Master’s Thesis

Interpretation of Anisotropic Solution X-ray Scattering by Means of Molecular Dynamics Simulation

Interpretation anisotropischer Röntgen-Streubilder von Lösungen mit Hilfe molekular-dynamischer Simulation

prepared by
Levin Brinkmann
from Münster

in the Computational Molecular Biophysics Group
at the Department of Molecular Structural Biology

Thesis period: 1st March 2014 until 13th September 2014

First Referee: Dr. Jochen Hub

Second referee: Prof. Dr. Tim Salditt
Abstract

Time-resolved wide-angle X-ray scattering (TR-WAXS) holds great promise to observe biomolecules while they function, in real time and in their natural environment. However, the structural interpretation of WAXS remains difficult because of the low information content of the patterns. Hence, anisotropy in the scattering pattern is expected to provide additional structural information, but calculating such patterns was so far impossible. Furthermore, current methods for structural interpretation of TR-WAXS often make use of the small scattering angles only.

Here, we report a new method to calculate anisotropic solution scattering patterns from molecular dynamics (MD) simulations. We demonstrate our method using CO dissociation of myoglobin as a test case. Qualitative agreement with the experiment was found. The main features of the anisotropic scattering pattern can be traced back to the initial displacement of the CO upon excitation and its migration between distinct cavities. In addition, we found slow relaxation processes independent of the CO migration. Here, we provide a structural interpretation of anisotropic and wide-angle isotropic TR-WAXS patterns, which have not been reported before. Scattering patterns are calculated with time-resolution spanning more then six orders of magnitude, from femtoseconds to 100 nanoseconds, readily giving a prediction of scattering pattern for ultrashort timescales, just to come into reach with the advent of free electron lasers.

The formalism used by Park et al. [The Journal of chemical physics, v. 130.13, No. 134114 (2009)] for the calculation of isotropic WAXS patterns was founded on a new theoretical ground and slightly modified, to be applicable for TR-WAXS patterns. Anisotropy of scattering patterns due to photo-selectivity was theoretically described and, for typical photo-selection probabilities, TR-WAXS patterns were proven to contain exactly two linearly independent components for each scattering angle. In Fourier space of the protein, these two components correspond to an isotropic average and to an average weighted by a second spherical harmonic. The latter was shown to decay exponentially due to rotational diffusion. We expect the information content of TR-WAXS to double by consideration of anisotropy. Advanced experiments with further increased information content are suggested.

We could show that the anisotropic component can be retrieved from a horizontal and a vertical cut of the 2D scattering pattern. In contradiction to earlier considerations, the radially averaged scattering pattern does contain an anisotropic contribution, too. Its decay due to rotational diffusion adds to the time evolution of the pattern and has to be considered in the interpretation of timescales and structural dynamics retrieving from TR-WAXS pattern.

This work is a case study on the interpretation of time-resolved crystallographic and WAXS experiments by the means of MD simulations. We believe our methods to be useful to unleash the potential of MD simulations to bridge these experimental methods, to guide their structural interpretation, thereby providing dynamics in atomistic detail and in a natural environment.
## Contents

1 Introduction 1

2 Solution X-ray Scattering 9
   2.1 Buffer subtraction .............................................. 10
   2.2 Time-resolved solution scattering .............................. 11
   2.3 Anisotropic solution scattering ................................ 13

3 Theory 15
   3.1 Molecular dynamics .............................................. 15
      3.1.1 Basics ....................................................... 15
      3.1.2 Leap frog integrator ........................................ 18
      3.1.3 Pressure and temperature coupling .......................... 19
      3.1.4 Replica exchange molecular dynamic ......................... 20
   3.2 Molecular scattering ............................................ 21
      3.2.1 The scattering amplitude ................................... 21
      3.2.2 Cromer-Mann parameter ..................................... 23
   3.3 Isotropic solution scattering ................................... 25
   3.4 Anisotropic time-resolved solution scattering ................. 34
   3.5 Rotational diffusion ............................................ 42
   3.6 Information content ............................................ 45

4 Implementation 49
   4.1 Spiral method .................................................... 49
   4.2 g_genenv .......................................................... 50
   4.3 mdrun .............................................................. 51
      4.3.1 Preparation ................................................... 52
      4.3.2 Calculation of scattering amplitude ....................... 53
   4.4 passrot ............................................................ 54
   4.5 Evaluation ........................................................ 55
## Contents

5 Computational Details 57

5.1 MD parameter ........................................ 57
5.2 CO model ............................................... 58
5.3 Dissociation process .................................. 59
5.4 Branches ............................................... 61
5.5 Rotational diffusion constant ....................... 64
5.6 Clustering ............................................... 65
  5.6.1 General remarks ................................. 66
  5.6.2 Clustering into Micro-states .................. 67
  5.6.3 Clustering into Marco-states ................. 67

6 Results and Discussion 71

6.1 CO diffusion ........................................ 71
  6.1.1 General remarks ................................. 71
  6.1.2 Kinetic model .................................... 72
6.2 Comparison with crystallographic data .............. 75
6.3 Time evolution of electron density and scattering pattern .... 76
6.4 Femtoseconds to picoseconds ....................... 77
6.5 One picosecond to one hundred picosecond .......... 79
6.6 Picoseconds to nanoseconds ....................... 81
  6.6.1 Electron density maps .......................... 81
  6.6.2 Radially averaged WAXS pattern ............. 84
  6.6.3 Difference pattern from horizontal and vertical cuts .... 87
6.7 Scattering pattern derived from cluster .......... 88
6.8 Backbone-restrained simulation ................... 90
6.9 Comparison with experimental data ............... 91

7 Conclusions and Outlook 95
1 Introduction

There is a natural interest in understanding life mechanistically on a molecular level. The basic principles of encrypting life in the deoxyribonucleic acid (DNA) are well understood despite of constant reevaluation. The genome of a number of organisms including humans have been decoded. The process of transcription of DNA into ribonucleic acid (RNA) as well as the mechanism of translation into proteins have been excessively studied. All this knowledge is of limited value without understanding the function of the biomolecules encoded.

Although certainly gaining the same attention, the process of folding and complex formation which result in a functional protein as well as the functional mechanisms of proteins and other biomolecules are much less understood. Proteins and RNA are building blocks and functional elements of the machinery of the cell. Their individual functional exercise within the cell depends on their interaction. The underlying mechanism is however only to be understood by their structure and dynamics. Changes in the shape alter the affinities between biomolecules. In addition, chemical reactions are mechanically triggered. Understanding the folding and the function of proteins on a mechanistic level has the potential of creating new momentum in medicine, pharmacy and biotechnology and is the precondition for a straight forward design of proteins.

There are numerous experimental and numerical methods used to explore the structure, dynamics and function of proteins. Crystallography and NMR have successfully resolved the structure of a large number of biomolecules in atomic detail. These techniques provide usually only a single or a few snapshots of all possible confirmation. At physiological temperature, biomolecules have a high flexibility and physical properties have to be calculated for an ensemble average.

Molecular dynamics (MD) simulation are numeric methods which provide such an ensemble of structures as well as information about the dynamics of the system. These simulations mimic physical laws of the dynamics of molecular systems. The time-dependent Schrödinger equation is able to describe molecular systems accu-
1 Introduction

rately. However, solving the Schrödinger equation exceeds modern computational capabilities even for small proteins. Thus several approximations are needed and the dynamics rely instead on an empirically parametrized model, which undermine the ab-initio character of the numerical method. For proteins the term MD is often used synonymously for simulations employing empirical force fields. State of the art force fields such as CHARMM [1], AMBER [2], OPLS [3] and GROMOS [2] have proven their performance in comparison with data obtained from NMR experiments [4–6]. They are widely used tools in molecular biology to study, for instance, the functional mechanism of proteins [7–9], their mechanical properties [10, 11], membrane protein permeation [12, 13] and folding pathways of proteins [14, 15]. The accuracy of the results depends on the sampling and the force field. Replica exchange MD (REMD) [16] as well as MD simulations based on Markov models [17, 18] have been used to improve sampling with a number of shorter simulations instead of a single long one.

For proteins, the relevant sizes are measured in Angstrom (Å=10^{-10} \text{m}). Here, experimental methods reach their limitations and can thus reveal relevant properties often only in an indirect manner. Numerical atomic models on the other hand, as pointed out, rely for biological system sizes on various approximations and may not be able to fully cover relevant timescales. It might thus be expected that many questions can only be answered in a combined approach, gathering information from various experimental methods as well as computational approaches. Computational techniques can assist in the interpretation of experimental results. Conversely, experimental results are needed for validating numerical approaches and underlying assumptions.

Since threedimensional structure is a result of the interplay of environment and genetic sequence, it is impossible to deduce structure from sequence only. Moreover, structure determination is hindered by the vast increase in possibilities when mapping from one to three dimensions. Therefore, experimental structures are typically used as starting point for MD simulations. On the contrary, MD simulations have been used for a refinement crystallographic [19, 20] and NMR [19–21] structures. Other experimental methods have been interpreted using MD simulations, like Förster resonance energy transfer [22] and force probing [11].

There is an increasing interest in using MD simulation for providing an atomistic interpretation of solution X-ray scattering. In solution X-ray scattering as well as crystallography proteins are radiated by a X-ray beam and their scattering is measured. However, solution X-ray scattering differs from the latter, in that molecules
are not crystallized but move freely in solution. The molecules are thus measured in their natural ensemble of conformations in solution, which may substantially differ from crystal structures [23]. Solution X-ray scattering shares this advantage with NMR, it is, however, unlike NMR, not limited to smaller proteins.

In comparison to crystallography and NMR the interpretation of solution X-ray scattering patterns suffers from low informational content of the one dimensional scattering curve. This curve has been estimated to contain 10-25 independent data points [24], depending on the size of the protein.

Nevertheless, solution X-ray scattering is highly sensitive to differences in structure of molecules [25]. Solution scattering is traditionally applied to studying protein responses to changes in the chemical environment [26, 27] and the dynamics of folding and unfolding [28, 29]. The key advantage of solution X-ray scattering is the ability to obtain data from any kind of probe and the relatively simple experimental setup allowing high throughput studies [25]. The recent increase in interest in this method is connected to advances in X-ray source technology.

Depending on the observed scattering angles, solution X-ray scattering experiments are either called small-angle x-ray scattering (SAXS) or wide-angle X-ray scattering (WAXS). High quality measurements in the WAXS regime have come possible by third generation synchrotrons, which now also allow time-dependent measurements [24]. Further advances are expected with the upcoming free electron lasers [30].

As the scattering pattern is a reciprocal image of the electron density, small scattering angles measure the general shape and higher scattering angles to secondary and tertiary structure. The methods used for the interpretation of solution X-ray scattering depend on the scattering angles. Numerous methods for the interpretation of solution X-ray scattering patterns up to intermediate angles have been developed [24, 31], but only a few can make use of wide scattering angles.

In the range of very small angles, information about the radius of gyration can be calculated using the Guinier’s equation [32]. Several ab initio methods have been developed to reconstruct three dimensional shapes from observed scattering patterns based on spherical harmonics [33] and on simulated annealing of dummy beat models [33, 34]. These methods are limited to shapes of low resolution. Rigid body models have been used to limit the degrees of freedom to global parameters like domain orientations to avoid overfitting of high resolution electron densities [35]. While the described methods are aiming for a reconstruction of the electron density
in real space from scattering patterns, other methods have been developed aiming for the contrary; to calculate expected experimental solution scattering curves from given atomistic structures.

Solution X-ray scattering is a contrast method where the scattering intensity of the solution is compared to the one of pure solvent. Its purpose is to cancel the effect of the bulk solvent, however, the solvent volume excluded by the protein must still be considered. Additionally, it is known that solvent molecules form solvation shells around proteins, the density of which exceed the one of bulk water by 10-15% [36, 37] (see Fig. 1.1.A). Methods for the calculation of WAXS scattering curves differ mainly in the treatment of (i) the solvent volume excluded by the protein, (ii) the solvation layer and (iii) in the method of spherical averaging and calculation of the Fourier transform.

Most methods threat the solvent implicitly assuming its homogeneity. Like others [38, 39] the popular CRY SOL [36] package accounts for the excluded volume by adding an additional factor to the proteins atomic form factors. There, as well as in other methods [38, 40, 41], the spherical average is computed in terms of a multiple expansion into spherical harmonics, which is efficient within the SAXS regime. In CRY SOL an envelope around the protein with uniform thickness is used to represent a homogenous solvation shell around the protein. Furthermore, the protein’s atomic scattering factors have been modified depending on their solvent-accessible surface [39] and the electron density has alternatively been calculated on a grid with corrections accounting for either the excluded volume or to the solvation shell [37, 42]. The natural solvation depends on the electrostatics of the protein surface and is far from uniform (see Fig. 1.1.A). Therefore a reasonable agreement of implicit solvent calculations with the experiment usually requires the fitting of 2-3 free parameters [43] with a risk of overfitting [44]. At higher scattering angles, finer structural differences can be resolved demanding a more accurate description of the proteins solvation shell.

When solvent scattering patterns are to be calculated from MD trajectories it is only natural to include the readily available information of the solvent and the excluded volume is, however, not trivial. Treating the excluded volume as extensions of the scattering factors like within the SASSIM package [40] may fail in the WAXS regime since this method yields an incorrectly structured density of the excluded volume [45]. High precision in the WAXS regime have been achieved using methods calculating the contribution of the excluded volume from a trajectory of pure solvent.
Figure 1.1: Myoglobin and its solvation shell (A) is shown. Isosurfaces of the average electron density calculated from 10000 structures are depicted in black for $400 \text{e} \cdot \text{nm}^{-3}$, red for $280 \text{e} \cdot \text{nm}^{-3}$ and blue $337.6 \text{ e} \cdot \text{nm}^{-3}$. The latter is calculated from the solvent electron density solely. Myoglobin within an envelope (orange) is depicted (B) including the q-vectors used for spherical averaging (yellow). Solvent molecules are explicitly shown within the envelope only.

[41–43] or alternatively using pair-distance distribution functions [46].

Park et al. [42] followed the first approach and derived an elegant formalism for calculating scattering intensities from two ensembles, one representing the solute in solution and one with the pure solvent. In both cases only those atoms are considered that lie within a certain distance from the protein respectively its ghost image in the pure solvent case. In their approach, the scattering intensity is calculated by a numerical spherical average. Chen and Hub [43] modified the method in a sense that they introduced a fixed envelope (see Fig. 1.1.B) including the solute and its solvation shell. Using a fixed envelope is essential for an average over a trajectory and it allows the calculation of heterogenous ensembles while at the same time reduces numerical cost and noise.

In this work the method presented by Chen and Hub is substantially extended in order to facilitate the calculation of anisotropic time-resolved WAXS (TR-WAXS) patterns.

TR-WAXS is related to time-resolved spectroscopy [47–50], where a stimulating laser pulse is followed by a probing laser pulse. By altering the time-delay between those, temporal resolution of the monitored process can be obtained. With the advent of sufficiently brilliant X-ray sources equivalent experiments can be made with a X-ray probing beam enabling time-resolved crystallography [51–53]. Only recently the same setup has been used for time-resolved solution X-ray experiments [54–57]. All these experiments aim for the observation of structural changes upon
an excitation event.

The ensemble of excited proteins is usually anisotropic, as the excitation probability typically depends on the orientation with respect to laser polarization. The idea of employing this effect, called photo-selection, to study rotational diffusion has already been described by Perrin in 1936 [58]. Based on this method, size, shape, flexibility and associative behavior of biomolecules [59] as well as local viscosity [60] were indirectly measured.

In solution X-ray scattering experiments, anisotropy in the probed ensemble manifests itself in a breaking of radial symmetry of the scattering pattern. This has been used to determine the order of organic polymer solutions [61, 62], but only recently such a breaking of symmetry due to photo-selection was reported in the context of TR-WAXS [63, 64]. The interpretation of these results were limited to the determination of rotational diffusion constants [63] and basic geometry [64]. Advanced methods are needed to make use of the radial component when interpreting TR-WAXS patterns, which constitutes one of the motivations for this thesis.

The system studied in this work is the protein myoglobin and its relaxation after dissociation of CO from its heme group. Being the first protein solved structural [65], myoglobin has served as a test case for numerical and experimental methods.

MD simulation have been performed to study dissociation [66–68], ligand diffusion (pathways) [69–73], the hydration shell [74] and heat dissipation [75]. In addition, quantum mechanical calculations of the dissociation process have been performed [76, 77].

Experimentally, time-resolved absorption spectroscopy has been used to study the excess energy upon dissociation [78, 79], heat dissipation [47, 78], rebinding of CO [80–82], the protein dynamics [79, 83] and CO diffusion pathways [78, 84]. Time-resolved crystallography has been performed with nanosecond [51, 85] and picosecond [52, 53] time resolution. More recently, isotropic [55, 56] and anisotropic [63] TR-WAXS experiments with 100ps time resolution were conducted, which are the focus of this study.

So far, TR-WAXS experiments have not gained the same attention as other time-resolved methods. This can be largely referred to a missing structural interpretation of these experiments and the low information content inherent in solution scattering experiments. In principal MD simulation could provide this structural interpretation. The obtainable amount of information can be enhanced by considering
anisotropy.

In this work a method is presented allowing the calculation of anisotropic time-resolved scattering patterns from MD simulations. It turns out that from anisotropic scattering patterns two independent components can be derived one of which is exponentially decaying due to rotational diffusion. The method is applied to the relaxation process of myoglobin after photodissociation of CO. Particular emphasis is put on controlling numerical noise and sampling bias. Results are compared to experimental data and prospects of upcoming experiments are pointed out. A possible structural interpretation of the results in terms of a kinetic model is presented as well.
2 Solution X-ray Scattering

The experimental setup for solution X-ray scattering is similar to the one of crystallography and both can generally be performed with identical hardware. A sample is irradiated by a high energetic monochromatic X-ray beam and the scattered radiation is collected by a two dimensional detector positioned in beam direction behind the target (see Fig. 2.1.A). In the case of crystallography, scattering is limited to single points obeying Bragg’s law. The absence of a crystallographic order and the therefore random orientation of the proteins results in a spherical averaging. As a result WAXS pattern are radially symmetric and typically radially averaged in favor of an improved statistic.

The scattering angle $\alpha$ is related to the momentum transferred to the photon. Their relationship in the case elastic scattering ($|k| = |k_0|$)

$$q = 2 \cdot k \cdot \sin \frac{\alpha}{2}$$

(2.1)

can be obtained geometrical (see Fig. 2.1.B). The momentum transfer $q$ will be used throughout this thesis as the physically more relevant measure. Depending on the momentum transfer solution X-ray scattering experiments are either called

Figure 2.1: (A) The experimental setup of solution X-ray scattering. A target is radiated by an X-ray beam, and a scattering pattern is measured by a detector. The geometry of the scattering is shown in (B) including the incoming ($k_0$) and the scattered wave vector ($k$), as well as the scattering vector $q$ representing the momentum transfer and the scattering angle $\alpha$. 

SAXS for momentum transfer up to \( q < 5 \text{ nm}^{-1} \) or WAXS for momentum transfer up to \( q < 25 \text{ nm}^{-1} \). In the course of this thesis the term WAXS may be used for solution X-ray scattering in general without differentiation towards SAXS in terms of the analyzed momentum range.

2.1 Buffer subtraction

To obtain information about the shape of a measured protein, solution X-ray scattering traditionally tried to match the scattering intensity of the protein itself and to remove all contributions of the solvent. For this reason, the scattering curve \( I_{\text{sol}}(q) \) of the protein in solution and a second one \( I_{\text{buf}}(q) \) is subtracted,

\[
I_{\text{prot}}(q) = I_{\text{sol}}(q) - (1 - \chi_s) I_{\text{buf}}(q) \{-\chi_s I_{\exp}(q)\}
\]

(2.2)

which is obtained from a measurement of the solvent without solute. Due to the excluded volume of the protein, the amount of solvent in both measurements differ. Here \( \chi_s \) is the volume fraction of the protein in the solution and the pre-factor \( 1 - \chi_s \) correspondingly matches with the volume fraction of solvent. The buffer subtraction also reduces contribution to the scattering curve originating from the experimental setup e.g. the curvet. However, as \( \chi_s \) differs from 0 they do not cancel out completely and an additional term (curved brackets) is needed to correct for the scattering \( I_{\exp}(q) \) of the experimental setup. In practice the latter correction is rarely done [24]. The detector in an scattering experiment measures the number of photons \( N_x(r_d) \) at a given radius \( r_d \). For absolute consistent results the concentration of the protein \( c \), the detector response \( DR(r_d) \) at given distance \( r_d \) from the detector center and the intensities \( I_{0\text{sol}}, I_{0\text{buf}}, I_{0\exp} \) of the x-ray beam during the measurements have to be taken into account.

\[
I_{\text{prot}}(q) = \frac{1}{c \cdot DR(r_d)} \left[ \frac{N_{\text{sol}}(r_d)}{I_{0\text{sol}}} - (1 - \chi_s) \frac{N_{\text{buf}}(r_d)}{I_{0\text{buf}}} - \chi_s \frac{N_{\exp}(r_d)}{I_{0\exp}} \right]
\]

(2.3)

The relation between different points on the detector and the momentum transfer \( q \) can either be calculated by geometric considerations following eq. 2.1 or by measurement of a standard sample. From equation 2.3 it becomes apparent that the
2.2 Time-resolved solution scattering

scaling of the scattering intensity depends on a number of factors. Especially the protein concentration and the corresponding volume fraction may vary from experiment to experiment or even within an experiment as the solution is usually pumped through the curvet during exposure to reduce radiation damage. In addition the intensity of the X-ray beam may fluctuate.

To avoid the ambiguity of the scaling often only the relative scattering intensities are considered and volume fraction can be dismissed if the excess intensity

\[
\Delta I(q) = I_{\text{sol}}(q) - I_{\text{buff}}(q)
\]  

is reported, which represents the electron density difference between solute and solvent.

This corresponds mathematically to \( \chi_s \) being equal to 0 in (2.2) and (2.3) making the third term therein redundant. The excluded volume is not obvious to determine due to the proteins solvation shell, whereas the latter equation is well defined. However its interpretation may be less intuitive. From a modeling point of view as well as from an experimental point of view the simple difference intensity (2.4) is easier to handle and may thus be more suitable for linking experiment and simulation.

An expression in terms of single protein contributions only holds for concentrations where protein-protein interaction do not play a role. While a repulsive protein-protein interaction usually results in a decrease at the very low angles, attractive protein-protein interactions usually results in an steep increase of the scattering intensity.

2.2 Time-resolved solution scattering

Unlike time-independent WAXS, TR-WAXS does not target the structure of the observed molecule, but instead structural changes after a triggering event, such as a photoinduced excitation of the solute. The difference in the experimental setup as depicted in Fig. 2.2 is an additional laser. The laser excites the sample and thereby induces structural changes, which are then measured by the succeeding X-ray beam. For cancelation of systematic errors, similar to the time independent case, a difference spectra is analyzed. However the protein solution is not considered in contrast to the pure solvent, but instead to the same protein solution prior to
Figure 2.2: The experimental setup of TR-WAXS. In addition to the X-ray beam, there is a laser, which excites the probe. The time-resolution is obtained by altering the delay between laser and X-ray beam. A possible laser excitation moment and excitation transition moment of the protein are depicted. Their relative angle determines the excitation probability.

excitation.

\[ \Delta I_{\Delta t}(q) = I_{t=t_0+\Delta t}(q) - I_{t<t_0}(q) \]  

(2.5)

The scattering intensity prior excitation \( I_{t<t_0}(q) \) is subtracted from the intensity after excitation \( I_{t=t_0+\Delta t}(q) \) which is obtained at time \( t = t_0 + \Delta t \). The time delay \( \Delta t \) between the excitation event \( t = t_0 \) and the X-ray beam at \( t = t_0 + \Delta t \) can be altered and thus the time-evolution of the scattering pattern measured. The time resolution is limited by the pulse length of excitation laser and X-ray beam. In practice, brilliance of the X-ray beam limits the time-resolution, because a certain amount of integrated intensity is needed for a sufficient signal to noise ratio. In favor of a higher intensity and thus a better temporal resolution, the X-ray source is usually not monochromatized.
2.3 Anisotropic solution scattering

As stated beforehand, scattering pattern of time-independent solution scattering are radially symmetric, due to the rotational freedom of the proteins. In time-resolved solution scattering, however, the excitation can induce anisotropy within the sample.

