Bachelor Thesis

Simulation of a Peptide in a Drift Tube Environment

by

Carlo von Carnap

at the Department of Theoretical and Computational Biophysics, Max Planck Institute for Multidisciplinary Sciences

Supervisor:	Dr. Daniel Szöllösi
First Reviewer:	Prof. Dr. Helmut Grubmüller
Second Reviewer:	Prof. Dr. Stefan Klumpp
Submission Date:	September 6, 2024

Contents

Introduction

Proteins are essential macromolecules in all living organisms. They are build as a linear chain of individual building blocks called amino acids, whose sequence and attributes such as charge, size, or hydrophilicity determine the folding of the polypeptide chain and thus the structure and function of the protein.

Proteins are involved in essentially all life-related processes, where they facilitate chemical reactions, contribute to the structure of cells, and aid in the transport of other molecules between the cellular compartments. Due to this crucial importance in the organisms functioning, several diseases like Type 2 Diabetes or Alzheimer's disease are a direct cause of protein defects. [1] Conversely, other diseases like certain cancers are characterised by changes in a cells protein composition [2], which is why techniques for separation and identification of polypeptide chains and detecting those protein overabundances are of major interest as a diagnostic tool.

One of those tools is **I**on **M**obility **S**pectrometry (IMS). In these measurements, charged molecules are accelerated by an electric field through a collision gas in the IMS device. Depending on their size and shape, these analyte ions collide more or less frequently with the molecules of the collision gas. The amount of collisions determines an ions mobility — a measure for the speed at which it drifts through the collision gas — which is in turn used for the separation of the analyte ions. [3] Here, digested proteins from a purified cell lysate are ionised by **E**lectrospray **I**onisation (ESI) and then separated by **T**rapped **I**on **M**obility **S**pectrometry (TIMS) [4] — a recent modification of IMS. Subsequently, the peptide ions are identified by **M**ass **S**pectrometry (MS) according to their molecular-weight-to-charge-ratio.

Analysing the eluted polypeptides had now revealed different ion mobilities for the same peptide, despite otherwise identical conditions. As a possible explanation, multiple peptide folds per amino acid chain had been suggested; however, this suggestion had so

far not been tested experimentally. Therefore, **M**olecular **D**ynamics (MD) simulations in vacuum conditions were performed using a peptide with unambiguous placement of charges. Simulating this peptide has then indeed shown two different types of conformations: an extended, helical fold with lower ion mobility, as well as a more spherical fold with higher ion mobility.

However, the experiment takes place in low pressure humid air of 2.7 mbar [4] and 30 % to 60 % relative humidity instead of vacuum. Therefore, we addressed the question whether the presence of dry and humid air influences the folding of the peptide using in silico methods.

Theory

2.1 Electrospray Ionization

Separation by TIMS requires a preceding ionisation of the peptides. This is done by electrospray ionisation (ESI), a technique that works under the mild laboratory conditions for pressure and temperature and is sketched in Figure 2.1. ESI proceeds by the formation of charged droplets, their desolvation, and the eventual formation of gas phase ions. Those ions can be of multiple charge states that vary in their amount of charge, as well as their charge placement. [5][6]

The ESI device consists of a capillary filled with sample solution. Assuming positively charged droplets as desired end product, the solution is deprived of negative charges by a positive capillary edge discharging the anions. Positively charged ions accumulate at the tip of the capillary, resulting in a so-called Taylor cone. Once the repulsion of those cations becomes stronger than the surface tension, positively charged droplets are emitted from the capillary and pulled towards a negatively charged counter-electrode. [5]
Due to evaporation of the solvent, the emitted droplets now lose mass until again the electrostatic repulsion exceeds the surface tension. Upon reaching this Rayleigh stability limit, the droplet dissolves into smaller droplets with increased charge density. [5]

Formation of Gas Phase Ions

The formation of gas phase peptide ions from the droplets is essential to the eventual fold of the peptide. For this process, two different pathways are proposed.
The ion emission model assumes that charged peptides are released from the droplet before major evaporation shrinkage. This emission is a result of the electrostatic repulsion at the droplets surface and eventually leads to extended peptide structures. [7]

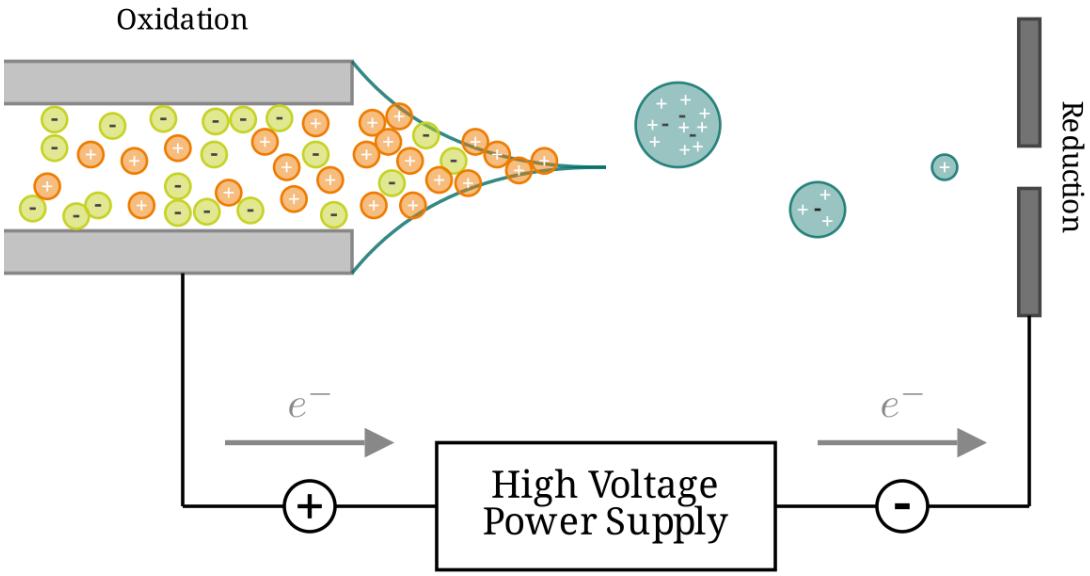


Figure 2.1: Schematic figure of the electrospray ionisation mechanism. The capillary on the left is filled with sample solution which is deprived of anions by the positive capillary charge. The tip at the capillary end is called Taylor cone and releases positively charged droplets once the electrostatic repulsion of the ions exceeds the surface tension. The droplets are then pulled to the negatively charged counter-electrode on the right while the solvent evaporates increasing its charge density.

The other proposed pathway is described by Raab et al. [8] In solution, the protein can relatively quickly overcome free energy barriers due to compensatory interactions with the solvent. Following the evaporation of the solvent, those barriers of the emerging gas-phase free energy landscape are becoming harder to pass and effectively fix the structure out of its equilibrium in its current (local) free energy minimum. [8]

2.2 Ion Mobility Spectrometry

Ion mobility spectrometry (IMS) is an experimental method used for the separation of ionised molecules based on their size and shape. It accelerates charged molecules in an electric field of strength E, where the air forces the ion to reach a terminal drift velocity v_d. The ratio between terminal drift velocity v_d and the strength of the electric field E defines the name-giving ion mobility, [3]

$$K = \frac{v_\mathrm{d}}{E}. \tag{2.2.1}$$

This quantity can now be normalised with respect to the environmental parameters in the drift tube, namely pressure p, and temperature T. With standardised p_0 and T_0, one receives the reduced ion mobility K_0, [3]

$$K_0 = \frac{p}{p_0}\frac{T_0}{T} \cdot K. \tag{2.2.2}$$

Under consideration of Equation 2.2.1, this means that a higher air pressure leads to lower terminal drift velocities, whereas higher temperatures lead to higher drift velocities.

