Simulations of ultrafast photon-matter interactions for molecular imaging with X-ray lasers

IBRAHIM ELIAH DAWOD

Dissertation presented at Uppsala University to be publicly examined in Häggsalen, Ångström, Lägerhyddsvägen 1, Uppsala, Thursday, 29 February 2024 at 13:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Lars Schäfer (Ruhr University Bochum, Center for Theoretical Chemistry).

Abstract
Eliah Dawod, I. 2024. Simulations of ultrafast photon-matter interactions for molecular imaging with X-ray lasers. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 2353. 95 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-2005-2.

Biological structure determination has had new avenues of investigation opened due to the introduction of X-ray free-electron lasers (XFELs). These X-ray lasers provide an extreme amount of photons on ultrafast timescales used to probe matter, and in particular biomolecules. The high intensity of the X-rays destroys the sample, though not before structural information has been acquired. The unique properties of the probe provide the unprecedented opportunity to study the un-crystallized form of biological macromolecules, small crystals of biomolecules and their dynamics.

In this work, we study processes in XFEL imaging experiments that could affect the achievable resolution of the protein structure in a diffraction experiment. Elastic scattering is the process which provides structural information and leaves the sample unperturbed. This interaction occurs far less often compared to damage inducing processes, such as photoabsorption, which leads to rapid ionization of the studied sample. By using density functional theory, we study the effect of ultrahigh charge states in small systems, such as amino acids and peptides, on the subsequent bond breaking and charge dynamics. Reproducible fragmentation patterns are studied in order to find features that could be understood in larger systems, such as proteins.

Biomolecules are dynamical systems, and the currently used pulse duration is not short enough to outrun the movement of the atoms. The diffraction patterns acquired in an experiment are therefore an incoherent sum of slightly different conformations of the same system. Water can help to reduce these structural variations, but the water molecules themselves will then be a source of noise. Using classical molecular dynamics, we study the optimal amount of water that should be used to achieve the highest resolution.

To simulate ultrafast molecular dynamics of larger systems such as proteins, we develop a hybrid Monte Carlo/molecular dynamics model. We utilize it to simulate the fragmentation dynamics of small proteins and investigate the possibility to extract structural information from the fragmentation patterns. For larger systems exposed to X-ray lasers, such as viruses and crystals, we develop a hybrid collisional-radiative and classical molecular dynamics approach. The method is used in several projects, both in theoretical studies and to support experiments conducted at XFEL facilities. In particular, we simulate the interaction of hexagonal ice with an X-ray laser, and show the structure makes a phase transition from the native crystal state to a plasma, while still partly retaining structural order. Furthermore, we note that the structural changes occur in an anisotropic manner, where different local structural configurations in ice decay on different time-scales.

Preliminary experimental results show this anisotropic dynamics in an X-ray pump-probe serial femtosecond X-ray crystallography experiment performed on I3C crystals. The real space dynamics as a function of probe delay given by our theoretical model and the experiment both show good agreement, where the iodine atoms exhibit correlated motion. The model is also used to calculate the expected atomic displacement and ionization in a hemoglobin crystal, revealing the time and length scales of the dynamics in the protein during the experiment.

Keywords: X-ray free-electron laser, molecular dynamics, radiation damage, plasma simulations, density functional theory, coherent diffractive imaging, protein structure, X-ray crystallography, single particle imaging

Ibrahim Eliah Dawod, Department of Physics and Astronomy, Chemical and Bio-Molecular Physics, Box 516, Uppsala University, SE-751 20 Uppsala, Sweden.

Dedicated to Shayma.

List of papers

This thesis is based on the following papers which are referred to in the text by their Roman numerals.

I **Femtosecond bond breaking and charge dynamics in ultracharged amino acids**
Oscar Grånäs, Nicusor Timneanu, **Ibrahim Dawod**, Davide Ragazzon, Sebastian Trygg, Petros Souvatzis, Tomas Edvinsson, and Carl Caleman.
Journal of Chemical Physics. **151**, 144307 (2019).

II **Imaging of femtosecond bond breaking and charge dynamics in ultracharged peptides**
Ibrahim Dawod, Nicusor Timneanu, Adrian P. Mancuso, Carl Caleman and Oscar Grånäs.
Physical Chemistry Chemical Physics. **24**, 1532-1543 (2022).

III MOLDSTRUCT**: modelling the dynamics and structure of matter exposed to ultrafast X-ray lasers with hybrid collisional-radiative/molecular dynamics**
Ibrahim Dawod, Sebastian Cardoch, Tomas André, Emiliano De Santis, Juncheng E, Adrian P. Mancuso, Carl Caleman, and Nicusor Timneanu
Submitted to Journal of Chemical Physics. arXiv:2401.03180. (2024)

IV **Anisotropic melting of ice induced by ultrafast non-thermal heating**
Ibrahim Dawod[1], Kajwal Kumar Patra[1], Sebastian Cardoch, H. Olof Jönsson, Jonas A. Sellberg, Andrew V. Martin, Jack Binns, Oscar Grånäs, Adrian P. Mancuso, Carl Caleman, and Nicusor Timneanu.
In review at ACS Physical Chemistry Au. (2024)

V **Radiation damage in a hemoglobin crystal studied with an X-ray free-electron laser**
Marina Galchenkova, **Ibrahim Dawod**, Janina Sprenger, Dominik Oberthur, Sebastian Cardoch, Emiliano De Santis, Oscar Grånäs, Nicusor Timneanu, Henry N. Chapman, Carl Caleman and Oleksandr Yefanov.
Manuscript. (2024)

---

[1]Contributed equally.

Reprints were made with permission from the publishers.

List of additional papers

Following papers are additional papers that the author has contributed to, but are not presented in the thesis.

VIII **Coherent diffractive imaging of proteins and viral capsids: simulating MS SPIDOC.**
T. Kierspel, A. Kadek, P. Barran, B. Bellina, A. Bijedic, M. N. Brodmerkel, J. Commandeur, C. Caleman, T. Damjanovic, **I. Dawod**, E. De Santis, A. Lekkas, K. Lorenzen, L. L. Morillo, T. Mandl, E. G. Marklund, D. Papanastasiou, L. Ramakers, L. Schweikhard, F. Simke, A. Sinelnikova, A. Smyrnakis, N. Timneanu, C. Uetrecht
Analytical and Bioanalytical Chemistry. **415**, 4209–4220 (2023)

IX **Ultrafast dynamics and scattering of protic ionic liquids induced by XFEL pulses**
Kajwal Kumar Patra, **Ibrahim Dawod**, Andrew V. Martin, Tamar L. Greaves, Daniel Persson, Carl Caleman and Nicusor Timneanu.
Journal of Synchrotron Radiation. **28**, 1296-1308 (2021)

X **Ultrafast X-ray laser induced explosion: how the depth influences the direction of the ion trajectory**
Emiliano De Santis, **Ibrahim Dawod**, Tomas André, Sebastian Cardoch, Nicusor Timneanu and Carl Caleman
Manuscript. (2024)

This work is in part based upon the licentiate thesis "Coherent imaging of ultrafast processes using X-ray lasers" by the same author presented in 2021 at Uppsala University.

Contents

1. Introduction

1.1 Biological macromolecules

Proteins are biological macromolecules which are essential for biology, providing key functions for systems such as viruses and cells. They are constructed from smaller molecules called amino acids, which by constructing peptide bonds can be assembled in different combinations into a polypeptide or a protein. Depending on the sequence of amino acids, proteins will acquire different functionalities. In order to understand this complex relationship, the structure must be determined [1].

The importance of understanding the relationship between structure and function has been known for a long time. In his famous lectures [2], physicist Richard P. Feynman discusses how particular sequences of molecules provide vastly different functions, giving the example of how enzymes are involved in different reactions depending on how they look. He points out that experimentally observing dynamical processes in biology is key to understand the function of a biomolecule [3]. Presently, structural biology has numerous methods to determine the structure of biological molecules [4]. However, studying dynamics with sufficient time and spatial resolution is difficult, and these requirements limit the particular method that can be used. In the next section, several of these methods are discussed, presenting their advantages and drawbacks.

1.2 Methods for structural determination

There are several well established methods for determining the structure of biological molecules [4], like nuclear magnetic resonance spectroscopy (NMR) [5], cryogenic electron microscopy (cryo-EM) [6] and X-ray crystallography [1]. Solved structures using these methods are commonly deposited to the Protein Data Bank (PDB) [7], which is a database that can be accessed for further study. X-ray crystallography has been the most successful method for a long time, providing almost 10000 new structures to the PDB per annum.

NMR studies of single proteins in solutions are performed by applying an external magnetic field, which the magnetic moments of the nuclei will align themselves with. Electromagnetic radiation of radio frequency wavelengths, are then used to probe the protein solution. The resulting frequency that the nuclei posses is compared to the frequency of the probe, which quantifies the chemical shift. This shift of an atom depends on the surrounding molecules,

and by correlating the chemical shift between different molecules one can extract structural information [5, 8]. The main advantage of this method is being able to study the dynamics of proteins in a native environment at atomic resolution, without the need for crystallization. However, the limit of the method is the size of the molecule, as it is not feasible to study large proteins [9].

Cryo-EM utilizes electron beams focused on cryogenically cooled single particle samples, where the elastically scattered electrons are recorded as a signal on a detector, and later used to reconstruct the target. Through new technological developments, such as better detectors, faster computers and new computational techniques, the method is presently able to retrieve close to atomic resolution structures [10]. One of the major drawbacks of the method however, is that the bio-sample is imaged while it is cryogenically cooled which is a non-native environment for such systems. Time-resolved functional studies are difficult to perform with high time-resolution, since the sample needs to be frozen quickly. However, one can study intermediate conformations of a biomolecule by inducing a dynamical process and freezing the sample at some time delay [11].

X-ray crystallography is a method for determining the structure of a material by using crystals. A crystal is matter built from a large number of copies of identical atoms or molecules which are arranged periodically. When illuminating the crystal with electromagnetic radiation, electrons in the sample can interact with the light in numerous ways. Elastic scattering is the process where the incoming photon changes its direction, with no loss of energy to the sample. The scattered waves from the atoms will interact by constructive or destructive interference, depending on the positions of the atoms in the crystal, thus providing structural information about the sample. This method has been successful for a long time, providing high resolution structures of a large number of biological macromolecules, mainly using synchrotrons sources. However, it can require larger than micrometer sized high quality crystals and many biological macromolecules are difficult to crystallize into such a form [12].

The introduction of X-ray free-electron lasers (XFELs) has made it possible to image biomolecules that can only be crystallized into dimensions of micro- or nanometers. These lasers currently operate at femtosecond timescales and are able to provide pulses with around 10^{12} X-ray photons, focused to spot sizes of a few 100s of nanometers to micrometers [13, 14]. The sample is destroyed by the intense probe, however the effects of damage can be mitigated using pulses with duration shorter than the typical timescale of atomic motion. Small crystals are studied by serial femtosecond crystallography (SFX), which is a method where crystallized biomolecules in a liquid jet are probed at room temperature by an XFEL [15].

For biomolecules that are hard or impossible to crystallize, it has been shown both theoretically [16] and experimentally [17] that one can use XFELs to probe them in the gas phase, and retrieve structural information. Even though the biomolecule is imaged in the gas phase, which is not its native environ-

12

ment, it has been proposed by theory that the structure is preserved in vacuum long enough for the system to be probed by an XFEL [18]. This method, named single particle imaging (SPI), has not yet reached atomic resolution, and there are numerous of challenges such as high background signal, radiation damage and structural heterogeneity among others that need to be addressed [19, 20]. The smallest object imaged in SPI to date is the 14 nm sized protein GroEL [21].

Fluctuation X-ray scattering (FXS) is an emerging method where an ensemble of non-interacting single bio-molecules in a solution are imaged by an XFEL, and the scattered signal on the detector is analyzed by angular correlations [22]. It is an extension of small angle X-ray scattering (SAXS), where only small angles of the elastically scattered photons are studied. The experiments have to date mainly been performed at synchrotron sources to study biomolecules. However, using a short XFEL pulse, FXS can achieve more structural information compared to SAXS, since each molecule preserves its orientation with respect to the beam during the pulse. The detected signal will therefore be an average of different orientations of the same particle [23].

The use of an XFEL enables the study of dynamics in biomolecules in a close to native environment, with a time resolution between femto- to microseconds or longer. In a pump-probe experiment, dynamical processes can be started using a pump laser, such as photosynthesis in photosystem II [24] or through radiation-induced damaged. The X-ray probe is then applied at different time delays, making it possible to follow the entire dynamical process. These experiments are impossible to conduct with conventional methods.

1.3 Motivation for this work

The results shown in this thesis are based on theoretical simulations and are compared to experiments studying ultrafast processes using high intensity X-ray lasers. We are particularly interested in developing tools to accurately and efficiently model these interactions, and extract useful information which can be used to improve future experiments. Our studies are an important part of efforts trying to simulate the start-to-end of XFEL experiments [25], as seen in figure (1.1). These tools simulate every single part of the experiment, and our focus is on the photon-matter interaction, diffractive imaging and reconstruction. Partners focus on simulating the propagation of the XFEL pulse through the beamline, the detector and how the sample is injected into the focus of the X-ray beam [25]. We are particularly interested in developing the method of bioimaging at XFELs. In the thesis, we present theoretical simulations studying processes that could affect the achievable resolution in SPI or SFX experiments at an XFEL. Solving a structure using an XFEL requires beamtime to perform experiments. It is usually difficult to be awarded beamtime, due competition in proposal schemes. To prepare for an experiment, theoretical
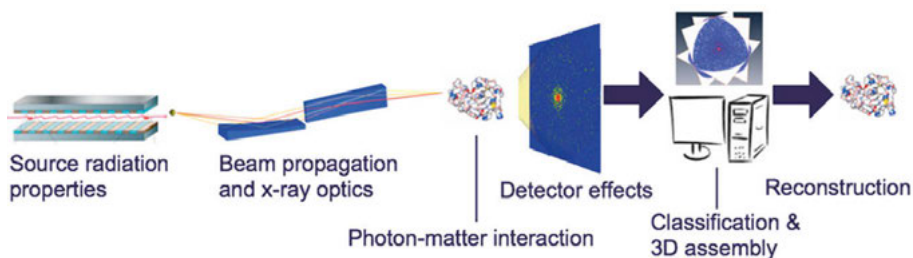
Figure 1.1. The start-to-end simulation pipeline. SIMEX [25] is a software suite where this pipeline has been developed. The focus of the thesis is to study the effects of photon-matter interaction on the reconstruction. (Image courtesy of SPB/SFX, European XFEL[2].)

studies can help to quantify optimal parameters that should be used in the experiment, in order to explore the most important parameters in any given experiment. Simulations could for instance be used to study a range of different pulse parameters in order to find those that minimize radiation damage [26, 27]. Prediction of the number of diffraction patterns required for reconstructing the structure at the target resolution can be done using simulations [28, 29], which affects how long one needs to collect data at the experiment [20]. Estimations of radiation damage in an experiment can help the interpretation of the retrieved structure, and one can potentially confirm if the structure has been solved based on correct assumptions of the average structure that is probed during the XFEL pulse. Finally, we explore the idea of tracking the trajectories of the fragments of single molecules induced by an XFEL, in order to extract structural information. Recent experiments on small molecules have shown that the structure can be reconstructed by following the Coulomb explosion dynamics induced by an X-ray laser [30].

1.4 Description of the thesis

This work has been written with the idea of being self-contained, where the theory needed to understand the subject and the majority of the important results in the papers are presented here. The thesis contains a description of the essential theory in chapter 2. Basics of photon-matter interaction and plasma physics is described in sections 2.1 and 2.2. In several papers, classical molecular dynamics as described in section 2.3 has been used. Building on the theory in the previous sections, the hybrid CR/MD and MC/MD models used to simulate photon-matter interaction and structural dynamics are presented in section 2.4. Paper I and II about amino acids and peptides utilize density functional theory, which is described in section 2.5. This also includes the method

---

[2]`https://www.xfel.eu/facility/instruments/spb_sfx/science_programme/`
`index_eng.html`

for analyzing the data in the density functional theory simulations. The theoretical basis for coherent diffraction imaging can be found in section 2.6, and procedures for analyzing the effects of radiation damage on the reconstructed electron density. In chapter 3, we present the papers and their results. The papers deal with systems of different sizes, in the following order: amino acids (paper I), polypeptides (paper II), water (paper III), ice (paper IV), clusters (paper III), protein crystals (paper III and V) and proteins in gas phase (paper VII and VIII). In addition, preliminary results from a project on a small molecule crystal is presented in section 3.5. We show both theoretical simulations, and experiments that they are compared to. Depending on the system size, different theoretical tools had to be developed. Chapter 4 briefly concludes the results, and discusses the current plans for future projects to pursue.

2. Theoretical background

2.1 Photon-matter interaction

When photons interact with matter, the electronic occupation of the atoms can be altered in several different processes. Photoionization occurs when an electron absorbs a photon with enough energy (or several photons in multi-photon ionization) to escape from a bound atomic state. A bound electron in a lower orbital can absorb a photon in order to make a transition to a higher bound atomic state. For an atom not in the ground state, like when an inner shell state is unoccupied, an electron occupying a higher energy level can fill the state followed by either fluorescence, where a photon is emitted, or by Auger-Meitner decay, where another electron is released from the atom. If another photon interacts with the atom during this excited state, stimulated emission can occur, resulting in fluorescence with a frequency equal to the incoming photon [31].

The mentioned processes are the basics of photon-matter interaction and in reality the dynamics is more complicated where the time-evolution of the system depends on intensity, wavelength and sample [32]. The following sections will focus on certain aspects of the interaction.

2.2 Plasmas and dynamics

Following sufficiently intense ionization in a sample due to the X-ray laser, free electrons are trapped by the charged ions, forming a nanoplasma [33]. In such a phase, the atoms' electronic structure can also be altered in processes involving free electrons. Collisional ionization corresponds to the interaction between a free electron and an atom, resulting in the ionization of the atom. The inverse process is also possible through electronic recombination, involving the same number of particles. The free electron must not necessarily result in ionization, but the bound electrons can instead be excited/de-excited between different energy levels [32, 34]. Furthermore, electrons that are free, can in the presence of an atom, either emit photons through bremsstrahlung, or absorb photons in the inverse process [35, 36].

All possible processes between the particles have their corresponding cross section, describing how often the interaction occurs per unit time, also called rates. To describe the dynamics of a system, the matrix $\mathbf{A}$ is defined containing rates for all interactions. The rates are dependent on both the ion and electron

temperature. They are coupled and evolved in time using a two-temperature model with one temperature for the ion and one for the electron and include processes such as collisions, conduction and hydrodynamics. By knowing the number densities of the different ions in the sample $\mathbf{y}$, where a particular atomic species can occupy several electronic states, the densities are evolved using the rate equation [37]

$$\frac{d\mathbf{y}}{dt} = \mathbf{A}\mathbf{y}. \tag{2.1}$$

The energy levels of an atom in the plasma can be altered due to the external potential from all other charged ions and free electrons. In a process called ionization potential depression (IPD), the energy required to ionize bound states is reduced. This can result in that previously bound high lying states will be in the continuum [32]. The calculation of the change in energy in IPD is commonly done using the Stewart-Pyatt formalism [38], where the result is based on a solution to the Poisson equation [32].

2.2.1 The Poisson equation

The Poisson equation is a partial differential equation which is the foundation of electrostatics. It describes how the electrostatic potential ϕ varies in space due to a charge distribution ρ and is formulated as [39]

$$\nabla^2 \phi(\mathbf{r}) = -\frac{n(\mathbf{r})}{\varepsilon_0}, \tag{2.2}$$

where ϕ is the potential, n the total charge density (containing both positively and negatively charged particles) and ε_0 the vacuum permittivity. For a system where the charge density is described by point particles $n(\mathbf{r}) = \sum_i q_i \delta(\mathbf{r} - \mathbf{R}_i)$ distributed in space with positions $\mathbf{R}_i$, one retrieves the Coulomb interaction

$$\phi(\mathbf{r}) = \sum_i \frac{q_i}{4\pi\varepsilon_0 |\mathbf{r} - \mathbf{R}_i|}. \tag{2.3}$$

Given a system with direct access to the ions' positions $n_{ion}(\mathbf{r}) = \sum_i q_i \delta(\mathbf{r} - \mathbf{R}_i)$ and the electron density $n_e(\mathbf{r}) = n_{e0}(\mathbf{r}) \exp(e\phi(\mathbf{r})/k_B T_e)$ assuming a Maxwell-Boltzmann distribution with temperature T_e, equation (2.2) is written as [40]

$$\nabla^2 \phi(\mathbf{r}) = -\left(\sum_i q_i \delta(\mathbf{r} - \mathbf{R}_i) + n_{e0}(\mathbf{r}) \exp(e\phi(\mathbf{r})/k_B T_e) \right). \tag{2.4}$$

This corresponds to a nonlinear equation which can be solved computationally in a self-consistent manner. Given the ion density n_{ion} and the initial electron density n_e on a grid $\mathbf{r} = (x, y, z)$, the corresponding potential $\phi(x, y, z)$ can be calculated numerically with finite differences. The value for the potential in the first calculation is then used to update the electron density by $n_e(\mathbf{r}) =$

$n_{e0}(\mathbf{r}) \exp(e\phi(\mathbf{r})/k_B T_e)$ and one can then solve the Poisson equation again. This is repeated until the root mean squared difference between two iterations is below a specified tolerance. Instead of using finite differences in real space, the potential can also be computed using the particle mesh Ewald algorithm [41]. This splits the Coulomb interaction into a short range and long range term. The short range part computes the interaction in real space, while the long-range term spreads the charges on a real space grid and uses Fourier transforms to solve the Poisson equation. The potential on each atom and the corresponding force can then be calculated, which will include the effect of the free electron gas adapted to the charged ions. This will have a screening effect on the electrostatic interaction between the ions [42].

Instead of numerical techniques, the Poisson equation can be simplified in different regimes in a plasma, which provides analytical expressions that can depend on the electron temperature, density and ion charge. In this work, two regimes are considered, the strongly and weakly coupled regimes. The parameter that defines the state of coupling is called the Coulomb coupling parameter Γ and is defined as the ratio of the average Coulomb interaction and kinetic energy of ions with temperature T_i [43]

$$\Gamma = \frac{E_C}{k_B T_i} = \frac{\langle Z_i e \rangle^2}{R_i K_B T_i},\tag{2.5}$$

with $R_i = (\frac{3}{4\pi n_i})^{1/3}$. When $\Gamma \ll 1$, the system is in the weakly coupled regime since the Coulomb interaction is low compared the kinetic energy. In the case of $\Gamma \gg 1$ the system is strongly coupled and the Coulomb interaction strongly affects the trajectory of the ions.

2.2.2 Ion sphere model

For the strongly coupled regime, the ion sphere (IS) model is used. The system is considered an ensemble of ions, each of which occupies a spherical volume represented by a radius R. This value is determined based on the radius where the free electron density neutralizes the volume. Given an ion with net charge z and a spherically symmetric electron density $n_e(r)$, the IS radius R is determined by the condition $z - 4\pi \int_0^R r^2 n_e(r) dr = 0$ [43]. This results in

$$R = \left(\frac{3z}{4\pi n_e} \right)^{1/3}.\tag{2.6}$$

The Poisson equation is simplified by the fact that only a single ion and the free electron density are considered in the ion sphere and is defined as [43]

$$\nabla^2 \phi_{IS}(r) = 4\pi \left(n_e(r) - z\delta(\mathbf{r}) \right),\tag{2.7}$$

which with the boundary condition $\phi_{IS}(r = R) = 0$ gives the following analytical expression

$$\phi_{IS}(r) = \frac{z}{r}\left[1 - \frac{r}{2R}\left(3 - \frac{r^2}{R^2}\right)\right]. \tag{2.8}$$

2.2.3 Debye shielding

For a system of positively charged ions, submerged in an ensemble of negatively charged particles (electrons), the Coulomb interaction between the ions will be altered. The electrons are attracted by the positively charged ions, which would eventually result in each ion being surrounded by a cloud of free electrons. This results in a reduction of the Coulomb interaction between the ions. The strength of this reduction is dependent on the Debye length λ_D, which can be derived using the Poisson equation (2.2). The mean field Poisson-Boltzmann equation for an ion z, assuming Maxwell-Boltzmann distributions for all particle species, is given by

$$\nabla^2\phi(r) = -\left(-n_e(r)e^{\frac{e\phi(r)}{k_B T_e}} + \bar{z}n_i(r)e^{-\frac{e\bar{z}\phi(r)}{k_B T_i}} + z\delta(\mathbf{r})\right) \tag{2.9}$$

where n_e is the free electron density, n_i the ion density of the average charge state $\bar{z}$ and T the temperatures. By applying the assumption of weak coupling $e\phi/k_B T \ll 1$, Taylor expanding the electron and ion densities and only keeping the linear terms, we get that the solution to the Poisson-Boltzmann equation is the screened Coulomb potential defined as

$$V(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}\exp\left(\frac{-r_{ij}}{\lambda_D}\right), \tag{2.10}$$

where the quantities q_i, q_j are the charges of ion i, j with relative distance r_{ij}. The corresponding screened force then becomes

$$F(r_{ij}) = -\frac{\partial V}{\partial r_{ij}} = \frac{q_i q_j}{4\pi\varepsilon_0}\left(\frac{r_{ij} + \lambda_D}{\lambda_D r_{ij}^2}\right)\exp\left(\frac{-r_{ij}}{\lambda_D}\right). \tag{2.11}$$

The potential and force converge to the regular Coulomb interaction when $\lambda_D \to \infty$. The Debye length λ_D is defined as [44]

$$\lambda_D = \sqrt{\frac{\varepsilon_0 K T_e}{n_e e^2}}, \tag{2.12}$$

where T_e is the free electron temperature and n_e is the free electron density. The concept of the Debye length is valid only if the parameter is much smaller than the dimensions of the system $\lambda_D \ll L$ and if there are a sufficient number of particles $n_e \lambda_D^3 \gg 1$ in the Debye sphere volume to make the concept statistically valid [44]. Solving the Poisson equation numerically in a self-consistent

manner can be more computationally demanding compared to simply using the screened potential in equation (2.10). However, since the Debye length is not always valid for all plasma phases, one must check its validity. Even though the Debye length might not be statistically valid in our simulations, we know that the electrons will have some screening effect on the electrostatic interaction between the ions. We therefore used the Debye screening potentials in several papers, since it is easy to implement and we noted that it agreed enough with a more complicated screening model. It has also been shown that using Debye screening to model radiation damage in XFEL experiments predicts the correct time-scales for bond-breaking for a disulfide bond in a protein crystal [45].

On the other hand, only using the Debye length restricts us to using a homogeneous electron density for the whole system. If the Poisson equation is solved self-consistently, the electron density is allowed to be in-homogeneous in space [42]. This is important to consider for systems with heavy atoms, since they can attract more electron density around them.

2.2.4 Hybrid ion-sphere and Debye screening model

When a sample is probed by X-rays and a plasma is initiated, the quantities of the plasma such as the Coulomb coupling constant can vary during and after the pulse. A potential which adapts to the current state of the plasma and that is valid for all regimes is therefore important to develop. For the adaptive potential, one can use a hybrid model combining the ion-sphere (IS) and Debye screening (DS) model discussed previously [43]. Given the ion z, ion number density n_i of species i with charge z_i and the electron density n_e we write the Poisson equation as

$$\nabla^2 \phi(r) = -4\pi \left[-n_e(r) + \sum_i z_i n_i(r) + z\delta(\mathbf{r}) \right]. \qquad (2.13)$$

We search for a transition point $r = r'$ where the IS model is valid below $r \leq r'$ and DS above ($r \geq r'$). In the IS interval $r \leq r'$, the electron density (assumed to be uniform over the interval ($n_e(r) = n_e$)) is greater than the ion density and one can approximate that $n_i/n_e \approx 0$. The solution to the Poisson equation in (2.13) is then for the interval $r \leq r'$

$$\phi_1(r) = \frac{c_0}{r} + c_1 - c_2 r^2, \qquad (2.14)$$

where the constants c_0 and c_1 will be given by the boundary conditions imposed at $r = r'$ and with $c_2 = \frac{ze^2}{2R^3} r^2$, where R is the ion-sphere radius. For the interval $r \geq r'$, the Coulomb potential given by the ion will be much smaller than the kinetic energies of the particles, which enables the weak-coupling

regime. Applying the approximation $e\phi/k_bT << 1$, the solution to equation (2.13) in the interval $r > r'$ becomes

$$\phi_2(r) = \frac{c_3}{r}e^{-\frac{r}{\lambda_D}},\tag{2.15}$$

where λ_D is the Debye length defined according to equation (2.12). By applying the boundary conditions $\phi_1(r) = \phi_2(r)$, $\phi_1'(r) = \phi_2'(r)$ and $\phi_1''(r) = \phi_2''(r)$, the unknown constants c_0, c_1, c_2 and c_3 are

$$c_0 = ze^2,$$
$$c_1 = -\frac{3z\lambda_D^2 e^2}{2R^3}\left[((\frac{R}{\lambda_D})^3 + 1)^{2/3} - 1\right],$$
$$c_2 = -\frac{ze^2}{2R^3}r^2,$$
$$c_3 = \frac{3z\lambda_D^2 e^2}{2R^3}r'r^{\frac{r'}{\lambda_D}}.\tag{2.16}$$

and the transition point $r = r'$ is determined as

$$r' = \lambda_D\left[((\frac{R}{\lambda_D})^3 + 1)^{1/3} - 1\right].\tag{2.17}$$

Given these constants, we have a hybrid potential $\phi(r) = \phi_1(r \leq r') + \phi_2(r \geq r')$ which adapts to the current state of the plasma. This expression of the Coulomb potential was used in the simulations, which allowed us to capture the changes of the plasma with time.