Excitation is the transition of a system from a lower into a higher energy level. The probability of a transition from a state A to a state B by interaction with an electromagnetic wave is proportional to the $\cos^2(\phi)$ of the angle $\phi$ between the corresponding transition moment $\mathbf{m}_{AB}$ and the electromagnetic waves polarization $\mathbf{I}$. Fig. 2.3 shows the sample after excitation. Preferably those proteins are excited (colored orange) with their excitation moment (thick arrow) aligned with the excitation moment (vertical). As only these excited protein undergo structural changes, the sample is anisotropic after the excitation event.

With the sample being anisotropic the scattering pattern is not necessarily radially symmetric. The excess intensity can now be defined either in terms of the scattering angle $\alpha$ and the azimuthal angle $\beta$

\[
\Delta I_{\Delta t}(\alpha, \beta) = I_{t=t_0+\Delta t}(\alpha, \beta) - I_{t<t_0}(\alpha, \beta)
\]  \hspace{1cm} (2.6)

or analog to (2.5) with the transferred momentum $q$

\[
\Delta I_{\Delta t}(q) = I_{t=t_0+\Delta t}(q) - I_{t<t_0}(q).
\]  \hspace{1cm} (2.7)
Here, however, $\mathbf{q}$ is a vector defining both the scattering angle $\alpha$ as well as the azimuthal angle $\beta$. Note that the second term in (2.6) is in fact still isotropic thus the replacement of $q$ with vector $\mathbf{q}$ is not essential. The anisotropic scattering intensity can be presented as a two dimensional intensity map. In practice, instead, either horizontal respectively vertical cuts or azimuthal averages of the scattering pattern are reported [63, 64]. The anisotropy decays with the timescale of rotational diffusion of the protein.
3 Theory

In this chapter the theory for a calculation of anisotropic TR-WAXS pattern from MD simulations is presented. In the course, the main principles of MD simulations alongside with some relevant specific details are presented, followed by a derivation of the calculation of scattering pattern form atomic coordinates. In the third section the method of Park et al. [42] for the calculation of WAXS pattern is revised to account for ultra short X-ray pulse, and in the fourth section to account for anisotropic TR-WAXS experiments. In the fifth section the effect of rotational diffusion is examined and this chapter closes with a discussion on the information content of WAXS pattern.

3.1 Molecular dynamics

3.1.1 Basics

A quantum mechanical (QM) calculations of a molecular system would involve solving the time-dependent Schrödinger equation for all nuclei and electrons. In contrast, in molecular dynamic (MD) simulation atoms are threatened as classical particles following newtons equation of motion. This corresponds to three key approximations, (i) the Born-Oppenheimer approximation, (ii) a classical treatment of the nuclei and (iii) an empirical force field describing the interactions.

The dynamics of a molecule can be described with the time-depedned Schrödinger equation

$$\hat{H}\psi(R, r) = i\hbar \frac{\partial \psi(R, r)}{\partial t}$$

with the molecular hamiltonian

$$\hat{H}$$
3 Theory

\[ \hat{H} = -\sum_i \frac{\nabla_i^2}{2M_i} - \sum_i \frac{\nabla_i^2}{2} + \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{i<j} \frac{Z_i Z_j}{|\mathbf{R}_i - \mathbf{R}_j|} + \sum_i \frac{Z_i}{|\mathbf{R}_i - \mathbf{r}_i|} \] (3.2)

(in atomic units) simultaneously acting on the nuclei and the electron degrees of freedom. These are the momentum \( T_n = \frac{\nabla_i^2}{2M_i} \) of the nuclei and the electrons, respectively, and their positions \( \mathbf{R}_i \) and \( \mathbf{r}_i \). The idea of the Born-Oppenheimer approximation is to separate the fast electronic degrees of freedom from the comparatively slow nuclei motions. The different timescales of their respective motions arise from the huge difference in the electron and the nuclei masses. In the Born-Oppenheimer approximation, the total wave-function

\[ \Psi(\mathbf{R}, \mathbf{r}) = \psi(\mathbf{R}, \mathbf{r})\phi(\mathbf{R}) \quad (3.3) \]

can then be separated into decoupled electronic \( \psi(\mathbf{R}, \mathbf{r}) \) and nuclear components \( \phi(\mathbf{R}) \). The Schrödinger equation can then be solved in two steps. First, solving the time-independent Schrödinger equation

\[ \hat{H}_e \psi(\mathbf{r}; \mathbf{R}) = E_e(\mathbf{R})\psi(\mathbf{r}; \mathbf{R}) \quad (3.4) \]

for the electronic problem with the nuclei essential clamped at positions \( \mathbf{R} \). Here, the electronic Hamiltonian \( \hat{H}_e \) corresponds to (3.2) with contribution of the nuclei kinetic energy \( T_n \) being removed. The solutions of (3.4) forms a potential energy surface for the Schrödinger equation of the nuclei

\[ [T_n + E_e(\mathbf{R})]\phi(\mathbf{R}) = i\hbar \frac{\phi(\mathbf{R})}{\partial t} \]. \quad (3.5) \]

The large mass of the nuclei allow a classic treatment of their dynamics

\[ -\nabla_1 E_e(\mathbf{R}) = M_1 \frac{\partial^2 \mathbf{R}_1}{\partial t^2} = M_1 \frac{\partial \mathbf{V}_1}{\partial t} \] \quad (3.6)

\[ \mathbf{V}_1 = \frac{\partial \mathbf{R}_1}{\partial t} \] \quad (3.7)
3.1 Molecular dynamics

The latter is the second fundamental approximation for most molecular calculations, which usually holds well with limitations for the dynamics of hydrogens and at low temperatures. Using the previous two approximations, the limiting computational effort is in solving the electronic problem. The computational cost of most QM methods scale cubic or worse with the number of atoms, which makes them unfeasible for large biomolecules. Essential linear scaling methods are in development [86, 87]. These usually face the dilemma of trading between acceptably scaling pre-factors and accuracy of the used approximations.

For these reasons most MD simulations of biological systems rely on molecular mechanical (MM) force fields. These forcefields approximate the potential energy surface of the electronic problem \( E_e(R) \) (now denoted as \( U(R) \)) with a number of analytic functions. There is no limit to the creativity for the set of functions used. They typically include terms for bonds, angles, dihedrals, improper dihedrals, a Lennard-Jonas potential and a Coulomb term:

\[
U(R) = U_{\text{bon.}} + U_{\text{ang.}} + U_{\text{dih.}} + U_{\text{imp. dih.}} + U_{\text{LJ}} + U_{\text{coul}}
\]

\[
= \sum_{\text{bon.}} \frac{k_i}{2} (l_i - l_{i,0})^2 + \sum_{\text{ang.}} \frac{f_i}{2} (\varphi_i - \varphi_{i,0})^2 + \sum_{\text{dih.}} \frac{V_i}{2} [1 + \cos (n\phi_i - \phi_{i,0})] + \sum_{\text{imp. dih.}} \kappa_i (\xi_i - \xi_{i,0})^2
\]

\[
+ \sum_{\text{pairs } i,j} \left[ 4\epsilon_{ij} \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi \varepsilon_0 r_{ij}}
\]

The bond stretching, bond angle and improper dihedral (this term is to describe out-of-plane vibrations) are modeled here as harmonical potentials. The dihedral term is modeled with a periodic potential. Other functional forms for the dihedrals are common. The last sum over all atom pairs includes the Lennard-Jonas potential, which maps effects of the Pauli repulsion as well as the attractive Van-der-Waals force, and the classic Coulomb interaction. In the CHARMM [1] forcefield the combination rules for calculating the Lennard-Jones interaction between two different atoms are \( \sigma_{ij} = \frac{1}{2}(\sigma_{ii} + \sigma_{jj}) \) and \( \epsilon_{ij} = \sqrt{\epsilon_{ii} \epsilon_{jj}} \) with \( \sigma_{ii} \) and \( \epsilon_{ii} \) being the parameters.
3 Theory

for the interaction of identical atoms. Note that the Lennart-Jones potential is not calculated for first and second neighbors and might be reduced for third neighbors. While the bonded terms scale approximately linear with the number of atoms in the system, the non-bonded would scale quadratically with the number of atoms. The Lennart-Jones potential converges with a power of six to zero and can therefore rather trouble-free being cut-off at long distances, effectively obtaining linear scaling, when using neighborlists [88]. The Coulomb interaction on the other hand is not guaranteed to converge. For a good approximation of the Coulomb interaction without quadratic scaling with the number of charges, the particle-mesh Ewald (PME) method is used in this work.

As suggested by its name the PME method is based on the Ewald summation [89]. The underlining idea is to split the Coulomb potential \( U_C(q) = U_{C,\text{short}}(q) + U_{C,\text{long}}(q) \) into a short-range and a long-range part (for instance by multiplying the individual terms in the summation with a Gaussian respectively an Gaussian error function). The short-range part converges fast in real space and can therefor effectively and accurate be calculated with a cut-off scheme. The long-range part does converge fast in reciprocal space. In the particle-mesh Ewald method, the charges are assigned to points on a 3D grid for the long-range part and the Fourier transform of this grid points is calculated. This allows fast computation using optimized 3D FFT algorithms and a scaling of \( N \log(N) \) with the number of atoms [88].

3.1.2 Leap frog integrator

For the integration of the equations of motion (3.6) and (3.7), GROMACS [88] makes use of the leap-frog algorithm. The key advantages of this integrator (beside of its simplicity) are that it is time-reversible and energy-conserving. The latter is important as a drift in the total energy is unphysical. Velocities are calculated

\[
V_I(t + \frac{1}{2}\Delta t) = V_I(t - \frac{1}{2}\Delta t) + \frac{\Delta t}{m} \nabla_I U(R)
\]

from the slope of the potential energy surface as well as from previous velocities. Positions are then calculated

\[
X_I(t) = X_I(t - \Delta t) + V_I(t - \frac{1}{2}\Delta t)
\]

form previous positions and velocities. The leap-frog integration is reasonable accurate if the time-step \( \Delta t = \frac{1}{8} T \) used corresponds to a fifth of the fastest oscillation.
3.1 Molecular dynamics

The fastest motions are bond vibrations and the second largest the hydrogen movements. To avoid these fast motions bond lengths can be constrained with the LINCS algorithm\[90\]. Removing the harmonic oscillation can be considered as an improvement as the of the bond-stretching vibrations are to be considered to be in their QM ground state\[88\]. The next fastest movements are the bond-angle movements of the hydrogen atoms. Using virtual sites where the hydrogen atom positions are calculated from the positions of neighboring heavy atoms increases the fastest oscillation period to 20 fs and thus allows time-steps of 4 fs \[88\].

3.1.3 Pressure and temperature coupling

A molecular dynamic simulation following the above equations yield a NVE (constant particle number, constant volume, constant energy) ensemble. In reality the system under investigation such as a protein within a solvent, is coupled to its surrounding. Thus a NPT (constant particle number, constant pressure, constant temperature) ensemble is usually the physical relevant. The exchange with the surrounding is implemented with coupling terms, which in the case of the the pressure acting on the box volume and in case of the temperature acting on the particle velocities. In this work, the velocity-rescaling thermostat was used. The kinetic energy $K$ is modified

$$dK = (K_0 - K) \frac{dt}{\tau_T} + 2 \sqrt{\frac{KK_0}{N_f}} \frac{dW}{\sqrt{\tau_T}}$$ \ (3.10)

in two ways\[88\]. The first term in (3.10) lets the kinetic energy within the system exponentially decay towards some reference value $K_0$ with time constant $\tau_T$. This first term corresponds to the to the Berendsen thermostat, while the second term is a stochastic term, with the number of degrees of freedom $N_f$ and $dW$ a Wiener process (a random variable) ensuring a correct canonical ensemble.

Likewise barostats for pressure control have been developed. In this work the Berendsen as well as the Parrinello-Rahman pressure coupling is used. The Berendsen barostat introduce a first order decay of the pressure like the first term in (3.10) does for the temperature. In this work the pressure is controlled by scaling all coordinates and correspondingly the box vectors by a factor

$$\mu = 1 - \frac{\Delta t}{3\tau_p} \beta(P_0 - P(t))$$
3 Theory

depending on the compressibility $\beta$ (typically $4.6 \cdot 10^{-10} Pa^{-1}$), the desired pressure $P_0$ and the actually pressure $P(t)$. The time-constant $\tau_p$ controls the strength of the coupling. The Berendsen scaling does not generate a true NPT ensemble, therefore in this work the Parrinello-Rahman pressure coupling is used as well. For the mathematical details I like to refer to [88, 91], the underlying principal is however to introduce a additional equation of motion of second order for the box vectors. An additional friction term controlled by the box vector dynamic is added to the equation of motion of the atoms. Parrinello-Rahman forces the pressure of the system to oscillate around the reference pressure and one obtain a correct NPT ensemble. If the system is far from equilibrium the Parrinello-Rahman coupling my result in large oscillations of the systems pressure, which may not be physically correct.

3.1.4 Replica exchange molecular dynamic

For MD simulations it is important that the system has visited a sufficiently large fraction of the accessible phase space. If the system inhibits large energy barriers the sampling within reachable simulation time may not be good enough. It is also not always possible to scale up parallelisation efficiently to speed up sampling. In replica exchange molecular dynamic (REMD) a number of replicas of the same system are simulated at the same time. The replicas may differ in different respects but a common approach is to simulate the replicas over a range of temperatures. At fixed time intervals the atom positions of neighboring replicas will be exchanged with a probability according to the metropolis algorithm. For each exchange attempt the difference in potential energy $\Delta U_i = U_i - U_{i+1}$ alternating for odd and even $i$ is calculated. An exchange probability of

$$p_{i,i+1} = \min[1, e^{-\frac{\Delta U_i}{k_B T}}]$$

for the exchange between replica $i$ and $i + 1$ will according to the metropolis algorithm assure the combined trajectory of each temperature to obey the correct canonical ensemble. The idea is that replicas of higher energy will be able to overcome energy barriers much faster then the lowest temperature, which is in general the one of interest. For sufficient mixing, the exchange probability should be considerably high and thus the temperature differences has to be sufficient low. REMD can speed up sampling, the trajectories obtained (for each temperature) are, however, not continues anymore, limiting the possibilities to study dynamics.
3.2 Molecular scattering

3.2.1 The scattering amplitude

MD simulation provides a trajectory of atomic coordinates. This section presents a derivation of scattering of an arbitrary target [92–94] and more specifically one defined by atomic coordinates. Although the derivation is general, the motivation is to describe the scattering of an X-ray beam with soft matter.

One may understand the process of scattering as the acceleration of an electron by an electromagnetic wave and a succeeding radiation as inherent for accelerated charged particles. In the following, it is abstracted from the underlying mechanism and the target will be described as an external potential $V(r)$ within the Schrödinger equation, excluding any mutual interaction between beam and target. The latter restriction does not allow the description of anomalous diffraction experiments, but it should be sufficient for our purposes. The Schrödinger equation is given by

$$\left(\nabla^2 + k^2\right)\psi(r) = \frac{2m}{\hbar^2} V(r)\psi(r), \text{ with } k^2 = \frac{2mE}{\hbar}. \quad (3.11)$$

The solution of this inhomogeneous differential equation is given as the sum of the solution of the corresponding homogeneous equation and a particular one for the inhomogeneous one. The homogenous solution is a plane wave $\phi_0(r) = A_0 \cdot e^{ik_0r}$ with $k_0 = \frac{2mE}{\hbar}$. For an inhomogenous solution it is useful to rewrite the the Schrödinger equation into its so called integral formulation

$$\psi(r) = \phi_0(r) - \frac{2m}{\hbar^2} \int \frac{e^{ik|r'-r|}}{|r - r'|} V(r')\psi(r')dr'. \quad (3.12)$$

The first term in (3.12) is the homogenous solutions and represents in a scattering experiment the incident wave. The second term consist of an integral over spherical waves originated at the target. Given that the scattering amplitude is small in comparison to the incident plane wave, i.e. given that the second term in (3.12) is much smaller then the first term in absolute value, the wave function within the integral can be replaced by the incoming wave function $\phi_0(r)$. In this so-called first Born approximation the total wave function is given by

$$\psi(r) = \phi_0(r) - \frac{2m}{\hbar^2} \int \frac{e^{ik|r'-r|}}{|r - r'|} V(r')\phi_0(r')dr' = \phi_0(r) + \phi_{\text{scat}}(r) \quad (3.13)$$

In a scattering experiment the first Born approximation holds if the scattering is
3 Theory

weak and therefore secondary scattering can be ignored.

For points far away from the target, e.g. if the size of the target is much smaller than the distance of the detector to the target, a further simplification can be done. The integral has then only non zero contributions for $r \gg r'$ (assuming the origin of the coordinate system at the target). It follows that

$$k \cdot |r - r'| = k \cdot \sqrt{r'^2 - 2rr' + r'^2} \approx k \cdot r \cdot \sqrt{1 - \frac{2rr'}{r'^2}} \approx k \cdot r \cdot (1 - \frac{rr'}{r'^2}) = kr - kr'$$

and

$$\frac{1}{|r - r'|} \approx \frac{1}{r \cdot (1 - \frac{rr'}{r'^2})} \approx \frac{1}{r}.$$  

Note that $k$ is the scattered wave vector in direction of $r$. Using the previous considerations, the scattered wave function (3.13) becomes

$$\psi(r) = -\frac{2\mu A_0 e^{ikr}}{\hbar^2} r \int e^{i(k-k_0)r'} V(r')dr'$$

which can be simplified to

$$\psi(r) = -\frac{2\mu A_0 e^{ikr}}{\hbar^2} \int e^{iqr'} V(r')dr' = \frac{e^{ikr}}{r} f(q)$$

introducing the momentum transfer $q = k - k_0$. The integral and prefactors can be condensed into the scattering amplitude $f(q)$ depending on the momentum transfer $q$. Each point on the detector corresponds to a specific momentum transfer according to (2.1) and the particle/photon flux there is the experimental measure. The flux at distance $r$ from the target is given by the wave function as

$$j(q) = \frac{\hbar}{m} \text{Im} [\phi_{\text{scat}}(r)^* \nabla \phi_{\text{scat}}(r)] = \frac{A_0 \cdot \hbar \cdot k}{m \cdot r^2} |f(q)|^2.$$  

Here we included only the scattering wave function (the second term of (3.14)), as the remainder of the incident plane wave at the focal point of the detector is in general not of interest and neither measured. The momentum transfer can be described by its absolute value $q$ and the corresponding solid angle $\Omega$. The flux of incoming particles/photons at a given solid angle element $d\Omega$ is given by

$$dJ = \frac{\hat{e}_\Omega \cdot \hat{n}}{r^2} dA = J_0 |f(k, \Omega)|^2 d\Omega.$$  

22
with the solid angle element $d\Omega$ corresponding to an area $\frac{dA}{r^2}$. The intensity per solid angle element measured by the detector is proportional to the flux and we thus obtain

$$I(q) \propto |f(q)|^2$$  \hspace{1cm} (3.16)

with a scattering amplitude

$$f(q) = \int e^{iqr'}V(r')dr'$$

being the Fourier transform of the target potential $V(r)$. In the derivation a classical target was assumed. X-ray primary scatter electrons and irrespectively of the detailed scattering process it seem to be a reasonable approximation to assume the target potential to be proportional to the electron density. This assumptions break down if excitation of the target has to be considered e.g. if the wave length of the X-ray beam corresponds to accessible energy levels within the target. This case will not be discussed here. It follows that the scattering amplitude is given by the Fourier transform of the targets electron density

$$f(q) = \int e^{iqr'}\rho(r')dr'.$$  \hspace{1cm} (3.17)

The scattering intensity is therefor proportional to the absolute squared of the Fourier transform of the targets electron density

$$I(q) \propto |\tilde{\rho}(q)|^2.$$  \hspace{1cm} (3.18)

### 3.2.2 Cromer-Mann parameter

For the calculation of scattering pattern from MD simulations the electron density is not directly available and is instead to be expressed in terms of atomic positions. In a first step the total electron density $\rho(r)$ of a system is formally split into individual contributions $f_i(r)$ of each atom

$$\rho(r) = \sum_{i=1}^{N} f_j(r - r_j).$$  \hspace{1cm} (3.19)
The electron density of the individual atoms $f_i(r)$ is described relatively to the position $r_i$ of the atomic core. Note that the electron density of neighboring atoms generally will overlap and the total density at a given point thus has to be calculated from their sum. Moreover, there is no obvious way to split the total electron density in a molecule into contributions of individual atoms. However, the fragmentation can be done completely arbitrary. The only requirement is to approximately obtain representations of the individual atomic electron densities, which can be generalized for corresponding atoms in different molecules.

It will turn out to be convenient to rewrite the summands using a convolution with a delta-function centered on the position of the corresponding atomic core.

\[
\sum_{i=1}^{N} f_j(r - r_j) = \sum_{i=1}^{N} \int_{-\infty}^{\infty} f_i(r') \delta(r' - (r - r_i)) \, dr' = \sum_{i=1}^{N} f_i(r) * \delta(r - r_i) \quad (3.20)
\]

When taking the Fourier image

\[
\tilde{\rho}(q) = \int e^{iqr} \sum_{i=1}^{N} f_i(r) * \delta(r - r_i) = \sum_{i=1}^{N} \tilde{f}_i(q) e^{-iqr_i} \quad (3.21)
\]

the convolution turns into a multiplication. The Fourier transform of the atomic electron density is described by a atomic form factor $\tilde{f}_i(q)$, which is independent from the atom position. The latter only enters as an additional phase factor.

Separating the internal electronic structure and the position of the atom is the main advantage of this description. To benefit from this separation commonly two assumptions are used. First, the atomic form factor are only calculated for each element and charge once. The detailed chemical structure within the protein is not taken into account. Second, the electronic structures of the atoms are approximated to be isotropic, thus the atomic from factor depends on the absolute momentum transfer $q$ only.

For parametrization the electronic density of each atom is usually approximated by four spherical Gaussians and a single delta function, all centered on the atomic core. The Fourier transform is then given by four Gaussians and a constant

\[
\tilde{f}_A(q) = \sum_{i=1}^{4} a_i^A \cdot e^{-(b_i^A q)^2} + c^A \quad (3.22)
\]

leading to 9 fitting parameters, the so called Cromer-Mann parameter.