The ion mobility itself is a direct result of the collisions of the respective ions with the tube gas, allowing to relate both quantities mathematically. Result is the **C**ollisional **C**ross **S**ection Ω (CCS), a quantity comparable across devices described by the Mason-Schamp equation [3]

$$\Omega = \frac{3}{16}\sqrt{\frac{2\pi}{\mu k_\mathrm{B} T}} \cdot \frac{ze}{N_0 K_0}, \tag{2.2.3}$$

valid for low field strengths. It uses the reduced ion mobility K_0 from Equation 2.2.2, as well as the ion properties μ (reduced mass of ion and collision gas molecules), N_0 (normalised gas density), and the total ion charge ze with elementary charge e. Further, T is the temperature and k_B the Boltzmann constant. The CCS is inversely proportional to the ion mobility K and is typically measured in units of square Ångstrom. [3]

2.2.1 Trapped Ion Mobility Spectrometry

Trapped ion mobility spectrometry (TIMS) is an IMS method that offers an increased ion selectivity by sequential elution of the separated ions along their CCS-to-charge-ratio. In contrast to conventional IMS, TIMS uses an air flow to carry the ions into the drift tube and holds them stationary by an increasing electric field of opposite direction. [4]

As shown in Figure 2.2, the tube channel itself consists of an entrance funnel receiving ions from the ESI, the TIMS tunnel comprising the two ion traps, and the exit funnel directing the ions to the subsequent mass-spectrometer. Ions are continuosly exiting the ESI capillary and forced into the TIMS tube by a charged deflection plate orthogonal to the tube channel, uncharged molecules do not enter the device. An oscillating axial iontrap confines the ions to the channels centre throughout the tube. [4]

The TIMS process can be described in three steps. First, the strong field of Trap 1

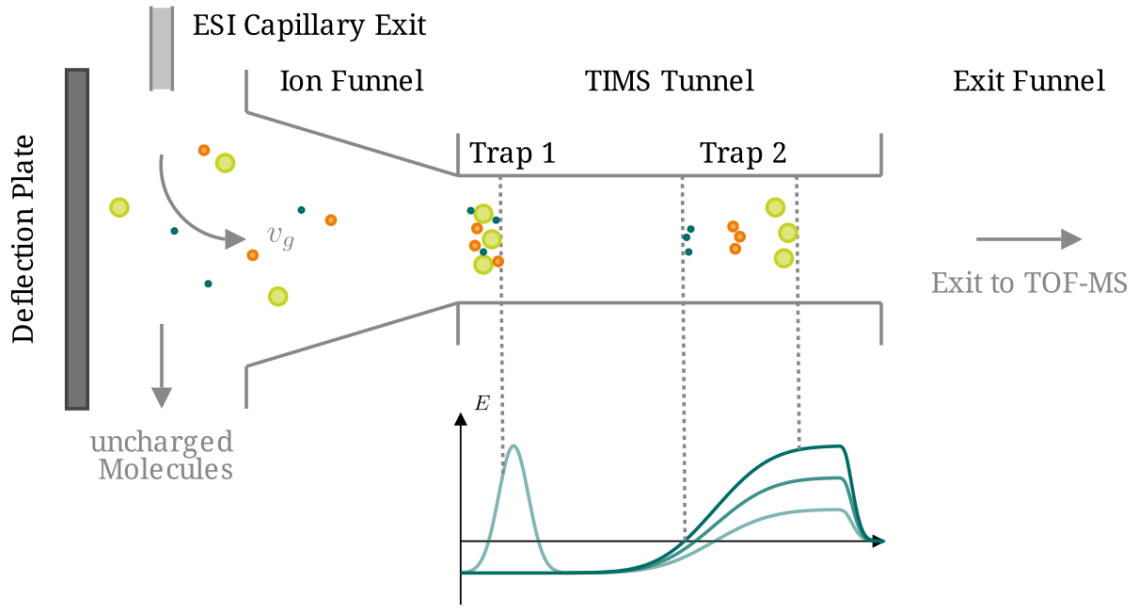


Figure 2.2: Schematic figure of a TIMS drift tube with the two electric field traps displayed in a separate coordinate system. Ions enter the tube upon electrospray ionisation after deflection by a charged plate, whereas uncharged molecules leave the device. The drift tube comprises the ion funnel for entering, the TIMS tunnel for separation of the ions and the exit funnel for elution and subsequent analysis by mass-spectrometry. It operates by **(1)** accumulating sufficiently many ions in Trap 1, before temporarily turning off the respective Trap 1 field to **(2)** separate the ions along their CCS-to-charge-ratio in the increasing field of Trap 2. For the elution phase **(3)**, the field of Trap 2 is gradually decreased to elute the ions sequentially, ensuring a high selectivity.

accumulates the ions until a sufficient amount is reached. Then, the field of Trap 1 is deactivated for a certain period of time and the ions are separated in the increasing electric field of Trap 2, where they find their respective equilibrium positions. For the elution, the field of Trap 2 is gradually decreased. During that period, the electric field of Trap 1 is turned back on to avoid the participation of newly inflowing ions in the elution process. After resetting the Trap 2 field the cycle can be restarted. [4]

The ion separation occurs in the gradually increasing field of Trap 2. Ions with higher CCS will reach equilibrium positions at higher field strengths in the TIMS tunnel, whereas ions with lower CCS find equilibrium positions at lower field strengths. Since this equilibrium position requires

$$v_\mathrm{d} + v_\mathrm{g} = 0 \tag{2.2.4}$$

according to Equation 2.2.1, their ion mobility K can be determined using the respective equilibrium field strength. An ions CCS can then be calculated analogous to conventional IMS using the Mason-Schamp Equation 2.2.3.

Computational Methods

3.1 Molecular Dynamics Simulations

3.1.1 Fundamentals of Molecular Dynamics Simulations

Molecular dynamics (MD) simulations solve the Newtonian equations of motion for the atoms of a (biological) system numerically for a certain period of time.

Technically, the equation supposed to be solved for all nuclei and electrons is the time-dependent Schrödinger-Equation

$$i\hbar\partial_t\psi\left(\mathbf{R},\mathbf{r}\right)=\hat{H}\psi\left(\mathbf{R},\mathbf{r}\right), \tag{3.1.1}$$

where $\psi\left(\mathbf{R},\mathbf{r}\right)$ is the respective wave function, $\mathbf{R}$ are the nucleus positions, and $\mathbf{r}$ are the electron positions for all respective atoms in the system. The Hamiltonian operator $\hat{H}$ represents the total energy of the system. However, solving this equation analytically is not possible for any larger molecular system. [9] As a result, from this purely quantum mechanical description two main approximations have to be made.

Born-Oppenheimer approximation

The Born-Oppenheimer approximation assumes the much heavier nuclei to move considerably slower than its surrounding electrons. With the electronic motion treated as instantaneous compared to the nucleus motion, it can be calculated separately from the latter for each configuration of fixed nuclei. The total wave function can now be written as an independent product of nuclear wave function $\psi_{\mathrm{n}}\left(\mathbf{R}\right)$, and electronic wave function $\psi_{\mathrm{e}}\left(\mathbf{r}\right)$, [9]

$$\psi(\mathbf{R}, \mathbf{r}) = \psi_{\mathrm{n}}(\mathbf{R})\, \psi_{\mathrm{e}}(\mathbf{r}). \tag{3.1.2}$$

Following this assumption, the electronic wave function $\psi_{\mathrm{e}}(\mathbf{r})$ can be described by the time-independent Schrödinger-Equation

$$\hat{H}_e(\mathbf{R}, \mathbf{r})\, \psi_{\mathrm{e}}(\mathbf{R}, \mathbf{r}) = E_e(\mathbf{R})\, \psi_{\mathrm{e}}(\mathbf{R}, \mathbf{r}), \tag{3.1.3}$$

where $\mathbf{R}$ are the fixed nuclei positions, $\mathbf{r}$ the dynamic electron positions, and $E_{\mathrm{e}}(\mathbf{R})$ the electronic eigenenergies.

Classical Nuclei

The heavier nuclei will now be assumed as point particles following classical mechanics, its Newtonian equations of motion are solved numerically.

3.1.2 Force Fields

The potential energy between the atoms in the system are described by a force field. In a classical all-atom force field the atoms are connected by unbreakable bonds and carry partial charges interacting electrostatically with each other. The parameters such as charge, or bond angles and potentials are either derived from the probability density of the electrons — obtained from the quantum mechanical solutions of Equation 3.1.3 — or by fitting to empirical data. [10]

Charmm36m Force Field

Here, for all simulations the `Charmm36m` force field was used. Originally, two state-of-the-art force fields were considered, namely `Amber99sb-ildn` and `Charmm36m`. However, simulations of a particular peptide only reproduced the expected multimodal size distribution for the latter.