The theory described above is the basis for the collisional-radiative (CR) software package CRETIN [34], which has been used to perform photon-matter calculations. For large samples where the ratio of escaping to trapped electrons is small, such as crystals and liquids, the CRETIN has been run to compute the expected photon-matter interaction. For smaller samples, where this ratio is larger, it has provided data for atomic transition rates and cross sections for running hybrid Monte Carlo/molecular dynamics (MC/MD) simulations.

2.3 Classical molecular dynamics

Classical molecular dynamics (MD) is a theoretical tool to study the dynamics of molecules based on Newtonian mechanics. It can be applied to systems that are generally to large to simulate with *ab-initio* theory, such as proteins and viruses. The main software used to perform classical molecular dynamics simulations in the projects presented is GROMACS [46, 47, 48, 49]. It is typically used for studying dynamics and function of proteins in native environments, and not in the highly charged conformations that are explored here. However, in the year 2000, GROMACS was used to show that ultrafast diffraction of a single biological molecule could be possible [16]. This started the field of bioimaging with XFELs.

2.3.1 GROMACS

The main idea of molecular dynamics is to use Newton's law to propagate the position $\mathbf{r}_i$ of an atom i, based on the force $\mathbf{F}_i = \sum_j \mathbf{F}_{ij}$ from all other particles j in the system

$$m \frac{\partial^2 \mathbf{r}_i}{\partial t^2} = \mathbf{F}_i, \tag{2.18}$$

where the force is derived from the potential V

$$\mathbf{F}_i = -\frac{\partial V(\mathbf{r}_1, \cdots, \mathbf{r}_N)}{\partial \mathbf{r}_i}. \tag{2.19}$$

The positions of all atoms at time $t + \Delta t$ can be calculated using an algorithm named *leap-frog*,

$$\mathbf{r}(t + \Delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \Delta t) + \mathbf{a}(t)\Delta t^2 + \mathcal{O}(\Delta t^4), \tag{2.20}$$

where the acceleration $\mathbf{a}(t)$ is calculated from the force on the atom. The potential which the force depends on is usually determined as the sum of bonded and nonbonded interactions

$$V_{bonded}(\mathbf{r}) = V_{bond} + V_{angle} + V_{dihedral}, \tag{2.21}$$

$$V_{nonbonded}(\mathbf{r}) = V_{LJ} + V_{Coulomb}. \tag{2.22}$$

Non-bonded potentials are calculated by the summation of two-body interactions, ignoring many-body effects [50],

$$V(\mathbf{r}_1, \cdots, \mathbf{r}_N) = \sum_{i<j} V_{ij}(\mathbf{r}_{ij}). \tag{2.23}$$

For instance, the Lennard-Jones (LJ) potential is,

$$V_{LJ}(r_{ij}) = \frac{C_{ij}^{(12)}}{r_{ij}^{12}} - \frac{C_{ij}^{(6)}}{r_{ij}^6} \tag{2.24}$$

with parameters $C_{ij}^{(12)}$ and $C_{ij}^{(6)}$ that are adjustable and depend on the two atoms involved i, j. The Coulomb interaction is defined as

$$V_{Coulomb}(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon_r r_{ij}}, \tag{2.25}$$

where q_i, q_j are net charges on the two atoms and ε_0, ε_r are the dielectric constants. Bonded interactions contain pair interactions, similar to the non-bonded potentials. However, they can also employ interactions with multi-body effects. An example of a two-body term is the potential for the covalent bond usually determined by the harmonic potential

$$V_{bond}(r_{ij}) = \frac{1}{2}k_{ij}(r_{ij} - b_{ij})^2, \tag{2.26}$$

where k_{ij} is the force constant and the equilibrium bond length b_{ij}. For systems that require bonds to break, which is expected for a sufficiently ionized system, the Morse potential is used [51, 48],

$$V_{bond}(r_{ij}) = D_{ij}(1 - \exp(-\beta_{ij}(r_{ij} - b_{ij})))^2. \tag{2.27}$$

Here, D_{ij} is the difference between the minimum of the potential energy curve and the energy when the two atoms are isolated. The term β_{ij} gives how fast the bond transitions from the equilibrium bond length b_{ij} to a dissociative state.

The parameters in the potentials are determined based on a specific force field used, which is developed using *ab-initio* theory and experiments. Commonly used force fields for simulating biomolecules are CHARMM [52], AMBER [53], OPLS [54] and GROMOS [55]. Different force-fields are suitable depending on the property of the system studied. They are developed by fitting the parameters of the force-field to reproduce experiments conducted to study a specific property, for example the specific heat capacity [56].

2.4 MOLDSTRUCT: A multiscale photon-matter model

To simulate photon-matter interaction induced by an ultrafast X-ray laser, we have developed a multiscale model MOLDSTRUCT to run photon-matter calculations coupled to classical MD. Two versions of the model exist, which differ in the way photon-matter interaction is calculated. The first hybrid CR/MD version is based on rate-equations and it is used for large samples, where the majority of free electrons are trapped. We utilize a CR model since it includes complex physical processes and models the system as a continuum, which removes restrictions to the sample size. By coupling it to MD, we can probe local dynamics of the atoms and follow their movement in detail. For the second version, a Monte Carlo algorithm is used and it allows for electrons to escape the sample. This makes it suitable to model smaller systems such as amino acids and proteins.

2.4.1 Hybrid CR/MD simulations

The hybrid CR/MD implementation is described by the flowchart in figure (2.2). Here, we use the CR code CRETIN [37, 34] and based on atomic mod-
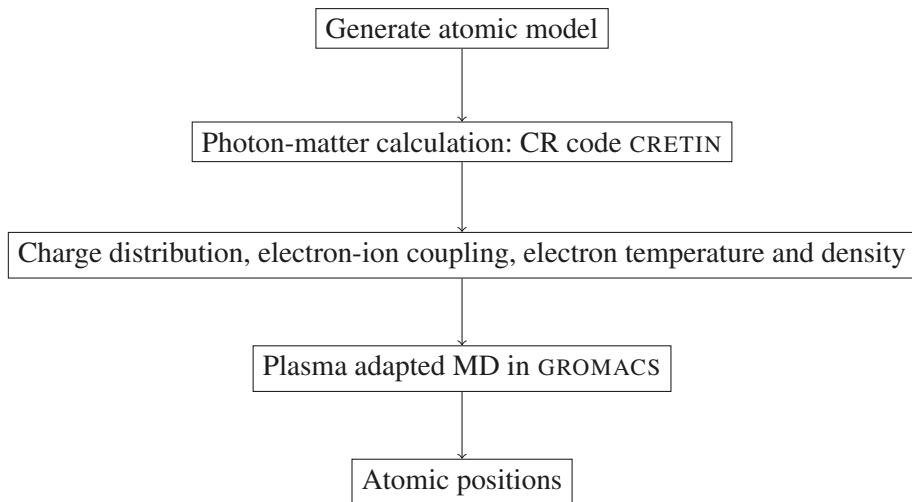
```
┌─────────────────────────┐
│  Generate atomic model  │
└─────────────────────────┘
            │
            ▼
┌───────────────────────────────────────────┐
│ Photon-matter calculation: CR code CRETIN  │
└───────────────────────────────────────────┘
            │
            ▼
┌────────────────────────────────────────────────────────────────────┐
│ Charge distribution, electron-ion coupling, electron temperature and density │
└────────────────────────────────────────────────────────────────────┘
            │
            ▼
┌──────────────────────────────────┐
│   Plasma adapted MD in GROMACS   │
└──────────────────────────────────┘
            │
            ▼
┌─────────────────────┐
│  Atomic positions   │
└─────────────────────┘
```

Figure 2.2. Flow chart of the hybrid CR/MD model.

els with varied complexity calculate the distribution of expected charge states, the free electron temperature and density. CRETIN does this by determining all transition rates for the current electronic occupations of the atoms, and propagates the rate equation (2.1) to determine the distribution of the electronic occupations of the next time-step. The Maxwell-Boltzmann distribution function is used to describe the electron dynamics and the temperature of the electrons and ions are evolved through a coupled two-temperature model. Structural changes to the sample is included through hydrodynamic expansion.

The data is acquired as a function of time and later used as an input to run MD in GROMACS, which was modified to read data from CRETIN [16, 48]. At each time-step, GROMACS iterates through the atoms, and assigns a charge state based on the data from CRETIN. The change in the electrostatic interaction due to the free electrons is modelled using screened Coulomb interactions where the plasma data is taken from the CR calculations. Since the Lennard-Jones (LJ) interaction in equation (2.24) is also mediated by the electromagnetic force, we modulated it by the same screening factor. The LJ force is dependent on the extension of the electronic cloud of the atom (ionic radius). We adapted the parameters C_6 and C_{12} based on the number of bound electrons in the two atoms involved. As electrons are removed from the atom, the ionic radius is reduced and the atoms will interact with a LJ interaction with a minimum shifted to shorter distances. For bonded interactions we utilize the Morse potential, which allows for bond breaking. Since large systems are

explored in this version, we utilize GROMACS version 4.5.4, which includes parallelization through domain decomposition. It divides the sample in cells given by the number of Central Processing Unit (CPU) cores available and the interactions within that cell are calculated by that CPU. The code was adapted to run photon-matter calculations in parallel.

The CR code is based on the independent atom approximation. This means that only atomic orbitals are used for calculating the cross sections for atomic transitions. This is an approximation which is deemed valid since we mainly use hard X-rays that primarily interact with inner shell orbitals. Furthermore, it does not incorporate chemical bonds and the properties of an ion (for instance charge state or temperature) are considered to be independent of its position [57, 58]. Since the CR calculations only provide information regarding the photon-matter interaction, MD is used to propagate the positions of the atoms in time. Earlier assumptions are that that all bonds are broken before the atoms have time to move due to the increase in Coulomb interaction [58, 59]. However, for low intensities, or in-homogeneous spatial distribution of the photons on the sample, a large fraction of molecules can still be neutral. Thus, one still needs to incorporate bonded interactions and dynamically reduce their strength as atoms involved in the interaction are ionized. Even though classical MD does not adapt the interactions to electronic changes, the advantage of using it for these calculations is that it allows for easy localisation of charge. For *ab-initio* simulations, the occupation of the orbitals are filled such that the lowest energy state of the system is reached. In order to occupy an excited energy state (to force a particular charge state on an atom) specific algorithms are required.

In the MD simulations, the energy transfer from the pulse to the kinetic energy of the ions is done through the Coulomb interaction. However, energy transfer also occurs through electron-ion collisions. This can be important to incorporate for low-ionized plasmas, since the effect of the energy transfer between electrons and ions on the subsequent dynamics will have more of an effect. This can be modelled using Langevin MD (also know as stochastic MD), which adds two terms two the right-hand side of equation (2.18). For each ion i, the Langevin equation evolves the position r_i according to

$$m\frac{\partial^2 \mathbf{r}_i}{\partial t^2} = \mathbf{F}_i - m_i \gamma_i \frac{\partial \mathbf{r}_i}{\partial t} + \mathbf{f}_i(t), \tag{2.28}$$

where γ_i is the electron-ion coupling constant extracted from the CR calculations and $\mathbf{f}_i$ a stochastic noise process with a Gaussian distribution. From the CRETIN code, a single value of the coupling parameter is given for all atoms, which is then multiplied by the number density of each atomic species to give separate coupling values. In GROMACS we can then separate the atom species by groups and apply the corresponding value for the coupling. The ion temperatures are then coupled to the electron temperature from the CR simulation at each time-step.

MOLDSTRUCT can be used for both crystalline and single particle samples. In the latter case, we calculated the minimum radius required to have no escaping free electrons. Given a net charge ρ for a spherical sample with radius R, the potential $\Phi(r)$ from the center of the sphere is [60]

$$\Phi(r) = \frac{\rho}{2\varepsilon_0}(R^2 - \frac{r^2}{3}).$$
(2.29)

Thus, we get the condition that the kinetic energy of the electron E_{kin} must at the boundary $r = R$ satisfy the following equation,

$$E_{kin} > \frac{\rho R^2}{2r\varepsilon_0}.$$
(2.30)

During first few femtoseconds of the CR calculations, the electrons will in reality be able to escape. We assume that an insignificant amount escape the sample, before the trapping potential of the sample reaches the kinetic energy of the electron.

2.4.2 Hybrid MC/MD simulations

For small samples such as amino acids, polypeptides and proteins where we expect a large amount of electrons to escape the sample, we developed a hybrid Monte Carlo/MD (MC/MD) model. In this scheme, we calculate the photon-matter interaction on-the-fly in GROMACS using an MC algorithm. The processes included are photoionization, fluorescence and Auger-Meitner decay. The rates are acquired from CRETIN, but any alternative model could be used to provide this data, like for instance FAC [61]. The positions of the atoms are propagated using classical MD in GROMACS, where the force-field is altered based on the electronic occupation of the atoms. For each time-step in the MD simulation, we apply algorithm (2.1). It is developed based on the assumption that each possible transition from initial atomic state i to final state j with rate r_{ij} can be described by a Poisson distribution

$$P(r_{ij}, k, t) = \frac{(r_{ij}t)^k e^{-r_{ij}t}}{k!}.$$
(2.31)

We can find the time until the next transition Δt_{ij} for all possible transitions $i \to j$ by extracting for how long no transitions occur ($k = 0$). By setting $k = 0$, we get

$$P(r_{ij}, 0, \Delta t_{ij}) = e^{-r_{ij}\Delta t_{ij}},$$
(2.32)

which can be equated to a random number u_{ij} for each transition sampled uniformly between 0 and 1, to determine the time Δt_{ij} as

$$\Delta t_{ij} = -\frac{\ln(u_{ij})}{r_{ij}}.$$
(2.33)

Algorithm 2.1 Monte-Carlo transition algorithm.

1: Initialize MC time-step $\Delta_{t_{MC}=0}$
2: Given N atoms
3: **for** $i = 1, 2, \ldots, N$ **do**
4:     Determine the rates r_{ij} for all possible transitions $i \to j$
5:     **for** $R = r_{i1}, r_{i2}, \ldots, r_{ij}$ **do**
6:         Sample random number $u_{ij} \in [0, 1]$ from uniform distribution
7:         Compute time until next transition $\Delta_{t_{MC_{ij}}} = -\frac{ln(u_{ij})}{R}$
8:     **end for**
9:     Determine minimum time-step $\Delta_{t_{MC_{ij}}}$
10:     Increment MC time $\Delta_{t_{MC}} = \Delta_{t_{MC}} + \Delta_{t_{MC_{ij}}}$
11:     **if** $\Delta_{t_{MC}} < \Delta_{t_{MD}}$ **then**
12:         Change atomic state i to j
13:         **goto** 4
14:     **else**
15:         **goto** 3
16:     **end if**
17: **end for**

The cross section for direct photoionization of hydrogen is low compared to higher atoms. For the MC/MD simulations, where we do not consider partial charges, the hydrogen atoms will therefore remain neutral. However, *ab-initio* simulations show that the electron from the hydrogen will be transferred to heavier atoms when the system becomes charged. Furthermore, when the +1 charged hydrogen leaves the molecule, the system will become more stable [62]. It is therefore important to include charge transfer processes in the classical simulations. For the large-scale CR/MD method, the hydrogen atoms will acquire a net charge due to collisional processes and through a charge exchange model [34]. For the MC/MD model, we implemented the classical over-barrier model (COB) [63, 30, 64], where the bound electron in the donor atom with binding energy E_p, feels a Coulomb potential which is a superposition of the interaction with the donor charge Q_D and the acceptor charge Q_A,

$$V(\mathbf{r}) = -\frac{Q_D + 1}{|\mathbf{r} - \mathbf{R}_D|} - \frac{Q_A}{|\mathbf{r} - \mathbf{R}_A|}, \tag{2.34}$$

with R_A and R_D being the position in space for the donor and acceptor atoms. At a certain critical distance R_c, the electron will be able to transfer from the donor to the acceptor atom. This distance depends on the charges Q_D, Q_A and the binding energy of the electron E_p, according to

$$R_c = \frac{Q_D + 1 + \sqrt{(Q_D + 1)Q_A}}{|E_p|}. \tag{2.35}$$

When the distance between the atoms is $R \leq R_c$, then a charge transfer occurs. For each time-step in the MC/MD simulation, algorithm (2.1) is applied first in order to determine each atom's electronic occupation independently. Thereafter, we compute the possible charge transfers in the system, according to equation (2.35). For each time-step we randomize the atom list for stochastic reasons, which gives variability in the charge transfer pathways.

2.5 *Ab-initio* molecular dynamics

For scenarios where the probe does not break all bonds in a system, and it is sufficiently small, detailed knowledge of the changes of the electronic structure due to ionization is important to understand the subsequent dynamics. Classical molecular dynamics is not sufficient to describe this case, since the parameters of the force fields are developed for a system in equilibrium. We therefore apply *ab-initio* molecular dynamics to study ultracharged small biomolecules, such as amino acids and peptides, which are the building blocks of proteins.

2.5.1 Hartree-Fock theory

The time-independent non-relativistic Schrödinger equation can be used to describe the electronic structure of a system. Given K number of atoms with positions $\mathbf{R}_K$ and N number of electrons with positions $\mathbf{r}_N$, the Schrödinger equation becomes [65]

$$\hat{H}\Psi(\mathbf{R}_1,\ldots,\mathbf{R}_N;\mathbf{r}_1,\ldots,\mathbf{r}_K) = E\Psi(\mathbf{R}_1,\ldots,\mathbf{R}_N;\mathbf{r}_1,\ldots,\mathbf{r}_K), \qquad (2.36)$$

where $\hat{H}$ is the Hamiltonian and Ψ is the wavefunction. The operators in the Hamiltonian are both the nuclear and electronic parts,

$$\hat{H} = \hat{H}_n + \hat{H}_e = \hat{T}_n + \hat{V}_{nn} + \hat{T}_e + \hat{V}_{ee} + \hat{V}_{ne}. \qquad (2.37)$$

The kinetic energies for the nuclei (n) and the electrons (e) are given by the operators $\hat{T}_n$ and $\hat{T}_e$. Coulomb interactions between the different particles are in the terms $\hat{V}$. This equation will generally be computationally demanding, since the number of parameters to solve for is large, even for a single atom. By applying the Born-Oppenheimer approximation, one uses the approximation that since the electrons are much lighter than atoms, their movement can be studied in a reference frame where the ions are considered stationary. The Schrödinger equation to solve for is then

$$\hat{H}_e\Psi_k^e(\mathbf{R}_1,\ldots,\mathbf{R}_N;\mathbf{r}_1,\ldots,\mathbf{r}_K) = E_k(\mathbf{R}_1,\ldots,\mathbf{R}_N)\Psi_k^e(\mathbf{R}_1,\ldots,\mathbf{R}_N;\mathbf{r}_1,\ldots,\mathbf{r}_K), \qquad (2.38)$$

which is computed for a given set of atomic positions $\mathbf{R}_1, \ldots, \mathbf{R}_N$. In this formulation, the effect of the nuclei on the electrons is seen as an external potential. Another approximation used it to assume that the wave-function can be written in terms of single particle states by constructing a determinant,

$$\Psi_k^e \approx \frac{1}{\sqrt{N}} \begin{vmatrix} \phi_1(\mathbf{r}_1) & \cdots & \phi_N(\mathbf{r}_1) \\ \vdots & & \vdots \\ \phi_1(\mathbf{r}_N) & \cdots & \phi_N(\mathbf{r}_N) \end{vmatrix}. \tag{2.39}$$

By finding the orbitals in the determinant that minimize the expectation value of the energy,

$$
\begin{aligned}
\langle \Psi_k^e | \hat{H}_e | \Psi_k^e \rangle &= \sum_{i=1}^{N} \sum_{\sigma=\pm\frac{1}{2}} \int d\mathbf{r} \phi_i^*(\mathbf{r}\sigma) \left[\frac{(-i\hbar\nabla_e)^2}{2m} - \sum_{\alpha=1}^{K} \frac{Z_\alpha e^2}{|\mathbf{R}_\alpha - \mathbf{r}|} \right] \phi_i(\mathbf{r}\sigma) \\
&+ \frac{1}{2} \sum_{i,j=1}^{N} \sum_{\sigma,\sigma'=\pm\frac{1}{2}} \int d\mathbf{r} \int d\mathbf{r}' \phi_i^*(\mathbf{r}\sigma)\phi_j^*(\mathbf{r}'\sigma') \frac{e^2}{|\mathbf{r}-\mathbf{r}'|} \phi_i(\mathbf{r}\sigma)\phi_j(\mathbf{r}'\sigma') \\
&- \frac{1}{2} \sum_{i,j=1}^{N} \sum_{\sigma,\sigma'=\pm\frac{1}{2}} \int d\mathbf{r} \int d\mathbf{r}' \phi_i^*(\mathbf{r}\sigma)\phi_j^*(\mathbf{r}'\sigma') \frac{e^2}{|\mathbf{r}-\mathbf{r}'|} \phi_j(\mathbf{r}\sigma)\phi_i(\mathbf{r}'\sigma')
\end{aligned}
\tag{2.40}
$$

where α is the nuclei index and σ is the spin, one retrieves Hartree-Fock (HF) equation. The single determinant in equation (2.39) enforces anti-symmetry, which means that the wavefunction becomes zero if two fermions (electrons) occupy the same quantum numbers in an atom. The Coulomb interaction in HF theory felt by an electron is based on the field from an electron cloud, rather than the position dependent Coulomb interaction between each electron. In order to describe the electronic structure in more detail, correlation effects must be included [65].

2.5.2 Density functional theory

Kohn-Sham (KS) proposed a different approach to model the electronic structure, where approximate correlation effects are incorporated. They formulated a theorem which states that the minimum energy of a system could be computed by the ground state electron density. The energy is then defined as a functional of the electron density [66]

$$E_0 = E[n_0(x,y,z)], \tag{2.41}$$

which is the basis of density functional theory (DFT). The electronic energy in DFT is determined by first calculating the expectation value of the energy for a model system, where the electrons are considered non-interacting,

$$E[n_0] = \langle T[n_0] \rangle + \langle V_{ne}[n_0] \rangle + \langle V_{ee}[n_0] \rangle. \tag{2.42}$$

The electron potential is defined here as the Coulomb interaction between charged clouds of densities

$$\langle V_{ee}[n_0]\rangle = \frac{1}{2} \int \int \frac{n_0(\mathbf{r})n_0(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}'. \tag{2.43}$$

The interaction between the electrons and the nuclei is defined by the term

$$\langle V_{ne}[n_0]\rangle = \int v(\mathbf{r})n(\mathbf{r})\mathrm{d}\mathbf{r} \tag{2.44}$$

and the kinetic energy of the non-interacting system is

$$\langle T[n_0]\rangle = \sum_{i=1}^{N} \int \phi_i^*(\mathbf{r}) \left(\frac{(-i\hbar\nabla)^2}{2m} \right) \phi_i(\mathbf{r})\mathrm{d}\mathbf{r}, \tag{2.45}$$

where ϕ_i are the KS molecular orbitals. The error in the electronic kinetic energy T and the repulsion V_{ee} compared to the real interacting system is modeled in the exchange correlation energy

$$E_{XC}[n_0] = \Delta\langle T[n_0]\rangle + \Delta\langle V_{ee}[n_0]\rangle. \tag{2.46}$$

The final form of the energy is then

$$E_{tot}[n_0] = \langle T[n_0]\rangle + \langle V_{ne}[n_0]\rangle + \langle V_{ee}[n_0]\rangle + E_{XC}[n_0], \tag{2.47}$$

which one can find the orbitals ϕ_i that determine the minimum energy. The result is that the KS orbital ϕ_i is determined from the single-particle equation

$$\left[-\frac{1}{2}\nabla_i^2 - \sum_A \frac{Z_A}{r_{iA}} + \int \frac{n(\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|} \mathrm{d}\mathbf{r} + \frac{\delta E_{XC}[n]}{\delta n(\mathbf{r})} \right] \phi_i = \varepsilon_i \phi_i, \tag{2.48}$$

with the corresponding electron density

$$n(\mathbf{r}) = \sum_{i=1}^{N} f_i \phi_i^*(\mathbf{r})\phi_i(\mathbf{r}), \tag{2.49}$$

where the occupation of the orbitals is usually determined according to the Fermi-Dirac distribution f_i. It is important to note that the orbitals determined in equation (2.48) are for the model system, and therefore do not correspond to real atomic orbitals. However, the idea is that they give the real ground state electron density in equation (2.49).

All *ab-initio* molecular dynamics simulations performed in this work have used the pseudopotential based code SIESTA [67]. For solving the atomic problem, RSPT (Relativistic Spin Polarized toolkit) has been utilized. It is a code based on DFT which solves the Dirac equation and therefore includes relativistic effects [68, 69, 70]. RSPT can calculate the atomic electron density for any arbitrary configuration of the electronic occupation that the user specifies, with the orbitals being relaxed given this configuration.

2.5.3 Tools for analyzing structure

In papers I and II, we study bond breaking as a function of ionization. The Souvatzis bond integrity parameter $\mathscr{B}$ [62] is used as a way to quantify if a bond is broken between two atoms A and B at any given time t, defined as

$$\mathscr{B}(A,B,t) = \frac{1}{N_{MD}} \sum_{i=1}^{N_{MD}} \frac{1}{\left(1 + e^{\lambda(|d_i[A,B](t) - d_i[A,B](0)| - 0.5)}\right)}. \tag{2.50}$$

It returns a value in the range [0, 1] where 0 means that the bond is broken while 1 that it is intact. The value of λ defines how fast $\mathscr{B}(A,B,t)$ goes from 1 to 0. $d_i(A,B)(t)$ is the distance between atom A and B at time t and N_{MD} corresponds to the number of simulations performed. Furthermore, we use the radius of gyration to study the expansion of the molecule [42]

$$R_g(t) = \sqrt{\frac{\sum_i |\mathbf{r}_i(t)|^2 m_i}{\sum_i m_i}}. \tag{2.51}$$

Here m_i is the mass of atom i, with a distance $|\mathbf{r}_i|$ from the center of mass of the molecule.