Cromer-Mann parameter calculated from Hartree-Fock calculations of individual
3.3 Isotropic solution scattering

atoms are tabulated in Ref. [95]. For force fields implemented in GROMACS these parameters are assigned to each atom within each amino acid, corresponding to the respective element and charge. A correction to the form factors of the water molecule $\tilde{f}_{H/O}(q) = \tilde{f}_{H/O}(q) \left(1 + \alpha_{H/O} e^{-q/2\delta}\right)$ with $\delta = 2.2\text{Å}$, $\alpha_O = 0.12$ and $\alpha_H = -0.48$ was derived from experimental data[96] to improve on the actual water molecule electron density in comparison to the one of the individual atoms. Inserting (3.21) into (3.17) and the latter into (3.16) we obtain the final result

$$I(q) \propto |\tilde{\rho}(q)|^2 = \left| \sum_i \tilde{f}_i(q) e^{-iqr_i} \right|^2 = \sum_i \sum_j \tilde{f}_i(q) \tilde{f}_j(q) e^{iq(r_i-r_j)}$$

(3.23)

for the numeric calculation of the scattering intensity of a single molecule.

From previous considerations it becomes apparent that the atomic form factors for a given system has to be calculated once for each atom type and each absolute value of $q$ only. The remaining computational challenge is then to calculate the phase factor within (3.23). For the Fourier transformation of $\tilde{\rho}(q)$ and thus the calculation of the scattering intensity for a single $q$-vector the numerical evaluation of $N$ sine and cosine are needed. In this section, the scattering of a single object was discussed. In WAXS experiments an ensemble of a large number of proteins in solution has to be considered.

3.3 Isotropic solution scattering

From MD simulation one usually obtains a trajectory of atomic positions of a single protein in solution. In a real experiment a large number of proteins are probed simultaneously. In the following section, it will be shown how under the assumption of neglectable protein-protein interactions the excess intensity as defined in (2.4) can be calculated from two trajectories. One trajectory of the protein in solution and a second trajectory of solvent only. The derivation will follow the work of Park et al. [42], with some derivation in the formulation to allow for a generalization to TR-WAXS.

Following (2.4), the excess intensity is given by the difference

$$\Delta I(q) = I_{\text{sol}}(q) - I_{\text{buff}}(q) = |\tilde{A}(q)|^2 - |\tilde{B}(q)|^2$$

(3.24)
3 Theory

Figure 3.1: I) An ensemble of proteins in solution is shown, corresponding to the term $A(r)$. The electron density is split into contributions of the proteins ($A_1$, $A_2$, $A_3$, ...) and of the bulk solvent $A_B$. II) The center of mass of all proteins ($A_I$) are mapped to the origin. Copies of the bulk term ($A_{BI}$) are mapped correspondingly. III: Molecules ($A_{Ω}$) of the same orientation are considered together and the corresponding bulk terms ($A_{ΩB}$) alongside.

of the scattering of the cuvet filled with the solution of the protein and the second is the cuvet filled with the buffer only. Using (3.18), the difference can be expressed in terms of the Fourier transform $\hat{A}(q)$ and $\hat{B}(q)$ of the electron density of both targets respectively.

This equation differs from the one presented by Park et. al[42] as here no time average of the corresponding terms is used. In time-independent WAXS not only a large number of proteins are probed at the same time, but in addition data is usually collected over a timespan in the range of seconds to minutes. In TR-WAXS experiments, however, typical exposure times are in the range of picoseconds [42]. Within these timescales, the autocorrelations of the solvation shell and even more the one of the protein degrees of freedom are still significant and, thus, using a time average would be misleading.

Instead it will turn out in the following that - although somewhat more cumbersome - averaging over the multiple proteins in the probe yields mathematically the same result. This is not surprising as a large number of instances of the same system corresponds to an average over time.

Note that to be precise in (3.24) a partial average over the exposure time needs to be added. As the final result is identical to the one obtained by Park et al. the temporal average is redundant and omitted here.

As the depicted in Fig.3.1.1 the electron density $A(r)$ of the cuvet filled with protein solution can be partitioned in space and written as as the sum of the electron density of the individual proteins $A_i$ including their vicinity to account for the
3.3 Isotropic solution scattering

proteins solvation layers and the rest of the system $A_B$

$$A(r) = A_B(r) + \sum_{i \in I} A_i(r - r_i).$$  \hspace{1cm} (3.25)

The rest of the system consist of the curvet and the bulk solvent, hence the subscript calligraphic $B$. The corresponding electron density has holes at the position of the protein, reminiscent of a swiss cheese, and is not to be confused with the pure bulk term $B$. Calligraphic $I = \{1, 2, ..., N\}$ denotes the set of index of all $N$ proteins measured. The electron density of the individual molecules are defined in respect to their corresponding center of mass $r_i$ for later convenience. There is in practice a difference how the included vicinity of the molecule is defined by Park et al. [42] and by Chen and Hub[43]. In the version used by Park et al. atoms within a shell of constant thickness around the molecule are considered. In Chen and Hub version an envelope relative to a reference structure is defined prior, and all atoms within this fixed envelope are considered in $A_i(r - r_i)$ (see Fig. 1.1). In any case, the envelope has to be taken large enough such that the solvent at the boundary can be considered as bulk-like. The hole in the bulk term $A_B(r)$ corresponding to molecule $A_i$ is referred to as its excluded volume.

The electron density of the curvet filled with the buffer can be partitioned and written in a similar way,

$$B(r) = B_B(r) + \sum_{i \in I} B_i(r - r_i)$$  \hspace{1cm} (3.26)

such that each $B_i(r - r_i)$ matches exactly the volume of the corresponding protein term in (3.25) defined by exactly the same envelope. Correspondingly $A_B(r)$ and $B_B(r)$ are covering the curvet and solvent molecules within the same space, with holes at the position of the proteins in $A(r)$. To simplify notation, the atoms within an envelope are referred to as a droplet irrespectively if taken from the solution $A(r)$ or the solvent only term $B(r)$.

Focussing on the positive part $|\tilde{A}(q)|^2$ in (3.24) first, the negative part may then be derived analogously. The positive part is expressed as the Fourier transform of
3 Theory

the electron density

\[ |\tilde{A}(q)|^2 = \int d\mathbf{r} d\mathbf{r}' \left\{ e^{-i\mathbf{q} \cdot (\mathbf{r} - \mathbf{r}')} A_B(\mathbf{r}) A_B(\mathbf{r}') + \sum_{i \in I} \sum_{j \in I} A_i(\mathbf{r} - \mathbf{r}_i) A_j(\mathbf{r}' - \mathbf{r}_j) \right\} \]

\[ + \sum_{i \in I} \left[ A_i(\mathbf{r} - \mathbf{r}_i) A_B(\mathbf{r}') + A_B(\mathbf{r}) A_i(\mathbf{r}' - \mathbf{r}_i) \right] \]  

(3.28)

With \( \tilde{A}'_i(q) \) being the Fourier transform of \( A_i(\mathbf{r}' - \mathbf{r}_i) \) and \( \tilde{A}_B(q) \) denoting the Fourier transform of \( \tilde{A}_B(\mathbf{r}) \), the previous equation (3.27) can be shortened to

\[ |\tilde{A}(q)|^2 = \tilde{A}_B(q) \tilde{A}_B^*(q) + \sum_{i \in I} \sum_{j \in I} \tilde{A}'_i(q) \tilde{A}'^*_j(q) + \sum_{i \in I} \left[ \tilde{A}'_i(q) \tilde{A}_B^*(q) + \tilde{A}_B(q) \tilde{A}'^*_i(q) \right] \]

(3.29)

with the asterisk denoting the complex conjugate and the prime indicating that the related electron densities are spread in space at the positions of the molecules.

The trajectory of a simulation contains one molecule only, which center of mass is translated to the origin before applying the envelope. In (3.27), the molecules’ electron density is already described in respect of their center of mass. For the Fourier transform

\[ \tilde{A}'_i(q) = \int d\mathbf{r} e^{-iq \cdot \mathbf{r}} A_i(\mathbf{r}' - \mathbf{r}_i) = \left[ \int d\mathbf{r} e^{-iq \cdot \mathbf{r}} A_i(\mathbf{r}') \right] \cdot e^{iq \cdot \mathbf{r}_i} = \tilde{A}_i(q) e^{iq \cdot \mathbf{r}_i}, \]  

(3.30)

it was shown in the previous section that center of mass position \( \mathbf{r}_i \) enters as a phase factor only. Beside this phase factor all proteins are now mapped to the same point in space as indicated in Fig. 3.1.II.

For the last sum in (3.27) it is convenient to shift the electron density of the bulk \( A_B(\mathbf{r}) \) likewise. However it needs to be shifted for each molecule differently. Therefore, for each droplet \( A_i \), a copy of the bulk density \( A_B(\mathbf{r}' - \mathbf{r}_i) = A_B(\mathbf{r}') \) is considered, such that the excluded volume corresponding to the molecule \( A_i \) within \( A_B(\mathbf{r}) \) is situated at the origin.

Analogous to (3.30), the translation manifest itself as a phase factor in the Fourier transform

\[ \tilde{A}_B(q) = \int d\mathbf{r} e^{-iq \cdot \mathbf{r}} A_B(\mathbf{r}') = \int d\mathbf{r} e^{-iq \cdot \mathbf{r}} A_B(\mathbf{r}' - \mathbf{r}_i) = \tilde{A}_B(q) e^{iq \cdot \mathbf{r}_i}. \]  

(3.31)
3.3 Isotropic solution scattering

It is important to note that all $\tilde{A}_{B_i}(q)$ originate from the same electron density, however their phases differ by the phase factor $e^{-iqr}$, which correspond to a translation. In Fig. 3.1, it is illustrated by blue and red dots that the vicinity of the different droplets (I) is now correspondingly mapped to the origin (II).

After plugging (3.30) and (3.31) into (3.29), we get

$$|\tilde{A}(q)|^2 = \tilde{A}_E(q)\tilde{A}_B^*(q) + \sum_{i\in I} \sum_{j\in I} \tilde{A}_i(q)\tilde{A}_j^*(q)e^{iq\Delta r_{ij}}$$

$$+ \sum_{i\in I} \left[ \tilde{A}_i(q)\tilde{A}_i^*(q) + \tilde{A}_i(q)\tilde{A}_{B_i}(q) + \tilde{A}_{B_i}(q)\tilde{A}_i^*(q) \right],$$

where the phases factors cancel out in the diagonal terms, corresponding to the same droplets. However, for the off-diagonal terms a phase factor remains corresponding to the difference vector $\Delta r_{ij} = r_i - r_j$.

The double sum in (3.32) can be related to a double average

$$\langle \tilde{A}_I(q) \langle \tilde{A}_I^*(q)e^{iq\Delta r} \rangle_{(\Delta r \neq 0)} \rangle.$$

The restriction to nonzero differences $\Delta r \neq 0$ is a consequence of the exclusive use of off-diagonal terms in (3.32). It is reasonable to assume that there is no correlation between the protein degrees of freedom and their relative position. Thus the average over the set of droplets $I$ factorizes into three individual factors

$$\langle \tilde{A}_I(q) \langle \tilde{A}_I^*(q)e^{iq\Delta r} \rangle_{(\Delta r \neq 0)} \rangle \approx \langle \tilde{A}_I(q) \rangle \langle \tilde{A}_I^*(q) \rangle \langle e^{iq\Delta r} \rangle_{(\Delta r \neq 0)}.$$ (3.33)

The phase factor in (3.33) is governed by the protein positions. Thus, unspecific protein interactions would play a role here. Assuming the absence of protein-protein interactions and low concentrations, there is no preferred inter-protein distance. Therefore, the last term

$$\langle e^{iq\Delta r} \rangle_{(\Delta r \neq 0)} \approx 0$$ (3.34)

in (3.33) and thus all off-diagonal terms in (3.32) are zero.

Building on the previous arguments we approximate
\[
\frac{|\tilde{A}(q)|^2}{N} \approx \frac{|\tilde{A}_B(q)|^2}{N} + \langle \tilde{A}_T(q)\tilde{A}_T^*(q) \rangle + \langle \tilde{A}_T(q)\tilde{A}_{BT}^*(q) \rangle + \langle \tilde{A}_{BT}(q)\tilde{A}_T^*(q) \rangle. \quad (3.35)
\]

The derivation presented can analogously be applied to the buffer term in (3.24) yielding

\[
\frac{|\tilde{B}(q)|^2}{N} \approx \frac{|\tilde{B}_B(q)|^2}{N} + \langle \tilde{B}_T(q)\tilde{B}_T^*(q) \rangle + \langle \tilde{B}_T(q)\tilde{B}_{BT}^*(q) \rangle + \langle \tilde{B}_{BT}(q)\tilde{B}_T^*(q) \rangle. \quad (3.36)
\]

Intuitively, it is clear that the scattering intensity of the curvet filled with solvent only \(|\tilde{B}(q)|^2\) does not depend on the particular realization of water molecule positions in the curvet. Mathematically this becomes apparent if one realizes that the scattering intensity in (3.23) only depends on the inter-atomic distances, but not on their absolute positions. It was shown that the scattering intensity can be calculated from pair distribution functions [46]. Due to the large number of solvent molecules in \(|\tilde{B}(q)|^2\), their pair distribution function converges already for a single realization and thus averaging over different realization of the solvent (such an average will be indicated by a subscript s) does not change the result

\[
|\tilde{B}(q)|^2 \approx \langle |\tilde{B}(q)|^2 \rangle_s. \quad (3.37)
\]

In conclusion, (3.37) shows that the left hand side of (3.36) is independent of the realization of the solvent and, thus, the right hand side must be independent of the realization, too. This in particular also applies for the the first term

\[
|\tilde{B}_B(q)|^2 \approx \langle |\tilde{B}_B(q)|^2 \rangle_s, \quad (3.38)
\]

as the other terms are readily averaged over different realization. There is no qualitative difference between \(|\tilde{A}_B(q)|^2\) and \(|\tilde{B}_B(q)|^2\), because the solvent in both cases can be considered as bulk-like, and therefore the same arguments applies analogously to the solution term

\[
|\tilde{A}_B(q)|^2 \approx \langle |\tilde{A}_B(q)|^2 \rangle_s. \quad (3.39)
\]

In both terms (3.39) and (3.38), the contributions of the curvet and thus also
the global occupied volume is the same. Also, the excluded volume of all molecules respectively their ghost images is identical. Because neither of them depends on the particular realization of the solvent they are identical

\[ |\tilde{B}_S(q)|^2 \approx \langle |\tilde{B}_S(q)|^2 \rangle_w = \langle |\tilde{A}_B(q)|^2 \rangle_w \approx |\tilde{A}_B(q)|^2. \] (3.40)

Note that with similar arguments used to dismiss the off-diagonal terms (3.33), the identity (3.40) can be generalized to cases, in which the molecules excluded volumes in \( |\tilde{A}_B(q)|^2 \) and \( |\tilde{B}_B(q)|^2 \) correspond in their shape and orientation, but not in their position. This will be the case for TR-WAXS presented in the next section.

With (3.35), (3.36) and (3.40) the difference intensity

\[ \Delta I(q) \propto |\tilde{A}(q)|^2 - |\tilde{B}(q)|^2 = \int dr dr' e^{-i\mathbf{q} \cdot (\mathbf{r} - \mathbf{r}')} D_I(\mathbf{r}, \mathbf{r}') \] (3.41)

can be written with

\[
D_I(\mathbf{r}, \mathbf{r}') = \langle A_I(\mathbf{r})A_I(\mathbf{r}') \rangle - \langle B_I(\mathbf{r})B_I(\mathbf{r}') \rangle + \langle A_I(\mathbf{r})A_{B_I}(\mathbf{r}') \rangle + \langle A_{B_I}(\mathbf{r})A_I(\mathbf{r}') \rangle
- \langle B_I(\mathbf{r})B_{B_I}(\mathbf{r}') \rangle - \langle B_{B_I}(\mathbf{r})B_I(\mathbf{r}') \rangle. \] (3.42)

Aiming for the relative scattering intensity up to a scaling factor, the pre-factor \( N \), corresponding to the number of proteins was omitted.

It is necessary to separate the rotational degrees of freedom of the protein - defined for instance by finding the rotational matrix which minimizes the RMSD to the reference structure of the envelope - and the remaining degrees of freedom of the molecule. As shown in Fig. 3.1.III the ensemble average is over molecules of identical orientation \( \Omega \) only. The average over the different orientations is explicitly formulated as an integration

\[ \Delta I(q) \propto \frac{1}{4\pi} \int d\Omega \int dr dr' e^{-i\mathbf{q} \cdot (\mathbf{r} - \mathbf{r}')} D_\Omega(\mathbf{r}, \mathbf{r}'). \] (3.43)

The difference between \( D_\Omega(\mathbf{r}, \mathbf{r}') \) and (3.42) is that in the latter it is averaged over the subset of molecules only, whose orientation can be described by the Euler angle \( \Omega \). The envelope and thus the space occupied by a droplet is defined with respect to the protein’s center of mass and its orientation. Thus, all \( A_\Omega(\mathbf{r}) \) occupy exactly the same space and consequently their excluded volume is pairwise congruent.
This allows to write the terms in (3.42) including the droplet as well as the bulk solvent, in terms of the product of their individual averages and an electron-density correlation $\alpha_{\text{BO}}(\mathbf{r}, \mathbf{r}')$:

$$
\langle A_\Omega(\mathbf{r})A_{\text{BO}}(\mathbf{r}') \rangle = \langle A_\Omega(\mathbf{r}) \rangle \langle A_{\text{BO}}(\mathbf{r}') \rangle + \alpha_{\text{BO}}(\mathbf{r}, \mathbf{r}')
$$

(3.44)

The correlation term $\alpha_{\text{BO}}(\mathbf{r}, \mathbf{r}')$ contains no autocorrelation, because the excluded volumes at the origin of $A_{\text{BO}}(\mathbf{r}')$ are pairwise congruent and exactly matching with the droplets $A_\Omega(\mathbf{r})$. The correlation term $\alpha_{\text{BO}}(\mathbf{r}, \mathbf{r}')$ is only significant within the correlation length of the solvent and therefore only in the boundary area of the envelope. The same reasoning applies to the bulk terms:

$$
\langle B_\Omega(\mathbf{r})B_{\text{BO}}(\mathbf{r}') \rangle = \langle B_\Omega(\mathbf{r}) \rangle \langle B_{\text{BO}}(\mathbf{r}') \rangle + \beta_{\text{BO}}(\mathbf{r}, \mathbf{r}')
$$

(3.45)

Under the assumption that the size of the envelope is tuned such that the solvent at its surface can be considered as bulk-like, there should be no contributions from the molecules solvation layers within $\alpha_{\text{BO}}(\mathbf{r}, \mathbf{r}')$ and it can be assumed that

$$
\alpha_{\text{BO}}(\mathbf{r}, \mathbf{r}') \approx \beta_{\text{BO}}(\mathbf{r}, \mathbf{r}').
$$

We thus obtain for the difference term

$$
D_\Omega(\mathbf{r}, \mathbf{r}') = \langle A_\Omega(\mathbf{r})A_\Omega(\mathbf{r}') \rangle - \langle B_\Omega(\mathbf{r})B_\Omega(\mathbf{r}') \rangle + \langle A_\Omega(\mathbf{r}) \rangle \langle A_{\text{BO}}(\mathbf{r}') \rangle + \langle A_{\text{BO}}(\mathbf{r}) \rangle \langle A_\Omega(\mathbf{r}') \rangle \\
- \langle B_\Omega(\mathbf{r}) \rangle \langle B_{\text{BO}}(\mathbf{r}') \rangle - \langle B_{\text{BO}}(\mathbf{r}) \rangle \langle B_\Omega(\mathbf{r}') \rangle.
$$

(3.46)

Although originating from the the same single bulk electron density $A_{\text{BO}}(\mathbf{r})$ the average $\langle A_{\text{BO}}(\mathbf{r}) \rangle$ is over translational copies. Under neglect of protein-protein interactions the latter can thus be approximated as homogenous

$$
\langle B_{\text{BO}}(\mathbf{r}) \rangle \approx \rho_s - \langle B_\Omega(\mathbf{r}) \rangle = \rho_s(1 - g_\Omega(\mathbf{r})),
$$

(3.47)

with the exception of the excluded volume, for length scales sufficient smaller then the curvet size. The length scales of the curvet are not relevant for the WAXS experiments as they correspond to momentum transfers which are too small to be measured.
In (3.47), the excluded volume is expressed in two different ways. Either as an average of the pure solvent droplet \( \langle B_\Omega(r) \rangle \) or in terms of a indicator function
\[
g(r) = \begin{cases} 
1 & \text{for } r \text{ within envelope} \\
0 & \text{for } r \text{ outside of envelope}
\end{cases}
\]
representing the shape of the envelope. The previous consideration applies for the bulk terms \( \langle A_{\Omega}(r) \rangle \) and \( \langle B_{\Omega}(r) \rangle \) of both, the solution and the pure solvent. Taking the Fourier transform the constant in (3.47) can be dismissed leading to
\[
\langle \tilde{A}_{\Omega}(q) \rangle = \langle \tilde{B}_{\Omega}(q) \rangle \approx - \langle \tilde{B}_{\Omega}(q) \rangle = - \rho_s \tilde{g}_\Omega(q), \text{ for } q \neq \bar{0}. \quad (3.48)
\]
Using the first representation as done by Park et al. [42] the difference electron density can now be written as
\[
\tilde{D}_\Omega(q) = \left| \tilde{A}_\Omega(q) \right|^2 - \left| \tilde{B}_\Omega(q) \right|^2 - \left| \tilde{A}_\Omega(r) \right| \langle \tilde{B}^*_\Omega(q) \rangle - \left| \tilde{B}_\Omega(q) \right| \langle \tilde{A}^*_\Omega(q) \rangle + \left| \tilde{B}_\Omega(q) \right| \langle \tilde{B}^*_\Omega(q) \rangle + \left| \tilde{B}_\Omega(q) \right| \langle \tilde{A}^*_\Omega(q) \rangle. \quad (3.49)
\]
In solution all molecular orientations are equally populated. In the experiment discussed so far there is no correlation between the internal degrees of freedom and the rotational degrees of freedom. Thus all \( \tilde{D}_\Omega(q) \) are equal up to a rotation \( \tilde{D}_\Omega(q) = \tilde{D}_0(\hat{R}_{\Omega\Omega} q) \). The rotational matrix \( \hat{R}_{\Omega\Omega} \) maps the orientation \( \Omega \) onto a single distinguished orientation \( 0 \). The set \( \hat{R}_{\Omega\Omega} q \) for all orientations corresponds to a sphere \( \{ \hat{R}_{\Omega\Omega} q, \forall \Omega \} = \{ q, \forall |q| = q \} = S(q) \) of radius \( q \).
Correspondingly the integrations over protein rotations \( \Omega \) can be replaced
\[
\int d\Omega \tilde{D}_\Omega(q) = \int d\Omega \tilde{D}_0(\hat{R}_{\Omega\Omega}(q)) = \int dq_\Omega \tilde{D}_0(q_\Omega)
\]
by an integration over a sphere of \( q \) vectors with radius \( q \). It becomes apparent that due to the spherical average the isotropic scattering pattern
\[
\Delta I(q) \propto \frac{1}{4\pi} \int_{|q| = q} dq_\Omega \tilde{D}_0(q_\Omega)
\]
does not depend on the vector \( q \) but solely on its absolute value. With rewriting the difference electron density (3.49) into a more compact form
3 Theory

\[
\tilde{D}_0(q) = |\langle \tilde{A}_0(q) \rangle - \langle \tilde{B}_0(q) \rangle|^2 + \left[ \langle |\tilde{A}_0(r)|^2 \rangle - |\langle \tilde{A}_0(r) \rangle|^2 \right] \\
- \left[ \langle |\tilde{B}_0(r)|^2 \rangle - |\langle \tilde{B}_0(r) \rangle|^2 \right]
\]  \hspace{1cm} (3.52)

The final result of Park et al. [42] is obtained. However, the average in the current derivation is by theoretical means an average over the ensemble of molecules within the solution, while the average in [42] is of temporal nature.

