH<sup>+</sup> and Ac-Ala<sub>8</sub>-LysH<sup>+</sup> compared to literature data. Also shown, the most stable conformations of monohydrated Ac-Ala<sub>5</sub>-LysH<sup>+</sup> and Ac-Ala<sub>8</sub>-LysH<sup>+</sup> from theory (PBE+vdW). Values and structures from Ref. [125].

and other proteins. The authors were able to decompose the effects responsible for the unfolding of highly-charged states in the gas phase into Coulomb repulsion and side chain to backbone interactions that interrupt backbone hydrogen bonding.

Experiments nowadays are able to provide more and more accurate data on thermochemical and structural properties of (micro)solvated biomolecules, but without the support of theoretical calculations, the understanding of the results is limited. It is not straightforward to obtain quantitative data for these systems from simulations, though. The difficulties are at least two fold: (i) One still has the high conformational freedom of the biomolecule itself, but now further complicated by the presence of ions and solvent which introduce an extra range of qualitatively different interactions to be modeled; (ii) It is known to be difficult to simulate even the solvent in isolation, with most quantum chemical methods failing to correctly describe overall structural properties like radial distribution functions, or diffusion coefficients [219, 220, 221, 222, 223, 224, 225], or the correct relative energies of hydrogen bonded structures [226, 227].

The main challenge is to correctly and thoroughly explore the potential energy surface (PES) *and* the entropic contributions to the free energy – even more important when related to the solvent. These simulations must involve an accurate evaluation of the potential energy and span a long time scale (or a huge volume of phase space). Unfortunately nowadays one can have either one or the other: an accurate evaluation of points in the PES can be achieved by the highest-level quantum chemistry methods but these are too computationally expensive to allow a thorough sampling of the PES, while empirical potentials allow a thorough sampling of the PES but do not provide quantitative estimates. It is also important to note that only describing the electronic structure of these systems is not enough – especially in connection with the solvent, the inclusion of nuclear quantum effects beyond the harmonic approximation is necessary [228, 229, 230, 231, 232, 233].

Nevertheless some successes from theory have been achieved for the microsolvation of model peptides, for example the Ac-Ala<sub>n</sub>-LysH<sup>+</sup> series already mentioned in this review. The groups of M. Bowers [209] as well as of M. Jarrold [208] performed IM-MS experiments for the monohydrated structures of a few conformers (different sizes) of this peptide series. In these experiments, they had access to equilibrium constants of the monohydration reaction, derived from the ratio between the intensity of the peaks corresponding to the bare and the monohydrated structures. Based on previous observations that more globular/compact structures had a lower propensity to adsorb one water molecule than helical ones, they concluded that the shortest helical member of this series would happen at  $n=8$  – without thorough theoretical support, it is difficult though to understand what is the atomistic mechanism for this difference in water adsorption propensity. In Ref. [125], S. Chutia and coworkers have performed extensive first principles conformational scans of  $n=5$  and  $n=8$  microsolvated by up to 5 water molecules. One conclusion is that the intramolecular hydrogen bonds of the self-solvated ammonium group, in both cases, are the most stable hydration sites. For one water molecule the most stable conformers are shown in Fig. 14, together with the calculated standard (Gibbs free) energy of formation  $\Delta G_0$  of the reaction. The agreement with experimental values is pretty good (also at other temperatures, shown in Ref. [125]). From the theoretical work, the authors concluded that the decrease in water adsorption propensity is not due to a radically different binding site, but instead only to modified internal free energy contributions (harmonic vibrational free energy) in the specific H<sub>2</sub>O adsorption site at the LysH<sup>+</sup> termination, in an example of how theory can help to gather a deeper understanding of experimental data. However, it is still a challenge for theory to be able to give even more reliable results for larger peptides surrounded by more solvent molecules.

It is thus pressing to build a tighter relationship between the quantum and the empirical world. While for water there is an appreciable effort to build better and more accurate potentials based on quantum mechanical calculations [234, 235, 236, 229], for the solvent-biomolecule (or ion-biomolecule) interaction these efforts are much less pronounced. An improvement in this area can be achieved precisely by performing these theory-experiment benchmarks of the stepwise build-up of solvation, and modifying empirical potentials according to this data.

## 6. Conclusions

The aim of this review was to give an overview on the interplay of experiment and simulation regarding the structure and dynamics of biomolecules in the gas phase. Given the scientific fields of the authors, the focus was clearly on first-principles calculations on peptides towards the computation of physical observables like vibrational spectra and collision cross sections. For flexible molecular systems, for which biomolecules are a prime example, a thorough search of the accessible conformational space is crucial before any attempt to compare simulated properties with their experimental counterpart. A

typical work flow is outlined in the following (and in Figure 9):

- (i) The exact chemical structure (connectivity of the atoms) of the molecular system has to be known. This includes knowledge about possible alternative protonation states (protomers). In cases where, for example, cations like  $\text{H}^+$  or  $\text{Na}^+$  are involved, their presence and location relative to the molecule has to be considered as well.
- (ii) An initial enumeration of structural candidates can be performed by the sampling of a computationally-cheap potential-energy surface (PES), for example of an empirical force field.
- (iii) As we have outlined in this review, the limited accuracy of force-field methods requires a refinement at the level of electronic-structure theory. This can be facilitated by using density-functional theory (DFT) methods or quantum-chemistry methods like Møller-Plesset perturbation theory (MP2). Higher-level methods, like coupled cluster, quantum Monte Carlo, or full configuration interaction, are computationally very demanding and thus normally limited to small systems and benchmark-type calculations.
- (iv) In order to remove a possible bias from the initial sampling of the force-field based PES, further exploration of the first-principles PES in the proximity of already located low-energy structures is advisable. This can be facilitated by, for example, (replica-exchange) *ab initio* molecular dynamics simulations.
- (v) Free-energy estimations in the harmonic approximation should be considered, not the least because they also offer a first glimpse at the vibrational spectrum of the molecular system. Further MD-based sampling can potentially be used to obtain more accurate thermodynamical observables (free energies, enthalpies, etc.). However, the size of structure space and the computational cost of the required converged simulations again restrict such approaches to either rather small or rigid molecular systems.
- (vi) The comparison to experiment serves as (i) validation of the method (search strategy and energy function) and (ii) as a way to add structural resolution to the experiment. Both can be achieved by the computation of physical observables, e.g. collision cross sections, vibrational spectra, optical spectra, *etc.*