In order to follow the atoms' charge states upon ionization, we employ Hirshfeld charge analysis [67, 71]. This scheme calculates the electron density of atom i in a molecule, by using its isolated atomic electron density. A weight for this atom is determined based on the sum of atomic densities of all atoms in the molecule

$$w_i(\mathbf{r}) = \frac{n_i(\mathbf{r})}{\sum_{i=1}^{N} n_i(\mathbf{r})}. \tag{2.52}$$

This is then used to determine the atom's charge state from the molecular electron density $n^{mol}(\mathbf{r})$

$$Q_i = Z_i - \int w_i(\mathbf{r}) n^{mol}(\mathbf{r}) \mathrm{d}V. \tag{2.53}$$

To analyze the structure of systems such as liquids and crystals, we have utilized the radial distribution function (RDF) in papers III and IV. This quantity, which can be extracted experimentally in scattering experiments, describes the probability of finding a particular configuration of the atomic positions. For atomistic simulations such as MD, we have direct access to every atom's position at each time-point. An atom's position $\mathbf{r}_i$ can mathematically be described as a delta function in space, $\delta(\mathbf{r} - \mathbf{r}_i)$. Given two particle species A and B, the RDF $g_{AB}(r)$ is defined as the number density $\langle \rho_B(r) \rangle$ of B with respect to its distance r_{ij} from each particle A, compared the average value with respect to all A particles $\langle \rho_B \rangle_{local}$ within the maximum radius r_{max} [48],

$$g_{AB}(r) = \frac{\langle \rho_B(r) \rangle}{\langle \rho_B \rangle_{local}} = \frac{1}{\langle \rho_B \rangle_{local}} \frac{1}{N_A} \sum_{i \in A}^{N_A} \sum_{j \in A}^{N_B} \frac{\delta(r_{ij} - r)}{4\pi r^2}. \tag{2.54}$$
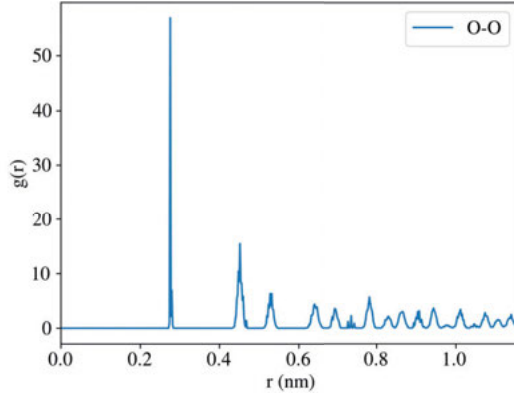
Figure 2.3. Example of the radial distribution function of oxygen (O-O), computed from the ice structure in paper IV using the GENICE code [72, 73].

An example of the RDF for hexagonal ice is shown in figure (2.3).

To resolve the dynamics of a system in terms of both angles and distances, we have utilized the pair-angle distribution function (PADF) [74], which is a many-body correlation function that depends on the distance between atoms and the angle between them, depicted in figure (2.4). It can be directly computed from trajectories produced by MD or DFT simulations. Alternatively, one can retrieve it from scattering data, which enables retrieval of structural information experimentally [74]. An example of a PADF map can be found in figure (3.25a).
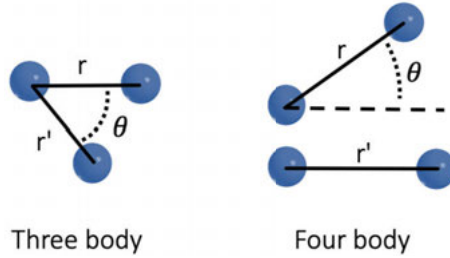

Figure 2.4. The three and four body terms which the PADF can analyze. Note that in this work, the distances r and r' are equal. Figure adapted from [75].

2.6 Coherent diffractive imaging

2.6.1 Scattering from anisotropic systems

We employ two ways of calculating the expected scattering signal from a sample. In the DFT simulations, the electron density $\rho(r, \mathbf{R}, t)$ at time t for the entire molecule is utilized to calculate the scattering by the molecular form factor

$$f(q, \mathbf{R}, t) = \int \rho(r, \mathbf{R}, t) e^{-i\mathbf{q} \cdot \mathbf{r}} dr \qquad (2.55)$$

where $\mathbf{R}$ is the set of nuclear coordinates which are time-dependent. The variable $q = |\mathbf{q}|$ is the momentum transfer defined as,

$$q = |\vec{q}| = |\vec{k}_{in} - \vec{k}_{out}| = \frac{4\pi \sin \theta}{\lambda}. \qquad (2.56)$$

Here λ is the wavelength of the photon and θ is half the angle between the incoming photon wave-vector $\vec{k}_{in}$ and the outgoing one $\vec{k}_{out}$. The momentum transfer q can be related to real space distances d according to

$$d = \frac{2\pi}{q} = \frac{\lambda}{2 \sin \theta}, \qquad (2.57)$$

which is also referred to as "resolution". In this case, the intensity becomes

$$I(q, t) \propto |f(q, \mathbf{R}, t)|^2. \qquad (2.58)$$

This model accounts for chemical bonding, since the calculated electron density in the DFT simulation employs molecular bonds. The accuracy of the resulting scattering is dependent on the quality of the electron density, calculated from the chosen *ab-initio* formulation [76]. For large systems such as proteins, both the *ab-initio* method used to calculate the electron density, and the Fourier transform in equation (2.55) would require extensive computational resources.

Instead, it is common to ignore chemical bonds and utilize an independent atom approximation [76, 77]. In this case, the total form factor for the molecule is

$$f(q, \mathbf{R}) = \sum_{i=1}^{N_{atoms}} f_i(q) e^{-i\mathbf{q} \cdot \mathbf{r}_i} \qquad (2.59)$$

and the atomic form factor $f_i(q)$ is calculated from the electron density of atom i

$$f_i(q) = \int \rho_i(r) e^{-i\mathbf{q} \cdot \mathbf{r}} dr. \qquad (2.60)$$

The electron density $\rho_i(r)$ gives a spherically symmetric form factor [78]

$$f_i(q) = \frac{4\pi}{q} \int r \rho_i(r) \sin(qr) dr. \qquad (2.61)$$

Since the electron density is only calculated for an atom in equation (2.61), a high level of theory can be used to solve the atomic problem, and RSPT was therefore used in the calculations.

The theory of scattering used in this work is based on the Born approximation. It assumes each electromagnetic plane wave to either pass through the sample unperturbed, or is elastically scattered one time by an atom. Multiple scattering of the plane waves are ignored [79]. By knowing the time dependent form factor $f_j(q,t)$ of atom j and its position $r_j(t)$ and how they change with time due to radiation damage, one can determine the scattered signal on a detector. In the independent atom and Born approximation, the intensity is formulated as [80]

$$I(q) = r_e^2 P \Delta \Omega \int I_0(t) \sum_{j=1}^{N} \sum_{j'=1}^{N} f_j(q,t) f_{j'}(q,t) \exp(-i\mathbf{q} \cdot (\mathbf{r}_j(t) - \mathbf{r}_{j'}(t))) \mathrm{d}t.$$
(2.62)

$I_0(t)$ is the pulse profile of the probe, r_e the electron radius, $\Delta\Omega$ a pixel's solid angle, $P(\vec{q})$ the polarization correction depending on the photon polarization and the position of the pixel.

2.6.2 Scattering from isotropic systems

A liquid can be seen as an isotropic system, given that the molecules involved in the liquid will on a macroscopic scale occupy all of the 3D rotation group. One can calculate the real space rotationally averaged intensity by integrating equation (2.62) over the solid angle [81].

The coherently scattered intensity of a sample from X-rays depends on the partial structure factor between atomic species α and β, $S_{\alpha,\beta}(q,t)$ and form factors $f_\alpha(q,t)$, $f_\beta(q,t)$ as a sum of a self-scattering and inter-molecular term

$$I(q,t) = I_{self}(q,t) + I_{inter}(q,t) = \sum_\alpha c_\alpha f_\alpha(q,t)^2 +$$
$$\sum_{\alpha,\beta} (2 - \delta_{\alpha,\beta}) c_\alpha c_\beta f_\alpha(q,t) f_\beta(q,t) S_{\alpha,\beta}(q,t).$$
(2.63)

The parameter α represents the number of different elements present in the sample, with c_α being the number density of element α. The atomic form factor $f_\alpha(q)$ is defined as the Fourier transform of the electron density of an atom, which depends on the electronic configuration. This configuration will change dynamically due to the photon-matter interaction. The form-factor $f_\alpha(q,t)$ can be computed by weighting all observed electronic configurations given by the CR simulations,

$$f_\alpha(q,t) = \sum_j w_{\alpha,j}(t) f_{\alpha,j}(t),$$
(2.64)

where $w_{\alpha,j} = [0,1]$ ($\sum_j w_{\alpha,j} = 1$) is the weight and $f_{\alpha,j}$ the form-factor of element α and the electronic configuration j. The electron density used to

compute the $f_{\alpha,j}$ was derived using wave-functions in RSPT [69]. The partial structure factors $S_{\alpha,\beta}$ can be calculated as the Fourier transform of the time dependent RDF $g_{\alpha\beta}(r,t)$ in equation (2.54) [82] as

$$S_{\alpha\beta}(q,t) = \frac{4\pi\rho_0}{q} \int_0^\infty r(g_{\alpha\beta}(r,t)-1)\sin(qr)\,\mathrm{d}r. \qquad (2.65)$$

2.6.3 Patterson function

The Patterson function is used to study the real space correlations of a sample, using only scattering intensities without the need of phases. This method is in particular suitable to extract positional information of heavy atoms, since they scatter more strongly compared to lighter atoms. Given real space positions defined by $\mathbf{u} = (u,v,w)$, crystal indices $\mathbf{h} = h,k,l$ the Patterson function is computed using the electron density $\rho(\mathbf{r})$

$$P(\mathbf{u}) = \int_h \rho(\mathbf{r})\rho(\mathbf{r}+\mathbf{u})\mathrm{d}\mathbf{r}. \qquad (2.66)$$

In a scattering experiment the data observed on the detector contains only information regarding intensities and lacks phase information. The Patterson function can be retrieved by Fourier transforming the structure factors $F_{\mathbf{h}}$, which are determined form the diffraction intensities I on the detector through $\sqrt{I} = F_{\mathbf{h}}$. By summing over all reflections $\mathbf{h} = h,k,l$, the result becomes

$$P(\mathbf{u}) = \mathscr{F}^{-1}(F_{\mathbf{h}}^2) = \sum_{\mathbf{h}=-\infty}^{\infty} F_{\mathbf{h}}^2 \exp(-2\pi i\mathbf{h}\cdot\mathbf{u}). \qquad (2.67)$$

Equation (2.67) can also be computed theoretically given direct access to the positions of the atoms and their electronic occupations. The structure factors $F_{\mathbf{h}}$ are calculated using the atoms' time-dependent positions $\mathbf{x}_i$ and atomic form-factors f_i for a given reflection $\mathbf{h}$ as

$$F_{\mathbf{h}} = \sum_i f_i(q)\exp(i\mathbf{h}\cdot\mathbf{x}_i). \qquad (2.68)$$

2.6.4 Computing the effect of radiation damage

The usual approach to reconstruct the electron density in X-ray crystallography of biomolecules is to utilize neutral form-factors for fitting the model to the data [1]. However, as the cross section for photoionization is generally much higher than coherent scattering, the pulse will scatter of a sample which is not the native one, which is particularly true for an XFEL [4, 78]. Furthermore, additional processes due to the free electrons will further induce electronic changes in the atoms.

We utilized the radiation damage data from CRETIN for an XFEL pulse interacting with a protein crystal to calculate the corresponding electron density that the pulse coherently scatters from during the pulse, employing the independent atom approximation. If we define the electron density at time t for element Z as $n_Z(r,t)$ with corresponding form-factor $f_Z(r,t)$, then the weighted form-factor during the pulse is given by

$$f_Z(q) = \int g(t) f_Z(q,t) dt, \tag{2.69}$$

where $g(t)$ is the pulse shape of the XFEL with duration T, and is normalized as $\int g(t) dt = T$. At each time-step, the form-factor $f_Z(r,t)$ weighted according to the damage calculations in CRETIN is computed, as described by equation (2.64).

Depending on the properties of the probe, the electron density will differ and thus the form-factors. The process of computing the effects of the probe in the scattering factors makes it possible to directly understand the effect in real space through a Fourier transform, which can be compared to the native electron density. This enables one to compute how large the variance in the experimental data must be in order to note differences.

2.6.5 Pearson correlation

In an experiment, diffraction patterns are acquired which provide information regarding the electron density in reciprocal space. The diffraction patterns calculated using the theory described in this section can be used to estimate the highest obtainable resolution of the structure from the data. We use Pearson correlation [78, 83] to calculate the similarity between two diffraction patterns. Given a pattern as calculated by equation (2.62), we first determine the radial profile $I(q)$. This is obtained from averaging the data of the pixels over the angle $\varphi = [0, 2\pi]$ for a specific q

$$I(q) = \frac{1}{2\pi} \sum_{\varphi}^{2\pi} I(q, \varphi). \tag{2.70}$$

Given two diffraction patterns, the Pearson correlation or Fourier ring correlation (FRC) is defined as,

$$FRC(q) = \frac{\sum_{\varphi}(I_1(q, \varphi) - I_1(q))(I_2(q, \varphi) - I_2(q))}{\sqrt{\sum_{\varphi}(I_1(q, \varphi) - I_1(q))^2} \sqrt{\sum_{\varphi}(I_2(q, \varphi) - I_2(q))^2}}. \tag{2.71}$$

This function returns a value in the range [-1, 1], where the two extremes correspond to full correlation, and 0 means no correlation.

3. Results and discussions

## 3.1 Femtosecond bond breaking and charge dynamics in ultracharged biomolecules

In papers I and II, we apply density functional theory to study bond breaking and charge dynamics, using SIESTA [67]. Paper I concerns several amino acids, while paper II studies di- and tripeptides. In the second paper, we also study the resulting scattering signal from an ensemble of aligned peptides, comparing the neutral case to the highest charge state.

In both studies, we investigate the dynamics as a function of ionization. The degree of net charge is defined by how many electrons e^- have been removed, normalized by the number of atoms in the molecule, N. We define a parameter for this ratio called $\bar{z}$ with unit (e^-/N), which has a value in the interval $[0, 1]$. The highest charge considered here is therefore the case where every atom has lost one electron on average. The electron removed corresponds to the one with the lowest binding energy. We assume that the photon-matter interaction has resulted in a charged system, where any excited states have relaxed, before any movement has occurred. The inner shell orbitals are therefore all occupied. Thus, the ionization is set at the first time-step, and thereafter the dynamics is evolved, where each value of $\bar{z}$ corresponds to a separate trajectory.

3.1.1 Amino acids (Paper I)

The amino acids investigated are shown in figure (3.1).
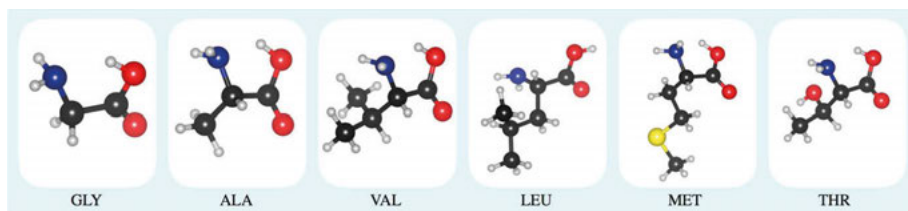
Figure 3.1. The six different amino acids studied, with glycine (GLY), alanine (ALA), valine (VAL), leucine (LEU), methionine (MET) and threonine (THR). The colors describe the atomic species, where carbon is in black, oxygen in red, nitrogen in blue, sulfur in yellow and hydrogen in white. Figure extracted from paper I.
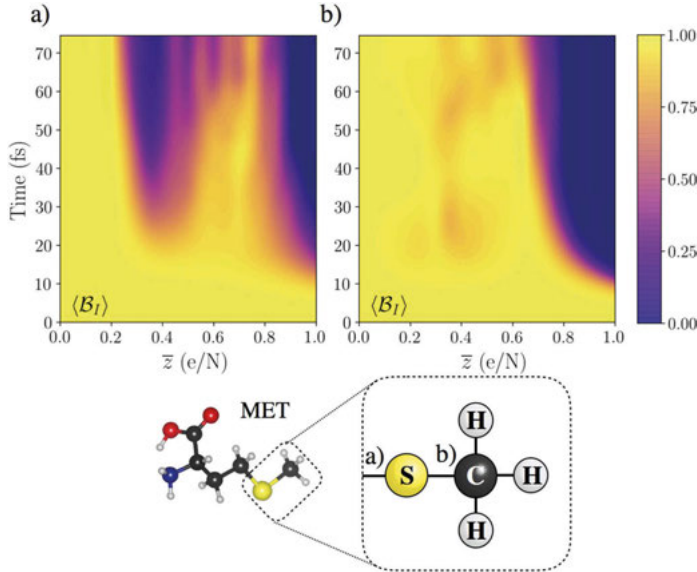
We present the result of the bond-integrity $\mathscr{B}(A,B,t)$ defined in equation (2.50), averaged over all six amino acids in figure (3.2). In the figure, a value of $\mathscr{B}(A,B,t) \approx 0$ means that the bond is broken, while $\mathscr{B}(A,B,t) \approx 1$ is a stable bond. It is clear that the C-C* bond in a) is the least stable in all amino acids for the 75 fs trajectory studied here, except for bonds to hydrogen in figure (3.3). For most charge states, this bond breaks after $t \approx 20$ fs. The bonds C-N and C-O are relatively stable until the highest charge states, where they start to break after approximately 15 fs. Details of the fragmentation dynamics in amino acids and changes due to the rapid loss of hydrogens are discussed in the next section together with the fragmentation in peptides.
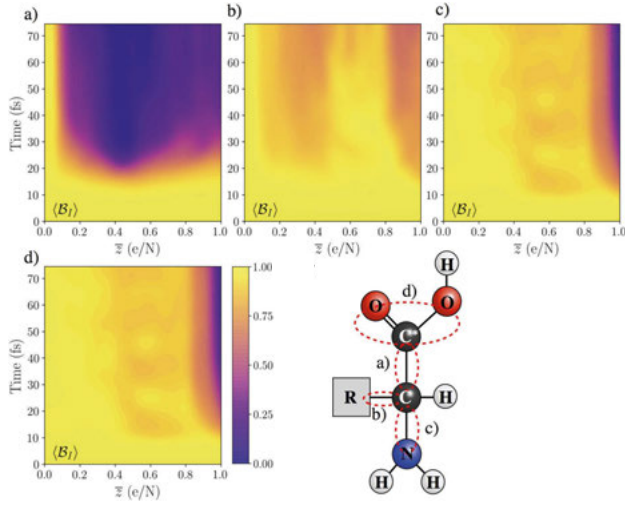
Figure 3.2. Average bond-integrity for all amino acids studied for multiple bonds. The schematic of the molecule shows which panel that the bond-integrity result of the particular bond is presented. The bond is intact for $\mathscr{B}(A,B,t) \approx 1$ and broken at $\mathscr{B}(A,B,t) \approx 0$ (see equation (2.50) for definition of bond-integrity). Figures extracted from paper I.
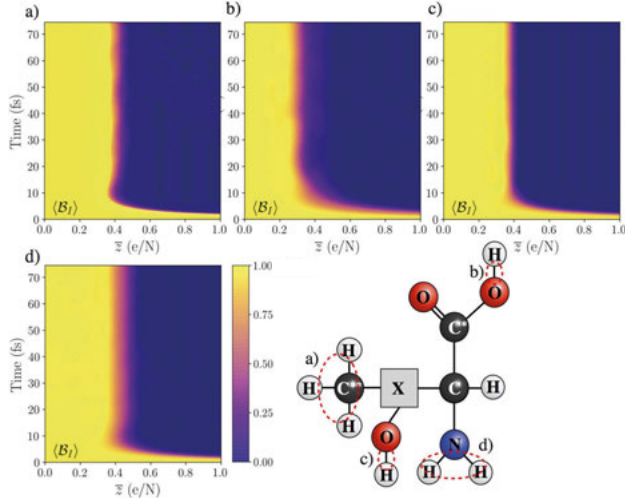
Figure 3.3. Bond-integrity for the bonds to hydrogen, averaged over all amino acids. The schematic of the molecule shows which panel that the bond-integrity result of the particular bond is presented. The bond is intact for $\mathscr{B}(A,B,t) \approx 1$ and broken at $\mathscr{B}(A,B,t) \approx 0$ (see equation (2.50) for definition of bond-integrity). Figures extracted from paper I.

Figure 3.4. Bond-integrity for methionine, which contains a sulfur atom. The schematic of the molecule shows which panel that the bond-integrity result of the particular bond is presented. In the panels, bonds to the sulfur (S) atom are shown. The sulfur binds to both the backbone in a), and the methyl group in b). The bond is intact for $\mathscr{B}(A,B,t) \approx 1$ and broken at $\mathscr{B}(A,B,t) \approx 0$ (see equation (2.50) for definition of bond-integrity). Figures extracted from paper I.
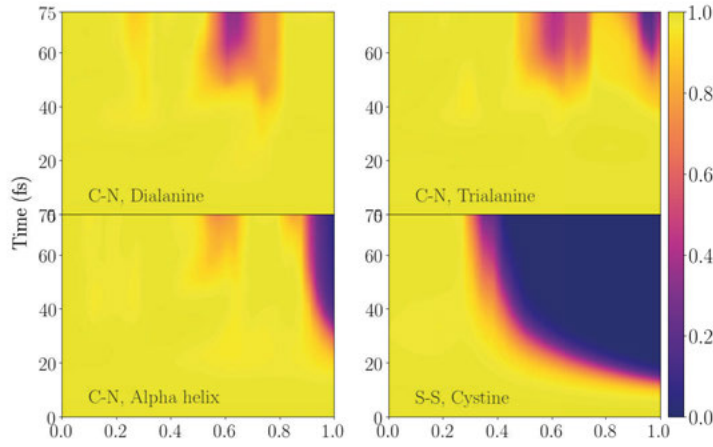
3.1.2 Peptides (Paper II)

Paper II uses the same framework as above, but the molecules in this work are di- and tripeptides, as shown in figure (3.5). Furthermore, the work presented here for the peptides is compared to the results of the amino acids.
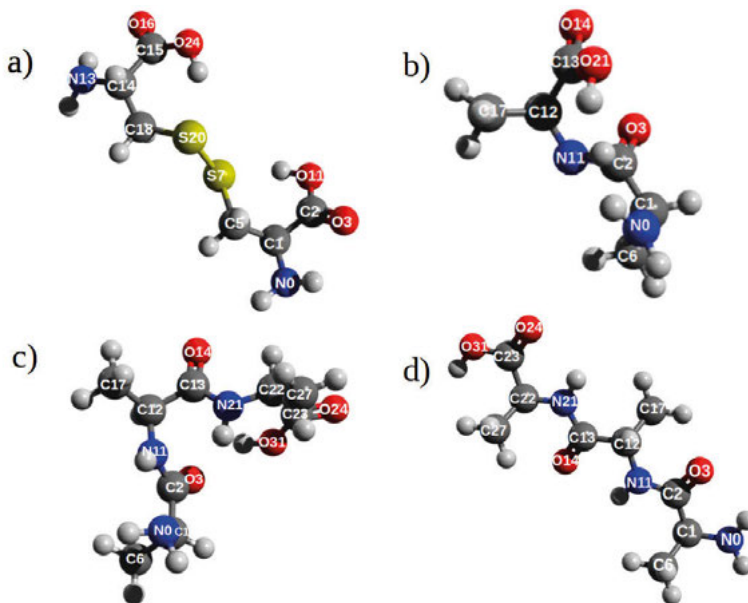
Figure 3.5. The polypeptides studied with a) cystine ($C_6H_{12}N_2O_4S_2$), b) dialanine ($C_6H_{12}N_2O_3$), c) trialanine ($C_9H_{17}N_3O_4$) and d) the alpha helix conformation of tri-alanine ($C_9H_{17}N_3O_4$). Hydrogen (H) is in white, oxygen (O) in red, sulfur (S) in yellow, carbon (C) in grey and nitrogen (N) in blue. The image was created using AVOGADRO [84]. The labels are used later on in the text. Figures extracted from paper II.

The molecules were constructed with AVOGADRO [84]. Linear conformation of the structures were created with angles $\phi = \psi = 180°$ [85, 86], where the angles are defined similar to the ones used in Ramachandran plots. The non-linear conformation of trialanine (called alpha helix) was created with angles $\phi = -60°$ and $\psi = -40°$. These molecules were chosen to study the effects of the number of amino acids in the peptide, the conformation and the chemical species on the subsequent bond breaking and charge dynamics.

Figure 3.6. The peptide bonds (C-N) for the alanine polypeptides, and the disulfide bond (S-S) in cystine. The bond is intact for $\mathscr{B}(A,B,t) \approx 1$ and broken at $\mathscr{B}(A,B,t) \approx 0$ (see equation (2.50) for definition of bond-integrity). Figure extracted from paper II.

Figure (3.6) provides a comparison between the peptide bonds in the trialanine molecules and the disulfide bond in cystine. It is clear that the disulfide bond is more sensitive to ionization. The start of the instability of cystine is around the degree of ionization corresponding to the where the bonds to hydrogen break, as seen in figure (3.8). After this point, the sulfur atoms' net charges start to increase quickly in figure (3.7). Fragmentation changes due to the loss of hydrogens was also noted for the amino acids. Comparing the peptide bonds of the alanine molecules, dialanine seems to be stable for the entire trajectory for $\bar{z} = 1$, in contrast to trialanine and the alpha helix. This is likely due to the fact that since $\bar{z} = 1$ in these molecules means a total ionization of +33, compared to +23 in dialanine, the Coulomb interaction on each atom will be higher. This will lead to stronger forces, even though the increase in size leads to more inertia.

In figure (3.4a), one can note that the bond to the sulfur has an interval $\bar{z} \approx [0.5, 0.8]$ where it stabilizes again. This particular interval of charge states correspond to those where the net-charge is larger or equal to the number of hydrogens. Thus, since the hydrogens leave as a proton, the remaining system is neutralized. This leads to more stability in the structure. We note that the re-stabilization due to the loss of charged hydrogens is a feature present for all molecules studied. Therefore, this behaviour is likely to be present in systems such as proteins. One can validate this neutralization in the peptides by studying the Hirshfeld charges defined by equation (2.53) of the atoms, as visualized in figure (3.7), where the positive charge of several of the atoms presented is reduced after the bonds to hydrogen are broken.
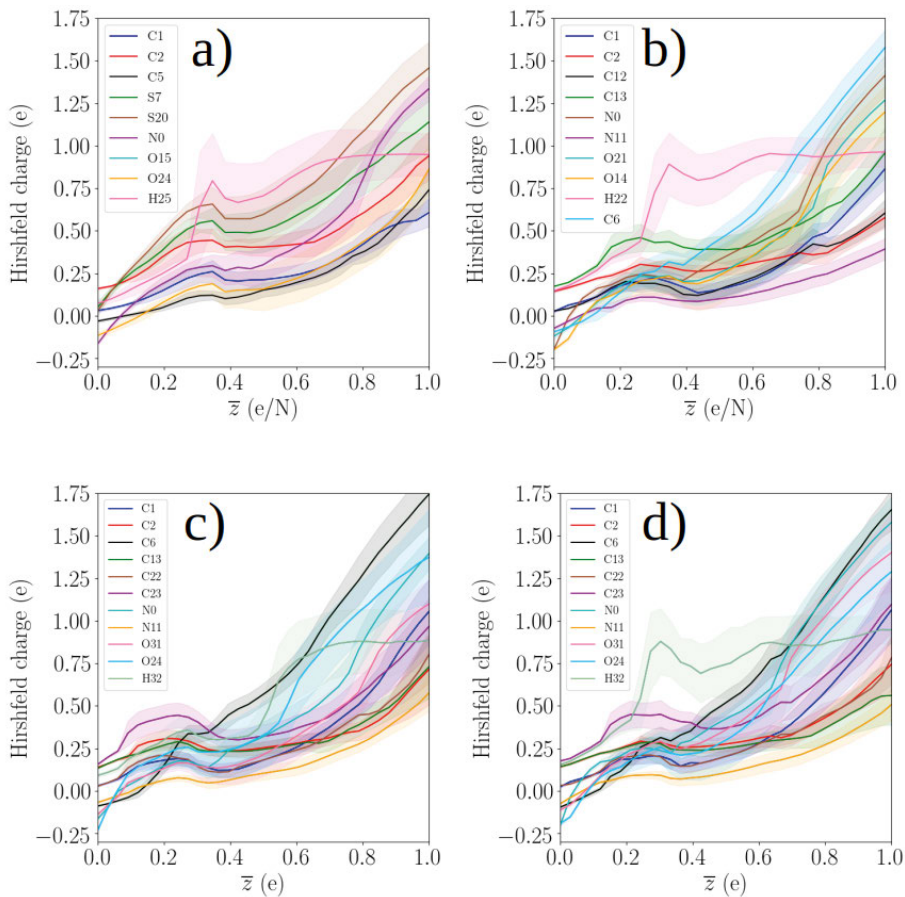
Figure 3.7. Hirshfeld charge dynamics as a function of ionization calculated using equation (2.53), for a) cystine b) dialanine c) alpha helix and d) trialanine. The label of the atom corresponds to those defined in the figure (3.5). The result is averaged over the entire trajectory and the number of simulations performed. Note that the increase in the Hirshfeld charge for the hydrogen atoms correlates with when the average charge of the system equals the number of hydrogen atoms in the system. This corresponds to when the hydrogen atoms leave the system as a proton. Figures extracted from paper II.
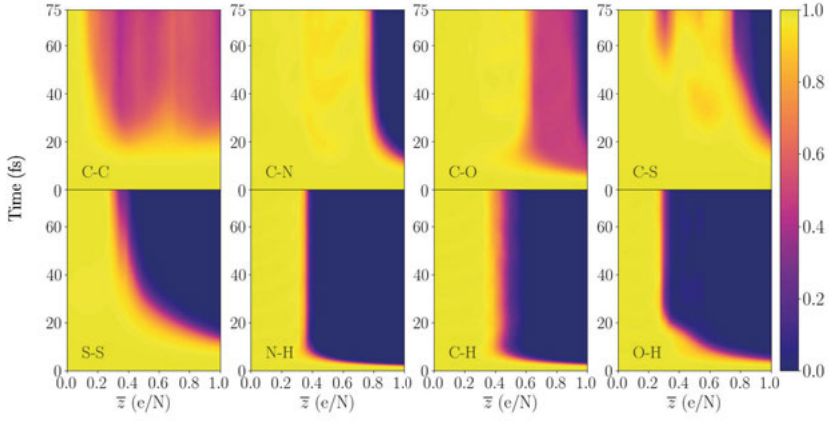
Figure 3.8. Bond-integrity of multiple bonds averaged over the bond type in cystine, as a function of ionization and time. The bond is intact for $\mathscr{B}(A,B,t) \approx 1$ and broken at $\mathscr{B}(A,B,t) \approx 0$ (see equation (2.50) for definition of bond-integrity). Figure extracted from paper II.
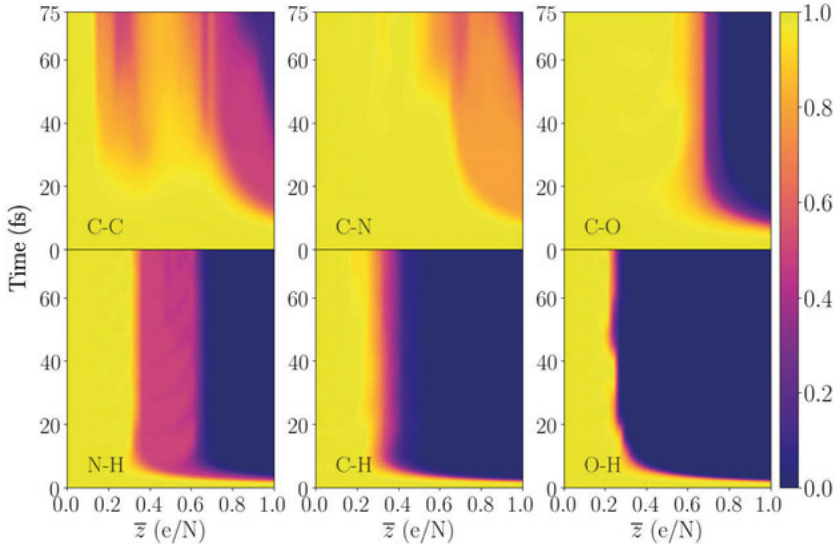

Figure 3.9. Bond-integrity of multiple bonds averaged over the bond type in trialanine, as a function of ionization and time. The bond is intact for $\mathscr{B}(A,B,t) \approx 1$ and broken at $\mathscr{B}(A,B,t) \approx 0$ (see equation (2.50) for definition of bond-integrity). Figure extracted from paper II.