This work aims for the calculation of WAXS pattern from MD simulations of a single molecule. Using the ergodic hypothesis, we can identify the average over the MD trajectory with the average over the ensemble of multiple molecules in (3.52). The averages in (3.52) are conducted over one molecule oriented in a specific direction and centered with their center of mass at the origin. Thus, before calculating the scattering factor for a given snapshot within the trajectory, a rotational and translational fit of the molecule is required. In the presented derivation additional effort was needed to cover ultrashort exposure times. These are of relevance for time-resolved solution scattering.

3.4 Anisotropic time-resolved solution scattering

In the isotropic case the scattering pattern of the solvent solely is subtracted from the pattern of the solution. Instead, for time-resolved solution scattering the contrast (2.4) between the scattering pattern of the protein before and the pattern of the protein after excitation is analyzed.

Thus, both \( A_\Omega \) and \( B_\Omega \) as in (3.24) correspond to the same solution of proteins after and before excitation, respectively. The difference spectra can be calculated analogous to the time-independent case,

\[
\Delta I(q, \varphi) \propto \frac{1}{4\pi} \int d\Omega \int d\mathbf{r} d\mathbf{r}' e^{-i\mathbf{q} \cdot (\mathbf{r} - \mathbf{r}')} \tilde{D}_\Omega(\mathbf{r}, \mathbf{r}')
\]  \hspace{1cm} (3.53)

\[
\tilde{D}_\Omega(\mathbf{r}, \mathbf{r}') = \langle A_\Omega(\mathbf{r}) A_\Omega(\mathbf{r}') \rangle - \langle B_\Omega(\mathbf{r}) B_\Omega(\mathbf{r}') \rangle - \left[ \langle A_\Omega(\mathbf{r}) \rangle - \langle B_\Omega(\mathbf{r}) \rangle \right] g_\Omega(\mathbf{r}')
\] \\
- \left[ \langle A_\Omega(\mathbf{r}') \rangle - \langle B_\Omega(\mathbf{r}') \rangle \right] g_\Omega^*(\mathbf{r}').
\]  \hspace{1cm} (3.54)
However, because $\langle B_{\Omega} \rangle$ does no longer correspond to pure solvent, it can no longer replace the bulk solvent terms $\langle \tilde{B}_{B\Omega}(\mathbf{q}) \rangle$. Instead, the analytic Fourier transform $\tilde{g}_{\Omega}(\mathbf{q})$ of the envelope is to be used, which is scaled by the solvent density $g_{\Omega}^{s}(r) = \tilde{g}_{\Omega}(\mathbf{q}) \cdot \rho_0$ for a shortened notion.

The ensemble of proteins before excitation is homogenous in a sense that it solely consist of proteins in their ground state. We can thus write the identity

$$\langle B_{\Omega}(\mathbf{r}) \rangle = \langle G_{\Omega}(\mathbf{r}) \rangle$$  \hspace{1cm} (3.55)

and

$$\langle B_{\Omega}(\mathbf{r}) B_{\Omega}(\mathbf{r'}) \rangle = \langle G_{\Omega}(\mathbf{r}) G_{\Omega}(\mathbf{r'}) \rangle$$  \hspace{1cm} (3.56)

with $\langle G_{\Omega}(\mathbf{r}) \rangle$ denoting the proteins in the ground state.

For the ensemble of proteins after the laser pulse, the situation is a little more complex. It is helpful to split the ensemble into two parts:

$$\langle A_{\Omega}(\mathbf{r}) A_{\Omega}(\mathbf{r'}) \rangle = p_\Omega \langle E_{\Omega}(\mathbf{r}) E_{\Omega}(\mathbf{r'}) \rangle + (1 - p_\Omega) \langle G_{\Omega}(\mathbf{r}) G_{\Omega}(\mathbf{r'}) \rangle$$  \hspace{1cm} (3.57)

and

$$\langle A_{\Omega}(\mathbf{r}) \rangle = p_\Omega \langle E_{\Omega}(\mathbf{r}) \rangle + (1 - p_\Omega) \langle G_{\Omega}(\mathbf{r}) \rangle ;$$  \hspace{1cm} (3.58)

one containing all excited proteins $\langle E_{\Omega}(\mathbf{r}) \rangle$ and an other $\langle G_{\Omega}(\mathbf{r}) \rangle$ containing all proteins remaining in their ground state. The pre-factor of these two contributions are governed by the probability of excitation $p_\Omega$, which depends in general on the orientation $\Omega$ of the protein for transitions with anisotropic probabilities. Note that we assume here that the probability of excitation does not depend on the internal degrees of freedom, which should be reasonable, at least for a large fraction of systems, with the exception of those atoms which are directly involved in the transition.

The ensemble of excited proteins $\langle E_{\Omega}(\mathbf{r}) \rangle$ will undergo conformational changes succeeding excitation, which are to be measured. Interestingly, in the difference
3 Theory

term

\[ D_\Omega^\gamma(r,r') = p_\Omega \left[ (E_\Omega(r)E_\Omega(r')) - \langle G_\Omega(r)G_\Omega(r') \rangle - \left( \langle E_\Omega(r) \rangle - \langle G_\Omega(r) \rangle \right) g_\Omega(r') - \\
- g_\Omega(r) \left( \langle E_\Omega(r') \rangle - \langle G_\Omega(r') \rangle \right) \right] \]

(3.59)

\[ = p_\Omega D_\Omega(r,r') \]

all contributions independent from the excitation probability \( p_\Omega \) cancel out and the latter can be separated from the unweighted difference term \( D_\Omega(r,r') \).

Before proceeding in the derivation it is worth adding some comments on entered assumptions and conditions. So far we assumed that each protein system either stays in the ground state \( G \) or gets excited into the state \( E \). Another common path for relaxation would be a direct decay into the ground state under emission of an photon of similar wave length compared to the one used for excitation. This would decrease the quantum yield, which acts as a factors within \( p_\Omega \) and thus effect only the absolute scaling. On the other hand fast radiationless decay is not covered by the theory as this would correspond to a number of thermally excited proteins electronically corresponding to the ground state. However as proteins in solution are very dissipative this should be relevant only on very short timescales.

On intermediate to long timescales, thermal heating of the solvent and the corresponding change in density plays a role. Thermal heating will effect both the pure bulk term (3.40) (which is absent in (3.59)) as well as the cross terms (the last terms in (3.59)). Unlike the latter, the first is an additional effect, which can be isolated from the remaining calculation. For the following derivation we will neglect the effect of heating and leave it for the discussion.

The angular excitation probability \( p_\Omega \) acts as weight when integrating (3.59)

\[ \Delta I(q) \propto \frac{1}{4\pi} \int d\Omega p_\Omega D_\Omega(q) \]

(3.60)

over all orientations \( \Omega \).

Analogous to the time-independent case, the integration over the rotational orientation of the molecule can be replaced by a integration over a sphere of \( q_\Omega \) vector. The vector however defines only two of the three degrees of freedom. The third degree of freedom corresponds to a rotation \( \psi \) around the vector \( q_\Omega \). In the time-independent case this could be ignored, because
3.4 Anisotropic time-resolved solution scattering

Figure 3.2: For a given momentum transfer \( q_0 \) in the molecule frame of reference (A) the excitation moment \( m \) is fixed and the X-ray beam \( k \) is located on a cone (transparent black). Possible laser polarization moments \( l_\beta \) for one specific possible X-ray beam direction (black) are depicted in yellow, possible laser polarization for any of the X-ray beams for two specific points on the detector on the horizontal \( l_H \) and the vertical \( l_V \) are depicted in orange respectively green. A top view of experiment frame of reference is shown in (B). For a defined momentum transfer \( q \) and \( q_0 \) in both reference frames, the orientation of the protein is not defined in respect of its rotation \( \psi \) around \( q \). The excitation moment \( m \) is thus somewhere on a cone (blue). The vector \( m \) and \( l \) are decomposition into components \( l_q, m_q \) parallel and \( l_\perp, m_\perp \) perpendicular to \( q \).

\[
D_0(q_\Omega) = \langle |E_0(q_\Omega)|^2 \rangle - \langle |G_0(q_\Omega)|^2 \rangle - 2 \Re \left[ \langle E_0(q_\Omega) \rangle - \langle G_0(q_\Omega) \rangle \right] \langle g_\Omega^* \rangle \tag{3.61}
\]

is independent of this rotation.

In equation (3.60) the integration over different orientation is define by an Euler angle \( \Omega \). In the isotropic case (3.51) it turn out to be convenient to express this integration in respect of the momentum transfer \( q_\Omega \) in the reference frame of the molecule. To do the same in the anisotropic case additional considerations are needed.

In Fig. 3.2.A the momentum transfer \( q_\Omega \) in the reference frame of the molecule is visualized by a red bar. For a given \( q_\Omega \) the corresponding incident wave vectors \( k \) is not unique, but lies on a cone; some instances are depicted as (transparent) black bars. The wave vector can be specified with an additional angle \( \psi \), corresponding to a rotation around \( q_\Omega \). Momentum transfer \( q_\Omega \) and \( k \) may define two vectors in the reference frame of the molecule. In Fig. 3.2.B the same two vectors are depicted in the reference frame of the experiment, however, the momentum transfer in the reference frame of the experiment will be denoted as \( q \). Two vectors fixed in two reference frames define the reference frame’s relative orientation. Here the momentum transfer \( q_\Omega \) and \( q \) and the wave vector \( k \) by the means of \( \psi \) are fixed in the molecule.
as well as in the experiment, they thus consequently define the orientation of the molecule. It was described before, that the wave vector is determined by $q$ and $\mathbf{\hat{A}}$ in the reference frame of the molecule and the wave vector is obviously fixed in the experiment. The important conclusion of this considerations is, that the molecules orientation is uniquely defined by $q$, $\mathbf{q}$ and $\mathbf{\hat{A}}$.

The integration over an Euler angle $\Omega$ in (3.60) can thus be replaced by an integration

$$\Delta I(q) \propto \frac{1}{8\pi^2} \int d\psi \int_{|q\mathbf{\Omega}|=q} dq_{\mathbf{\Omega}} p(q_{\mathbf{\Omega}}, q, \psi) \tilde{D}_0(q_{\mathbf{\Omega}})$$

(3.62)

over the momentum transfer $q_{\mathbf{\Omega}}$ and an angle $\psi$ in the molecule for a fixed momentum transfer in the experiment $q$. So far the angle $\psi$ was not referenced. However the integration over the full circle in (3.62) will turn this not to be necessary and thus $\psi$ will simply denote the degree of freedom of the rotation around the $q_{\mathbf{\Omega}}$ and $q$ vector. The excitation probability $p(q_{\mathbf{\Omega}}, q, \psi)$ is a function of the molecules orientation specified by $q_{\mathbf{\Omega}}$, $q$ and $\psi$. In (3.62) only the excitation probability $p(q_{\mathbf{\Omega}}, q, \psi)$ depends on $\psi$ but not the difference term $D_0(q_{\mathbf{\Omega}})$ and we can therefore separate the corresponding integration

$$\Delta I(q) \propto \frac{1}{8\pi^2} \int_{|q\mathbf{\Omega}|=q} dq_{\mathbf{\Omega}} \tilde{D}_0(q_{\mathbf{\Omega}}) \int d\psi p(q_{\mathbf{\Omega}}, q, \psi).$$

(3.63)

The probability in the second integral $p(q_{\mathbf{\Omega}}, q, \psi)$ can be an arbitrary function of the proteins orientation; however, for the following derivation we assume a particular form. The excitation probability for photo-selection is often described to be proportional to the $\cos^2$ of the angle $\phi$ between the laser polarization $l$ and the excitation dipole moment $m$. Here we generalize $p(q_{\mathbf{\Omega}}, q, \psi) = p_{\text{lm}} \left( (l \cdot m(q_{\mathbf{\Omega}}, q, \psi))^2 \right)$ to be an arbitrary function of the $\cos^2$ of a vector $l$ fixed in the reference frame of the experiment and the other one, $m(q_{\mathbf{\Omega}}, q, \psi)$, being fixed in the reference frame of the molecule. The latter thus effectively depends on the orientation of the molecule and therefore on $q$, $q_{\mathbf{\Omega}}$ and $\psi$.

At this point one may recall that the vector $q$ describes two angles, the scattering angle $\alpha$, related to its norm $q$, and the azimutal angle $\beta$. The spherical average is for each absolute value of $q$ individual, and one may thus think of $q$ as a constant within this derivation. On the other hand, the azimutal angle described the radial point on the detector. It is important to note that the same $q_{\mathbf{\Omega}}$ and $\psi$ denote different orientation of the molecule depending on $q$ i.e. the azimutal angle $\beta$. This is the
origin of the radial asymmetric scattering pattern. The value of the second integral will act as a weight for the first integral, determining the contribution of \( \tilde{D}_0(q_\Omega) \) for a specific \( q_\Omega \) to the intensity \( \Delta I(q) \) at a specific point on the detector. It is the weight of the mapping from the three-dimensional reciprocal image \( \tilde{D}_0(q_\Omega) \) to the two-dimensional projection on the detector \( \Delta I(q) \).

The second integral

\[
\frac{1}{2\pi} \int d\psi p(q_\Omega, q, \psi) = p_{lm} \left( \left\langle (\mathbf{l} \cdot \mathbf{m})^2 \right\rangle_{\psi}^{q_\Omega q(q, \varphi)} \right) \tag{3.64}
\]

correspond to the average over all orientation of the molecule complying with fixed momentum transfers \( q \) and \( q_\Omega \) in their respective frame of reference. The vector \( \mathbf{m} \) is fixed in the molecule, however, the vector \( q_\Omega \) and \( q \) do not define the rotation \( \psi \) of the molecule around \( q \). With \( q_\Omega \) and \( q \) given, the vector \( \mathbf{m} \) is thus in the reference frame of the experiment only specified on a cone, depicted blue in Fig. 3.2. The average \( \left\langle (\mathbf{l} \cdot \mathbf{m})^2 \right\rangle_{\psi}^{q_\Omega q(q, \varphi)} \) in (3.64) is the average squared scalar product of the vector \( \mathbf{m} \) on the blue cone and the vector \( \mathbf{l} \).

For the calculation of this average it is convenient to represent both \( \mathbf{m} \) and \( \mathbf{l} \) as the sum (3.65) of their part \( \mathbf{l}_q \) and \( \mathbf{m}_q \) parallel to \( q \) respectively their part \( \mathbf{l}_\perp \) and \( \mathbf{m}_\perp \) perpendicular to \( q \).

\[
\left\langle (\mathbf{l} \cdot \mathbf{m})^2 \right\rangle = \left\langle ((\mathbf{l}_q + \mathbf{l}_\perp) \cdot (\mathbf{m}_q + \mathbf{m}_\perp))^2 \right\rangle \tag{3.65}
\]

\[
= \left\langle (\mathbf{l}_q \cdot \mathbf{m}_q + \mathbf{l}_\perp \cdot \mathbf{m}_\perp)^2 \right\rangle \tag{3.66}
\]

\[
= \left\langle (l_q m_q + l_\perp m_\perp \cos(\psi))^2 \right\rangle \tag{3.67}
\]

\[
= \left\langle l_q^2 m_q^2 + 2l_q m_q l_\perp m_\perp \cos(\psi) + l_\perp^2 m_\perp^2 \cos^2(\psi) \right\rangle \tag{3.68}
\]

\[
= l_q^2 m_q^2 + \frac{1}{2} \left( 1 - l_q^2 \right) \left( 1 - m_q^2 \right) \tag{3.69}
\]

\[
= \left( \frac{1}{2} l_q^2 - \frac{1}{6} \right) \left( 3m_q^2 - 1 \right) + \frac{1}{3} \tag{3.70}
\]

The parallel respectively perpendicular parts are orthonormal by construction, and thus the cross terms \( \mathbf{l}_q \cdot \mathbf{m}_\perp \) and \( \mathbf{l}_\perp \cdot \mathbf{m}_q \) are both zero (3.66). The parallel parts are projected on the same vector and hence their dot product equals the product of their norm (3.67). The perpendicular parts \( \mathbf{l}_\perp \) and \( \mathbf{m}_\perp \) are on a plane orthogonal to the vector \( q \). Their scalar product can be calculated as a product of their norms and the cosine of their angle \( \psi_{lm} \). However, \( \mathbf{m}_\perp \) is, due to the free rotational degree of freedom \( \psi \), free to be oriented in any direction on the mentioned plane, which
corresponds to the base of the blue cone in Fig. 3.2. The angle defining the rotation
\( \psi \) of the molecule around \( q \) was not referenced so far and we can without loss of
generality reference this angle \( \psi = \psi_{lm} \) in respect of \( l_\perp \) and \( m_\perp \). Integration of
(3.69) over \( \psi \) from 0 to \( 2\pi \) eliminates finally the free parameter \( \psi \) and alongside all
cross terms arising form expanding the square root (3.68). The two factors \( m_q^2 \) and \( l_q^2 \) are left, corresponding to the projection of the vector \( m \) and \( l \) on the momentum
transfer \( q_\Omega \) and \( q \). This projection is independent of the reference frame, and can be
calculated in the more convenient one, respectively. \( m_q^2(q_\Omega) \) is thus to be calculated
in the reference frame of the protein and will depend on \( q \). \( l_q^2 \) on the other hand is to be calculated in the reference frame of the experiment and depends on
the point on the detector parametrized by \( q \).

Plugging (3.70) into (3.64) and the latter into (3.62) we obtain

\[
\Delta I(q) \propto \int_{|q_\Omega|=l} d\mathbf{q}_\Omega p_{lm} \left( \left( \frac{1}{2} l_q^2 - \frac{1}{6} \right) (3m_q^2 - 1) + \frac{1}{3} \right) D_0(q_\Omega).
\]  

(3.71)

For a excitation probability proportional to the \( \cos^2 \) of the angle between laser
polarization \( l \) and excitation dipole moment \( m \) the function \( p_{lm}(x) = x \) is simply the
identity. Myoglobin, which is in the focus of this work, is however a circular absorber.
In this case the excitation probability can be described by \( p_{lm}(x) = 1 - x \), with \( m \)
representing a vector perpendicular to the excitations moments. Both functions are
linear and as \( l_q^2 \) does not depend on the \( q_\Omega \) and it can therefore be pulled in front
of the integral (3.62). The anisotropic difference scattering intensity for a circular
absorber is given as the sum

\[
\Delta I(q) \propto -\frac{1}{6} (3l_q^2(q) - 1) \frac{1}{4\pi} \int_{|q_\Omega|=l} d\mathbf{q}_\Omega \left[ (3m_q^2(q_\Omega) - 1) D_0(q_\Omega) \right] + \frac{2}{3} \frac{1}{4\pi} \int_{|q_\Omega|=q} d\mathbf{q}_\Omega \tilde{D}_0(q_\Omega)
\]

\[
= -\frac{1}{6} (3l_q^2(q) - 1) \Delta I_m(q) + \frac{2}{3} \Delta I_{iso}(q)
\]

(3.72)

of two contributions, the isotropic difference scattering intensity \( \Delta I_{iso}(q) \) and the
weighted difference scattering intensity \( \Delta I_m(q) \). The basis used in (3.72) is in general
arbitrary. The basis chosen here has the advantage of corresponding to the spherical
harmonics \( Y_0^0 \) and \( Y_2^0 \) and thus being orthogonal. The weight

\[
m_q^2(q_\Omega) = \left( \frac{m_0 \cdot q_\Omega}{q} \right)^2
\]

(3.73)
corresponds to the $\cos^2$ of the angle between $\mathbf{q}_\Omega$ and the vector $\mathbf{m}$ in the reference orientation of the protein indicated by the subscript $0$. The pre-factors of both contributions depend on $l_2^2(q)$ and thereby on the $\cos^2$ of the angle between $\mathbf{q}$ in the reference frame of the experiment and the vector $\mathbf{m}$. It is useful to split $l$ and $q(q, \varphi)$ into their parts parallel and perpendicular to the X-ray beam (here in line with the x-axis).

\[
\begin{align*}
l_2^2(q) &= \left(\frac{1 \cdot q}{q}\right)^2 \\
&= \left(\frac{l_x q_x}{q} + \cos(\varphi) \frac{l_y q_y q_y}{q}\right)^2 \quad \text{(3.74)} \\
&= \frac{l_x^2 q_x^2}{q^2} + 2 \cos(\varphi \frac{l_x q_x l_y q_y}{q^2} + \cos^2(\varphi) \frac{l_y^2 q_y^2}{q^2} \quad \text{(3.75)} \\
&= \frac{l_x^2 q_x^2}{4k^2} + 2 \cos(\varphi \frac{l_x l_y q_y}{2k} \sqrt{q^2 - \frac{q^4}{4k^2} + \cos^2(\varphi) \frac{l_y^2 q_y^2}{4k^2}} \quad \text{(3.76)} \\
&= \frac{l_x^2 q_x^2}{8k^2} - \frac{1}{6} \Delta I_{\text{iso}}(q) + \frac{2}{3} \Delta I_{\text{iso}}(q) \quad \text{(3.77)}
\end{align*}
\]

It is to be kept in mind that $q$ is substituting the scattering angle $\theta$ and the azimuth angle $\varphi$. Geometric considerations $\frac{q}{2k} = \cos \alpha = \frac{q_x}{q}$ allows to express the parallel part $q_x = \frac{q^2}{4k}$ in terms of the norm $q$ and the X-ray beams wave number $k$. The perpendicular part is then given by $q_y = \sqrt{q^2 - q_x^2} = \sqrt{q^2 - \frac{q^2}{4k^2}}$. In a typical setup, the laser beam is perpendicular to the X-ray beam and can be assumed to be in line with the vertical z-axis. It follows that $l_z = 0$ and the azimuth angle $\varphi$ is to be referenced by the horizontal y-axis. It is easily seen, that for a laser polarization parallel to the X-ray beam, the difference intensity

\[
\Delta I^\parallel(q) = -\left(\frac{q^2}{8k^2} - \frac{1}{6}\right) \Delta I_{\text{iso}}(q) + \frac{2}{3} \Delta I_{\text{iso}}(q) \quad \text{(3.78)}
\]

does not depend on the azimuthal angle $\varphi$, thus yielding an isotropic scattering pattern. However, for all other orientations $\Delta I(q)$ depends on $\varphi$, thus yielding an anisotropic pattern. Table 3.1 contains the difference intensity for laser polarization orthogonal $\Delta I^\perp(q)$ to the X-ray beam as well as for circular polarization $\Delta I^\theta(q)$.

Azimuthal averages, which have been reported experimentally[63], can easily be obtained by corresponding integration of (Table 3.1.1). Once the isotropic $\Delta I_{\text{iso}}(q)$ and the weighted $\Delta I_{\text{m}}(q)$ difference scattering intensity are obtained, it is an easy task to calculate the experimental difference scattering pattern using table 3.1.
### 3 Theory

\[
\Delta I^2 = \begin{cases} 
I_y = 1, I_z = I_x = 0 & \frac{1}{2} \left[ \Delta I^H + \Delta I^V \right] \\
\Delta I(q) & 1
\end{cases}
\]

<table>
<thead>
<tr>
<th>( \Delta I^2 )</th>
<th>( \Delta I^3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Delta I(q) )</td>
<td>( \Delta I^H(q) - \frac{1}{3} \Delta I^V(q) )</td>
</tr>
<tr>
<td>( \Delta I(q) )</td>
<td>( \Delta I^V(q) - \frac{1}{3} \Delta I^H(q) )</td>
</tr>
</tbody>
</table>

Table 3.1: The difference intensities for a circular absorber for a horizontal and circular laser polarization (columns). The general formula is presented (1) (note: \( c_q^2 = \cos^2(q) \)). The intensities of horizontal (2) and vertical (3) cuts, their difference (4) and a radial average (5) are tabulated.

is important to note that the two-dimensional scattering pattern in the anisotropic case does not contain the full two-dimensional information content, but instead can be constructed by a linear combination of two one-dimensional components.