Each simulation represents an approximation to reality and inherently produces errors. The gas phase is a clean-room environment and gas-phase experiments can produce accurate and sharp data that represents a challenge to theory and simulation. We would dare to say that the higher signal-to-noise ratio that is present in condensed-phase experiments might actually cover some of the involved systematic errors in the theoretical description. This highlights the importance of the gas phase as an ideal environment for validating energy functions and simulation techniques.

An important point that we can conclude is that it is not sufficient to focus on a single or a few structures, given the complex dynamics observed in the gas phase (and even more so in solution). Most of the larger sources of uncertainties in the theoretical

treatment have to do with an insufficient or still inaccurate treatment of dynamics. If an accurate free-energy surface could be accessed and sampled, most of the remaining problems would be solved. This would allow, for example, the correct prediction of the conformational ensembles observed in ion-mobility measurements (CCS/ATD) or in vibrational spectroscopy. In addition, it would give access to reliable barriers and a natural inclusion of anharmonic effects in vibrational spectra. In order to reach this goal, we need to compute potential energies and forces including the correct physics, which then need to be sampled faster and for long time scales. We note that the correct physics may go even beyond just grasping the physics of the electronic structure but also the quantum nature of the nuclei, which can cause much stronger anharmonicities (as shown in this review) and change considerably effective barrier heights. Going even further, for these highly anharmonic and high-dimensional systems, in many situations the dynamics of nuclei and electrons are coupled. These non-adiabatic effects are truly difficult to treat from a theoretical point of view in these structures.

The efficient exploration of conformational space for high-dimensional flexible systems in an accurate manner thus poses one of the most pressing issues in this field. For it to be solved, either the accuracy of force fields must be improved, or the computational limitations of first-principles methods, when it comes to larger length scales and longer time scales, needs to be lifted. Possible routes that can be followed in methodological developments involve, for example, better parametrization of force fields based on the increasing number of first-principles data present in the literature, development of smarter free energy evaluation methods that can deal with fewer statistical sampling, and/or even better scaling of first-principles codes in massively parallel architectures. As these issues are already recognized by the community, several efforts in all fronts are paving the way to treat larger systems with state-of-the-art accuracy (e.g. Refs. [76, 100, 188, 237, 238, 239, 240, 241] and many others).

Nevertheless, as it has been shown in this review, both the time and length scale currently accessible to first-principles methods already allow an accurate treatment of systems with hundreds of atoms in simulations. On the experimental side, it is routinely possible to transfer large biomolecules, e.g. large proteins and even complexes, to the gas phase by electrospray ionization and to study them by mass spectrometry and ion mobility-mass spectrometry (IM-MS) [11]. However, with the size of the molecular systems, vibrational spectroscopy investigations get hindered by more and more congested spectra. A promising route that is currently being followed to circumvent this problem is to measure conformer selective spectra by either using (i) UV/IR double-resonance techniques and (ii) pre-selecting conformers by using IM-MS. A way to get sharper spectra is to measure them at low temperatures for example by using either cold-ion traps [64, 65] or helium droplets [66, 67]. Conformational selection and cold-ion spectroscopy can also be combined.

The investigation of biomolecules in the gas phase is a dynamically growing field and a constant challenge to experimentalists and theorists alike. The constant developments and improvements of experimental techniques trigger the use of more and

more sophisticated simulations and *vice versa*. As such this line of research pushes our understanding of the very basics of biomolecular structure formation and dynamics. For the development of simulation methods, the precise data that can be obtained from gas-phase experiments is ideal to develop and test new methodologies that will also have an impact in condensed-phase simulation.

## 7. Acknowledgements

We are grateful to our current and previous colleagues in the “Bio Group”: Sucismita Chutia, Matti Ropo, Jürgen Wiefierink, Franziska Schubert, Markus Schneider, Mateusz Marianski, Adriana Supady and the previous group leader Volker Blum. We acknowledge the continuous support of our work by Matthias Scheffler and the inspiring discussions on the experimental side, especially Gert von Helden, Kevin Pagel, Lucy Voronina, Tom Rizzo, and Knut Asmis. M.R. acknowledges funding from the German Research Foundation (DFG) under project RO 4637/1-1 and office space in the group of Prof. David Manolopoulos.