The average bond-integrity for all bonds in the peptides cystine and triala-nine are shown in figures (3.8) and (3.9). The C-C bonds in the peptides are more stable than the same bonds in the amino acids. In particular, the bond re-stabilizes when bonds to hydrogens are broken around $\bar{z} = 0.4$. One can therefore conclude that as the polypeptide chain grows with the number of additional amino acids, the stability of this bond will increase.The C-O bond becomes significantly more sensitive to breaking for all peptides, while the C-N bond increases in instability mostly in cystine and the alpha helix compared to the amino acids. Finally, we see that most bonds to hydrogen all break at around $\bar{z} = 0.4$, similar to the amino acids in figure (3.3). The results of the
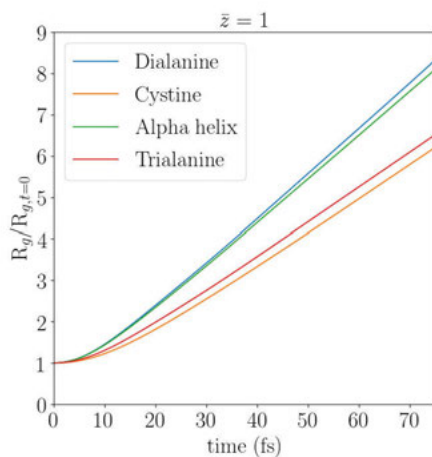
Figure 3.10. The normalized radius of gyration (R_g) for all molecules studied. The result is an average of 10 different trajectories, for the highest ionization $\bar{z} = 1$. Lower degrees of ionization provided a similar result. Figure extracted from paper II.

radius of gyration defined by equation (2.51), depicted in figure (3.10), show that the alpha helix and dialanine explode on shorter time-scales compared to the other molecules. We note that there is a dependence on the conformation of the molecule on the dynamics of explosion, since the alpha helix dissociates faster compared to trialanine. Even though the bond-integrity for trialanine and the alpha helix are relatively similar for most bonds, especially for $\bar{z} = 1$ as seen in for instance the peptide bond in figure (3.6), the radius of gyration is different. We theorize that the difference in the rate of expansion is due to the difference in the conformation. Since the alpha helix is more densely packed, there will be a higher density of charged atoms. The Coulomb potential, which will dictate the result of the dynamics, is dependent on the relative positions of the atoms. This leads to higher forces in the alpha helix, and therefore a more violent explosion.

The conclusion of this work, is that the bond-integrity is not greatly affected by the conformation of the same molecule. The chemical species in the system, and the particular sequence of amino acids provide the largest variance. However, the expansion of the molecule is mostly dependent on the size of the system and its particular fold.

### 3.1.3 Imaging of aligned biomolecules in the gas phase

The study of the dynamics in reciprocal space, comparing the neutral and highest charged state ($\bar{z} = 1$) is done by calculating diffraction patterns using CONDOR in paper II [77]. The intensity for each time step is calculated by extracting the electron density in the trajectory from the DFT simulations. For each time-step, the data is added incoherently to simulate the time-integrated signal on the detector. The 10 different trajectories are aligned in space the same way relative to the incoming beam. Thereafter, the 10 different diffraction patterns are incoherently added. This corresponds to the diffraction from an ensemble of perfectly aligned molecules in the gas phase, where we have presumed that each molecule in the ensemble are separated enough such that the scattered signal do not interfere.

We provide a comparison of the diffraction patterns from the trajectories with neutral charge ($\bar{z} = 0$) and those with ($\bar{z} = 1$) for momentum transfers $q = [0, 0.6]$ Å^{-1}, in figure (3.11). The features of the patterns show that the native structure provides a larger difference between the highest and lowest signal, compared to the damaged one, for all molecules. The charged molecule results in a different curve for the integrated intensities in figure (3.12). This is true at time $t = 0$ fs and for the signal integrated over the entire trajectory, $t = 30$ fs. By normalizing the signal at $q = 0$, the number of electrons are the same. The difference is therefore due to the changes in the electron density. We note that higher resolutions have more signal for the charged structure at $t = 0$ fs, which concludes that the electrons are more localized in space. This is because for a charged molecule, the remaining electrons will be less screened by the nuclei. Furthermore, at this point in time, the atoms are still in their initial positions. This type of change in the electronic structure will always exist, independent if the pulse is short enough to outrun the timescale of atomic motion. The relocation of the electron density at short timescales can also be noted in the time-resolved intensity at each value of the momentum transfer q, as seen in figure (3.13). We note that for all the molecules in this figure, the scattering intensity drops for the damaged trajectories for the range around $q < 0.2$ Å^{-1}, but at $q > 0.2$ Å^{-1} it is higher.

Finally, it seems feasible to detect the difference between the conformations in reciprocal space. Since the electrons in the alpha helix are more densely distributed, there will be a smaller difference between the low and high resolution signal, as can be seen by comparing c) with e) and d) with f) in figure (3.11).
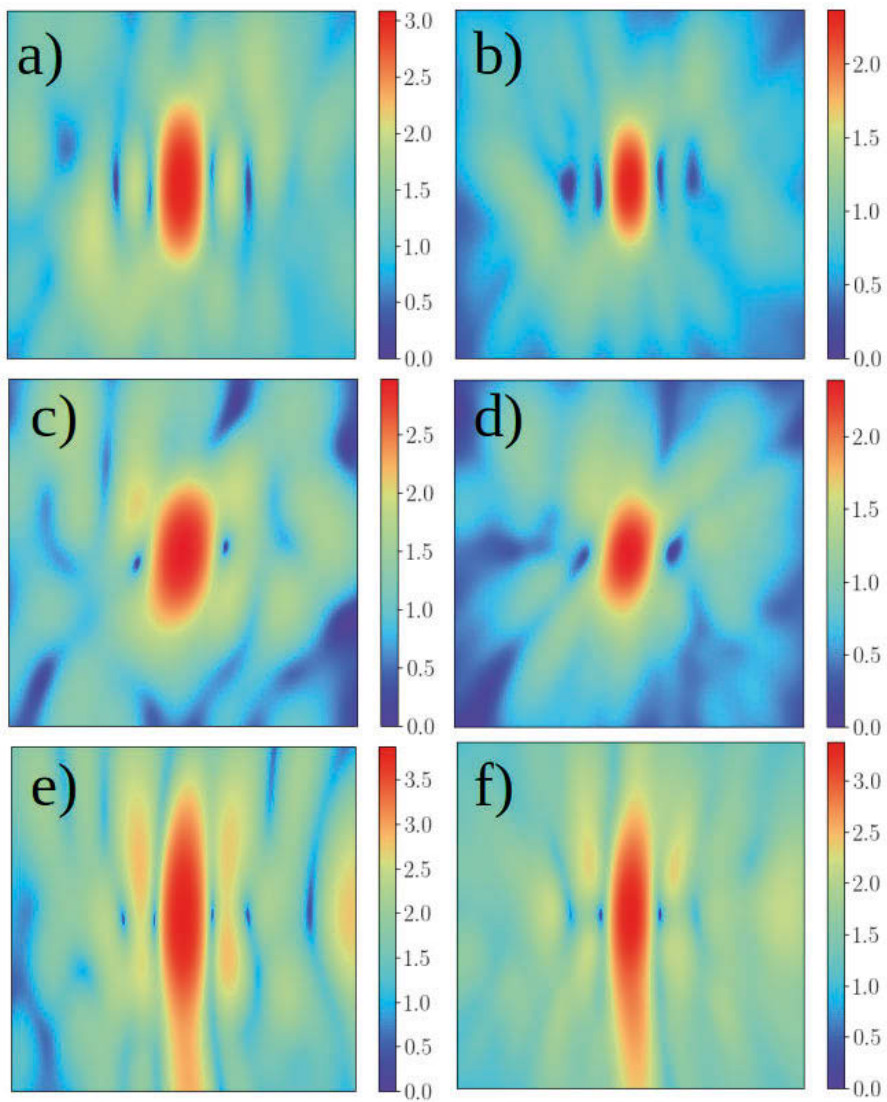
Figure 3.11. Diffraction for three of the peptides. (Left) undamaged ($\bar{z} = 0$) diffraction patterns and (right) damaged ($\bar{z} = 1$). Cystine is shown in a), b), the alpha helix in c), d) and trialanine in e), f). The maximum momentum transfer is around $q = 0.6$ Å^{-1}, which is a spatial resolution of $d \approx 1.66$ Å. The unit-less values in the colorbar correspond to the logarithm of the ratio of the maximum value of the diffraction pattern and its minimum value. Figures extracted from paper II.
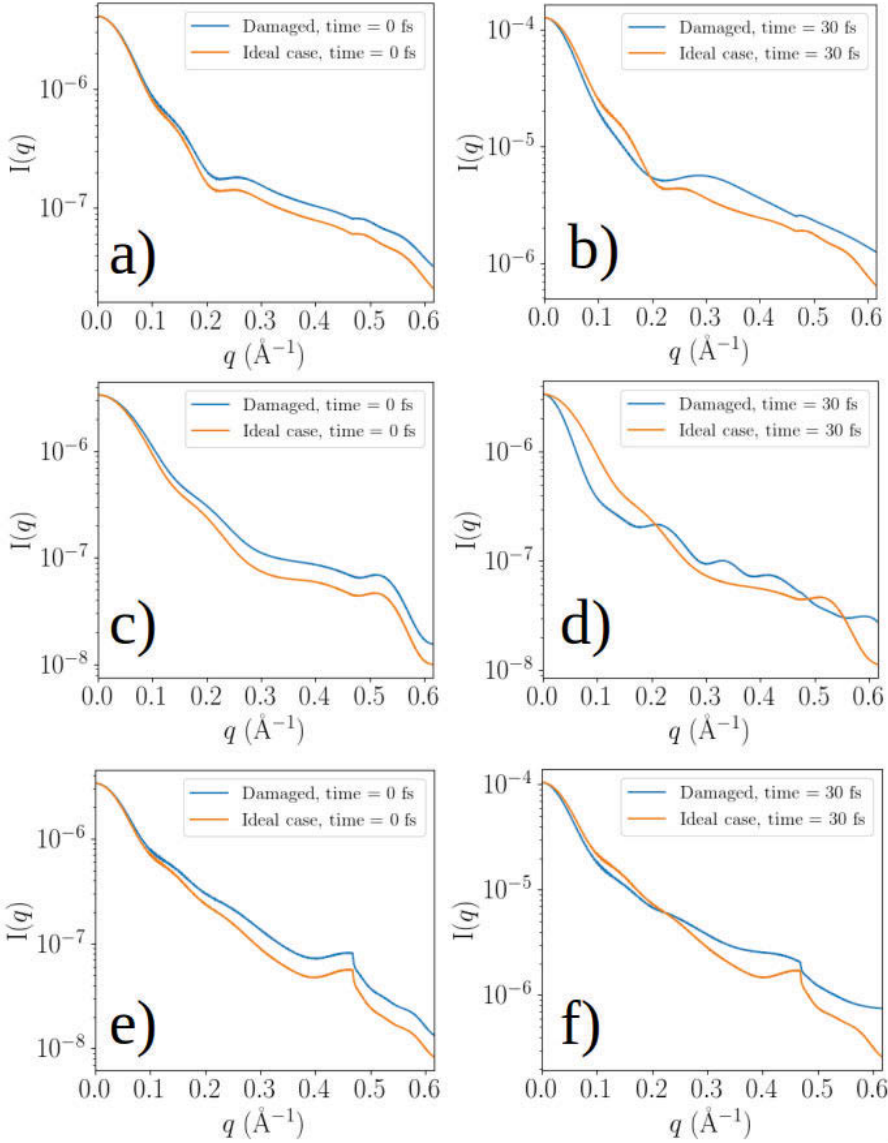
Figure 3.12. (Left) Integrated intensity comparing the damaged ($\bar{z} = 1$) case to the neutral ($\bar{z} = 0$) one at time 0. (Right) The same but after 30 fs of the trajectory. (Top) Cystine, (middle) alpha helix and (bottom) trialanine. The damaged curve has been normalized such that it is equal to the undamaged one at $q = 0$. Figures extracted from paper II.
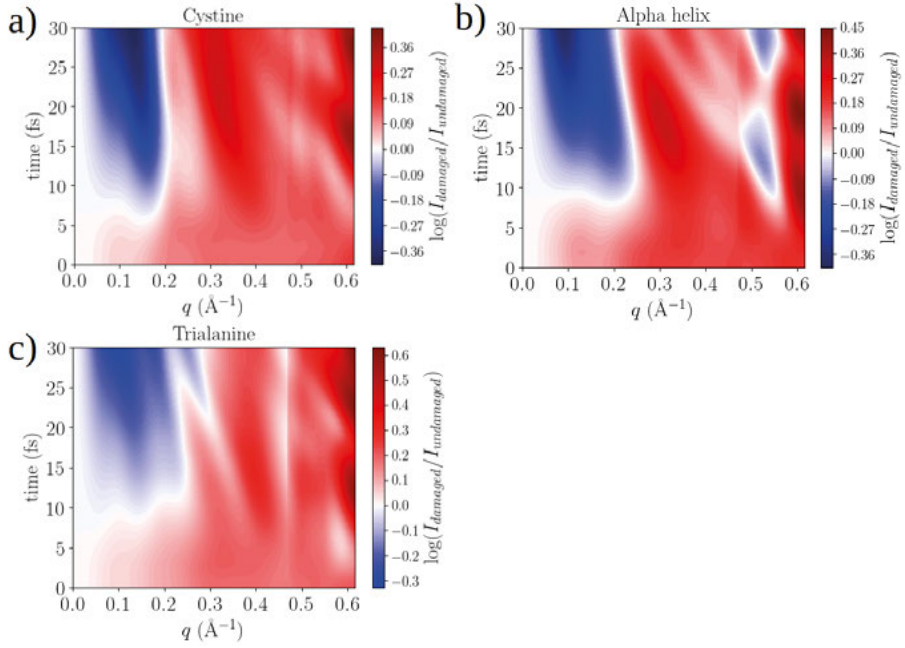
Figure 3.13. Comparison of the damaged intensity ($\bar{z} = 1$) to undamaged ($\bar{z} = 0$), shown as a logarithm of the ratio ($\frac{I_{damaged}}{I_{undamaged}}$). The panels show the ratio as a function of time (time-resolved intensities) and momentum transfer q, for three different peptides. The intensities shown here are closely related to figure (3.14) where the time-resolved intensities have been integrated. Normalization has been done to ensure the damaged and undamaged intensities are the same for $q = 0$. The plot maps the regions in momentum space where the scattering intensities for the damaged molecules are higher relative to the undamaged (at the same normalization) and could provide a way to distinguish between damage and undamaged cases. Figures extracted from paper II.

3.2 MOLDSTRUCT: modelling the dynamics and structure of matter exposed to ultrafast X-ray lasers (Paper III)

This work presents a newly developed tool MOLDSTRUCT to study photon-matter interaction using hybrid collisional-radiative calculations with classical molecular dynamics. The model is compared to two experimental studies using XFELs, one where scattering from non-thermally heated bulk water was studied and one where disulfide bond breaking due to an X-ray pump-probe scheme was investigated. In both cases, our model shows good agreement. We then apply the model to study fragmentation dynamics and coherent imaging in a cluster in the context of radiation damage in single particle imaging.

(a)

(b)
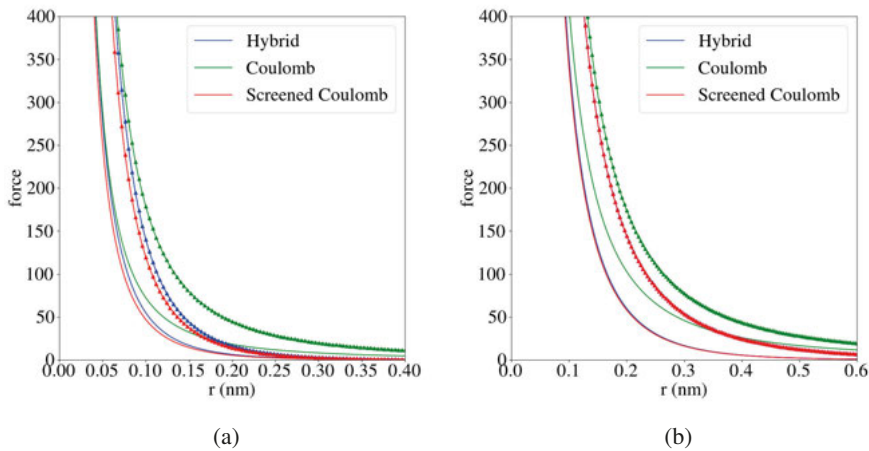
Figure 3.14. Coulomb force as a function of distance and different time-points during the interaction of an XFEL pulse and bulk water. This is done in for an intensity a) 10^{18} Wcm^{-2} and b) 5×10^{19} Wcm^{-2} using the hybrid screening (in blue), Debye screening (in red) and regular Coulomb model (in green). The pulse has a photon energy of 6860 eV and is 75 fs long. Figures extracted from paper III.

Figure (3.14) shows a comparison between the hybrid screening potential defined in equations (2.14, 2.15 and 2.16), Debye screening model as defined in equation (2.10) and the standard Coulomb interaction. It can be noted that the Debye screening model reduces the Coulomb interaction more than the hybrid model, mostly for the lower intensity in a). For the higher intensity in b), the difference is reduced. Interestingly, the transition point in the hybrid model between the ion-sphere (IS) and Debye screening (DS) model decreases as a function of time for the lower intensity, and increases instead with time for the higher intensity. This means that for lower intensities for this photon energy and sample, the system transitions more towards the ion-sphere model, while for the higher intensity, the system transitions to the Debye screening model.

3.2.1 Comparison to experiments

We achieve good agreement when comparing the experimental scattering curves [87] from bulk water with the simulations in figure (3.15). The results are obtained with intensities lower than reported in the experiment and the ratio of the intensity between the two pulses is also reduced. We note in figure (3.16), which shows a large range of intensities and the corresponding scattering curves, that a new peak forms as the intensity is increased. Since there is no indication of this in the long experimental curve, one could think that the experiment does not have sufficient intensity to induce this type of structural change. Alternatively, the model can be overestimating the charge states that the real experiment sees. This would lead to larger forces and more rapid structural disorder.
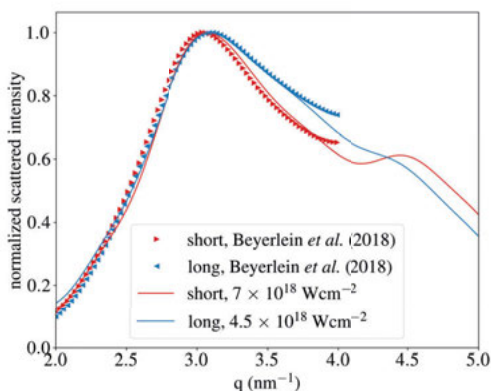

Figure 3.15. Simulations of scattering from bulk water compared to experiment [87], for the short (25 fs) and long (75 fs) pulse duration using the best matching intensities. Figure extracted from paper III.

Apart from studying liquids, the model can be used to study radiation damage in the context of structure determination using SFX. We compare the results of the model to an SFX experiment conducted using an X-ray pump-probe scheme seen in figure (3.17). We explored three different intensities, and note that depending on the time-delay, different intensities agree with the experimental data-points. This could be attributed to the experiment having lower intensities in the interaction region, compared to what was reported in the article. Another possible reason could be due to Bragg termination [88, 89], where parts of the crystal which are initially exposed to the peak of the spatial pulse profile are destroyed and do not contribute to the Bragg spots. By using a low peak intensity, the part of the system which still contributes to Bragg spots, and is exposed to the tail of the spatial profile, will not provide enough signal over the background. Thus, by using a higher peak intensity, the tail of the spatial profile increases and contributes signal over the noise.
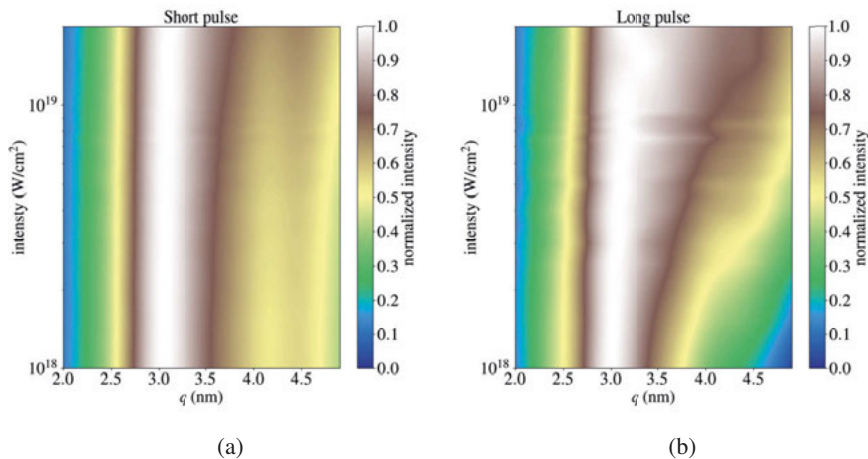
Figure 3.16. Integrated scattering as a function of intensity and momentum transfer for the a) short pulse (25 fs) and b) long pulse (75 fs). Figures extracted from paper III.
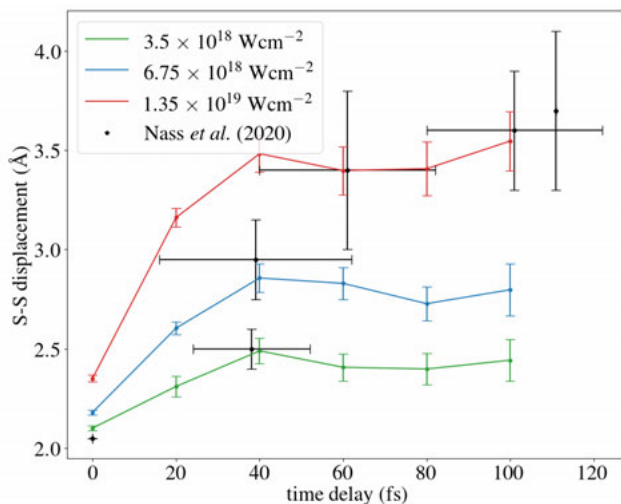

Figure 3.17. Pulse weighted displacement between the disulfide bond (S-S) as a function of probe delay, computed for multiple intensities. The data is averaged over all disulfide bonds in the Cystine amino acids. Nass *et al.* (2020) corresponds to the experimental data [45] and is extracted through visual inspection. Figure extracted from paper III.

3.2.2 Application of model: radiation damage in SPI

To run hybrid CR/MD simulations of a cluster, we calculated the size of the cluster required to trap the free electrons produced by the pulse. This is presented in figure (3.18) and shows that by using a cluster with radius larger

than 42 Å, the free electrons are trapped by the potential of the cluster within 2 fs. This allows the use of a plasma description to model the dynamics of the cluster. In the simulations, we utilized a cluster with a radius of ≈ 5.5 nm. To show the code's capabilities in the context of SPI, we study radiation dam-
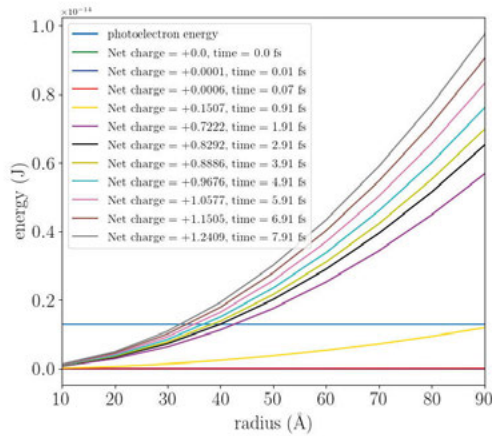

Figure 3.18. Comparison of photoelectron energy and electrostatic energy from the cluster during the photon-matter interaction. The electrostatic energy from the cluster is computed as a function of the net charge density and the radius of the sample. The legend indicates the net charge of the system at the corresponding time-point as given by the CR simulation. Figure extracted from paper III.

age in a methane cluster. This system is used as a model for more relevant macromolecules like proteins and viruses. We conduct simulations based on a scheme which could be used in an experiment to study radiation damage. By first exposing the methane cluster to an X-ray pump and after some time-delay to an X-ray probe, the effects of radiation damage can be followed. We use a pulse with an intensity of 10^{19} Wcm^{-2}, a photon energy of 8 keV and FWHM of 15 fs. The mass density of methane calculated from averaging during each probe delay is found in figure (3.19), and contains a comparison between including electron-ion coupling and when it is omitted.