With the help of table 3.1 the isotropic component can for example be extracted from the horizontal and vertical scattering intensities

\[
\Delta I_{iso}(q) = \frac{3}{2} \left[ \frac{1}{6} \Delta I^H(q) + \left( \frac{1}{3} - \frac{q^2}{8k^2} \right) \Delta I^V(q) \right] \tag{3.79}
\]

with a laser polarization perpendicular to the x-ray beam. Likewise, the weighted component can be obtained

\[
\Delta I_m(q) = \frac{1}{2} \left[ \Delta I^H(q) - \Delta I^V(q) \right].
\]

Hence, the difference term \( \Delta I^H(q) - \Delta I^V(q) \) is proportional to the weighted component. This makes the difference term an experimentally convenient measure of the anisotropy content of an WAXS experiment and suggest this to be the primary reported one as done by Cho et al. [64].

#### 3.5 Rotational diffusion

As described before, the excitation event induces an anisotropy into the solution as only preferable oriented proteins are excited. The proteins in solution are tumbling
3.5 Rotational diffusion

do to stochastic movements of the solvent and previous preferable orientation of the excited proteins will vanish. Diffusion affects both the translational degrees of freedom and equally the rotational degrees of freedom. The orientational distribution of excited proteins \( f(\Omega, t) \) with the proteins orientation defined by \( \Omega \) is anisotropic in the event of excitation. By rotational diffusion, this distribution equilibrates until isotropy is reached. The related process is rotational diffusion and can be described by the rotational version of Fick’s second law

\[
\frac{\partial p(\Omega, t)}{\partial t} = \nabla^2_{\Omega} \hat{D} p(\Omega, t) \tag{3.80}
\]

which differ form the translational version only as the differential operator \( \nabla^2_{\Omega} \) here acts on the rotational degrees of freedom. The diffusion tensor \( \hat{D} \) is the second moment of angular displacements. In this work we want to focus on the special case of isotropic diffusion i.e. the rotation diffusion is the same for any given rotational axis within the protein. In the isotropic case we obtain a diffusion constant

\[
D = \frac{1}{2} \lim_{\Delta t \to 0} \frac{\langle \Delta \Omega^2 \rangle}{\Delta t} \tag{3.81}
\]

from infinitesimal small angular displacements \( \Delta \Omega \) [97]. The time limit refers to timescales, which are long enough for the protein internal degrees of freedom to be averaged, but short enough for the proteins overall angular displacement to be infinitesimal. (3.80) can then be transposed

\[
\frac{1}{D} \frac{\partial p(\Omega, t)}{\partial t} = \nabla^2_{\Omega} p(\Omega, t) \tag{3.82}
\]

such that the right hand side corresponds to the Laplace’s equation. Laplace’s spherical harmonics \( Y^m_l \) form a orthogonal set of solutions to the angular Laplace’s equation [98]. Note that the spherical harmonics do only include two of the three rotational degrees of freedom. Here we will make us of one only anyway. The spherical harmonics are eigenfunctions

\[
\nabla^2_{\Omega} Y^m_l = -l(l + 1)Y^m_l \tag{3.83}
\]

of the \( \nabla^2_{\Omega} \) operator with eigenvalue \(-l(l + 1)\). Expanding the orientational probability distribution in terms of spherical harmonics
$$p(\Omega, t) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} f_{lm}(t) Y_l^m$$

(3.84)

let us easily find the solution of (3.82) as

$$p(\Omega, t) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} f_{lm} Y_l^m e^{-t/\tau_l}$$

(3.85)

with \(f_{lm}\) being the population of the spherical harmonics at \(t = 0\) and

$$\tau_l = \frac{1}{D_l(l+1)}$$

(3.86)

being the time-constant of the decay of the spherical harmonics amplitude [58]. For our purposes it is relevant how the orientation distribution proportional to \(\cos^2\) evolves. The initial probability \(p(\Omega, t) = \cos^2(\theta_\Omega)\) can be expressed as a linear combination of two spherical harmonics,

$$p(\Omega, 0) = \frac{2\sqrt{\pi}}{3} Y_0^0 + \frac{4\sqrt{\pi}}{3\sqrt{3}} Y_2^0$$

(3.87)

with \(Y_0^0 = \frac{1}{2\sqrt{\pi}}\) and \(Y_2^0 = \frac{\sqrt{5}}{4\sqrt{\pi}} (3\cos^2 \theta - 1)\). The solution of (3.82) under the given constrained (3.87) is

$$p(\Omega, t) = e^{-6D_l t} \left( \cos^2 \theta - \frac{1}{3} \right) + \frac{1}{3}$$

showing a decay with a time constant of \(\tau_D = \frac{1}{D_6}\). For \(\cos^2 \theta\), we can enter \((1 \cdot m)^2\)

and obtain a time evolution of its average

$$\left\langle (1 \cdot m)^2 \right\rangle_t = \left\langle e^{-6D_l t} \left( (1 \cdot m)^2 - \frac{1}{3} \right) + \frac{1}{3} \right\rangle = e^{-6D_l t} \left( \left\langle (1 \cdot m)^2 \right\rangle - \frac{1}{3} \right) + \frac{1}{3}$$

over the free rotational degree of freedom \(\psi\). The previous result for the scalar product (3.70) can be inserted
\[
\langle (\mathbf{l} \cdot \mathbf{m})^2 \rangle_t = e^{-6Dt} \left( m_q^2 \left( \frac{3}{2} l_q^2 - \frac{1}{2} \right) + \left( -\frac{1}{2} l_q^2 + \frac{1}{2} \right) - \frac{1}{3} \right) + \frac{1}{3}
\]

\[
= e^{-6Dt} \left( \frac{1}{2} l_q^2 - \frac{1}{6} \right) \left( 3m_q^2 - 1 \right) + \frac{1}{3}.
\]

Thus we obtain for the difference scattering pattern

\[
\Delta I(q, \varphi, t) \propto e^{-6Dt} \left( \frac{1}{2} l_q^2 - \frac{1}{6} \right) \Delta I_m(q, t) + \frac{2}{3} \Delta I_{iso}(q, t) \tag{3.88}
\]

considering rotational diffusion in the case of a circular absorber. This equation shows that contribution of the anisotropic component \( \Delta I_m(q, t) \) as well as radial asymmetry of the scattering pattern, which is determined by \( l_q^2 \), decay exponential with a time constant of \( \tau_D = \frac{1}{6D_r} \). Using (3.88) and table 3.1 it is a straightforward task to obtain the relevant difference intensities for different laser polarization. Interestingly, this exponential decay is present irrespectively of the laser polarization and even in the case of an radial average, however in different strength. The smallest contribution can be found in the case of the radial average and circular polarization. To completely remove the effect of the anisotropic decay, the isotropic component \( \Delta I_{iso}(q) \) can be extracted according to (3.79).

This section was limited to isotropic rotational diffusion. In general and especially for non globular proteins the diffusion will be anisotropic. One then may obtain a three dimensional diffusion vector. Anisotropic rotational diffusion has been studied in the context of anisotropic fluorescence and it was found that the diffusion process results in up to 6 distinct timescales [97], determined by the three dimensional diffusion vector. These timescales corresponds to different spherical harmonics describing orientational distributions of the corresponding component. For the calculation of corresponding scattering pattern additional intensities are to be calculated and weighted according to these spherical harmonics.

### 3.6 Information content

The most established way to estimate the information content of WAXS scattering curves is based on the number of Shannon channels[99]. From the calculation of \( I(q) \) as a Fourier transform of \( D(r, r') \) we can conclude that the highest frequency within \( I(q) \) is limited by the largest distance \( \Delta r = r - r' \). The largest relevant distant is
given by the solute diameter $D_{\text{max}}$ resulting in a maximum frequency of $f_{\text{max}} = \frac{D_{\text{max}}}{2\pi}$. The Nyquist-Shannon sampling theorem[100] states that the full information content can be obtained when sampling a function equidistant with $\Delta q = \frac{1}{2f_{\text{max}}}$ depending on the maximum frequency $f_{\text{max}}$ within the signal. The largest relevant distant of an WAXS pattern is given by the solute diameter $D_{\text{max}}$ and, thus, sampling $I(q)$ equal distant with spacings of $\Delta q = \frac{\pi}{D_{\text{max}}}$ should be sufficient. The total number of necessary sampling points and in reverse the total number of independent data points is thus

$$N_q = (q_{\text{max}} - q_{\text{min}})D_{\text{max}}/\pi.$$ 

This corresponds e.g. for myoglobin ($D_{\text{max}} = 3.5nm$) to $\sim 27$ independent data points in the WAXS regime ($q_{\text{max}} = 25 \text{nm}^{-1}$). Estimating the information content with the number of Shannon channels has be questioned. First of all representing the function $I(q)$ with a corresponding finite number of Shannon channels gives meaning full results only within the observed $q$-range[24]. Physically meaningful asymptotical behavior is not considered as additional constrains. Furthermore usually much more data points then theoretical estimated Shannon channels are measured. This over-sampling has argued to allow reconstruction beyond the limit of Shannon channels, similar as done for image analysis (‘super-resolution’) [24]. Until these arguments are theoretical well-founded the number of Shannon channels remains the relevant measure of the information content of solution x-ray pattern.

It was shown in Subsection 3.4, that anisotropic scattering pattern can be calculated from two components and visa versa these two components can be obtained from anisotropic scattering pattern; one equally weighted over all orientations $\Delta I_{\text{iso}}(q)$ and the other $\Delta I_{m}(q)$ weighted according to a spherical harmonic with $l = 2$ quantum number. Since the weight of both components are orthonormal, we do not expect any systematic correlation between the two components.

To estimate the $q$-range, in which we may get a non zero anisotropic component, on should consider that alternative to the here presented quadrature over a sphere, scattering pattern can be calculated from a series of spherical harmonics. The weight of the corresponding spherical harmonics is given by the corresponding Bessel function $j_l(x)$. The first maximum of $j_2(x)$, the relevant Bessel function for the earlier mentioned spherical harmonic $Y_2^0$, is at $x_1 = 3.34$. As a lower limit we can conclude that the second harmonics becomes relevant for $q > \frac{3.3}{D_{\text{max}}}$. With this rough estimate we expect to observe an non zero anisotropic component in the case
of myoglobin for \( q > 0.9 \text{ nm}^{-1} \). Assuming the same information density in the anisotropic component thus \( \sim 26 \) independent data-points can be expected. The arbitrary scaling absorbs one degree of freedom of the isotropic pattern and we can estimate that the use of anisotropic information can increase the information content by a factor of two.

For experimentals, it is, however, relevant how many of that theoretically available information can be obtained above noise level. In this respect, the anisotropic decay due to rotational diffusion is becoming relevant. In addition, with increasing \( q \), the radial asymmetry decreases (see table 3.1) and the signal to noise ratio of double difference spectra \( \Delta I_H(q) - \Delta I_V(q) \) can be expected to decrease.
4 Implementation

Fig. 4 shows the workflow for the calculation of an anisotropic scattering pattern from two trajectories G and E. Two in-house developed software tools as well as a modified GROMACS main binary (mdrun) are successively used. The latter was recently presented [43] and calculated scatter pattern were shown to match experimental data with high accuracy[43].

In a first step (i) both trajectories are combined and a single envelope is constructed by the tool \texttt{g\_genenv} containing the protein and its solvation shell. The envelope’s scattering amplitude is calculated and the envelopes shape is passed to \texttt{mdrun}. The latter is then (ii) used to accumulate an averaged scattering amplitude. Finally (iii) the anisotropic scattering pattern is calculated from the scattering amplitudes of the two trajectories G and E as well as the envelope by a Python script called \texttt{passrot}.

For a robust error estimation the trajectories are usually binned into smaller fragments, and the average scattering amplitude of these fragments is then individually calculated in step (ii). This allow for an accurate error estimates as well as an estimate of the sampling quality of the trajectories. The scattering amplitudes are calculated in \texttt{mdrun} and \texttt{g\_genenv} for the same sphere of \textit{q}-vectors, before the spherical integration is performed in step (iii). Thus, first the generation of the sphere of \textit{q}-vectors will be discussed and then the function of each tool will be elaborated in more detail.

4.1 Spiral method

For the numerical quadrature of a sphere by a number of \textit{q}-vectors it is essential that these are equally distributed to prohibit unintentional weighting of certain directions. Here the spiral method[101] is implemented for obtaining set of \textit{J} vectors on a sphere, which has successfully been used in the context of spherical averaging scattering amplitudes[42]. Accordingly, \textit{q}-vectors are placed at equal distance on a spiral with \[ \theta_j = \sqrt{\pi} J \arccos((2j - 1 - J)/J) \] and \[ \phi_j = \sqrt{\pi} J \arcsin((2j - 1 - J)/J) \]
defining the solid angle of $q_j$ in the usual spherical coordinates. In the current implementation, the same set of vectors is used for all absolute values of $q$ only differing in their norm.

### 4.2 $g\_genenv$

The derivation presented in the theory section is based on separating the protein and its solvation layer from the bulk solvent. The envelope represent the boundary surface of this separation. Numerically, it is feasible to represent the envelope as a polygon mesh. In our implementation it is constructed from an icosphere, which itself is build from an icosahedron. Starting form the 20 equal-sided triangles of the icosahedron, its faces are iteratively subdividing into four equal-sided triangles. Four
4.3 mdrun

A spin off of the GROMACS main binary mdrun was developed in-house to calculate difference spectra from MD simulations of ordinary WAXS experiments[43] according to (3.51) and (3.52) in a stand alone fashion. For this purpose beside the trajectory of the solution and the envelope file also a trajectory of a pure water box needs to be passed as an input. For the purpose of this work, however, the GROMACS extension was slightly modified to calculated solely the averaged scattering amplitude as well as the average of its absolute squared as needed for a separate
calculation of the difference intensity according to (3.61). In a first step, each frame needs to be prepared, before the scattering amplitude can be calculated.

4.3.1 Preparation

In an unconstrained MD simulation the protein can move and rotate freely. Because the average is to be taken for the protein in a specific orientation a number of steps are needed to place the protein within the envelope in an optimal manner. This includes a shift of atoms into the correct unit cell and a rotational and translational movement of the system. This has to be done such that the solute is in its optimal position within the envelope, and such that the solvent completely fills the envelope. The previous considerations are fulfilled by the following protocol:

1. As the protein may in the course of the MD partly move across periodic boundaries the protein is made whole in a first step. From possible mirror images of each atom the one is chosen with minimize the distant to a preselected reference atom. A suitable reference atom is suggested by \( g_{genenv} \) before construction of the envelope.

2. The center of mass of the solute is moved to the COM of the reference structure and the rotational matrix which applied to the solute atoms minimize the RMSD to the reference structure \( x_{\text{ref}} \) is calculated.

\[
\min_{R_{\omega' \rightarrow \omega}} (x_{\text{ref}} - R_{\omega' \rightarrow \omega} x)^2
\]

The rotational matrix \( R_{\omega' \rightarrow \omega} \) rotates the protein from is orientation \( \omega' \) into the distinguished orientation \( \omega \) needed for an optimal placement within the fixed envelope.

3. It is essential that the envelope is completely filled with water and that no periodic images of solvent atoms are present, as taking the Fourier transform in this case would course artifacts. To test whether the envelope fits into the unit cell box it is first rotated using the inverse rotational matrix \( R_{\omega \rightarrow \omega'}^{-1} \) in order to be in the same orientation then the protein. The midpoint \( r_{es}(\omega') \) of the smallest sphere enclosing the envelope is calculated using the Miniball algorithm by Gärtner[102]. The envelope is then translated \( r_{box} - r_{es}(\omega') \) such that the enclosing sphere’s midpoint coincide with the midpoint \( r_{box} \) of the simulation box. Subsequently for each vertices of the envelope is checked
wether it is within or outside of the unit cell. If a single vertex is situated outside of the unit cell the code exits with an error message as artifacts are expected.

4. All atoms of solute and solvent are translated in such a way that the COM of the solute coincide with the optimal position for the enclosing sphere midpoint \( r_{\text{box}} - r_{\text{es}}(\omega') \). At this point all solute atoms should be inside one unit cell. Otherwise, an additional check throws an error suggesting that something went wrong in step 1 and a better reference atoms should be chosen. Next, solvent atoms are translated along the box vectors to completely fill the zero unit cell.

5. In a next step all atoms are translated such that the COM of the solute is centered at the origin and subsequent rotated by \( R_{\omega' \rightarrow \omega} \) for the protein being oriented into the reference orientation.

6. The protein and the surrounding solvent are now in the correct position for the envelope, which is defined in respect to the proteins COM. The previous steps assure that the envelope will be completely filled with solvent atoms. The last step is now to select those solvent atoms which are within the envelope as the solvation layer. All atoms which are further apart from the COM then the furthest envelope vertex are disregarded. For all other atoms, those icosphere face is selected whose normalized midpoint vector has the the larges scalar product with the vector to the atoms position. This way, the pyramidal volume of the envelope is found which could contain the atom. The atom is then added to the solvation layer if it is situated in this pyramidal volume.

### 4.3.2 Calculation of scattering amplitude

In a next step the fourier transform of the solute and its solvation layers electron density will be calculated from (3.20). This involves a sum over all \( N_{\text{atom}} \) atoms for each of the \( N_q \) \( q \)-vector. The atomic scattering factors are first calculated and stored for each atom and \( q \)-vector. The exponential within each summand can then be calculated with a cosine for the real part and a sine for imaginary part. Thus, in total the calculation of \( N_{\text{atom}} \cdot N_q \) sine and as many cosine are required for each frame.

As a numerical example the calculation of the scattering factor for \( N_q = N_{\text{qabs}} \cdot N_{\text{sphere}} = 20 \cdot 1500 = 30000 \) \( q \)-vectors for an envelope including \( N_{\text{atom}} = 50000 \) atoms
4 Implementation

takes $\sim 10.5s$ on a single CPU (Intel Xeon E5-2670 2.6Ghz; Santa Clara, CA) for a single frame. This speed is realized using single instruction/multiple data (SIMD) instructions, allowing the simultaneous calculation of four single precision floats on the mentioned architecture. Threat parallelisation is implemented allowing practical linear scaling for typical core numbers (e.g. 16 CPUs on the used machine).

In addition, portability of this calculation on graphical processing units (GPU) was investigated. The same problem took $\sim 510ms$ on a single Nvidia GTX 770 graphic card and one can therefore expect a single GPU to outspeed a 16-20 core node. Different strategies for an efficient memory management were investigated, without gaining any speedup and thus the trigonometric calculations themselves can be considered as the bottle neck. As the standard GROMACS code performs calculation on the GPU as well the memory management has to be investigated in more detail before implementation into GROMACS can be done.

The scattering amplitude as well as its absolute value squared for each q-vector is averaged over the input trajectory. The GROMACS spin off is able to directly calculate difference scattering spectra using (3.51) and (3.52). For this purpose, a trajectory containing solvent only needs to be provided. For the calculation of TR-WAXS patterns, scattering amplitude of excited as well as ground state confirmations need to be calculated separately. For both trajectories the calculated scattering amplitude as well as its absolute squared is stored and passed to passrot.

4.4 passrot

Although being able to calculate isotropic difference spectra with a pure solvent background (as relevant for time-independent experiments), the main purpose of passrot is the calculation of isotropic and anisotropic difference spectra from time-dependent solution x-ray scattering. As input parameters, the electron density of the solvent, the laser polarization and the transition moment of the relevant excitation in respect of the reference structure have to be provided. To include isotropic rotational diffusion, the diffusion constant as well as the time delay needs to be provided. The tool was written in Python and performs the following major steps:

1. The scattering amplitude and its absolute value squared of the ground state, of the excited state and of the envelope are read from input files. If error estimates are to be calculated, multiple uncorrelated averages for both states have to be provided, calculated from $N_{\text{chunks}}$ chunks of the the full data. The
input files should match, i.e. the first file for the excited state and the first file of the ground state (and so on for the second, third, ...) should correspond to trajectories obtained from the same starting position for a realistic error estimate. The input files are checked for consistency, namely that the provided \( q \)-vectors in all files correspond exactly.

2. After this safety check, the weighted \( \Delta I_m(q) \) and unweighted \( \Delta I_{iso}(q) \) scattering intensities are calculated. Therefore, first the difference structure factor \( \tilde{D}(q) \) is calculated as in (3.61) for each \( q \)-vector. The results are accumulated twice, once without pre-factor and an other time with a pre-factor \( m^2(q) \) (see (3.73)) depending on the angle between the \( q \)-vector and a vector \( m \) defined in the proteins reference frame. After normalization, the two accumulated values correspond to \( \Delta I_{iso}(q) \) and \( \Delta I_m(q) \), respectively. For the calculation of an error estimate, both values are calculated independently for each pair of input-files. The intensities \( \Delta I^i(q) \) calculated from all pairs are then averaged to \( \Delta I(q) \) and its standard error is calculated according to

\[
SE_{\Delta I}(q) = \sqrt{\frac{1}{N_{\text{chunks}}} \sum_{i=1}^{N_{\text{chunks}}} (\Delta I(q) - \Delta I^i(q))^2}.
\]

3. From this data, multiple output files are written showing a 2D scattering pattern as well as 1D scattering curves corresponding to the rows of table 3.1. The scattering curves are written with and without consideration of rotational diffusion. Isotropic scattering pattern are additionally calculated for different subsets of the data. This allows a details analysis of the possible autocorrelation between different chunks.

### 4.5 Evaluation

The derivation of Section 2.3 is based on the change of the spherical integration from the reference frame of the experiment to the reference frame of the solute. The first will be called active integration, as it involves an active rotation of solute, the latter correspondingly passive integration.

To test their identity in theory as well as to test their numerical identity in the current implementation, we followed both paths for a very small test system of four randomly positioned iodine atoms. The positions of the atoms where chosen to
avoid any symmetry axis in the system. Our GROMACS extension is capable of calculating the excess intensity for the $q$-vectors as they correspond to the points on a 2D detector screen without performing any spherical integration. The spherical integration was instead performed externally by first generating a set of structures all corresponding to the same object of four atoms, however, with different orientations. The orientation generated followed the spiral method and an additional rotation around the calculated solid angle.

The 2D scattering pattern obtained from all orientations were averaged and compared to the one obtained by passive rotation as described in the previous sections. For the passive rotation a sphere of 1500 $q$-vectors was used.

Fig. 4.2 shows the RMSD between the intensities obtained by passive and active rotation relative to the absolute intensity plotted over the absolute $q$ value for increasing number of active rotations. The plot reveals two things. First the scattering intensity of the active rotation indeed converges towards the values for the passive rotation with increasing number of orientations. In the case of $50 \cdot 750$ orientations the error is less then 0.01% for almost all $q$ values. Second the relative error increases exponentially in the region from $q = 5$ to $q = 15$. This can be referred to the decreasing scattering intensity, but it also shows that for higher $q$ values more accurate sampling of the $q$-sphere is necessary. Indeed the number of $q$ vectors should be taken as $\sim (Dq)^2$ [43].
5 Computational Details

5.1 MD parameter

The GROMACS 4.6.x [88, 103] toolkit was used for all MD simulation throughout this work. The initial crystal structure of horse heard myoglobin was taken from the Protein Data Bank (code 1DWR) [104]. Hydrogens were added with the *Automatic PSF Builder* of VMD[105]. Virtual sides were used for the hydrogens and bonds were constrained with the LINCS [90] algorithm (see section 3.1.2). Due to the virtual sides for the hydrogens and the constrained bonds time-steps of 4fs could be used. A modified CHARMM[1] forcefield called CHARMM22*[106] with TIP3P as a water model was used. Two sodium counter ions were added to neutralize the system.