## 8. References

- [1] Jarrold M 2000 *Annu. Rev. Phys. Chem.* **51** 179–207
- [2] Jarrold M F 2007 *Phys. Chem. Chem. Phys.* **9**(14) 1659–1671 see also references therein.
- [3] Wyttenbach T and Bowers M T 2009 *Chem. Phys. Lett.* **480** 1–16
- [4] Simons J 2003 *C. R. Chim.* **6** 17–31
- [5] Simons J, Jockusch R, ÇarÇabal P, Hünig I, Kroemer R, Macleod N and Snoek L 2005 *Int. Rev. Phys. Chem.* **24**
- [6] Simons J 2009 *Mol. Phys.* **107** 2435–2458
- [7] De Vries M S and Hobza P 2007 *Annu. Rev. Phys. Chem.* **58** 585–612
- [8] Schermann J P 2008 *Spectroscopy and Modeling of Biomolecular Building Blocks* (Amsterdam: Elsevier)
- [9] Rijs A M and Oomens J 2015 Ir spectroscopic techniques to study isolated biomolecules *Gas-Phase IR Spectroscopy and Structure of Biological Molecules (Topics in Current Chemistry vol 364)* ed Rijs A M and Oomens J pp 1–42
- [10] Zhou M, Dong X, Baldauf C, Chen H, Zhou Y, Springer T A, Luo X, Zhong C, Gräter F and Ding J 2011 *Blood* **117** 4623–4631
- [11] Marcoux J and Robinson C 2013 *Structure* **21** 1541–1550
- [12] Abi-Ghanem J and Gabelica V 2014 *Phys. Chem. Chem. Phys.* **16** 21204–21218
- [13] Arcella A, Dreyer J, Ippoliti E, Ivani I, Portella G, Gabelica V, Carloni P and Orozco M 2015 *Angew. Chem. Int. Ed.* **54** 467–471
- [14] Werz D B, Ranzinger R, Herget S, Adibekian A, von der Lieth C W and Seeberger P H 2007 *ACS Chem. Biol.* **2** 685–691
- [15] Karas M, Bachmann D, Bahr U and Hillenkamp F 1987 *Int. J. Mass Spectrom.* **78** 53 – 68 ISSN 0168-1176
- [16] Fenn J, Mann M, Meng C, Wong S and Whitehouse C 1989 *Science* **246** 64–71
- [17] Aebersold R and Mann M 2003 *Nature* **422** 198 – 207
- [18] Polfer N C and Oomens J 2009 *Mass Spectrometry Reviews* **28** 468–494 ISSN 1098-2787
- [19] Smalley R E, Wharton L and Levy D H 1977 *Accounts of Chemical Research* **10** 139–145
- [20] Meyer T, Gabelica V, Grubmüller H and Orozco M 2013 *Wiley Interdisciplinary Reviews: Computational Molecular Science* **3** 408–425 ISSN 1759-0884

- [21] Barran P E, Polfer N C, Campopiano D J, Clarke D J, Langridge-Smith P R, Langley R J, Govan J R, Maxwell A, Dorin J R, Millar R P and Bowers M T 2005 *Int. J. Mass Spectrom.* **240** 273–284
- [22] Benesch J L P and Robinson C V 2009 *Nature* **462** 576–577
- [23] de Vries M S and Hobza P 2007 *Annu. Rev. Phys. Chem.* **58** 585–612
- [24] Rizzo T R, Stearns J A and Boyarkin O V 2009 *Int. Rev. Phys. Chem.* **28** 481–515
- [25] Gaigeot M P 2010 *Phys. Chem. Chem. Phys.* **12**(14) 3336–3359 see also references therein.
- [26] Simons J P 2004 *Phys. Chem. Chem. Phys.* **6**(10) E7
- [27] McDaniel E W, Martin D W and Barnes W S 1962 *Rev. Sci. Instrum.* **33** 2–7
- [28] Mason E A and Schamp Jr H W 1958 *Ann. Phys.* **4** 233–270
- [29] Kohtani M, Jones T C, Schneider J E and Jarrold M F 2004 *J. Am. Chem. Soc.* **126** 7420–7421
- [30] Tkatchenko A, Rossi M, Blum V, Ireta J and Scheffler M 2011 *Phys. Rev. Lett.* **106**(11) 118102
- [31] Shelimov K B and Jarrold M F 1996 *J. Am. Chem. Soc.* **118** 10313–10314
- [32] Warnke S, Baldauf C, Bowers M T, Pagel K and von Helden G 2014 *J. Am. Chem. Soc.* **136** 10308–10314
- [33] Pierson N A, Valentine S J and Clemmer D E 2010 *J. Phys. Chem. B* **114** 7777–7783
- [34] Pierson N A, Chen L, Valentine S J, Russell D H and Clemmer D E 2011 *J. Am. Chem. Soc.* **133** 13810–13813
- [35] Servage K A, Silveira J A, Fort K L and Russell D H 2014 *J. Phys. Chem. Lett.* **5** 1825–1830
- [36] Stedwell C N, Galindo J F, Roitberg A E and Polfer N C 2013 *Annu. Rev. Anal. Chem.* **6** 267–285
- [37] Heine N and Asmis K R 2015 *Int. Rev. Phys. Chem.* **34** 1–34
- [38] Borysow J, Moraldi M and Frommhold L 1985 *Mol. Phys.* **56** 913–922
- [39] Oomens J, van Roij A J A, Meijer G and von Helden G 2000 *Astrophys. J.* **542** 404
- [40] Heine N, Yacovitch T I, Schubert F, Brieger C, Neumark D M and Asmis K R 2014 *J. Phys. Chem. A* **118** 7613–7622
- [41] Yacovitch T I, Heine N, Brieger C, Wende T, Hock C, Neumark D M and Asmis K R 2013 *J. Phys. Chem. A* **117** 7081–7090
- [42] Weymuth T, Jacob C R and Reiher M 2010 *J. Phys. Chem. B* **114** 10649–10660
- [43] Rossi M, Blum V, Kupser P, von Helden G, Bierau F, Pagel K, Meijer G and Scheffler M 2010 *J. Phys. Chem. Lett.* **1** 3465–3470
- [44] Jaeqx S, Oomens J, Cimas A, Gaigeot M P and Rijs A M 2014 *Angew. Chemie.* **126** 3737–3740
- [45] Tanabe K, Miyazaki M, Schmies M, Patzer A, Schütz M, Sekiya H, Sakai M, Dopfer O and Fujii M 2012 *Angew. Chem. Int. Ed.* **51** 6604–6607
- [46] Fricke H, Funk A, Schrader T and Gerhards M 2008 *J. Am. Chem. Soc.* **130** 4692–4698
- [47] Stearns J A, Seaiby C, Boyarkin O V and Rizzo T R 2008 *Phys. Chem. Chem. Phys.* **11** 125
- [48] Zabuga A V and Rizzo T R 2015 *J. Phys. Chem. Lett.* **6** 1504–1508
- [49] Rossi M, Scheffler M and Blum V 2013 *J. Phys. Chem. B* **117** 5574–5584
- [50] Chin W, Mons M, Dognon J P, Piuzzi F, Tardivel B and Dimicoli I 2004 *Phys. Chem. Chem. Phys.* **6** 2700–2709
- [51] Garand E, Kamrath M Z, Jordan P A, Wolk A B, Leavitt C M, McCoy A B, Miller S J and Johnson M A 2012 *Science* **335** 694–698
- [52] Dian B C, Clarkson J R and Zwier T S 2004 *Science* **303** 1169–1173
- [53] Compagnon I, Oomens J, Meijer G and von Helden G 2006 *J. Am. Chem. Soc.* **128** 3592–3597
- [54] Rijs A M, Ohanessian G, Oomens J, Meijer G, von Helden G and Compagnon I 2010 *Angewandte Chemie International Edition* **49** 2332–2335
- [55] Blom M N, Compagnon I, Polfer N C, von Helden G, Meijer G, Suhai S, Paizs B and Oomens J 2007 *Journal Of Physical Chemistry A* **111** 7309–7316
- [56] Barnes L, Schindler B, Allouche A R, Simon D, Chambert S, Oomens J and Compagnon I 2015 *Physical Chemistry Chemical Physics* DOI: 10.1039/C5CP02079D
- [57] Nicely A L and Lisy J M 2011 *Journal Of Physical Chemistry A* **115** 2669–2678
- [58] Nicely A L, Miller D J and Lisy J M 2009 *Journal of the American Chemical Society* **131** 6314–