3.1.3 Simulation Setup in GROMACS

For all MD simulations the simulation package GROMACS 2023.6 was used. Each simulation ran over 550 ns with a step size of 1 fs using the leap-frog integration scheme in double precision. The high precision and the unusually low step size were necessary to ensure the stability of the simulation. For the initial 10 ns the peptide was fixed at a temperature of 600 K to overcome the energy barriers separating the different conformations.

The subsequent quenching process occurred over 500 ns down to 305 K, where the temperature was fixed for another 40 ns. During the runs in air, non-protein molecules were held at 305 K throughout the simulation. For the temperature coupling, velocity rescaling [11] was used with a coupling constant of $\tau = 0.1$ for both coupling groups. Pressure coupling was turned off because the near vacuum environment requires a constant volume. The translational centre of mass velocity was removed every 100 simulation steps for peptide and air molecules separately. The short-range calculations for the neighborlist, electrostatics, and van-der-Waals forces were cut off within a radius of 30 nm, no long-range interactions are calculated due to the overall charge of the system. Because of the high relative velocities of the molecules, the neighborlist was updated every step to prevent air molecules from passing the peptide without detecting a collision. A cubic simulation box with sidelength 100 nm was used. The box size allows for a sufficient sample size of air molecules while keeping the simulation time at appropriate levels of 100 ns/day.

3.1.4 Starting Structure

For the simulations a peptide with unambiguous placement of charges has been selected, its sequence reads

DFGYGVEEEEEEAAAAGGGVGAGAGGGCGPGGADSS**K**P**R**.

The peptides charge has been identified as +3 via mass-spectrometry, meaning that all protonation sites are protonated and the charges are carried by the N-Terminus, as well as the two positively charged amino acids — Lysine **K** and Arginine **R** — respectively.

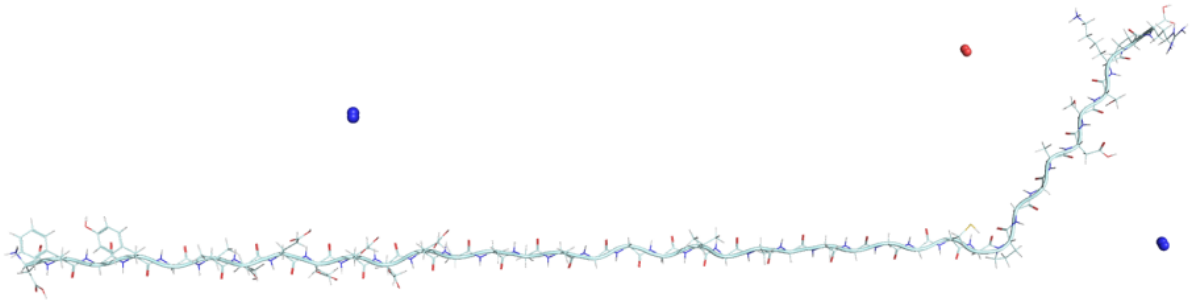


Figure 3.1: Extended starting structure of the peptide with surrounding gas molecules. The blue molecules are nitrogen molecules, the red one is an oxygen molecule.

From this sequence — because the actual structure is unknown — the extended structure in Figure 3.1 was created and used for all simulation runs in order to minimise the bias due to the starting structure.

Air Composition

To the cubic simulation box of 100 nm side length, 51 Nitrogen, and 13 Oxygen molecules were added in order to mimic the atmospheric composition under a pressure of 2.7 mbar. [4] Carbon dioxide as well as noble gases were neglected.

For the simulation runs in humid air, further one or two water molecules were added to the box to recreate the relative humidities of 30 % and 60 % at room temperature, respectively. [12]

3.2 Analysis of the Trajectories

3.2.1 Calculation of the Radii of Gyration

Whereas in ion mobility spectrometry the collisional cross section (CCS) is the determining factor for the ion separation, due to its easier calculation, the radius of gyration was used as a proxy for the CCS. This is convenient due to the qualitative nature of the measured ion mobilities and the quadratic relationship (which is strictly monotonic, Appendix Figure 6.2) between CCS and radius of gyration. [13][14]

The radius of gyration was calculated using the `gmx gyrate` function of GROMACS. With the atom positions $\mathbf{r}_i$, their respective masses m_i, and the peptide centre of mass $\mathbf{r}_{\text{com}}$, the radius of gyration R_{gyr} for the peptide is calculated as [15]

$$R_{\text{gyr}} = \sqrt{\frac{\sum_i m_i \|\mathbf{r}_i - \mathbf{r}_{\text{com}}\|^2}{\sum_i m_i}}. \tag{3.2.1}$$

The histograms for the radius of gyration contain 100 histogram bins from 0.7 nm to 1.8 nm each.

3.2.2 Calculation of the Solvent Accessible Surface Area

Another proxy for the CCS is the **S**olvent**A**ccessible **S**urface **A**rea (SASA). For convex peptide shapes the SASA has a linear relationship with the average projected area, assuming a uniform orientation distribution of the peptide. The average projected area itself (the mean shadow size of the peptide) is proportional to the CCS, because SASA and calculated CCS show a linear relationship (Appendix Figure 6.2). The calculation of SASA assumes the surface of the molecule to be consisting of spheres while a probing sphere is rolled over the surface to access the available area. For the calculation of the

solvent accessible surface area the GROMACS function `gmx sasa` was used with standard probe size of 0.14 nm and 24 dots per sphere. GROMACS itself uses the **D**ouble **C**ubic **L**attice **M**ethod (DCLM). [16][17]

The histograms were created with 46 bins from 75 nm^2 to 120 nm^2.

3.2.3 Calculation of the Collisional Cross Section

The Collisional Cross Sections for the final vacuum structures were calculated using IMoS (**I**on **Mo**bility **S**pectrometry Suite). [18]

3.2.4 Bootstrapping for Error-Estimation in the Histograms

The uncertainties for each bin of the radius of gyration and solvent accessible surface area distributions were calculated via bootstrapping. From the 1000 final values, 1000 ones were drawn randomly with replacement to generate a bootstrapping sample. Overall, 1000 bootstrapping samples were generated and the standard deviation between the histogram bin sizes n_i^{bin} for each interval *bin* across the samples i was calculated via [19]

$$\sigma_n^{\text{bin}} = \sqrt{\frac{1}{N}\sum_i \left(n_i^{\text{bin}} - \langle n_i^{\text{bin}}\rangle_i\right)^2}, \tag{3.2.2}$$

where $\langle n_i^{\text{bin}}\rangle_i$ is the average amount of values in one interval between the bootstrapping samples.

3.2.5 Calculation of the Transition Rate

A transition between spherical and helical fold is defined as the passing of the radius of gyration threshold $R_{\text{gyr}} = 1\,\text{nm}$ from one frame to another, separated by 1 ns. The radius of gyration of 1 nm was chosen because it is a very low probability state for this peptide, marking an energy barrier between the spherical ($R_{\text{gyr}} \approx 0.8\,\text{nm}$), and helical fold ($R_{\text{gyr}} \approx 1.4\,\text{nm}$). The transition rate is the total amount of transitions per time step in all trajectories divided by the number of trajectories, which is 1000 in this case.

3.2.6 Calculation of the Collisions with Air

A collision with air is a close encounter between peptide and at least one gas molecule, calculated by the GROMACS function `gmx mindist`. The distance cut-off was chosen to be 1 nm, close enough for encounters to affect the trajectories of both, peptide and gas

molecules. The air encounters were calculated for an exemplary run in dry air, identical to the main runs, though with more frequent output of positions and velocities (both every 1 ps).

3.2.7 Calculation of the Fraction of Peptides with Water in their close Proximity

For the fraction of peptides with at least water molecules in their close proximity, the minimal distance between any water molecule and any peptide atom was calculated using the corresponding GROMACS command `gmx mindist`. The proximity threshold was set to 2 nm.