It can be concluded the mass density changes with increasing probe delay. We note an increase in the density at larger distances from the center of mass and a decrease closer. This is due to the expansion of the cluster. The hydrogen atoms are the fastest atoms leaving the molecule. Previously, we have shown [62] that the hydrogen atoms are important for the stability for a system. By using pulse parameters which favor ionization of the hydrogen atoms, one could potentially reduce the displacement of the heavier atoms which scatter the strongest. Investigating the effect of the electron-ion coupling, we see the expansion of the system occurs slightly faster with electron-ion coupling. This is because additional energy is transferred to the ions and they therefore reach

higher kinetic energies. However, on the time-scales explored here, which are typical for SPI, we do not see a significant difference for the resolution used. We further studied the fractional ion yield in figure (3.20) to get detailed un-
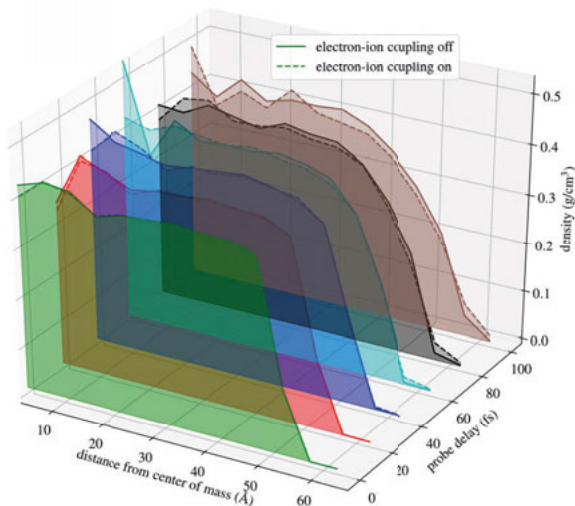

Figure 3.19. Mass density of a 5.5 nm methane cluster as a function of distance, probe delay and distance from the center of mass. The density for each probe delay is averaged during the probe pulse duration. A resolution of 5 Å was used to sample the radius of the sphere. The dotted line corresponds to a simulation with electron-ion coupling, and filed is without. Figure extracted from paper III.

derstanding of the effects of electron-ion coupling on the fragments produced. For a short probe delay of 0 and 20 fs in figures (3.20a) and (3.20b), we see a relatively small difference between the result with electron-ion coupling on and off. For the longer time delay of 40 fs in (3.20c) this difference is more noticeable. This concludes that for short enough time scales, the result will not be dependent on energy transfer from the electrons to the ions. The difference when using coupling comes from the electrons adding additional energy into the system and therefore allows for new fragments to form. It is clear that when electron-ion coupling is not included, the X-ray pulse is the major factor for a change of the ion yield. This is indicated by the convergence of the ion yields as the pulse terminates, and a change as the probe hits the sample at 40 fs in figure (3.20c).
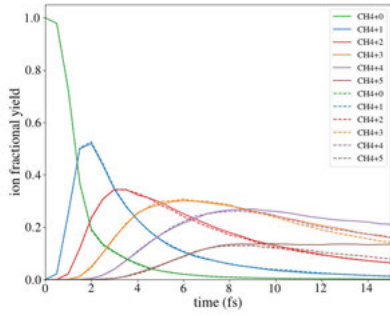
(a) 0 fs probe delay.
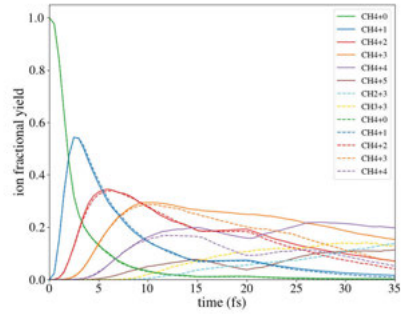
(b) 20 fs probe delay.
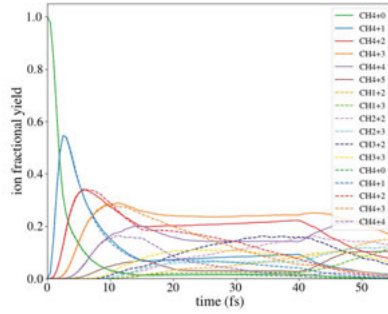
(c) 40 fs probe delay.

Figure 3.20. Time-evolution of the fractional ion yield of the methane cluster for different probe delays. The legend shows how many carbon/hydrogen atoms are in the fragment together with the net charge. The dotted lines correspond to simulations with electron-ion coupling, while the solid lines are without. The sudden change for the ion yields at 20 fs in b) and 40 fs in c) is due to the introduction of the probe pulse in the simulation. Figures extracted from paper III.

3.3 Anisotropic melting of ice induced by ultrafast non-thermal heating (Paper IV)

This work explores the interaction of an XFEL and hexagonal ice. We follow the changes to the structural and electronic dynamics induced by the X-rays using the newly developed hybrid model MOLDSTRUCT. The dynamics is analyzed in real space using pair angle distribution functions (PADFs) and in reciprocal space through elastic scattering. The results show that a crystalline plasma is formed, where the initial structure of the crystal is maintained for several tens of femtoseconds, while turning into a plasma phase. Furthermore, the PADFs, which show the frequency of key pairs of angles and distances, disappear on different time-scales, revealing an anisotropic melting process.
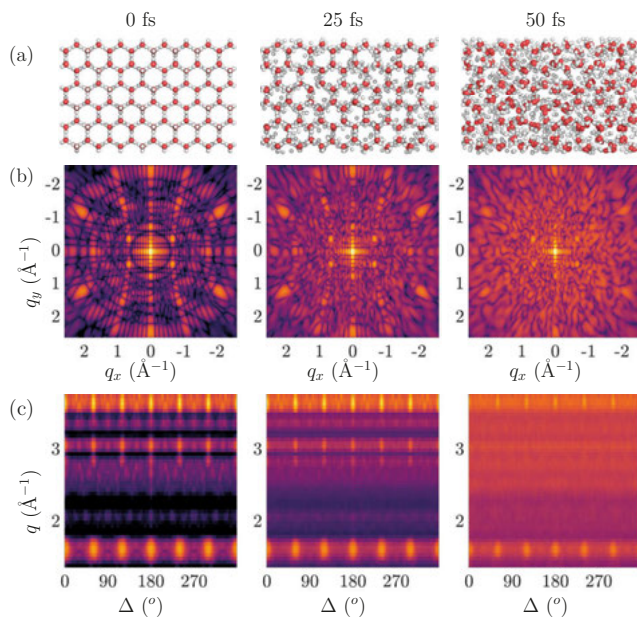
Figure 3.21. Structural dynamics in hexagonal ice induced by an XFEL pulse, simulated with the CR/MD code. The intensity of the probe is 1×10^{18} Wcm^{-2} and the photon energy is 8 keV. (a) Real space dynamics of ice at 0, 25 and 50 fs, with oxygen in red and hydrogen in grey. (b) Signal in reciprocal space from a single-shot at the same time-points, including both structural and electronic damage. (c) Correlation of intensity as a function of azimuthal angle and momentum transfer q. Figures extracted from paper IV.

To connect the theoretical study to future experiments, we determine the time-resolved elastic scattering of the sample. This is done by computing the trajectory given by the MOLDSTRUCT model and extracting the RDF of oxygen-oxygen (O-O), oxygen-hydrogen (O-H) and hydrogen-hydrogen (H-H). At each time-point we also determine the electronic occupation weighted form-factor of each atomic species according to equation (2.64). The RDF and the

weighted form-factors are then used to compute the scattering given by equation (2.63). An example of the RDF and the corresponding isotropic inter-molecular scattering from the sample is shown in figure (3.22). We see in the
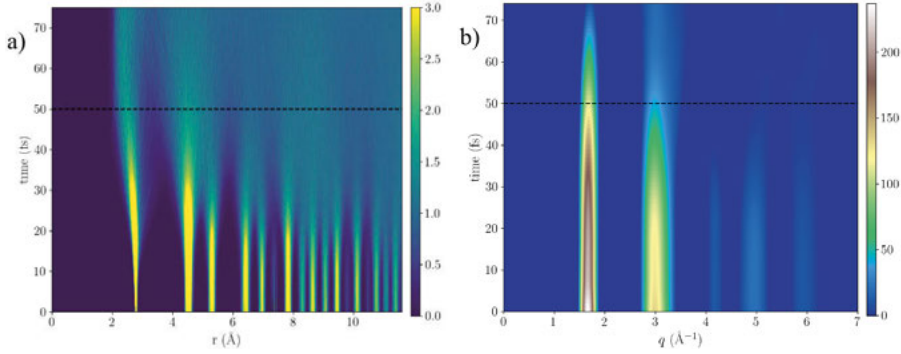
Figure 3.22. Radial distribution function for O-O in ice as a function of time, given by an XFEL pulse of 50 fs, with an intensity of 10^{18} Wcm^{-2} and photons with 8 keV energy. b) Time evolution of the inter-molecular intensity $I_{inter}(q,t)$ defined as a Fourier transform of the RDF in equation (2.63), which includes contribution from atomic displacement and changes in the form-factors due to electronic damage. Scattering due the hydrogen atoms contributes weakly. The results in both figures are averaged over multiple trajectories with different charge distributions. The colour scale is in arbitrary units. Figures extracted from paper IV.

RDF that the first O-O peak starts to decay substantially compared to the initial value after around 30 fs. Similarly, we note that the scattering signal in figure (3.22b) still provides a strong signal relative to the initial time-point for all the available peaks. Figure (3.23) shows the scattering from higher XFEL

(a)   (b)
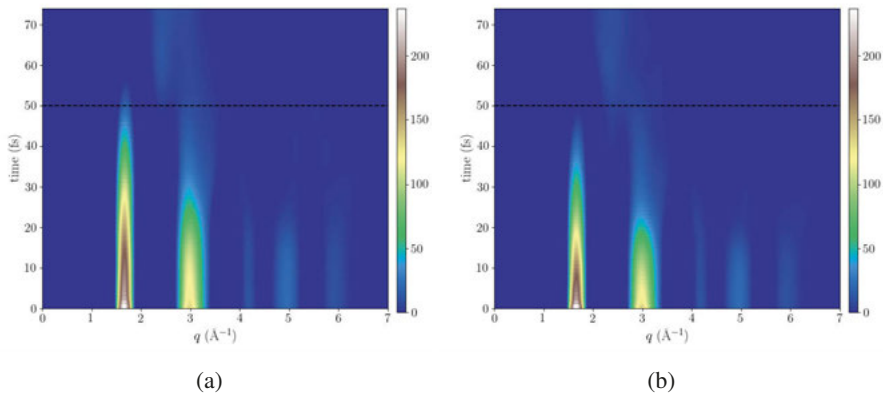
Figure 3.23. Inter-molecular intensity for a) 5×10^{18} Wcm^{-2} and b) 10^{19} Wcm^{-2} b) Time evolution of the inter-molecular intensity $I_{inter}(q,t)$ defined as a Fourier transform of the RDF in equation (2.63). Figures extracted from paper IV.

intensities of 5×10^{18} Wcm^{-2} in a) and 10^{19} Wcm^{-2} in b). One can note apart from the dynamics being faster, that there is a new peak emerging for both intensities at around 2.9 Å^{-1}, indicating a new structural phase is appearing.

To determine the time-scales of phase transition from crystalline to the plasma phase, we tracked the signal of the first three peaks in the scattering for all XFEL intensities. The time-point for when the intensity was reduced to half of the initial intensity was extracted and is shown in figure (3.24) in the top three lines, as function of incoming intensity.
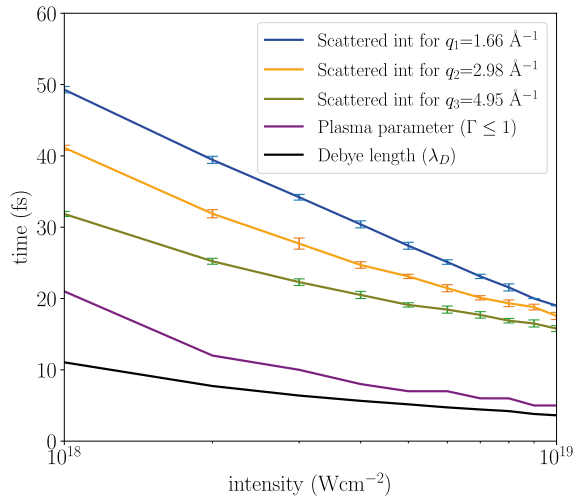

Figure 3.24. Time evolution of the phase transition from the crystalline to plasma phase for ice with respect to the intensity of the XFEL. Starting from the bottom, the first (black) line indicates when the Debye length (λ_D) reaches its asymptotic value. The second line corresponds to when the plasma parameter reaches $\Gamma = 1$, and the three final lines give the points in time when the scattered intensities at the particular q values that reach half their initial intensity. These data-points are averaged over several simulations with alternating distributions of charge, with the error-bar indicating the standard deviation. The figure shows the time-scale for the phase transition from the native crystalline state to the plasma phase, depending on the incident intensity. From the time between $\Gamma = 1$ and the blue line, one can observe a transient state of matter which is simultaneously a crystal and in a plasma phase. Figure extracted from paper IV.

The other two lines at the bottom of figure (3.24) correspond to quantities relating to the state of the plasma. When the X-ray laser interacts with the sample, electrons will be released and a plasma will be formed. The electron cloud will screen the charged ions, resulting in a decrease of the Coulomb interactions. The reduction is quantified based on the Debye length, as described in the method section. In the beginning of the simulation, no screening will exist and

therefore the screening length can be considered infinite. The time-point in the simulation where the Debye length reaches its asymptotic value is shown by the black line, which is much faster than typical atomic motion during the same time interval. The rate of phase transition from the crystalline state to the liquid phase is heavily reduced by incorporating screening.
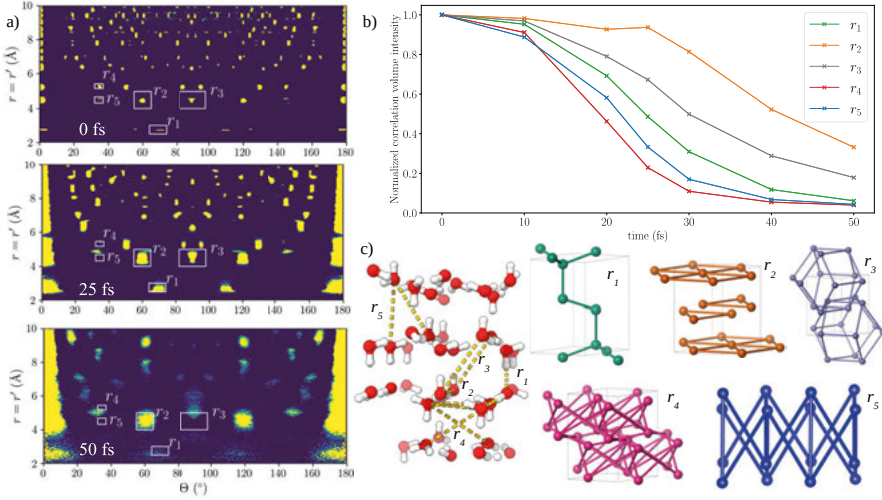
Figure 3.25. (a) The PADFs for $r = r'$ of ice exposed to an XFEL with intensity 10^{18} Wcm^{-2} at 0, 25 and 50 fs. Yellow corresponds to high intensity of particular values of pair-angles, where the different geometries observed are labelled r_{1-5} and shown in c). (b) Normalized intensities for the peaks r_{1-5} indicating the stability of the geometrical arrangements. (c) Visualization of the geometries corresponding to arrangements of oxygen atoms (red spheres) contributing to the peaks in a). Figure extracted from paper IV.

In order to extract information regarding the structural dynamics beyond the RDF, we analyzed the MD trajectories using the PADF [90, 91, 74]. It reveals the likelihood of having particular repeating many-body structures in the sample, based on angles and distances. In figure (3.25a) the PADFs for 0, 25 and 50 fs are shown for the lowest intensity of 10^{18} Wcm^{-2}, where the yellow color indicates high signal and dark low signal. Several points are enclosed by a white box, which are particular structures r_{1-5} shown in figure (3.25c) that we followed during the MD trajectory. The first time-point in (3.25a) shows the many-body correlations that are available in the native structure of hexagonal ice. As the pulse interacts with the sample and time evolves, we see for 25 and 50 fs that the signal of the peaks start to decrease. Some of them eventually disappear, while others are still visible at 50 fs, but have started transitioning to other pair-angles. By looking at r_1 we note that there is an indication of correlated motion between the atoms, where the angle is kept, but the dis-
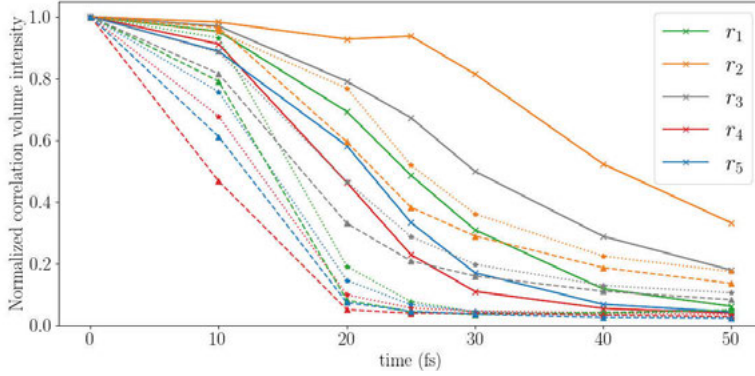
Figure 3.26. Correlation intensity for PADFs given by three intensities of 1×10^{18} Wcm^{-2} (solid line), 5×10^{18} Wcm^{-2} (dotted line) and 10^{19} Wcm^{-2} (dashed line). The same trend where $r_{1,2,3}$ are kept longer and $r_{4,5}$ disappear faster is seen for all intensities. Figure extracted from paper IV.
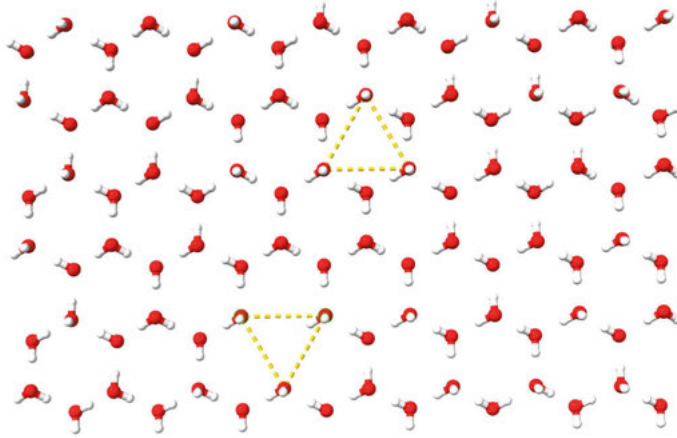
Figure 3.27. A subset of the r_2 PADF contact showing that some hydrogens are directed similarly. These pair-angles are preserved the longest up to 25-40 fs, depending on the intensity. Figure extracted from paper IV.

tance increases. This means that the PADF can reveal information about the direction of the phase transition.

In figure (3.25b), we show the integrated intensity of the peak, normalized to the signal at the initial time. The structures disappear at different rates on the femtosecond time scale, thus revealing the order to disorder transition occurs in an anisotropic manner. Different crystalline planes disappear at different rates, where the hexagonal planes in r_2 are preserved for longer compared to

r_4, which corresponds to a structure spanning between different planes. Additionally, r_2 contains hydrogen bonds with similar direction as seen in figure (3.27). This could result in correlated motion and thus longer conservation of the structure. We note that the shortest r_1, which occupies in-plane and out of the plane, disappear faster than r_2 and r_3. Compared to r_3, which also spans multiple planes, r_1 occupies short distances, which tend to disappear faster than longer distances. This result is in agreement with other X-ray imaging experiments, which have shown that high values of momentum transfer diminish fastest [92, 93], which in real space means small length scales. The r_4 contact contains 4-bodies and disappears more rapidly than the 3-body contact r_5. New parts of the PADF space can also be seen to be occupied during the interaction of the pulse, like the peak that is created near r_4 and r_5. A transient geometric configuration is created, which could be due to the force field pushing the state into a new minima. Investigating the integrated peak as a function of time for higher intensities in figure (3.26), the conclusion is that the anisotropic part of the phase transition can be seen for a higher intensity, however the transition is faster.

The model is built on the assumption that the probe turns the sample into a plasma in the first few femtoseconds, before any significant atomic movement has occurred. Thus, the force field used in this study involves interactions only based on spherical symmetry. Therefore the dynamics is largely dependent on the native geometry of the system and the dynamics of the phase transition could be predicted from it. Experimentally, the dynamics could be followed using an X-ray pump-probe setup available at an XFEL facility. In this experimental setup, an X-ray pulse pumps the sample, resulting in the start of structural dynamics. After some time-delay, another X-ray pulse probes the sample providing a snapshot of the dynamics of the sample. By varying the time-delay, one can follow the structural evolution of the sample in detail.

3.4 Radiation damage in a hemoglobin crystal exposed to an ultrafast X-ray laser (Paper V)

We performed theoretical studies where a hemoglobin crystal is probed by a short XFEL pulse, in order to compare to experimental results. In the experiment performed at the Linac Coherent Light Source (LCLS) [94], a hemoglobin crystal was imaged with two pulse durations, 3 fs and 10 fs with different fluences. Since the intensity that probes the sample varies between each shot, we use a relative intensity between the pulses in our theoretical study to quantify the radiation damage. The electron density of hemoglobin was reconstructed from experimental data using standard crystallographic procedures, where it was concluded that all the different pulses gave the same electron density within the resolution of the experiment.
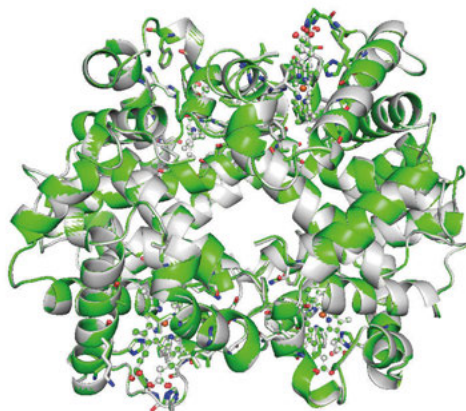
Figure 3.28. The two experimentally determined structures superimposed, retrieved from the 3 fs pulse in grey and 10 fs in green. The two structures show negligible differences and will therefore not provide different interpretations of the function of the protein. Figure extracted from paper V.

As radiation damage is expected to be present in these experiments, our goal is to study how the structure changes due to the pulse, and to quantify if the differences in the structure between the pulses would be possible to observe experimentally. In figure (3.29), the reconstructed atomic model of hemoglobin is shown for the two pulse durations, which show negligible differences. The approach is to use data for the expected damage in CRETIN, in order to estimate the atomic displacement of the structure using the hybrid CR/MD model. The data from the CR simulations was also used in RSPT [69] to calculate the spherically symmetric electron density of each atomic species and statistically significant electronic occupation in the hemoglobin crystal. This was then used to compute the resulting pulse and electronic occupation weighted form-factors.
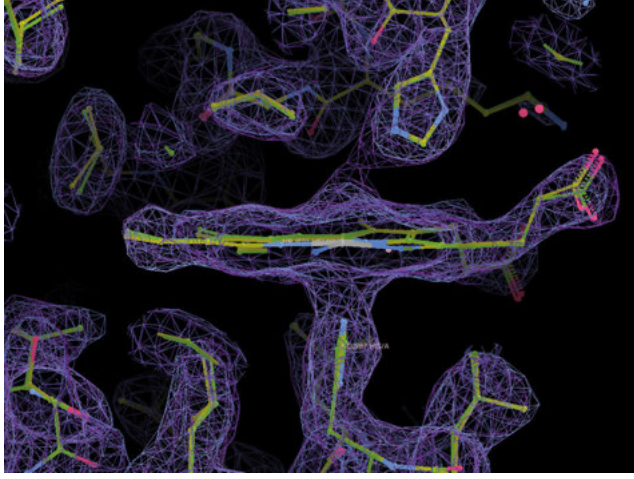
Figure 3.29. $2F_O - F_C$ electron density map (blue 3 fs, violet 10 fs). It provides means of quantifying features that are within the experimentally determined model (F_O) but that is not incorporated in the calculated (F_C) model and vice versa. Figure extracted from paper V.

3.4.1 Radiation damage affected scattering factors

To theoretically model the form-factors that will be observed in the experiment, we first compute the expected photon-matter interaction using CRETIN. In the simulations, the photon energy was set to 7.13 keV. Both pulse durations had the following fluences $F_0 = 1 \times 10^4$ J/cm^2, $4.3F_0$ and $17.4F_0$. In our simulations, we utilized a longer pulse of 15 fs, compared to the 10 fs reported in the experiment. This is because we wanted make sure not to underestimate the atomic displacement in the experiment. For every atomic species, the distribution of each electronic configuration during the pulse duration was obtained. The form-factor weighted by the Gaussian pulse profile and the electronic configurations according to equation (2.69) was obtained. The result for the different pulses and a subset of the atoms is shown in figure (3.30). The figure shows that for longer, more intense pulses, the damage is more severe. It is clear that there is a difference for the form-factor between the pulses, with the largest difference between the 3 fs pulse and the 15 fs pulse with fluence $17.4F_0$. In order to detect this in an experiment, one would need to have the sensitivity to detect the difference in the electron density presented here.

Our results show that there is a difference in the structure that the probe sees, depending on the pulse used. However, if the experiment does not provide sufficient information to detect this, then fitting a model to the experimental data based on undamaged form-factors should result in no detectable differences in the reconstructed electron density.

(a) Iron.

(b) Sulfur.

(c) Oxygen.

(d) Carbon.
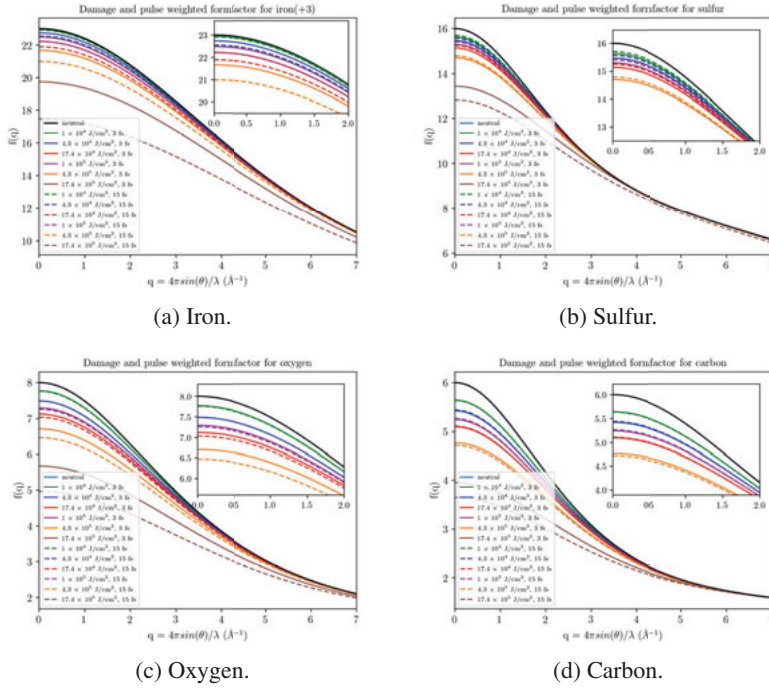
Figure 3.30. Pulse and electronic occupation-weighted form-factors for a subset of the atoms in hemoglobin. The figure shows the effective form-factor that the pulse scatters from and gives an indication if the electronic damage could be differentiated between the pulses. The figure in a) is extracted from paper V.

3.4.2 Atomic displacement from hybrid CR/MD

We perform hybrid CR/MD simulations to asses the degree of motion through the root mean squared displacement (RMSD) of the atoms in the hemoglobin crystal due to the X-ray pulse. The crystal structure used for the simulations is depicted in figure (3.31).
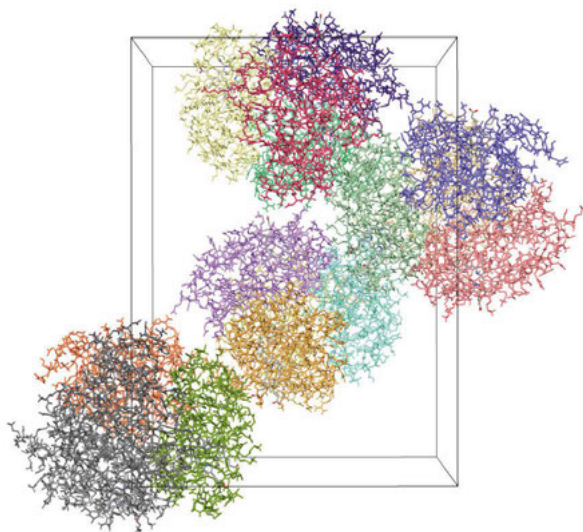
Figure 3.31. A single unit cell of the hemoglobin crystal with $P2_12_12_1$ symmetry, visualized using UCSF ChimeraX [95]. The cell parameters are: $a = 62.51$ Å, $b = 80.97$ Å, $c = 111.48$ Å and $\alpha = \beta = \gamma = 90°$.