- 6315
- [59] Vaden T D, de Boer T S J A, Simons J, Snoek L C, Suhai S and Paizs B 2008 *J. Phys. Chem. A* **112** 4608–4616
- [60] Vaden T D, Gowers S A N and Snoek L C 2009 *J. Am. Chem. Soc.* **131** 2472–2474
- [61] Papadopoulos G, Svendsen A, Boyarkin O V and Rizzo T R 2011 *Faraday Discuss.* **150** 243–255
- [62] Voronina L and Rizzo T R 2015 *Phys. Chem. Chem. Phys.* Advance Article
- [63] Warnke S, Seo J, Boschmans J, Sobott F, Scrivens J H, Bleiholder C, Bowers M T, Gewinner S, Schöllkopf W, Pagel K and von Helden G 2015 *J. Am. Chem. Soc.* **137** 4236–4242
- [64] Burke N L, Redwine J G, Dean J C, McLuckey S A and Zwier T S 2015 *Int. J. Mass Spectrom.* **378** 196–205
- [65] Wassermann T N, Boyarkin O V, Paizs B and Rizzo T R 2012 *J. Am. Soc. Mass Spectrom.* **23** 1029–1045
- [66] Toennies J P and Vilesov A F 2004 *Angew. Chem. Int. Ed.* **43** 2622–2648
- [67] Flórez A I G, Ahn D S, Gewinner S, Schöllkopf W and von Helden G 2015 *Phys. Chem. Chem. Phys.* **17** 21902–21911
- [68] Ballard A J, Martiniani S, Stevenson J D, Somani S and Wales D J 2015 *WIREs Comput. Mol. Sci.* **5** 273–289
- [69] Penev E, Ireta J and Shea J E 2008 *J. Phys. Chem. B* **112** 6872–6877
- [70] Rossi M, Chutia S, Scheffler M and Blum V 2014 *J. Phys. Chem. A* **118** 7349–7359
- [71] Valdes H, Pluhackova K, Pitonak M, Rezac J and Hobza P 2008 *Phys. Chem. Chem. Phys.* **10** 2747–2757
- [72] Tzanov A T, Cuendet M A and Tuckerman M E 2014 *J. Phys. Chem. B* **118** 6539–6552
- [73] Baldauf C, Pagel K, Warnke S, von Helden G, Koksche B, Blum V and Scheffler M 2013 *Chem. Eur. J.* **19** 11224–11234
- [74] Rossi M, Tkatchenko A, Rempe S B and Varma S 2013 *Proc. Natl. Acad. Sci. USA* **110** 12983
- [75] Vitalini F, Mey A S J S, Noé F and Keller B G 2015 *J. Chem. Phys.* **142** 084101
- [76] Riplinger C and Neese F 2013 *J. Chem. Phys.* **138** 034106
- [77] Liakos D G, Sparta M, Kesharwani M K, Martin J M L and Neese F 2015 *J. Chem. Theory Comput.* **11** 1525–1539
- [78] Genovese L, Neelov A, Goedecker S, Deutsch T, Ghasemi S A, Willand A, Caliste D, Zilberberg O, Rayson M, Bergman A and Schneider R 2008 *J. Chem. Phys.* **129** 014109
- [79] Haynes P D, Mostof A A, Sklyaris C K and Payne M C 2006 *J. Phys. Conf. Ser.* **26** 143
- [80] Blum V, Gehrke R, Hanke F, Havu P, Ren X, Reuter K and Scheffler M 2009 *Comp. Phys. Comm.* **180** 2175–2196
- [81] Clark S J, Segall M D, Pickard C J, Hasnip P J, Probert M J, Refson K and Payne M C 2005 *Z. Kristallogr.* **220** 567–570
- [82] Hutter J, Iannuzzi M, Schiffmann F and VandeVondele J 2014 *WIREs Comput. Mol. Sci.* **4** 15–25
- [83] Valdes H, Pluhackova K, Pitonak M, Rezac J and Hobza P 2008 *Phys. Chem. Chem. Phys.* **10** 2747–2757
- [84] DiStasio R A, Steele R P, Rhee Y M, Shao Y and Head-Gordon M 2007 *J. Comput. Chem.* **28** 839–856
- [85] Rossi M 2011 *Ab initio study of alanine-based polypeptide secondary-structure motifs in the gas phase* Ph.D. thesis Technical University Berlin and Fritz Haber Institute URL <http://opus.kobv.de/tuberlin/volltexte/2012/3429/>
- [86] Schwabe T and Grimme S 2007 *Phys. Chem. Chem. Phys.* **9**(26) 3397–3406
- [87] Klimeš J and Michaelides A 2012 *J. Chem. Phys.* **137** 120901
- [88] Perdew J, Burke K and Ernzerhof M 1996 *Phys. Rev. Lett.* **77** 3865–3868
- [89] Tkatchenko A and Scheffler M 2009 *Phys. Rev. Lett.* **102** 073005
- [90] Tkatchenko A, DiStasio R A, Car R and Scheffler M 2012 *Phys. Rev. Lett.* **108**(23) 236402
- [91] DiStasio Jr R A, Gobre V V and Tkatchenko A 2014 *J. Phys. Condens. Matter* **26** 213202
- [92] Adamo C and Barone V 1999 *J. Chem. Phys.* **110** 6158–6170