First, the system was equilibrated with position constrains on the heavy atoms for 1ns and using a Berendsen thermostat [88]. Then another 20ps followed without constrains and the same thermostat. The following productive simulations made use of the Parrinello-Rahman thermostat [88]. The most important parameters used for all productive simulations (if not stated otherwise) are summarized in Table 5.1.

As a base, a single 2.5$\mu$s trajectory and a 240 ns REMD trajectory were calculated.

For the REMD calculations 56 replicas with temperatures between 298.15 K and 400.02 K were used. The temperatures of the replicas were calculated using a web-tool from Patriksson and van der Spoel [107] aiming a exchange probabilities of 0.25. The actual exchange probabilities however turned out to be 0.15. There were exchange attempts every 1.2 ps. The different replicas mixed well within the first 10 ns and constantly over the course of the whole simulation of 240 ns.

The eigenvalues for the covariance matrix of the combined trajectory of the single and the REMD calculation were calculated. Fig. 5.1.B shows the projections of the first five eigenvectors on both trajectories individually. Both simulations span the same range for all of these eigenvectors, however, the 2D projection of the first two
eigenvectors in Fig. 5.1.A shows that the phase space of both simulations differ. For the first four eigenvectors the single simulations shows only 1-2 period(s), indicating a incomplete sampling of the phase space. The first eigenvector corresponds to a collective motion, which increases the proteins compactness and includes the arrangement of Glu83 into the regular structure of the E helix. This movement occurred over a timespan of 2 µs in the single calculation, and within 70 ns in REMD simulation.

5.2 CO model

As a part of the story is to model the diffusion of a CO molecule within myoglobin, the modeling of the electronic structure of the CO itself can expected to be crucial. Staub et al. [67] developed a three-side model for CO with charges on the atomic cores as well as on the center of mass, thus including a quadruple moment. The parameters have been fitted to ab initio calculations of CO within a number of different solvents [67]. The three-side model have been used by different studies since then [70, 71].

For this work the parameters as shown in table 5.2 derived by Staub et al. where included into the CHARMM22* forcefield.
5.3 Dissociation process

The heme-CO within myoglobin can be excited into its $\pi \pi^*$ state by yellow and green light\cite{55, 78}. The corresponding transitions are known to be perfectly polarized in the plane of the heme \cite{108}. Heme is thus a circular absorber with two orthogonal components within the plane of equal contribution\cite{109}. Therefore, (3.72) can be applied for the calculation of anisotropic TR-WAXS pattern.

Experimentally the quantum yield of dissociation was estimated as 90\%\cite{81} and 95\%\cite{110}. 3\%\cite{110} and 4\%\cite{111} of rebinding was observed on timescales of hundreds of nanoseconds. Excited proteins, which do not dissociate CO, will still have some contribution to the scattering pattern on very short time scales due to the dissipating heat. In this work this effect was neglected and rebinding was not considered, because of the relatively rare occasions of both effects.

The experimentally the enthalpy difference between MbCO and Mb-CO was mea-
sured as $88.7 - 107.5 \text{ kJ/mol}$. Experiments reported by Kim et al. [55, 63] used an excitation laser with a wavelength of 532 nm deploying a photon energy of $224.9 \text{ kJ/mol}$. In this work, we assumed the excess energy deposited into the system in terms of kinetic energy to be $128.2 \text{ kJ/mol}$, calculated as the difference between the photon energy and the enthalpy increase.

The protocol for the modeling of the dissociation process used in this work is the following. First, (i) the bond between the carbon and the iron is removed. Then (ii) the excess energy due to the removal of that bond is calculated and (iii) the missing energy is added by scaling the kinetic energy of the porphorin ring.

Upon removal of the bond C-Fe bond (i), new non-bonded interactions e.g. Lennard-Jones interaction, which have been excluded previously, do now play a role. This include all first and second neighbors of the CO, i.e., the interactions C-Fe, C-N and O-Fe. Due their proximity this results in a step increase in potential due to non-bonded interactions, which is found to be $81 \text{ kJ/mol}$ on average with a standard deviation of $14 \text{ kJ/mol}$.

It was suggested that an electronic state with an energy difference of $\geq 136 \text{ kJ/mol}$ compared to the ground state is involved in the dissociation process[78]. The heme would relax into this intermediate state releasing $\geq 89 \text{ kJ/mol}$ in heat. The remaining energy $\sim 38 \text{ kJ/mol}$ exceeding the energy difference between MbCO and Mb-CO was suggested to be deployed as kinetic energy between the heme’s iron and the CO[78].

The potential difference of $81 \text{ kJ/mol}$ found in our calculations is expected to quickly convert into kinetic energy, mainly of the CO and the Fe, but also of the porphorin ring. Subsequently It may thus be expected that the kinetic energy of the CO and the Fe exceed somewhat the experimentally expected value of $\sim 38 \text{ kJ/mol}$, however it should correspond in the order of magnitude. The kinetic energy freed up in the CO bond is thus solely originated form this Lennard-Jones terms. A similar protocol was used before [68, 112] and an additional repulsive term as used elsewhere.
Table 5.3: Simulation settings after dissociation of CO. The timescales of the couplings are: Berendsen: $\tau_P = 5.0$; Parrinello-Rahman: $\tau_P = 1.0$ and V-rescale: $\tau_T = 5.0$.

<table>
<thead>
<tr>
<th>simulation time</th>
<th>5ps</th>
<th>5ps - 45ps</th>
<th>45ps - 5045ps</th>
<th>5045ps - 30000ps</th>
</tr>
</thead>
<tbody>
<tr>
<td>time-step</td>
<td>0.5fs</td>
<td>4fs</td>
<td>4fs</td>
<td>4fs</td>
</tr>
<tr>
<td>$P$ coupling</td>
<td>-</td>
<td>Berendsen</td>
<td>Parrinello-Rahman</td>
<td>Parrinello-Rahman</td>
</tr>
<tr>
<td>$T$ coupling</td>
<td>-</td>
<td>V-rescale</td>
<td>V-rescale</td>
<td>V-rescale</td>
</tr>
<tr>
<td>output frequency</td>
<td>1fs</td>
<td>1ps</td>
<td>10ps</td>
<td>100ps</td>
</tr>
</tbody>
</table>

[113] is not needed.

In this work, to ensure the corrected total energy to be deployed, additional steps follow. (ii) The potential difference due to the non-bonded interactions was compared to the expected excess energy (128.2 kJ/mol). If the earlier difference exceed the later value, the corresponding frame is not considered. However, this happens in less then 1% of all cases. Otherwise the velocities of the porphorin ring are scaled up to match the total excess energy. On average 41 kJ/mol are added this way.

The protocol used is only able to reproduce the order of magnitude of the CO kinetic energy. The dynamics of the first fs are thus to be interpreted with care. For longer timescales the initial kinetics of the CO is reported to be of modest relevance[72]. The total energy deployed into the system can be expected to accurately match the experiment.

For the first 5 ps the dissociated system was calculated deterministically without pressure and temperature coupling to model the heat dissipation. For the following 40 ps Berendsen barostat was used to avoid oscillations while cooling the heated system, and after 45 ps the system was simulated as a NPT ensemble using a Parrinello-Rahman barostat. The corresponding settings are summarized in table 5.3.

5.4 Branches

For the calculation of the TR-WAXS pattern according to (3.62) and (3.61), an ensemble of structures $\mathcal{G}$ corresponding to the ground state and a number of ensembles $\mathcal{E}_t$ corresponding to the system at different delay times $t$ after the excitation
5 Computational Details

Figure 5.2: A: An ensemble of excited protein structures $\mathcal{E}_t$ is generated by starting numerous independent MD simulations (branches) from a single long trajectory. The ensemble correspond to the ground state $\mathcal{G}_0$ is directly taken from this. B: The ensemble of excited proteins $\mathcal{E}_t$ is generated by the same branches then in A, for the ground state however, a delay time depended ensemble $\mathcal{G}_t$ is now generated like the excited one. Structures of both ensembles corresponding to the same delay time are compared.

are needed.

To obtain the latter, an ensemble, including structure and velocities, of the ground state $\mathcal{G}_0$ is taken (i), then (ii) for each instance the force field is changed into the excited version, i.e., the CO bond is removed (see Sec. 5.3) and a MD simulation is started (iii).

Fig. 5.2.A shows how MD simulations of Mb-CO are (with exceptions see Sec. 5.3) branch off at equal time intervals a single trajectory of MbCO. The ensemble $\mathcal{E}_t$, corresponding to a delay time of $t$ in the experiment, is then taken from the trajectories of all branches at delay time $t$ after branching off.

A straight-forward way to for the calculation of difference scattering pattern would be to use the ensemble $\mathcal{E}_t$ for the excited state and the ensemble $\mathcal{G}_0$ from the initial trunk trajectory for the ground state (see Fig. 5.2.A). An alternative way of sampling would be to calculate for each branch with force field in the excited version a second branch with the force field in the initial version, obtaining for each delay time two ensemble $\mathcal{G}_t$ and $\mathcal{E}_t$ (see Fig. 5.2.B). As shown in Fig. 5.2, for the calculation of the difference scattering patterns, frames from both branches with corresponding delay times are used.

If $\mathcal{G}_0$ corresponds to the equilibrium ensemble, $\mathcal{G}_0$ and $\mathcal{G}_t$ are equal and the is no difference between the two approaches. As shown in Sec. 5.1, however, for myoglobin it was not possible to obtain a perfect equilibrated ensemble even on microsecond timescales. As a result, $\mathcal{G}_0$ and $\mathcal{G}_t$ were not identical. In comparison to $\mathcal{G}_0$ the additional simulation time of the branch improves the the equilibration of the $\mathcal{G}_t$. 

62
ensemble. This difference is important for the calculation of difference scattering pattern.

To check this effect from a single simulation of 1 \(\mu\)s, 500 times three branches were diverted. One with the CO bonded to Mb \(\mathcal{G}_t\) and two with CO diffusing free, denoted \(\mathcal{E}_t\) and \(\mathcal{E}'_t\). Velocities were recalculated in one of the latter cases to prevent numerical identity. Fig. 5.4 shows the root mean squared of the difference intensity from comparing ensemble \(\mathcal{G}_0\) and \(\mathcal{G}_t\) in red and from ensemble \(\mathcal{E}_t\) and \(\mathcal{E}'_t\) in green. In both case ensembles corresponding to the same force field are compared, thus for perfect equilibrated systems the difference intensity should be zero. Indeed, with increasing number of branches considered (x-axis) the intensity decreases. There is no difference to be seen for a delay time of 3 ns. After 25 ns, however, the intensity in the \(\mathcal{G}_0\) and \(\mathcal{G}_t\) case converges to a higher level.

This bias can be related to the difference in the sampling. Consequently one can expect that the difference pattern calculated from \(\mathcal{G}_0\) and \(\mathcal{E}_t\) includes features which do not only related to the dissociation process, but also to the different sampling quality. To avoid the latter contribution, always two branches were simulated in this work, one \(\mathcal{E}_t\) corresponding to the excited state and one \(\mathcal{G}_t\) corresponding to the ground state, and difference scattering pattern are calculated from these.

For the scattering pattern presented in this work, MD simulations have been started from 1300 starting structures, two branches each. An overview of the used simulations, which accumulate to 100 \(\mu\)s, are given in Table 5.4. All scattering pattern presented in this work have been calculated as depicted in 5.2.B, using two branches.

When scattering pattern were to be calculated form different cluster, for each
5 Computational Details

Figure 5.4: The root mean squared scattering intensity comparing either two bonded ensembles as depicted in Fig. 5.3.A (filled circles) or two non-bonded ensemble obtained as depicted in Fig. 5.3.B. The horizontal axis denotes the number branches used for calculating the mean.

frame in a cluster of the non-bonded simulations, exactly the corresponding frame from a bonded simulation was considered; thereby eliminating a possible bias due to different sampling for these calculations as well.

5.5 Rotational diffusion constant

The rotational diffusion constant can be calculated directly from a trajectory using the discretized version of (3.81), here shown for the diffusion tensor $D_{ij}$

<table>
<thead>
<tr>
<th>batch</th>
<th>trunk</th>
<th>trunk window</th>
<th>#branches</th>
<th>length</th>
<th>constrains</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>single</td>
<td>5ns to 1µs</td>
<td>200</td>
<td>100ns</td>
<td>no</td>
</tr>
<tr>
<td>B</td>
<td>single</td>
<td>1µs to 2µs</td>
<td>500</td>
<td>30ns</td>
<td>no</td>
</tr>
<tr>
<td>C</td>
<td>REMD</td>
<td>25ns to 250ns</td>
<td>500</td>
<td>20ns</td>
<td>no</td>
</tr>
<tr>
<td>D</td>
<td>REMD</td>
<td>5ns to 250ns</td>
<td>10000</td>
<td>95ps</td>
<td>no</td>
</tr>
<tr>
<td>E</td>
<td>constrained</td>
<td>5ns to 250ns</td>
<td>100</td>
<td>20ns</td>
<td>backbone</td>
</tr>
</tbody>
</table>

Table 5.4: Overview over the batches of simulations performed including the simulation initial structures where taken from (trunk), which part the of these was used (t. window), the number of branches, their length and whether constrains where used.
5.6 Clustering

\[ D_{ij} = \frac{1}{2\Delta t} \int \Theta_i \Theta_j p(\Theta, \Delta t) d\Theta \]  
\[ = \frac{1}{2T} \sum_t \Theta_i(t_k) \Theta_j(t_k) \]  

with \(\Theta_i(t_k)\) denoting the rotation from time-step \(k\) to time-step \(k+1\) and \(T\) the total timespan analyzed. The eigenvalues of the diffusion tensor \(D_{ij}\) are the three rotational diffusion constants for the principal axes of the system.

In our simulations the diffusion constants \(D_1 = 0.027\) ns\(^{-1}\), \(D_2 = 0.005\) ns\(^{-1}\) and \(D_3 = 0.004\) ns\(^{-1}\) were obtained for myoglobin. This corresponds to rotational diffusion time-constants (3.86) of 6 ns, 37 ns respectively 41 ns. Experimentally constants of \(\sim 15\) ns [63] within a SAXS experiment and 15 ns[114] respectively 15-20 ns[115] in NMR measurements are obtained.

Considering the shape of Myoglobin being roughly the one of an oblate ellipsoid, one would expect a single fast and two slow rotational timescales, which is exactly what we found. For the anisotropic solution scattering measurements of myoglobin the rotation around the heme’s normal is arbitrary and the rotation perpendicular most relevant for the decay of anisotropy. The angles between the heme’s normal and the principal axes are 66\(^\circ\), 57\(^\circ\) respectively 43\(^\circ\). The first diffusion constant is thus not only the fastest, but also the one most relevant for the scattering pattern considering its axis.

From the computational data a decay of the anisotropy in the order of the fastest time-constant, rather then in the order of the other two larger time-constants is expected, in agreement with the experimental data. However, the obtained timescales indicate, that the diffusion process should be described by a set of diffusion constants rather then by a single one.

5.6 Clustering

The calculation of scattering pattern for a single structure has little meaning as it would primarily be governed by highly flexible groups e.g. side chains. In this work scattering features of particular structural confirmations are to be elaborated by clustering the data first and then calculating an average scattering pattern corresponding to each cluster.
5  Computational Details

5.6.1 General remarks

Clustering can be based on structural considerations. The interest is however often in the relevant timescales. These are governed by the transitions over energy barriers and a special separation may not necessarily resemble a separation based on the energy landscape. Thus, an alternative clustering may differentiate between conformations separated by an energy barrier, and, thus, forms cluster which contain only comparably low internal energy barriers.

In an ideal case of such a model, the probability of a transition from a conformation in one cluster to a conformation in another cluster, does only depend on the cluster involved, but not particular conformations (at least on relevant timescales). Such a model, in which the probability of a transition from one state to any other does not depend on the realization within the state is called a Markov model. Within the theory, this premise is formulated such that the probability of transitions only depend on the current state, but not on the future or the past of the system.

In this work, the data was clustered using tools developed for the construction of Markov models. These tools cluster not only based on separation in space, but also make use of the available information about the dynamics of the system within the trajectories. Markov models offer a simplified description of the dynamics. However, in the context of this work, we only aim for a smart clustering algorithm.

Constructing Markov models from raw data is usually done in two steps. First (i), the raw data is assigned to micro-states based on a given metric. Using their transition matrix, the micro-states are then (ii) merged to a smaller set of macro-states, such that the dynamics can still be described as Markovian. This is possible, if the energy barriers between different confirmations within the macro-states are small compared to the barrier between the macro-states. Then, the probability density of the confirmations in each state can assume to equilibrate within the timescales relevant for transitions between the macro-states and, consequently, the dynamics can be described by a Markov model.

In this study the diffusion of CO can be assumed to be relevant for structural changes. Thus, the trajectories of batch B (see Tab. 5.4) where clustered based on the CO position. The metric used in this work is the RMSD of the CO atoms after fitting the backbone atoms onto a reference structure.
5.6.2 Clustering into Micro-states

The trajectories were clustered into micro-states using an approximate k-center algorithm as implemented in the MSMbuilder package [116]. The algorithm iteratively aims for the minimal and optimal subset of states (here corresponding to the CO position), such that each state of the full dataset, is within a given radius \( \xi \) to at least one state of the subset. In this case, distances were measured in the RMSD and the threshold \( \xi \) corresponded to a RMSD of 0.1 nm.

The threshold \( \xi \) should be small enough to separate important states. On the other hand, it is important that each micro-state contains several snapshots of the trajectories to ensure that the transition matrix is based on a sufficient statistics.

It is thus advisable to use the algorithm on a fraction of the full trajectories and the remaining frames are then added to the respectively closest of the micro-states[117]. Especially, if the phase space is sparsely occupied in some areas (like in the bulk with CO freely diffusing in the solvent) the statistic of the states within this area is usually bad with almost every confirmation corresponding to a micro-state initially.

In such cases, applying additional data to a fixed number of sparse micro-states can help to obtain sufficient statistics. Here all frames with CO being separated from the backbone by more than a specific distance are lumped into a single micro-state to improve the kinetic interconnection of those micro-states corresponding to the CO outside of the protein.

With the described method a total of 1.6M frames from 500 trajectories a 31 ns where clustered into 1686 micro-states initially using 5% of the data and then adding the rest.

5.6.3 Clustering intoMacro-states

Subsequently, the micro-states were merged into macro-states using a more robust version of the Perron cluster cluster analysis (PCCA) algorithm called PCCA+ [116, 118]. The main idea of PCCA is to iterative split the system into macro-states based on the eigenvectors of the transition matrix with the largest eigenvalues[119].

The transition matrix is calculated from a single trajectory by comparing the micro-state of the frame time \( t \), with the micro-state at time \( t + l \), with \( l \) being the lag-time (here \( l = 200 \text{ps} \)). A joint transition matrix is obtained by accumulating the once of all individually trajectories.

The largest eigenvalues correspond to the slowest transitions. Therefore PCCA
splits the system into macro-states which still resample the slowest dynamics of the system.

In the specific case of the diffusion of CO within myoglobin, there are number of pockets, into which the CO diffuses within slow timescales, which show, however, only low occupation and are thus of little relevance for the scattering pattern. The important macro-states can thus only be resolved using a larger number of macro-states. In practice, we initially derived 16 macro-states. Only 6 of these macro-states had an occupation of more then 1% averaged over the analyzed timespan, and most other macro-states an occupation of less then 0.1%. The latter are then merged with the one of these 6 main states, which they are closet connected to, judged from the transition matrix. Finally 6 macro-states as depicted in Fig. 5.5 are obtained.

The number of 16 initial macro-states is derived from a screening including a clustering using 5 to 25 macro-states. For 5 to 9 macro-states three of them exceed the threshold of 1%, for 10 and 11 there are four, for 12 to 15 there are five and for 16 to 25 there are six. This suggest that with six main macro-state some saturation
5.6 Clustering

is reached.

These macro-states are also highly consistent throughout the screening. The same procedure have been used for slower lag-times without substantial differences. The same 6 macro-states are contained in a model of 17 macro-states with a lag-time of 50 ps and a model with 18 macro-states with a lag-time of 20 ps.
6 Results and Discussion

Myoglobin has been extensively studied, both experimentally and numerically. In the following two sections, the dynamic of the MD model will be illuminated. The evolution the occupation of different clusters allows one to draw conclusions about relevant timescales, and the evolution of the electron density difference can be compared to crystallographic results. The remaining main part of this chapter is focused on the interpretations of (anisotropic) scattering pattern. In the following, scattering patterns ranging form femtosecond to nanosecond time-resolution will be discussed and put into relation to difference maps of the electron density. The calculated scattering patterns are also compared to experimental data.

6.1 CO diffusion

6.1.1 General remarks

After dissociation from the heme, CO is known to defuse through multiple pockets of myoglobin. Some of these pockets were first observed as binding xenon and are hence named Xe1-Xe4 with increasing xenon affinity [83].

In the simulations, CO was found in the Xe1 (red), Xe2 (blue) and Xe4 (orange) pocket (see Fig. 5.5.A), but not in the Xe3 pocket. Instead, the CO diffused rather freely in its vicinity within the protein (pink) in a few instances (maximal 3% of all trajectories). The distal (green) and the bulk (yellow) state complete the six marco-states, which were obtained in the clustering described in Sec. 5.6.

CO has been detected within the distal and the Xe1 pocket in crystallographic studies [53, 85, 104]. Occupation of the Xe2 and the Xe4 pocket has not been observed directly in wild-type myoglobin; however, it has been observed in a mutant [52, 53]. To our knowledge, there is no experimental evidence for an occupation of the Xe3 pocket. MD simulations of crystals have shown occupation of the the Xe3 cavity[71, 120]. However, in those simulations the Xe1, Xe2, Xe4 and distal
6 Results and Discussion

pockets remain the most relevant. In our simulation the distal and the Xe1 pocket showed the highest occupation, together accounting for more than 60% within the observation range (>30ns).

In ~ 14% of all trajectories CO was observed to leave the protein. No reverse entry of CO were observed. In ~ 20% of all instances the CO left the protein from the distal pocket, in ~ 30% the CO left from the Xe2 pocket and in ~ 40% of the cases the exit was from the Xe1 pocket. In a few instances, escape through the other pockets have been observed. This results corresponds somewhat with numerical results based on a partially frozen deoxydated structure estimating an escape from the distal and Xe1/Xe2 pocket by probabilities of ~ 12.5% and ~ 42.5% [73], respectively.

There has been a debate about the CO escape pathways [121]. Extensive mutation studies suggest that the distal pocket is the only entry and exit pathway [122]. However, other studies suggest, alternatively, multiple escape pathways [73, 123, 124], which are consistent with our findings. An early model includes four states [125] and assumes an escape pathway though the consecutive passing though three intermediate states B, C, D before leaving the protein into the solvent S. Here A denotes the bonded state MbCO. The four-state model is still used for the interpretation of experimental timescales [55, 56].

Our data, however, does not support the four state model as ligand exit was observed from all pockets. All studies agree, that the CO diffuses between different cavities after dissociation and before escaping into the solvent.