In order to perform the MD simulations, the CHARMM36 force field [96] was used to construct the topology files. The required amount of solvation was computed in an iterative manner. First we used the *solvate* command in GRO-MACS to acquire a starting number of solvents. Thereafter *insert-molecules* was used to add 50 water molecules by randomly placing t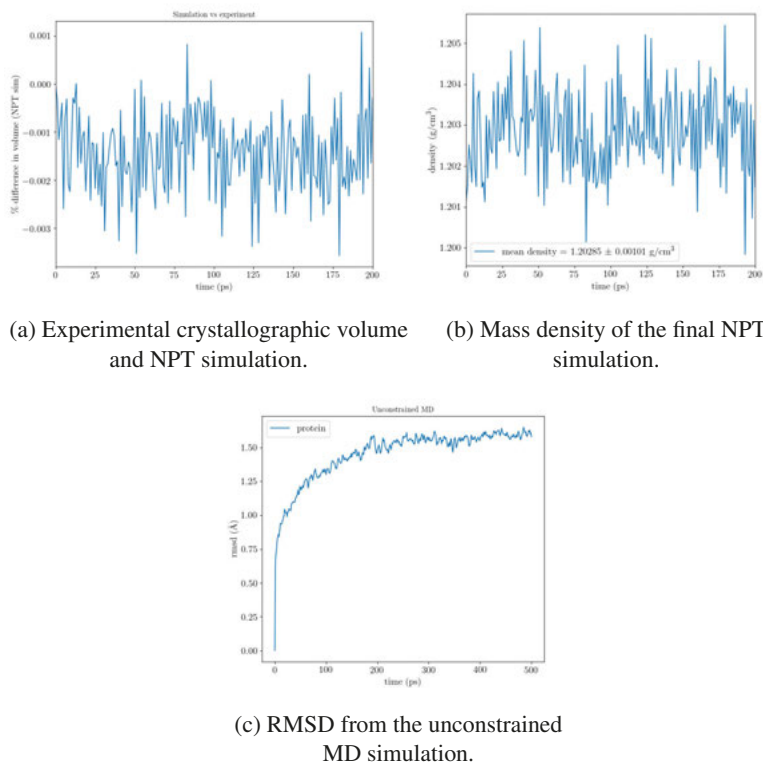hem in the system, given that each atom in the water molecule has a distance to another atom larger than the sum of their van der Waals radii. We minimized the energy of the system using steepest descent. An NVT simulation with a temperature of $T = 300$ K was computed using velocity re-scaling [97] and a subsequent NPT run at atmospheric pressure of $P = 1$ bar was run using Parrinello-Rahman pressure coupling [98]. We monitored the volume during the NPT run and compared the average volume during the simulation to the experimental crystal volume. We repeated the insertion of another set of molecules if the average volume during the NPT run was not within a small tolerance of the experimental crystal volume. The final results of the solvation procedure are shown in figure (3.32), which conlcuded that 8309 water molecules provided a conver-

gence of the volume. This equated to a mass density of 1.203 ± 0.001 g/cm^3, which was used as an input in the CR calculations.

(a) Experimental crystallographic volume and NPT simulation.

(b) Mass density of the final NPT simulation.

(c) RMSD from the unconstrained MD simulation.

Figure 3.32. Results from solvation and equilibriation.

To mimic the entire protein crystal, we simulate the unit cell in figure (3.31) with periodic boundary conditions in all dimensions. The RMSD in figure (3.33) shows that the more intense pulse induces a more rapid increase of the RMSD as expected. Even though the longer 15 fs pulse could reach a relatively large RMSD (≈ 0.8 Å) at the end of the pulse, the scattered data will be an incoherent addition of the different conformations of the protein seen during the pulse, modulated by its shape. Furthermore, as the RMSD increases, the number of bound electrons reduces, leading to less scattering from the sample. This fact in combination with that the number of photons reduces as the RMSD increases means the data recorded on the detector will be biased towards a sample that has an RMSD that is closer to the initial value. Even if the probe would largely scatter from the sample with RMSD ≈ 0.8 Å, the achieved resolution in the experiment was 2.4 Å, which means this displacement is not measurable.

In conclusion, our theoretical results show that we can quantify damage in the structure, with a difference in the form-factors in figure (3.30) and

(a) 3 fs FWHM.
(b) 15 fs FWHM.
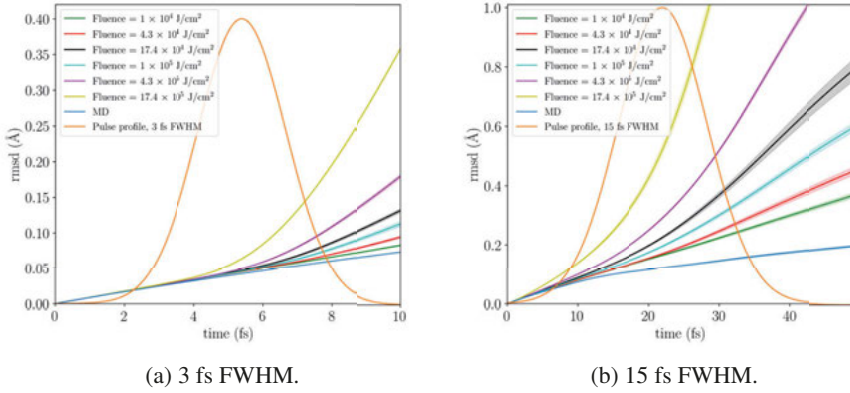
Figure 3.33. Root mean squared deviation of the heme group for increasing intensities, with FWHM 3 fs (left) and 15 fs (right). "MD" is a simulation without damage. The standard deviation comes from averaging multiple trajectories. The Gaussian pulse is centred at 5 and 25 fs in panels (a) and (b). For these simulations, we utilized the Langevin equation, with an electron-ion coupling and electron temperature averaged over the CR simulation. Figures extracted from paper V.

in the atomic displacement in figure (3.33) for the estimated pulses used in the experiment. The experiment concludes that there is no quantitative difference between the reconstructed models. This agrees with the theoretical results presented, since the spatial resolution achieved in the experiment was larger than the RMSD shown here. Furthermore, the biggest difference in the form-factor does not seem to be detectable in the experiment, likely because the intensity is not high enough in the experiment. This means that our hybrid CR/MD model which uses CRETIN for the photon-matter calculations and a plasma-adapted force field in GROMACS, provides conclusions that agree with the experiment. Finally, future experiments that achieve atomic resolution will reconstruct in presence of radiation damage. These experiments will need theoretical studies as presented here in order to quantify the amount of damage, and if this can affect the interpretation of the structure and therefore the function of the biomolecule. For instance, quantifying the displacement due to the probe could be useful information in the reconstruction process. We also note that both experiment and theory show that using 10-15 fs pulses, gives the same structural interpretation as shorter pulses of 3 fs. This motivates the use of longer pulses, which are generally easier to produce.

3.5 Correlated movement of heavy atoms with X-ray pump-probe serial femtosecond crystallography

In this work, femtosecond X-ray pump-probe of an iodine containing crystal (I3C) is investigated. The project contains a comparison of preliminary experimental results[1] and theory. The theoretical simulations were conducted using the MOLDSTRUCT model as described in the method section. The Patterson function defined by equation (2.67) was calculated from the experiment using the obtained scattering data, while the theoretical pattern was determined using trajectories from the hybrid CR/MD code. In this case, structure factors $F_{\mathbf{h}}$ were calculated from a $3 \times 3 \times 3$ supercell (shown in figure (3.34)) of the I3C unit-cell where the form-factors were weighted based on the CR simulations and a Gaussian pulse profile.

(a)   (b)
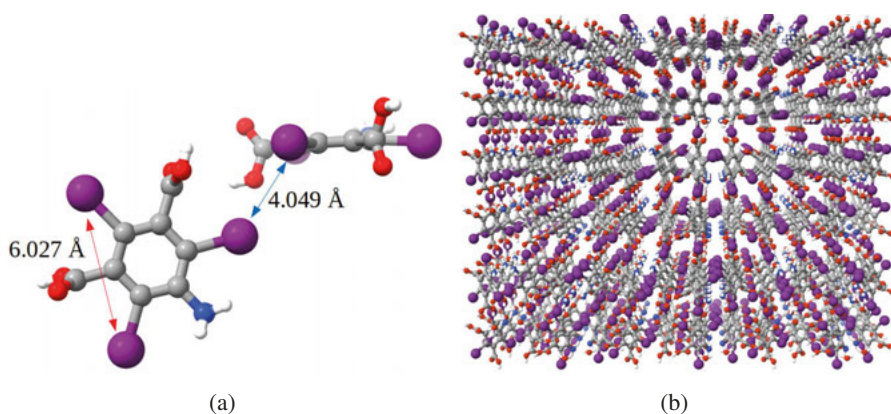
Figure 3.34. a) The I3C molecule and the closest iodine atom from a neighbouring molecule. The blue line corresponds to the shortest intermolecular distance observed in the system, which is shorter than the intramolecular distances seen within the molecule (red lines). b) A $3 \times 3 \times 3$ super-cell of the unit-cell of I3C. The unit-cell satisfies Pbca symmetry, which results in each unit-cell containing 8 copies of the molecule in a) given by the 8 symmetry operations available in this space group.

The experiment is described in figure (3.35). It was conducted by probing an I3C crystal with a two-colour X-ray pump-probe scheme using the XFEL at the LCLS. The probe arrived at the sample with different time delays, which in the experiment were 0, 15, 25 and 55 fs. In order to distinguish the diffraction signal from the pump and probe, a slight difference in the photon energies of the pulses was used. X-ray pump photons were absorbed by a Zinc filter covering half of the detector, since Zinc has an edge just below the pump photon energy. The probe pulse had a photon energy lower than the Zinc edge and the complete diffraction was recorded on the detector.

---

[1]Thanks to Chufeng Li at the Center for Free-Electron Laser Science (CFEL) for conducting the experimental data analysis and calculating Patterson maps from the simulations.
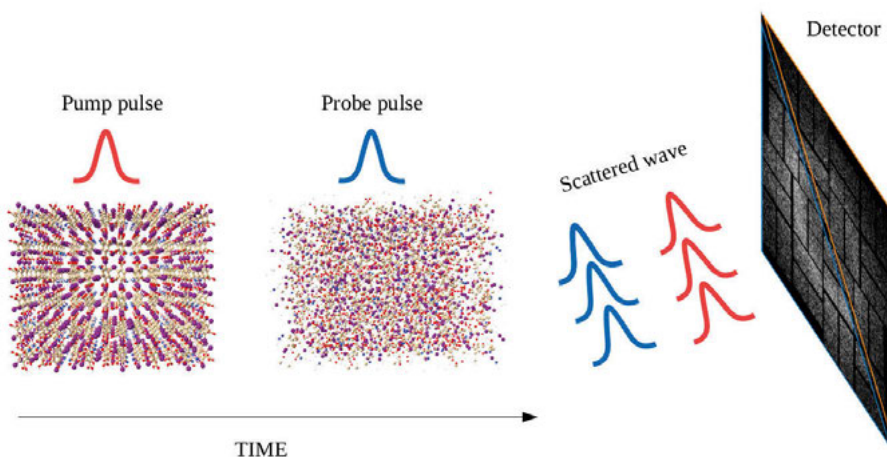
Figure 3.35. The experimental setup for X-ray pump-probe studies. From the left, the native I3C crystal is probed by a pump pulse, which induces structural changes. By systematically increasing the probe delay, one can attain the time-evolution of how this change occurs.
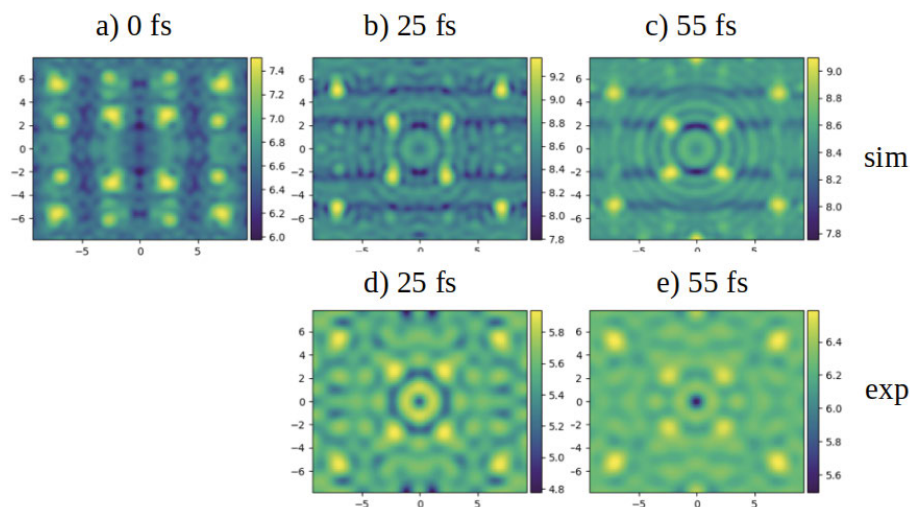
Figure 3.36. The Patterson function as a function of probe delay for the simulation in the top panel, and the experiment in the bottom panel. The probe delays are (a) 0 fs, (b) 25 fs using a fluence of 1.7×10^{18} Wcm^{-2} and (c) 55 fs using a fluence of 8×10^{17} Wcm^{-2}. For the bottom panel (d) is the experimental data for a 25 fs time delay and (e) 55 fs time delay.

The Patterson functions computed in figure (3.36) and (3.37) provide unbiased real space information, since no assumptions are made about the model in the experiment. Other common crystallographic methods tend to for instance as-

68
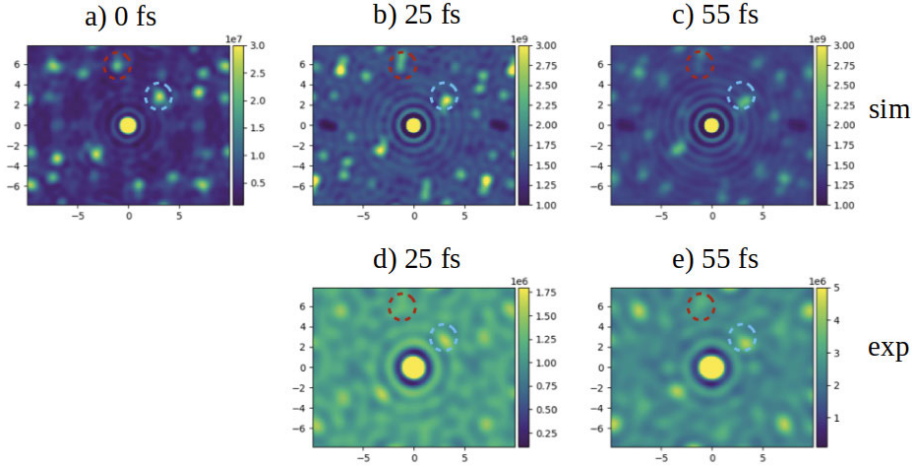
Figure 3.37. The Patterson map as a function of probe delay for the simulation in the top panel, and the experiment in the bottom panel. This data is extracted from the plane which contains the Benzene ring, and the closest I-I intermolecular distance (4.049 Å) shown in figure (3.34a). The probe delay is (a) 0 fs, (b) 25 fs with an intensity of 1.7×10^{18} Wcm^{-2} and (c) 55 fs with an intensity of 8×10^{17} Wcm^{-2}. (d) Experimental data at 25 fs time delay and e) experimental data at 55 fs time delay. The Patterson vector seen in the red dashed circle corresponds to the intramolecular distance of 6.027 Å in figure (3.34a) and the blue circle is the intermolecular distance of 4.049 Å.

sume that the form-factors are the native ones. However, in XFEL experiments the form-factors can strongly deviate depending on the intensity of the laser.

Comparing the results from experiment and simulations in figure (3.36), both show that certain correlations are lost, while others are maintained. Furthermore, the simulations predict, similar to the experiment, that the correlation that do survive have similar correlated motion of atoms. We note in figure (3.37), the Patterson peaks that survive the longest correspond to the ones that have the shortest I-I intermolecular distances, indicated by the blue line in figure (3.34a) and the blue circle in figure (3.37). The intramolecular distance in figure (3.34a) indicated by the red line is larger than the blue line and therefore disappears faster in figure (3.37). It can also be seen that the simulated and experimental data show similar changes for the shortest intermolecular distance (highlighted by the blue circle). The reason for this is that the Coulomb force is dependent on distance as $F \sim \frac{1}{r^2}$, and the shortest distance between the intermolecular iodine atoms of ≈ 4 Å results in twice as large force, compared to the intramolecular distance of ≈ 6 Å. Furthermore, there are multiple other atoms between the iodine atoms in the Benzene ring, which influences the interaction between these iodines. If the total force on the two closest iodine atoms is similar in the different unit cells, their motion will be reproducible

between the unit-cells and therefore noted in the Patterson map. Since the

(a) (b)
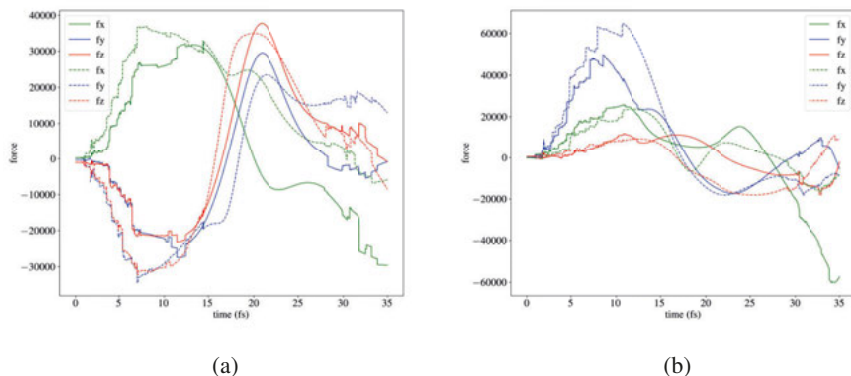
Figure 3.38. a) The force for one iodine (I1) within an I3C molecule (filled lines) with the same atom in a different unit-cell (dashed lines) and b) the force for one iodine (I3) within an I3C molecule (filled lines) with the same atom in a different unit-cell (dashed lines). Similar forces for the same atom in the different unit-cells means that the motion will be reproducible and therefore noted in the Patterson map. The legend shows the corresponding component of the force.

simulations provide similar trends as a function of probe delay, we can use the theoretical model to understand the mechanism in the experimental results. By inspecting the force on each atom we can provide the reason for the correlated motion observed. We extracted the force from a single trajectory for the iodine atoms in the first molecule in the PDB file, and their translated copies in the other unit-cells. Since the supercell contained 27 unit-cells, each iodine atom had 27 copies. Thereafter, we determined the correlation matrix (of size 27×27) and extracted those iodine atoms that correlated above a value of $|0.8|$. From the subset of correlated unit-cells, an example of two iodine atoms (called I1 and I3) in a pair of unit-cells is shown in figure (3.38). We note that the same iodine atom in two pairs of unit-cells can have correlated motion, with several components of the forces having similar time-evolution.

To further understand the molecular dynamics as a function of probe delay, we computed the PADF. The PADF space in figure (3.39) is occupied in a discrete manner at 0 fs probe delay, indicating the crystalline order is largely maintained. As one increases the probe delay, intensity in several regions of the PADF space are lost, while certain key angles and distances such as the I1-I3 intermolecular distance of ≈ 4 Å are increasingly occupied. This indicates the transition to a new structural order, which was also seen for our simulations on hexagonal ice exposed to an XFEL in Paper IV.

In conclusion, model independent structural information (Patterson functions) can be obtained from X-ray pump-probe experiments to study radiation damage dynamics. Both experiment and theory show consistent behavior,
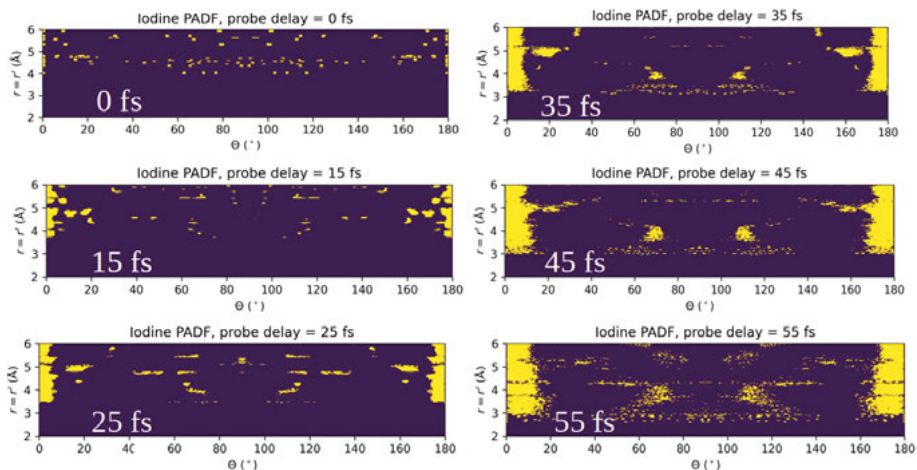
Figure 3.39. The PADF of the I3C super-cell, as exposed to an XFEL pulse with varying probe delay. The intensity used here is 10^{18} Wcm^{-2}, which is in between the intensities (8×10^{17} Wcm^{-2} and 1.7×10^{18} Wcm^{-2}) used in the Patterson results in figures (3.36) and (3.37).

they show correlated motion between pairs of iodine atoms, correlation between pairs degrades at different rates and the theoretical model can predict the time-scale of radiation damage with the experimental parameters.

The results presented provide a more complete picture of radiation damage. Previously, it has been thought that radiation damage only leads to destructive effects in the structure. For instance, in a crystalline sample, the translational order is rapidly destroyed due to the X-ray pulse. We show here that the Bragg diffraction signals do not disappear with the highest resolution first, and then the lower resolution. Instead, the reflections originating from the ordered structure remain, with the intensities changing even for the largest probe delay of 55 fs. The result of Bragg peaks being visible is that the periodicity is still present. The changes in the intensities are due to the structural changes induced by the pulse, which alters the interaction between the atoms and therefore induces movement. Given that the forces on the same atoms in multiple unit cells can be reproducible, this makes their interference visible in diffraction space.

Finally, the experimental data is limited to 1.4 Å resolution which only allows the study of correlated motion between the iodine atoms. Future data collection with higher resolution will allow us to observe the lighter atoms and reconstruct the charge states of the atoms.

3.6 Macromolecule classification using X-ray laser induced fragmentation (Paper VI)

This project explores the fragmentation dynamics of biological macromolecules using an X-ray laser to asses if structural information can be attained. Using the described MC/MD model to simulate photon-matter interaction, we studied the dynamics of fragmentation in single biological macromolecules exposed to an X-ray laser, to understand if one can obtain information regarding the structure.

3.6.1 Benchmark of MC/MD model to *ab-initio* simulations

The MC/MD method was first benchmarked to previous DFT simulations of ionized peptides. These showed that the charge transfer of an electron from the hydrogen atoms to the heavier atoms played a great role in the stability of the system. Thus, we aimed to test the charge transfer model used to see if it could qualitative reproduce the DFT simulations. Since a net-charge was put at the start of the DFT simulations, we aimed to reproduce this in the MC/MD model by exposing the system to a very short pulse. In the simulations, we used a Gaussian pulse with full width half maximum (FWHM) of 3 fs. Several different photon energies and intensities were used, which provided variability in the way we distributed charge on the molecule. By inspecting the mean charge of the system, we could make it converge within the first few femtoseconds, as seen in figure (3.40).
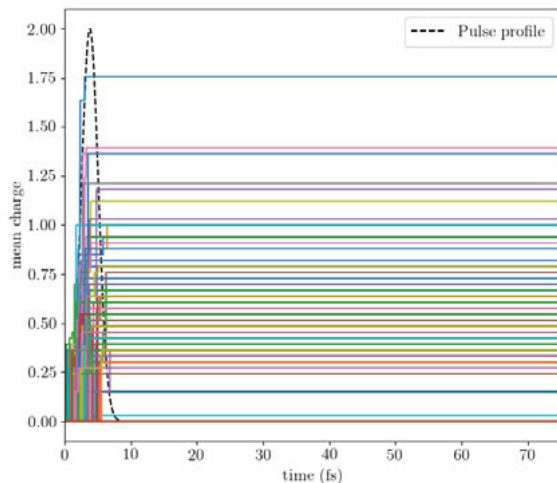

Figure 3.40. Mean charge as a function of time for pulses with different fluences and a FWHM of 3 fs. Note that only simulations that had charge ≤ 1 were considered in the bond-integrity plots below. It can be noted that the molecule attains the converged charge state in the first few femtoseconds of the simulation.

To compare how the structure evolves due the probe when including charge transfer or not, we provide a visualization of the trajectories in figure (3.41). Including charge transfer in the top panel provides a more stable system, where most bonds between non-hydrogen atoms are intact, compared to the bottom panel without charge transfer. Several hydrogen atoms leave the molecule as a proton resulting in the neutralization of the system. Most notably, we note that all oxygen atoms leave the molecule if we do not have charge transfer. For each simulation, we computed the bond integrity as given by equation
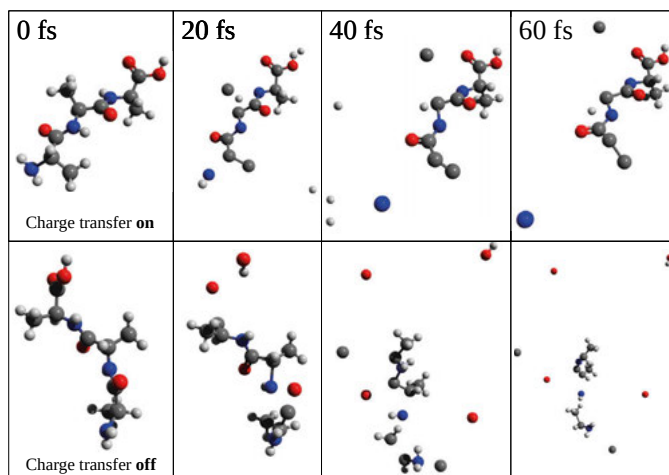
Figure 3.41. Trajectories of trialanine when exposed to an XFEL pulse of 3 FWHM, resulting in a mean charge of ≈ 0.6. Top panel corresponds to a simulation with charge transfer, and bottom is without. Figure extracted from paper VI.

(2.50) as a function of time and the mean charge during the simulation. We present the result in figure (3.42), where (3.42a) includes charge transfer and (3.42b) does not. Comparing to the bond integrity for trialanine given by the DFT simulations in figure (3.9) we note by omitting charge transfer in the MC/MD simulations, we get completely different dynamics involving bonds to hydrogen. In this case, the hydrogen atoms will remain neutral as they have low photoionization cross section, and will therefore most likely stay bonded. On the other hand, if charge transfer is included, we get much better agreement with the DFT simulations where the bonds to hydrogen atoms are broken as one increases the mean charge on the system.
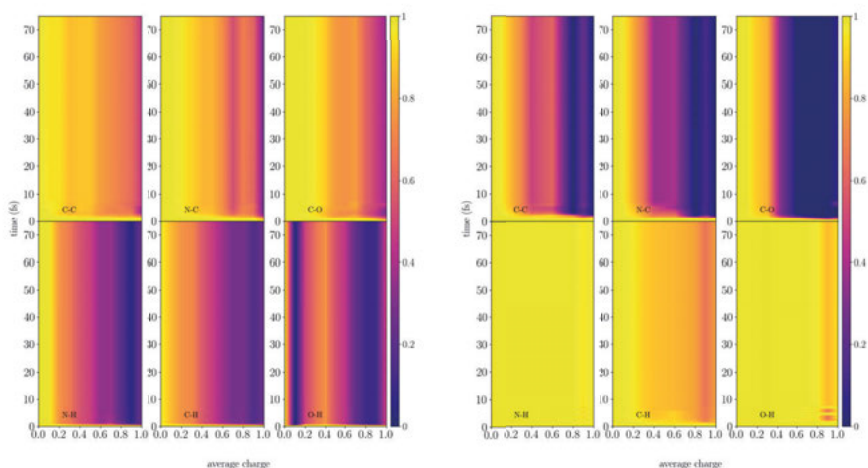
(a) With charge transfer.   (b) Without charge transfer.

Figure 3.42. Bond-integrity plot from the MC/MD simulations. Each plot is averaged over multiple bonds and simulations. Top figure is including charge transfer from the hydrogen atoms to the heavier atoms. The bottom figure does not include charge transfer. This figure can be directly compared to the result given by the DFT simulations in figure (3.9). Figure extracted from paper VI.