- [93] Jorgensen W L, Maxwell D S and Tirado-Rives J 1996 *J. Am. Chem. Soc.* **118** 11225–11236
- [94] Hornak V, Abel R, Okur A, Strockbine B, Roitberg A and Simmerling C 2006 *Proteins* **65** 712–725
- [95] Mackerell A D, Feig M and Brooks C L 2004 *J. Comput. Chem.* **25** 1400–1415
- [96] MacKerell A D, Bashford D, Bellott M, 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, Wiorkiewicz-Kuczera J, Yin D and Karplus M 1998 *J. Phys. Chem. B* **102** 3586–3616
- [97] Ren P and Ponder J W 2003 *J. Phys. Chem. B* **107** 5933–5947
- [98] Ponder J W and Case D A 2003 Force fields for protein simulations *Protein Simulations (Advances in Protein Chemistry* vol 66) ed Daggett V (Academic Press) pp 27 – 85
- [99] Xie Y, Schaefer H F, Silaghi-Dumitrescu R, Peng B, Li Q, Stearns J A and Rizzo T R 2012 *Chem. Eur. J.* **18** 12941–12944
- [100] Schubert F, Rossi M, Baldauf C, Pagel K, Warnke S, von Helden G, Filsinger F, Kupser P, Meijer G, Salwiczek M, Koksche B, Scheffler M and Blum V 2015 *Phys. Chem. Chem. Phys.* **17**(11) 7373–7385
- [101] Schubert F 2014 *Conformational equilibria and spectroscopy of gas-phase homologous peptides from first principles* Ph.D. thesis Free University Berlin and Fritz Haber Institute URL [http://www.diss.fu-berlin.de/diss/receive/FUDISS\\_thesis\\_000000097733?lang=en](http://www.diss.fu-berlin.de/diss/receive/FUDISS_thesis_000000097733?lang=en)
- [102] Shi Y, Xia Z, Zhang J, Best R, Wu C, Ponder J W and Ren P 2013 *J. Chem. Theory Comput.* **9** 4046–4063
- [103] Baldauf C, Günther R and Hofmann H J 2003 *Helv. Chim. Acta* **86** 2573–2588
- [104] Baldauf C, Günther R and Hofmann H J 2006 *J. Org. Chem.* **71** 1200–1208
- [105] Baldauf C and Hofmann H J 2012 *Helv. Chim. Acta* **95** 2348–2383
- [106] Wu Y D and Wang D P 1998 *J. Am. Chem. Soc.* **120** 13485–13493
- [107] Wu Y D and Wang D P 1999 *J. Am. Chem. Soc.* **121** 9352–9362
- [108] Möhle K, Günther R, Thormann M, Sewald N and Hofmann H J 1999 *Biopolymers* **50** 167–184
- [109] Swendsen R H and Wang J S 1986 *Phys. Rev. Lett.* **57** 2607–2609
- [110] Hukushima K and Nemoto K 1996 *J. Phys. Soc. Jpn.* **65** 1604–1608
- [111] Okabe T, Kawata M, Okamoto Y and Mikami M 2001 *Chem. Phys. Lett.* **335** 435–439
- [112] Seibert M M, Patriksson A, Hess B and van der Spoel D 2005 *J. Mol. Biol.* **354** 173–183
- [113] van der Spoel D and Seibert M M 2006 *Phys. Rev. Lett.* **96** 238102
- [114] Lei H and Duan Y 2007 *Curr. Opin. Struct. Biol.* **17** 187–191
- [115] Wales D J and Doye J P K 1997 *J. Phys. Chem. A* **101** 5111–5116
- [116] Pappu R V, Hart R K and Ponder J W 1998 *J. Phys. Chem. B* **102** 9725–9742 URL <http://dasher.wustl.edu/tinker>
- [117] Wales D J and Doye J P K 1997 *J. Phys. Chem. A* **101** 5111–5116 URL <http://www-wales.ch.cam.ac.uk/GMIN>
- [118] Hartke B 1995 *Chem. Phys. Lett.* **240** 560–565
- [119] Clark D E and Westhead D R 1996 *J. Computer-Aided Mol. Des.* **10** 337–358
- [120] Nair N and Goodman J M 1998 *J. Chem. Inf. Comput. Sci.* **38** 317–320
- [121] Damsbo M, Kinnear B S, Hartings M R, Ruhoff P T, Jarrold M F and Ratner M A 2004 *Proc. Natl. Acad. Sci. USA* **101** 7215–7222
- [122] Supady A, Blum V and Baldauf C 2015 *arXiv:1505.02521 [q-bio]* ArXiv: 1505.02521
- [123] Fafoom - flexible algorithm for optimization of molecules URL <https://github.com/adrianasupady/fafoom>
- [124] Hartke B 1998 *Theor. Chem. Acc.* **99** 241–247
- [125] Chutia S, Rossi M and Blum V 2012 *J. Phys. Chem. B* **116** 14788–14804
- [126] Schubert F, Pagel K, Rossi M, Warnke S, Salwiczek M, Koksche B, von Helden G, Blum V, Baldauf C and Scheffler M 2015 *Phys. Chem. Chem. Phys.* **17** 5376–5385