3.2.8 Hierarchical Clustering

All final structures of the four different setups — vacuum, dry air, humid air at 30 % and 60 % relative humidity — were structurally compared using the GROMACS C-alpha RMSD (**R**oot **M**ean **S**quare **D**eviation) function `gmx rms`, resulting into a 4000×4000 matrix. This RMSD matrix was then used as a distance matrix for the hierarchical clustering. The clustering itself was performed according to Ward's method of minimum variance, which uses the euclidean norm d_{ij} between two RMSD values $RMSD_i$ and $RMSD_j$ for clustering,

$$d_{ij} = \|RMSD_i - RMSD_j\|^2. \tag{3.2.3}$$

Results and Discussion

4.1 Quenching in dry Air

To test the hypothesis of multiple folds per sequence, 1000 MD simulations of the peptide had been performed in vacuum. In these quenching simulations, the system was rapidly cooled from 600 K to 305 K over a time span of 500 ns as described in Subsection 3.1.3. It had been shown that the peptide indeed folds into two different structures, namely a more spherical and a helical one.

Our goal was now to examine the influence of more realistic simulation environments on the peptide folding. Therefore, we recreated the quenching simulations with an appropriate number of air molecules in the simulation box as described in Subsection 3.1.4. The final structures of the 1000 simulations in dry air were then compared to the 1000 vacuum runs with respect to their radius of gyration R_{gyr} and solvent accessible surface area (SASA). Both of these quantities serve as valid proxies for the collisional cross section (CCS) (Appendix Figure 6.2) and are therefore used due to their easier calculation.

4.1.1 Distribution of the Radii of Gyration

The final radii of gyration for the 1000 simulations in each condition — vacuum and dry air — were calculated according to Subsection 3.2.1. In order to see a potential influence of the dry air conditions on the folding, those the radii of gyration for both simulation environments are compared in the histogram in Figure 4.1. The errors were approximated via bootstrapping as described in Subsection 3.2.4.

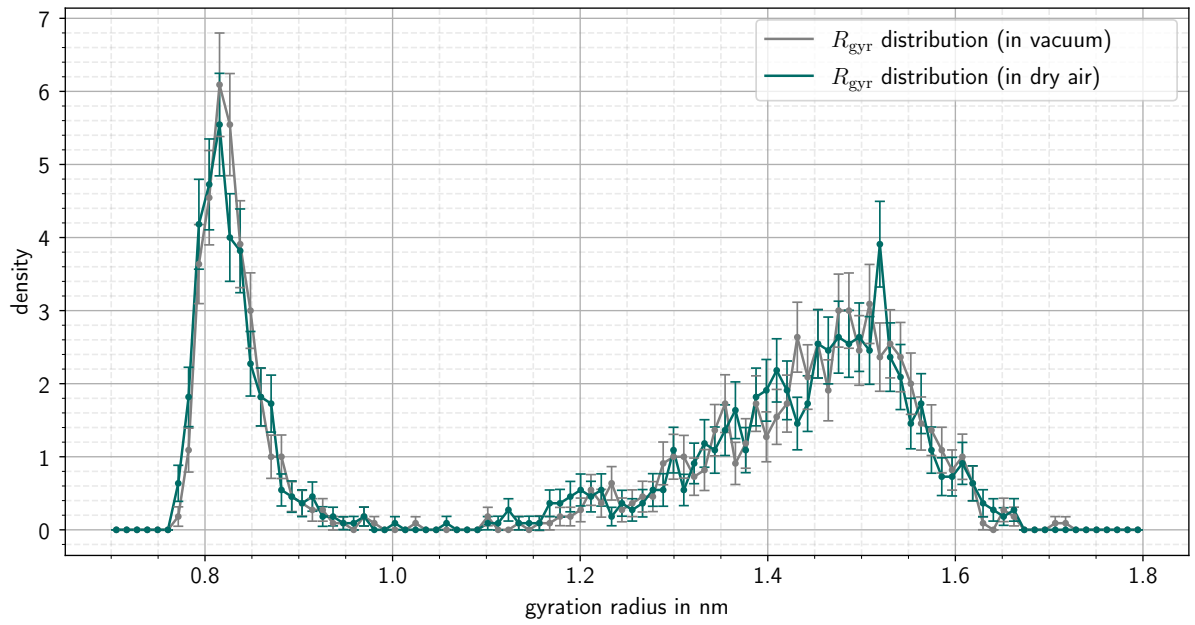
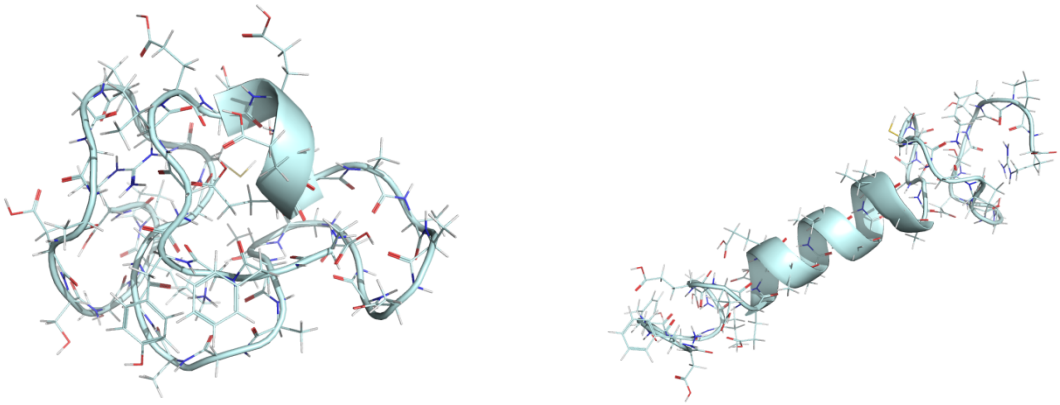


Figure 4.1: Radius of gyration distribution of 1000 quenching runs of the examined peptide in vacuum and dry air. The first maximum at around 0.8 nm represents the final structures folding into the spherical structure, the second maximum at around 1.5 nm the ones folding helically.



(a) spherical fold **(b)** helical fold

Figure 4.2: Visualisations of the two peptide folds using exemplary structures from the last frame of quenching simulations in dry air. The structure in **(a)** has a radius of gyration of 0.80 nm and represents the spherical fold, **(b)** has a gyration radius of 1.42 nm and represents the helical fold.

Overall, there are two major R_{gyr} clusters for both simulation environments: one around 0.8 nm, and one around 1.5 nm. Those radii of gyration correspond to a spherical fold with smaller radius of gyration, and a helical fold with bigger radius of gyration. Exemplary structures of both folds can be seen in Figure 4.2. The smaller radius of gyration

cluster has a sharper distribution around its most frequent value, whereas the bigger radius of gyration is distributed more widely. This can have the following reasons — the terminal ends of the peptide chain in the helical fold are more flexible and thus result into stronger variations of the radius of gyration variation — as well as a mathematical artefact resulting from the way the radius of gyration is calculated — Equation 3.2.1 weighs further deviations from the centre of mass of the molecule stronger.

4.1.2 Distribution of the Solvent Accessible Surface Area

The relation between the CCS and the radius of gyration is not linear, whereas it is with the SASA. Therefore, as a parallel analysis, the solvent accessible surface area of the peptide was calculated as described in Subsection 3.2.2. The corresponding histogram is shown in Figure 4.3, displaying the computed values for vacuum and dry air environments, as well as the errors estimated by bootstrapping.