3.6.2 Classifying proteins from Coulomb explosions

The trajectories of the ions determined from the MC/MD calculation were used to compute an ion heat map, a 2D representation on a sphere of where the ions end up after a long time. Previous studies have shown that information regarding the molecule's orientation at time of exposure can be obtained from the fragmentation pattern [99]. The trajectories of the ions are connected to their initial location in the native structure, which one could theorize it provides distinct fragmentation patterns. This idea was studied by comparing the ion heat maps for three cases: i) a monomer and dimer of the same protein, ii) two conformations of the same protein, an asymmetric and symmetric fold and iii) a stretched conformation of a protein and its compact form. The heat maps were analyzed using two unsupervised dimensionality reduction techniques called principal component analysis (PCA) and t-distributed stochastic network embedding (t-SNE). These algorithms enable conversion of high-dimensional data into a lower-dimensional space while retaining maximum information. Reduction of data in PCA is done by projecting it onto the two principal components of the covariance matrix. On the other hand, t-SNE assigns data near in a high-dimensional space to also be near in a lower-dimensional space. Initially, the data is projected into a 2D space and then adjusted within

the lower-dimensional space to minimize the Kullback-Leibler divergence employing gradient descent.

Both algorithms were used to reduce the dimension of the ion heat map by projecting it to a two-dimensional space (figure (3.43)). Each ion map provides a single point in this two-dimensional space. After many simulations with different starting configurations and damage scenarios for each protein, multiple distinct clusters of points can be seen in the reduced space. In order to determine the groups that are available among all points, k-means clustering is used. Comparing to the true labels of each point, we note that a dimensionality reduction technique with clustering algorithms allow us to classify macromolecules that have been exposed to the X-ray laser.

(a) Classification using PCA   (b) Classification using t-SNE.
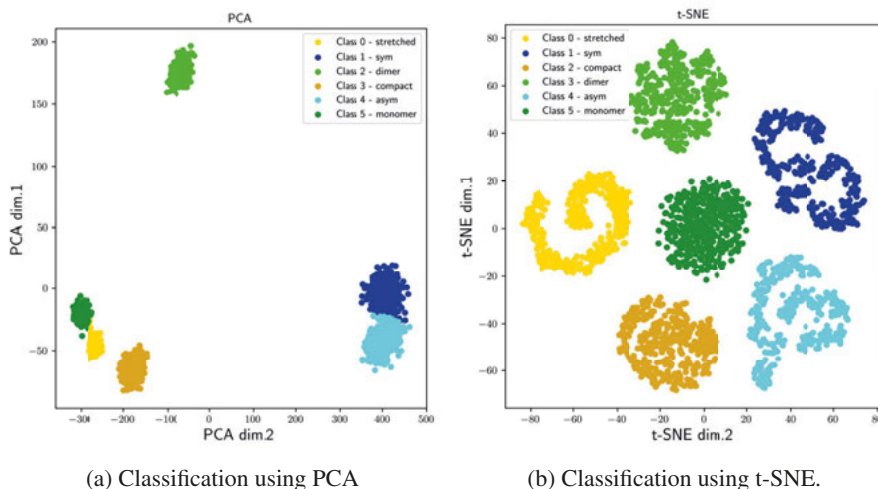
Figure 3.43. Classification of the heat maps from X-ray induced fragmentation of several proteins. The original protein orientation is kept fixed and considered known. When the orientation is randomized, only the monomer/dimer can be classified without additional information. Figure extracted from paper VI.

3.7 Structural heterogeneity in single particle imaging using X-ray lasers (Paper VII)

Radiation damage as explored in the summarized papers above, is not the only aspect of imaging with XFELs that can limit the obtainable resolution. As SPI is dependent on averaging similar diffraction patterns, it is important to understand how structural heterogeneity of biomolecules can be reduced. This work was therefore focused on examining the effects of water on the structural heterogeneity in the proteins ubiquitin and lysozyme, similar to previous work on viruses [83]. The goal was to understand if an optimal water layer could be embedded in a protein, in order to make their signal in diffraction space more similar.

Three different water layers were examined with different thickness, 0 Å, 3 Å and 6 Å, which were initially simulated with four different temperatures 200 K, 250 K, 300 K and 350 K. The production MD simulations were performed in vacuum without temperature coupling, using GROMACS version 4 [100]. For the simulations with solution, the water will cluster itself at specific areas in the protein. An example of a solvated lysozyme is shown in figure (3.44). This corresponds to a single instance of all possible spatial configurations that the water layer can occupy. From the MD simulations, we com-
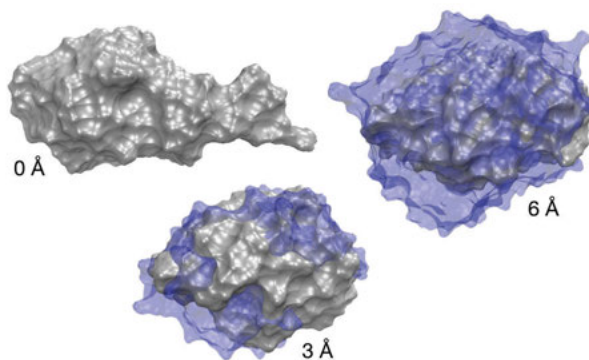
Figure 3.44. Lysozyme (grey) with snapshots of possible configurations of the different water layers (blue). Figure extracted from paper VII.

puted diffraction patterns. Using equation (2.71) for the Fourier ring correlation (FRC), we correlate diffraction patterns from the same set of temperature and water layer simulations. We used 550 different patterns for each set, which were divided in half and then correlated. This was done 100 times, in order to asses the dependence of the result on the way one divides the set. The result is presented in figure (3.45).

Figure 3.45. The mean and standard deviation of the FRC as a function of temperature and water layer. Figure extracted from paper VII.

The cut-off of $FRC(q) = 0.5$ was used to determine the highest obtainable resolution given the data. We used the form of equation (3.1) with the two parameters b and c to fit the FRC data in figure (3.45),

$$FRC(q) = \frac{1}{1 + \exp(bq + c)}. \tag{3.1}$$

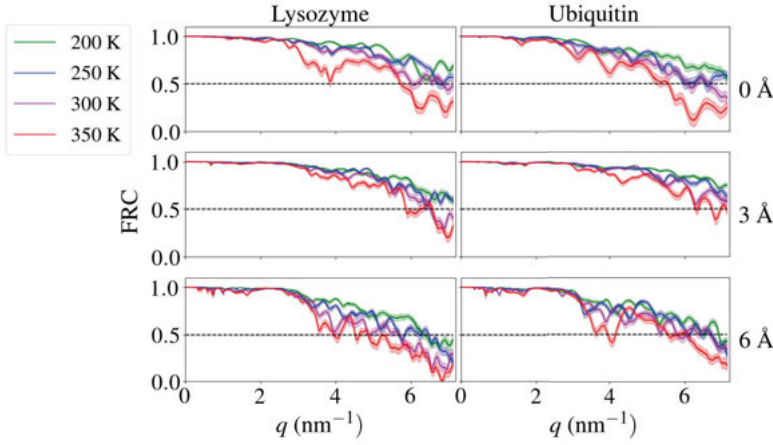The result of the fits are shown in figure (3.46).
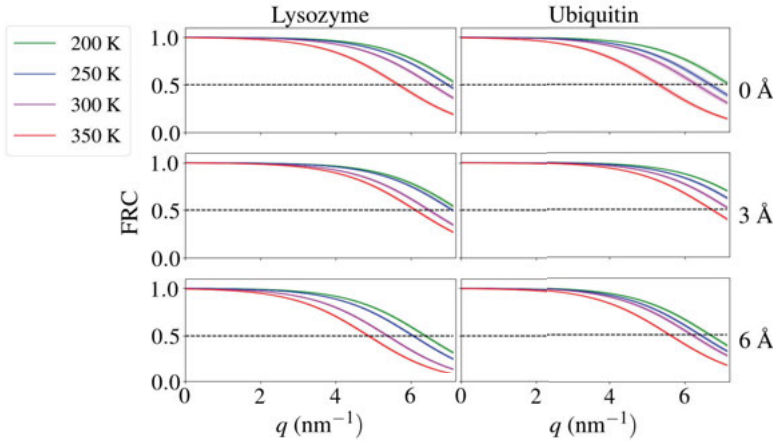
Figure 3.46. The mean and standard deviation for the fit of the FRC as a function of temperature and water layer. Figure extracted from paper VII.

Thereafter, we extracted the resolution q where $FRC(q) = 0.5$, and then used this to calculate the resolution limits in figure (3.47).
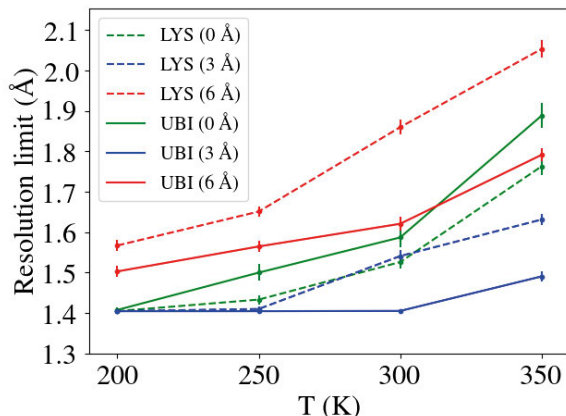
Figure 3.47. The calculated resolution limit as a function of temperature. Mean values of the 100 repeated correlation calculations are shown, with the error bars representing the standard deviation. Figure extracted from paper VII.

The figure shows that for both the proteins, a water layer of around 3 Å is the optimal choice. We note that at 200 K, a water layer of 6 Å will only provide additional noise in the diffraction pattern. At this temperature, there is not enough heterogeneity that the water layer will suppress. It can be seen that there is no difference between the 3 Å water layer compared to none. As the temperature is increased, and more structural variation is introduced, the 3 Å case performs better than 0 Å.

In these simulations, we have only studied the effects of coherent scattering, where the non-structural signal comes from water-water and water-protein interference. However, the process of incoherent scattering, which does not provide structural information, is also part of a realistic system [101]. By incorporating incoherent scattering, the extra water molecules would induce even more noise in the diffraction patterns, leading to a reduction in the correlation between the patterns. Furthermore, other important sources of noise in an XFEL experiment have been omitted, such as background and radiation damage [102, 103].

In conclusion, the study shows there is an optimal amount of water that can reduce the structural heterogeneity of the protein, and thus provide higher resolutions, compared to a naked molecule. However, there is a limit to how much solvent should be used, since these additional molecules will induce further noise in the diffraction pattern.

4. Conclusion and outlook

The work presented in this thesis investigates the ultrafast dynamics of biological molecules of different size scales (amino acids, peptides, protein crystals, liquids and molecular clusters) due to different levels of ionization and conformations observed in XFEL experiments. The simulation codes that were used are dependent on the system size, and we have therefore explored different domains of physics, such as plasma physics, classical molecular dynamics and density functional theory. For the smaller biomolecules, an established first-principle code was used. We showed that the dynamics of the hydrogen atoms are important for the stability of the system. We concluded that charge transfer from the hydrogen atoms to the heavier atoms in the system was important to include. This fact was also taken into account when we modelled photon-matter interaction with a newly developed Monte Carlo/classical MD code, by incorporating charge transfer in a phenomenological way.

In order to simulate photon-matter interaction in large macromolecules and crystals we developed a hybrid collisional-radiative/molecular dynamics model, which is publicly available for all to use. It was used in several projects, both in theoretical work and for supporting XFEL experiments. We used the model to conduct a theoretical study to explore photon-matter interaction in hexagonal ice. The results show that due to the XFEL, the system transitions into a crystalline plasma, where some crystalline order is maintained while still being in a plasma phase. Additionally, we note the phase transition from the solid to liquid occurs in an anisotropic manner, where different parts of the system change on different time-scales.

Using the same model, we aimed to understand an X-ray pump-probe experiment on I3C crystals, and could reveal correlated motion between the heavy iodine atoms. This motion was maintained for the longest time-delay of 55 fs, which indicates that the motion of the atoms is not random. Furthermore, we conducted simulations for an experiment involving SFX of hemoglobin crystals using two sets of pulse parameters. Our findings indicate that, considering the given parameters and achievable resolution, it was not possible to detect the incurred damage. In both projects, preliminary experimental results show that the simulations provide good agreement with the data.

To study small systems such as amino acids, peptides and smaller proteins, a hybrid Monte Carlo/MD model was developed. It was used to simulate photon-matter interaction in different biological macromolecules, where we studied the fragmentation to extract structural information. It was possible to connect the dynamics of the fragmentation to structural information through the use of classification algorithms.

In summary, we have developed different theoretical methods for studying photon-induced dynamics due to an X-ray laser in matter at different length scales. The model was compared to two different experiments and showed good qualitative agreement.

Several future research questions that involve the developed methods in this thesis can be explored. By integrating the models into the photon-mater module of the SIMEX pipeline, an entire imaging experiment could be simulated at an XFEL. Furthermore, we have shown that our model is applicable for studying structure and dynamics in SFX experiments. It will therefore be a useful tool for experiments where ultrafast structural dynamics is followed and scattering factors rapidly change during the pulse. The models could also be used during an experiment, as a part of the online analysis. By integrating them with a diffraction code such as CONDOR [77], analysis of the effects of the used pulse parameters in an experiment on the acquired diffraction signal could be estimated. The effect of radiation damage during an experiment can be minimized or accounted for in the structural determination based on the developed methods here.

Finally, the number of photons delivered by an XFEL might not change significantly, but new mirrors will allow for even higher intensities on the sample [104]. This would allow the exploration of matter exposed to ultrahigh intensities, which could show exciting new dynamics to be explored. In combination with the introduction of sub-femtosecond pulses all the way to the attosecond regime, one could potentially directly image the electron's motion with diffraction.

5. Author contributions

The work discussed here is a result from collaborating with other scientists. In all papers, I was involved in discussions and writing. More specifically, my contributions to the papers are the following:

Paper I: Conducted the time-dependent-DFT benchmark simulations shown in the supplementary information. Contributed to writing at the final stage.

Paper II: Conducted all simulations, DFT and coherent diffractive imaging, and did most of the writing with support from the co-authors. I am the corresponding author.

Paper III: Developed and implemented the MOLDSTRUCT code in GROMACS required to run the simulations. Conducted all simulations, and did most of the writing with support from the co-authors. I am the corresponding author.

Paper IV: Developed and implemented the MOLDSTRUCT code in GROMACS required to run the simulations. Conducted all final MD and powder diffraction simulations, and analyzed the MD using the PADF. Was involved in the majority of the writing with support from the co-authors. I am the corresponding author.

Paper V: Developed and conducted the hybrid CR/MD and the calculation of the damaged form-factors with data taken from CR simulations performed by co-authors. Wrote the photon-matter interaction part of the paper in collaboration with co-authors.

Paper VI: Developed and implemented the first iteration of the MC code in GROMACS and supervised the use of the code in the project. Conducted the MC/MD simulations for trialanine and benchmarked it to previous DFT simulations. Contributed to the writing and was involved in the discussions.

Paper VII: Conducted the final diffraction simulations and their analysis. Analyzed the final MD simulations and contributed to writing of the paper.

6. Sammanfattning på svenska

Ljus används av människor för att observera i naturen och beroende på våglängden kan man fotografera allt från makroskopiska objekt till enskilda atomer. För att avbilda biologiska system som proteiner, virus och celler behövs mycket korta våglängder. Dessa biologiska system består av mindre beståndsdelar, amino-syror, som genom peptidbindningar bygger upp de större systemen och ger proteinet specifika funktioner. Genom att bestämma strukturen på biomolekylen kan man därför förstå sig på dess funktion. 1942 blev Rosalind Franklin den första att avbilda DNA med en röntgenkälla och sedan dess har biologisk avbildning med en röntgenkälla utvecklats enormt. Under en lång tid bestämde man strukturen på biologiska partiklar genom att kristallisera systemet och sedan avbildade man det med en röntgenkälla, som en synkrotron. Problemet nu är att inte alla biopartiklar är möjliga att kristallisera. År 2000 presenterades en teoretisk studie där en ny metod för att avbilda biomoleykler som inte kan kristallieras föreslogs. Genom att använda en högintensiv och ultrasnabb röntgenlaser, många gånger starkare än konventionella synktronkällorna, så skulle man kunna avbilda enskilda biomolekyler utan kristallisering. Nackdelen är att molekylerna exponeras med så mycket energi att strukturen ändras under tiden lasern avbildar systemet. Detta medför att det slutgiltiga rekonstruerade systemet kanske inte är den riktiga strukturen. Destruktiva förändringar av material på grund av ljus benämns som strålskada. Hur mycket strukturen ändras beror på laserns egenskaper, vilket medför att det är viktigt att studera de laserparameterar som minimerar strålskada. Kan man förstå sig på strålskada i detalj, skulle man inte bara minimera det, men också möjligtvis använda det för att extrahera information om strukturen.

Användning av teoretiska studier ger möjlighet att studera interaktionen mellan ljus och material utan att behöva göra dyra experiment, men också för att förstå dem experimentella resultaten. I denna avhandling har teoretiska och numeriska studier gjorts av strålskada, med motivationen att förbättra avbildning av biologiska molekyler med en röntgenkälla. Studien påbörjades med att studera strålskada i aminosyror och peptider, för att förstå hur biomolekylers byggstenar påverkas av strålskada. Givet de små molekylerna möjliggörs användandet av kvantmekaniska beräkningar av strukturförändringen. Strukturförändringen givet en viss joinseringsnivå studerades i detalj genom att undersöka tidsskalan för när de kemiska bindningarna bröts. Effekten av strålskadan studerades även i den reciproka rymden, för att kunna jämföra med experiment. Ett viktigt resultat var att väte-atomernas dynamik spelade stor roll för systems stabilitet. Det visade sig att väteatomer ofta blev joniserade

på grund av elektronöverföring till en tyngre atom. När väteatomerna lämnade molekylen medförde detta att den kvarvarande molekylen neutraliserades och blev mer stabil.

För att studera interaktionen mellan en röntgenlaser och större system som proteiner, vätskor och kristaller utvecklades en avancerad modell. Den bygger på en kombination av plasmafysik och klassisk molekyldynamik. Två versioner av modellen har utvecklats, där den första modellerar stora system som blir till plasma och där majoriteten av de fria elektroner stannar kvar i systemet. Denna version användes för att studera interaktionen mellan en röntgenlaser och kristallen is. Signalen i den reciproka rymden studerades och visade att den kristallina strukturen kunde bibehållas under en stor del av pulslängden, samtidigt som kristallen var i en plasmafas. Utöver detta observerades en icke-isotropisk fasövergång från den kristallina strukturen till den flytande fasen, där olika kristallografiska plan försvann på olika tidsskalor. Detta visades med hjälp av en parvinkelfördelningsfunktion, vilket är en korrelationsfunktion som beror på både avståndet mellan atomerna och vinkeln mellan dem. Slutsatsen blev att den lokala strukturen har stor påverkan på stabiliteten.

Modellen användes också för att stödja två experiment gjorda vid en röntgenlaser. I det första experimentet studerades strålskada i en hemoglobinkristall, som strukturbestämdes med hjälp av en frielektronlaser. För att bestämma hemoglobins struktur användes två olika pulslängder, en kort puls på 3 femtosekunder och en längre på 10 femtosekunder. Resultatet var att givet upplösningen på strukturen, så gick det inte att notera skillnader på strukturerna som bestämdes med dem olika pulsparametrarna. Med hjälp av modellen som utvecklats, simulerades strålskadan givet de olika pulslängderna och intensiteter. Slutsatsen var att givet parametrarna som användes i experimentet, kommer man inte kunna se skillnad på den korta och långa pulsen, då strålskadan i den längre pulsen inte är större än upplösningen i experimentet.

I det andra experimentet studerades strålskadeutvecklingen i kristaller innehållande jodatomer. Experimentet gjordes genom att en första puls med röntgenfotoner inducerar strukturförändringar. Därefter användes en andra puls med varierande tidsfördröjning, som endast skiljer sig minimalt med avseende på fotonenergin, för att observera strukturutvecklingen. Genom att analysera diffraktionsmönstret kan man extrahera information gällande relativa positioner för de tunga jodatomerna, eftersom dem medför starkast spridning. Detta genomfördes med hjälp av Pattersonfunktionen, som extraherades från både experimentet, och utifrån simuleringar gjorda med den beskriva modellen. Simuleringarna och experimentet påvisade liknande beteende gällande jodatomernas rörelse. Resultaten visade att det fanns en korrelerad rörelse mellan specifika par av jod-atomer, där olika korrelationer försvinner på olika tidsskalor. Tidsskalan för hur snabbt dessa korrelationer försvann som teorin förutsåg stämde bra överens med experimentet.

Slutligen studerades hur man skulle kunna använda strålskada för att få information gällande strukturen hos en molekyl. Tanken är att exponera, till exempel ett enskilt protein, med en röntgenlaser och studera fragmenterings-förloppet. Olika proteiner skulle medföra unika sätt som de fragmenterar och därmed kunna kopplas till strukturen. Med hjälp av modellen som utveck-lades, simulerades ultrasnabba interaktionen mellan olika biologiska makro-molekyler och en röntgenlaser. Fragmenteringen studerades och genom an-vändandet av klassifiseringsalgoritmer kunde det visas att man kan urskilja biologiska molekyler.

Baserat på metoderna som utvecklats i detta arbete, skulle man kunna un-dersöka användandet av attosekundpulser för molekylär avbildning. Dessa pulser motsvarar tidsskalan för elektronens bana runt atomkärnan, och skulle möjlig-göra att man kan studera detta genom till exempel diffraktion.

7. Acknowledgement

GROMACS reminds you: "Thanks to all peeps!" (Ibbe)

4.5 years of work presented in thesis would not see the light of day without a lot of nice and smart people. It would also not been possible without people funding science. I am thankful to the **European XFEL** who funded my entire PhD thesis, and **Uppsala University** for providing the graduate education, research environment and computational resources such as the **DaVinci** computer cluster.

I would like to say thanks to my supervisors, **Adrian** who started this shared PhD project, together with **Nic** and **Calle**[1]. Thanks for the wonderful opportunity to do science and for helping me create something that I can be proud of. Without your guidance throughout these years both related to work and life outside of it, this would have been impossible. I've had a lot of fun with all of you, whether it has been discussing potentials and crashing simulations, shooting lasers in the computer or in real life (not just experiments!), hangout at Samarkand[2], drinking or playing MTG. I wish you the very best in the future! Thanks to **Oscar** as well, for all your help throughout the PhD with everything related to DFT.

For my fellow MOLDSTRUCT students in Uppsala: **Sebastian**, **Tomas**, **Pamela**, **Harald** and not in Uppsala: **Mandl**. Thanks for great company with many fun conversations over coffee. I am looking forward to reading your theses, and wish you the best of luck with it. Keep up the fun in the office!! Special thanks to **Sebastian** and **Tomas** for your help with finalizing all the papers in time for my thesis. And don't forget: never stop the print statements, and please keep the MC/MD code going! A big thanks goes to **Emiliano** aka the master of GROMACS, for working hard all these years on many projects and for having fun at the same time. Thanks for helping with the manuscripts to get them ready for my thesis. I wish you the best of luck with your future research. I am sorry you had to suffer all those print statements..

Thanks to **Olof** for still being interested in my work and randomly calling me at times to check in. It is great to see that you got to do a lot of science even outside of academia. I promise you we will finally publish I3C! Thanks to **Ludde** for the daily snaps and keeping in touch.

The **biophysics network** is thanked for fruitful discussions and Good luck to all the other PhD students in the network with your theses. Good luck to

---

[1]No thanks to Baal?

[2]Sebastian's favorite place!

all PhD students at MolKond and keep up the memes. Thanks to **Kajwal** for great work on ice and PAF.

I spent half my PhD life in Hamburg where I had the opportunity to meet a lot of nice people. Thanks to all members of **SPB**. Special thanks to **Jun** for your help throughout the years and all the MTG/DnD people for some fun moments. Wishing all of the members of **SPB** the best of luck in the future. I also had the opportunity to interact with people at CFEL. Special thanks to **Chufeng** for many nice dinners and conversations. Thanks for all your incredibly hard work on I3C, and I am confident it will turn out to be a really awesome paper.

I would like to say thanks to all of my **friends** outside of work, for a lot of great times these years that made me relax from all of my crashing simulations. Thanks to all of my **family** and **relatives** for great support, and thanks to **Jakob** for making such a nice cover! Finally, thanks for everything **Shayma**, for your support, help, patience and love. You are the best that happened these years! Love you.

References

[1] Bernhard Rupp. *Biomolecular crystallography: principles, practice, and application to structural biology*. Garland Science, New York, 2010.

[2] Richard P. Feynman, Robert B. Leighton, and Matthew L. Sand. *The Feynman lectures on physics*. 1963.

[3] Bernhard Rupp. *Biomolecular Crystallography*. 2009.

[4] Richard Henderson. The potential and limitations of neutrons, electrons and x-rays for atomic resolution microscopy of unstained biological molecules. *Quarterly reviews of biophysics*, 28:171–93, 06 1995.

[5] K. Wuthrich. Protein structure determination in solution by nuclear magnetic resonance spectroscopy. *Science (American Association for the Advancement of Science)*, 243(4887):45–50, 1989.

[6] Yifan Cheng, Nikolaus Grigorieff, Pawel A. Penczek, and Thomas Walz. A primer to single-particle cryo-electron microscopy. *Cell*, 161(3):438–449, 2015.

[7] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The Protein Data Bank. *Nucleic Acids Research*, 28(1):235–242, 2000.

[8] Mark J. Howard. Protein nmr spectroscopy. *Current biology*, 8(10):R331–R333, 1998.

[9] Dominique P. Frueh, Andrew C. Goodrich, Subrata H. Mishra, and Scott R. Nichols. Nmr methods for structural studies of large monomeric and multimeric proteins. *Current Opinion in Structural Biology*, 23(5):734–739, 2013.

[10] Ka Man Yip, Niels Fischer, Elham Paknia, Ashwin Chari, and Holger Stark. Atomic-resolution protein structure determination by cryo-em. *Nature*, 587(7832):157–161, 2020.

[11] Joachim Frank. Time-resolved cryo-electron microscopy: Recent progress. *Journal of structural biology*, 200(3):303–306, 2017.

[12] Petra Fromme, Sebastien Boutet, and Mark S. Hunter. *X-Ray Free Electron Lasers: A Revolution in Structural Biology*. Springer, New York, 2019.

[13] Henry N. Chapman, Carl Caleman, and Nicusor Timneanu. Diffraction before destruction. *Philosophical transactions. Biological sciences*, 369(1647), 2014.

[14] Adrian P Mancuso, Andrew Aquila, Lewis Batchelor, Richard J Bean, Johan Bielecki, Gannon Borchers, Katerina Doerner, Klaus Giewekemeyer, Rita Graceffa, Oliver D Kelsey, et al. The single particles, clusters and biomolecules and serial femtosecond crystallography instrument of the european xfel: Initial installation. *Journal of synchrotron radiation*, 26(3):660–676, 2019.

[15] Henry Chapman, Petra Fromme, Anton Barty, Thomas White, Richard Kirian, Andrew Aquila, Mark Hunter, Joachim Schulz, Daniel DePonte, Uwe

Weierstall, R Doak, Filipe Maia, Andrew Martin, Ilme Schlichting, Lukas Lomb, Nicola Coppola, Robert Shoeman, Sascha Epp, Robert Hartmann, and John Spence. Femtosecond x-ray protein nanocrystallography. *Nature*, 470:73–7, 02 2011.

[16] Richard Neutze, Remco Wouts, David van der Spoel, Edgar Weckert, and Janos Hajdu. Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature*, 406(6797):752–757, 2000.

[17] Tomas Ekeberg, Dameli Assalauova, Johan Bielecki, Rebecca Boll, Benedikt J Daurer, Lutz A Eichacker, Linda E Franken, Davide E Galli, Luca Gelisio, Lars Gumprecht, et al. Observation of a single protein by ultrafast x-ray diffraction. *Light: Science & Applications*, 13(1):15, 2024.

[18] Erik G. Marklund, Daniel S. D. Larsson, David van der Spoel, Alexandra Patriksson, and Carl Caleman. Structural stability of electrosprayed proteins: Temperature and hydration effects. *Physical chemistry chemical physics*, 11(36):8069, 2009.