6.1.2 Kinetic model

Fig. 5.5.B shows the occupation of the six states for the first 30ns after dissociation. Initially, the CO is located in the distal pocket (green). Within hundreds of picoseconds, the Xe4 (orange) pocket is filled reaching its maximum occupation (8% of all trajectories) after 0.9 ns. An occupation of the Xe1 (red) and Xe2 (blue) pocket can be observed within nanoseconds. The occupation of the Xe2 pocket was initially exceeding the one of the Xe1 pocket until 2.1 ns. An almost constant escape of the CO from the protein can be observed (yellow). The diffusive state (pink) plays only a minor role.

The occupancies presented in Fig. 5.5.B were fitted to a kinetic model. The model obeys the differential equation $\dot{X} = KX$. From the transition matrix of the underlying Markov model used for clustering (see Sec. 5.6) relevant transitions were
selected. Here, the vector $\mathbf{X}$ represents the population of the different states and, the matrix

$$
\hat{K} = \begin{bmatrix}
-k_{4D} - k_{42} - k_S & 0 & k_{D4} & k_{24} & 0 \\
0 & -k_{F2} - k_S & 0 & k_{2F} & 0 \\
k_{4D} & 0 & -k_{D4} - k_S & 0 & 0 \\
k_{42} & k_{F2} & 0 & -k_{24} - k_{2F} - k_{21} - k_S & k_{12} \\
0 & 0 & 0 & k_{21} & -k_{12} - k_S \\
\end{bmatrix}
$$

defines transitions between the different states with $k_{AB}$ indicating the transition from state $A$ to state $B$. The components of the vector $\mathbf{X}$ corresponds to the macrostates in the order Xe4, diffusive, distal, Xe2, Xe1 and are shortened to 4,F,D,2,1 in the indexes. To simplify the model we used a single rate $k_S$ for the escape from all pockets into the solvent. The obtained rates are visualized in Fig. 5.5.C.

The timescale for the transition from the distal into the Xe4 pocket is $1/k_{D4} = 1.7\text{ ns}$. The reverse transition was found to occur on faster timescales in the order of hundreds of picoseconds, resulting in a modest occupation of the Xe4 pocket. MD simulations of crystalline MbCO showed a similar timescale of 1.1\text{ ns}[71] for the forward transition. In those simulations, however, the backwards transition into the distal pocket was found on a similar timescale, resulting in a considerable larger occupation of the Xe4 pocket. MD simulations of MbCO reveal a timescale of 4-5\text{ ns}, however using a weak statistics [70].

Experimentally, there has not been found a consistent timescale for transition between the distal and the Xe4 pocket, which one may identify with the B - C transition of the four state model. Indirect measure estimate it with 180\text{ ps}[125]. Other studies are missing any timescale of similar length[126]. In time resolved crystallography, the occupation of the distal pocket decay within a nanosecond to 45% without leaving any further trace of the remaining CO[85]. This matches with our and earlier numerical results [71], suggesting a disorder of CO between the distal and the Xe4 pocket within a nanosecond.

The further transitions into the pockets Xe2 and Xe1 occur on timescales of $1/k_{42} = 1\text{ ns}$ and $1/k_{21} = 1.1\text{ ns}$, receptively. Numerical results for the later transitions have not been reported. However we found ratio of $\sim 1 : 2$ between the population of Xe1 and Xe2, which is in line with an earlier reported one of $\sim 1 : 3$ in a simulation of a crystal [71].

The implied timescales of the dynamics, represented by the transition matrix’s
Figure 6.1: Comparison of the electron density difference (green positive; red/pink negative) 100ps after dissociation obtained numerical (A; this work) and experimentally (B; [53]). Differences in the electron density can be observed in the area of the heme, the E Helix and the B Helix. Most prominent is a negative feature at the former binding site of CO (CO) and a positive feature in the distal pocket (CO*). The molecules are oriented with the NA and NB nitrogens of the heme in the front.

Eigenvalue composition, may be more suitable for a comparison with experimentally measured timescales. The smallest non-zero eigenvalue of $\hat{K}$ is $-0.08 \text{ns}^{-1}$ corresponding to a timescale of 12.5 ns. Its eigenvector describe the transport into the Xe1 pocket.

Experimentally, timescales of $\sim 50 \text{ ns}$ [127], $\sim 70 \text{ ns}$ [55] and $180 \text{ ns}$ [126] have been reported, which may match with the transition into the Xe1 pocket. Transition into the Xe1 pocket on a tenth of nanoseconds scale have been shown in time-resolved crystallographic studys. In those studies, 65% of all dissociated CO has been found in the Xe1 pocket after 100 ns, and the crossover point between the occupation of the distal and the Xe1 pocket was found after $\sim 20 \text{ ns}$ [85]. This matches well with the crossover point found in this study after 12ns. Transition from the Xe2 pocket into the diffusive state follow a time constant of $\geq 125 \text{ ns}$ and the escape into the solvent is on a timescale of $\sim 200 \text{ ns}$. For the latter, timescales of 700 ns have been reported [126, 127], somewhat matching our result.

74
6.2 Comparison with crystallographic data

Time-resolved crystallographic structures have been resolved allowing for comparison with the here obtain simulation-based data.

For the numerical calculation of electron density maps, for each time delay, 10000 frames of batch C (see Tab. 5.4) were fitted in respect of their backbone and a three-dimensional histogram of number densities of all atom types individually was calculated on as 0.05 nm spaced grid. To account for the different atom sizes, the density map of each atom type was filtered by a gaussian with a sigma corresponding to its Van-der-Waals radius. The number densities were then scaled according to the atomic number and accumulated to a total electron densities. A difference map was obtained by subtraction of correspondingly generated density maps in the bonded and the non-bonded case.

Electron density difference maps from simulation for time delays of 100 ps and 1 ns are shown in Fig. 6.1 and Fig. 6.2 respectively alongside with experimental maps from two different time-resolved crystallographic studies[53, 85].

Isosurfaces of the numerical results for an excess respectively a deficiency of 240 e·nm⁻¹ are visualized in green and red. Hence, movements of the protein structure are expected from red to green.
Most prominent at both delay times is the vacancy left by the CO-labeled CO and a new feature labeled CO* presumably corresponding to the photo-dissociated CO. This new feature is located about 0.15 nm apart from the oxygen of the CO in the bonded structure, which corresponds to the experimentally found distance of 0.218 nm [85].

Apparent in the numerical maps already at 100 ps but more pronounced at 1 ns, is a movement of the entire heme group. The heme group is moved downwards, tilded down at the side of the NB nitrogen and additionally shifted into the latter direction. Negative features above the heme group and positive features below indicate this movement. The movement of the heme group pushes the F helix downwards, mediated through the His93 residue.

The same movement of the F helix was observed in crystallographic experiments [53, 85]. In the experiment, however, it was related to a displacement of the iron out of the plane of the heme. A displacement of the Fe downward of 0.03 nm [128] respectively 0.036 nm [129] was observed experimentally. In the crystallographic density maps of Schotte et al. [53] (see Fig. 6.1.B) this displacement manifest itself in a positive feature below the original iron position. There a tilt around the NB/ND nitrogens and a dooming of the heme group pushing its outer parts upwards was found.

In the simulation, on the other hand, the heme remains planar, resulting in a clear discrepancy between experiment and simulation. In experiment the rise of the hemes porphyrin ring is dragging Leu104 and Leu107 and the attached G helix upwards, which is not seen in the simulation.

There is agreement between experiment and simulation in the movement of His64 and Val68 in the direction of the vacancy left by the CO. This mediates a downwards movement of the E helix (see Fig. 6.1 and 6.2). Another consistent finding is that the CO in the newly occupied distal pocket (CO*) pushes Leu29 and the B helix upwards.

### 6.3 Time evolution of electron density and scattering pattern

In the following the time evolution of electron density of myoglobin as observed within the MD simulations will be presented. Thereby the change in electron density in real space is presented along with the corresponding anisotropic TR-WAXS
6.4 Femtoseconds to picoseconds

In batch D (see Tab. 5.4) 10000 branches of 95ps have been calculated. For femtosecond time-resolution, averaging over different confirmations of the same branch is not feasible because of the high autocorrelation of atomic positions on this timescale. For a sufficient sampling, a high number of branches are thus needed.

The density maps as well as scattering patterns have been calculated taking a single confirmations from each branch. Fig. 6.3 show the density maps in the femtosecond range, which were calculated as described in the Sec. 6.2. Here, the density was, however, smoothed with a Gaussian filter with a width of $\sigma = 0.05 \text{ nm}$ for the solid surface and $\sigma = 0.2 \text{ nm}$ for the transparent surface. In the reciprocal image this would damp contributions of momentum transfer larger then $q > 20 \text{ nm}^{-1}$ respectively $q > 5 \text{ nm}^{-1}$. While the solid surface represent the maximum resolution of WAXS experiments, the translate surface should give an impression of the density difference measured in the SAXS regime. The iso value is $5.6 \text{ e nm}^{-3}$ for the transparent surface and $200 \text{ e nm}^{-3}$ for the solid ones.
Figure 6.4: WAXS pattern of myoglobin with fs time delays, both (A) radial averaged for circular polarized excitation laser and (B) the horizontal/vertical difference for a polarization perpendicular to the X-ray beam, are depicted.

The radially averaged (RA) scattering pattern for a circular polarized excitation laser and the horizontal-vertical intensity difference (HV) for a laser polarization perpendicular to the X-ray beam are shown in Fig. 6.4. The RA pattern are calculated using the right column of the fifth row of tab. 3.1 and contains the least anisotropic contribution. The HV patterns are calculated using the left column of the fourth row of tab. 3.1 and contains the anisotropic component only. Therefore these two components are suitable for examine anisotropic scattering pattern.

The first feature to emerge in the density maps is a decrease in density between the bonded position of the CO (marked CO) and the heme and a increase of the density above the CO (marked CO*), which is forming a mushroom shape within the first 80 fs. This clearly originated from the CO being dissociated and shooting upwards. Both features increase exclusively up to a time delay of 80 fs. Simultaneously, a
6.5 One picosecond to one hundred picosecond

peak at \( q \approx 7.5 \text{ nm}^{-1} \), negative in the RA patterns and positive in the HV patterns, arrises (marked CO). This peak reaches its minimum (most negative) in the RA pattern at 80–160 fs, when the CO vacancy has its largest size, presumably formed by the CO, which is, accelerated by the dissociation, kicking surrounding atoms away. It is then constant from 320 fs on, on a less negative level, when supposable the initial kinetics of the CO dissipated into the protein.

After 160 fs, the red translucent isosurface increases significant and even more after 320 fs. The decrease in density in this area can be attributed to a shock wave or a locally increased temperature of the protein around the heme. IR[78] and Raman[47] spectroscopy as well as MD simulations [75] suggest heat dissipation on two timescales one in the range of picoseconds and a second in the range of tenth of picoseconds. Form the density map one can not differentiate between a collective and a random motion. The fast arise and decline of the feature argues for the collective motion of a shock wave.

The shock wave manifest itself in the very low \( q \)-range with a step peak in both pattern. Interestingly, the peak in the HV pattern is delayed. This may be attributed to the shape of the pressure wave starting spherical at 160 fs and thus missing a anisotropic contribution, but then dissipating mainly in horizontal direction at 640 fs.

A local displacement of the E and B helix upwards, as observed at time delays between 160 fs and 640 fs, may also be a result of the initial kinetics. After 1 ps the shock wave vanish in real space (see Fig. 6.3) as well as the corresponding feature in the scattering pattern (see Fig. 6.4).

In the 640 fs snapshot, the positive CO* feature moved from above the negative CO vacancy, to a the right, occupying a pocket below Leu29. The RA pattern shows now a double-well with a new minima at \( q \approx 4 \text{ nm}^{-1} \) and a new maxima at \( q \approx 6 \text{ nm}^{-1} \), which are correspondingly marked as CO* in Fig. 6.4. Both pattern show modest derivations at \( q \)-values higher then \( q \approx 10 \text{ nm}^{-1} \) at femtosecond times, which are difficult to trace back to specific rearrangements.

6.5 One picosecond to one hundred picosecond

Within the following 100 ps four different structural changes appear almost simultaneously in the electron density maps (see Fig. 6.5). This are (i) a downward movement of the E helix, (ii) a upward movement of the B helix, (iii) a downward
6 Results and Discussion

Figure 6.5: Maps of the difference electron density as in Fig. 6.3, here however for time delays between 1ps and 90ps.

movement of the heme and (iv) a downward movement of the F helix.

In the scattering patterns (see Fig. 6.6), eight features (I-VIII) can be observed in the same time span. The first well (I) decreases considerably between 1ps and 2ps and fluctuates then in its depth. The peak (II) and the second well (III) decreases (becomes more negative) in the discussed timespan. In the HV signal there is a decrease in the amplitude for \( q < 6 \text{ nm}^{-1} \) (VI) and an increase of the peak at \( g \approx 7.5 \text{ nm}^{-1} \) (VII). The RA scattering patterns show a constant decrease in intensity in the range for \( 9 \text{ nm}^{-1} < q < 12.5 \text{ nm}^{-1} \), which results in a feature (IV) similar to the first hermite function at \( g \approx 12.5 \text{ nm}^{-1} \). In the signal corresponding to higher momentum transfer a lot of fluctuations appear. Between 20ps and 50ps the scattering intensity at \( q \approx 18 \text{ nm}^{-1} \) is considerable above average (V), however this trend does not appear at 90ps. In HV pattern a significant decrease in intensity at \( 17.5 \text{ nm}^{-1} \) (VIII) can be observed during the here discussed timespan.

Assigning certain features in real space to their counterpart in the scatter pattern
is difficult form the data provided, as numerous changes happen simultaneously. In general, the shape of the HV pattern appears much more consistent even for higher q-values, while the RA pattern shows large fluctuations. Relative movements of two helixes may result in changes of the distance of numerous atoms in proximity possibly causing this fluctuations. It can be recorded that the most prominent features of the scattering pattern already appeared within the first picoseconds.

6.6 Picoseconds to nanoseconds

6.6.1 Electron density maps
Figure 6.7: Maps of the difference electron density as in Fig. 6.3, here however for time delays between 100 ps and 50 ns. The positive electron density difference of the CO only are included as blue wireframes. Here the transparent and solid isosurfaces represents electron density differences of $11.2 \, \text{e nm}^{-3}$ respectively $280 \, \text{e nm}^{-3}$.
6.6 Picoseconds to nanoseconds

Electron density differences from 500 branches of batch C (see Tab. 5.4) are shown in Fig. 6.7, with increased isosurfaces compared to the earlier presented maps (Fig. 6.3/6.5), are presented for time delays between 100 ps and 10 ns. Therein also time delays of 30 ns and 50 ns are presented, calculated from batch B and A, respectively (see Tab. 5.4).

Interestingly, with the different scaling, the density maps almost match the ones on the picosecond scale (compare Fig. 6.5 and 6.7). The same features can be observed; (i) a down movement of the E helix, (ii) an upward movement of the B helix, (iii) a downward movement of the heme and (iv) a downward movement of the F helix. These features, however, seem to saturate on different timescales.

The downward movements of the B helix (iii), for instance, occurs mainly up to a delay time of 3 ns. At 50 ns this feature decreases considerable. This observation clearly matches with the occupation of the Xe4 pocket (see Fig. 5.5). The positive contribution of the difference electron density arising from the CO only is represented by blue wireframes in Fig. 6.7, directly showing how the decrease of occupation of the Xe4 pocket leads to the decay of the (iii) feature.

After 3 ns, CO reaches the Xe1 pocket and simultaneously the F helix is pushed down. This may be a consequence of the space needed by the CO between the heme and the F helix. The CO in the Xe1 pocket can be seen as solid green surface after 10 ns and 30 ns, but not after 50 ns. Occupations for this delay time have not been calculated, but as more and more CO escape into the solvent occupation of the Xe1 pocket is expected to decay. After 50 ns, the F helix seemed to move partly back to its original position due to this decaying occupation.

The movement of the heme and the E helix reaches is maximum after 10 ns and little changes can be observed afterwards. To judge whether this observation can be traced back to the diffusion of CO, density difference maps are shown in Fig. 6.8 for three different timespans, calculated from frames with the CO in the distal pocket only (see Fig. 6.8).

These maps show the same movement of the heme and the E helix compared to Fig. 6.7, calculated considering all frames. However, after 25-30 ns no decay of the displacement of the B helix can be observed and, in contrast to the maps calculated from all frames (see Fig. 6.7), the displacement of the F helix does not further increase between a delay time of 10 ns and 30 ns. This indicates, that the displacement of the heme and the E helix had their origin in relaxations processes independent from the CO diffusion and that the displacement of B and F helixes
are actually a consequence of the CO diffusion.

At 30 ns and 50 ns additional a increased density can be observed at a position, where no protein atoms can be expected (marked SOL in Fig. 6.7/6.8). Thus, here a change of the solvation layer can be observed. The increase in density can not be attributed to CO escape, as the same feature can be seen in the difference maps calculated from distal pocket frames only.

One may also note that although there is a considerable density arising from the CO at different areas within the protein at the later times (indicated with the blue wireframe in Fig. 6.8), only the CO occupation in the distal and the Xe1 pocket are also represented in the in total electron density difference (indicated by green solid surfaces). This matches with crystallographic, studies which solely found contribution of the CO arising in these two pockets [53, 85] and support the therein made argument that other occupied pockets may be not detectable due to disorder.

### 6.6.2 Radially averaged WAXS pattern

The numerically obtained RA and HV scattering pattern show only small changes between 100 ps and 20 ns. This corresponds to experimental patterns [55, 56], in which main features are already present for the first observed delay time after 100 ps. To emphasise the evolution, RA and HV scattering patterns for time delays up to 20 ns are presented in Fig. 6.9 relative to the scattering pattern obtained after 100 ps. For later delay time no data is available from batch C, which was used for the earlier times. Calculating WAXS pattern for larger times delays from a different dataset, would make it impossible to differentiate between an evolution of the pattern and features arising due to differences in sampling and starting structures. Therefore time delays > 20 ns were here omitted.
Additionally, scattering pattern from the structures of the different cluster (see Fig. 6.10) and from structures with CO in the distal pocket for different timespans have been calculated (see Fig. 6.11); the latter based on the same data then Fig. 6.8. This pattern may allow a judgement wether features are more likely to origin from CO diffusion independent structural changes or from CO diffusion dependent structural changes.

On nanosecond timescales, rotational diffusion becomes important. For the RA pattern in Fig. 6.9 this effect has been included according to (3.88) using a diffusion constant of $0.011 \text{ ns}^{-1}$, corresponding to a timescale of 15 ns. For the HV pattern rotational diffusion changes the pre-factor, but not the relative intensities and it was thus not considered.

The directly calculated RA pattern (see Fig. 6.9.A) shows variation in the intensity...
6 Results and Discussion

Figure 6.10: WAXS pattern obtained from the different clusters, calculated as described in Sec. 5.6.

below $q \approx 4.5 \text{ nm}^{-1}$ (I) in course of time, however no trent can be observed. Both
the scattering pattern form the distal pocket (see Fig. 6.11.A) as well as the different
clusters show variation (see Fig. 6.10.A), prohibiting an interpretation.

The RA pattern of the whole ensemble (see Fig. 6.9.A) shows a distinct decrease
in the intensity at $q \approx 5.5 \text{ nm}^{-1}$ (II), which start to arise for delay times larger than
3 ns. The time-resolved RA pattern of the distal cluster (see Fig. 6.11.A) shows only
small fluctuations here and in the momentum range $5 \text{ nm}^{-1} < q < 14 \text{ nm}^{-1}$. The
RA scattering pattern of the different clusters in Fig. 6.10 reveal a considerably
lower intensity for the Xe1 and Xe2 pockets compared to the other clusters. One
might thus attribute this feature to the increasing occupation of these two pockets.

At $q = 8 \text{ nm}^{-1}$ (III) and $q = 11 \text{ nm}^{-1}$ (IV) there is a increase in intensity starting
somewhat after 1 ns respectively 3 ns, but then with the main contribuiton between
3 ns and 10 ns. Based on the cluster’s scattering pattern (see Fig. 6.10.A), the latter (IV) can attributed to the occupation of the Xe1 pocket, while the origin of the earlier (III) is less clear.

At \( q \approx 15.5 \text{ nm}^{-1} \) (V) the RA pattern features a well, followed by an step peak at \( q \approx 17 \text{ nm}^{-1} \) (VI) and another well at \( q \approx 18 \text{ nm}^{-1} \) (VII). The depth of the first ditch is well manifest after 320 ps and indicates only a modest further decrease, but large fluctuations. Because the second wells (VII) occur similarly in the time-resolved distal pocket’s RA pattern (see Fig. 6.11.A) and there are no large differences in the species associated pattern (see Fig. 6.10.A), this feature may be indepentend of the CO diffusion.

The peak (VI) on the other hand has clear counterparts in the scattering curve of the Xe4 and the Xe1 pocket (see Fig. 5) and the well (V) in the Xe2 pocket, may these features may be a consequence of their respective occupation.

In the very high \( q \)-range (VIII) an increase of the scattering intensity, overlayed with considerable noise, can be observed in comparision to the value after 100 ps. The time-resolved pattern calculated from the distal cluster exhibits a similar increase (see Fig. 6.11.A), suggesting a this feature to be diffusion-independent.

### 6.6.3 Difference pattern from horizontal and vertical cuts

The changes of the HV scattering pattern depicted in Fig. 6.9 is, considered over the whole range of momentum transfer, more regular then the RA pattern and may remind to the tail of a cardinal sine function for the later time delays. In its time evolution, however, several independent features can be identified.

After 320 ps, a minimum, at \( q \approx 3 \text{ nm}^{-1} \) (IX) and a maximum (X) at \( q \approx 5.5 \text{ nm}^{-1} \) appears . At subsequent times, the well deepens and the peak rises and shifts to lower \( q \)-vectors. Simultaneously, a second peak (XI) arrises at \( q \approx 7.5 \text{ nm}^{-1} \). Based on the pattern of the different clusters in Fig. 6.10.B the minima (IX) as well as the maxima (X) can be assigned to the occupation of the xenon pockets (Xe1, Xe2, Xe3).

At higher \( q \)-values, three wells can be observed. A wide one at \( q \approx 10 \text{ nm}^{-1} \) (XII) which arises between 1 ns and 10 ns, a second one at \( q \approx 14 \text{ nm}^{-1} \) (XIII) and a third one at \( q \approx 17.5 \text{ nm}^{-1} \) (XVI). The second well (XIII) can be attributed to the emptying of the distal pocket. In the corresponding scattering pattern (see Fig. 6.10) a peak can be observed for the distal pocket and a well for all other clusters. Neither the time evolution of the distal pockets scattering pattern, nor the one of the
results and Discussion

Figure 6.11: WAXS pattern corresponding to different delay time ranges, obtained from frames with the CO in the distal cluster only. The pattern corresponding to a delay time of 5ns to 10ns was used as a baseline.

Individual cluster give a hint about the origin of the other two wells (XII),(XVI).

Taken together, a number of distinct features where found to arise in both, the RA and the HV patterns. Assigning these features to specific conformational transitions was partly possible, but often the precise causality remained vague. One main reason is the overlap of diffusion-dependent and diffusion-independent dynamics.