- [127] Ropo M, Baldauf C and Blum V 2015 *arXiv:1504.03708 [physics, q-bio]* URL <http://arxiv.org/abs/1504.03708>
- [128] Valdes H, Pluhackova K, Pitonak M, Rezac J and Hobza P 2008 *Phys. Chem. Chem. Phys.* **10** 2747–2757
- [129] Bhattacharya S, Levchenko S V, Ghiringhelli L M and Scheffler M 2014 *New J. Phys.* **16** 123016
- [130] Price M, Dennis O and William J 2001 *J. Comput. Chem.* **22** 1340–1352
- [131] Ceriotti M, Tribello G A and Parrinello M 2011 *Proc. Natl. Acad. Sci. USA* **108** 13023–13028
- [132] Ceriotti M, Tribello G A and Parrinello M 2013 *J. Chem. Theory Comput.* **9** 1521–1532
- [133] Gasparotto P and Ceriotti M 2014 *J. Chem. Phys.* **141** 174110
- [134] Neese F 2012 *WIREs Comput. Mol. Sci.* **2** 73–78
- [135] Auckenthaler T, Blum V, Bungartz H J, Huckle T, Johanni R and Kr L 2011 *Parallel Computing* **37** 783 – 794
- [136] NOMAD repository DOI: 10.17172/NOMAD/20150526220502 URL <http://nomad-repository.eu>
- [137] Antoine R and Dugourd P 2011 *Phys. Chem. Chem. Phys.* **13** 16494–16509
- [138] Hutter J 2006 Excited-state dynamics in finite systems and biomolecules *Time-Dependent Density Functional Theory (Lecture Notes in Physics vol 706)* ed Marques M, Ullrich C, Nogueira F, Rubio A, Burke K and Gross E U (Springer Berlin Heidelberg) pp 217–226 ISBN 978-3-540-35422-2
- [139] Mercier S R, Boyarkin O V, Kamariotis A, Guglielmi M, Tavernelli I, Cascella M, Rothlisberger U, and Rizzo T R 2006 *J. Am. Chem. Soc.* **128** 16938–16943
- [140] Guglielmi M, Doemer M, Tavernelli I and Röthlisberger U 2013 *Faraday Discuss.* **163** 189–203
- [141] Wyttenbach T, von Helden G, Batka J J, Carlat D and Bowers M T 1997 *J. Am. Soc. Mass. Spectrom.* **8** 275–282
- [142] Shvartsburg A A and Jarrold M F 1996 *Chem. Phys. Lett.* **261** 86–91
- [143] Mesleh M F, Hunter J M, Shvartsburg A A, Schatz G C and Jarrold M F 1996 *J. Phys. Chem.* **100** 16082–16086
- [144] Bleiholder C, Wyttenbach T and Bowers M T 2011 *Int. J. Mass Spectrom.* **308** 1–10
- [145] Wyttenbach T, Bleiholder C and Bowers M T 2013 *Anal. Chem.* **85** 2191–2199
- [146] Marklund E G, Degiacomi M T, Robinson C V, Baldwin A J and Benesch J L P 2015 *Structure* **23** 791–799
- [147] Herrmann C and Reiher M 2007 First-principles approach to vibrational spectroscopy of biomolecules *Atomistic Approaches in Modern Biology (Topics in Current Chemistry vol 268)* ed Reiher M (Springer Berlin Heidelberg) pp 85–132 ISBN 978-3-540-38082-5
- [148] Herrmann C, Neugebauer J and Reiher M 2007 *New J. Chem.* **31**(6) 818–831
- [149] Wilson E, Decius J and Cross P 2003 *Molecular Vibrations: the theory of infrared and Raman vibrational spectra*. new ed (Dover)
- [150] McQuarrie D 2000 *Statistical Mechanics* 1st ed (University Science Books)
- [151] Craig I R and Manolopoulos D E 2004 *J. Chem. Phys.* **121** 3368–3373
- [152] Cao J and Voth G A 1994 *J. Chem. Phys.* **100** 5106–5117
- [153] Borisow J, Moraldi M and Frommhold L 1985 *Mol. Phys.* **56** 913–922
- [154] Ramirez R, Lopez-Ciudad T, Kumar P and Marx D 2004 *J Chem Phys* **121** 3973–3983
- [155] Berne B J and Pecora R 2000 *Dynamic Light Scattering - With Applications to Chemistry, Biology, and Physics* new ed (Dover)
- [156] Roy T K and Gerber R B 2013 *Phys. Chem. Chem. Phys.* **15** 9468–9492
- [157] Barone V, Biczysko M and Bloino J 2014 *Phys. Chem. Chem. Phys.* **16** 1759–1787
- [158] Calvo F, Basire M and Parneix P 2011 *J. Phys. Chem. A* **115** 8845–8854
- [159] Basire M, Parneix P, Calvo F, Pino T and Bréchnignac P 2009 *J. Phys. Chem. A* **113** 6947–6954
- [160] Roy T K, Kopysov V, Nagornova N S, Rizzo T R, Boyarkin O V and Gerber R B 2015 *ChemPhysChem* **16** 1374–1378
- [161] Gaigeot M P and Sprik M 2003 *J. Phys. Chem. B* **107** 10344–10358