[19] A. Aquila, M. S. Hunter, R. B. Doak, R. A. Kirian, P. Fromme, T. A. White, J. Andreasson, D. Arnlund, S. Bajt, T. R. M. Barends, M. Barthelmess, M. J. Bogan, C. Bostedt, H. Bottin, J. D. Bozek, C. Caleman, N. Coppola, J. Davidsson, D. P. DePonte, V. Elser, S. W. Epp, B. Erk, H. Fleckenstein, L. Foucar, M. Frank, R. Fromme, H. Graafsma, I. Grotjohann, L. Gumprecht, J. Hajdu, C. Y. Hampton, A. Hartmann, R. Hartmann, S. Hau-Riege, G. Hauser, H. Hirsemann, P. Holl, J. M. Holton, A. Homke, L. Johansson, N. Kimmel, S. Kassemeyer, F. Krasniqi, K.-U. Kuhnel, M. Liang, L. Lomb, E. Malmerberg, S. Marchesini, A. V. Martin, F. R. N. C. Maia, M. Messerschmidt, K. Nass, C. Reich, R. Neutze, D. Rolles, B. Rudek, A. Rudenko, I. Schlichting, C. Schmidt, K. E. Schmidt, J. Schulz, M. M. Seibert, R. L. Shoeman, R. Sierra, H. Soltau, D. Starodub, F. Stellato, S. Stern, L. Struder, N. Tîmneanu, J. Ullrich, X. Wang, G. J. Williams, G. Weidenspointner, U. Weierstall, C. Wunderer, A. Barty, J. C. H. Spence, and H. N. Chapman. Time-resolved protein nanocrystallography using an x-ray free-electron laser. *Optics express*, 20(2):2706–2715, 2012.

[20] Johan Bielecki, Filipe R. N. C. Maia, and Adrian P. Mancuso. Perspectives on single particle imaging with x rays at the advent of high repetition rate x-ray free electron laser sources. *Structural Dynamics*, 7(4):040901, 2020.

[21] Tomas Ekeberg, Dameli Assalauova, Johan Bielecki, Rebecca Boll, Benedikt J. Daurer, Lutz A. Eichacker, Linda E. Franken, Davide E. Galli, Luca Gelisio, Lars Gumprecht, Laura H. Gunn, Janos Hajdu, Robert Hartmann, Dirk Hasse, Alexandr Ignatenko, Jayanath Koliyadu, Olena Kulyk, Ruslan Kurta, Markus Kuster, Wolfgang Lugmayr, Jannik Lübke, Adrian P. Mancuso, Tommaso Mazza, Carl Nettelblad, Yevheniy Ovcharenko, Daniel E. Rivas, Amit K. Samanta, Philipp Schmidt, Egor Sobolev, Nicusor Timneanu, Sergej Usenko, Daniel Westphal, Tamme Wollweber, Lena Worbs, P. Lourdu Xavier, Hazem Yousef, Kartik Ayyer, Henry N. Chapman, Jonas A. Sellberg, Carolin Seuring, Ivan A. Vartanyants, Jochen Küpper, Michael Meyer, and Filipe R.N.C. Maia. Observation of a single protein by ultrafast x-ray diffraction. *bioRxiv*, 2022.

[22] Ruslan P. Kurta, Jeffrey J. Donatelli, Chun Hong Yoon, Peter Berntsen, Johan Bielecki, Benedikt J. Daurer, Hasan DeMirci, Petra Fromme, Max Felix

Hantke, Filipe R. N. C. Maia, Anna Munke, Carl Nettelblad, Kanupriya Pande, Hemanth K. N. Reddy, Jonas A. Sellberg, Raymond G. Sierra, Martin Svenda, Gijs van der Schot, Ivan A. Vartanyants, Garth J. Williams, P. Lourdu Xavier, Andrew Aquila, Peter H. Zwart, and Adrian P. Mancuso. Correlations in scattered x-ray laser pulses reveal nanoscale structural features of viruses. *Physical review letters*, 119:158102, 2017.

[23] Erik Malmerberg, Cheryl A. Kerfeld, and Petrus H. Zwart. Operational properties of fluctuation X-ray scattering data. *IUCrJ*, 2(3):309–316, 2015.

[24] Christopher Kupitz, Shibom Basu, Ingo Grotjohann, Raimund Fromme, Nadia A. Zatsepin, Kimberly N. Rendek, Mark S. Hunter, Robert L. Shoeman, Thomas A. White, Dingjie Wang, Daniel James, Jay-How Yang, Danielle E. Cobb, Brenda Reeder, Raymond G. Sierra, Haiguang Liu, Anton Barty, Andrew L. Aquila, Daniel Deponte, Richard A. Kirian, Sadia Bari, Jesse J. Bergkamp, Kenneth R. Beyerlein, Michael J. Bogan, Carl Caleman, Tzu-Chiao Chao, Chelsie E. Conrad, Katherine M. Davis, Holger Fleckenstein, Lorenzo Galli, Stefan P. Hau-Riege, Stephan Kassemeyer, Hartawan Laksmono, Mengning Liang, Lukas Lomb, Stefano Marchesini, Andrew V. Martin, Marc Messerschmidt, Despina Milathianaki, Karol Nass, Alexandra Ros, Shatabdi Roy-Chowdhury, Kevin Schmidt, Marvin Seibert, Jan Steinbrener, Francesco Stellato, Lifen Yan, Chunhong Yoon, Thomas A. Moore, Ana L. Moore, Yulia Pushkar, Garth J. Williams, Sebastien Boutet, R. Bruce Doak, Uwe Weierstall, Matthias Frank, Henry N. Chapman, John C. H. Spence, and Petra Fromme. Serial time-resolved crystallography of photosystem ii using a femtosecond x-ray laser. *Nature*, 513(7517):261–265, 2014.

[25] Chun Hong Yoon, Mikhail Yurkov, Evgeny Schneidmiller, L. Samoylova, Alexey Buzmakov, Zoltan Jurek, Beata Ziaja, Robin Santra, N. Loh, Th Tschentscher, and Adrian Mancuso. A comprehensive simulation framework for imaging single particles and biomolecules at the european x-ray free-electron laser. *Scientific Reports*, 6:24791, 04 2016.

[26] H. Olof Jönsson, Nicuşor Tîmneanu, Christofer Östlin, Howard A. Scott, and Carl Caleman. Simulations of radiation damage as a function of the temporal pulse profile in femtosecond X-ray protein crystallography. *Journal of Synchrotron Radiation*, 22(2):256–266, 2015.

[27] H. Olof Jönsson, Christofer Östlin, Howard A. Scott, Henry N. Chapman, Steve J. Aplin, Nicuşor Tîmneanu, and Carl Caleman. Freedam - a webtool for free-electron laser-induced damage in femtosecond x-ray crystallography. *High Energy Density Physics*, 26:93 – 98, 2018.

[28] Timur E. Gureyev, Alexander Kozlov, Andrew J. Morgan, Andrew V. Martin, and Harry M. Quiney. Effect of radiation damage and illumination variability on signal-to-noise ratio in X-ray free-electron laser single-particle imaging. *Acta Crystallographica Section A*, 76(6):664–676, 2020.

[29] I. Poudyal, M. Schmidt, and P. Schwander. Single-particle imaging by x-ray free-electron lasers - how many snapshots are needed? *Structural Dynamics*, 7(2):024102, 2020.

[30] Rebecca Boll, Julia M Schäfer, Benoît Richard, Kilian Fehre, Gregor Kastirke, Zoltan Jurek, Markus S Schöffler, Malik M Abdullah, Nils Anders, Thomas M Baumann, et al. X-ray multiphoton-induced coulomb explosion images

complex single molecules. *Nature Physics*, 18(4):423–428, 2022.

[31] S. P. Hau-Riege. *High-Intensity X-Rays-Interaction with Matter: Processes in Plasmas, Clusters, Molecules and Solids.* John Wiley Sons, 2012.

[32] Yuri Ralchenko, Hyuh-Kyung Chung, Howard Scott, Stephanie Hansen, G. Ferland, Robin Williams, Olivier Peyrusse, Vladimir Novikov, Christopher Fontes, James Colgan, Joseph Abdallah, and Jr. *Modern Methods in Collisional-Radiative Modeling of Plasmas*, volume 90. 03 2016.

[33] Carl Caleman, Francisco Junior, Oscar Grånäs, and Andrew Martin. A perspective on molecular structure and bonding-breaking in radiation damage in serial femtosecond crystallography. *Crystals*, 10:585, 07 2020.

[34] Howard A. Scott. Cretin—a radiative transfer capability for laboratory plasmas. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 71(2-6):689–701, October 2001.

[35] Joachim Oxenius. *Kinetic Theory of Particles and Photons: Theoretical Foundations of Non-LTE Plasma Spectroscopy*, volume 20. Springer, 2012.

[36] O. Jönsson. *Ultrafast Structural and Electron Dynamics in Soft Matter Exposed to Intense X-ray Pulses*. PhD thesis, Department of Physics and Astronomy, Uppsala University, 2017.

[37] H.A. Scott and R.W. Mayle. Glf - a simulation code for x-ray lasers. *Applied Physics B*, 58:35–43, 1994.

[38] John C Stewart and Kedar D Pyatt Jr. Lowering of ionization potentials in plasmas. *Astrophysical Journal, vol. 144, p. 1203*, 144:1203, 1966.

[39] Francis Chen. *Introduction to Plasma Physics and Controlled Fusion*. 2016.

[40] Magnus Bergh, Nicusor Timneanu, and David van der Spoel. Model for the dynamics of a water cluster in an x-ray free electron laser beam. *Phys. Rev. E*, 70:051904, Nov 2004.

[41] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An n log (n) method for ewald sums in large systems. *The Journal of chemical physics*, 98(12):10089–10092, 1993.

[42] Magnus Bergh, Nicusor Tîmneanu, and David van der Spoel. Model for the dynamics of a water cluster in an x-ray free electron laser beam. *Phys. Rev. E*, 70:051904, Nov 2004.

[43] Michael S Murillo and Jon C Weisheit. Dense plasmas, screened interactions, and atomic ionization. *Physics Reports*, 302(1):1–65, 1998.

[44] Francis F. Chen. *Introduction to Plasma Physics and Controlled Fusion*. Springer International Publishing, Cham, third edition.

[45] Karol Nass, Alexander Gorel, Malik Abdullah, Andrew Martin, Marco Kloos, Agostino Marinelli, Andrew Aquila, Thomas Barends, Franz-Josef Decker, R. Doak, Lutz Foucar, Elisabeth Hartmann, Mario Hilpert, Mark Hunter, Zoltan Jurek, Jason Koglin, Alexander Kozlov, Alberto Lutman, Gabriela Kovacs, and Ilme Schlichting. Structural dynamics in proteins induced by and probed with x-ray free-electron laser pulses. *Nature Communications*, 11:1814, 04 2020.

[46] H.J.C. Berendsen, D. van der Spoel, and R. van Drunen. Gromacs: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 91(1):43–56, 1995.

[47] David Van Der Spoel, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E.

Mark, and Herman J. C. Berendsen. Gromacs: Fast, flexible, and free. *Journal of computational chemistry*, 26(16):1701–1718, 2005.

[48] David van der Spoel, Erik Lindahl, Berk Hess, Aldert Buuren, Emile Apol, Pieter Meulenhoff, Peter Tielemann, Alfons Sijbers, K. Anton Feenstra, Rudi Drunen, and Herman Berendsen. *Gromacs User Manual Version 3.3*. 01 2004.

[49] Erik Lindahl, Berk Hess, and David van der Spoel. Gromacs 3.0: A package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling*, 7:306–317, 2001.

[50] C. Caleman. *Towards Single Molecule Imaging - Understanding Structural Transitions Using Ultrafast X-ray Sources and Computer Simulations*. PhD thesis, Department of Cell and Molecular Biology, Molecular biophysics, Uppsala University, 2007.

[51] Philip M. Morse. Diatomic molecules according to the wave mechanics. ii. vibrational levels. *Phys. Rev.*, 34:57–64, Jul 1929.

[52] Alex D MacKerell Jr, Donald Bashford, MLDR Bellott, Roland Leslie Dunbrack Jr, Jeffrey D Evanseck, Martin J Field, Stefan Fischer, Jiali Gao, H Guo, Sookhee Ha, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The journal of physical chemistry B*, 102(18):3586–3616, 1998.

[53] Yong Duan, Chun Wu, Shibasish Chowdhury, Mathew C. Lee, Guoming Xiong, Wei Zhang, Rong Yang, Piotr Cieplak, Ray Luo, Taisung Lee, James Caldwell, Junmei Wang, and Peter Kollman. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of Computational Chemistry*, 24(16):1999–2012, 2003.

[54] William L. Jorgensen and Julian Tirado-Rives. The opls [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, 110(6):1657–1666, 1988.

[55] Chris Oostenbrink, Alessandra Villa, Alan E. Mark, and Wilfred F. Van Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *Journal of computational chemistry*, 25:1656–1676, 2004.

[56] David van der Spoel, Paul J van Maaren, and Carl Caleman. Gromacs molecule & liquid database. *Bioinformatics*, 28(5):752–753, 2012.

[57] Alexander Kozlov, Andrew Martin, and Harry Quiney. Hybrid plasma/molecular-dynamics approach for efficient xfel radiation damage simulations. *Crystals*, 10:478, 06 2020.

[58] Zoltan Jurek, Sang-Kil Son, Beata Ziaja, and Robin Santra. *XMDYN* and *XATOM*: versatile simulation tools for quantitative modeling of X-ray free-electron laser induced dynamics of matter. *Journal of Applied Crystallography*, 49(3):1048–1056, Jun 2016.

[59] T Tachibana, Z Jurek, H Fukuzawa, K Motomura, K Nagaya, S Wada, Per Johnsson, M Siano, S Mondal, Y Ito, et al. Nanoplasma formation by high intensity hard x-rays. *Scientific reports*, 5(1):10977, 2015.

[60] Stefan P Hau-Riege, Richard A London, and Abraham Szoke. Dynamics of biological molecules irradiated by short x-ray pulses. *Physical Review E*,

69(5):051906, 2004.

[61] Ming Feng Gu. The flexible atomic code. In *AIP Conference Proceedings*, volume 730, pages 127–136. American Institute of Physics, 2004.

[62] Oscar Grånäs, Nicusor Timneanu, Ibrahim Eliah Dawod, Davide Ragazzon, Sebastian Trygg, Petros Souvatzis, Tomas Edvinsson, and Carl Caleman. Femtosecond bond breaking and charge dynamics in ultracharged amino acids. *The Journal of Chemical Physics*, 151(14):144307, 2019.

[63] Kirsten Schnorr, Arne Senftleben, Moritz Kurka, A Rudenko, Georg Schmid, Thomas Pfeifer, Kristina Meyer, Matthias Kübel, Matthias F Kling, YH Jiang, et al. Electron rearrangement dynamics in dissociating i 2 n+ molecules accessed by extreme ultraviolet pump-probe experiments. *Physical review letters*, 113(7):073001, 2014.

[64] Phay J Ho, Dipanwita Ray, C Stefan Lehmann, Adam EA Fouda, Robert W Dunford, Elliot P Kanter, Gilles Doumy, Linda Young, Donald A Walko, Xuechen Zheng, et al. X-ray induced electron and ion fragmentation dynamics in ibr. *The Journal of Chemical Physics*, 158(13), 2023.

[65] Eberhard Engel and Reiner M. Dreizler. *Density functional theory: an advanced course*. Springer, New York;Heidelberg [Germany];, 1. aufl. edition, 2011.

[66] Errol Lewars. *Computational chemistry: introduction to the theory and applications of molecular and quantum mechanics*. Springer, 2011.

[67] José M Soler, Emilio Artacho, Julian D Gale, Alberto García, Javier Junquera, Pablo Ordejón, and Daniel Sánchez-Portal. The SIESTA method for ab initio order-N materials simulation. *Journal of Physics: Condensed Matter*, 14(11):2745, 2002.

[68] D D Koelling and B N Harmon. A technique for relativistic spin-polarised calculations. *Journal of Physics C: Solid State Physics*, 10(16):3107–3114, 1977.

[69] John M. Wills, Mebarek Alouani, Per Andersson, Anna Delin, Olle Eriksson, and Oleksiy Grechnyev. *The Full-Potential Electronic Structure Problem and RSPt*, pages 47–73. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[70] Daniel A. Rehn, John M. Wills, Torey E. Battelle, and Ann E. Mattsson. Dirac's equation and its implications for density functional theory based calculations of materials containing heavy elements. *Phys. Rev. B*, 101:085114, Feb 2020.

[71] F. L. Hirshfeld. Bonded-atom fragments for describing molecular charge densities. *Theoretica Chimica Acta*, 44(2):129–138, 1977.

[72] Masakazu Matsumoto, Takuma Yagasaki, and Hideki Tanaka. Genice: Hydrogen-disordered ice generator. *Journal of Computational Chemistry*, 39:61–64, 2017.

[73] Masakazu Matsumoto, Takuma Yagasaki, and Hideki Tanaka. Novel algorithm to generate hydrogen-disordered ice structures. *Journal of Chemical Information and Modeling*, 61:2542–2546, 2021.

[74] Andrew V. Martin. Orientational order of liquids and glasses *via* fluctuation diffraction. *IUCrJ*, 4(1):24–36, Jan 2017.

[75] Patrick Adams, Jack Binns, Tamar L. Greaves, and Andrew V. Martin. The sensitivity of the pair-angle distribution function to protein structure. *Crystals*,

10(9), 2020.

[76] Thomas Northey, Nikola Zotev, and Adam Kirrander. Ab initio calculation of molecular diffraction. *Journal of Chemical Theory and Computation*, 10:4911–4920, 11 2014.

[77] Max Hantke, Tomas Ekeberg, and Filipe Maia. Condor: A simulation tool for flash x-ray imaging. *Journal of Applied Crystallography*, 49, 2016.

[78] C. Östlin. *Simulations of Biomolecular Fragmentation and Diffraction with Ultrafast X-ray Lasers*. PhD thesis, Department of Physics and Astronomy, Uppsala University, 2019.

[79] J. J. Sakurai and Jim Napolitano. *Modern quantum mechanics*. Addison-Wesley, Reading, Mass, 2. edition, 2011.

[80] Henry N. Chapman, Petra Fromme, Anton Barty, Thomas A. White, Richard A. Kirian, Andrew Aquila, Mark S. Hunter, Joachim Schulz, Daniel P. DePonte, Uwe Weierstall, R. Bruce Doak, Filipe R. N. C. Maia, Andrew V. Martin, Ilme Schlichting, Lukas Lomb, Nicola Coppola, Robert L. Shoeman, Sascha W. Epp, Robert Hartmann, Daniel Rolles, Artem Rudenko, Lutz Foucar, Nils Kimmel, Georg Weidenspointner, Peter Holl, Mengning Liang, Miriam Barthelmess, Carl Caleman, Sebastien Boutet, Michael J. Bogan, Jacek Krzywinski, Christoph Bostedt, Sasa Bajt, Lars Gumprecht, Benedikt Rudek, Benjamin Erk, Carlo Schmidt, Andre Hoemke, Christian Reich, Daniel Pietschner, Lothar Strueder, Guenter Hauser, Hubert Gorke, Joachim Ullrich, Sven Herrmann, Gerhard Schaller, Florian Schopper, Heike Soltau, Kai-Uwe Kuehnel, Marc Messerschmidt, John D. Bozek, Stefan P. Hau-Riege, Matthias Frank, Christina Y. Hampton, Raymond G. Sierra, Dmitri Starodub, Garth J. Williams, Janos Hajdu, Nicusor Timneanu, M. Marvin Seibert, Jakob Andreasson, Andrea Rocker, Olof Joensson, Martin Svenda, Stephan Stern, Karol Nass, Robert Andritschke, Claus-Dieter Schroeter, Faton Krasniqi, Mario Bott, Kevin E. Schmidt, Xiaoyu Wang, Ingo Grotjohann, James M. Holton, Thomas R. M. Barends, Richard Neutze, Stefano Marchesini, Raimund Fromme, Sebastian Schorb, Daniela Rupp, Marcus Adolph, Tais Gorkhover, Inger Andersson, Helmut Hirsemann, Guillaume Potdevin, Heinz Graafsma, Bjoern Nilsson, and John C. H. Spence. Femtosecond X-ray protein nanocrystallography. *Nature*, 470(7332):73–78, 2011.

[81] Jonas A Sellberg. *X-ray scattering and spectroscopy of supercooled water and ice*. PhD thesis, Department of Physics, Stockholm University, 2014.

[82] Jean-Pierre Hansen and I. R McDonald. *Theory of Simple Liquids*. Elsevier Science Technology, London, 2006.

[83] F. Wang, E. Weckert, B. Ziaja, D. S. D. Larsson, and D. van der Spoel. Coherent diffraction of a single virus particle: The impact of a water layer on the available orientational information. *Phys. Rev. E*, 83:031907, Mar 2011.

[84] Marcus Hanwell, Donald Curtis, David Lonie, Tim Vandermeersch, Eva Zurek, and Geoffrey Hutchison. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of cheminformatics*, 4:17, 2012.

[85] Michael Feig. Is alanine dipeptide a good model for representing the torsional preferences of protein backbones? *Journal of Chemical Theory and Computation*, 4(9):1555–1564, 2008.

[86] Vladimir Mironov, Yuri Alexeev, Vikram Khipple Mulligan, and Dmitri G.

Fedorov. A systematic study of minima in alanine dipeptide. *Journal of Computational Chemistry*, 40(2):297–309, 2019.

[87] Kenneth R. Beyerlein, H. Olof Jönsson, Roberto Alonso-Mori, Andrew Aquila, Saša Bajt, Anton Barty, Richard Bean, Jason E. Koglin, Marc Messerschmidt, Davide Ragazzon, Dimosthenis Sokaras, Garth J. Williams, Stefan Hau-Riege, Sébastien Boutet, Henry N. Chapman, Nicuşor Tîmneanu, and Carl Caleman. Ultrafast nonthermal heating of water initiated by an x-ray free-electron laser. *Proceedings of the National Academy of Sciences*, 115(22):5652–5657, 2018.

[88] Anton Barty, Carl Caleman, Andrew Aquila, Nicusor Timneanu, Lukas Lomb, Thomas A. White, Jakob Andreasson, David Arnlund, Saša Bajt, Thomas R. M. Barends, Miriam Barthelmess, Michael J. Bogan, Christoph Bostedt, John D. Bozek, Ryan Coffee, Nicola Coppola, Jan Davidsson, Daniel P. DePonte, R. Bruce Doak, Tomas Ekeberg, Veit Elser, Sascha W. Epp, Benjamin Erk, Holger Fleckenstein, Lutz Foucar, Petra Fromme, Heinz Graafsma, Lars Gumprecht, Janos Hajdu, Christina Y. Hampton, Robert Hartmann, Andreas Hartmann, Günter Hauser, Helmut Hirsemann, Peter Holl, Mark S. Hunter, Linda Johansson, Stephan Kassemeyer, Nils Kimmel, Richard A. Kirian, Mengning Liang, Filipe R. N. C. Maia, Erik Malmerberg, Stefano Marchesini, Andrew V. Martin, Karol Nass, Richard Neutze, Christian Reich, Daniel Rolles, Benedikt Rudek, Artem Rudenko, Howard Scott, Ilme Schlichting, Joachim Schulz, M. Marvin Seibert, Robert L. Shoeman, Raymond G. Sierra, Heike Soltau, John C. H. Spence, Francesco Stellato, Stephan Stern, Lothar Strüder, Joachim Ullrich, X. Wang, Georg Weidenspointner, Uwe Weierstall, Cornelia B. Wunderer, and Henry N. Chapman. Self-terminating diffraction gates femtosecond x-ray nanocrystallography measurements. *Nat. Photonics*, 6:35–40, 2012.

[89] Carl Caleman, Nicuşor Tîmneanu, Andrew V. Martin, H. Olof Jönsson, Andrew Aquila, Anton Barty, Howard A. Scott, Thomas A. White, and Henry N. Chapman. Ultrafast self-gating bragg diffraction of exploding nanocrystals in an x-ray laser. *Opt. Express*, 23(2):1213–1231, Jan 2015.

[90] Jack Binns, Connie Darmanin, Cameron M. Kewish, Sachini Kadaoluwa Pathirannahalge, Peter Berntsen, Patrick L. R. Adams, Stefan Paporakis, Daniel Wells, Francisco Gian Roque, Brian Abbey, Gary Bryant, Charlotte E. Conn, Stephen T. Mudie, Adrian M. Hawley, Timothy M. Ryan, Tamar L. Greaves, and Andrew V. Martin. Preferred orientation and its effects on intensity-correlation measurements. *IUCrJ*, 9(2):231–242, Mar 2022.

[91] Andrew V Martin, Alexander Kozlov, Peter Berntsen, Francisco Gian Roque, Leonie Flueckiger, Saumitra Saha, Tamar L Greaves, Charlotte E Conn, Adrian M Hawley, Timothy M Ryan, et al. Fluctuation x-ray diffraction reveals three-dimensional nanostructure and disorder in self-assembled lipid phases. *Communications Materials*, 1(1):40, 2020.

[92] Felix Lehmkühler, Francesco Dallari, Avni Jain, Marcin Sikorski, Johannes Möller, Lara Frenzel, Irina Lokteva, Grant Mills, Michael Walther, Harald Sinn, et al. Emergence of anomalous dynamics in soft matter probed at the european xfel. *Proceedings of the National Academy of Sciences*, 117(39):24110–24116, 2020.

[93] Fivos Perakis, Gaia Camisasca, Thomas J Lane, Alexander Späh, Kjartan Thor

Wikfeldt, Jonas A Sellberg, Felix Lehmkühler, Harshad Pathak, Kyung Hwan Kim, Katrin Amann-Winkel, et al. Coherent x-rays reveal the influence of cage effects on ultrafast water dynamics. *Nature communications*, 9(1):1917, 2018.

[94] Paul Emma, R Akre, J Arthur, R Bionta, C Bostedt, J Bozek, A Brachmann, P Bucksbaum, Ryan Coffee, F-J Decker, et al. First lasing and operation of an ångstrom-wavelength free-electron laser. *nature photonics*, 4(9):641–647, 2010.

[95] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. Ucsf chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, 2004.

[96] Robert B. Best, Xiao Zhu, Jihyun Shim, Pedro E. M. Lopes, Jeetain Mittal, Michael Feig, and Alexander D. MacKerell. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ_1 and χ_2 Dihedral Angles. *Journal of Chemical Theory and Computation*, 8(9):3257–3273, September 2012. Publisher: American Chemical Society.

[97] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of chemical physics*, 126(1), 2007.

[98] Michele Parrinello and Aneesur Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics*, 52(12):7182–7190, 1981.

[99] Christofer Östlin, Nicusor Tîmneanu, H. Olof Jönsson, Tomas Ekeberg, Andrew V. Martin, and Carl Caleman. Reproducibility of single protein explosions induced by x-ray lasers. *Phys. Chem. Chem. Phys.*, 20:12381–12389, 2018.

[100] Berk Hess, Carsten Kutzner, David van der Spoel, and Erik Lindahl. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theor. Comp.*, 4:435–447, 2008.

[101] J. E, M. Stransky, Z. Jurek, C. Fortmann-Grote, L. Juha, R. Santra, B. Ziaja, and A. P. Mancuso. Effects of radiation damage and inelastic scattering on single-particle imaging of hydrated proteins with an x-ray free-electron laser. *Scientific report*, 2021.

[102] C. Östlin, N. Timneanu, C. Caleman, and A. V. Martin. Is radiation damage the limiting factor in high-resolution single particle imaging with x-ray free-electron lasers? *Structural Dynamics*, 6(4):044103, 2019.

[103] A. Aquila, A. Barty, C. Bostedt, S. Boutet, G. Carini, D. dePonte, P. Drell, S. Doniach, K. H. Downing, T. Earnest, H. Elmlund, V. Elser, M. Gühr, J. Hajdu, J. Hastings, S. P. Hau-Riege, Z. Huang, E. E. Lattman, F. R. N. C. Maia, S. Marchesini, A. Ourmazd, C. Pellegrini, R. Santra, I. Schlichting, C. Schroer, J. C. H. Spence, I. A. Vartanyants, S. Wakatsuki, W. I. Weis, and G. J. Williams. The linac coherent light source single particle imaging road map. *Structural Dynamics*, 2(4):041701, 2015.

[104] Saša Bajt, Mauro Prasciolu, Holger Fleckenstein, Martin Domaracký, Henry N Chapman, Andrew J Morgan, Oleksandr Yefanov, Marc Messerschmidt, Yang Du, Kevin T Murray, et al. X-ray focusing with efficient high-na multilayer laue lenses. *Light: Science & Applications*, 7(3):17162–17162, 2018.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 2353

Editor: The Dean of the Faculty of Science and Technology