6.7 Scattering pattern derived from cluster

By multiplying the scattering pattern of the different clusters (see Fig. 6.10) with the occupation of the respective cluster at given delay times (see Fig. 5.5.B), a time evolution of a scattering pattern can be constructed. Fig. 6.12 presents the
6.7 Scattering pattern derived from cluster

Figure 6.12: A time-resolved WAXS pattern is constructed by multiplying the scattering pattern of the different cluster (see Fig. 6.10) with their individual relative occupation (see Fig. 5.5.B). Short delay times are scaled up in an inlay.

time-evolution of the RA respective HV scattering pattern. The expected scattering pattern after 100 ps has been subtracted in the same way as it was done for the direct calculations. Feature II,III,IV and IV of the RA TR-WAXS pattern directly calculated form the trajectories (see Fig. 6.9.A) are equally prominent in the cluster derived TR-WAXS pattern (see Fig. 6.12.A). However, other features (V,VI, VI) are missing.

In the HV pattern, some features (VII, VIII and IX) can be recognized (compare Fig. 6.12 and 6.9) and other not (XI, X). Interestingly here there is a positive peak at $q \approx 19\text{ nm}^{-1}$ (XV), which is missing in the direct calculation of the scattering pattern (Fig. 6.9).

On nanosecond timescales the constructed patterns appear to consist of a single component only with increasing amplitude. In Fig. 6.12 the pattern for time delays
6 Results and Discussion

Figure 6.13: The RA and HV pattern for the constrained simulation for a time delay of 100ps are shown.

up to 1 ns are additionally plotted with higher detail. On these timescales the scattering pattern clearly consist of multiple independent components. The main features in the period from 100 ps to 1 ns are, however, the same as in the following period from 1 ns to 20 ns and mainly the amplitude is increasing.

6.8 Backbone-restrained simulation

Independent calculation were setup restraining the alpha carbon atoms as well as the hemes porphorin ring. Branches of 20 ns length were calculated form 100 independent starting configurations. Observation of combined trajectories show that the same main pockets (distal, Xe1, Xe2 and Xe4) are occupied on timescales of about 10 ns. Noise in the data does not allow conclusions about the time evolution on nanosecond scales and thus only the scattering pattern for a time delay of 100 ps is shown in Fig. 6.13.

The characteristic double well, (I) and (III), at low q-values and the distinct peak (VI), found in the free simulation, can be observed. This suggest their origin in the electron density difference of the CO or in the side chains in the vicinity of CO binding sides. The distinct feature (IV) of the free simulation (see 100 ps in Fig. 6.14) is missing in the restrained simulation, suggesting that this feature might arise due to a relaxation of the backbone or a movement of the heme in the free simulation.
6.9 Comparison with experimental data

Figure 6.14: Overlay of the RA experimental scattering curve (black) corresponding to a time delay of 100ps[55], its numerical counterpart (blue) and the numerical scattering curve (violet) after 3ps of simulation time. For the experimental data the curve after subtraction of the solvent heat contribution was used (see supplementary data to Ref. [55]). The scale of the experimental curve was chosen to support comparison.

The most prominent feature of the HV pattern is a negative peak (XI), which is in agreement with the free simulation. Other peaks at $q \approx 11$ nm$^{-1}$, $q \approx 13$ nm$^{-1}$ and $q \approx 18$ nm$^{-1}$ correspond as well, however the relative amplitudes differ.

6.9 Comparison with experimental data

TR-WAXS experiments of myoglobin were performed by Cho et al. [56] using myoglobin from sperm whale and Kim et al. [55, 63] using myoglobin extracted from horse heart. The general shape of the scattering pattern of both studies is very similar, yet there are differences in the time evolution. As the protein structure used in this work corresponds to horse heart, and Kim et al. also reported anisotropic scattering data [63], in the following a comparison with their results will be presented.

Figure 6.14 presents the RA scattering curves form our calculation together with an experimental curve[55]. For the experimental curve the contribution of solvent heating was subtracted[55], as the numerical calculations do not account for this effect. The scale of the experimental curves is arbitrary and thus a scale was chosen which facilitates the comparison.

The experimental and the numerical data agree qualitatively concerning a double-well with minimal at $q = 4$ nm$^{-1}$ and $q = 7.5$ nm$^{-1}$, and the latter being the deeper
Results and Discussion

Figure 6.15: Time evolution of the experimental RA WAXS pattern. The numerical data was slightly filtered, horizontal shifted to match the intensity at $q = 20.7\,\text{nm}^{-1}$\[55] and the data obtained for a delay time of 100ps subtracted.

one. However, in the experiment the well’s depth differ only marginally, while in the simulations the the second peak is clearly more pronounced than the first peak. In the simulations, the intensity difference reaches positive values between the two wells, while in the experiment the barrier between the wells is roughly half the height of the wells depth.

The RA pattern from the free simulations show a distinct feature (IV), a well followed by a peak, which starts to emerge from 3ps onwards. This feature is dominant in the pattern for the delay time of 100ps, but it is absent in the experimental data. For this reason, Fig. 6.14 also presents the scattering pattern form the simulation after 3ps, matching better with the experimental data in the range from $q = 9\,\text{nm}^{-1}$ to $q = 15\,\text{nm}^{-1}$, with a small ditch at $q \approx 12.5\,\text{nm}^{-1}$ in both cases. At $q = 17\,\text{nm}^{-1}$ the experimental data show a peak, followed by an minima at $q = 19\,\text{nm}^{-1}$ and a monotone increase afterwards. The simulations show large fluctuations for this frequencies, however a peak at $q = 17\,\text{nm}^{-1}$ and a minima at $q = 19\,\text{nm}^{-1}$ is consistent with the scattering curve after 90ps.

To compare the time evolution of the scattering patterns, scattering pattern for the experimental data, in which the contribution prior 100ps time delay is subtracted, are plotted in Fig. 6.15. Prominent is a decrease in the intensity at $q \approx 7\,\text{nm}^{-1}$. In our simulation this feature is missing. Instead, there is a steep decrease (II) at $q \approx 5\,\text{nm}^{-1}$ (see Fig. 6.9). Some agreement (comp. Fig. 6.15 and Fig. 6.9) can be
6.9 Comparison with experimental data

Figure 6.16: The experimental (A) [63] and numerical (B) HV scattering pattern of myoglobin. For experimental data both is shown, the raw data in the background and the one filtered with a Gaussian. The numeric delay times differ slightly by about $\sim 10\%$ from the experimental once shown in the legend.

observed for momentum transfers between $q \approx 9\,\text{nm}^{-1}$ and $q \approx 13\,\text{nm}^{-1}$, showing a increase of intensity on nanosecond timescales, which is more pronounced at feature III and IV. The largest change in the experimental data occurs at $q \approx 17\,\text{nm}^{-1}$, with a broad increase until 50 ns. Here the simulations show a prominent peak (IV), which is however embedded into two wells (V/VII).

Anisotropic TR-WAXS pattern have been measured [63]. Figure 6.16.A shows the HV scattering pattern for different delay times. The raw data (thin lines in the background) show large fluctuations. We filtered the data by a convolution with an Gaussian (thick lines) to unveil the underlying signal. The HV pattern obtained from simulations are shown in Fig. 6.16.B for the same q-range. Here, rotational diffusion is taken into account according to (3.88).
6 Results and Discussion

Qualitative agreement between simulation and experiment can be observed. In both cases we observe maxima at \( q \sim 4 \text{ nm}^{-1} \) and \( q \sim 7.5 \text{ nm}^{-1} \), with the latter being the global one and minima at \( q \sim 2.5 \text{ nm}^{-1} \) and \( q \sim 5 \text{ nm}^{-1} \), with the first being the global one. There is a third minima at \( q \sim 9 \text{ nm}^{-1} \) and a maxima at \( q \sim 10.5 \text{ nm}^{-1} \) exclusively in the numerical data. The shape of the main peak in the experimental data however suggest to be a superposition of two peaks, with the second one roughly at \( q \sim 10 \text{ nm}^{-1} \), consistent with the numerical results.

The time evolution of both, the experimental and the numerical data is primary governed by a decay due to rotational diffusion. Though in both cases the amplitude of the main peak increases within the first nanoseconds, before rotational diffusion becomes relevant for a time delays of 5.62 ns and later.
7 Conclusions and Outlook

Current and upcoming X-ray sources allow measurement of TR-WAXS curves with picosecond time resolution. The interpretation of TR-WAXS curves has usually been limited to the calculation of implied timescales and the compliance with established kinetic models[55, 56]. TR-WAXS pattern have been interpreted with numerical models for low momentum transfers[57, 130]. In this study a new method is presented, utilizing MD simulations for the calculation of TR-WAXS pattern up to high momentum transfers and taking into account anisotropic contributions; both having been out of reach previously.

Theoretical foundation

It was shown that the method of Park et al. [42] for the calculation of WAXS pattern, can be easily modified to account for TR-WAXS experiments. It was verified that their derivation can be based on an ensemble average over multiple molecules within solution, instead of a time average and thus the method is in principal suitable for the timescales of TR-WAXS, which are too short for a converged time average of an individual molecule. The only difference is in the contribution of the bulk, which instead of being taken from a simulation of a water droplet, had to be replaced by an analytic term representing the excluded volume.

In the following it was theoretically confirmed that photo-selection can break the radial symmetry of TR-WAXS pattern. WAXS patterns can be calculated as the spherical average of the Fourier transform of the electron density. For an excitation probability, which is proportional to the $\cos^2$ of the angle between excitation moment and laser polarization, it could be shown that TR-WAXS pattern consist of exactly two linear independent components for each absolute value of $q$. One component is obtained as the isotropically weighted spherical average, and the other component is weighted corresponding to a second spherical harmonic.

In a typical experiment, the anisotropically weighted contribution can be extracted by subtracting the scattering intensity of the vertical radial line from the one of
the horizontal radial line. There are no indications suggesting a difference in the information content of the two contributions and we therefore suggest that anisotropic scattering double the information content as compared to isotropic scattering.

Of practical importance may be that the anisotropically weighted contribution does not completely cancel out when taking a radial average of the scattering pattern. Numerical calculations of the iodine suggested the opposite[131], which may be due to the very simple structure of iodine, just containing two atoms.

This finding is relevant for the interpretation of timescales of the evolution TR-WAXS pattern. In Fig. 7.1, the radially averaged scattering pattern is depicted, with and without consideration of rotation diffusion. In a experiment, it might be difficult to differentiate between structural changes and rotational diffusion. In a study of a radially averaged TR-WAXS pattern of myoglobin a time-scale, of 9 ns was found[55]. A similar timescale was not reported before in other experiments. Rotational diffusion, which was reported on time-scales of $15 - 20$ ns [63, 114, 115], offers an alternative to an explanation in terms of internal structural changes.

Furthermore, when ab initio algorithms are used to calculate the expected electron density from TR-WAXS experiments on nanosecond instead of microsecond timescales, for which such calculations have been done [57], rotational diffusion and anisotropy needs to be accounted for, either by calculating the the isotropic component first following (3.79), or by incorporating the effect of anisotropy into the modeling.
Calculation and evaluation of anisotropic TR-WAXS of myoglobin

Anisotropic scattering pattern of the diffusion process of CO within myoglobin were calculated using a modified version of the GROMACS simulations toolkit in conjunction with a self-written Python script, for delay times spanning more then six orders of magnitude, from femtoseconds to 100 nanoseconds.

The calculated scattering pattern show the main features of the experimental curves. These are, for the radially averaged scattering pattern, a double-well in the SAXS regime and a peak in the WAXS regime at \( q \approx 17\,\text{nm}^{-1} \). Anisotropic scattering pattern have only been reported for \( q < 12\,\text{nm}^{-1} \). Within this region we found good agreement between our simulation and the experiment concerning, both, the relative amplitude and the positions of minima and maxima.

Our calculation of anisotropic scattering pattern allow for a possible interpretation of these features. The double-well in the radial averaged pattern arises within picoseconds and is of even larger amplitude in a backbone-restrained simulation. This suggests that the main contribution of the double well can be directly referred to displacement of the CO and local related conformational changes. The shape of the experimentally found anisotropic pattern was established in simulations within a picosecond, too, and could similarly been related to the CO displacement.

The peak at \( q \approx 17\,\text{nm}^{-1} \) in the radially average pattern (see Fig. 6.9) arised within nanoseconds only, substantially delayed compared to the experiment. Scattering pattern, based on a clustering of the data, suggest this peak to be related to an occupation of the Xe4 and/or Xe1 pocket. Examine electron density difference maps suggested that the occupation of the Xe4 and Xe1 pocket induces a displacement of the B helix and F helix, respectively. The same peak was, however, observed in a simulation restraining the backbone. This feature should thus be only related to local displacements.

The experimental data shows a considerable increase in the height of this maximum in the corse of time, which could now be related to the diffusion from the distal pocket into the Xe4 and Xe1 pockets. A prediction of the evolution of the scattering pattern due to the diffusion process, based on the population of different clusters was presented. Beside the discussed features they showed a decrease in intensity at \( q \approx 7\,\text{nm}^{-1} \) in agreement with the experiment; the well in the experiment is however considerable narrower.
The anisotropic scattering pattern showed multiple pronounced features at higher scattering angles, encouraging experiments measuring anisotropic scattering pattern over a larger range of momentum transfer. In particular, we consistently found a decreasing intensity at $q \approx 14$ nm$^{-1}$ related to the diffusion form the distal into the xenon pockets.

For the radial averaged pattern, however, multiple featured were observed, which were not reported experimentally. As a proof of concept, these features could be traced back as partially arising in the diffusion process, partially been related to a relaxation of the protein independent from the diffusion process, and partly due to structural changes within the first 100 ps.

From its nature as a contrast method TR-WAXS pattern can be expected to be highly sensitive to small conformational changes. To obtain numerical a structural ensemble, which map the nature with high accuracy both is needed, an accurate model, i.e. force field, and a good sampling. In the following both issues will be discussed.

**Model and Sampling**

By comparison to crystallographic density maps, our simulations clearly reproduce the displacements of the E, F and B helix, but they fail to describe a reported displacements of the G helix. The latter was suggested to be related a conformational change of the heme into a domed shape$^{[53]}$. Crystallographic studies show a displacement of the heme’s ion $^{[53, 85, 128, 129]}$, resulting in a domed structure, which could not be reproduced numerically in this work nor in an earlier study$^{[72]}$. A dissociation involves the breaking of a chemical bond. Such a change in the electronic structure is difficult to map with classical force fields, and possibly a new set of parameters is required for the dissociated case. A bias in the force field might explain some of discrepancy of the scattering pattern.

Some features, present in the simulations, but not in the experiment, arise parallel to conformational changes of the protein backbone and the displacement of the heme. Repeating the calculations with a domed version of the heme may answer causes of some of the discrepancies between simulation and experiment. Consequently, the here presented method may open up an additional source of empirical data for the refinement and benchmarking of molecular force fields. Within the context of this work, a potential refinement would include the heme parameters after excitation.

For TR-WAXS already small bias in observed ensembles may cause artifacts in
the scattering pattern. In this work, the patterns were calculated from an ensemble corresponding to the ground state and a second one corresponding to the excited state. On the short time-scales a large number of simulations were needed to converge the calculated patterns; on nanosecond time-scales, additional calculations of the ground state were required to avoid artifacts from an unequal sampling of the two ensembles. With the sampling method presented here, sufficient low baseline fluctuations were obtained with substantial computational cost. The performed calculations for this work took combined about 12 computing years on current 8-core CPUs.

The calculated data contain features of considerably higher frequency, compared to the experimental curves, with amplitudes clearly exceeding calculated uncertainties. This qualitative disagreement is not necessary to be related to the ensembles considered, but additional effects are to be considered. In the calculation of the scattering pattern a monochromatic X-ray beam, an ideal detector, no anomalous scattering and the absence of secondary scattering was assumed. All the mentioned effects may flatten features and effectively smooth the scattering pattern. For the experiments considered, the X-ray beam is known to have a energy peak with a width of 3%[55]. The convolution of the pattern with the energy spectra may be calculated for a more detailed comparison.

For correct diffusion rates, an accurate parameterization of the CO molecule can be expected to be crucial. In short time scales considerable deviations of the dynamics have been reported [113, 132]. A fluctuating three side model [113] and a multipol model [132] have been suggested as superior to the here used quadropolar model. The timescales observed, although partially corresponding in their order of magnitude, indeed deviate from the once experimentally found.

**Fitting of Numerical Results to Experimental Data**

In this work the evolution of scattering pattern has been predicted from scattering pattern, calculated from different states. The occupation of the different states have been fitted to a kinetic model. An interesting question to address is to find the most likely model, based on the scattering pattern of the different states and the experimental data. Or in other words, which kinetic model reproduces numerically a scattering pattern that matches the one found experimentally. MD simulations might not always show the correct dynamics. However, they often sample the relevant phase space. With the here outlined method the strength of both, simulation
and experiment could be combined with the former offering the structural interpretation and the latter a verification as well as the relevant time scales.

For the interpretation of TR-WAXS scattering such an approach might be superior to other methods, attempting a direct fit of the structure, because of the general risk of overfitting due to the limited information content of WAXS scattering curves. E.g., fitting a rigid body model to the low q range of a TR-WAXS data by simulated annealing suggested a movement of the A helix [131]. This was neither observed in crystallographic experiments [53, 85] nor in our MD simulations. However, being close to the N-terminal, the A helix may have easily just being the most flexible one, allowing numerically the best fitting, rather then its displacement physically being the most likely.

The numerical data obtained allows some additional remarks to make. Although myoglobin has been extensively studies, there is no consents on the diffusion pathway(s). Using an advanced algorithm, which takes advantage of the dynamical information of a MD simulation, we found six states in total, four main ones within the protein, corresponding to the distal, Xe1, Xe2 and Xe4 pocket. An additional state with CO diffusing sparsely in the protein and one corresponding to the CO escaped into the solvent complete the picture. The distal pocket in proximity to the binding site and the xenon pockets Xe1, Xe2 and Xe4 known from early crystallographic studies[83]. Our calculations suggest that the latter two, although occupied, do not necessary show sufficient excess electron density to appear in crystallographic density maps [53, 85].

The exit of CO form the Xe1 pocket was claimed to be ruled out by mutation studies [122] and, instead, the distal pathway was suggested as the only entry and exit pathway[122]. Our simulations suggest in contrast, in line with earlier studies [73, 124], multiple escape pathways. We believe that the mutation studies do not rule out multiple pores[123], because the main argument, that blocking of the Xe1 pocket does not significantly change the global escape rate[123], is equally consistent with multiple escape pathways.

Further Outlook

In this work two independent components within the scattering pattern were derived. Although exceeding the scope of this work, in the following we like to elaborate whether these two components represent the maximum of information possible to extract from a TR-WAXS scattering experiment.
In equation (3.72) the difference intensity depends on two differently weighted spherical averages. The form of the anisotropic one \( \Delta I_m(q) \) is related to the excitation probability. For a different excitation probability one may still expect two components, however, the anisotropic component will contain different weights, depending on the underlying orientational probabilities. This alone may change, but not necessarily increase the available information content. If, however, an experiment could be designed, which would enable one to control the orientational probabilities, anisotropic components with different weights could be calculated.

For each linear independent weighting additional information content can be expected. In the following, two mechanism, anisotropic rotation diffusion and excitation saturation, will be presented, which alter the orientational probability and may be controlled in an experiment.

In the case of anisotropic rotational diffusion, it should be clear that the orientational probabilities, for delay times at which rotational diffusion is relevant, depend on the three rotational diffusion constants. It has been shown that the anisotropic diffusion after photo-excitation can be described as a linear combination of six orthogonal spherical functions[133]. The corresponding linear combination may be varied by changing of the individual rotational diffusion constants, for example by adding a tail to the protein and/or changing the viscosity of the solvent. One may expect that up to six independent components can be obtained. This method would only be feasible for timescales where rotational diffusion is becoming relevant, but have not yet smeared out the anisotropy completely.

In contrast the method making use of excitation saturation may be feasible for short timescales. The probability of photo-excitation with a single photon is proportional to \( \cos^2 \) of the angle between transition moment and laser polarization. For high intensities, saturation has to be taken into account and the excitation probability must be altered to \( f(\phi) = 1 - e^{-a \cos^2 \phi} \), where \( a \) is a constant that is proportional to the laser intensity. Expressing \( f(\phi) \) as a series, it becomes apparent that with increasing \( a \) higher moments of \( \cos^2 \phi \) become relevant. This effect has been argued to enable the detection of higher-order moments in anisotropic fluorescence experiment[134], although quantitative determination was questioned. In the context of solution X-ray scattering higher moments may correspond to higher spherical harmonics. Measuring TR-WAXS pattern for different intensities may allow to extract these contributions.

The underlying idea is closely related to STED super-resolution experiments[135],
Conclusions and Outlook

in which the distribution of excited molecules is shrinked by a second de-excitation beam in saturation. Correspondingly, one may think of a second high power laser pulse, perpendicular to the initial one, de-exciting the sample prior structural changes and leaving only molecules excited and consequently undergoing structural changes, which are increasingly well defined in their orientation. Very recently exactly this method has been used for the high-resolution detection of fluorophore orientation[136]. Light-induced relaxation of heme after photodizotiation has indeed been reported [82]. The effect of heating during relaxation has to be considered.

On may note that, although additional information content may be obtained, in both methods one axis of free rotation remains, covering the full three dimensional information.

Final remarks

In this work the potential of anisotropic solution scattering was presented and it was shown how MD simulation can provide useful insight for the interpretation of these experiments. The required theory was derived, and scattering patterns of the dissociation process of myoglobin were calculated. Qualitative agreement with the experiment allowed the interpretation of main features of the WAXS patterns. However, remaining discrepancies from the experiments render any molecular structural interpretation still difficult. On the femtosecond scale, the scattering patterns appear much more distinct. It will be interesting how the method performs in comparison with experiments on the femtosecond scale, which will shortly be possible with the upcoming free electron laser. Additionally, we are curious whether the anisotropic features predicted at higher scattering angles can be confirmed experimentally.
Bibliography


Bibliography


[26] B. A. Seaton, J. F. Head, D. M. Engelman, and F. M. Richards. Calcium-induced increase in the radius of gyration and maximum dimension of calmod-


Bibliography


Bibliography


112


Bibliography


Acknowledgement

First of all, I like to specially thank Dr. Jochen Hub for giving me the possibility to work on a project, which not only rendered nicely this thesis, but also allowed for creativity and personal emphasis. Having a topic head-to-head with the on going research within the group motivated me a lot. From the very first day on I got the feeling of being a full member of the group. The permanent interest to follow and elaborate any idea from their very early state on, resulted in numerous interesting and fruitful discussions, which I enjoyed a lot.

Furthermore, I like to thank to Prof. Tim Salditt for taking over the position of the second referee.

Particular I like to thank Jannes for the very helpful and detailed discussion about the manuscript. I like the thank my group members Celsa, Chris, Igor, Kalina, Neha and Pocker for helping me out anytime and making sure the is no wrong question possible. In particular I like to highlight our lively lunch conversations, ranging form politics over worlds cultures to science.

Of course I like to thank my awesome friends I had around me within the last year and who made sure that there was plenty in my life beside of science.

I also like to thank my dad, Bernard, for infecting me with his scientific curiosity. Finally I like to thank my mom, Ulla, for her solid support and for all we share.
Erklärung nach §18(8) der Prüfungsordnung für den Bachelor-Studiengang Physik und den Master-Studiengang Physik an der Universität Göttingen:

Hiermit erkläre ich, dass ich diese Abschlussarbeit selbständig verfasst habe, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen wurden, als solche kenntlich gemacht habe.

Darüberhinaus erkläre ich, dass diese Abschlussarbeit nicht, auch nicht auszugsweise, im Rahmen einer nichtbestandenen Prüfung an dieser oder einer anderen Hochschule eingereicht wurde.

Göttingen, den September 13, 2014

(Levin Brinkmann)