- [162] Gaigeot M P, Martinez M and Vuilleumier R 2007 *Mol. Phys.* **105** 2857–2878
- [163] Gaigeot M P and Spezia R 2015 Theoretical methods for vibrational spectroscopy and collision induced dissociation in the gas phase *Gas-Phase IR Spectroscopy and Structure of Biological Molecules (Topics in Current Chemistry vol 364)* ed Rijs A M and Oomens J (Springer International Publishing) pp 99–151
- [164] Cimas A, Vaden T D, de Boer T S J A, Snoek L C and Gaigeot M P 2009 *J. Chem. Theory Comput.* **5** 1068–1078
- [165] Gregoire G, Gaigeot M P, Marinica D C, Lemaire J, Schermann J P and Desfrancois C 2007 *Phys. Chem. Chem. Phys.* **9**(24) 3082–3097
- [166] Sediki A, Snoek L C and Gaigeot M P 2011 *Int. J. Mass Spectrom.* **308** 281–288
- [167] Parneix P, Basire M and Calvo F 2013 *J. Phys. Chem. A* **117** 3954–3959
- [168] Ghiringhelli L M, Gruene P, Lyon J T, Rayner D M, Meijer G, Fielicke A and Scheffler M 2013 *New J. Phys.* **15** 083003
- [169] Hammer N I, Diken E G, Roscioli J R, Johnson M A, Myshakin E M, Jordan K D, McCoy A B, Huang X, Bowman J M and Carter S 2005 *J. Chem. Phys.* **122** 244301
- [170] Kohtani M and Jarrold M F 2004 *J. Am. Chem. Soc.* **126** 8454–8458
- [171] Pendry J 1980 *J. Phys. C Solid State* **13** 937–944
- [172] Blum V and Heinz K 2001 *Comput. Phys. Commun.* **134** 392–425
- [173] Rossi M, Ceriotti M and Manolopoulos D E 2014 *J. Chem. Phys.* **140** 234116
- [174] Morrone J A and Car R 2008 *Phys. Rev. Lett.* **101**(1) 017801
- [175] Marx D, Tuckerman M E, Hutter J and Parrinello M 1999 *Nature* **397** 601–604
- [176] Li X Z, Walker B and Michaelides A 2011 *Proc. Natl. Acad. Sci. USA* **108** 6369–6373
- [177] Pérez A, Tuckerman M E, Hjalmarsen H P and von Lilienfeld O A 2010 *J. Am. Chem. Soc.* **132** 11510–11515
- [178] Feynman R P and Hibbs A R 1964 *Quantum Mechanics and Path Integrals* (New York: McGraw-Hill)
- [179] Tuckerman M 2010 *Statistical Mechanics: Theory and Molecular Simulation* (Oxford, UK: OUP Oxford)
- [180] Cao J and Voth G A 1994 *J. Chem. Phys.* **101** 6168–6183
- [181] Rossi M, Liu H, Paesani F, Bowman J and Ceriotti M 2014 *J. Chem. Phys.* **141** 181101
- [182] Ceriotti M, More J and Manolopoulos D E 2014 *Comput. Phys. Commun.* **185** 1019–1026
- [183] Kohtani M, Kinnear B and Jarrold M F 2000 *J. Am. Chem. Soc.* **122** 12377–12378
- [184] Martens J K, Compagnon I, Nicol E, McMahon T B, Clavaguéra C and Ohanessian G 2012 *J. Phys. Chem. Lett.* **3** 3320–3324
- [185] Fox S, Wallnoefer H G, Fox T, Tautermann C S and Skylaris C K 2011 *Journal of Chemical Theory and Computation* **7** 1102–1108
- [186] Fox S J, Pittock C, Tautermann C S, Fox T, Christ C, Malcolm N O J, Essex J W and Skylaris C K 2013 *J. Phys. Chem. B* **117** 9478–9485
- [187] Freddolino P L, Harrison C B, Liu Y and Schulten K 2010 *Nat. Phys.* **6** 751–758
- [188] Shaw D, Dror R, Salmon J, Grossman J, Mackenzie K, Bank J, Young C, Deneroff M, Batson B, Bowers K, Chow E, Eastwood M, Ierardi D, Klepeis J, Kuskin J, Larson R, Lindorff-Larsen K, Maragakis P, Moraes M, Piana S, Shan Y and Towles B 2009 Millisecond-scale molecular dynamics simulations on Anton *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis* pp 1–11
- [189] Piana S, Lindorff-Larsen K and Shaw D E 2012 *Proc. Natl. Acad. Sci. USA* **109** 17845–17850
- [190] Piana S, Lindorff-Larsen K and Shaw D E 2013 *Proc. Natl. Acad. Sci. USA* **110** 5915–5920
- [191] Noé F, Horenko I, Schütte C and Smith J C 2007 *J. Chem. Phys.* **126** 155102
- [192] Shirts M R, Pitera J W, Swope W C and Pande V S 2003 *J. Chem. Phys.* **119** 5740–5761
- [193] Lane T J, Shukla D, Beauchamp K A and Pande V S 2013 *Curr. Opin. Struct. Biol.* **23** 58–65
- [194] Schwantes C R, McGibbon R T and Pande V S 2014 *J. Chem. Phys.* **141** 090901
- [195] Bowman G R, Pande V S and Noé F 2014 Introduction and Overview of This Book *An*

- Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation (Advances in Experimental Medicine and Biology no 797)* ed Bowman G R, Pande V S and Noé F (Springer Netherlands) pp 1–6 ISBN 978-94-007-7605-0 978-94-007-7606-7
- [196] Larson S M, Snow C D, Shirts M, P V S and Pande V S Folding@home and genome@home: Using distributed computing to tackle previously intractable problems in computational biology
- [197] Kohlhoff K J, Shukla D, Lawrenz M, Bowman G R, Konerding D E, Belov D, Altman R B and Pande V S 2014 *Nat Chem* **6** 15–21
- [198] Torrie G M and Valleau J P 1977 *J. Comp. Phys.* **23** 187–199
- [199] Kong X and Brooks III C L 1996 *J. Chem. Phys.* **105** 2414–2423
- [200] Bartels C and Karplus M 1997 *J. Comput. Chem.* **18** 1450–1462
- [201] Young W S and Brooks III C L 1996 *J. Mol. Biol.* **259** 560–572
- [202] Bonomi M, Branduardi D, Bussi G, Camilloni C, Provasi D, Raiteri P, Donadio D, Marinelli F, Pietrucci F, Broglia R A and Parrinello M 2009 *Comput. Phys. Commun.* **180** 1961 – 1972
- [203] Tribello G A, Bonomi M, Branduardi D, Camilloni C and Bussi G 2014 *Comput. Phys. Commun.* **185** 604 – 613
- [204] Kebarle P 2000 *Int. J. Mass Spectrom.* **200** 313–330
- [205] Warnke S, von Helden G and Pagel K 2013 *J. Am. Chem. Soc.* **135** 1177–1180
- [206] Kohtani M, Jarrold M F and Wee S 2004 *J. Phys. Chem. B* **108** 6093–6097
- [207] Kohtani M and Jarrold M F 2002 *J. Am. Chem. Soc.* **124** 11148–11158
- [208] Kohtani M and Jarrold M F 2004 *J. Am. Chem. Soc.* **126** 8454–8458
- [209] Liu D, Wyttenbach T and Bowers M T 2004 *Enzyme Microb. Technol.* **236** 81–90
- [210] Headrick J M 2005 *Science* **308** 1765–1769
- [211] Mizuse K and Fujii A 2011 *Phys. Chem. Chem. Phys.* **13** 7129
- [212] Asmis K R 2003 *Science* **299** 1375–1377
- [213] Heine N, Fagiani M R, Rossi M, Wende T, Berden G, Blum V and Asmis K R 2013 *J. Am. Chem. Soc.* **135** 8266–8273
- [214] Hammer N I, Diken E G, Roscioli J R, Johnson M A, Myshakin E M, Jordan K D, McCoy A B, Huang X, Bowman J M and Carter S 2005 *J. Chem. Phys.* **122** 244301
- [215] Guasco T L, Elliott B M, Johnson M A, Ding J and Jordan K D 2010 *J. Phys. Chem. Lett.* **1** 2396–2401
- [216] Nagornova N S, Rizzo T R and Boyarkin O V 2012 *Science* **336** 320–323
- [217] Cocinero E, Stanca-Kaposta E, Dethlefsen M, Liu B, Gamblin D, Davis B and Simons J 2009 *Chem. Eur. J.* **15** 13427–34
- [218] Mayorkas N, Rudić S, Cocinero E J, Davis B G and Simons J P 2011 *Phys. Chem. Chem. Phys.* **13** 18671
- [219] DiStasio R A, Santra B, Li Z, Wu X and Car R 2014 *J. Chem. Phys.* **141** 084502
- [220] Grossman J C, Schwegler E, Draeger E W, Gygi F and Galli G 2004 *J. Chem. Phys.* **120** 300
- [221] Schwegler E, Grossman J C, Gygi F and Galli G 2004 *J. Chem. Phys.* **121** 5400
- [222] Fernández-Serra M V and Artacho E 2004 *J. Chem. Phys.* **121** 11136
- [223] VandeVondele J, Mohamed F, Krack M, Hutter J, Sprik M and Parrinello M 2005 *J. Chem. Phys.* **122** 014515
- [224] Lee H S and Tuckerman M E 2007 *J. Chem. Phys.* **126** 164501
- [225] Ben M D, Schönherr M, Hutter J and VandeVondele J 2013 *J. Phys. Chem. Lett.* **4** 3753–3759
- [226] Santra B, Michaelides A and Scheffler M 2007 *J. Chem. Phys.* **127** 184104
- [227] Santra B, Michaelides A, Fuchs M, Tkatchenko A, Filippi C and Scheffler M 2008 *J. Chem. Phys.* **129** 194111
- [228] Marx D, Tuckerman M E, Hutter J and Parrinello M 1999 *Nature* **397** 601–604
- [229] Habershon S, Markland T E and Manolopoulos D E 2009 *J. Chem. Phys.* **131** 24501
- [230] Habershon S, Fanourgakis G S and Manolopoulos D E 2008 *J. Chem. Phys.* **129** 74501
- [231] Schmitt U W and Voth G A 1999 *J. Chem. Phys.* **111** 9361
- [232] Paesani F and Voth G A 2010 *J. Chem. Phys.* **132** 014105

- [233] Vendrell O, Gatti F and Meyer H D 2007 *J. Chem. Phys.* **127** 184303
- [234] Liu H, Wang Y and Bowman J M 2015 *The Journal of Chemical Physics* **142** 194502
- [235] Voora V K, Ding J, Sommerfeld T and Jordan K D 2013 *The Journal of Physical Chemistry B* **117** 4365–4370
- [236] Medders G R, Babin V and Paesani F 2014 *Journal of Chemical Theory and Computation* **10** 2906–2910
- [237] Hartke B and Grimme S 2015 *Phys. Chem. Chem. Phys.* **17** 16715–16718
- [238] Fox S J, Dziedzic J, Fox T, Tautermann C S and Skylaris C K 2014 *Proteins* **82** 3335–3346
- [239] Pinski P, Riplinger C, Valeev E F and Neese F 2015 *J. Chem. Phys.* **143** 034108
- [240] Schiffmann F and VandeVondele J 2015 *J. Chem. Phys.* **142** 244117
- [241] Hasnip P J, Refson K, Probert M I J, Yates J R, Clark S J and Pickard C J 2014 *Phil. Trans. R. Soc. A* **372** 20130270–20130270