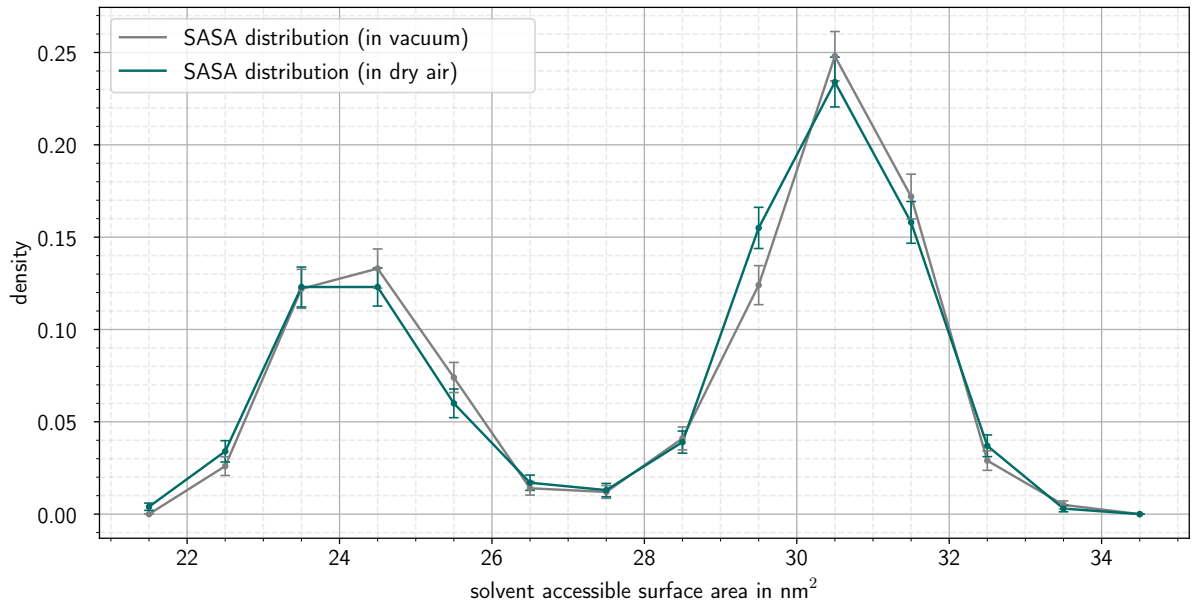


Figure 4.3: Solvent accessible surface area calculated for all 1000 final structures and displayed as a histogram. The structures of less surface accessible to the solvent around 24 nm^2 are the spherical folds, the structures around 31 nm^2 the helical ones.

Again, the histogram shows two types of folds, one has an area accessible to the solvent of around 24 nm^2, the other one of around 31 nm^2. The spherical fold is the one with less surface area exposed to the solvent, the helical fold the one with the larger surface area. The helical fold does also appear to be more abundant among the final structures, a result that might be confirmed by integration over either the R_{gyr} or the SASA maxima. Similar to the radius of gyration histogram, there is no noticeable difference between

the simulation conditions. Therefore, these findings are in line with those from the R_{gyr} analysis.

However, because collisions between collision gas and peptide do occur (Appendix Figure 6.1), this absence of structural differences might either be due to the lack of influence of the collisions on the folding directly, or because the effect is too small to be noticed in only 1000 simulations. Even though in Trap 2 the molecules of the device gas hit the peptide preferably from one side, this effect can be neglected. First, due to the significantly higher thermal speed of the gas (around 400 m/s) compared to the drift velocities (below 80 m/s), and secondly due to the rotation of the peptide. The option to improve the simulation by running it in a constant electric field — accelerating the peptide in one direction and keeping the collision gas stationary to reproduce the asymmetric collisions — would therefore probably not yield different results.

4.1.3 Stability of the Folds

In order to see how stable the two folds of the peptide are under influence of the quenching temperature, a transition rate was calculated according to Subsection 3.2.5. It is shown in Figure 4.4 and counts the number of transitions between spherical and helical fold between the recorded time frames (separated by 1 ns) and per peptide.
After an initial increase of the transition rate over the first 50 ns, the rate decreases during the quenching process.

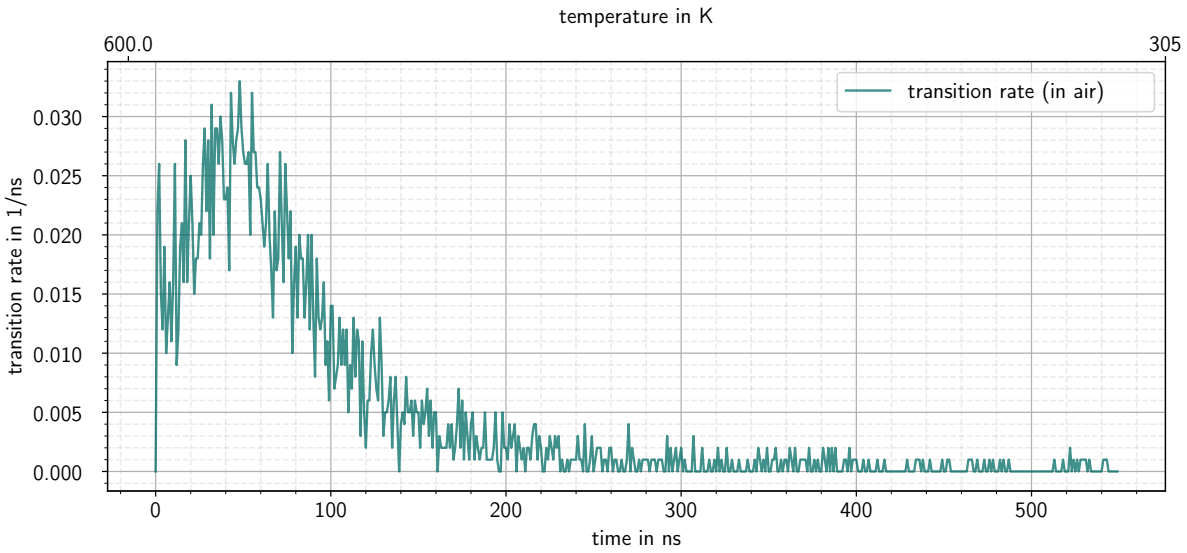


Figure 4.4: Transition rate between spherical and helical fold of the peptide. The rate is defined as the number of transitions per peptide between the simulation frames (recorded each 1 ns) for the simulation in dry air. After an initial increase of the transition rate in the first 50 ns, the transition rate decreases with the decrease in temperature.

The initial increase in the transition rate can be attributed to the formation of the folds — transitions can only occur after a spherical fold been reached from the initial extended structure. According to Figure 4.4 this process takes about 50 ns. The following decrease is expected due to the temperature coupling — the lower the system energy, the less frequent the free energy barrier between the two folds is going to be passed. The non-zero transition rate at lower temperatures (after around 400 K) is a result of a single structure which is fixed at a radius of gyration around the threshold. Consequently, with the transition rate vanishing for peptides clearly assigned to one of the main two folds, those folds can be assumed as stable at these lower temperatures.

4.2 Quenching in humid Air

To examine whether humidity affects the folding of the peptide and therefore the final conformations of the quenching, we added one and two water molecules to the simulation box, respectively, as described in Subsection 3.1.4. These conditions correspond to approximately 30 % and 60 % relative humidity (or 0.8 and 1.6 water molecules per box) that were assumed as potential lower and upper boundary. Again, the system was simulated 1000 times for each environment and the final radii of gyration and the solvent accessible surface areas were compared.

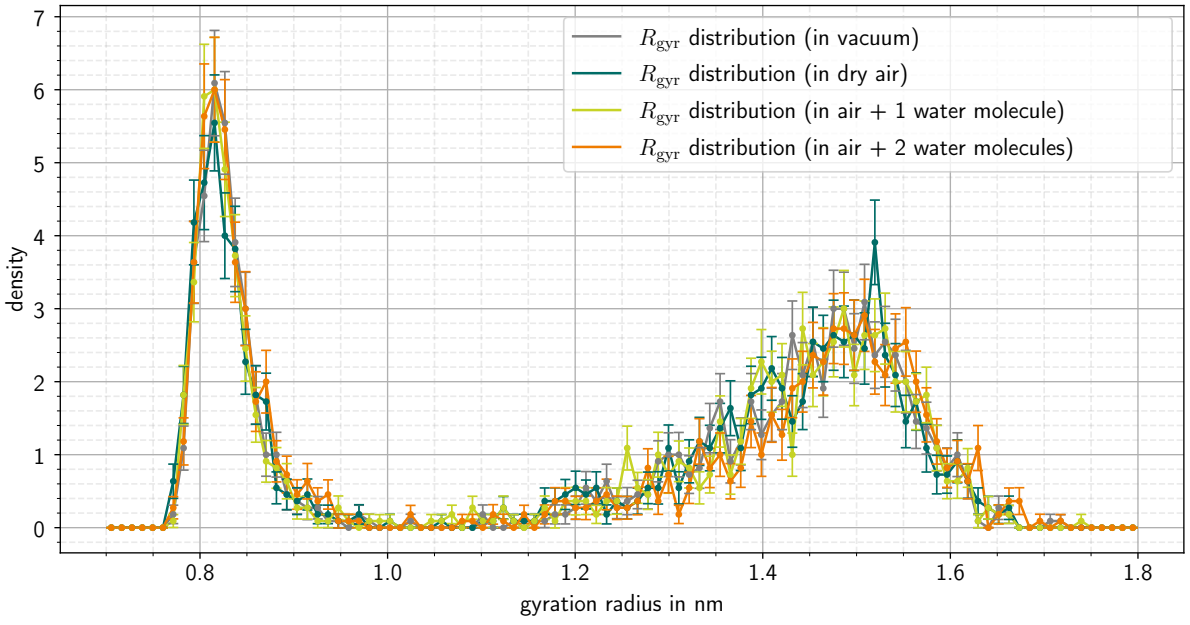


Figure 4.5: Radius of gyration distribution comparison between the final structures of 1000 quenching runs in vacuum, dry air, as well as humid air resembling 30 %, and 60 % relative humidity, respectively. There are no visible differences in the radius of gyration distributions between the structures under different condition.

4.2.1 Distribution of the Radii of Gyration

The histograms of the radii of gyration of all simulations including vacuum, as well dry and humid air conditions are shown in Figure 4.5. Again, there are no significant differences in the R_{gyr} distribution. Surprisingly, even adding one or two water molecules did not change the radius of gyration distribution of the folds.

4.2.2 Distribution of the Solvent Accessible Surface Area

Analogous to the radius of gyration distribution, the SASA distribution is shown for all simulations in Figure 4.6.
Similar to the radius of gyration as a metric, there are no noticeable differences in the SASA between the simulation environments.

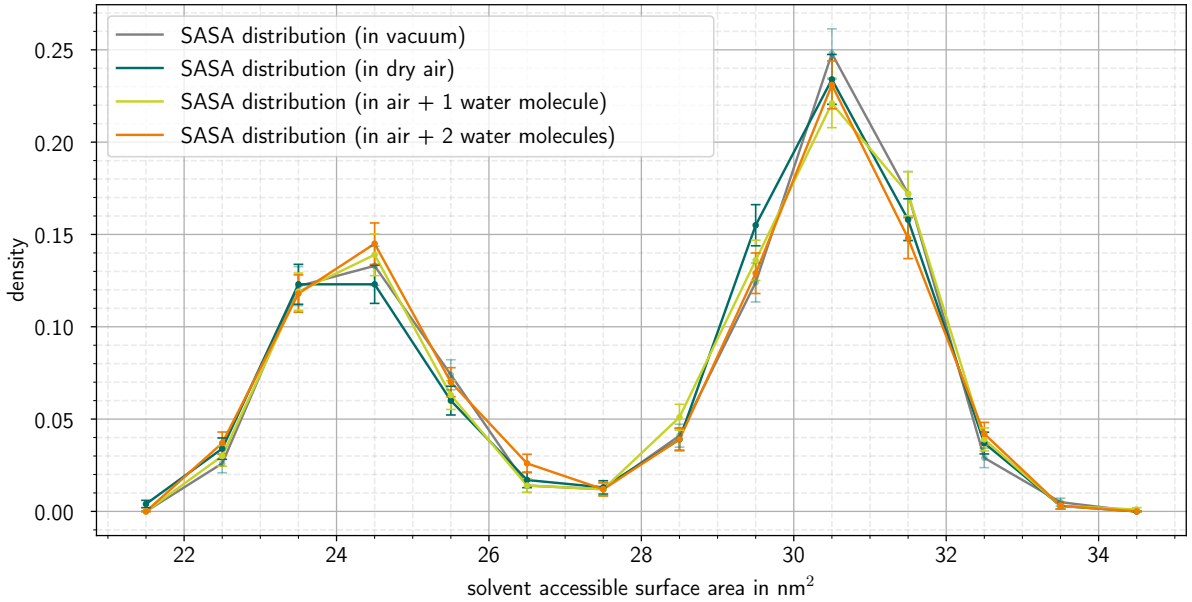


Figure 4.6: Solvent accessible surface area distribution comparison between the final quenching structures in vacuum, dry air, and humid air. The structures of less surface accessible to the solvent around 24 nm^2 are the spherical folds, the structures around 31 nm^2 the helical ones.

4.2.3 Encounters of Peptide and Water Molecules

These absences of differences in radius of gyration and SASA distributions were not expected, because water as a polar molecule is assumed to strongly interact with the peptide. As a result of that, we quantified the interactions of peptide and water molecules by calculating the fraction of peptides with at least one water molecule in their proximity

(within 2 nm distance). This fraction was calculated according to Subsection 3.2.7 and is displayed over the simulation time in Figure 4.7.

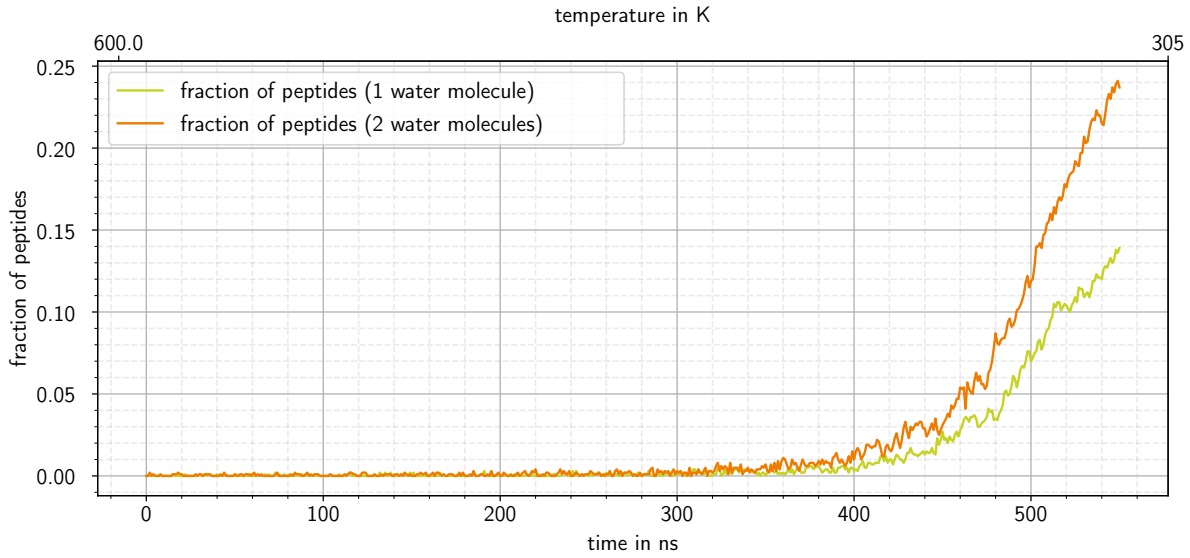


Figure 4.7: Fraction of peptides with at least one water molecule within 2 nm distance to the peptide over the course of the 550 ns of the simulation. Only for temperatures of around 400 K or lower do water molecules remain stable in the peptides close proximity. For the simulation with two water molecules in the simulation box, there are statistically more peptides with close water molecules.

It can be seen that only after 300 ns water molecules stay in the proximity of a significant fraction of the peptides. At the final frame, around 25 % of the peptides with two water molecules in the simulation box have at least one water molecule next to them, for the peptides with one water molecule this percentage is at around 15 %. This can be explained by the higher kinetic energy of the water molecules in the earlier stage of the simulation, only with lower kinetic energy the water molecules can form stable interactions with the peptide. Combining this result with the above transition rate now gives a possible explanation for the lack of influence of the water molecules on the final folds: significant water interactions are only stable after the final fold has been reached.

4.3 Hierarchical Clustering based on RMSD

Previously, the radius of gyration and the solvent accessible surface area were examined as an estimate for the collisional cross section. Even though there were no differences across environments in these metrics, there might be structural differences hidden from these proxies. To discover possible enrichments of certain conformations due to the respective simulation environments, we performed a hierarchical clustering based on the root mean

square deviation (RMSD) between all final structures as explained in Subsection 3.2.8.

However, there was no enrichment of certain conformations beyond the expected statistical deviations, meaning the environment does not influence the structures of the peptide.

Conclusion

Prior to this work, multiple collisional cross section (CCS) for individual peptides had been detected in TIMS experiments. The straightforward explanation of different folds per peptide had then been confirmed by MD simulations in vacuum. However, because the actual tube conditions differ from vacuum, we gradually adjusted the simulation environment by first adding air, and then additional one and two water molecules to mimic humidity. Goal was to see how the different simulation evironments influence the existence and the distribution of the two folds previously found in vacuum.
This was done by comparing the distribution of the radius of gyration R_{gyr} as well as the distribution of the solvent accessible surface area (SASA) of the final structures of 1000 simulation runs for each simulation condition. The radius of gyration and SASA were used as proxies for the CCS due to their easier calculation.

It was found that neither, low pressure air, nor additional water molecules within the simulation box affected the final structures. This was unexpected, because we were able to show that air does collide with the peptide and that the water molecules also spend a considerable amount of time within close proximity of the peptide.
The absence of an effect of the additional water molecules on the folding can be explained by the timing of the interactions: water molecules spend a considerable amount of time in the peptides proximity only after the eventual fold has been reached and became stable. For the lack of influence of the air on the folding, two explanations can be proposed. Either, the density of the molecules in the simulation box is too low to affect the peptide folding, or the individual interactions of the unpolar air and peptide are too weak to have any effect at all at any density.

Overall, it can be concluded that the sequence of a peptide has more influence on the folding than moderate variations in the simulation environment. Even though it might be possible to detect minor differences in the R_{gyr} and SASA distributions by increasing the sample size to more than 1000 simulation runs, these differences will likely be within the

current statistical fluctuations of the distributions and therefore be comparably small. Hence, the consistent distributions across environments can be seen as reliable.
To see the degree to which the fold distribution remains stable under the influence of the environment, the simulation box conditions can be further varied to more extreme, unphysical conditions as a follow-up. Increasing the pressure and thereby the density of the air will answer the above question whether collisions with air are capable to influence the folding at all. In case of proteins in solution it is known that the folding differs strongly to vacuum or low pressure air environments. Here, a gradual increase in water molecules in the simulation box can determine the critical amount at which the folding is affected.

References

[1] Abhisek Mukherjee et al. "Type 2 diabetes as a protein misfolding disease". In: *Trends in Molecular Medicine* 21.7 (July 2015), pp. 439–449. ISSN: 1471-4914. DOI: 10.1016/j.molmed.2015.04.005. URL: https://doi.org/10.1016/j.molmed.2015.04.005.

[2] Yangying Zhou et al. "Proteomic signatures of 16 major types of human cancer reveal universal and cancer-type-specific proteins for the identification of potential therapeutic targets". In: *Journal of Hematology & Oncology* 13.1 (Dec. 2020), p. 170. ISSN: 1756-8722. DOI: 10.1186/s13045-020-01013-x. URL: https://doi.org/10.1186/s13045-020-01013-x.

[3] James N. Dodds and Erin S. Baker. "Ion Mobility Spectrometry: Fundamental Concepts, Instrumentation, Applications, and the Road Ahead". In: *Journal of The American Society for Mass Spectrometry* 30 (2019), pp. 2185–2195. ISSN: 1879-1123. DOI: 10.1007/s13361-019-02288-2. URL: https://doi.org/10.1007/s13361-019-02288-2.

[4] Karsten Michelmann et al. "Fundamentals of Trapped Ion Mobility Spectrometry". In: *Journal of the American Society for Mass Spectrometry* 26.1 (2015). PMID: 25331153, pp. 14–24. DOI: 10.1007/s13361-014-0999-4. eprint: https://doi.org/10.1007/s13361-014-0999-4. URL: https://doi.org/10.1007/s13361-014-0999-4.

[5] Andries P Bruins. "Mechanistic aspects of electrospray ionization". In: *Journal of Chromatography A* 794.1 (1998), pp. 345–357. ISSN: 0021-9673. DOI: https://doi.org/10.1016/S0021-9673(97)01110-2. URL: https://www.sciencedirect.com/science/article/pii/S0021967397011102.

[6] Simon J. Gaskell. "Electrospray: Principles and Practice". In: *Journal of Mass Spectrometry* 32.7 (1997), pp. 677–688. DOI: https://doi.org/10.1002/(SICI)1096-9888(199707)32:7<677::AID-JMS536>3.0.CO;2-G. eprint: https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/pdf/10.1002/%28SICI%291096-9888%28199707%2932%3A7%3C677%3A%3AAID-JMS536%3E3.0.

CO%3B2-G. URL: https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291096-9888%28199707%2932%3A7%3C677%3A%3AAID-JMS536%3E3.0.CO%3B2-G.

[7] Joshua A. Silveira et al. "From Solution to the Gas Phase: Stepwise Dehydration and Kinetic Trapping of Substance P Reveals the Origin of Peptide Conformations". In: *Journal of the American Chemical Society* 135.51 (2013). PMID: 24313458, pp. 19147–19153. DOI: 10.1021/ja4114193. eprint: https://doi.org/10.1021/ja4114193. URL: https://doi.org/10.1021/ja4114193.

[8] Shannon A. Raab et al. "Protons Are Fast and Smart; Proteins Are Slow and Dumb: On the Relationship of Electrospray Ionization Charge States and Conformations". In: *Journal of the American Society for Mass Spectrometry* 32.7 (2021). PMID: 34151568, pp. 1553–1561. DOI: 10.1021/jasms.1c00100. eprint: https://doi.org/10.1021/jasms.1c00100. URL: https://doi.org/10.1021/jasms.1c00100.

[9] Andrew R. Leach. *Molecular Modeling: Principles and Applications*. 2nd ed. Pearson Education, 2001. ISBN: 0-582-23933-8.

[10] Qian Mao et al. "Classical and reactive molecular dynamics: Principles and applications in combustion and energy systems". In: *Progress in Energy and Combustion Science* 97 (2023), p. 101084. ISSN: 0360-1285. DOI: https://doi.org/10.1016/j.pecs.2023.101084. URL: https://www.sciencedirect.com/science/article/pii/S036012852300014X.

[11] Giovanni Bussi, Davide Donadio, and Michele Parrinello. "Canonical sampling through velocity rescaling". In: *The Journal of Chemical Physics* 126.1 (Jan. 2007), p. 014101. DOI: 10.1063/1.2408420. eprint: https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.2408420/15393604/014101_1_online.pdf. URL: https://doi.org/10.1063/1.2408420.

[12] *Transport Information Service (TIS). Climate/humidity table*. URL: https://www.tis-gdv.de/tis_e/misc/klima-htm/ (visited on 05/08/2024).

[13] Alexander Kulesza et al. "Bringing Molecular Dynamics and Ion-Mobility Spectrometry Closer Together: Shape Correlations, Structure-Based Predictors, and Dissociation". In: *The Journal of Physical Chemistry B* 122.35 (2018). PMID: 30068075, pp. 8317–8329. DOI: 10.1021/acs.jpcb.8b03825. eprint: https://doi.org/10.1021/acs.jpcb.8b03825. URL: https://doi.org/10.1021/acs.jpcb.8b03825.

[14] Florent Calvo et al. "Statistical Analysis of Ion Mobility Spectrometry. II. Adaptively Biased Methods and Shape Correlations". In: *Journal of the American Society for Mass Spectrometry* 23.7 (2012). PMID: 22573497, pp. 1279–1288. DOI:

10.1007/s13361-012-0391-1. eprint: https://doi.org/10.1007/s13361-012-0391-1. URL: https://doi.org/10.1007/s13361-012-0391-1.

[15] *GROMACS documentation. Radius of gyration and distances*. URL: https://manual.gromacs.org/documentation/nightly/reference-manual/analysis/radius-of-gyration.html (visited on 06/27/2024).

[16] *CompChems. How to compute the Solvent Accessible Surface Areas (SASA) with GROMACS*. Oct. 30, 2022. URL: https://www.compchems.com/how-to-compute-the-solvent-accessible-surface-areas-sasa-with-gromacs/#solvent-accessible-surface-area-sasa (visited on 07/02/2024).

[17] Frank Eisenhaber et al. "The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies". In: *Journal of Computational Chemistry* 16.3 (1995), pp. 273–284. DOI: https://doi.org/10.1002/jcc.540160303. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.540160303.

[18] *IMos*. URL: https://www.imospedia.com/imos/ (visited on 08/01/2024).

[19] *NumPy documentation. numpy.std*. URL: https://numpy.org/doc/stable/reference/generated/numpy.std.html (visited on 06/27/2024).

Appendix

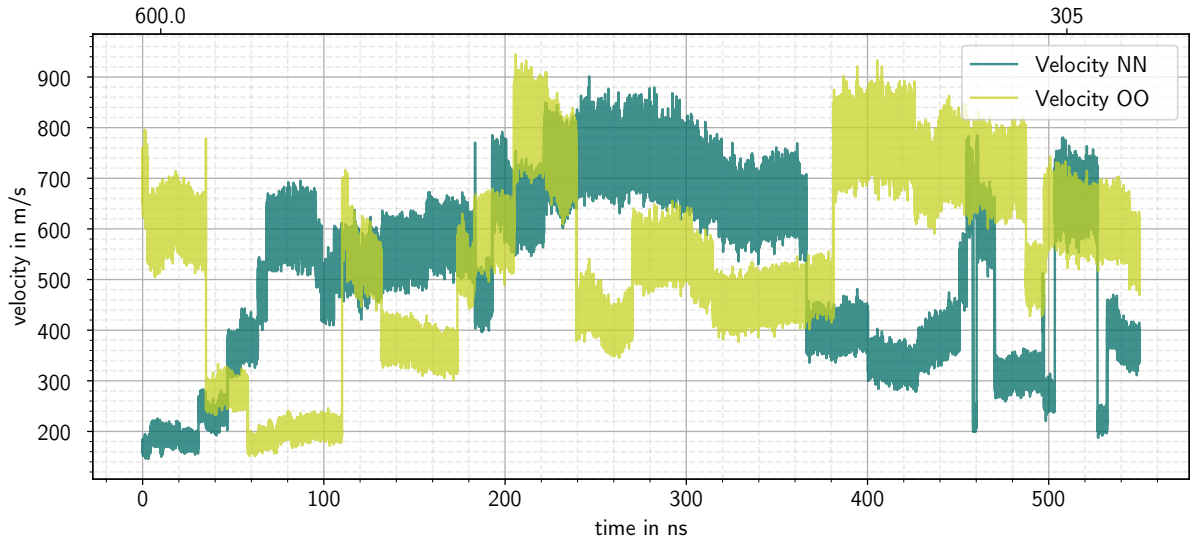


Figure 6.1: Undirected velocities in m/s of one exemplary nitrogen molecule, and one exemplary oxygen molecule over one simulation run. The changes in velocity are caused by collisions with the peptide and show that collision or close encounters between peptide and air molecules do occur. Counting the close encounters according to Subsection 3.2.6 results into 246 collisions with nitrogen molecules and 62 collisions with oxygen molecules over the course of the entire simulation.

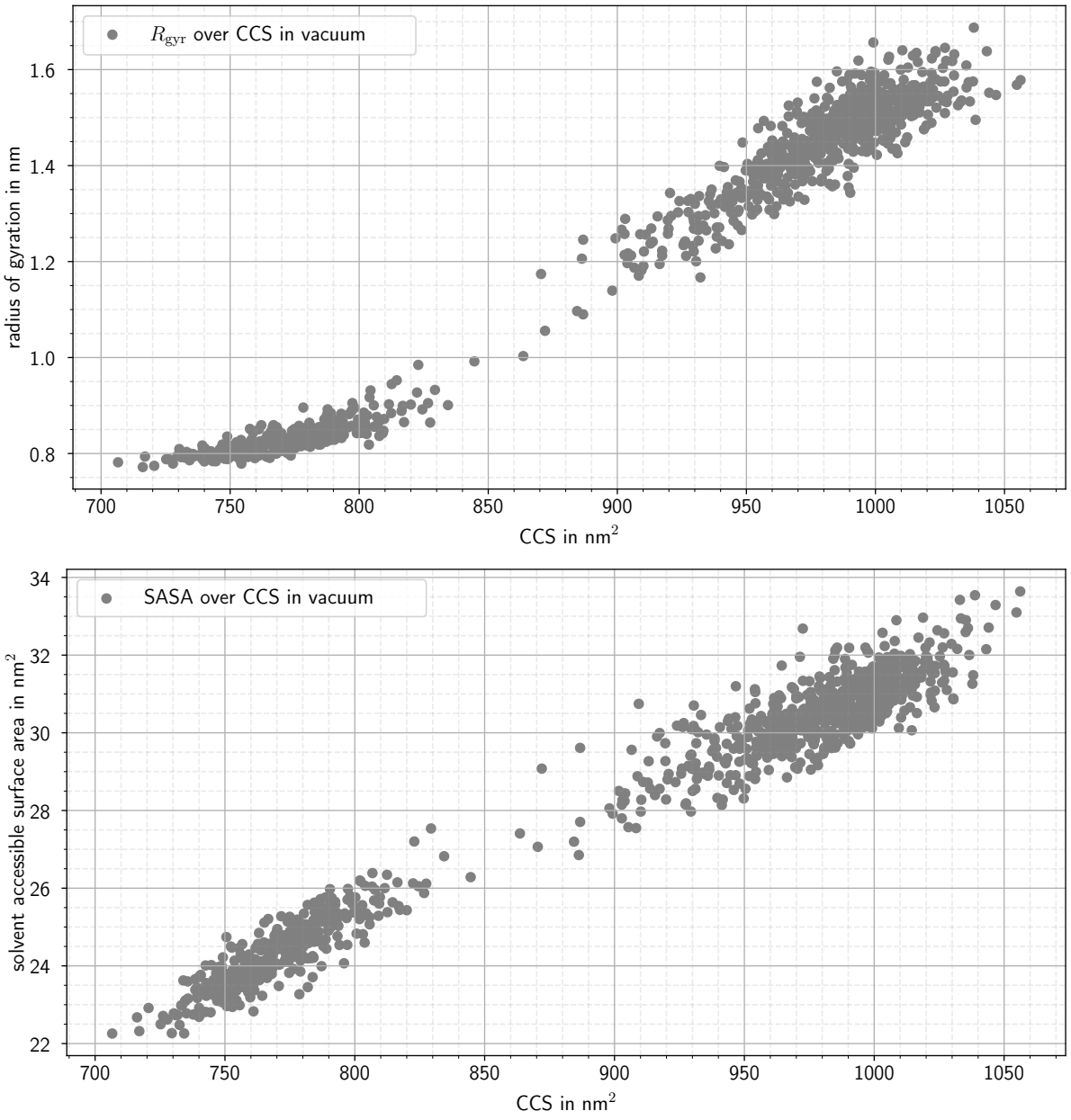


Figure 6.2: Radii of Gyration R_{gyr} and solvent accessible surface areas (SASA) plotted against their corresponding values for the collisional cross section for the simulations in vacuum. The radius of gyration shows an approximately quadratic relation with the collisional cross section, the SASA a linear one. This is in line with the dimensionalities: R_{gyr} is measured in nm, SASA and CCS both have the unit nm^2. R_{gyr} and SASA both having a monotonistic relationship with the collisional cross section confirms their suitability as a CCS proxy as each CCS value has an unambiguous R_{gyr} or SASA counterpart.

Acknowledgement

First of all, I would like to thank Professor Helmut Grubmüller for the opportunity to work in his group. Entering Biophysics has been a long term goal for me dating back to the time before I started studying physics.
Further, I want to thank Professor Stefan Klumpp for being the second referee of the thesis.

Finally, I am very grateful to Daniel for convincing me to work with him and investing all those hours into helping me. It was a short but very enjoyable time and I could not have hoped for a better introduction to MD simulations.